HAL
open science

Étude de la relation entre l'annotation des formes et des fonctions prosodiques en anglais britannique contemporain

Saandia Ali

To cite this version:

Saandia Ali. Étude de la relation entre l'annotation des formes et des fonctions prosodiques en anglais britannique contemporain. Linguistique. Université de Provence - Aix-Marseille I, 2010. Français. NNT : . tel-00460431

HAL Id: tel-00460431

https://theses.hal.science/tel-00460431

Submitted on 1 Mar 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.

UNIVERSITÉ AIX-MARSEILLE I – Université de Provence
U.F.R. LACS
Formation doctorale : Langage et Parole
Discipline : phonétique anglaise

Nº attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

THÈSE
Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ AIX-MARSEILLE I
Présentée et soutenue publiquement
Le 22 janvier 2010
Par

Saandia ALI

Étude de la relation entre l'annotation des formes et des fonctions prosodiques en anglais britannique contemporain

Sous la direction de
Monsieur Daniel HIRST
Directeur de recherche CNRS, UMR6057
Laboratoire Parole et Langage - Université de Provence

JURY

Elisabeth Delais-Roussarie, Directeur de recherche CNRS à l'Université Paris 7.
Jill House, Teaching Fellow, University College London.
Daniel Hirst, Directeur de recherche CNRS à l'Université de Provence.
Sophie Herment, Maître de conférences à l'Université de Provence.
Gerry Knowles, Senior Lecturer, University of Lancaster and University of Malaya.

C’est le ton qui fait la chanson

Ose ton chant

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de thèse Daniel Hirst qui m'a fait confiance pour travailler sur un sujet qui lui est cher. Je le remercie chaleureusement pour son aide tout au long de ces années de thèse, pour ses nombreux conseils et encouragements et pour son attitude toujours rassurante dans les moments d'angoisses et de doutes.

Je remercie les chercheurs qui m'ont fait l'honneur d'accepter de faire partie de mon jury de thèse. Je remercie tout particulièrement Jill House et Elizabeth Delais-Roussarie qui ont aimablement accepté de rapporter ma thèse.

Je remercie mes enseignants du *Département d'étude du monde anglophone,* qui ont assuré ma formation initiale en tant que linguiste. Je pense tout particulièrement à Michel Ginésy et à Sophie Herment qui m'ont donné en premier lieu le goût de la phonétique et de l'intonation de l'anglais.

Je remercie également les membres du *Laboratoire Parole et Langage* qui m'ont permis de mener à bien mes recherches. Cette thèse a aussi pris forme au fil de rencontres et de discussions qui ont été très enrichissantes. Je remercie notamment Noël Nguyen et Maria Paola D'Imperio qui ont gentiment accepté de me recevoir et de discuter de mon travail.

Merci à Stéphane Rosy pour toutes ses explications et pour le temps qu'il m'a consacré lors de mes premiers essais expérimentaux.

Un grand merci à Isabelle Marlien pour ses relectures inestimables ainsi que pour sa gentillesse, ses conseils et encouragements tout au long de ces années de thèse.

Je tiens aussi à remercier Yi Xu et Jill House, qui ont tous deux très gentiment accepté de me recevoir à Londres pour discuter de mon travail.

Je tiens enfin à remercier mes proches et mes amis qui ont tous énormément compté pour moi tout au long de cette aventure.

Un immense merci à ma mère qui m'a été d'une aide inestimable dans la réalisation de ma thèse. Je la remercie d'avoir mis ses talents d'informaticienne au service de la linguistique et d'avoir rendu possible le travail expérimental de cette thèse. Je la remercie pour toutes les heures passées à travailler à mes côtés sur tous les aspects informatiques de ma thèse. Je la remercie aussi tout simplement pour sa présence et pour avoir toujours cru en moi.

Merci à mon père pour son soutien, sa gentillesse, ses paroles de sagesse qui sont toujours salutaires dans les moments difficiles.

Merci à Dan pour sa présence, sa patience, son écoute et sa grande générosité.

TABLE DES MATIÈRES

REMERCIEMENTS *5*

TABLE DES MATIÈRES *7*

TABLE DES ILLUSTRATIONS *13*

INTRODUCTION *23*

Motivation et présentation du sujet de l'étude **25**

Objectifs et méthodologie **28**

Plan de la thèse **29**

PARTIE I *31*

CONTEXTE SCIENTIFIQUE **31**

Chapitre 1 ..... **33**

Fonctions généralement attribuées à la prosodie **33**

1.1 La fonction lexicale **36**

1.2 La fonction grammaticale **37**

1.3 Prosodie et discours : la fonction discursive **41**

1.4 Prosodie et information : la fonction de focalisation **41**

1.5 Prosodie, attitude et émotion : la fonction expressive **43**

1.6 Fonction identificatrice **44**

1.7 Conclusion **45**

Chapitre 2 ..... **47**

Représentation des formes prosodiques de l'anglais : les différents modèles **47**

2.1 Les modèles britanniques des formes prosodiques **49**

2.1.1 The « tune approach » 50

2.1.2 The « English nuclear tones approach » 53

2.1.3 « Prosodic features » 62

2.1.4 Problèmes non résolus dans les approches britanniques de l'intonation 66

2.2 Le modèle autosegmental **69**

2.2.1 Origines et fondements de la théorie autosegmentale 69

2.2.2 Les modèles standards américains 70

2.3 La question de l'alignement tonal **77**

2.3.1 Prémisses à l'analyse de l'alignement tonal chez les premiers théoriciens de l'intonation 78

2.3.2 Facteurs influençant l'alignement des tons mélodiques 79

2.3.3 Le domaine d'association des tons 80

2.3.4 Résultats des différents travaux de recherche 82

2.4 Conclusion **89**

Chapitre 3 ..... **91**

Principaux systèmes d'annotation prosodique de l'anglais : illustration de la relation entre formes et fonctions prosodiques **91**

3.1 Les systèmes de transcription des formes prosodiques **93**

3.1.1 Les portées musicales 93

3.1.2 Les niveaux 94

3.1.3 Le « crazy type » et les tracés curvilignes 95

3.2 Les systèmes représentant formes et fonctions prosodiques **96**

3.2.1 Les têtards 96

3.2.2 Un système de transcription intégrée 97

3.2.3 Tonetic Stress Marks : TSM 98

3.2.4 TOBI 100

3.2.5 Etude comparative des notations TSM et TOBI 101

3.2.6 Les limites et critiques apportées à TOBI 102

3.3 Vers une annotation distincte des formes et fonctions prosodiques **104**

3.3.1 Fondations théoriques du modèle quadratique 104

3.3.2 Annotation automatique des formes prosodiques : INTSINT et MOMEL 105

3.3.3 Annotation des fonctions prosodiques : StarBI et IF 109

3.4 Conclusion **112**

Chapitre 4 ..... **115**

Différentes approches de l'articulation forme/fonction **115**

4.1 Les approches théoriques : « Top down approaches » **117**

4.1.1 L'approche syntaxique ou grammaticale 117

4.1.2 Approches discursives et pragmatiques 120

4.1.3 L'approche compositionnelle 123

4.2 Les approches expérimentales: « Bottom up approaches » **128**

4.2.1 L'approche perceptive 128

4.2.2 Les approches physiologiques et articulatoires 134

4.2.3 L'approche probabiliste 142

4.2.4 L'analyse par synthèse 144

4.3 Conclusion **154**

CONCLUSION GENERALE **157**

PARTIE II ***163***

TRAVAIL EXPERIMENTAL **163**

Chapitre 5 ..... **167**

Choix des corpus et de l'annotation **167**

5.1 Le corpus EUROM1 ..... **169**

5.2 Le corpus Aix-MARSEC **171**

5.3 Développements de l'annotation du corpus EUROM1 ..... **172**

Chapitre 6 ..... **175**

Analyse par synthèse des contours intonatifs de l'anglais britannique **175**
6.1 Méthode d'analyse **177**
6.2 Modèles d'analyse par synthèse de l'intonation de l'anglais **178**
6.2.1 Le modèle « nul » 178
6.2.2 Le modèle « IU » 179
6.2.3 Le modèle « accent» 180
6.2.4 Le modèle « terminal » 181
6.2.5 Le modèle « emphatique » 182
6.3 Evaluation des résultats **184**

Chapitre 7 ..... **187**

Phases d'optimisation des modèles de synthèse **187**
7.1 Premières phases d'optimisation : modèle des pauses **189**
7.1.1 Optimisation de la séquence de tons INTSINT 189
7.1.2 Optimisation de l'alignement des points cibles 193
7.2 Optimisation des représentations formelles au niveau de l'unité intonative **193**
7.3 Optimisation des représentations formelles au niveau des unités tonales **195**
7.3.1 Optimisation de l'évaluation objective des modèles 195
7.3.2 Optimisation de la séquence de tons en fonction de l'unité tonale 202
7.3.3 Optimisation de l'alignement des points cibles 205
7.4 Optimisation de la représentation des fonctions **207**
7.4.1 Détections des frontières d'IU dans les modèles des pauses et des IU 207
7.4.2 Détection des types de TU 213
7.4.3 Détection des frontières d'IU dans le modèle des TU 216
7.4.4 Résultats des détections des fonctions 219
7.5 Conclusions **225**
7.5.1 Evaluation des modèles 225
7.5.2 Modèles des formes prosodiques 228
7.5.3 Détection automatique et réinterprétation des fonctions prosodiques 229

Chapitre 8 ..... **231**

Optimisation de la relation forme/fonction à partir du corpus Aix-MARSEC **231**
8.1 Introduction **233**
8.2 Hypothèses et questions de départ **233**
8.2.1 Les questions liées aux fonctions prosodiques 233
8.2.2 Les questions liées aux représentations formelles 235
8.3 Annotations fonctionnelles **236**
8.3.1 Conversion des TSM en IF : vers un IF « British School » 236
8.3.2 Unité nucléaire et accents mélodiques : un IF « britannique adapté » 246
8.4 Annotation des formes **249**
8.4.1 Optimisation du codage INTSINT 249

8.5 Description des modèles à tester 254
8.6 Résultats de l'optimisation des représentations formelles modèle par modèle 264
8.6.1 Le modèle des pauses 264
8.6.2 Le modèle des IU 265
8.6.3 Les modèles basés sur les unités tonales 265
8.6.4 Evaluation des modèles d'optimisation des séquences INTSINT 271
8.7 Analyse et évaluation des résultats de l'optimisation de l'alignement 274
8.7.1 Alignement de P1 dans les modèles basés sur l'unité tonale 274
8.7.2 Effet de la position du TU sur l'alignement de P1 277
8.7.3 Effet de la position du nucléus sur l'alignement de P1 282
8.7.4 Alignement de P1 en fonction du contour 285
8.7.5 Alignement de P2 dans les modèles basés sur l'unité tonale 290
8.7.6 Alignement de P2 en position finale en fonction du contour 297
8.7.7 Alignement des points cibles dans les modèles basés sur l'unité tonale élargie (TTU) 301
8.7.8 Effet de la position du TTU sur l'alignement de P1 dans les modèles basés sur l'unité tonale élargie 303
8.7.9 Effet de la position du nucléus sur l'alignement de P1 dans les modèles basés sur l'unité tonale élargie 306
8.7.10 Alignement de P2 dans un TTU 308
8.7.11 Alignement de P3 dans un TTU 310
8.7.12 Conclusion provisoire 311
8.7.13 Evaluation objective des modèles d'optimisation de l'alignement 312
8.8 Evaluation des modèles finaux 314
8.9 Grammaires formes/fonctions 319
8.9.1 Grammaires à partir du modèle 11 319
8.10 Optimisation des représentations fonctionnelles dans les différents modèles 329
8.10.1 Détection des types d'IU 329
8.10.2 Détection des types de TU 331
8.10.3 Détection jointe des types de TU-IU 345
8.10.4 Détection des frontières de TTU 347
8.10.5 Détection des frontières d'unités intonatives (IU) 349
8.10.6 Conclusion provisoire sur les détections de fonctions 354
8.11 Retour sur les données et évaluation des modèles par analyse par synthèse 355
8.11.1 Hypothèses expliquant les échecs de modélisation 355
8.11.2 Relevé de problèmes de modélisation à partir des modèles 11 et 24 356
8.11.3 Exemple de détections de fonctions problématiques 363

***Chapitre 9* 367**

Conclusion et synthèse des résultats 367
9.1 Annotations fonctionnelles 371

9.2 Codage des formes et alignement tonal 373
9.3 Tentatives de reconnaissance automatique des fonctions prosodiques 375
9.4 Perspectives et limites 376

RÉFÉRENCES BIBLIOGRAPHIQUES..... 379

ANNEXES..... 393

Annexe 1. Grammaires de l'articulation forme/fonction dans les différents modèles..... 395
1.1. Grammaires du modèle 24 (TTU et INTSINT) 397
1.2. Grammaires du modèle 62 (TU et INTSINT3)..... 398
1.3. Grammaires du modèle 42 (TU et INTSINT2)..... 400
1.4. Grammaires du modèle 52 (TU et INTSINT1)..... 402
1.5. Grammaires du modèle 46 (TTU et INTSINT2) 404
1.6. Grammaires du modèle 66 (TTU et INTSINT3) 406

Annexe 2. Exemples de fichiers résultats de l'optimisation des tons 409
2.1. Résultats optimisation des tons pour modèle 11, IU non terminal, TU S et A 411
2.2. Résultats optimisation des tons pour modèle 11, IU non terminal, TU N et !NM 412
2.3. Résultats optimisation des tons pour modèle 11, IU terminal, TU S et A 413
2.4. Résultats optimisation des tons pour modèle 11, IU terminal, TU N et !NM. 414
2.5. Résultats optimisation des tons pour modèle 11, IU emphatique non terminal, TU S et A..... 415
2.6. Résultats optimisation des tons pour modèle 11, IU emphatique non terminal, TU N et !N..... 416
2.7. Résultats optimisation des tons pour modèle 11, IU emphatique terminal, TU S et A 417
2.8. Résultats optimisation des tons pour modèle 11, IU emphatique terminal, TU N et !N..... 418

Annexe 3. Description de l'application 419
3.1. Le menu 421
3.2. Le paramétrage..... 422
3.3. Calcul de l'étendue, de la clé et des valeurs INTSINT 434
3.4. Le programme d'optimisation des formes et de l'alignement et de détection des types de TU et/ou IU et des frontières..... 438
3.5. Outil pour la visualisation et l'écoute des résultats de la modélisation 446

Annexe 4. Schéma récapitulatif de l'application 449
4.1. Calcul des valeurs d'INTSINT et des paramètres « key » et « range » 451
4.2. Optimisation des formes / détection des frontières et des fonctions..... 452
4.3. Préparation des données pour les statistiques 456

ANNEXES CD-ROM 457

Annexe 5. Annexes sons 457

5.1. Extraits du corpus EUROM1 : locuteurs fa, fe, ff, fg..... 457
5.2. Extraits du corpus MARSEC : locuteurs G..... 457
5.3. Exemples de résultats de la resynthèse : premiers essais avec les « modèles de synthèse » 457
5.4. Exemples de résultats de la resynthèse : modèles EUROM1..... 457
5.5. Exemples de résultats de la resynthèse : modèles MARSEC..... 457

Annexe 6. Scripts*..... *457
6.1. Scripts de calcul de l'étendue, de la clé et des valeurs d'INTSINT 457
6.2. Scripts divers d'optimisations des tons, de l'alignement et détection des fonctions 457
6.3. Scripts divers de préparation des données pour statistiques..... 457

RÉSUMÉS*..... *459
Summary 461
Résumé..... 463

TABLE DES ILLUSTRATIONS

Table des figures

Figure 1 : Illustration des variations distinctives de la quantité dans les voyelles et consonnes en finnois. Extrait de Hirst (2006, p.168). 36
Figure 2 : Les deux contours intonatifs de base 1 et 2 suivant Jones (1972). 50
Figure 3 : Exemple de contour 1 emphatique extrait de Jones (1972, p. 298-299)..... 51
Figure 4 : Contour 1 avec une emphase d'intensification, extrait de Jones (1972, p. 309)..... 51
Figure 5 : Exemple de contour 2 emphatique complexe. 52
Figure 6 : Exemple de contour 1 avec une parenthèse basse. 52
Figure 7 : Exemple de contour 2 avec une parenthèse finale, extrait de Jones (1972, p. 317). 52
Figure 8 : Chute-montée divisée selon Kingdon (1958). 58
Figure 9 : Exemple de chute-montée interrompue par une syllabe inaccentuée selon Kingdon 59
Figure 10 : Exemple de chute-montée interrompue par un accent partiel selon Kingdon. 59
Figure 11 : Système de diacritiques utilisé pour décrire les contours intonatifs selon Crystal(1969, p.213). 63
Figure 12 : Exemple de structure récursive dans le modèle de Hirst (1987) 68
Figure 13 : Différents types d'accents mélodiques dans le modèle de Pierrehumbert (respectivement : un pic, un sommet et un mouvement mélodique). Extrait de Pierrehumbert (1980)..... 74
Figure 14 : Grammaire des contours intonatifs donnée dans Pierrehumbert (1980, p. 29)..... 75
Figure 15 : Notation de Joshua Steele (1775) (extrait de Knowles, G & al. 1996)..... 93
Figure 16 : Transcription de Trager et Smith (1951). 94
Figure 17 : Transcription de Liberman et Sag (1974) pour le contour intonatif de la contradiction. 95
Figure 18 : Transcription de Svartvik et Quirk (1980)..... 96
Figure 19 : Transcription en TSM d'un passage d'une minute extrait du corpus MARSEC.... 99
Figure 20 : TOBI pitch accents 100
Figure 21 : Exemple de courbe de F0 et courbe spline quadratique obtenue avec MOMEL.. 106
Figure 22 : Le système INTSINT..... 108
Figure 23 : Configurations correspondant aux symboles relatifs..... 108
Figure 24 : Signal et F0 pour « If you can't lift it ask Peter to help you » (extrait de Hirst 2005, p. 339)..... 111
Figure 25 : Représentation fonctionnelle IF sous forme d'un arbre phonologique (extrait de Hirst 2005, p. 340). 111
Figure 26 : Contour stylisé selon la méthode IPO. 129
Figure 27 : Transcription simple et riche du prosogramme (extrait de Mertens 2004)..... 133

Figure 28 : Illustration du processus de génération des contours de F0 dans le modèle Fujisaki (extrait de Fujisaki 2002). 136
Figure 29 : Application du modèle Fujisaki au japonais (Fujisaki 2006). 137
Figure 30 : Génération des contours de F0 dans le modèle PENTA (extrait de Xu 2005). 141
Figure 31 : Courbe de F0 brute (ligne continue) et courbe modélisée (en pointillés) pour le passage « I have a problem with my water softener » du corpus EUROM1 153
Figure 32 : Codage INTSINT du passage illustré dans la figure 3.1..... 154
Figure 33 : TextGrid G0101B avec l'étiquetage initial du corpus. 172
Figure 34 : Passage FAo30072 avec l'annotation fonctionnelle utilisée dans l'analyse. 174
Figure 35 : Un extrait du TextGrid correspondant au premier IU du passage ci-dessus. 174
Figure 36 : La courbe de F0 générée par MOMEL superposée au résultat du modèle « nul ». 179
Figure 37 : La courbe de F0 générée par MOMEL superposée au résultat du modèle IU. 180
Figure 38 : La courbe de F0 générée par MOMEL superposée au résultat du modèle accent. 182
Figure 39 : La courbe issue du modèle emphatique superposée à la courbe issue de MOMEL. 184
Figure 40 : Courbe issue du processus d'optimisation (en bleu) comparée à MOMEL d'origine (en vert) et TextGrid représentant les niveaux d'annotation et d'analyse pour le fichier FAO40073. 192
Figure 41 : Courbes optimisées à partir de la corrélation seule ou de l'écart quadratique moyen pour le fichier G0110G, comparées à MOMEL. 197
Figure 42 : Courbes optimisées à partir de la corrélation seule (en bleu) ou de l'écart quadratique moyen (en marron) comparées à MOMEL (en vert) pour le fichier G0105G..... 198
Figure 43 : Courbes optimisées à partir de la corrélation seule (en bleu), de l'écart quadratique moyen (en marron) pour le fichier G0105B comparées à MOMEL (en vert). 198
Figure 44 : Comparaison des valeurs de corrélation et de l'écart quadratique moyen pour l'ensemble des séquences de tons testées au niveau d'un TU sur G0101B. 200
Figure 45 : Comparaison des valeurs de corrélation et de l'écart quadratique moyen pour l'ensemble des séquences de tons testées au niveau d'un TU sur G0101B toutes corrélations. 200
Figure 46 : Courbes optimisée à partir de la corrélation seule (en bleu), de l'écart quadratique moyen (en marron) et du « score » (en rose) pour le fichier G0105B comparées à MOMEL (en vert)..... 202
Figure 47 : Courbe modélisée au niveau de la pause (en bleu) comparée à la courbe de MOMEL originale (en vert) pour un passage du fichier FAo30072. 208
Figure 48 : Courbe MOMEL modélisée après la détection des frontières (en bleu) comparée à la courbe de MOMEL d'origine. 208
Figure 49 : Courbe de MOMEL originale (en vert) comparée à la courbe modélisée à partir de la détection des frontières dans le modèle BD1 (en bleu) et TextGrid correspondant. 211
Figure 50 : Courbe modélisée après détection des frontières dans le modèle 3 (en bleu) comparée à MOMEL (en vert)..... 212

Figure 51 : Courbe modélisée après détection des frontières dans le modèle 5 (en bleu) comparée à MOMEL (en vert) pour le fichier FFO00792. 220
Figure 52 : Courbe modélisée après détection des frontières dans le modèle 5 (en bleu) comparée à MOMEL (en vert) pour le fichier FFO60788. 221
Figure 53 : Courbe modélisée après détection des frontières dans le modèle 7 (en bleu) comparée à MOMEL (en vert) pour le fichier FEo2075. 222
Figure 54 : Comparaison entre deux courbes modélisées après détections des fonctions dans les modèles 7 (en bleu) et 8 (en rose)..... 224
Figure 55 : Courbes issues de la détection des fonctions dans les modèles 5 et 6. 225
Figure 56 : TextGrid G0301B montrant les différentes rangées d'annotation 239
Figure 57 : Courbe de MOMEL et annotation fonctionnelle pour un passage du fichier G0102B..... 242
Figure 58 : Exemple de conversion automatique des TSM en IF pour un passage extrait du fichier G0201B. 244
Figure 59 : Exemple de conversion automatique des TSM en IF pour un passage extrait du fichier G0504G. 245
Figure 60 : TextGrid extrait du ficher G0401g montrant les différents modèles fonctionnels et la courbe de MOMEL d'origine (en vert) comparée à la courbe issue du premier modèle fonctionnel (en bleu). 248
Figure 61 : Histogramme de la valeur des tons H, U, h et en fonction de l'étendue pour INTSINT1 et 2..... 251
Figure 62 : Histogramme de la valeur des tons H et U en fonction de l'étendue pour INTSINT1 et 3. 252
Figure 63 : Représentation des contours montants en fonction du type de codage INTSINT et de la position dans l'étendue. 252
Figure 64 : Écart quadratique moyen entre les courbes d'INTSINT, INTSINT2 et INTSINT 3 et la courbe de MOMEL d'origine. 253
Figure 65 : Histogramme de la répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la première syllabe ou de la durée du TU. 275
Figure 66 : Histogramme des meilleurs points d'alignement dans les modèles en valeur..... 277
Figure 67 : Variation de l'alignement de P1 en % de la durée de la première syllabe et en fonction de la position du TU. 278
Figure 68 : Variation de l'alignement de P1 en % de la durée du TU en fonction de la position du TU..... 279
Figure 69 : Variation de l'alignement de P1 par rapport au début du TU et en fonction de la position du TU..... 280
Figure 70 : Variation de l'alignement de P1 par rapport au début du de la voyelle et en fonction de la position du TU. 281
Figure 71 : Variation de l'alignement de P1 en % de la durée de la syllabe et en fonction de la position du nucléus. 283
Figure 72 : Variation de l'alignement de P1 par rapport au début du TU et en fonction de la position du nucléus. 284

Figure 73 : Variation de l'alignement de P1 par rapport au début de la voyelle et en fonction de la position du nucléus. 285
Figure 74 : Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour HL et en fonction de durée de la syllabe accentuée. 286
Figure 75 : .Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour LH et en fonction de la durée de la syllabe accentuée. 287
Figure 76 : Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour HL et en fonction de la position du TU. 288
Figure 77 : .Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour LH et en fonction de la position du TU. 288
Figure 78 : Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour HL et en fonction de la position du TU. 289
Figure 79 : .Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour LH et en fonction de la position du TU. 290
Figure 80 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1 ou M. 291
Figure 81 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position F avec point d'ancrage la cible précédente ou un pourcentage de la durée du TU. . 293
Figure 82 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position F avec point d'ancrage la dernière syllabe ou la dernière voyelle. 294
Figure 83 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la cible précédente ou un pourcentage de la durée du TU. . 295
Figure 84 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la cible précédente ou un pourcentage de la durée du TU. . 296
Figure 85 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage la syllabe accentuée. 297
Figure 86 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage la voyelle accentuée. 298
Figure 87 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage le point cible précédent. 298
Figure 88 : Histogramme des meilleurs points d'alignement de P2 en position F en fonction du contour avec comme point d'ancrage la syllabe accentuée. 299
Figure 89 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage la voyelle accentuée. 300
Figure 90 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour et du point cible précédent. 300
Figure 91 : Répartition des meilleurs points d'alignement pour P1 dans les TTU2 avec point d'ancrage la première syllabe, première voyelle ou début du TTU. 302
Figure 92 : Variation de l'alignement de P1 en fonction de la durée de la première syllabe et de la position du TTU. 303
Figure 93 : Variation de l'alignement de P1 en fonction du début du TTU et en fonction de la position du TTU. 304

Figure 94 : Variation de l'alignement de P1 par rapport au début de la voyelle et en fonction de la position du TTU. 305
Figure 95 : Variation de l'alignement de P1 en fonction de la durée de la syllabe et de la position du nucléus. 306
Figure 96 : Variation de l'alignement de P1 en fonction de la première voyelle et de la position du nucléus. 307
Figure 97 : Répartition des meilleurs points d'alignement pour P2 en fonction de la cible tonale précédente. 308
Figure 98 : Variation de l'alignement de P2 en fonction de l'attaque et de la durée de la dernière syllabe du TTU. 309
Figure 99 : Variation de l'alignement de P3 en fonction de la fin du TTU et de la fin de la dernière syllabe du TTU. 310
Figure 100 : Courbe générée à partir des modèles 11 et 24 pour un passage de MARSEC. 317
Figure 101 : Modélisation d'un contour de chute+montée à partir des TT1 (modèle 16) et TTU2 (modèle 24). 318
Figure 102 : Grammaire des IU de type non terminal pour le modèle 11. 320
Figure 103 : Grammaire des IU de type terminal pour le modèle 11. 320
Figure 104 : Grammaire des IU de type emphatique non terminal pour le modèle 11. 323
Figure 105 : Grammaire des IU de type emphatique terminal pour le modèle 11. 323
Figure 106 : Grammaire des IU de type interrogatif non terminal pour le modèle 11. 325
Figure 107 : Grammaire des IU de type interrogatif terminal pour le modèle 11. 325
Figure 108 : Grammaire des IU de type non terminal pour le modèle 24. 327
Figure 109 : Grammaire des IU de type terminal pour le modèle 24. 327
Figure 110 : Courbe issue du modèle de détection des frontières de TTU comparée à MOMEL. 349
Figure 111 : Exemple de frontières détectées dans le modèle 3. 351
Figure 112 : Exemple de frontières détectées dans le modèle 11. 353
Figure 113 : Exemple de modélisation problématique d'un TU A 358
Figure 114 : Exemple de modélisation problématique d'un TU A dans un IU emphatique 359
Figure 115 : Exemple de problème de modélisation du TU A dû aux paramètres d'alignement sur G0302B. 360
Figure 116 : Comparaison des modèles 11 et 24 pour le même exemple. 360
Figure 117 : Exemple de registre étendu localement non modélisé sur le fichier G0103B. 361
Figure 118 : Exemple de contour vocatif non modélisé sur le fichier G0501G. 362
Figure 119 : Exemple de contour montant + grand chute associée à un IU [!| sur G0302B. 363
Figure 120 : Exemple de nucléus confondu avec un accent rythmique postnucléaire 364
Figure 121 : Exemple de nucléus confondu avec un accent rythmique postnucléaire dans les modèles 11 et 24 365
Figure 122 : Exemples de détection de fonction pour un énoncé ambigu 366

Table des tableaux

Tableau 1 : Corrélation moyenne pour chaque modèle et par type d'unité intonative (tous, [+, [|). 185
Tableau 2 : Optimisation des séquences de tons INTSINT entre deux pauses. 190
Tableau 3 : Extrait du fichier récapitulatif des séquences de tons testées entre deux pauses. . 191
Tableau 4 : Résultat pour FA : meilleures formes par type de frontière 191
Tableau 5 : Meilleures séquences de tons et alignements pour chaque locuteur. 193
Tableau 6 : Meilleures séquences de tons INTSINT pour le locuteur FA avec modélisation au niveau de l'IU. 194
Tableau 7 : Meilleures séquences de tons et alignements pour le locuteur FA avec modélisation au niveau de l'IU. 194
Tableau 8 : Extrait du tableau montrant les séquences de ton INTSINT testées au niveau d'un TU et les valeurs de corrélation et de l'écart quadratique moyen obtenues. 199
Tableau 9 : Différents types de TU annotés en position initiale 203
Tableau 10 : Différents types de TU annotés en position médiane et finale 203
Tableau 11 : Meilleures séquences de tons INTSINT au niveau des TU. 205
Tableau 12 : Les huit meilleurs alignements pour les séquences de tons optimales dans les TU A1 et AM. 206
Tableau 13 : Exemple des combinaisons de frontières successivement essayées lors de la détection des frontières d'IU. 209
Tableau 14 : Résultats de la détection des frontières dans le modèle des pauses (BD1) et dans le modèle des IU (BD3). 210
Tableau 15 : Tous les types de TU possibles dans les modèles sans contrainte en fonction de la position du TU dans l'IU. 215
Tableau 16 : Tous les types de TU possibles avec la règle du nucléus. 215
Tableau 17 : Un exemple des traitements effectués afin de placer une frontière intermédiaire sur un passage du TextGrid FEO10704. 218
Tableau 18 : Résultats des détections de frontières dans les modèles 1 à 8. 219
Tableau 19 : Matrice de confusion des types de TU annotés et des types de TU détectés tous modèles confondus. 223
Tableau 20 : Matrice de confusion des types de TU annotés et des types de TU détectés pour les modèles 5, 6, 7 et 8. 223
Tableau 21 : Classement des modèles d'optimisation des tons en fonction de la corrélation, du score et de l'écart quadratique moyen. 226
Tableau 22 : Classement des modèles d'optimisation de l'alignement en fonction de la corrélation, du score et de l'écart quadratique moyen. 226
Tableau 23 : Classement des modèles de détection des frontières en fonction de la corrélation, du score et de l'écart quadratique moyen. 227
Tableau 24 : Classement des modèles de détection des proéminences en fonction de la corrélation, du score et de l'écart quadratique moyen. 227
Tableau 25 : Annotation des frontières en TSM et en IF. 237

Tableau 26 : Équivalences entre les TSM et IF. 238
Tableau 27 : Type de TU en fonction de la position dans l'IU et de la position du nucléus. 239
Tableau 28 : Exemples de l'annotation fonctionnelle complète pour un nucléus. 240
Tableau 29 : Matrice de confusion mettant en valeur les corrections manuelles effectuées au niveau des unités intonatives. 241
Tableau 30 : Matrice de confusion comparant la conversion automatique des TU et la version corrigée. 243
Tableau 31 : Matrice de confusion pour les TU annotés manuellement et la conversion directe des TSM. 246
Tableau 32 : Types de TTU1 en fonction de la position dans l'IU. 247
Tableau 33 : Types de TTU2 en fonction de la position dans l'IU. 248
Tableau 34 : Tons possibles en fonction de la proportion de l'étendue et des différents codages d'INTSINT. 251
Tableau 35 : Tableau récapitulatif de tous les modèles utilisés. 254
Tableau 36 : Mode d'alignement / Point d'ancrage pour le modèle 6. 256
Tableau 37 : Mode d'alignement / Point d'ancrage pour le modèle 7. 256
Tableau 38 : Mode d'alignement / Point d'ancrage pour le modèle 8. 257
Tableau 39 : Mode d'alignement / Point d'ancrage pour le modèle 9. 257
Tableau 40 : Mode d'alignement / Point d'ancrage pour le modèle 11. 258
Tableau 41 : Mode d'alignement / Point d'ancrage pour le modèle 12. 258
Tableau 42 : Mode d'alignement / Point d'ancrage pour le modèle 13. 259
Tableau 43 : Mode d'alignement / Point d'ancrage pour le modèle 14. 259
Tableau 44 : Mode d'alignement / Point d'ancrage pour le modèle 15. 260
Tableau 45 : Mode d'alignement / Point d'ancrage pour le modèle 16. 260
Tableau 46 : Mode d'alignement / Point d'ancrage pour le modèle 22. 261
Tableau 47 : Mode d'alignement / Point d'ancrage pour le modèle 23. 261
Tableau 48 : Paramètres d'optimisation de l'alignement tonal en fonction du type de point d'ancrage et de l'étape. 263
Tableau 49 : Paramètres d'optimisation des séquences de tons entre deux pauses. 264
Tableau 50 : Meilleures séquences de tons INTSINT pour le modèle 1. 264
Tableau 51 : Meilleures séquences de tons retenues par type d'IU pour optimiser l'alignement. 265
Tableau 52 : Extrait du tableau R montrant les résultats de l'optimisation des séquences par TU et IU dans le modèle 11. 266
Tableau 53 : Les 10 meilleurs séquences par type de TU dans le modèle 11 pour un IU [|. 267
Tableau 54 : Meilleures séquences pour les TU A et S dans un IU [!| (modèle 11). 267
Tableau 55 : Meilleures séquences pour les TU N et !N dans un IU [+ (modèle 11). 268
Tableau 56 : Meilleures séquences de tons au niveau des TTU dans le modèle 16. 268
Tableau 57 : Meilleures séquences de tons au niveau des TTU2 dans le modèle 23 pour un IU [!+. 269
Tableau 58 : Meilleures séquences de tons au niveau des TTU N dans le modèle 16. 269
Tableau 59 : Meilleures séquences de tons au niveau des TTU A dans le modèle 24. 270

Tableau 60 : Classement de tous les modèles en fonction de la corrélation et du type d'IU. 271
Tableau 61 : Classement de tous les modèles en fonction du score et du type d'IU. 272
Tableau 62 : Classement de tous les modèles en fonction de l'écart quadratique moyen et du type d'IU. 273
Tableau 63 : Répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la syllabe accentuée ou de la durée de l'unité tonale. 275
Tableau 64 : Répartition des meilleurs points d'alignement : alignement par rapport au début de l'unité tonale et par rapport à l'attaque de la première voyelle. 276
Tableau 65 : Répartition des meilleurs points d'alignement pour P1 en fonction de la position du TU. 277
Tableau 66 : Répartition des meilleurs points d'alignement pour P1 en % de la durée du TU et en fonction de la position du TU. 279
Tableau 67 : Répartition des meilleurs points d'alignement pour P1 par rapport au début du TU et en fonction de la position du TU. 280
Tableau 68 : Répartition des meilleurs points d'alignement pour P1 par rapport au début de la voyelle accentuée et en fonction de la position du TU. 281
Tableau 69 : Répartition des meilleurs points d'alignement de P1 en % de la durée de la syllabe en fonction de la position du nucléus. 282
Tableau 70 : Répartition des meilleurs points d'alignement en fonction de la position du nucléus et en pourcentages cumulés. 282
Tableau 71 : Répartition des meilleurs points d'alignement pour P1 par rapport au début du TU et en fonction de la position du nucléus. 283
Tableau 72 : Répartition des meilleurs points d'alignement pour P1 par rapport au début de la voyelle et en fonction de la position du nucléus. 284
Tableau 73 : . Variations de l'alignement de P1 dans les contours HL, LH, HLH par rapport à la syllabe accentuée et en fonction de la position du TU. 286
Tableau 74 : Variations de l'alignement de P1 par rapport au début du TU en fonction de la position du TU. 287
Tableau 75 : Variations de l'alignement de P1 par rapport au début de la voyelle en fonction de la position du TU. 289
Tableau 76 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1 et M. 291
Tableau 77 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position F avec point d'ancrage la cible précédente ou la durée du TU. 292
Tableau 78 : Répartition des meilleurs points d'alignement pour P2 en fonction de la dernière syllabe ou de la dernière voyelle. 294
Tableau 79 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la cible précédente ou la durée du TU. 295
Tableau 80 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la dernière syllabe ou la dernière voyelle. 296
Tableau 81 : Répartition des meilleurs points d'alignement pour P1 dans les TTU2 avec comme points d'ancrage la première syllabe, première voyelle ou début du TTU. 301

Tableau 82 : Répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la première syllabe et de la position du TTU. 303
Tableau 83 : Répartition des meilleurs points d'alignement pour P1 par rapport au début du TTU et en fonction de la position du TTU. 304
Tableau 84 : Répartition des meilleurs points d'alignement pour P1 par rapport au début de la voyelle et en fonction de la position du TTU. 305
Tableau 85 : Répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la syllabe et de la position du nucléus. 306
Tableau 86 : Répartition des meilleurs points d'alignement pour P1 en fonction de la première voyelle et de la position du nucléus. 307
Tableau 87 : Répartition des meilleurs points d'alignement pour P2 en fonction de la cible tonale précédente. 308
Tableau 88 : Répartition des meilleurs points d'alignement pour P2 en fonction de la durée de la dernière syllabe du TTU. 309
Tableau 89 : Répartition des meilleurs points d'alignement pour P3 en fonction de la fin du TTU et de la fin de la dernière syllabe du TTU. 310
Tableau 90 : Classement des modèles en fonction de la corrélation moyenne et du type d'IU. 312
Tableau 91 : Classement des modèles en fonction de l'écart quadratique moyen et du type d'IU. 313
Tableau 92 : Classement des modèles en fonction du score moyen du type d'IU. 313
Tableau 93 : Classement des modèles en fonction de la corrélation moyenne. 314
Tableau 94 : Classement des modèles en fonction de l'écart quadratique moyen. 315
Tableau 95 : Classement des modèles en fonction du score moyen. 315
Tableau 96 : Classement des modèles basés sur les variantes d'INTSINT en fonction de la corrélation moyenne. 316
Tableau 97 : Classement des modèles basés sur les variantes d'INTSINT en fonction de l'écart quadratique moyen . 316
Tableau 98 : Classement des modèles basés sur les variantes d'INTSINT en fonction du score moyen. 316
Tableau 99 : Matrice de confusion des types d'IU annotés et détectés pour les modèles des TU. 330
Tableau 100 : Matrice de confusion des types d'IU annoté et détectés pour les modèles en TTU2. 330
Tableau 101 : Matrice de confusion des types de TU annotés et détectés dans les modèles 6 à 14. 331
Tableau 102 : Matrice de confusion des types de TTU annotés et détectés dans les modèles TTU2. 332
Tableau 103 : Résultats de la détection du nucléus par modèle et en pourcentage. 332
Tableau 104 : Résultats de détection du nucléus en position initiale. 333
Tableau 105 : Résultats de détection du nucléus en position médiane. 333
Tableau 106 : Résultats de détection du nucléus en position finale. 334
Tableau 107 : Résultats de détection du nucléus emphatique par modèle. 334

Tableau 108 : Résultats de détection du nucléus emphatique en position initiale. 335
Tableau 109 : Résultats de détection du nucléus emphatique en position médiane. 335
Tableau 110 : Résultats de détection du nucléus emphatique en position finale. 336
Tableau 111 : Résultats de détection du TU A par modèle. 336
Tableau 112 : Résultats de détection du TU A en position initiale. 337
Tableau 113 : Résultats de détection du TU A en position médiane. 337
Tableau 114 : Résultats de détection du TU S par modèle. 338
Tableau 115 : Résultats de détection du TU S en position initiale. 338
Tableau 116 : Résultats de détection du TU S en position médiane. 339
Tableau 117 : Résultats de détection du TU S en position finale. 339
Tableau 118 : Résultats de détection du TTU N. 340
Tableau 119 : Résultats de détection du TTU N en position initiale. 341
Tableau 120 : Résultats de détection du TTUN en position médiane. 341
Tableau 121 : Résultats de détection du TTU N en position 1F. 341
Tableau 122 : Résultats de détection du TTU N en position finale. 342
Tableau 123 : Résultats de détection du TTU !N. 342
Tableau 124 : Résultats de détection du TTU !N en position 1F. 343
Tableau 125 : Résultats de détection du TTU !N en position médiane. 343
Tableau 126 : Résultats de détection du TTU !N en position finale. 343
Tableau 127 : Résultats de détection du TTU A. 344
Tableau 128 : Résultats de détection du TTU A en position initiale. 344
Tableau 129 : Résultats de détection du TTU A en position médiane. 344
Tableau 130 : Résultats de détection du TTU A en position finale. 345
Tableau 131 : Matrice de confusion des types d'IU annotés et détectés dans le modèle 11. . 346
Tableau 132 : Matrice de confusion des types d'IU annotés et détectés dans le modèle 24. . 346
Tableau 133 : Résultats de détection des types de proéminence dans les modèles britanniques adaptés. 347
Tableau 134 : Résultats de détection des frontières de TTU dans le modèle 24. 348
Tableau 135 : Résultats de détection des frontières de TTU en fonction du type de TTU et de la position dans l'IU. 348
Tableau 136 : Résultats de détection des frontières d'IU en fonction du type d'IU dans les modèles 1 et 3. 350
Tableau 137 : Résultats de détection des frontières d'IU en fonction du type d'IU pour modèle 11. 352
Tableau 138 : Répartition des notes obtenues en fonction du type de TU modélisé dans le modèle 11. 357

INTRODUCTION

Motivation et présentation du sujet de l'étude

Plus d'un siècle d'études et de recherches sur l'intonation et nous sommes loin d'avoir fait le tour du sujet et d'arrêter de nous interroger sur cette faculté passionnante qu'est la parole.

On reconnaît communément à la prosodie un rôle très important dans la communication orale. On sait que la prosodie, et en particulier l'intonation qui est le sujet de notre étude, permet de véhiculer du sens en exprimant notre attitude par rapport à ce que l'on dit, en laissant transparaître notre émotion, en attirant l'attention sur telle ou telle partie du message etc. On sait que la façon dont on dit quelque chose peut être aussi importante que ce que l'on dit, comme le laisse entendre la remarque suivante : « *It's not what you said, it's the way you said it* ». Et dans de nombreuses situations de communication, c'est l'intonation utilisée qui est responsable de cet écart entre le contenu d'un énoncé d'un point de vue sémantique et l'interprétation que l'on en fait.

L'intonation interagit donc de façon complexe avec les autres canaux de la communication verbale que sont la syntaxe, la sémantique, la gestuelle, pour créer du sens et l'on peut s'interroger sur la façon dont elle s'intègre dans ce prisme complexe. Si dans certains cas précis il est aisé d'identifier le rôle de l'intonation, dans d'autres cas il est difficile de distinguer ce qui est véhiculé par l'intonation seule par rapport aux autres canaux de la communication verbale.

Travailler sur l'articulation entre la représentation des formes et des fonctions de l'intonation, c'est donc chercher à mettre en relation les manifestations phonétiques de l'intonation (les contours de F0, les mouvements de la voix) avec l'expression de fonctions communicatives comme la marque de l'ironie ou de l'étonnement. Ceci revient alors à confronter le niveau de l'expression et du contenu (cf. Di Cristo, 2004) : quelles formes utilise-t-on pour quelles fonctions ? dans quel but ? Cela revient aussi à confronter différents niveaux d'analyse de la prosodie allant du niveau le plus concret représenté par l'étude des variations de la courbe de la fréquence fondamentale par exemple, au niveau le plus abstrait représenté par l'élaboration de catégories phonologiques.

C'est une problématique qui a été au cœur des recherches portant sur l'intonation de l'anglais depuis fort longtemps déjà. Palmer (1924) définissait déjà l'intonation comme une science visant à expliquer la signification des différents contours intonatifs par opposition à la « tonétique » qui s'occupait essentiellement de les décrire et de montrer comment les contours s'enchaînent dans un énoncé :

> *« The science which is concerned with the nature and meaning of this tone-play is called intonation. That part which is concerned chiefly with the tone-curves irrespective of their meanings has been called Tonetics. »* Palmer (1924, p.3)

De multiples analyses et approches de la relation forme/sens au niveau de la prosodie et de l'intonation en particulier ont été développées par la suite : des approches principalement théoriques cherchant à mettre en relation chaque accent mélodique avec une unité de sens (comme l'approche compositionnelle de Pierrehumbert, 1990), des approches globales attribuant une signification particulière à des contours entiers seulement, des approches physiologiques s'appuyant sur des caractéristiques articulatoires ou des « codes biologiques » pour expliquer la relation au sens etc. (c'est ce que nous verrons en détail dans le premier chapitre de cette thèse).

Les approches et interprétations du rôle de la prosodie et par là-même de l'intonation sont donc multiples et variées et cela montre qu'il reste du chemin à parcourir afin de mieux comprendre le fonctionnement de l'intonation.

C'est l'analyse que fait Hirst (2005). Il affirme ainsi qu'il n'existe pas de consensus à nos jours quant à la façon d'expliquer et d'analyser la façon dont la prosodie contribue à la création du sens malgré le caractère central et essentiel de cette question et met en valeur le problème complexe que cela représente :

> *« There is, today, no general consensus on the way in which the prosody of an utterance contributes to its meaning. The last six chapters of Couper-Kuhlen (1986), constituting about half of the book, give a well-documented account of the various ways in which the problem of the meaning of intonation has been approached in the literature. The fact that intonation meaning can be approached in so many different ways, for results which, as Couper-Kuhlen (p 209) admits:*
>
> > *"are rather modest indeed"*
>
> *seems to indicate that we have still not got properly started on the analysis of intonational meaning. As Cruttenden (1986:184) puts it:*
>
> > *"It is not yet even clear what sorts of meanings are involved". »* Hirst (2005, p.335)

Hirst (2005) avance aussi qu'il n'y a pas de consensus non plus sur la façon de représenter le rôle de l'intonation d'un point de vue phonologique et même sur les apports sémantiques qui lui sont attribués. C'est cet état de fait qui a motivé notre choix de mener une étude sur la relation entre l'annotation des formes et fonctions prosodiques.

Pouvoir coder et représenter le rôle de l'intonation de façon consensuelle dans un système d'annotation reste encore un défi à relever et l'on remarquera que celui-ci s'inscrit dans la mouvance de la recherche actuelle sur l'intonation.

C'est un défi qui a son importance pour l'enseignement des langues étrangères et dans le développement des nouvelles technologies du langage telles que la reconnaissance automatique de la parole ou les systèmes de synthèse, pour ne citer que quelques exemples.

Si l'usage de la prosodie pour véhiculer du sens semble être universel, la façon dont les formes sont reliées aux fonctions est très certainement spécifique à chaque langue et doit être connue de l'apprenant d'une seconde langue. Le développement du codage de l'intonation et plus précisément l'automatisation de celui-ci peut aussi viser à améliorer les systèmes de traitement automatique de la parole.

Les travaux récents de Vainio et *al* (2009) illustrent bien cette idée : en partant d'une annotation fonctionnelle manuelle celui-ci a montré que l'on peut améliorer la synthèse de la parole multilingue.

Les travaux de Xu (2005, 2006) s'inscrivent aussi dans cette mouvance qui prend conscience de l'importance de l'explicitation des fonctions communicatives de l'intonation pour mieux décrire les systèmes intonatifs de diverses langues. Dans une thèse récente, son élève Liu (2009) met ce principe en application en proposant une étude de l'intonation de l'anglais et du mandarin basée sur une approche fonctionnelle. Elle cherche alors à mettre en relation des aspects fonctionnels comme l'accent lexical, le focus et le type d'énoncé avec des variations spécifiques de F0. La remarque qui introduit sa conclusion résume bien notre idée de l'importance de l'analyse des fonctions prosodiques :

> *« Much controversy in the literature results from lack of recognition of various communicative functions. As an effort to achieve a better understanding of intonation systems of Mandarin Chinese and General American English, this dissertation explores the contribution of lexical tone/stress, focus and sentence-type in declarative and interrogative sentences in the two languages, with an approach that prioritizes the functional aspect of speech [..].* » Liu (2009, p.148)

Nous nous proposons alors dans cette thèse de développer notre propre approche fonctionnelle de l'intonation de l'anglais britannique dont nous décrirons les caractéristiques principales dans la section suivante.

Objectifs et méthodologie

L'objectif principal de cette étude est de développer une méthode de recherche permettant de mettre en relation l'annotation des formes et fonctions prosodiques de façon automatique et d'étudier les règles de dérivation permettant de prédire un niveau d'annotation en fonction de l'autre. Nous espérons qu'une telle approche nous permettra de proposer un modèle de l'articulation forme/fonction à partir de l'étude de corpus et de développer l'annotation fonctionnelle de l'intonation.

Notre approche comporte certaines particularités. Tout d'abord nous utilisons deux systèmes d'annotation distincts pour les informations formelles et fonctionnelles suivant Hirst (2005), et par opposition à la plupart des systèmes d'annotation existant tels que TOBI et les TSM (« Tonetic Stress Marks »), qui représentent en général les deux types d'information.

Puis, nous basons notre approche sur le moins d'aprioris théoriques possibles en utilisant une annotation fonctionnelle minimale représentée par le système IF (« Intonative Features »/ « Intonation Functions », Hirst 1977 et 2005). Celui-ci suffit d'après Hirst (2005) à capturer les aspects fonctionnels contribuant à l'interprétation syntaxique d'un énoncé. Ce système permet d'annoter des niveaux de proéminence (inaccentué, accentué, nucléaire, nucléaire emphatique) et des catégories de frontières (frontières terminales et non terminales). On essaiera ensuite de développer cette annotation de base en la confrontant à l'étude de corpus.

Les formes prosodiques sont annotées automatiquement en utilisant l'algorithme MOMEL-INTSINT. L'algorithme MOMEL fournit une représentation phonétique des données alors que le codage INTSINT en donne une représentation phonologique de surface (cf. Hirst, Di Cristo & Espesser, 2000). La thèse cherchera donc à mettre en relation la représentation fonctionnelle avec ces deux niveaux d'analyse des représentations formelles.

Enfin on utilisera l'analyse par synthèse comme outil privilégié afin de tester différents modèles de l'intonation. Il s'agira ainsi de générer des représentations formelles à partir d'une représentation minimale des fonctions prosodiques et de comparer le résultat obtenu avec les données observées. En observant les manques et erreurs de prédiction, on espère ensuite pouvoir réévaluer et enrichir la représentation fonctionnelle et voir si cela permet une meilleure reproduction des données.

Plan de la thèse

Cette thèse se décompose en deux parties principales : la première partie est une revue de lecture sur l'intonation de l'anglais britannique et permet d'approfondir notre réflexion sur le sujet de cette étude : elle constitue le contexte scientifique de la thèse. La seconde partie décrit le travail expérimental que nous avons ensuite développé.

La première partie de la thèse se décompose en quatre chapitres : le premier chapitre est dédié à la description des fonctions généralement associées à la prosodie, le second à la description des principaux modèles des formes prosodiques, le troisième fait une revue des systèmes d'annotation prosodique utilisés en anglais et le quatrième est consacré aux différentes approches de l'articulation forme/fonction c'est-à-dire aux différentes méthodes de recherche.

Le chapitre 1 aborde la question des fonctions attribuées à la prosodie de façon générale. Nous faisons dans cette partie un premier inventaire des termes fonctionnels associés à la prosodie dans toutes les langues avant de nous concentrer sur la littérature anglophone. Six fonctions spécifiques sont ensuite décrites et analysées : la fonction lexicale, grammaticale, focale, discursive, expressive et identificatrice. Ce chapitre permet de mettre en valeur la diversité des analyses fonctionnelles de l'intonation chez les différents auteurs et la complexité de la tâche que l'on se propose d'effectuer.

Le second chapitre aborde naturellement la question des formes prosodiques de l'anglais. Nous présentons d'abord la façon dont les contours intonatifs ont été décrits et étudiés dans deux modèles théoriques opposés : l'école britannique de l'intonation et les théories autosegmentales. Nous décrivons donc l'opposition classique entre la théorie des contours et la théorie des niveaux en observant la façon dont les variations de F0 sont analysées dans les deux approches d'un point de vue formel. Enfin dans une troisième sous-section, nous abordons la question de l'alignement tonal qui est une problématique formelle plus récente (datant des années 80).

Dans le troisième chapitre, nous présentons un court historique des systèmes d'annotation utilisés en anglais britannique en partant des portées musicales jusqu'aux systèmes les plus utilisés aujourd'hui (TOBI et les TSM). Notre fil conducteur reste dans ce chapitre l'illustration de la relation entre l'annotation des formes et fonctions prosodiques dans chaque système. Nous faisons aussi une évaluation critique des systèmes les plus utilisés tels que TOBI et expliquons nos propres choix expérimentaux en termes de systèmes d'annotation.

Le quatrième chapitre analyse les différentes approches de l'étude de l'articulation forme/fonction. Nous comparons dans ce chapitre différentes méthodes de recherche comme les approches théoriques (dites « Topdown ») qui se basent sur le niveau du discours ou de la pragmatique pour expliquer la relation des formes au sens, ou les approches plus expérimentales (dites « Bottom up ») qui se basent sur les caractéristiques de la perception ou au contraire de la production et des contraintes articulatoires. C'est le lieu d'une réflexion sur les différentes

méthodes de recherche que l'on peut utiliser pour réaliser notre étude en fonction des exemples que l'on trouve dans la littérature.

La deuxième partie de cette thèse présente notre travail expérimental. Cette partie est constituée de cinq chapitres :

Dans le chapitre 5, nous présentons brièvement les deux corpus que nous avons utilisés. Nous présentons d'abord le corpus EUROM1 qui est la base de nos premières tentatives de modélisation de la relation forme/fonction. Nous décrivons aussi les développements de l'annotation fonctionnelle qui a été effectuée manuellement. Nous présentons ensuite le corpus Aix-MARSEC dont nous utiliserons aussi une partie pour nos travaux.

Le chapitre 6 fait état de notre première tentative d'analyse par synthèse des contours intonatifs de l'anglais britannique. Nous expliquons d'abord les tenants et aboutissants de notre méthode de recherche. Puis cinq modèles explicites de plus en plus complexes de l'articulation forme/fonction sont présentés et évalués par analyse par synthèse.

Le chapitre 7 relate les étapes d'optimisation des paramètres des premiers modèles développés à partir de l'étude du corpus EUROM1. Nous décrivons l'élaboration d'une série d'outils permettant d'optimiser les représentations formelles en fonction des représentations fonctionnelles. Une série de scripts est élaborée afin de mettre en relation les étiquettes fonctionnelles fournies avec des contours mélodiques canoniques codés avec INTSINT. C'est dans ce chapitre que les questions d'ordre technique sont abordées : comment extraire des représentations formelles constantes à partir de l'annotation fonctionnelle ? comment modéliser les variations de l'alignement des points cibles ? comment évaluer la qualité de nos modèles ? À l'issue du développement de notre méthode de travail, nous présentons aussi un premier modèle expérimental de l'articulation forme/fonction.

Le chapitre 8 est le lieu de l'approfondissement de notre travail de recherche à partir de la base méthodologique développée autour du corpus EUROM1. On s'attaque à la partie « fiction » du corpus MARSEC qui constitue une base de données de parole de type expressif ce qui ouvre la voie à un enrichissement de notre modèle de l'annotation des formes et fonctions prosodiques. L'annotation fonctionnelle est fournie automatiquement par une conversion des TSM déjà présents dans le corpus, en IF. La méthode d'analyse par synthèse est ensuite utilisée pour tester plusieurs interprétations des fonctions de l'intonation qui constituent des adaptations du système IF mais aussi plusieurs façons de modéliser les représentations formelles en termes de codage des tons mélodiques et de leur alignement. À l'issue de ces étapes expérimentales, nous proposons une série de grammaires de l'articulation forme/fonction ainsi qu'une série d'expériences visant à utiliser ces généralisations pour la détection automatique des informations fonctionnelles à partir des représentations formelles.

Le chapitre 9 constitue la conclusion générale de la thèse et présente une synthèse des résultats obtenus à partir des différentes expériences menées sur les corpus EUROM1 et MARSEC.

PARTIE I

CONTEXTE SCIENTIFIQUE

Chapitre 1

Fonctions généralement attribuées à la prosodie

Le terme « fonction » de la prosodie fait référence au rôle joué par la prosodie dans la communication orale. Comme le résume très bien E. Couper-Kuhlen (1986) dans son ouvrage le terme fonction peut se définir comme suit :

> *« Role of a given element or system in language communication »* (p. 110)

Dans le cas de l'étude de la prosodie, il s'agira donc d'étudier le rôle des différents éléments qui constituent le système prosodique comme les tons, l'intonation, l'accent, le tempo et les pauses. Un des problèmes qui se pose immédiatement, à la lecture des multiples ouvrages traitant des fonctions prosodiques, est la multiplicité des termes fonctionnels utilisés.

Répertorions d'abord les termes fonctionnels trouvés dans les références bibliographiques de ce mémoire (Couper-Kuhlen 1986, Faure 1962, Jassem 1952). On obtient la liste suivante :

> *« fonction passive, fonction active, fonction expressive (liée à l'expression des émotions ou bien des attitudes), fonction grammaticale, fonction de discours, syntaxique, de délimitation, vocative, textuelle, indexicale, identificatrice, marquage de proéminence, focalisation et mise en perspective, culminative, illocution, esthétique, évocatrice, optimisation de la production, de la perception, activation des inférences, fonction mécanique, fonction d'identification de différents styles de discours. »*

On peut reconnaître dans cette liste au moins deux types de fonctions différentes : les fonctions de type structurel (Post *et al.*, 2007) et les fonctions de type non structurel, ce qui équivaut à faire la distinction entre le rôle linguistique joué par la prosodie en interaction avec la grammaire, la syntaxe et la sémantique et les apports ou informations extra textuelles qu'elle peut apporter. Au moyen de l'intonation, on ne se contente pas de venir souligner ou au contraire de minimiser l'importance de tel mot ou syllabe, mais on peut aussi véhiculer ce que l'on ressent par rapport à ce que l'on dit en utilisant un contour ironique ou une intonation marquant le détachement, pour ne citer que deux exemples. On retrouve cette distinction entre les fonctions paralinguistiques et linguistiques chez de nombreux auteurs tels que Ladd (1996) et Fujisaki (1997) :

> *« Prosody is the systematic organization of various linguistic units into an utterance or a coherent group of utterances in the process of speech production. Its realization involves both segmental and suprasegmental features of speech and serves to convey not only linguistic information but also paralinguistic and non linguistic information. »* Fujisaki (1997, p. 28)

Il reconnaît donc deux aspects fondamentaux en prosodie : il y a d'abord l'aspect linguistique puis l'aspect paralinguistique. La prosodie permet d'exprimer une information symbolique représentée par des symboles discrets qui se combinent entre eux selon des règles caractéristiques. D'autre part la prosodie permet de véhiculer ses informations non contenues dans le matériau verbal se cumulant avec l'information linguistique pour ajouter du sens (fonction identificatrice, expressive…). La notion d'information paralinguistique est toutefois assez controversée, on s'en tiendra ici à la définition donnée par Ladd (1996). Celui-ci considère que les éléments prosodiques dits paralinguistiques participent à l'interprétation de l'énoncé, mais ne sont pas organisés en codes linguistiques élaborés.

Dans ce chapitre, nous nous proposons de faire une présentation succincte des principales fonctions prosodiques étudiées. On parlera de fonction lexicale, grammaticale, discursive, de focalisation, de la fonction expressive et de la fonction identificatrice.

Dans le chapitre 4 portant sur l'articulation forme/fonction, nous reviendrons en détails sur certaines analyses fonctionnelles seulement introduites dans ce chapitre en nous concentrant cette fois-ci sur l'intonation de l'anglais.

1.1 La fonction lexicale

Nous utilisons ce terme suivant Hirst (2006) qui décrit cette fonction en partant de l'observation de la prosodie de plusieurs langues telles que le finnois, le vietnamien, le japonais, le russe et l'anglais. Les facteurs prosodiques permettent, dans certaines langues, d'identifier des mots, d'en changer le sens ou la nature grammaticale. En finnois, la durée est un paramètre prosodique qui permet de distinguer des mots comme : /taka/, /takka/, /taakka/, /takaa/, /takkaa/, /taakkaa/ (cf. Hirst 2006, p.168) qui ont tous une signification différente (« *back-, fireplace, burden* » suivis par le cas du partitif pour chacun de ces mêmes mots). La quantité est donc une caractéristique distinctive lexicale utilisée en finnois (cf. figure 1).

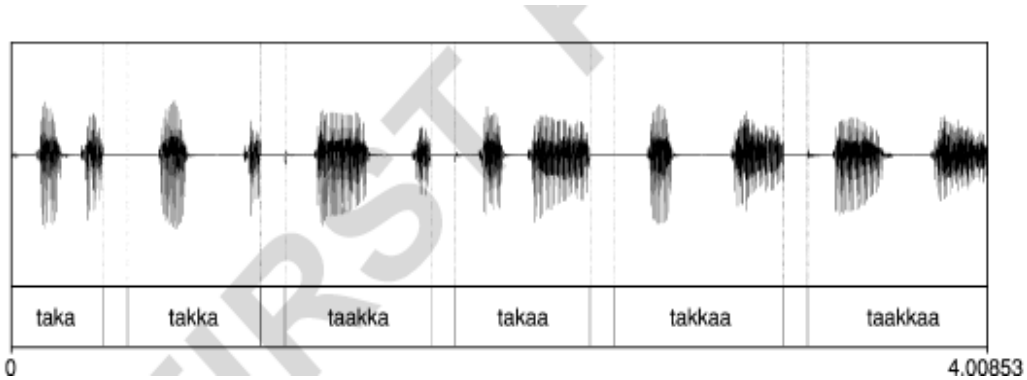


Figure 1 : Illustration des variations distinctives de la quantité dans les voyelles et consonnes en finnois. Extrait de Hirst (2006, p.168).

Hirst (2006) prend aussi les exemples des langues à tons comme le vietnamien. En vietnamien, on utilise en effet la fréquence fondamentale pour distinguer les différentes significations de la syllabe /ma/. Celle-ci peut être prononcée avec six tons différents (avec un ton statique aigu, une petite montée, une montée glottalisée, une petite chute, une chute glottalisée et un ton descendant–montant). Ce mot peut signifier respectivement : fantôme, mère, cheval, maïs, tombe, jeune pousse de riz.

En anglais, ce sont les paramètres de durée, d'intensité et de fréquence fondamentale associés à la production de l'accent de mot, qui participent à la réalisation de la fonction lexicale. Par une localisation distincte de l'accent de mot, on peut changer la nature grammaticale de certains mots comme les paires minimales similaires à « *'import* » *et « im'port* ». Avec un accent initial, on fait référence au nom « *import* » alors qu'un accent sur la deuxième syllabe fait référence à l'action d'importer, c'est-à-dire au verbe « *to import* ». Cela n'est pas une règle systématique en anglais, mais toute une série de mots illustre cette fonction lexicale de l'accent de mot. Ce sont en général des mots dissyllabiques qui désignent un nom quand leur première syllabe est accentuée et un verbe si la deuxième syllabe est accentuée, mais on trouve aussi des cas où l'accent distingue un adjectif d'un verbe (cf. « *'perfect* » (adj.) *vs* « *per'fect* » (v.)).

Nous avons donc vu des exemples du rôle de la prosodie au niveau de l'identification des mots. Dans le reste de ce chapitre, nous présenterons les fonctions prosodiques appliquées à un niveau d'analyse supérieur à celui du mot tel que les syntagmes, les énoncés ou le discours. Ce sont des fonctions dites « supra-lexicales » ou « non lexicales » (cf. Hirst et Di Cristo, 1998).

1.2 La fonction grammaticale

Traditionnellement, on associe l'étude de l'intonation de l'anglais ou des phénomènes prosodiques à la grammaire. C'est ce qu'Halliday (1970) appelle « speech function » ou « communicative function ». Le terme grammatical est à mettre en relation ici avec les contrastes intonatifs liés à la modalité ou aux types d'énoncés. Ainsi, on avance qu'un contour montant utilisé à la place d'un contour descendant transforme une affirmation en une question, un ordre en une suggestion ou requête :

> |*William's having* `<u>*dinner*</u>.|| (Contour descendant sur *« dinner »* = il s'agit ici d'un constat)
>
> *vs*
>
> |*William's having* <u>*dinner*</u>.|| (Montant = *Is William having dinner? Are you sure?)*

L'affirmation syntaxique apparente est remise en question par l'intonation finale montante.

| *Sing it one more `time*.|| (D = on a affaire à un ordre)

| *Sing it one more time*.|| (M = l'ordre est atténué par la montée finale et est plutôt perçu comme une requête ou une suggestion)

Ce lien étroit établi entre l'intonation et la grammaire se retrouvait aussi déjà dans les travaux d'Armstrong et Ward (1931) ou O'Connor & Arnold (1961). Ainsi le contour numéro un dit « tune 1 » (contour descendant) d'après Armstrong et Ward est utilisé dans les cas suivants :

1. Les affirmations définies, communes, certaines :

   I had to do it.
2. Les questions ouvertes (qui requièrent une réponse autre que oui ou non) :

   When? How did it happen?
3. Les orders : *Tidy up your room.*
4. Les phrases exclamatives : *How easy that was!*

Et le contour numéro deux (tune 2) ou contour montant, dans les cas suivants :

1. Les affirmations non certaines :

   It seems rather complicated .
2. Les questions fermées auxquelles on répond par oui ou non (Yes/No questions) :

   Are you OK?
3. Les requêtes :

   Don't trouble to answer it.
4. Les énoncés incomplets:

   When she came back home, they had already left (Armstrong et Ward, 1931, p. 9)

Selon cette approche, la prosodie tient essentiellement le rôle d'un marqueur de ponctuation orale, qui indique, comme le fait la virgule, que la proposition n'est pas terminée ou indique, comme le point d'interrogation, que le locuteur est en train de poser une question, etc. C'est un indicateur du type d'acte de langage entrepris par le locuteur (il est en train de poser une question, de mettre en doute une affirmation, etc.).

On regroupe aussi dans cette catégorie grammaticale les liens de la prosodie avec la syntaxe. On dit alors que la position du nucléus a une fonction syntaxique et permet par là même d'expliciter les structures syntaxiques ambiguës comme le montre l'exemple ci-dessous :

|*George has directions to `FOLlow*||

vs

|*George has di`RECtions to follow*|| dans Couper-Kuhlen (1986, p. 111)

Dans le premier énoncé, on a dit à George de suivre quelqu'un ou quelque chose alors que dans le second exemple George doit suivre un certain nombre d'instructions. La construction syntaxique « *directions to follow* » peut être dérivée soit de « *directions that George must follow* » (les instructions que George doit suivre) soit de « *George must follow the directions* » (George doit suivre les instructions).

De la même manière, le placement des frontières d'unités intonatives peut désambiguïser un énoncé :

| *Those who sold `quickly |made a `profit* ||

vs

| *Those who `sold |quickly made a `profit* || Dans Roach (1991, p.174)

Dans le premier exemple, il y a une frontière après « *quickly* » dont la première syllabe est le nucléus. Cela signifie que ceux qui ont vendu rapidement ont fait des bénéfices. Dans le deuxième exemple, le nucléus est porté par « *sold* » et une frontière est placée après ce mot. On peut traduire l'énoncé de cette façon : ceux qui ont vendu (leurs actions par exemple) ont fait des bénéfices rapidement. Le changement de sens est induit par le rôle de l'adverbe « *quickly* » qui dans le premier cas porte sur le verbe « *sold* » et dans le deuxième sur le verbe « *made* ».

Cette analyse du rôle de l'intonation fait toutefois l'objet de nombreuses critiques. Certains chercheurs (cf. Bolinger 1965) avancent que ce lien étroit entre la grammaire et l'intonation n'est pas obligatoire et souvent même accidentel :

> *« Encounters between intonation and grammar are casual not causal. »*
> Bolinger (1965)

Hirst (1998) s'oppose à cette conception du rôle de l'intonation et en propose une analyse plus précise en se posant la question suivante : le choix du contour mélodique final peut-il vraiment transformer une affirmation en une question ?

Pour beaucoup de linguistes le contour final utilisé pour exprimer la continuité et l'interrogation est le même : il s'agit de la petite montée ou montée inférieure « low rise ».

Quand une distinction est effectuée, on observe une opposition entre la petite et la grande montée (« low rise » vs « high rise »). La petite montée serait utilisée dans les affirmations inachevées ou véhiculant une implication particulière et la grande montée correspondrait à un énoncé interrogatif d'après Kingdon (1958).

Palmer avance que la petite montée ne s'utilise que dans les affirmations et les ordres par opposition aux questions.

O'Connor et Arnold (1961) affirment que la petite montée est le contour le plus utilisé dans les questions fermées de type «Yes/No», mais avancent que pour transformer une affirmation en une question, il faut une grande montée.

D'après Hirst (1998) cette conception du rôle de l'intonation est problématique : en effet un énoncé perçu comme une question dans un contexte particulier peut ne plus être perçu comme tel dans un tout autre contexte. Il mentionne alors l'exemple des travaux de Brown, Currie et Kenworthy (1980) portant sur l'anglais d'Édimbourg. Ceux-ci ont montré que lorsque l'on demande à un sujet de produire un contour interrogatif hors contexte, il est hautement probable qu'il utilise un contour qu'il n'utiliserait que très rarement dans le discours spontané.

Bolinger (1986) et Couper-Kuhlen (1986) ont aussi suggéré qu'il était erroné d'établir un lien direct entre le contour mélodique final et le type d'énoncé. En effet, il est possible de trouver des questions sans contour final montant et inversement.

Hirst (1998) en conclut qu'il faut distinguer le type d'énoncé ou la modalité au niveau syntaxique de l'acte de langage au niveau pragmatique. Il prend alors les exemples suivants, opposant différentes structures syntaxiques et la possibilité d'utiliser un contour montant :

1(a) *He bought something*.
1(b)**He bought anything.* Hirst (1998, p. 64-65)

Dans le premier exemple, il est possible d'utiliser un contour montant et par là même d'obtenir un énoncé de type interrogatif même si la structure syntaxique est celle d'une affirmation. Avec l'utilisation de « anything » normalement associé à une modalité interrogative ou à la négation, l'énoncé est incorrect.

Si l'on prend les mêmes exemples en respectant la structure syntaxique de la question, le contour montant peut être utilisé indifféremment du fait que l'on emploie le mot « something » ou « anything » :

Did you buy something? vs Did you buy anything?

D'après Hirst (1998), l'intonation ne permet pas de transformer une affirmation en une question mais serait donc plutôt un moyen de montrer ou de signaler qu'une affirmation syntaxique est utilisée de façon pragmatique comme une requête d'information.

D'ailleurs, les questions avec une forme déclarative et un contour descendant sont très courantes :

"You manage in the distance alright?" Hirst (1998, p. 65)

Au-delà de la phrase ou de l'énoncé, on reconnaît aussi un lien entre la prosodie et le discours. C'est ce que Couper-Kuhlen (1986) appelle la fonction discursive ou textuelle.

1.3 Prosodie et discours : la fonction discursive

Cette fonction fait référence à l'étude de la relation entre plusieurs énoncés et la place de l'intonation au cœur de la discussion ou interaction, on préfère ici parler de fonction discursive. Prenons l'exemple suivant extrait de Couper-Kuhlen (1986) :

> The h|lecture was CANcelled|| the h|the speaker was ILL||
> The h|lecture was CANcelled| the |speaker was ILL.|| Brazil, Courthard et Johns (1980, p. 31)

Alors que dans le premier exemple, on a affaire à deux propositions a priori indépendantes, dans le second les deux propositions sont liées. Le redémarrage de F0 ou « reset » dans le premier exemple (noté h) signale la présence de deux propositions distinctes. Dans le second cas, on n'observe pas de variation significative de registre et la seconde proposition s'inscrit donc dans la continuité de la première, on en déduit un lien de cause à effet : la conférence a bien été annulée car le conférencier était malade.

Cette fonction est aussi associée à la gestion de la conversation, c'est-à-dire que l'on utilise l'intonation pour signaler les tours de parole par exemple.

C'est une étiquette fonctionnelle utilisée de façon assez large et diverse par les théoriciens de l'intonation tout comme la fonction grammaticale.

On peut remarquer par exemple que Roach (1991) parle de fonction discursive pour désigner l'exploration des contextes supérieurs à la phrase ou énoncé comme Couper-Kuhlen et souligne la fonction « subordination » dont on a montré un exemple plus haut, mais englobe aussi d'autres aspects fonctionnels sous cette même appellation. Il met ainsi la fonction de discours en relation avec les « actes de langage » : l'intonation dans cette interprétation permettrait d'indiquer ce que le locuteur fait : est-il en train de poser une question, de défier son interlocuteur, de conseiller ou d'encourager celui-ci ? Cette interprétation est proche de ce que l'on a appelé la fonction grammaticale plus haut. Il ajoute aussi à cette catégorie fonctionnelle les phénomènes de focalisation que nous présenterons dans la prochaine partie.

1.4 Prosodie et information : la fonction de focalisation

On entend par fonction de focalisation le fait que la prosodie permet de mettre en valeur le matériau verbal. Ceci revient donc à parler d'accentuation au niveau de la phrase et aussi de la place de l'accent tonique principal ou du nucléus. L'accent de mot (« word stress ») est présenté comme une caractéristique indépendante de l'intonation puisqu'il fait partie de la nature des

mots. Cependant, on avance aussi que l'intonation est portée par les syllabes accentuées d'une unité intonative, c'est ce qui a été appelé l'accent de phrase (« sentence stress ») ou fonction accentuelle. Si au niveau du mot, on a toujours une syllabe forte par opposition à des syllabes plus réduites, on retrouve aussi ce phénomène au niveau d'un énoncé qui contiendra une alternance de temps forts et de temps faibles, les premiers étant localisés au niveau des syllabes accentuées des mots concernés. En anglais, on dispose de deux mots différents pour faire référence à ces deux types d'accents : « stress » *vs* « accent ». Le placement de ces accents de phrase a un rôle linguistique et influence l'interprétation d'un énoncé. L'interprétation la plus neutre mène à accentuer les mots lexicaux ou à contenu par opposition aux mots grammaticaux tels que les propositions, articles et autres, comme on peut le voir dans l'exemple ci-dessous :

| *I'd <u>like</u> to <u>know</u> <u>where</u> he's <u>going</u>* |

Dans cet exemple, les mots lexicaux « *like, know, where et going* » sont mis en valeur et de ce fait c'est la syllabe accentuée de chacun de ces mots qui porte la proéminence. Le dernier de ces accents est généralement défini comme le nucléus dans une interprétation non contrastive.

À partir de l'étude de ces phénomènes accentuels au niveau de l'énoncé, une théorie du rôle de l'intonation dans la structuration de l'information a été développée.

Cette théorie a ses origines dans les travaux de l'école de Prague et en Angleterre dans les travaux de Halliday (1970) :

« In English information structure is expressed by intonation. » (p. 162)

Alors que les fonctions grammaticales ou types de phrases sont marquées par le choix des contours mélodiques, la structure de l'information dépendrait, elle, de la place des proéminences.

La place des proéminences et plus particulièrement du centre intonatif ou nucléus permet de distinguer les informations connues (« given information ») des informations nouvelles (« new information ») comme le montrent les exemples suivants tirés de Cruttenden (1986) :

I left on SAturday.

vs

I LEFT on Saturday.

Dans le premier exemple, le centre intonatif est placé sur la première syllabe de « *Saturday* ». On peut imaginer que l'on répond ici à la question « *When did you leave* ? »

Dans cet exemple, l'information nouvelle est représentée par « *Saturday* » car le reste était déjà contenu dans la question, donc connu. Dans le deuxième exemple on pourrait imaginer répondre à la question : « *Did you come back on Saturday* ?». Cela supposerait en effet que l'information

« *on Saturday* » soit déjà connue et que l'information nouvelle soit représentée par *« left* » qui contraste avec l'action « *come back* ».

On reconnaît aussi le rôle des frontières de groupes intonatifs dans la structuration de l'information comme le montrent les exemples suivants tirés de Crystal (1975) :

There's a MAN|| *in the GARden*||

vs

There's a MAN in the garden.||

vs

There's a MAN [in the GARden] || Crystal (1975, p. 27)

Le premier exemple contient deux unités d'informations par opposition au second qui en contient une seule. Dans le premier cas, on a deux centres intonatifs alors que dans le second il n'y en a qu'un. Enfin le dernier exemple a été analysé comme un énoncé contenant une unité d'information principale suivie d'une unité qui lui serait subordonnée.

Après avoir présenté les principales fonctions structurelles de la prosodie, nous nous intéresserons à la fonction expressive et à la fonction identificatrice.

1.5 Prosodie, attitude et émotion : la fonction expressive

La plupart des chercheurs s'accordent à dire que le rôle principal de la prosodie et plus particulièrement de l'intonation est de transmettre les émotions ou attitudes du locuteur au moment où il s'exprime. Un même énoncé peut être dit de multiples façons avec un ton « colérique », « ironique », « triste », « détaché » etc. Pike (1945) avance ainsi que la signification de l'intonation s'ajoute de façon temporaire au sens des mots qui est, lui, intrinsèque et stable :

> « *Rather than being a stable inherent part of words, it's a temporary addition to their basic form and meaning. Rather than being carried by permanent consonants and vowels, it's carried by a transitory extrinsic pitch contour. Rather than contributing to the intrinsic meaning of a word, it's merely a shade of meaning added to or superimposed upon that intrinsic lexical meaning, according to the attitude of the speaker […].* » Pike (1945, p. 21)

Par l'intermédiaire de l'intonation on peut donc exprimer des émotions, marquer notre attitude par rapport à ce que l'on avance. Ce n'est pas seulement ce que l'on dit, c'est-à-dire le choix des mots ou de la syntaxe qui permet d'interpréter un énoncé mais aussi la façon dont on le dit qui se retrouve au moins en partie au niveau prosodique. Prenons l'exemple suivant cité dans O'Connor & Arnold (1961).

On peut opposer deux façons de dire « *thank you* » en anglais britannique :

`*Thank you.* (HL ou contour descendant)

 Thank you. (LH contour montant)

Dans le premier cas, un contour descendant est utilisé. On peut imaginer la situation suivante : un ami très cher vous offre un cadeau très spécial et vous lui exprimez votre sincère gratitude.

Dans le second exemple, le mouvement de F0 est montant et la connotation tout à fait différente. Comme l'explique O'Connor, c'est une façon plus anodine de dire merci qui pourrait très bien s'appliquer à l'exemple d'un chauffeur de bus remerciant les passagers après qu'ils aient acheté leur ticket, chose considérée comme normale. Dans ce cas, on parlera d'attitude.

Certains vont plus loin dans l'analyse de la fonction expressive en distinguant deux niveaux d'expression : celui des émotions et celui des attitudes, comme l'avance Faure (1962) :

> *« Inner states need not be restricted exclusively to emotions which escape from the speaker uncontrollably, but that they may also be subject to will, reflection and various different motivations without necessarily involving emotive agitation. »* (p. 98)

Le terme émotion fait donc référence aux états émotionnels non contrôlés ou d'ordre purement physiologique par opposition au terme attitude qui désigne ici une expression consciente et volontaire ayant un objectif communicatif. C'est une distinction partagée par les psychologues tels que Scherer (1979) qui parlent d'émotion en termes de « mécanisme d'externalisation physiologique » par opposition à l'attitude qui est l'expression d'un « comportement communicatif en interaction sociale utilisé de façon stratégique » (p. 524).

1.6 Fonction identificatrice

Enfin, il a été prouvé qu'il est possible, par le biais de l'intonation, d'identifier les locuteurs en tant qu'individus ou en tant que membres d'un groupe social. C'est ce que l'on appelle la fonction d'indexation ou d'identification.

L'appartenance à la gent féminine ou masculine, par exemple, prédispose à utiliser certains contours intonatifs ou un registre particulier de F0 (Brend 1972). Certains types de contours intonatifs sont particuliers à une classe d'âge comme le « baby talk ». Les différences régionales sont identifiables dans l'intonation et l'on peut reconnaître les particularités de l'intonation des personnes exerçant un type de profession particulier (les prêtres, les commerçants).

Chaque individu peut avoir une intonation qui lui est spécifique et qui le rend reconnaissable par ses proches.

1.7 Conclusion

Les six aspects fonctionnels de la prosodie présentés ici ne sont pas exhaustifs et ne font pas non plus l'objet d'un consensus parmi les théoriciens de l'intonation. Non seulement, comme nous l'avons remarqué plus haut, il existe une grande diversité de termes fonctionnels appliqués à la prosodie, mais de plus les termes les plus largement utilisés ne font pas forcément dans chaque théorie exactement référence à la même signification. La fonction grammaticale de la prosodie peut être interprétée de différentes façons. Paul Tench (1996) fait par là référence au rôle de la prosodie dans la désambiguïsation syntaxique d'un énoncé. Peter Roach (1991) préfère à ce terme l'appellation de fonction de discours mettant en relation l'information prosodique avec les actes de langage. Les liens entre prosodie et pragmatique ne sont pas définis de manière consensuelle non plus.

D'autre part, les différentes fonctions évoquées ne sont pas totalement distinctes les unes des autres ou discrètes. Ainsi il est difficile de distinguer la fonction grammaticale de la fonction expressive d'une attitude comme l'avance Couper-Kuhlen (1986) :

> *« The fundamental mistake is the belief that grammatical and attitudinal functions are discrete and non-overlapping. »* (p.182)

Il serait caricatural d'essayer de distinguer des énoncés purement grammaticaux et d'autres exprimant essentiellement l'attitude du locuteur. Couper-Kuhlen suppose donc que ces deux aspects fonctionnels sont présents dans tous les énoncés :

> *« We communicate our attitude towards all utterance, even if this is to indicate as far as possible that we have no attitude. »*
>
> *« It is only reasonable to conclude that both grammatical and attitudinal factors are present in every utterance. »* (p. 182)

Enfin, on sait aussi que si la prosodie véhicule du sens et participe à l'interprétation d'un énoncé, elle n'est pas l'unique support de l'expression du sens, mais interagit avec le lexique et le contexte :

> *« It appears that the attitudinal role of intonation depends on pragmatic information. […] We're dealing with a three-way interaction between intonation, lexis and context. »* (p.183-184)

L'étude des fonctions de la prosodie se révèle donc complexe, car la prosodie peut véhiculer simultanément des significations diverses et certains signaux prosodiques ne sont pas signifiants par eux-mêmes et doivent se combiner avec d'autres signaux (verbaux par exemple) pour transmettre du sens (cf. Di Cristo, 2004).

Ainsi quand on étudie le rôle de la prosodie dans la communication orale, on se confronte aux problématiques suivantes :

- De quel type de fonction ou signification la prosodie est-elle responsable ?
- Comment ces fonctions sont-elles exprimées (fonctions/formes) ?
- Comment la prosodie interagit-elle avec les autres canaux de la communication orale que sont les mots (ou lexique), la syntaxe, le contexte ?

Dans ce travail de recherche, c'est principalement à la deuxième question présentée ici que nous tenterons de répondre. On cherchera donc à analyser et à décrire l'articulation entre formes et fonctions ou forme et contenu. Il s'agira alors de prendre comme point de départ une représentation fonctionnelle minimale de l'intonation et d'essayer de la mettre en relation avec des représentations formelles afin d'établir une grammaire intonative et des règles de dérivation de l'information fonctionnelle à partir de l'information formelle. Par conséquent, le chapitre suivant sera consacré à la présentation des principaux systèmes de représentation des formes prosodiques en anglais britannique à partir desquels le travail expérimental de cette thèse sera développé.

Chapitre 2

Représentation des formes prosodiques de l'anglais : les différents modèles

Les théoriciens de l'intonation anglaise se sont de tout temps attachés à décrire le système intonatif de l'anglais en identifiant ses composants ou formes aussi bien que ses différents apports sémantiques.

On distingue deux grands modèles rendant compte du système intonatif anglais : les théories de la « British School of Intonation » et le modèle autosegmental associé à une approche américaine de l'intonation. Ces deux modèles s'opposent du point de vue des descriptions de l'intonation qu'ils proposent et de leur analyse de la façon dont l'intonation contribue à l'interprétation des énoncés. On parle notamment de l'opposition entre la théorie des contours et celle des niveaux (« contours *vs* levels » ou « levels *vs* configurations » cf. Bolinger, 1951).

Nous consacrerons ce chapitre à la question des représentations formelles dans ces différents modèles. Il s'agira donc de passer en revue les contours ou patrons intonatifs décrits dans la littérature avant de s'intéresser à une problématique formelle plus récente qui est celle de l'alignement tonal. Nous présenterons dans un premier temps les principaux modèles britanniques de l'intonation de l'anglais pour ensuite présenter le modèle autosegmental métrique.

2.1 Les modèles britanniques des formes prosodiques

Cette section est fondée sur une revue de lecture des ouvrages des principaux théoriciens de l'intonation anglaise : Jones (1909 et 1972), Palmer (1924), Armstrong et Ward (1931), Jassem (1952), Kingdon (1958), Faure (1962), Halliday (1970), O'Connor & Arnold (1961 et 1973).

Ceux-ci ont proposé diverses études descriptives des formes intonatives de l'anglais aboutissant à des inventaires des formes existantes et de leurs usages. On distingue tout d'abord deux approches descriptives : l'approche globale ou « the whole tune approach » et l'approche des tons nucléaires ou « nuclear tones approach ». D'après ces deux approches, les unités susceptibles d'être significatives en anglais sont des mouvements de la voix comme les chutes ou les montées. Ces différents mouvements sont classés selon leur configuration (forme) et d'autre part selon leur rôle dans la construction prosodique d'un énoncé (fonction). Les méthodes employées pour établir ces différents inventaires et modèles sont essentiellement auditives, ce qui vaut à cette théorie d'être qualifiée d'impressionniste (cf. Ladd, 1996) par opposition aux méthodes d'approche instrumentales et acoustiques. Ces modèles, notamment celui proposé par O'Connor & Arnold, sont très utilisés encore aujourd'hui pour leurs qualités didactiques.

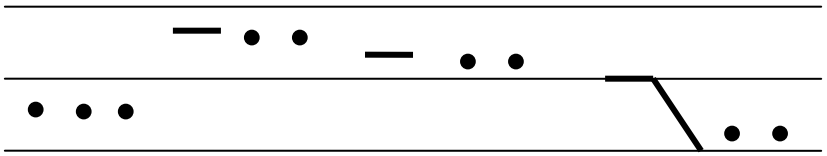
2.1.1 The « tune approach »

Cette approche de l'intonation anglaise est représentée par les travaux du professeur Daniel Jones (1972) ainsi que les travaux d'Armstrong et Ward (1931).

D'après ces chercheurs, l'intonation peut être représentée de manière efficace au moyen de deux contours intonatifs fondamentaux appelés contour 1 et contour 2 (« tune 1 and tune 2 »). Le premier fait référence à un contour mélodique descendant et le second à un contour globalement descendant mais se terminant par un contour ascendant.

Le terme « tune » correspond au contour mélodique qui s'applique à un énoncé ou une proposition entière comme le montrent les exemples suivants :

Tune 1



Tune 2

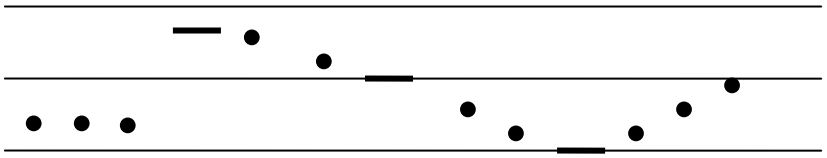


Figure 2 : Les deux contours intonatifs de base 1 et 2 suivant Jones (1972).

Jones propose une description du développement de ces contours globaux.

Le contour 1 peut comporter des syllabes inaccentuées en position initiale (représentées par des points ci-dessus). Celles-ci sont produites dans un registre de voix assez bas et gardent la même hauteur (le registre est représenté par les trois lignes horizontales). Les syllabes accentuées (représentées par les tirets) quant à elles, forment une séquence de notes descendantes, allant décroissant. Quand l'énoncé contient plus d'une syllabe accentuée, la chute commence à une hauteur quasiment égale à celle des syllabes initiales inaccentuées et se termine au niveau de la

limite inférieure de la tessiture du locuteur. La hauteur du début de la chute dépend du nombre de syllabes accentuées précédentes et de leur hauteur.

Le contour 2 se distingue du premier par son mouvement final qui est cette fois-ci une montée. Les syllabes accentuées et inaccentuées forment une séquence descendante jusqu'à la note la plus basse pour un locuteur donné. Les syllabes qui suivent ce point le plus bas forment une séquence montante.

À partir de ces deux patrons mélodiques, on peut établir des patrons divers permettant d'exprimer des significations différentes comme l'emphase. Jones (1972) distingue deux types d'emphase : l'emphase contrastive et l'emphase d'intensification. Le premier type d'emphase permet de mettre en opposition un mot avec un autre (que ce dernier soit sous-entendu ou mentionné précédemment). Jones ajoute que cela peut permettre d'introduire une idée nouvelle ou inattendue (cf. exemple ci-dessous). L'emphase d'intensification équivaut à l'ajout dans une phrase d'adverbes tels que « très, extrêmement, beaucoup de » (« very, extremely, a great deal of »).

Ainsi, un contour 1 emphatique peut se décliner de deux façons : s'il s'agit de l'expression d'un contraste on aura une séquence de syllabes accentuées montantes suivies d'une grande chute :

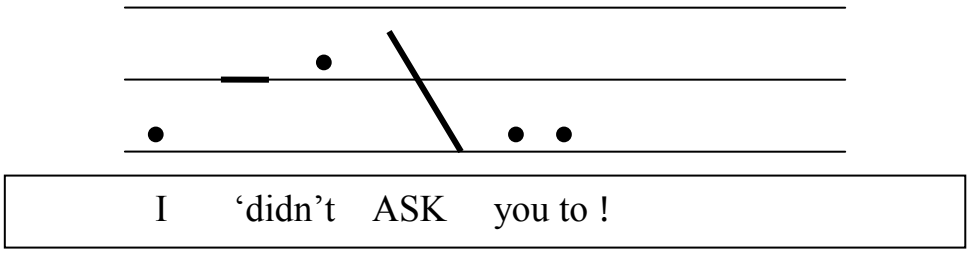


Figure 3 : Exemple de contour 1 emphatique extrait de Jones (1972, p. 298-299).

Dans le cas d'une emphase exprimant l'intensification, c'est la différence de hauteur qui la marquera :

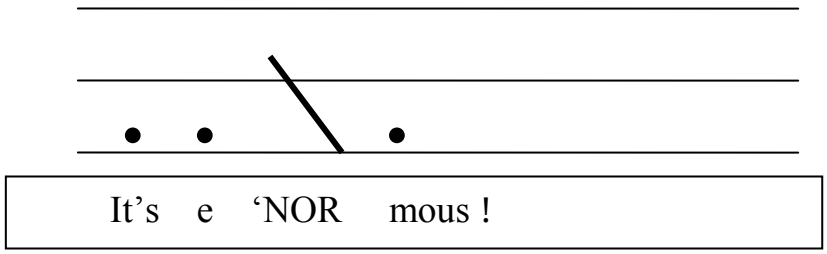


Figure 4 : Contour 1 avec une emphase d'intensification, extrait de Jones (1972, p. 309).

L'emphase transforme aussi le contour 2 en un contour mélodique complexe composé d'une chute et d'une montée finale :

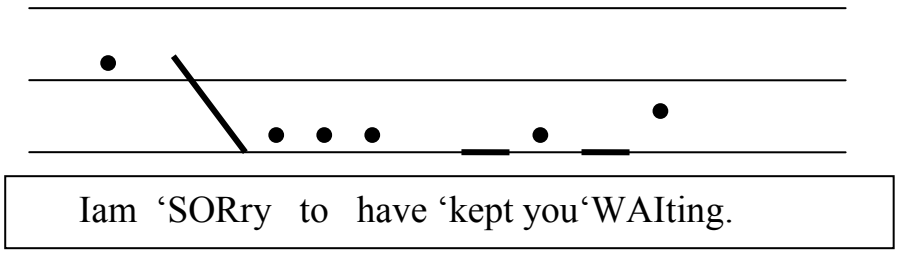


Figure 5 : Exemple de contour 2 emphatique complexe.

De plus, Jones mentionne l'existence de contours dits « spéciaux » qui sont à mettre en relation avec des types d'énoncés spécifiques comme les parenthèses ou les interjections pour n'en citer que deux exemples.

Une parenthèse produite à la fin d'un énoncé marqué par un contour 1 aura un contour intonatif grave et monotone présenté comme une continuation du ton grave atteint après la dernière syllabe accentuée :

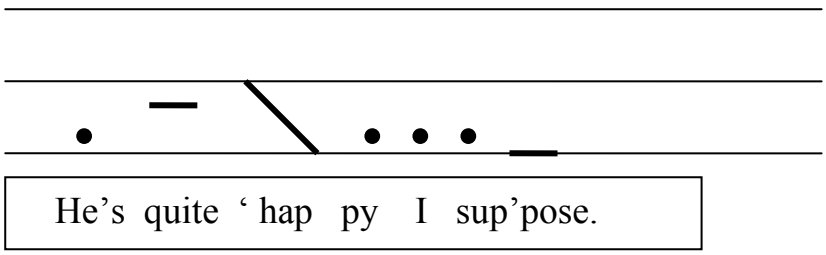


Figure 6 : Exemple de contour 1 avec une parenthèse basse.

Une parenthèse apparaissant à la suite d'un contour 2 sera marquée par un contour montant final :

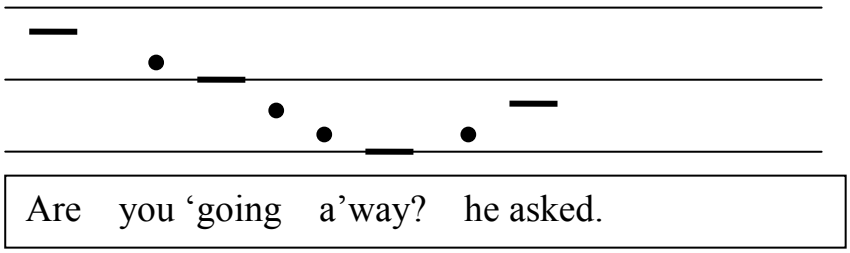


Figure 7 : Exemple de contour 2 avec une parenthèse finale, extrait de Jones (1972, p. 317).

Armstrong et Ward (1931) partagent aussi cette analyse de l'intonation d'un point de vue global. Ils développent une grammaire des contours mélodiques de l'anglais à partir des deux contours fondamentaux I et II et l'appliquent à l'enseignement de l'anglais comme langue seconde :

> *« They have shown, quite correctly in my opinion, that most sense-groups are said with one of two fundamental tunes or with other tunes which are formed by modifying the fundamental tunes according to definite principles. »* Extrait de Jones (1972, p. 279)

D'autre part, on identifie déjà, dans l'analyse de Jones, une division entre plusieurs éléments du contour mélodique : des syllabes inaccentuées en début ou fin d'énoncé, un corps accentué adoptant un mouvement cohérent et un mouvement mélodique final ne touchant que quelques syllabes. Ce sont ces différents éléments du contour mélodique qui sont à la base des représentations formelles développées dans l'approche des contours nucléaires.

2.1.2 The « English nuclear tones approach »

Depuis les travaux de Palmer (1924) lui-même inspiré par Klinghart et Klemm (1920), on définit la structure des patrons intonatifs comme suit : chaque contour intonatif peut être décomposé en trois parties : la tête, le pivot, la queue, ce qui correspond en anglais aux termes de « Head, Nucleus, Tail ». Ces éléments apparaissent au niveau d'un énoncé ou groupe de sens (« sense-group » chez Jones et O'Connor & Arnold, 1961). Un groupe de sens se définit, d'après Palmer, comme une suite de mots formant un énoncé ne comportant qu'un seul centre intonatif ou maximum de proéminence c'est-à-dire le nucléus ou pivot. Celui-ci est porté par la syllabe accentuée du mot le plus proéminent dans un groupe intonatif.

Toutes les syllabes suivant le nucléus dans un même groupe intonatif sont appelées queue ou « tail ». Toutes les syllabes précédant le nucléus dans un groupe intonatif sont appelées « tête » ou « head ».

Kingdon adapte cette terminologie dans son ouvrage en 1958. Le terme « tête » prend une définition plus réduite, il désigne la première syllabe accentuée dans un groupe intonatif : c'est la tonique de tête. Ensuite, il ajoute une catégorie appelée « pré-tête » ie « prehead » regroupant toutes les syllabes inaccentuées précédant la tonique de tête. Enfin, toutes les syllabes situées entre la tonique de tête et le nucléus sont appelées « corps » ie « body ». O'Connor & Arnold n'utilisent pas le terme « body » et le terme « tête» regroupe donc la tonique de tête et le corps. Si la décomposition du contour intonatif en différentes unités est partagée par la plupart des théoriciens de l'école britannique de l'intonation, la définition de chacune de ces unités et le domaine d'application de celles-ci sont donc variables, comme le montre l'exemple ci-dessous :

I didn't need to |*take*| *it.* (Palmer)

Head | N | Tail

I |*di* |*dn't need to* |*take*| *it.* (Kingdon)

PH| H| Body | N | Tail

I |*didn't need to* |*take*| *it.* (O'Connor & Arnold)

PH| Head |N | Tail

La définition de chaque élément dans la plupart de ces approches est fondée sur la distinction entre plusieurs niveaux de proéminence que Cruttenden (1986) résume de la façon suivante :

- L'accentuation primaire (« primary accent/stress ») : qui concerne l'accent ou schéma mélodique principal appelé le pivot ou noyau (« nucléus ») ;
- L'accentuation secondaire : comprenant les accents mélodiques non nucléaires (« Head » et/ou « body ») ;
- L'accentuation tertiaire : comprenant les proéminences locales essentiellement marquées par des variations au niveau de la durée, ou longueur d'une syllabe ou au niveau de l'intensité (« body ») ;
- Le niveau inaccentué : ensemble des mots inaccentués appartenant au groupe intonatif concerné (« prehead » / « tail »).

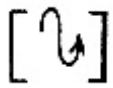
À partir de cette décomposition des contours intonatifs, les théoriciens de l'anglais ont développé des inventaires des formes prosodiques de l'anglais qui sont à nouveau très variables d'un auteur à l'autre. Prenons les modèles de Palmer (1924), O'Connor & Arnold (1961), Kingdon (1958) et Halliday (1970) comme exemple.

Palmer considère que l'on peut représenter le système intonatif anglais à l'aide de quatre types d'accents nucléaires :

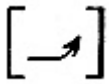
- Le ton descendant (« falling tone) ;

- La montée supérieure (« high rise ») ;

- La chute-montée (« fall-rise ») ;



- La montée inférieure (« low rise »).

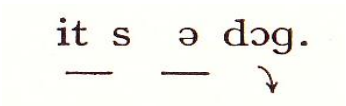


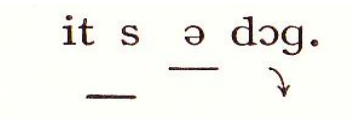
Tous ces contours possèdent aussi des formes variées quand une forme d'emphase est exprimée. Palmer parle alors de « chutes intensifiées » quand le contour descendant ou la chute est précédée d'une légère montée. Pour les autres contours, c'est le type de tête qui leur est associé qui permet d'exprimer une emphase.

Le terme tête désigne à la fois toutes les syllabes précédant le nucléus et le contour qu'elles portent. Il existe alors quatre types de « têtes » d'après Palmer :

- La tête inférieure (« inferior head ») : les syllabes de tête apparaissent dans un registre médium avant la chute ; elles ne sont jamais plus hautes que l'attaque du nucléus ;

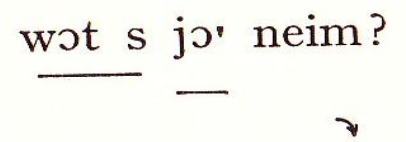
  It's a dog





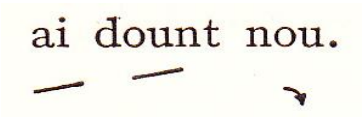
- La tête supérieure (« superior head ») : les syllabes de tête commencent dans un registre aigu et forment une séquence descendante jusqu'à la chute ; les syllabes sont plus hautes que l'attaque du nucléus ;

  What's your name ?



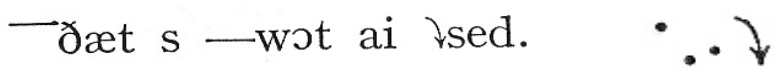
- La tête montante (« scandent head ») : elle contient une montée progressive des médiums vers les aigus pour atteindre la hauteur maximale qui sera le point de départ de la chute ;

  I don't know !



- La tête hétérogène (« heterogeneous head ») : c'est une combinaison des trois autres types définis.

That's what I said.



L'exploration des combinaisons possibles de ces différents accents nucléaires et toniques de tête au niveau de la phrase intonative aboutit à une grammaire intonative de l'anglais.

O'Connor & Arnold (1961), mentionnent six contours nucléaires :

- « Low fall » (chute inférieure) qui va du registre médium jusqu'aux graves ;

- « High fall » (grande chute) qui commence dans les aigus et se termine dans les graves ;

- « Rise fall » (contour montant-descendant) ;

- « Low rise » (montée inférieure) qui part du registre grave et s'étend jusqu'aux médiums ;

- « High rise » (montée supérieure) qui part du registre médiums et s'étend jusqu'aux aigus ;

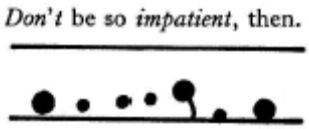
- « Fall-rise » (chute- montée).

Nous voyons que le contour montant-descendant est identique à la chute intensifiée de Palmer. De plus les deux inventaires contiennent la montée supérieure mais ce terme ne semble pas être utilisé de la même façon. La montée supérieure de Palmer est décrite comme un mouvement ascendant commençant dans la partie grave de la tessiture du locuteur et s'étendant jusqu'au registre aigu alors que la montée supérieure de O'Connor & Arnold va des médiums jusqu'aux aigus. Pour Palmer, c'est le point d'arrivée de la montée qui est distinctif puisque les petites et grandes montées commencent dans les graves alors que pour O'Connor & Arnold, c'est le point

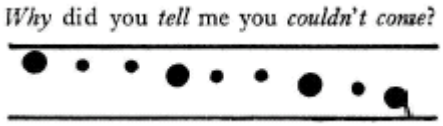
de départ dans les graves ou médiums qui est significatif. Enfin, seule la grande chute n'apparaît pas dans l'inventaire de Palmer.

Les contours prénucléaires sont classés ainsi :

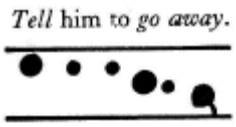
- La tête basse (« low head ») où toutes les syllabes sont prononcées dans un registre grave ;



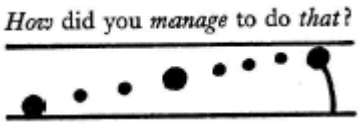
- La tête haute (« High head ») où toutes les syllabes sont prononcées dans un registre aigu ;



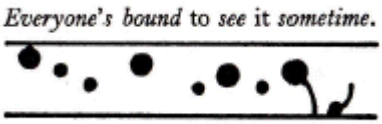
- La tête descendante (« falling head » ou « downstepping head») où les syllabes accentuées forment un contour descendant progressivement ;



- La tête montante (« rising head» ou « upstepping head »);



- La tête montante-descendante (« sliding head »).



Il semble que chaque type de tête puisse trouver un équivalent dans le système de Palmer. La tête basse appartiendrait à la classe des têtes inférieures, la tête haute à celle des têtes supérieures, la tête montante à celle des mêmes têtes montantes, la tête montante-descendante à celle des têtes hétérogènes. La tête descendante peut être associée à la classe des têtes supérieures car celles-ci peuvent décrire un contour descendant quand elles contiennent plusieurs syllabes.

Chacun de ces types de « tête » peut être précédé d'une « pré-tête » inférieure (dans le registre grave) ou supérieure (dans le registre aigu). Ce n'est pas la seule terminologie utilisée par O'Connor & Arnold pour décrire leur modèle des formes prosodiques de l'anglais. Les éditions

successives ont montré l'apparition notamment de termes de plus en plus imagés pour décrire les contours prénucléaires en particulier (cf. « *the jack knife* », « *the dive* », « *the take off* »).

Kingdon (1958) présente aussi son interprétation du système intonatif de l'anglais britannique basée sur la distinction entre les tons statiques (« static tones ») et les tons dynamiques (« kinetic tones ») . Les premiers sont associés aux syllabes inaccentuées ou seulement « partiellement accentuées » alors que les seconds sont associés aux syllabes pleinement accentuées. D'après Kingdon, on utilise les tons statiques pour mettre en valeur un mot donné sans y attacher un sentiment particulier par opposition à l'utilisation des tons dynamiques.

> *« The static tones are the level tones, accompanied by stress, which are used on the words to which it is desired to give prominence in the sentence, but to which no particular feeling is attached. »* Kingdon (1958, p. 4)

Il décrit les accents mélodiques prénucléaires et nucléaires avec une terminologie identique qui est très proche du système d'O'Connor & Arnold (1961) mais y ajoute aussi la possibilité d'avoir des tons statiques en position nucléaire. On a donc des tons :

- Statiques graves (« the low level tone »);
- Statiques aigus (« the high level tone »);
- Montants supérieurs (« high rising tone » identique à celui de O'Connor & Arnold) ;
- Montants inférieurs (« low rising tone ») ;
- Descendants : grands et petits correspondant à la grande chute et petite chute chez O'Connor & Arnold ;
- Descendants-montants (divisés ou non : on peut les distinguer en notant chute-montée pour un contour non divisé et chute+montée pour un contour divisé).

Le système de Kingdon permet de décrire le contour descendant-montant plus précisément. Il fait une distinction entre deux grands types de chutes-montées : la chute-montée entière (non divisée) et la chute-montée divisée ou chute + montée. Une chute-montée non divisée est réalisée sur un mot monosyllabique. Une chute-montée divisée peut être de trois sortes :

- Divisée si elle s'étend sur deux mots adjacents ;

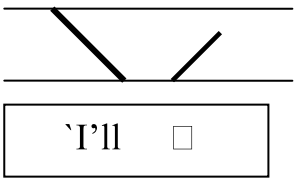


Figure 8 : Chute-montée divisée selon Kingdon (1958).

- Interrompue si les mots qui portent les tons principaux sont séparés par une à plusieurs syllabes inaccentuées ;

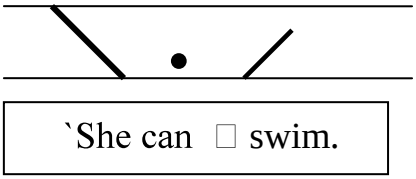


Figure 9 : Exemple de chute-montée interrompue par une syllabe inaccentuée selon Kingdon

- Interrompue par des accents partiels : parmi les mots séparant ceux qui portent les tons principaux, un à plusieurs, sont partiellement accentués (cette accentuation est réalisée par un ton statique grave).

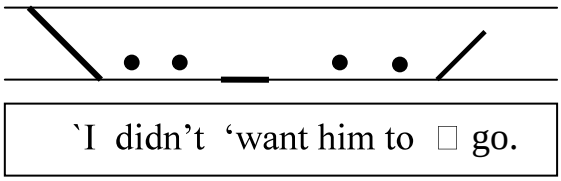


Figure 10 : Exemple de chute-montée interrompue par un accent partiel selon Kingdon.

La notion d'accent partiel permet aussi à Kingdon d'ajouter deux types de pré-têtes en plus des pré-têtes statiques graves ou aiguës. Quand un accent partiel est présent avant la tête, la pré-tête peut être ascendante ou descendante.

Enfin, le modèle de Halliday (1970) se distingue des modèles précédents par sa terminologie et une approche plus fonctionnelle de la description des contours intonatifs de l'anglais. Il ne parle pas de contours de tête, pré-tête et de nucléus mais plutôt de tons pré-toniques et toniques. Les tons ou contours mélodiques peuvent être de deux sortes : il y a les contours primaires et les contours secondaires.

Le ton ou contour primaire est l'élément essentiel de base d'un énoncé et est déterminé par le contour porté par la syllabe tonique ou nucléaire. De ce point de vue, le modèle de Halliday est identique aux modèles précédents. Il définit sept tons primaires (cinq étant simples et deux composés) :

- Le ton 1 : un ton descendant (« falling tone ») ;
- Le ton 2 : une montée supérieure ou un contour légèrement montant-descendant (« falling-rising, pointed ») ;
- Le ton 3 : la petite montée (« low rise ») ;
- Le ton 4 : descendant-montant correspondant à la chute-montée de Kingdon (« falling-rising, rounded ») ;

- Le ton 5 : montant-descendant (« rising-falling, rounded ») ;
- Le ton 13 qui est composé du ton 1 et du ton 3 : il est donc descendant puis montant, les deux mouvements pouvant être séparés par une à plusieurs syllabes ;
- Le ton 53 qui est donc montant-descendant-montant.

Les cinq premiers tons primaires sont identiques aux tons nucléaires représentés dans les approches précédentes. Le ton 13 correspond à la chute-montée divisée décrite par Kingdon et le ton complexe 53 contenant trois mouvements mélodiques est particulier à l'inventaire de base de Halliday. On peut aussi remarquer que Halliday (1970) ne prend en compte ni les tons statiques (cf. Kingdon) ni la grande chute (cf. O'Connor et Arnold) dans son inventaire des tons primaires.

Toutefois, la particularité du modèle de Halliday (1970) vient surtout de sa définition des contours secondaires. Ceux-ci ne sont pas simplement les contours associés aux syllabes non nucléaires c'est-à-dire à la tête ou la pré-tête. Ils peuvent en fait s'appliquer tout autant à la syllabe tonique qu'aux syllabes pré-toniques. Ce sont des « subdivisions » du contour primaire, déterminées par le mouvement de ce dernier. Alors que le ton primaire est l'élément essentiel de base pour un énoncé donné, le ton secondaire représente les différentes variétés, les différentes formes que peut prendre le mouvement primaire. Ainsi le ton 1 (ton descendant) peut être affecté des tons secondaires suivants :

- Le ton secondaire « grande chute » qui implique que la chute va des aigus vers les graves (« high to low ») ;
- Le ton secondaire « chute moyenne » : la chute va des médiums jusqu'aux graves ;
- Le ton secondaire « chute étroite » : la chute va des mediums-graves aux graves.

À partir du ton primaire descendant, trois types ou variétés de chutes peuvent donc être décrites dans le modèle de Halliday.

Les tons secondaires sont alors notés sous forme de diacritiques qui s'ajoutent au contour primaire. Au niveau des syllabes pré-toniques, ils permettent de décrire les équivalents des différents types de tête dans les approches précédentes. On obtient alors trois grands types de tons secondaires pré-toniques :

1. Le ton secondaire régulier (« even ») ou neutre qui est présent par défaut pour chaque énoncé. Il correspond aux contours de tête bas (« low head »), montant (« upstepping-head ») ou descendant (« downstepping- head ») ;
2. Le ton secondaire irrégulier (« uneven ») qui décrit un contour descendant-montant. (« falling-rising head ») ;

3. Le ton secondaire de type « liste » (ou « listing ») correspondant à une suite de petites montées.

Chaque énoncé contient donc au moins le ton secondaire prétonique « neutre » dans le modèle de Halliday. L'utilisation de tout autre ton secondaire est motivée par la volonté d'apporter une signification différente à l'énoncé :

> *«There is always one secondary tone in each set which is called neutral […] The other secondary tones are used when there is a definite motive, a particular meaning to be expressed; in other words when there is 'good reason'[…]»* Halliday (1970, p. 14)

Il ne s'agit pas là d'un compte-rendu exhaustif de tous les modèles des formes prosodiques de l'anglais et nous aurions pu décrire beaucoup d'autres modèles. Mais notre objectif est autre : il s'agit surtout de souligner les points communs entre ces différents modèles. Ainsi nous avons vu que les formes prosodiques sont classées et distinguées en fonction des niveaux de proéminence (« Prehead/ Head/ Nucleus/ Tail »). On distingue aussi les tons statiques des tons dynamiques. Les premiers étant souvent associés aux syllabes inaccentuées alors que les toniques principales et surtout le nucléus, sont typiquement décrits comme des tons dynamiques. Au niveau du contour nucléaire, certains théoriciens distinguent par exemple deux types de chutes, celles dont le point de départ se fait dans les aigus et celles dont le point de départ se fait dans les médiums (« low *vs* high fall / low rise *vs* high rise »). Enfin, toutes les descriptions formelles présentent des tons mélodiques simples (chutes, montées) et complexes (chutes-montées). La citation suivante résume ainsi la vision britannique des formes prosodiques :

> *« Intonation is either level, rising or falling. The level tone is not used much in speech. Besides these simple tones there are compound tones. (…) The level tone may be either low or high in pitch, and the others may begin in a high or low pitch. »* Sweet (1892)

Malgré le succès des approches britanniques de l'intonation, du point de vue de l'enseignement, celles-ci peuvent être remises en question au niveau de leurs fondations scientifiques. Comme on l'a montré dans cette partie, l'analyse de la structure des contours intonatifs ou bien les inventaires des formes prosodiques ne font pas l'objet d'un consensus entre les différents auteurs. On peut s'interroger sur les raisons pour lesquelles tel ou tel chercheur décide de réduire le système intonatif de l'anglais à cinq tons nucléaires plutôt qu'à six ou à huit etc. ? Il nous semble alors important de distinguer l'une de ces approches des tons nucléaires, qui est celle de Crystal (1969) car elle représente à nos yeux la première étude scientifique de l'intonation de l'anglais.

2.1.3 « Prosodic features »

L'approche développée à la suite des travaux de Crystal (1969) est aussi adoptée notamment par Quirk & *al.* (1972) et Couper-Kuhlen (1986). Celle-ci se détache de l'approche didactique évoquée dans la partie précédente :

> *« Despite its eminent suitability for didactic purposes, the tune approach to intonation imposes a rather rigid frame on the melodic line. »*
> Couper-Kuhlen (1986)

On passe ici d'une approche souvent dite « auditive » et « impressionniste » à une approche expérimentale et scientifique. En effet pour la première fois, un grand corpus d'anglais est utilisé pour mener une recherche sur l'intonation de l'anglais. Crystal se distingue des autres théoriciens de l'approche des contours nucléaires bien qu'il en adopte la représentation formelle de base. Son unité tonale est composée des mêmes éléments que ses prédécesseurs :

« TU (Tone Unit) = (PH) (H) Nucleus (Tail) »

(Seul le nucléus est essentiel comme l'indiquent les parenthèses autour des autres éléments).

Il mentionne toutefois les incohérences des systèmes prosodiques précédents tels que celui d'O'Connor & Arnold dans lequel « grande chute » signifie que la chute commence dans le registre aigu et s'étend jusqu'au registre grave du locuteur alors que la grande montée commence dans les médiums et s'étend jusqu'aux aigus.

De façon générale, il constate qu'il y a un accord sur l'existence des petites chutes, des petites et grandes montées, de la chute-montée et de la montée-chute, mais beaucoup plus de problèmes quand il s'agit de la grande chute, du ton statique, ou des tons complexes comme les chutes-montées divisées.

Le système de Crystal est basé sur les contrastes trouvés dans les données et prend en compte tous les contours présentés auparavant pour en faire une synthèse et les enrichir. Il postule l'existence de trois types de tons nucléaires : les tons nucléaires simples, composés et complexes.

2.1.3.1 Les tons nucléaires simples

Les tons nucléaires simples n'ont qu'une seule direction : ils sont marqués par une chute ou une montée sous la forme d'un « glide » et s'il y a des syllabes de queue, celles-ci sont maintenues dans la continuité du mouvement initié par le nucléus.

Crystal (1969) rejette la distinction entre « grande et petite chute ». Il pense en effet que cette distinction est très réductrice au regard des variations de registre multiples qu'il observe dans ses données. Il n'est pas utile de distinguer tous les niveaux de hauteur possibles non plus, mais Crystal donne au registre une importance toute particulière en le traitant comme un système

prosodique spécifique qui peut s'appliquer aux contours nucléaires mais aussi à tous les autres points de l'énoncé.

Il développe alors un système de diacritiques permettant de qualifier les variations de registre : il annote la hauteur du point de départ des mouvements mélodiques ainsi que leur amplitude. Tout d'abord le point de départ du mouvement peut être dans un registre aigu (« High »), médium ou grave. Ensuite le mouvement peut être ample ou étroit ou normal (« width = wide, narrow, normal »). Couper-Kuhlen (1986) ajoute aussi que la pente de la chute ou de la montée peut être forte (« steep »), progressive (« gradual ») ou plate (« flattened »). Enfin le mouvement peut se terminer dans un registre grave ou non grave (« low *vs* non low »). Le système de description des contours intonatifs est représenté de la façon suivante dans Crystal (1969) :

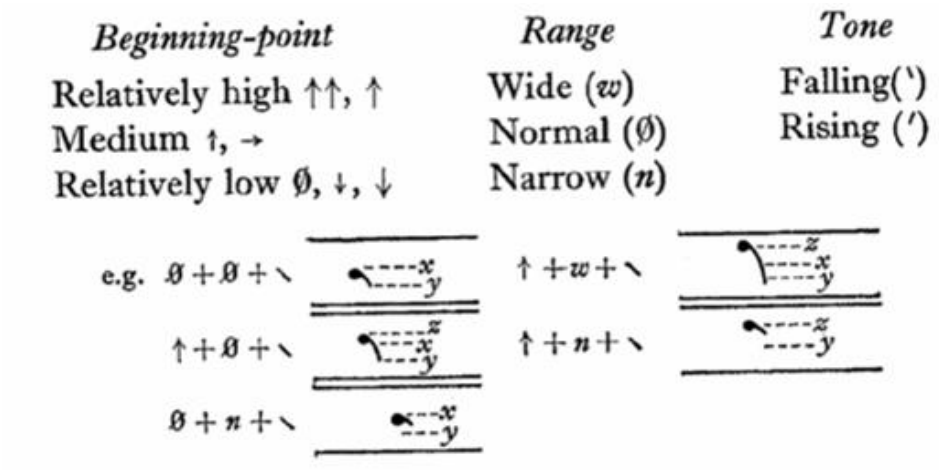


Figure 11 : Système de diacritiques utilisé pour décrire les contours intonatifs selon Crystal(1969, p.213).

Le modèle de Crystal donne un statut particulier aux tons statiques. Dans la plupart des modèles britanniques, les tons statiques ne sont pas considérés comme des accents potentiellement nucléaires. On estime en général que lorsqu'un ton statique est utilisé à la place d'un ton dynamique, on perçoit un énoncé interrompu ou inachevé. Or Bolinger (1945), mentionne des exemples d'utilisation des tons statiques en position nucléaire qui ne s'accompagnent pas d'un effet inachevé mais qui ajoutent plutôt une connotation d'ennui, de monotonie. Crystal confirme cet argument par de nombreux exemples tirés de son corpus, dans lesquels on perçoit une frontière d'unité intonative mais pas de contour dynamique avant celle-ci. Crystal donne des exemples dans lesquels le ton statique aigu est utilisé comme une version marquée de la petite montée (p. 214). On trouve aussi des cas où le ton statique fonctionne comme l'équivalent d'une chute. Ceci laisse à penser que les tons statiques ont un statut fonctionnel particulier et ont donc leur place dans un système représentatif des tons nucléaires de l'anglais. Le ton nucléaire chez Crystal ne se limite donc pas aux tons dynamiques et permet de se détacher de la définition purement dynamique du nucléus :

> Trim (1959, p. 28)*: [If levels are to be allowed as nuclear] "then the usual definition of the nucleus in kinetic terms would have to be abandoned".* Crystal (1969, p. 215)

2.1.3.2 Les tons nucléaires complexes et composés

Les tons nucléaires complexes sont marqués par un changement de direction de F0 au niveau de la syllabe avec un seul pic de proéminence. Ce sont les chutes-montées et les montées-chutes. Le premier élément est toujours le plus proéminent d'un point de vue phonétique.

Les tons composés se distinguent des tons complexes car ils contiennent deux proéminences. On ne pourrait pas supprimer le deuxième élément qui est essentiel pour l'interprétation de l'énoncé. La distribution phonétique des tons mélodiques est bien distincte dans les deux cas et l'on peut trouver des exemples où, l'utilisation de l'un ou de l'autre modifie le sens comme dans le cas d'une chute-montée non divisée opposée à une chute-montée divisée :

> | *vYou don't 'know*| *vs* | *`You don't □ know* |
>
> Extrait de Crystal (1969, p. 220)

Dans le premier cas, il y a une seule proéminence mélodique principale et l'on peut interpréter l'énoncé comme il suit : « si toi tu ne sais pas qui le sait ? ». Dans le second exemple, on observe deux proéminences mélodiques (la première sur « You » et la seconde sur « know »), on peut interpréter l'énoncé de la façon suivante : « Tu ne sais pas » (alors pourquoi dis-tu que tu sais ?).

On observe donc huit contours mélodiques dans le modèle de Crystal :

1. La chute (`) ;
2. La montée (□) ;
3. La chute-montée (v) ;
4. Le contour montant-descendant (^) ;
5. La chute + montée + chute ;
6. La montée + chute + montée ;
7. La chute + montée où les deux mouvements sont séparés par des syllabes atones dans les graves (`…□) ;
8. La montée + chute.

2.1.3.3 Les contours de tête

Crystal propose aussi une révision des classifications des contours de tête à partir de l'étude de données expérimentales. D'après Crystal, les descriptions des types de têtes trouvées dans la littérature sont très réductrices. Ce que l'on appelle la norme (cf. « stepping head ») n'est pas vraiment la norme en termes fréquentiels puisque ce type de tête n'apparaît que dans 30% des

cas. Ce contour de tête « idéal » serait essentiellement caractéristique de l'anglais écrit lu à haute voix ou des énoncés isolés utilisés pour l'enseignement de l'anglais :

> *« The ideal stepping head may well be characteristic of some kinds of written English being read aloud, or of set of examples in a pedagogical context […]. »* Crystal (1969, p. 232)

Il faudrait donner plus d'importance aux autres types de tête comme la tête montante ou les têtes contenant plusieurs schémas telle que la tête montante-descendante. Il y a aussi des types de tête plus susceptibles d'être trouvés avec un type de ton nucléaire particulier, les combinaisons les plus courantes toutefois varient selon le style de parole (conversation, lecture…). Plus la fin de la tête est haute, moins on a de chance d'avoir un ton nucléaire montant après, pour ne citer qu'un exemple.

Crystal définit donc quatre types de têtes :

1. La tête tombante (« falling head ») ;
2. La tête ascendante (« rising head ») ;
3. La tête montante-descendante (« rising falling head ») ;
4. La tête descendante-montante (« falling-rising head ») .

2.1.3.4 Les contours de pré-tête

Enfin les contours de pré-tête sont aussi étudiés : Crystal observe que ces contours peuvent s'étendre sur un maximum de cinq syllabes équivalent à quatre mots pour les plus longs. Toutes ces syllabes sont inaccentuées sauf quand un nom, un verbe ou un adjectif apparaît dans la pré-tête. On parle alors de la présence d'un léger accent faisant ressortir une syllabe par rapport aux autres tout en maintenant un niveau mélodique identique :

> *« It's then pronounced with a slight inherent stress, so that it's louder than the surrounding unstressed syllables, but it remains on the same pitch-level as these, i.e, it's not pitch prominent. »* Crystal (1969, p. 234)

Il définit cinq contours de pré-tête :

- La norme : un niveau de hauteur légèrement inférieur à la première syllabe tonique (onset) : les syllabes peuvent être statiques ou décrire une montée progressive vers la tête ;
- La pré-tête haute : les syllabes sont plus hautes que l'attaque de la tête ;
- La pré-tête très haute (« extra high ») ;
- La pré-tête moyenne qui a le même niveau de hauteur que la tête ;.
- La pré-tête très grave (« extra-low prehead »).

Le type de pré-tête dépend du registre d'attaque de la tête. Sa structure grammaticale est prévisible : ce sont essentiellement des mots grammaticaux.

Cette approche semble donner une représentation plus complète et précise des formes prosodiques de l'anglais. Elle s'accompagne d'une description phonétique très fine des contours mélodiques. Au moyen du système de description du registre (cf. figure ci-dessus), on peut décrire quinze types de chutes différents. Cependant on pourrait reprocher à ce modèle d'être trop détaillé et de surgénérer des contours qui ne sont pas toujours distinctifs. Dans l'optique de créer une grammaire de l'intonation mettant en relation les différentes formes et les fonctions, est-il possible d'utiliser une description aussi fine au niveau acoustique ? est-ce même pertinent au niveau perceptif ?

2.1.4 Problèmes non résolus dans les approches britanniques de l'intonation

Nous avons présenté dans ce premier chapitre, les deux principaux modèles britanniques des formes prosodiques. Dans l'optique d'une utilisation de ces modèles comme un point de départ du travail expérimental de ce mémoire, nous aborderons quelques problématiques qui nous semblent non résolues dans ces approches. Certaines d'entre elles ont d'ailleurs poussé les chercheurs à se tourner vers une autre approche des formes prosodiques que l'on présentera dans la seconde partie de ce chapitre.

Tout d'abord, on a déjà vu que l'approche britannique n'est pas consensuelle et uniforme, mais donne lieu au contraire à plusieurs interprétations du système intonatif de l'anglais. Que l'on se place du point de vue de la structure prosodique (définition des éléments prosodiques : « head, body, nucleus ») ou bien du point de vue des inventaires des formes, il semble complexe d'aboutir à un système de représentation unique. Les différents inventaires ont pu être qualifiés trop minimalistes ou bien trop complexes (comme le système de Crystal). On peut se poser la question des critères et modalités de description des contours mélodiques. Faut-il donner de l'importance au point de départ des mouvements mélodiques, à leur amplitude, à la valeur de F0 atteinte en fin de mouvement ?

Nous pensons que l'une des façons de valider ou invalider ces systèmes est de les confronter à l'étude de grands corpus de parole d'une part et d'autre part d'adopter une approche fonctionnelle. Nous partageons ainsi le point de vue de Cruttenden (1986) qui avance que l'on devrait se limiter à décrire les contours intonatifs exprimant des significations différentes de façon évidente.

> *« […] in practice, we limit ourselves to describing just those variations which most obviously carry major differences in meaning. »* Cruttenden (1986, p. 50)

De plus, les modèles britanniques sont basés sur des contraintes théoriques fortes qui font l'objet de multiples critiques. L'une d'entre elles est liée à la place du nucléus. Celui-ci est

l'élément central et essentiel de ces approches. On a déjà vu qu'il était défini comme le dernier accent mélodique d'un énoncé. Cela signifie pour ces chercheurs qu'il ne peut y avoir d'autres accents mélodiques après celui-ci. Les seuls accents pouvant apparaître après le nucléus sont alors de nature rythmique, c'est-à-dire qu'ils n'impliquent pas de proéminence mélodique mais plutôt un allongement ou une augmentation de l'intensité. Les accents prénucléaires sont alors fondamentalement distincts des accents postnucléaires d'après les théoriciens de l'école britannique de l'intonation. Cette théorie devient problématique quand on se penche sur le cas de la chute-montée interrompue. Ce contour est décrit comme l'association d'une chute sur la syllabe nucléaire, d'une série de syllabes partiellement accentuées ou non, prononcées dans un registre grave et d'une montée mélodique finale à partir du dernier accent. Ce cas a amené certains auteurs à s'interroger sur le statut accentuel de ce dernier mouvement mélodique et par là même du statut du nucléus.

Prenons l'exemple suivant (tiré de Crystal, 1969) :

> | *Iam* □ *SOrry about the* □ *BOOkcase* | (où on a une chute sur sorry et une montée sur bookcase)

On peut trouver au moins six analyses différentes de ce cas dans la littérature :

1. Le nucléus est sur « *sorry* » et le reste est le « tail ». L'accent sur « *book* » est considéré comme rythmique (interprétation classique) ;
2. Il y a une proéminence non nucléaire sur « *sorry* » et le nucléus est porté par « *bookcase* » (les indices acoustiques montrent pourtant que la plus forte proéminence est portée par « *sorry* ») ;
3. Crystal décrit cette séquence comme un « compound tone » (1969, p. 218-220). O'Connor & Arnold l'appellent « compound tune » (1973, p. 28). Il y a donc deux proéminences, mais l'une reste plus importante que l'autre ;
4. Halliday (1967, p. 13-18, 1970, p. 12) propose de considérer un groupe intonatif contenant deux nucléus (« double-tonic tone group » *ie* « a single tone-unit with two nuclei ») ;
5. Brazil rejette l'analyse d'Halliday et affirme que s'il y a deux nucléus, il y a par conséquent deux groupes intonatifs ;

Hirst (1987, p. 236-237), interprète cet énoncé à l'aide d'une structure prosodique récursive. Les contradictions relevées dans les analyses précédentes viendraient du fait que l'on se place dans un cadre purement linéaire où l'on a la structure suivante : « | *I am °sorry a'bout the °bookcase*+|| ». Hirst réconcilie Halliday et Brazil en considérant une unité intonative supérieure au niveau hiérarchique dans laquelle s'enchâssent deux autres sous-unités intonatives. Cela permet d'expliquer la structure des énoncés avec parenthèses et adverbes. On a une frontière d'unité intonative après « *sorry* », mais l'énoncé en entier est aussi analysé comme une seule unité intonative à un niveau hiérarchique supérieur.

Ci-dessous, un exemple de ce type d'analyse pour la phrase « *I saw George the other day* » :

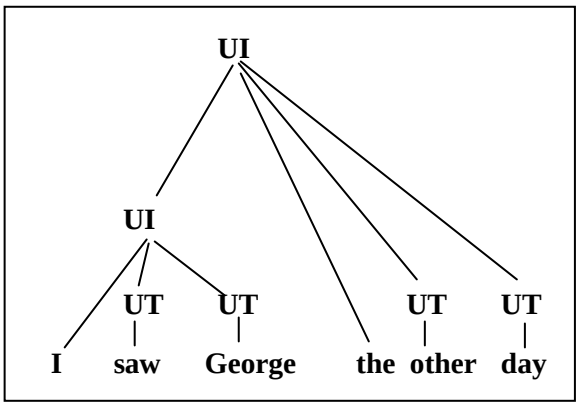


Figure 12 : Exemple de structure récursive dans le modèle de Hirst (1987).

Nous ne proposons pas de solution particulière à ce problème à ce stade, mais le considérons plutôt comme une question ouverte nécessitant une exploration expérimentale. La chute-montée interrompue ou « chute + montée » n'est pas le seul cas problématique relevé à partir des théories britanniques. Cruttenden (1986) mentionne aussi la possibilité d'une chute mélodique de petite ampleur après la chute principale portée par le nucléus (cf. exemple p. 42). Dans ces cas, la chute postnucléaire est considérée comme légèrement plus proéminente qu'un accent rythmique mais reste moins proéminente que la chute principale. Cruttenden fait d'ailleurs la même analyse du cas de la chute-montée interrompue et remarque que les contours montants sont généralement moins proéminents que les contours descendants. Il s'appuie sur des tests de perception pour prouver que c'est la chute qui a précédence sur le mouvement final qui est minimisé par la présence de celle-ci. Il en conclut aussi que le concept du nucléus devient moins problématique si on le définit comme l'accent mélodique le plus proéminent d'un énoncé (qu'il en soit le dernier ou pas) :

> *« […] I shall consistently use the notion of nucleus to describe that pitch accent (usually the last) which generally stands out as the most prominent in each of the typical tonal sequences within intonation-groups. »* Cruttenden (1986, p.43)

La question de la nature du nucléus mène aussi à remettre en question la distinction chez les théoriciens britanniques entre les accents de type rythmique et les accents mélodiques. Martine Grice (1995) qui a tenté d'appliquer les modèles formels britanniques à l'étude de l'italien de Palerme, s'oppose à cette distinction qu'elle juge vague et imprécise. Elle s'intéresse tout particulièrement à l'interdépendance entre le nucléus et la queue dans les théories britanniques. Elle avance tout comme Couper-Kuhlen, que l'on n'accorde pas assez d'importance aux syllabes postnucléaires qui sont considérées comme dépendantes mélodiquement du nucléus à partir duquel leur contour pourrait être prédit alors que le contour nucléaire peut s'étendre sur les syllabes de queue.

> *« The tail […] has long been considered the least significant part of the tone-unit, because the pitch contour of its syllables is said to be determined by the type of nucleus to which it is attached […] However this ignores the fact that nuclei - unless they are monosyllabic - are more often than not spread over the tail or part of it, so that the tail actually contributes to the shape and identification of the nucleus. »* Couper-Kuhlen (1986, p. 86)

Grice remarque aussi que le terme nucléus n'est pas employé de façon cohérente puisqu'il désigne parfois la syllabe nucléaire, le mot qui est le nucléus, ou bien le contour porté par celui-ci. Elle propose alors de modifier le domaine du nucléus en créant l'unité nucléaire. Elle divise le contour en trois parties principales : la pré-tête, la tête et l'unité nucléaire qui comprend la syllabe nucléaire ainsi que toutes les syllabes de queue qui suivent jusqu'à la fin de l'énoncé :

> *« However, the analysis is simplified if we see the nuclear tone as an abstract pitch movement which must be synchronized with a segmental layer made up of both the nuclear syllable and the tail. […] Both the nuclear syllable and the tail is determined by the direction of the nuclear tone. »* Extrait de Grice (1995, p. 19)

À l'issue d'une tentative d'analyse de l'intonation des contours interrogatifs dans l'italien de Palerme, Grice conclut toutefois que l'approche britannique ne suffit pas à rendre compte de la spécificité de l'italien de Palerme et se tourne vers l'approche autosegmentale qui fait l'objet de la prochaine partie de ce chapitre.

2.2 Le modèle autosegmental

Par opposition au modèle précédent, ce ne sont pas les contours mélodiques qui sont ici les bases d'une description des patrons intonatifs anglais mais plutôt une série de « tons mélodiques ». On ne tient plus compte ici des mouvements et configurations décrits par la fréquence fondamentale mais plutôt des cibles tonales atteintes par celle-ci à un moment donné. Les mouvements sont considérés comme des transitions vers des cibles qui seront significatives.

2.2.1 Origines et fondements de la théorie autosegmentale

L'analyse traditionnelle des systèmes des langues à tons, qui a été entreprise par certains chercheurs tels que Woo (1944), Leben (1976), Goldsmith (1976 et 1979), est à l'origine de la genèse de la phonologie autosegmentale.

Grâce à ses études sur les langues africaines, Goldsmith (1976) apparaît comme un précurseur :

Les tons ne sont plus représentés sous la forme de diacritiques associés à des segments mais comme des unités autonomes situées sur une rangée[1] particulière (cf. Di Cristo 2004). Cette ligne des tons est parallèle à celle des mots et syllabes ou segments.

> *« Tones are represented on a separate tier - the tonal tier- and on this tonal tier each segment is specified for tone and nothing else. »* Goldsmith (1979)

Goldsmith définit le terme « tier » comme une suite de mouvements ou gestes d'un point de vue acoustique. Un segment est une unité indivisible constituant une unité minimale dans cette représentation phonologique. La rangée tonale représente les mouvements du larynx en relation avec les mots, la rangée non tonale représente les mouvements de la voix.

Il existe des lignes d'association entre ces différents niveaux, d'un point de vue purement acoustique, elles sont représentées par la simultanéité dans le temps ou la coarticulation *ie* « co-registration ». Cette théorie comporte donc une contrainte de linéarité : la rangée tonale et la rangée segmentale sont synchronisées. On passe ainsi progressivement d'une approche holistique (contours) à une approche compositionelle (niveaux) de l'intonation pour reprendre les termes utilisés par Bolinger (1958). Il avance que les possibilités d'associations multiples entre la rangée tonale et la rangée segmentale permettent de traiter les chutes et montées comme des séquences de tons hauts et bas tout comme Goldsmith (1979) :

> *« The possibility of many - to - one associations between one tier and another opens up the possibility of treating rising and falling tones as sequences of level tones (high, low, mid) associated with a single vowel. »* Goldsmith (1979)

On s'intéressera dans la section suivante aux développements de cette approche dans le cadre de l'étude de l'anglais américain.

2.2.2 Les modèles standards américains

2.2.2.1 Les structuralistes américains

L'analyse des phénomènes prosodiques en termes de niveaux et non de contours dans l'étude de l'anglais américain, remonte aux travaux de Pike (1945), Wells (1945), Trager et Smith (1951) et Liberman (1978).

---

[1] Nous employons le terme « rangée » pour traduire le terme anglais « tier » en suivant l'exemple de H. Sil dans sa thèse de doctorat récente (septembre 2009). Nous préférons ce terme à celui de « tire » imité de l'anglais et utilisé par François Dell et Jacques Durand

Selon ces descriptions, le système intonatif de l'anglais peut être défini grâce à quatre niveaux mélodiques :

- « Low » (1) ;
- « Mid » (2) ;
- « High » (3) ;
- « Extra high » (4).

À ces niveaux mélodiques sont ajoutés trois types de frontières (« terminal junctures »). Celles-ci correspondent à la forme ou direction prise par la fréquence fondamentale à la fin d'un énoncé c'est-à-dire à partir de la dernière syllabe (cf. Cruttenden, 1986). On peut donc avoir une configuration finale descendante (#), montante (//) ou restant à niveau (/). On obtient l'illustration suivante extraite de Cruttenden (1986, p. 38) :

3 Usually // 2 John goes to 3 London #

Cet exemple contient deux groupes intonatifs. La première syllabe du mot « *usually* » est produite dans un registre aigu (« 3 ») et ce mot est suivi par une pause marquée par un contour montant (« // »). Le second groupe intonatif commence dans un registre médium puisqu'il est coordonné à l'énoncé précédent et se termine par une grande chute sur la première syllabe de « *London* » que l'on reconnaît par la combinaison du niveau aigu « 3 » au niveau de l'attaque et par la frontière finale descendante « # ».

Liberman (1978) présente dans sa thèse de doctorat, son interprétation du système intonatif de l'anglais en termes de niveaux en se distinguant de ses prédécesseurs. Il reproche aux approches précédentes de manquer d'abstraction et de ne pas rendre compte des relations complexes existant entre les contours mélodiques, l'accent lexical et le phrasé d'une part et l'interaction de ces trois éléments avec la grammaire d'autre part :

> *« The task of a theory of intonation is defined as the representation of stress, tune and phrasing, and the explication of their interactions with each other and with the rest of the grammatical system. »* Liberman (1978, p. 1)

La mise en relation des contours mélodiques et des accents lexicaux se fait au travers d'une grille métrique qui permet d'interpréter le matériel verbal en termes de syllabes fortes et faibles (« strong and weak »). Les étiquettes « fortes » et « faibles » ne font pas référence à un degré de proéminence absolu et ne donnent pas lieu à une interprétation phonétique directe. Il s'agit d'une annotation de la structure abstraite des énoncés :

> *« The association of text (toneless linguistic material) and tune is accomplished by a metrical system, which assigns metrical patterns to text and tune, establishes congruence between these patterns in any given case, and specifies possible alignments of the congruent patterns with a metrical grid. Metrical patterns are*

viewed as abstract structuring of complex events; metrical grids as abstract structuring of time. » Liberman (1978, p. 1)

Le modèle de représentation des formes prosodiques de Liberman est basé sur des formes sous-jacentes (« underlying form ») qu'il met en relation avec des formes de surface qui en sont l'interprétation phonétique.

Il définit alors quatre « phonèmes prosodiques » (cf. Liberman, 1978, p. 86) :

1. High (H) [+High,-Low] ;
2. High-mid (H-M) [+High, +low] ;
3. Low-mid (L-M) [-low, -High] ;
4. Low (L) [-High, +low].

Ces quatre niveaux sont définis de façon abstraite ou phonologique. Liberman précise que leur réalisation phonétique est influencée par les phénomènes d'expansion ou compression du registre. À partir de ces phonèmes prosodiques, il décrit par exemple les contours intonatifs associés à la surprise ou la redondance ainsi que le contour vocatif (« calling tune »).

La forme sous-jacente du contour de la surprise est la suivante :

W	S	W
[-High, +low]	[+High]	[-High, +low]
T1	T2	T3

Extrait de Liberman (1978, p. 101) où W = Weak, S = Strong et T = Tone

Et peut permettre de générer les formes de surface suivantes :

I gave it to Samuel.

L H-M L-M

I gave it to Samuel.

L H-M L

Dans le premier cas, on observe une montée de F0 suivie d'une chute finale atteignant un registre médium (Low-mid) alors que dans le second exemple la chute finale atteint le registre des graves. Ces deux contours appartiennent à la même famille phonologique dans les termes utilisés par Liberman (1978) mais leur réalisation peut être variée et induire des nuances de sens subtiles.

On peut générer à partir de la même structure de base la famille des contours vocatifs comme le montre l'exemple suivant :

What a fantastic baseball game!

L H-M H-M

Extrait de Liberman (1978, p. 105)

Dans ce dernier cas la chute finale n'atteint pas le registre le plus grave ce qui, d'après Liberman, laisse percevoir un certain inachèvement par opposition aux chutes se terminant par un ton « L » ayant un caractère plus terminal.

Ces systèmes ont été lourdement critiqués, car on considérait que le choix de quatre niveaux pouvait être à l'épreuve de l'étude d'un corpus soit trop riche, soit trop limité. Bolinger (1951) affirme que l'on peut ainsi prédire plus de contraste qu'il ne paraît en exister et que certains contrastes ne sont au contraire pas pris en considération. Les variations de F0 sont ainsi parfois confondues avec de simples variations de registre.

Ces problèmes ont été résolus par l'avènement d'une théorie de l'intonation basée sur deux niveaux distinctifs « Haut » et « Bas » (« High and Low »). Celle-ci s'est largement inspirée d'ailleurs des travaux de Liberman (1978) et de ceux de Gösta Bruce (1977).

2.2.2.2 Le modèle de Pierrehumbert (1980)

Le modèle de l'intonation de l'anglais développé par Pierrehumbert dans sa thèse en 1980, est donc basé sur l'opposition entre deux tons : « High and Low », H et L. La mélodie est interprétée comme une suite d'évènements discrets et les transitions qui les relient prennent un rôle secondaire. L'auteur propose un modèle de représentation phonologique de l'intonation de l'anglais et fournit les règles qui permettent de dériver, à partir de cette représentation abstraite, une représentation phonétique.

Son modèle est composé de trois volets principaux : d'abord une grammaire des contours intonatifs bien formés au niveau de l'unité intonative, ensuite une représentation phonologique du texte de l'énoncé sous forme de grilles métriques identiques à celles de Liberman et Prince (1977), enfin une série de règles qui permettent d'aligner les contours et le texte sur la base de la structure du contour et de la représentation métrique du texte.

La grammaire des contours intonatifs de Pierrehumbert (1980) repose sur trois catégories de « morphèmes » : les accents mélodiques (« pitch accents »), les accents de syntagme (« phrase accents ») et les accents de frontière (« boundary tones »).

Un contour bien formé est constitué d'une séquence de tons H et L. Ils représentent un à plusieurs accents mélodiques qui sont soit monotones soit bitonals.

Ces tons sont associés soit aux syllabes accentuées (« pitch accents ») soit aux frontières droite et gauche du groupe intonatif (« boundary tones »). L'alignement avec la syllabe accentuée permet une distinction catégorielle entre les contours H* L (où H est aligné avec l'attaque de la syllabe accentuée) et HL* (où L est aligné avec l'attaque de la syllabe accentuée). Le ton étoile est attaché à la syllabe accentuée alors que le ton non-étoile peut être sujet à prolongation c'est-à-dire s'étendre sur une portion plus grande.

Les tons H et L prennent aussi des valeurs relatives tout au long d'un énoncé et sont soumis aux effets globaux de déclinaison ou d'augmentation de registre (« upstepping/ downstepping effect »).

La représentation phonétique des accents mélodiques peut prendre la forme d'un pic, d'un sommet ou d'un mouvement mélodique comme le montre l'illustration suivante :

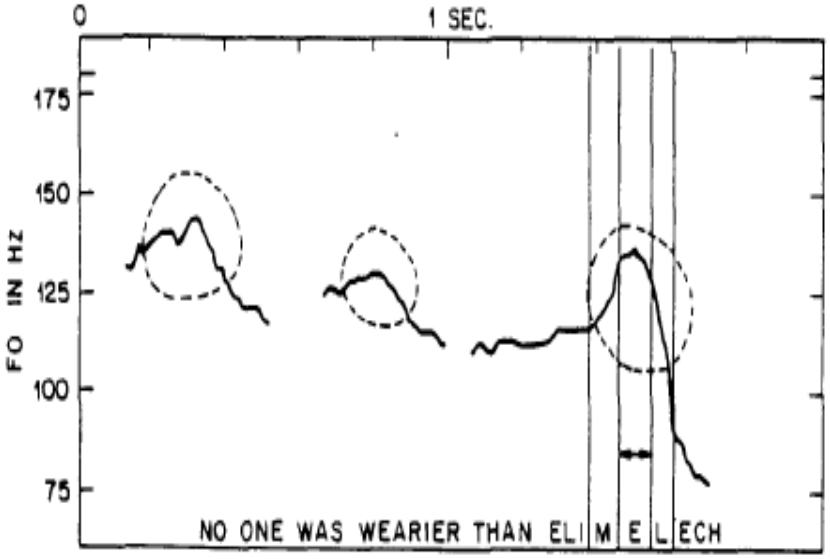


Figure 13 : Différents types d'accents mélodiques dans le modèle de Pierrehumbert (respectivement : un pic, un sommet et un mouvement mélodique). Extrait de Pierrehumbert (1980).

La grammaire développée dans Pierrehumbert (1980) est représentée dans la figure ci-dessous. Pour chaque morphème intonatif (ton de frontière, accent mélodique, accent de syntagme), on a une liste des tons possibles puis par un système d'arcs et de flèches, toutes les combinaisons possibles entre les différents tons sont présentées.

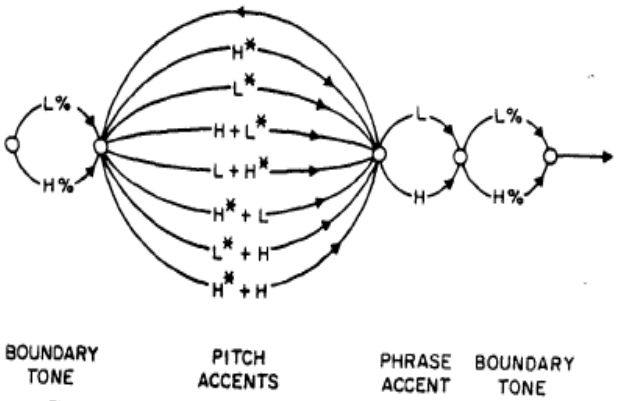


Figure 4: The grammar of the phrasal tunes of English given in Pierrehumbert (1980).

Figure 14 : Grammaire des contours intonatifs donnée dans Pierrehumbert (1980, p. 29).

Les contours définis contiennent tous au moins un accent mélodique suivi d'un ton de phrase et d'un ton de frontière (tous deux étant obligatoires). Comme le remarque Ladd (1996), il n'y a pas de structure comparable à la tête et au nucléus dans le modèle de Pierrehumbert. Toutes les combinaisons possibles entre un accent mélodique et un ton de frontière sont en effet autorisées de sorte qu'il n'y a pas de différence entre les accents prénucléaires et les accents nucléaires à l'exception de leur position. D'après Pierrehumbert, le nucléus est simplement le dernier accent de l'énoncé, ce qui représente un point de rupture fondamentale avec la tradition britannique.

À partir de cette grammaire, un mot dissyllabique tel que « *Anna* » peut être prononcé avec cinq contours différents illustrés dans les annexes de la thèse de Pierrehumbert (1980, p.255). Tous ces contours mélodiques ont la même structure métrique, c'est-à-dire que le premier accent mélodique est associé à la première syllabe de « *Anna* ». On peut établir une comparaison superficielle de ces contours avec les tons nucléaires des approches britanniques. On peut donc distinguer :

- Un contour descendant : H*L-L% ;
- Un contour descendant-montant : H*L-H% ;
- Un contour montant-descendant-montant : L*+H-L-H% ;
- Un contour montant : L*H-H% ;
- Un descendant plus progressif : H*+L-H-L% .

Il ne semble pas possible de trouver un équivalent direct à la distinction entre les grands mouvements mélodiques ou les plus petits (cf. « low fall » *vs* « High fall ») si bien qu'il est plus facile de comparer ces accents mélodiques à ceux définis dans le modèle de Crystal (1969) (« falling, falling-rising, rising-falling-rising, rising », cf. section 2.1.3 sur « prosodic

features »). Le dernier contour peut être mis en relation avec le contour vocatif défini par Liberman (1978) qui représente le deuxième contour descendant par le ton « low-mid ».

Ladd (1996) propose un tableau de correspondance entre les accents nucléaires de l'approche britannique et de l'approche de Pierrehumbert (1980). Il dénombre vingt-huit combinaisons possibles à partir des sept accents mélodiques, des deux tons de phrase et des deux tons de frontière. Il trouve un ton nucléaire équivalent pour chacun de ces contours et en conclut que les contrastes de base définis dans les deux approches sont valides. Il émet toutefois des réserves quant à l'utilité de traduire l'approche américaine dans les termes de l'approche britannique : les deux modèles étudient des phénomènes différents comme les contours de tête chez les Britanniques et la distinction entre les contours dits grands et petits (« high fall *vs* low fall »). De plus le modèle de Pierrehumbert a pour but de proposer une nouvelle façon de représenter les contrastes intonatifs plutôt que de proposer une nouvelle façon de représenter les mêmes modèles de l'intonation :

> *« It is however pointless to attempt to state a complete correspondence between the two systems. For one thing, the British tradition covers a range of divergent analyses of certain phenomena, notably the heads or patterns of prenuclear accents, but also a distinction between for example, high fall and low fall. […] For another thing, Pierrehumbert's approach is intended to provide a new foundation for analysing intonational distinctions, not merely a new notation for the same old description. »* Extrait de Ladd (1996, p. 83)

L'analyse développée dans ce modèle a connu de nombreuses modifications à la suite des travaux menés par Pierrehumbert après la publication de sa thèse. Nous prendrons l'exemple des modifications proposées par Ladd (1983).

2.2.2.3 Modifications du modèle original de Pierrehumbert

Bien qu'ayant adopté l'approche autosegmentale métrique de l'intonation de l'anglais, Bob Ladd souligne l'existence d'incohérences et de choix ambigus dans la représentation de Pierrehumbert et propose une révision de ce modèle. Ladd avance que le modèle de Pierrehumbert est trop complexe et ne parvient pas à représenter certaines similitudes entre les différentes classes de contours mélodiques.

La révision du modèle s'articule autour de deux problématiques essentielles : la représentation des phénomènes d'abaissement (« downstep ») et la nature de la phrase intermédiaire (« intermediate phrase »).

Les phénomènes d'abaissement sont considérés comme des modifications phonétiques affectant certaines séquences de tons. Ainsi dans une séquence d'accents de type H*…H*, le second accent est réalisé au même niveau que le premier alors que dans une séquence de type H*+ L…H*, le second accent mélodique est abaissé. Cette analyse est empruntée à l'analyse des langues ayant un ton lexical ou mélodique comme le japonais. Ladd (1983) ne partage pas

ce point de vue et affirme que les phénomènes d'abaissement devraient être considérés comme des choix intonatifs indépendants.

Ladd propose alors de considérer l'abaissement comme une caractéristique spécifique qui peut être présente ou absente indépendamment du choix du type d'accent mélodique. On peut avoir une séquence sans abaissement notée H*…H* comme dans l'approche de Pierrehumbert mais aussi un effet d'abaissement pour un même type de séquence noté H*… !H*.

D'autre part, Ladd émet des réserves quant à la justification du ton de phrase dans le modèle de Pierrehumbert. Ce ton est inspiré des études menées par Gösta Bruce sur le suédois. En effet, Bruce (1977) a montré qu'un contour intonatif du suédois de Stockholm pouvait être analysé comme une séquence d'un à plusieurs accents lexicaux suivis d'un pic mélodique localisé au niveau de la fin du mot focalisé de chaque phrase et finissant par une chute mélodique finale. Pierrehumbert a ainsi avancé qu'en anglais l'accent mélodique était aussi suivi de deux types de tons : le ton de phrase et le ton de frontière. Ceci est reflété dans la représentation du contour montant-descendant-montant. La première montée est analysée comme un accent mélodique L*+H suivi d'un ton de phrase L- qui reflète la chute et un ton de frontière H% représentant la montée. Cette analyse devient problématique quand il s'agit de représenter des contours mélodiques plus ordinaires comme les équivalents des tons nucléaires montants supérieurs et montants inférieurs (« high rise » et « low-rise »). Ladd (1996) fait ainsi remarquer que dans une séquence du type H*…H…H%, il n'est pas possible de justifier la distinction d'un ton de phrase du point de vue phonétique. Il admet la présence d'un autre ton après le « ton étoile » dans certains cas, mais émet des doutes quant au fait que ce ton soit interprété comme un élément phonologique toujours présent dans un énoncé. Il pourrait ne s'agir parfois que du second élément d'un accent mélodique bitonal.

Enfin, l'essentiel de ce modèle est codifié dans le système de transcription prosodique TOBI (Tone and Break Indices) que nous présenterons dans une prochaine partie.

Dans les deux premières parties de ce chapitre, nous avons passé en revue les différents modèles de représentations des formes prosodiques dans les approches britannique et américaine. Nous nous sommes focalisée dans un premier temps sur les types de contours intonatifs ou configurations étudiées dans chaque approche. La dernière partie de ce chapitre sera consacrée à une autre problématique formelle, celle de l'alignement de ces différents contours.

2.3 La question de l'alignement tonal

Le terme « alignement tonal » fait référence à la synchronisation dans la parole, entre la « rangée » des tons et la « rangée » des segments. La production de syllabes ou segments est en effet simultanée à la production de la mélodie ou des cibles tonales. Le phénomène d'alignement revêt un caractère universel, car présent dans toutes les langues et spécifique, car on suppose qu'il varie selon les langues et peut aussi contribuer à distinguer les variétés d'une même langue. L'étude de l'alignement tonal contribue à enrichir toute tentative de modéliser le

système intonatif d'une langue donnée et peut même avoir une application directe en synthèse vocale. Nous nous poserons alors les questions suivantes : comment les tons sont-ils associés aux segments ? quels sont les points de repère segmentaux qui permettent de prédire l'alignement des cibles mélodiques ?

Tout d'abord, les chercheurs ne se sont penchés sur la question de l'alignement à proprement parler qu'à partir des années quatre-vingt. Toutefois, bien que les théoriciens de l'école britannique de l'intonation ne parlent pas directement d'alignement tonal, on peut déduire de leurs descriptions du système intonatif de l'anglais quelques généralités en la matière.

2.3.1 Prémisses à l'analyse de l'alignement tonal chez les premiers théoriciens de l'intonation

Tous les contours décrits dans l'approche britannique sont associés aux syllabes accentuées d'un énoncé. Jones (1972) décrit le « contour 1 » à partir des syllabes accentuées qui forment une séquence de notes descendantes tout comme la plupart des autres théoriciens. Chaque contour nucléaire dans les modèles de Palmer (1924) ou O'Connor & Arnold (1961) est associé au début de la syllabe accentuée. Le traitement des contours de queue est aussi un bon exemple de la description des phénomènes d'alignement. Prenons le cas de la chute-montée divisée dans le modèle de Kingdon (1958). Si le contour n'est pas interrompu par des accents partiels dans la queue, on a une chute sur la syllabe nucléaire et une montée au niveau du second accent mélodique. Si des accents partiels apparaissent entre le premier accent mélodique et le second, le contour est étendu sur ces syllabes et reste à un niveau grave à partir de la fin de la chute (cf. p. 23). Pour d'autres tels que Gimson (2001), la chute-montée divisée est décrite différemment dans le cas où le nucléus n'est pas suivi de syllabes accentuées : la montée commence au début de la première syllabe inaccentuée et se répartit sur les autres jusqu'à la fin de l'énoncé. Nous avons donc là quelques indications de la façon dont l'alignement des contours est interprété dans les modèles britanniques.

C'est toutefois grâce au développement de la théorie autosegmentale que les phénomènes d'alignement tonal ont pu être étudiés plus précisément. La représentation séparée des évènements tonals et segmentaux chez Goldsmith (1979) permet d'analyser les contours prosodiques comme des séquences de tons qui sont alignés avec une voyelle unique.

Enfin, il est important de faire la distinction entre ce que l'on appelle l'association et l'alignement. D'après Ladd (1996), le concept d'« association » fait référence à une opération phonologique alors que le terme « alignement » désigne une propriété phonétique du chronométrage relatif des évènements dans les contours de F0 et des évènements de la ligne segmentale.

Ainsi, en observant les différentes approches britanniques et autosegmentales, on peut déduire les concepts d'association des tons qui en découlent, mais ces théories ne suffisent pas à rendre

compte du phénomène phonétiquement parlant. Un précurseur de la recherche expérimentale sur l'alignement tonal est Gösta Bruce (1977).

La thèse de G.Bruce donne suite à de multiples travaux de recherche sur le phénomène de l'alignement en montrant que sa structure sous-jacente est à l'origine de la distinction catégorielle de deux types d'accents nucléaires en Suédois. En effet, G.Bruce compare deux types d'accents mélodiques suédois dans sa thèse qu'il se propose de distinguer à un niveau phonétique. Dans le premier accent, le sommet mélodique H* est aligné plus tôt que dans le second : le premier est réalisé avant l'attaque de la syllabe alors que dans le second cas, le pic est aligné avec le noyau de la syllabe accentuée.

Dans l'optique d'une description phonétique de l'alignement tonal, nous passerons en revue les principaux travaux effectués dans les années quatre-vingt. Cette revue de lecture sera axée autour de trois aspects : les facteurs influençant l'alignement des points cibles, le domaine d'association des cibles tonales et quelques résultats plus précis concernant l'ancrage segmental des points cibles.

2.3.2 Facteurs influençant l'alignement des tons mélodiques

Dans son commentaire sur les travaux de recherches sur l'alignement des sommets prénucléaires en américain par Pierrehumbert et Silverman, Bruce (1990) dresse un bilan quasiment exhaustif des facteurs qui influencent l'alignement des cibles tonales en mentionnant :

- La composition tonale (analyse phonologique des accents mélodiques : accent mono ou bitonal, lié ou non, cible ou mouvement…) ;
- Le contexte prosodique : les frontières de mots, phrases ou énoncés, l'organisation rythmique (groupes rythmiques et conflits d'accents adjacents), la focalisation (position pré-focale, focalisée, post-focale), l'environnement tonal (interaction à l'intérieur et entre les accents mélodiques successifs = « tonal crowding »), les variations de registre (locales ou globales, les différences graduelles d'emphase dues au degré d'implication du locuteur), l'intonation globale de l'énoncé (absence ou présence de modifications dues au type d'énoncé qui peut être interrogatif ou déclaratif) ;
- Le contexte segmental (les différences de longueur intrinsèque de la voyelle) ;
- Le tempo (rapide, normal, lent).

Silverman (1987) propose une analyse en trois axes qui reflète la répartition des facteurs révélés par Bruce en distinguant :

- Le plus petit niveau qui correspond aux perturbations segmentales et à la micro prosodie ;
- Le niveau intermédiaire : l'intonation de la phrase et le choix des groupes qui constitueront en séquence la phrase intonative ;

- Le niveau le plus élevé : le niveau du paragraphe ou du discours.

À chacun de ces niveaux, on peut faire correspondre un ensemble de facteurs susceptibles d'influencer l'alignement des tons.

La bonne connaissance de ces facteurs permet d'éliminer les variations graduelles pour essayer de se concentrer sur les phénomènes sous-jacents stables dans l'alignement. Une autre question se pose alors : quelle est l'unité de base la plus pertinente pour l'étude des phénomènes d'alignement ?

2.3.3 Le domaine d'association des tons

On a vu dans la partie précédente que les tons mélodiques sont couramment associés à la syllabe accentuée : ainsi dans Laver (1994) on trouve l'affirmation suivante :

> *« Tone is a feature of syllables, pitch accents of words and intonation of phrases or sentences. »*

Ladd (1996) affirme aussi que les accents mélodiques doivent s'associer avec les syllabes proéminentes dans la chaîne segmentale et les accents ou tons de frontière avec les extrémités des domaines prosodiques de tailles variées.

À travers les différents travaux et à la lumière des méthodes expérimentales, ces affirmations évoluent et se précisent. Cette question est loin d'être triviale, car elle revient à se demander comment les différents chercheurs procèdent concrètement, comment ils effectuent les mesures qui conditionnent les résultats qu'ils obtiennent ensuite.

Quels sont les points de repère de référence dans la chaîne segmentale ? Le début de la syllabe entière ? Le début de la voyelle ? Le début du voisement ?

Pierrehumbert et Steele (1989) prennent le début de la syllabe accentuée comme point de référence de l'alignement de F0. Leur étude porte sur la configuration de F0 dans le contour montant-descendant-montant (« rise-fall-rise ») en anglais américain et ils se proposent de distinguer deux configurations différentes : (L+H*LH%) et (L*HLH%). Dans le premier cas la syllabe nucléaire est alignée avec le pic mélodique H alors que dans le second, elle s'aligne avec l'attaque de la montée L.

House (1989) étudie trois types d'accents nucléaires : la grande chute « high fall », la chute-montée « fall-rise » et la petite montée « low rise». Le début des différents mouvements de F0 (début de la chute, de la montée) a été localisé dans différentes parties de la rime syllabique. Le domaine identifié dans cette étude comme étant pertinent pour la réalisation du contour mélodique est la partie voisée de la rime, qui commence au début du voisement de la voyelle et continue jusqu'à la fin de la périodicité de la syllabe.

Gussenhoven et Rietveld (1995) affirment que l'alignement des cibles mélodiques n'est pas seulement sensible à des facteurs internes à la syllabe accentuée mais aussi au domaine extra syllabique :

> *« The alignment of pitch targets of intonation contours is sensitive to factors both within the accented syllable as well as in the subsequent portions of the utterance. »* Gussenhoven et Rietveld (1995)

Hirst (1983) montre que l'étude des structures et catégories dans la représentation prosodique conduit à observer que dans certains exemples des syllabes ne sont associées à aucune valeur mélodique alors que dans d'autres cas, elles sont associées à trois valeurs distinctes. On suppose généralement que les valeurs mélodiques sont connectées au patron accentuel de l'énoncé. Toutefois, on trouve aussi des unités mélodiques associées à des syllabes non accentuées. Il n'y a pas de correspondance directe entre les catégories phonologiques et les éléments tonals. Au-delà de la syllabe, on peut s'intéresser au pied métrique (pied d'Abercombrie : Ssss ou S = syllabe accentuée et s = syllabe inaccentuée).

Cette thèse est aussi soutenue par House et Wichmann (1996) dans leurs travaux sur l'alignement tonal en anglais britannique. Elles proposent d'associer les tons nucléaires anglais au niveau du pied métrique. Dans la réalisation phonétique, il y a de grandes chances pour que les sommets mélodiques de F0 coïncident avec les syllabes accentuées, mais même si elles ne le font pas, c'est la syllabe accentuée qui sera perçue comme proéminente. Cette suggestion n'est pas nouvelle alors, car Pierrehumbert et Beckman (1988) décrivent l'accent comme une « propriété du pied métrique attirée par la syllabe de tête ou syllabe forte ».

Möbius (1995), travaillant sur la modélisation de F0 pour la synthèse, avance que la courbe associée à un accent est la propriété du groupe accentuel en entier, défini comme ceci : « syllabe accentuée + quelques syllabes de queue inaccentuées ». Les exemples qu'il donne correspondent aux limites du pied d'Abercombrie (1964), car toutes les syllabes accentuées reçoivent une configuration mélodique.

On a donc bien vu que le domaine traditionnellement associé au ton est la syllabe accentuée, mais, si à un niveau phonologique abstrait cette réponse est tout à fait acceptable, elle se heurte à des difficultés lorsqu'il s'agit d'évaluer le phénomène phonétiquement parlant. Comment expliquer que des tons mélodiques soient associés à des syllabes non accentuées ou que le sommet mélodique H* soit parfois réalisé avant l'attaque de la syllabe accentuée elle-même ? La première analyse est donc trop imprécise et ne rend pas compte de la réalité phonétique du phénomène d'alignement. De nombreux théoriciens s'en sont rendu compte dans leurs expériences et proposent une autre solution. Il s'agit d'associer le domaine des tons à une unité plus large qui est le pied métrique constitué d'une syllabe accentuée de tête suivi d'un nombre indéfini de syllabes inaccentuées.

Après avoir défini les facteurs influençant l'alignement des tons et le domaine pertinent pour l'étude de la réalisation de ceux-ci, reste l'observation des différents travaux de recherches eux-

mêmes et la mise en valeur des résultats obtenus. Quelles sont les principales découvertes sur les phénomènes d'alignement à ce jour ? Peut-on en retirer des principes généraux caractéristiques de ce phénomène qui aideront à effectuer la recherche conduite dans ce travail ? C'est l'objet de la section suivante.

2.3.4 Résultats des différents travaux de recherche

Les différents auteurs se sont d'abord attachés à localiser les maxima de F0 dans le temps par rapport aux segments. Ces maxima locaux ont été étudiés en position nucléaire et non nucléaire. On s'est aussi penché sur des accents mélodiques particuliers comme la chute, la montée pour localiser et tenter de mettre à jour des régularités dans leur formation. Nous pouvons donc commencer par passer en revue les différents travaux portant sur l'alignement de H*.

2.3.4.1 Alignement de H* en position nucléaire

En ce qui concerne la langue espagnole, Pilar Prieto, Jan van Santen et Julia Hirschberg ont mené un projet commun en 1994. Il s'agissait de l'étude d'un corpus oral de 810 énoncés déclaratifs lus par deux locuteurs parlant l'espagnol mexicain. Les phrases contenaient toutes neuf syllabes-cibles associées à un sommet mélodique haut (H*) et placées dans différents contextes prosodiques (à la fin d'un syntagme intonatif), et l'on changeait systématiquement la position de la syllabe dans le mot et la distance en syllabes qui sépare les syllabes accentuées des syllabes inaccentuées. Les données ont montré qu'il était important de prendre en compte les variations de durée intra syllabique, les facteurs prosodiques, la nature des frontières et les « clash » accentuels c'est-à-dire les configurations où l'on trouve plusieurs syllabes accentuées dans un environnement proche ou plus précisément quand la syllabe nucléaire est suivie d'autres syllabes accentuées.

Les mesures extraites des expériences menées furent les suivantes :

- La durée des segments comprenant la syllabe accentuée et les syllabes inaccentuées qui la suivent ;
- La valeur maximale de F0 dans l'accent mélodique cible ;
- La distance mesurée en unité de temps entre le maximum de F0 et l'attaque de la syllabe accentuée (pour évaluer les délais possibles) ;
- La valeur minimale de F0 au début de la montée ;
- La distance en unité de temps entre le début de la montée de F0 et l'attaque de la syllabe accentuée ;
- Est-ce que la syllabe accentuée suivant la cible accentuelle étudiée portait aussi un accent mélodique ?

À la lumière des résultats des tests sur l'effet de la durée de la syllabe accentuée, il apparaît que l'allongement de l'attaque et de la voyelle provoque un déplacement du ton H* qui se décale

vers la droite. Ainsi le début de la rime semble être un point de référence pertinent pour l'alignement des éléments de F0.

De plus, de façon générale, la proximité d'une frontière de groupe intonatif majeur implique plus d'effet de rétraction vers la gauche de H* que celle d'une frontière mineure ou intermédiaire. De même les sommets de F0 ont tendance à se rétracter vers la gauche alors que leurs syllabes associées se rapprochent de la fin du mot.

En 1995, T.Rietveld et C.Gussenhoven publient les résultats de leurs recherches portant sur l'alignement des cibles de F0 dans la synthèse de la parole et les effets de la structure syllabique. Ils travaillent sur le contour « flat hat » (LH*L) en néerlandais, qui peut avoir une chute finale avancée ou tardive. Ils pensent que c'est la variation de la structure de la syllabe accentuée qui en est le principal moteur. Ils étudient plus précisément l'influence des segments sur l'alignement de H*. Les segments voisés dans l'attaque de la syllabe et dans la coda semblent décaler la cible H* tantôt vers la gauche, tantôt vers la droite alors que les segments non voisés dans l'attaque exercent une attraction vers la gauche. Leur originalité vient de l'utilisation d'un facteur non mentionné jusqu'ici nommé le « centre-p » ou « p-centre ». Ce facteur fait référence à la perception des phénomènes d'alignement. Le centre p est le point d'ancrage que l'on décèle intuitivement pour un accent mélodique. L'étude des données montre que ce centre est sensible aux variations ayant lieu au niveau de l'attaque et de la coda. Ainsi une attaque voisée correspond à un centre p tardif alors qu'une attaque plus longue tend à provoquer l'avancement du centre p.

Cependant les derniers résultats diminuent la pertinence de ce point d'ancrage. D'après les auteurs de cette recherche l'alignement de H* varie plutôt en fonction de la longueur de l'attaque et de la présence ou absence dans l'attaque et la coda de consonnes voisées ou non voisées.

2.3.4.2 Alignement de H* : H* en position prénucléaire

Une des premières recherches sur le sujet est l'œuvre de Silverman et Pierrehumbert (1990). Le schéma étudié est le suivant : H* HL. Deux adultes de sexe masculin ont enregistré le corpus composé de noms du type : « Ma Lemm, Mom le Mann, Mamalie, Mamalie Lemonick et Mama Lemonick ». Ces noms ne contiennent que des consonnes voisées sonores pour minimiser les perturbations segmentales à un niveau micro-prosodique de F0.

En résumé des résultats obtenus, on peut avancer que : à la fois la vitesse de production de la parole et le contexte droit influencent le placement de H* mais de façon qualitativement différente. Quand une syllabe est allongée du fait que l'on parle plus doucement, H* apparaît plus tard aussi. En revanche, quand cela est dû au contexte prosodique droit, la dernière partie de la syllabe subit un allongement nettement plus conséquent et H* apparaît plus tôt dans la rime syllabique. En ce qui concerne la distinction généralement admise entre les accents nucléaires et prénucléaires, on affirme ici que la distinction n'a pas lieu d'être. Tout comme dans les accents nucléaires, H* est aligné proportionnellement à la durée de la syllabe associée

plutôt qu'à une distance fixe dans la voyelle. Dans les deux positions, le débit de parole et le contexte droit exercent différents effets. Dans les deux cas, H* est aligné plus tard quand le nombre de syllabes suivant la syllabe portant l'accent augmente.

Dans les travaux de House et Wichmann (1996), la distinction entre les deux types d'accents est aussi mentionnée. Leur étude s'appuie sur l'analyse d'extrait du journal de la BBC appartenant au corpus SEC (Spoken English Corpus). Le début du schéma intonatif ou IO (intonational onset) correspondait à la première syllabe accentuée de chaque groupe intonatif tels qu'ils étaient notés dans la transcription du corpus. Pour chaque IO la location de H* dans le temps a été calculée comme un pourcentage de la durée totale de la syllabe accentuée. Dans quelques cas le maximum de F0 associé à la syllabe accentuée se situait plus tard dans le pied donnant alors des valeurs supérieures à 100%. Tous les facteurs autres que le tempo, identifiés pour leur effet de rétraction ou avancement du sommet furent pris en compte. Les facteurs définis par leur tendance à rétracter le sommet vers la gauche étaient les suivants :

- Une voyelle brève ;
- Le nombre de consonnes constituant l'attaque de la syllable ;
- La présence d'un segment voisé dans l'attaque ;
- La présence de deux accents consécutifs ;
- Les frontières du mot précédent ;
- Les frontières intonatives précédentes.

Les facteurs reconnus pour pousser le sommet vers la droite étaient :

- Une longue voyelle ou une diphtongue ;
- Une coda de la syllabe sonore ;
- Un pied polysyllabique.

Au niveau segmental, tous les facteurs attirant le sommet vers la gauche ont confirmé leur effet dans l'expérience. Ensuite la structure du discours et du développement du thème ou « topic » ont montré des effets comparables en exerçant une poussée vers la gauche au point que le sommet pouvait apparaître à l'extérieur des limites de la syllabe accentuée elle-même.

En ce qui concerne la distinction établie entre l'alignement des accents nucléaires et celui des accents non nucléaires, les auteurs de ces travaux proposent une autre interprétation. Il a été dit que le ton H* est aligné plus tard dans la syllabe accentuée quand il s'agit d'un accent non nucléaire et serait avancé dans le cas où il se trouverait en position finale. On suppose ici que les différences d'alignement qui ont été expliquées par la distinction entre les contours nucléaires et non nucléaires pourraient plus probablement s'expliquer par les effets de la position initiale et de la position finale (« initiality *vs* finality »).

Enfin les chercheurs ne se sont pas contentés d'étudier l'alignement du maximum de F0, mais se sont aussi intéressés aux contours intégralement décrits par F0, ce qui a permis notamment de révéler des points fixes d'ancrage en ce qui concerne l'alignement du minimum L* dans les montées.

2.3.4.3 Alignement des contours mélodiques complets

Bruce (1990) souligne l'importance d'aller plus loin dans la recherche sur l'alignement en s'intéressant aussi au chronométrage du mouvement mélodique en entier. En étudiant les deux types d'accents suédois : accent I = (H)L* et accent II = H*L, Bruce a constaté que c'était bel et bien la différence d'alignement des maxima et minima de F0 dans ces accents qui différenciait les deux sortes d'accents. Toutefois il ajoute que quand les contours de F0 pour l'accent H*LH sont étudiés dans différents environnements sans aucune référence segmentale autre que le début de la syllabe accentuée, le mouvement H*LH a un alignement stable. Alors que le ton H* a un alignement plus tardif dans les contextes droits plus longs, le mouvement H*LH est étonnamment constant dans son alignement.

Prieto *et al.* (1995) se sont aussi penchés sur la question dans leur recherche sur l'alignement en espagnol. Ils voulaient observer le comportement du mouvement en entier lorsqu'on le soumettait à des variations prosodiques contextuelles, quand le sommet H* se rétractait. On observe ici un effet de déclin du mouvement mélodique épousant la diminution du maximum de F0 lorsque l'on passe d'une position initiale à une position médiane puis une position finale. En revanche les valeurs minimales L semblent bien ancrées au niveau de l'attaque de la syllabe, ces valeurs semblent aussi diminuer plus on s'approche de la fin de l'énoncé. Le minimum de F0 de L* dans une montée est localisé précisément au niveau de l'attaque de la syllabe, ou quelques millièmes de secondes après. On ajoute aussi que la pente décrite dans une montée dépend du déplacement vers la droite du maximum de F0.

Enfin, des travaux de recherche sur l'alignement en néerlandais et en anglais ont été entrepris et les résultats en ont été publiés le 28 novembre 2003. Ceux-ci correspondent à l'aboutissement d'un projet de recherche nommé « target alignment project » financé par l'ESRC. Il s'agissait alors d'étudier en néerlandais et en anglais l'alignement des contours montants-descendants, les chutes ainsi que les montées. Le corpus était en partie constitué de textes lus et préparés puis de dialogues orientés (« The […] Map Task », Anderson *et al.*, 1991). Dans les accents montant-descendant, la cible minimale s'alignait avec l'attaque de la syllabe accentuée. De même, en néerlandais, L* se trouve être sensible à la location des frontières des mots. En néerlandais, on a découvert que si toutes les syllabes suivant la syllabe nucléaire étaient réduites, alors la montée finale dans les schémas H*LH commençait au début de la dernière syllabe. En revanche, si une syllabe forte apparaît après la syllabe nucléaire, le début de la montée s'aligne avec l'attaque de cette syllabe : tôt dans la syllabe s'il n'y a plus d'autres syllabes et plus tard si cette syllabe est suivie d'une autre syllabe accentuée.

Ainsi on a vu au travers de ces travaux que les facteurs définis par G. Bruce influençaient bien l'alignement des maxima et des minima de F0. L'alignement du maximum H* ne semble pas

être ancré de manière régulière en un point fixe du pied métrique, bien qu'il soit attiré par la tête du pied. Sa localisation varie selon le contexte prosodique droit et gauche ainsi que selon les facteurs micro-prosodiques. En revanche plusieurs chercheurs s'accordent à dire que le minimum L* de F0 précédant une montée a un point d'ancrage invariant qui est l'attaque de la syllabe accentuée et ceci dans plusieurs langues dont l'espagnol et l'anglais. De plus G. Bruce constate qu'en suédois, alors que le ton H* est inconstant, le mouvement mélodique intégral H*LH est aligné de façon stable, ce qui laisse penser qu'il est pertinent d'étudier l'alignement d'un schéma mélodique dans son ensemble et pas seulement comme une suite d'éléments discrets indépendants les uns des autres.

Les résultats de nos propres travaux expérimentaux sur l'alignement des contours creusés ou descendants-montants en anglais britannique (cf. Ali, 2004) offrent aussi quelques points de repère.

Deux hypothèses d'alignement sont mises à l'épreuve dans l'analyse statistique : la première postulant que la cible tonale étudiée est localisée à une distance fixe d'une unité « point de repère », la seconde postulant au contraire que la cible tonale étudiée s'aligne proportionnellement à la durée d'une unité segmentale. Les résultats sont les suivants.

D'une part les cibles tonales H1 (début de la chute) et H2 (fin de la montée finale) s'alignent de la même façon. Chacune de ces cibles peut être localisée par rapport à un point d'ancrage semblant fixe : l'attaque de la rime pour le pic mélodique H1 et la fin du groupe accentuel pour H2.

D'autre part, L1 et L2 représentant la fin de la chute et l'attaque de la montée montrent un comportement différent. L1 s'aligne en fonction de H1 auquel il est très fortement lié. L2 est d'une part très lié au maximum final qui explique la plupart de ses variations, et d'un autre côté, son alignement varie en fonction de la durée du groupe accentuel. Il est ainsi plus évident de prédire l'alignement du contour en entier que de se concentrer sur les différents points cibles ce qui rejoint les résultats de Bruce (1990) mentionnés plus haut.

Au-delà de l'étude de l'alignement des pics mélodiques dans différentes langues ou des configurations mélodiques dans leur intégralité, la recherche actuelle se pose des questions plus profondes. Il s'agit non pas seulement de faire un exposé exhaustif des différents cas mais de tenter de comprendre les mécanismes profonds qui définissent le phénomène d'alignement.

D'autres contraintes ont donc été mises en valeur grâce notamment à l'étude des mécanismes de production des tons (Xu et Wang 2001) et des limites de la perception (D'Imperio 2002a).

2.3.4.4 Mécanismes et contraintes articulatoires

C'est essentiellement dans les travaux de recherche entrepris par Yi Xu en 2001 sur le Mandarin, que ce thème de recherche est développé.

Ce dernier souligne le fait suivant : alors que l'on a prouvé que certains éléments de F0 sont alignés de façon stable avec des segments, comme les frontières syllabiques, rares sont ceux qui se sont intéressés aux mécanismes provoquant les schémas d'alignement observés. Pour comprendre les phénomènes d'alignement, Xu et Wang (2001) soutiennent qu'il est impératif d'être conscient des contraintes articulatoires qui pèsent sur la formation des contours de F0 dans le discours. En particulier, ils évoquent la vitesse maximale à laquelle F0 peut changer de valeur, ce qui limite la vitesse de production des mouvements de F0. De plus, la coordination des mouvements au niveau du larynx et au-delà de celui-ci contraint la façon dont les syllabes et les tons peuvent être alignés. En d'autres termes, le locuteur n'a pas autant de liberté que l'on pourrait le croire en ce qui concerne la façon dont il peut aligner les tons et les éléments segmentaux.

Vitesse maximale de changement de valeur de F0

Les premières expériences furent entreprises par Ohala et Ewan (1973) et Sundberg (1979). La méthode appliquée consistait à demander à des locuteurs de passer d'un niveau de F0 à un autre le plus rapidement possible. Avec cette méthode, il était difficile de déterminer le début et la fin du changement de niveau. Dans une étude récente, Xu et Sun (2000) ont appliqué une autre méthode : des sujets produisaient des changements de niveau rapides en imitant un modèle d'ondulation de F0 conçu par les chercheurs. Ces ondulations correspondaient à une vitesse que l'homme ne peut pas reproduire car trop rapide. Cela a permis de mesurer la durée entière de chaque changement mélodique. Il fallut aux sujets deux fois plus de temps que le modèle pour produire ces changements.

Il apparut alors que la vitesse maximale de changement variait de façon linéaire (du moins pour des changements mélodiques supérieurs à deux demi-tons).

Cette relation peut être représentée comme suit :

$S = 10{,}8 + 5{,}6\ d$ pour un mouvement montant ;
$S = 8{,}9 + 6{,}2\ d$ pour un mouvement descendant.

Où S est la vitesse moyenne de changement de niveau en demi-tons par seconde et d la valeur du changement de hauteur en demi-tons. Xu et Wang (2001) ont ainsi conclu qu'un minimum de 100 ms serait nécessaire pour passer d'un ton à un autre.

Coordination des mouvements laryngés et supra laryngés

Nul doute que les mouvements du spectre sont contrôlés séparément dans la parole. Toutefois cela ne signifie pas que les mouvements sont indépendants les uns des autres. La question que l'on se pose est la suivante :

Un locuteur peut-il librement ajuster l'alignement au niveau micro-prosodique des contours de F0 relativement à la chaîne segmentale ? Sinon, quelles sont les contraintes qui empêchent le changement d'alignement de se produire ?

En mandarin, le ton et la syllabe sont synchronisés, mais les expériences montrent que des variations sont possibles (alignement avec l'attaque consonantique ou la rime.)

On peut imaginer que dans un contexte tonal conflictuel et à cause des limites de la vitesse de changement mélodique, le locuteur ait à effectuer des réajustements pour faciliter la transition. Or les tests prouvent que cela ne se produit pas. Ainsi c'est la taille de la chute ou de la montée qui est compromise : il doit donc y avoir une contrainte plus forte qui empêche les réajustements. Cette contrainte peut être d'origine perceptive. Ou bien une autre explication pourrait reposer sur le fait que l'articulation des tons et des syllabes sont des mouvements concurrents contrôlés par un centre nerveux identique.

Ainsi Xu (2002) met en relief deux sources de contraintes sur la production de F0 : la vitesse de changement mélodique et la coordination entre les mouvements laryngés et supra laryngés. Toutefois si ces découvertes semblent être pertinentes dans le cas du mandarin, D'Imperio (2002) souligne le fait qu'elles se heurtent à des paradoxes notables. Comment rendre compte avec ces contraintes des différences d'alignement entre les différentes langues spécifiques ? D'Imperio propose de regarder les points d'ancrage sous un autre point de vue : celui de la perception.

2.3.4.5 Perception de l'alignement des cibles

D'Imperio (2002a) propose une alternative à l'analyse strictement acoustique de l'alignement. Cette analyse stipule que les cibles tonales sont des constructions issues de la perception, qui sont dépendantes des contraintes psycho-acoustiques et des contraintes spécifiques à la langue étudiée.

D'après les résultats des expériences menées dans cette recherche, il apparaît que, dans certains cas, la position des cibles tonales est difficile à déceler (cf. le cas des plateaux). Parfois, ce sont les perturbations occasionnées par les consonnes non voisées par exemple qui masquent la cible. Dans ces cas-là, l'étude de l'interaction entre les contraintes psycho-acoustiques et la perception peut être pertinente. Les expériences menées sur des configurations complexes de F0 comme les chutes, les montées, ont montré que les mouvements n'étaient pas perçus dans leur intégralité et que la cible de F0 perçue correspondait à la valeur de F0 aux deux tiers de la montée ou de la chute. De plus, quand un mouvement mélodique est précédé ou suivi d'un plateau, la valeur de F0 détectée est celle du plateau. La recherche sur l'alignement tonal du point de vue de la perception pourrait aussi aider à mieux comprendre la production.

D'après ses travaux portant sur l'italien napolitain, D'Imperio affirme que la structure de la syllabe conditionne la perception des cibles mélodiques. L'étude porte sur l'alignement des cibles avec les syllabes ouvertes et fermées. Les résultats confortent l'hypothèse de l'existence

d'un point d'ancrage fixe au niveau de la frontière gauche de la syllabe. La seconde variable mise en jeu est la distinction entre les énoncés interrogatifs et les énoncés affirmatifs. L'hypothèse testée était la suivante : les différences de structure syllabique dans les deux cas ne provoqueraient pas de déplacement des cibles vers d'autres frontières. Cette hypothèse fut en effet vérifiée. Cela permet de supposer que l'élément perceptif est déjà influent au niveau de la production de F0. De plus il y a un fossé entre les résultats trouvés pour l'italien et ceux correspondant à d'autres langues telle que le néerlandais. Par exemple en néerlandais, la présence de segments voisés dans la coda et dans l'attaque modifie la perception des points cibles dans le schéma montant-descendant (en chapeau). Cela confirme l'importance de la spécificité de chaque langue.

Ainsi, on constate que l'alignement de F0 répond non seulement à des contraintes liées à l'articulation et la production mais aussi à des contraintes d'ordre perceptif et dépendantes de la langue sur laquelle on travaille.

2.4 Conclusion

Dans ce second chapitre, nous nous sommes intéressée à la représentation des formes de l'intonation du point de vue des contours décrits dans la littérature et de leur alignement. Nous nous sommes étendue tout particulièrement sur les approches de la tradition britannique en ce qui concerne la description des contours intonatifs, car ils constituent une référence naturelle sur laquelle on reviendra au cours de notre travail expérimental. Nous avons soulevé des problématiques non résolues à notre connaissance telles que la nature du nucléus et l'analyse des accents postnucléaires dans le contour de chute+monté, qui nous intéresseront aussi du point de vue expérimental.

L'approche autosegmentale a été présentée par la suite et a permis de diriger notre attention vers la question de l'alignement tonal. Les résultats de différents travaux que nous avons cités sur l'alignement permettent d'obtenir quelques points de repère très utiles pour une tentative de modélisation de ces phénomènes. On retiendra notamment le point d'ancrage du point cible «L» au niveau de l'attaque de la syllabe accentuée dans le cas des contours montants, puis la rime comme point de repère important dans l'alignement du pic H. Nous prendrons aussi en compte les travaux portant sur le domaine d'association des tons permettant d'analyser l'alignement au niveau d'une unité supérieure à la syllabe, c'est-à-dire le pied métrique. Enfin, on pourra prendre pour référence la vitesse minimale de changement de F0 et porter une attention particulière aux limites de la perception.

Dans la section suivante, nous nous intéresserons aux systèmes de transcription de la prosodie qui représentent un aspect fondamental de ce travail de recherche. Les systèmes d'annotation actuels sont divers et variés, ils reflètent la multiplicité des théories phonologiques appliquées à la prosodie dont on a parlé dans ce chapitre. On essaiera de dégager les qualités et limites de ces systèmes d'annotation prosodique pour enfin proposer une approche expérimentale qui pourrait

aboutir à les tester ou à créer un nouveau système d'annotation prosodique à partir du travail expérimental mené sur un grand corpus.

Chapitre 3

Principaux systèmes d'annotation prosodique de l'anglais : illustration de la relation entre formes et fonctions prosodiques

De nombreux systèmes d'annotation prosodique ont été développés en vue de décrire et d'analyser l'intonation de l'anglais. On peut distinguer les systèmes de type « iconique » ou pour être plus précis les systèmes basés sur une description des formes prosodiques développés très tôt dans l'histoire de la recherche sur la prosodie, des systèmes plus abstraits ajoutant une représentation de l'information fonctionnelle et appliqués à l'étude de grands corpus ainsi qu'à l'avancée des technologies du langage (synthèse, reconnaissance vocale etc.). Dans ce chapitre, nous présenterons brièvement les premiers systèmes de transcription de l'anglais avant de nous étendre plus longuement sur les deux systèmes d'annotation les plus utilisés à grande échelle : les TSM (« Tonetic Stress Marks ») et TOBI. Enfin dans une troisième section, nous présenterons les systèmes de transcription que nous utiliserons dans la partie expérimentale de ce mémoire et expliquerons ce qui a motivé notre choix.

3.1 Les systèmes de transcription des formes prosodiques

3.1.1 Les portées musicales

C'est le plus ancien système de transcription de la prosodie que l'on retrouve notamment dans les travaux de Joshua Steele (1775) et Daniel Jones (1909).

Prenons la notation en note de musique utilisée par Joshua Steele (1775) comme exemple (cf. figure ci-dessous).

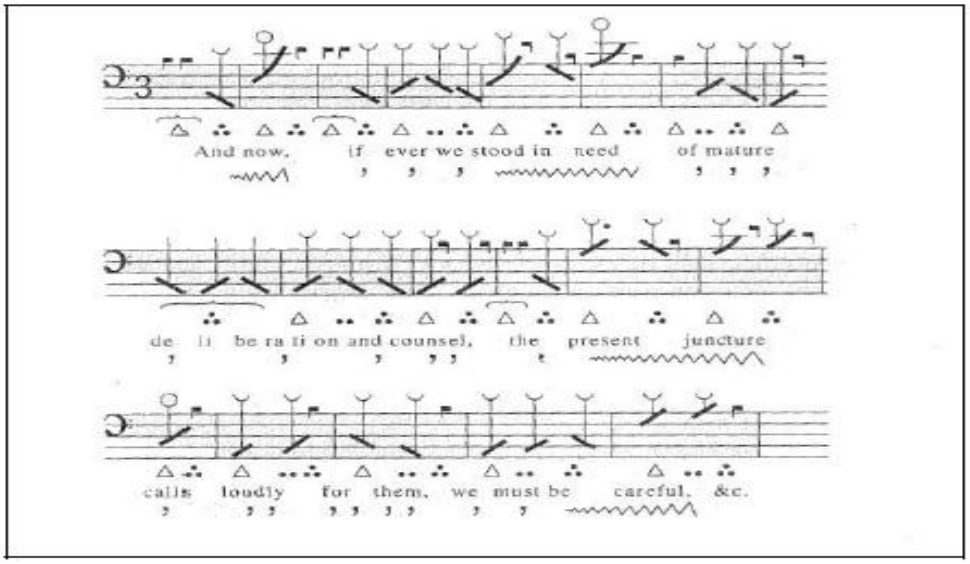


Figure 15 : Notation de Joshua Steele (1775) (extrait de Knowles, G & al. 1996)

La voix parlée est alors traitée de la même façon que la voix chantée bien qu'il existe de grandes différences de tessiture, d'étendue vocale entre les deux modes d'expression et malgré les différentes manières de passer d'une note à une autre. La tenue des notes, le « sustain » ou le vibrato sont spécifiques à la voix chantée.

Ce système a été considéré par la suite comme non approprié pour la description de l'intonation car il impose des limites non justifiées dans la voix parlée. Le demi-ton par exemple est le plus petit intervalle trouvé en musique occidentale alors que la parole peut présenter des écarts inférieurs et pourtant perceptibles. L'utilisation d'un tel système est laborieuse et ne permet pas de prendre en compte l'information fonctionnelle. Il s'agit d'un système purement descriptif. Il a toutefois inspiré d'autres systèmes de transcription basés sur la description des formes comme celui des niveaux.

3.1.2 Les niveaux

Le système utilisé par Trager et Smith (1951) représente un niveau d'abstraction supérieur aux portées musicales. L'intonation est représentée au moyen de quatre niveaux de hauteur notés sur la ligne segmentale sous forme de chiffres allant de 1 à 4. Ces niveaux sont interprétés comme des entités relatives plutôt que des valeurs absolues. La durée reste aussi largement variable.

Ce système se limite aux niveaux jugés pertinents au niveau linguistique et définit ainsi trois types de frontières représentant les contours intonatifs trouvés en fin d'énoncé : la chute rapide (« quick fall »), la montée finale (« terminal rise »), le maintien à niveau (« sustention of pitch »).

/ ²hæ̂w+də+ðèy+³stə́diy²| ²næ̂w+wìyv+gât+ðèhr+³búks¹#/
/ ²hæ̂w+də+ðèy+³stə́diy¹| ¹næ̂w+wìyv+gât+ðèhr+búks#/
/ ²hæ̂w+də+³ðéy+stə̂diy¹| ¹næ̂w+wìyv+gât+ðèhr+búks#/
/ ³hǽw+də+ðèy+stə̂diy¹| ¹næ̂w+wìyv+gât+ðèhr+búks||/
/ ³hæ̂w+də+ðèy+stə́diy²||²næ̂w+wìyv+gât+ðèhr+búks||/

Figure 16 : Transcription de Trager et Smith (1951).

À nouveau ce système de transcription reste difficile à utiliser pour l'annotateur et difficile à interpréter par le lecteur, car il ne s'agit pas d'un système iconique. On se tourne alors vers deux exemples de systèmes de transcription à la fois iconiques et faciles d'utilisation : « le crazy type » et les tracés curvilignes.

3.1.3 Le « crazy type » et les tracés curvilignes

À travers ses multiples publications, Bolinger (1958) base sa théorie de l'accent mélodique sur une annotation iconique utilisant les caractères alphabétiques. La hauteur relative des syllabes exprime alors les variations de hauteurs de F0 comme on peut le voir dans l'exemple ci-dessous :

lady had
dropped
i
A *t*

L'article « *a* » est prononcé dans un registre grave puis la première syllabe de « *lady* » est prononcée dans un registre aigu maintenu jusqu'à l'accent suivant porté par « *dropped* » et qui marque le début du contour descendant.

Cette annotation est donc facile à interpréter et plus pratique du point de vue du lecteur. Il n'y a pas d'idées préconçues en termes de niveaux, car ceux-ci peuvent être potentiellement infinis. Cependant l'information fonctionnelle, c'est-à-dire ici la place de l'accent de mot, n'est pas clairement indiquée et le lecteur doit avoir accès aux connaissances d'un locuteur natif. L'indication des maxima et minima de F0 n'est pas explicite non plus.

Les tracés curvilignes utilisés entre autres, par Liberman et Sag (1974) peuvent faire l'objet des mêmes critiques. La représentation se fait par une ligne continue située au-dessus de la ligne orthographique. La hauteur graphique représente aussi la hauteur relative de F0. Le lien avec le texte est indiqué de façon plus indirecte que dans la notation de Bolinger :

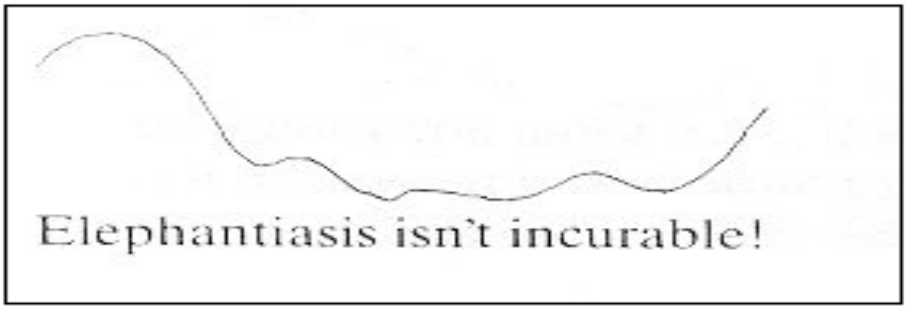


Figure 17 : Transcription de Liberman et Sag (1974) pour le contour intonatif de la contradiction.

L'accent de mot et la valeur des frontières hautes et basses de F0 ne sont pas explicités. Toutefois d'autres chercheurs (Couper-Kuhlen, 1986 p.107 ; Sherer et Wollmann, 1977) ont utilisé une version alternative de ce système dans laquelle la ligne était épaissie par endroits pour indiquer l'accent lexical. La place de l'information fonctionnelle va par la suite être de

plus en plus importante dans les systèmes de transcription prosodiques comme on le verra pour les Têtards, les TSM ou TOBI.

3.2 Les systèmes représentant formes et fonctions prosodiques

Nous présenterons dans cette section quatre systèmes de transcription utilisés pour l'étude de l'intonation de l'anglais : la notation en têtards, le système de transcription intégrée de Faure (1948 et 1962), les marques tonétiques et TOBI.

3.2.1 Les têtards

Ce type de représentation est très répandu et est utilisé aujourd'hui dans l'enseignement de l'anglais (cf. Ginésy 1995). En effet, sa nature iconique en fait un système aisé à interpréter pour les apprenants. Il utilise des points et tirets de différentes sortes, encadrés par deux lignes représentant les limites supérieures et inférieures du registre d'un locuteur.

Armstrong et Ward (1931) utilisent des tirets pour représenter les syllabes accentuées et des points pour les syllabes inaccentuées. Svartvik et Quirk (1980) analysent l'intonation en termes de nucléus et tonique de tête et utilisent des points larges pour les syllabes accentuées et des points plus petits pour les syllabes inaccentuées. Les points larges peuvent avoir des queues représentant un « glide » sur une syllabe :

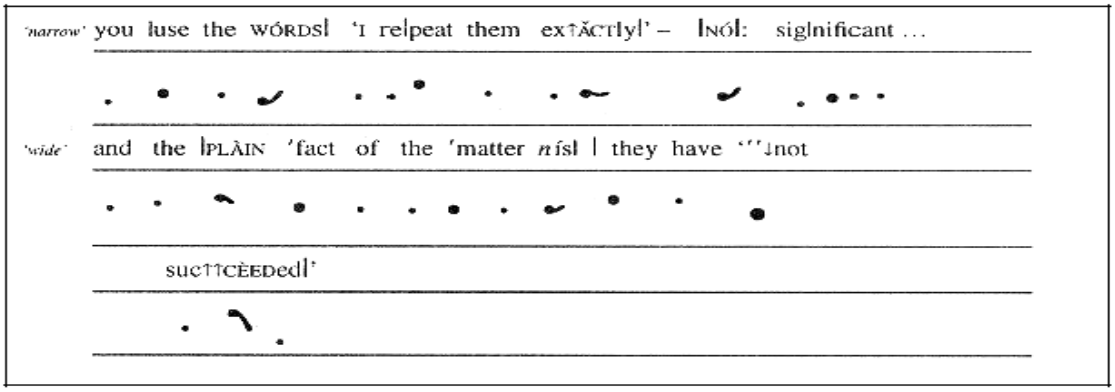


Figure 18 : Transcription de Svartvik et Quirk (1980).

C'est une représentation équivalente au « crazy-type » ou aux tracés curvilignes à l'exception du fait que la mélodie est divisée en unités discrètes correspondant aux syllabes. Au niveau de la représentation formelle, les glissandos et les sauts mélodiques sont distingués sans apriori à propos de la fonction de ceux-ci. De plus, une représentation graphique du registre est proposée grâce aux lignes supérieures et inférieures. Ces lignes correspondent en effet au maximum et minimum de F0 pour un locuteur donné. C'est une avancée considérable par rapport aux

modèles précédents, mais l'interprétation acoustique des fréquences maximum et minimum n'est toujours pas spécifiée.

Faure (1962) propose aussi un système de transcription basé sur la représentation de l'information formelle et fonctionnelle, mais choisit d'intégrer son annotation au texte comme on le verra dans la section suivante.

3.2.2 Un système de transcription intégrée

D'après Faure, aucun système de transcription ne peut fournir une représentation rigoureuse « *des durées, intensités et hauteurs manifestées par une voix donnée au cours d'un énoncé quelconque* » (p.171).

Faure veut donc développer une transcription de type phonologique qui ne décrira pas les variations ou variantes d'un même contour qui ne seraient pas distinctives :

> *« Les oppositions significatives, qu'elles soient démarcatives, accentuelles ou distinctives, sur l'un quelconque des trois plans du signifié psychologique, retenant seules notre attention. »* Faure (1962, p. 171).

La spécificité de ce système de transcription vient du fait qu'il s'intègre au texte si bien que toutes les informations nécessaires à la compréhension sont présentes :

> *« La transcription s'intègre au texte imprimé, afin que ce dernier constitue pour l'œil, comme pour l'oreille et pour la pensée, une unité cohérente, solidement structurée, dans laquelle sont étroitement intégrés à la fois les éléments phonématiques et des éléments musicaux. »* Faure (1962, p.172).

L'annotation des contours mélodiques principaux est associée aux syllabes accentuées comme dans les têtards. Les signes diacritiques sont tout simplement introduits dans le texte imprimé, au niveau du début de chaque syllabe accentuée :

> *« He can dis \solve the _parliament. »*

De plus, le système de Faure présente des enrichissements considérables au niveau de la représentation des variations de registre. Il distingue cinq niveaux de registre significatifs :

1. Le niveau suraigu pour annoter un ton statique (« ¯When ») et le point de départ ou le point d'arrivée d'un contour mélodique : « \ When (descendant) » ou « /When (montant) » ;
2. Le registre aigu noté par un trait horizontal ou un point de départ ou d'arrivée de mouvement mélodique au niveau du sommet des majuscules : « ¯When » ;
3. Le registre médium noté par un trait horizontal ou un point de départ ou d'arrivée au niveau des minuscules : « –When » ;

4. Le registre grave noté par un trait horizontal ou un point de départ ou d'arrivée de mouvement mélodique sur la ligne de l'écriture : « _When » ;
5. Le registre infra-grave noté en dessous de la ligne d'écriture : « __When ».

Les quatre premiers niveaux correspondent aux niveaux définis pas Pike, mais la figuration est plus parlante que les chiffres. Le cinquième niveau est associé à la fonction accentuelle et est très répandu en anglais britannique d'après Faure. Le fait de produire ce ton qui se situe nettement au-dessous du grave normal pour une voix donnée, exige un effort important qui lui confère une valeur nettement accentuelle. On peut l'illustrer dans l'exemple suivant :

« -Did you _like it ? »

C'est le saut vers le bas effectué sur « like » qui rend ce mot proéminent au niveau de cet énoncé.

Enfin, les systèmes les plus utilisés dans les travaux récents portant sur l'anglais sont les TSM et TOBI.

3.2.3 Tonetic Stress Marks : TSM

Les TSM ont été développées à partir des régularités découvertes dans les représentations en têtards. Kingdon (1958), O'Connor & Arnold (1961), Crystal (1969) et Svartvik & Quirk (1980) font usage de cette codification.

Les TSM sont des marques diacritiques de formes variées, ajoutées au texte avec deux types de barres pour marquer les frontières. Cette codification a été créée en pensant aux besoins des apprenants étrangers à l'origine, plutôt que pour véhiculer ou représenter une théorie de l'intonation en particulier. L'accent lexical est représenté de manière explicite et il n'y a pas de nombre fixe de niveaux mélodiques. Le registre n'est pas explicitement représenté dans la codification originale, mais Crystal (1969) utilise par exemple les guillemets pour indiquer un registre élargi ou compressé sur une portion de texte et dans une autre version des TSM, les signes « > » et « < » sont utilisés pour indiquer un registre élargi ou compressé localement. Les différentes versions de TSM proposées reposent sur les convictions théoriques de leurs auteurs.

La plupart des versions distinguent les accents nucléaires des accents non nucléaires (les « glides » n'apparaissent qu'en position nucléaire). Il y a en général un seul accent nucléaire par unité intonative (porté par la dernière syllabe accentuée mélodiquement). On distingue des syllabes accentuées du point de vue mélodique (accented), des syllabes accentuées du point de vue rythmique (stressed) et des syllabes inaccentuées.

D'autre part, on distingue deux niveaux de frontières prosodiques : les frontières majeures et les frontières mineures. La frontière majeure est accompagnée d'une pause remarquable, elle

marque la fin d'un énoncé complet. La frontière mineure peut apparaître à un niveau syntaxique inférieur, elle peut n'être marquée que par une rupture rythmique (cf. Roach, 1994 et Knowles, 1996).

Le système comprend donc un groupe intonatif majeur marqué par deux barres à la fin de celui-ci (//) et un groupe intonatif mineur noté « / ».

Ces groupes intonatifs ne sont pas divisés à nouveau en structures inférieures, mais chaque groupe doit contenir au moins une syllabe accentuée. L'inventaire des contours nucléaires dans ce système est le suivant :

- Low level : _
- Low fall :
- Low rise :
- Low fall-rise : \,
- High rise-fall : /'
- High level : ~
- High fall : \
- High rise : /
- High fall-rise : `/
- Low rise-fall : '\

L'utilisation des adjectifs « haut » ou « bas » signifie que le point de départ du mouvement jugé « haut » a une fréquence supérieure à celle du mouvement associé à la syllabe précédente et inversement pour un mouvement jugé « bas » ou « inférieur ». On distingue donc ici des grandes et petites chutes ou montées.

Le corpus MARSEC (cf. Auran *et al.* 2004) originellement appelé SEC (Spoken English Corpus) a été transcrit prosodiquement par Gerry Knowles et Brionny Williams (1995), en utilisant ce système et une méthode auditive. Cette transcription prosodique fournit une base de recherche remarquable sur l'intonation de l'anglais britannique et nous l'utiliserons dans la partie expérimentale de ce mémoire (cf. figure ci-dessous).

/`going to the \,shore | on the `first *morning of the `/holiday || the ~young \,English boy | ~stopped at a _turning of the \,path | and ~looked _down at a ~wild and _rocky \bay || and `/then | *over to the \crowded `beach | he ~knew so \well | from `other *years || his `mother *walked on in `front of him | *carrying a _bright *striped `bag | in one `hand || her `/other *arm | *swinging \loose | was very `white | in the \sun ||

Figure 19 : Transcription en TSM d'un passage d'une minute extrait du corpus MARSEC.

Enfin le système d'annotation prosodique le plus largement utilisé aujourd'hui est le système TOBI issu des théories américaines de l'intonation.

3.2.4 TOBI

Cette codification a d'abord été formulée par Pierrehumbert (1980) pour la partie tonale et par Silverman *et al.* (1992) pour les indices de frontières. C'est une codification inspirée en partie par le modèle de Pierrehumbert développé dans sa thèse de doctorat que l'on a présenté dans le chapitre 1.3. Pierrehumbert a utilisé TOBI pour annoter un corpus d'anglais journalistique et pour une expérience de synthèse à partir de cette annotation. C'est un système d'annotation standard applicable à une variété particulière de l'anglais : l'anglais américain. La première version de ce système d'annotation est appelée « MAE_TOBI » (Mainstream American English, cf. Jun 2005). Il a aussi été appliqué avec succès à l'anglais britannique méridional et à l'anglais australien.

Une transcription effectuée ainsi comporte six parties obligatoires :

1. L'enregistrement d'un énoncé quelconque ;
2. Le contour brut de la fréquence fondamentale ;
3. Une transcription autosegmentale du contour intonatif ;
4. Une transcription orthographique de chaque mot de l'énoncé ;
5. Une annotation des degrés de frontières perçus (« Break indices ») indiquée au niveau de la fin d'un mot ;
6. Une annotation de commentaires divers qui peut comprendre l'indication des hésitations, interruptions etc.

On a déjà mentionné les primitives de ce système qui sont les tons H et L (« high *vs* low » ou « higher *vs* lower », cf. chapitre 1.3), qui sont combinés en accents mélodiques. Le symbole «*» marque l'alignement de la cible tonale avec la syllabe accentuée correspondante sur la ligne orthographique.

Dans les accents mélodiques contenant deux tons, le signe « - » indique le ton qui suit l'accent (« trailing tone ») ou celui qui précède l'accent (« leading tone »).

Ci-dessous un inventaire des cinq tons mélodiques de base retenus par Beckman et Hirschberg dans les conventions d'annotation de ETOBI (English TOBI) :

H*	'peak accent'
L*	'low accent'
L+H*	'scooped accent'
L*+H	'rising peak accent'
H+!H*	'clear step down onto the accented syllable'

Figure 20 : TOBI pitch accents

On peut ajouter un symbole diacritique « ! » à chacun de ces accents pour marquer les phénomènes d'abaissement ou « downstep ». La modification proposée par Ladd (1983), (cf. chapitre 1.3) a finalement été intégrée dans cette version du système TOBI.

Les deux domaines intonatifs pris en compte sont identiques à la théorie d'origine : on distingue la phrase intermédiaire et la phrase intonative.

Au niveau de la phrase intermédiaire, on peut avoir les tons de phrase suivants : L- et H-.

Au niveau de la phrase intonative, quatre combinaisons de tons de frontière sont possibles :

1. L-L%
2. L-H%
3. H-H%
4. H-L%

D'autre part, il y a quatre niveaux de frontières entre les mots transcrits sur la ligne orthographique : de 0 à 4. Le degré 0 correspond à la frontière existant dans un groupe clitique c'est-à-dire entre deux mots prononcés en un seul mot comme « did you » prononcé « didya ». Le niveau 1 marque la frontière entre deux mots en l'absence de frontière prosodique évidente. Les niveaux 3 et 4 correspondent respectivement aux niveaux de phrase intermédiaire et phrase intonative. Le niveau 2 correspond à une frontière prosodique évidente qui serait perceptivement moins marquée que les niveaux 3 et 4.

Dans ce système tout comme dans les TSM, les informations fonctionnelles et formelles sont représentées conjointement. Les tons H et L font référence à la forme de F0, l'astérisque (*), le pourcentage et le tiret, représentent l'information fonctionnelle prise en compte (niveau de frontière et accent lexical).

L'intérêt évident représenté par ce genre de système d'annotation pour l'étude de grand corpus a poussé certains chercheurs à essayer d'établir des équivalences entre les deux systèmes principaux que l'on vient de présenter comme on le verra dans la section suivante.

3.2.5 Étude comparative des notations TSM et TOBI

À l'heure actuelle, seul Peter Roach (1994) a essayé d'établir des règles de conversion automatique entre le système TOBI et les TSM. Il ne prétend pas toutefois trouver des équivalences exactes entre les deux systèmes, car ceux-ci représentent deux théories de l'intonation distinctes.

Les TSM utilisées dans le SEC (Spoken English Corpus) permettent de distinguer dix accents nucléaires alors que l'on en compte seulement cinq dans TOBI. Deux niveaux de frontières sont pris en considération dans les TSM alors que dans TOBI on en compte quatre. D'après

Roach (1994), bien qu'une entreprise de conversion automatique semble difficile, une telle tentative permet de mieux comprendre les deux systèmes et l'information qu'ils représentent.

Tout d'abord, d'après les travaux de Roach (1994), on peut considérer une équivalence entre le niveau de frontière noté 4 dans le système TOBI et la frontière mineure dans les TSM. C'est-à-dire que la frontière mineure serait équivalente à une frontière d'unité intonative dans TOBI (« full intonation phrase boundary »). Ceci étant, le niveau de frontière majeure ne trouve donc pas d'équivalent dans le système TOBI. Roach propose alors d'ajouter un niveau 5 au système TOBI qu'il met ensuite à l'épreuve dans une série de tests de perception. À l'issue de ceux-ci, il conclut qu'un système d'annotation scalaire ou graduel semble plus approprié pour décrire ce que perçoivent les sujets, ce qui conduirait à un système à cinq niveaux de frontières plutôt que seulement deux dans les TSM. Toutefois il remarque aussi que les quatre niveaux de frontières de TOBI sont définis de manière bien spécifique et qu'une analyse plus approfondie devrait être menée sur cette question.

Une équivalence partielle est aussi établie pour l'annotation des phénomènes d'abaissement dans les deux systèmes. La flèche descendante serait donc le symbole le plus proche du diacritique « ! ». Cependant, la flèche peut être utilisée dans tout contexte alors que dans TOBI, on utilise ce signe que pour indiquer un ton H qui serait plus bas que la valeur par défaut.

Enfin, un problème de taille se pose lorsque l'on se propose de trouver un équivalent à la chute-montée non finale dans l'annotation de Pierrehumbert. En effet, si Roach parvient à trouver un équivalent à tous les contours nucléaires britanniques en se plaçant au niveau de la frontière de l'unité intonative, c'est-à-dire à l'aide de l'accent mélodique, de l'accent de syntagme et du ton de frontière combinés**,** il n'en est pas de même au niveau du ton de phrase. Au niveau de la frontière intermédiaire, la solution évidente de l'accent H* + LH n'est pas permise. Une chute+montée ne peut apparaître qu'avant une frontière d'unité intonative. Une autre expérience de perception est entreprise et montre qu'en anglais britannique, la chute+montée non finale a sa place, ce qui laisse penser que les systèmes de transcription doivent être adaptés à chaque langue ou même dialecte étudié.

3.2.6 Les limites et critiques apportées à TOBI

Bien que largement adopté par la communauté des chercheurs en linguistique ce système d'annotation prosodique fait l'objet de quelques remarques et critiques. Dix ans après la création de ce système, certains chercheurs font un bilan sur les apports et les limites de TOBI.

Tout d'abord certains chercheurs soulignent le fait que c'est un système développé à partir d'une étude de l'anglais américain. Il ne s'agit donc pas d'un alphabet prosodique international permettant de rendre compte des phénomènes linguistiques appartenant à d'autres langues. Des précautions sont de rigueur avant de tenter d'appliquer ce système à d'autres langues d'où la création d'un « Greek TOBI » et d'un « German TOBI ».

Ensuite ce système est l'expression d'une théorie phonologique particulière que tous n'adoptent pas. Les difficultés rencontrées dans la tentative de conversion menée par Roach montrent que chaque système se base sur des aprioris théoriques qui restent pour certains encore à mettre à l'épreuve dans l'étude d'un grand corpus de données.

La théorie de l'explication du sens apporté par la prosodie ou « intonational meaning » telle qu'elle est développée à partir de TOBI est aussi remise en cause par Dainora (2002) et par Calhoun (2004). Selon cette dernière cette théorie conduit à une simplification exagérée des faits prosodiques et de leur contribution au sens : c'est une critique apportée à l'approche compositionnelle mais aussi au système de représentation lui-même. D'après Calhoun, TOBI est beaucoup trop simple et ne suffit pas pour capturer les variations de sens causées par l'intonation. Elle cite des travaux qui montrent que l'expression des émotions telle que la colère, la tristesse dépend surtout de l'intensité ou de la durée qui est un indice des dispositions affectives (nervosité, hostilité…) :

> *« We should view intonational meaning as being conveyed by multiple acoustic cues (*F0*, intensity, duration, spectral tilt) interacting in a multi-dimensional intonational structure; in the absence of an agreed set of intonation abstract categories, we should concentrate on perceptual evidence which consistently links particular changes in acoustic features to a particular change in interpretation in a given discourse and phonetic context; statistical modelling is a good way to test the predictions of this type of approach. »* Calhoun (2004)

D'après Calhoun, le sens est donc véhiculé par plusieurs couches d'indices acoustiques comme l'intensité, la durée, la qualité de voix et l'on ne dispose pas actuellement d'un système d'annotation représentant ces différents indices. Les différents accents mélodiques décrits dans TOBI ne suffisent pas à rendre compte du rôle de l'intonation dans l'interprétation des énoncés.

Wightman (2002), dresse aussi un bilan de l'utilisation de TOBI dix ans après sous l'intitulé « TOBI or not TOBI ». Wightman réévalue l'efficacité de TOBI à la lumière des applications de l'étude de la prosodie dans les technologies du langage comme la reconnaissance vocale ou la synthèse. Il souligne les besoins fondamentaux tels que la rapidité de transcription, le coût et la valeur des annotateurs experts, l'utilisation de données authentiques et montre que TOBI ne parvient pas à y répondre. Tout d'abord, une transcription en TOBI demande un temps considérable, car même les annotateurs entraînés ont besoin de 200 fois plus de temps que le temps réel ce qui va à l'encontre des attentes de commercialisation des technologies du langage.

Ensuite le taux d'accord entre les annotateurs est discutable. On obtient un taux d'accord très élevé quand il s'agit de comparer la présence et l'absence de tons de frontière 50% quand il s'agit du type d'accent en question. Et ceci dans des conditions idéales avec des annotateurs experts.

Or Wightman remarque aussi que les systèmes développés à partir de TOBI et mettant l'accent sur les évènements perçus tel que le projet « Verbmobil » obtiennent de meilleurs taux d'accord.

Il en conclut alors qu'une solution aux problèmes rencontrés avec le système TOBI est de se focaliser sur l'annotation de ce que l'on entend :

« Label what you hear. »

Wightman nous encourage à ne pas annoter ce que l'on peut obtenir automatiquement grâce à des algorithmes tels que les contours de la fréquence fondamentale, ou ce que nous dictent les théories de l'intonation. Hirst (2005) partage ce point de vue et va plus loin en traduisant le problème soulevé par Wightman en termes de formes et fonctions. Selon Hirst, quand Wightman parle de se focaliser sur ce que l'on perçoit, il fait référence au message, à l'information fonctionnelle par opposition à la description des formes ou contours de F0 qu'il juge de nos jours inutile et qui est à l'origine des mauvais taux d'accord entre annotateurs. Il ajoute que « les machines sont efficaces pour annoter les formes alors que les hommes le sont pour annoter les fonctions ». C'est ainsi que Hirst (2005) propose une annotation distincte des formes et fonctions prosodiques dans le but de remédier aux problèmes de TOBI d'une part et d'autre part, d'arriver à une meilleure compréhension de l'articulation entre le niveau formel et le niveau fonctionnel. Nous adhérons à cette proposition qui se concrétise par notre choix d'utiliser deux systèmes d'annotation distincts : MOMEL-INTSINT au niveau formel et IF au niveau fonctionnel. C'est donc ce cadre d'analyse qui sera présenté dans les parties suivantes.

3.3 Vers une annotation distincte des formes et fonctions prosodiques

3.3.1 Fondations théoriques du modèle quadratique

Avant de décrire l'algorithme MOMEL (modélisation de la mélodie) et le système d'annotation INTSINT, nous présenterons la théorie de l'intonation qui en est à l'origine (cf. Hirst, Di Cristo et Espesser 2000).

D'après ces auteurs, la meilleure façon de développer le système de l'intonation d'une langue est de maintenir un dialogue entre les données empiriques et les théories linguistiques. Il s'agit de formuler des hypothèses à partir de données empiriques et de pouvoir tester celles-ci sur de nouvelles données. Ils proposent donc une approche à la fois « bottom up » et « top down ». Ils développent ainsi un système d'analyse et de représentation de la prosodie permettant potentiellement de confronter les différentes théories existantes à l'étude de données expérimentales.

Ce système de représentation repose sur la distinction entre plusieurs niveaux d'analyse de la prosodie. Tout d'abord, comme on l'a déjà mentionné plus haut, ils proposent de distinguer les représentations fonctionnelles qui permettent de coder l'information nécessaire à l'interprétation de la prosodie d'un énoncé, des représentations formelles qui codent l'information nécessaire pour la prononciation d'un énoncé.

L'opposition entre phonétique et phonologie permet aux auteurs de décrire les représentations formelles de façon plus approfondie que leurs prédécesseurs. Entre le niveau le plus abstrait (appelé niveau phonologique sous-jacent) et le niveau le plus concret (le niveau acoustique ou physique), ils postulent l'existence de deux autres niveaux de représentation : le niveau phonétique et le niveau phonologique de surface. Le niveau phonétique représente les phénomènes continus et variables dont on extrait les contraintes universelles sur la perception et production des sons. Il se distingue du niveau acoustique, car on présuppose dans cette approche que toutes les langues obéissent aux mêmes contraintes physiques (cf. Di Cristo 1985, p. 31). Pour expliciter cette distinction, les auteurs donnent l'exemple de la durée qui est souvent considérée comme un paramètre acoustique à l'image de la fréquence fondamentale et de l'intensité. Or par opposition à ces deux derniers paramètres, il n'est pas possible d'extraire automatiquement, une courbe de la durée identique aux courbes F0 ou de l'intensité. La durée n'est pas un paramètre purement acoustique et afin de mesurer et d'analyser ses variations, il est nécessaire de s'appuyer sur des frontières ou composants phonologiques comme les syllabes ou les phonèmes. La durée appartient donc au niveau de représentation phonétique dans cette approche. Ce niveau phonétique est modélisé pour la fréquence fondamentale avec l'algorithme MOMEL comme on le verra dans la section suivante.

Le niveau phonologique de surface représente des catégories discrètes distinctes avec lesquelles on peut décrire les phénomènes de surface à travers les langues. Il est codé par le système INTSINT que nous décrirons plus bas. L'hypothèse qui supporte ce niveau de représentation est que l'on peut décrire la prosodie de langues différentes en utilisant un inventaire de symboles restreint qui serait comparable à l'IPA (International Phonetic Alphabet). Cette hypothèse a été confirmée dans Hirst et Di Cristo (1998) où les patrons intonatifs de vingt langues différentes étaient présentés.

Dans la suite de ce chapitre, nous rentrerons dans plus de détails sur cette analyse de l'intonation en présentant MOMEL, INTSINT et IF.

3.3.2 Annotation automatique des formes prosodiques : INTSINT et MOMEL

3.3.2.1 Modélisation de la mélodie avec MOMEL

L'appellation MOMEL fait référence à la modélisation de la mélodie. Cet algorithme a été créé par Hirst et Espesser (1993) et repose aussi sur les travaux de recherche menés dans la thèse de doctorat de Hirst (1987). D'après Di Cristo et Hirst (1986), une représentation de la fréquence fondamentale devrait être dépourvue de ses caractéristiques micro-segmentales si bien que la courbe résultant d'une telle modélisation devrait être similaire à celle que l'on obtiendrait pour une phrase ne contenant que des sonorantes (cf. « Molly may marry Larry »). Une courbe de F0 est donc modélisée comme une superposition entre deux composantes : la composante micro-prosodique qui correspond aux variations ou perturbations causées par la nature des phonèmes (consonne sourde causant un pic dans les aigus), et une composante macro-prosodique reflétant le choix du contour intonatif pour un énoncé donné. Le principal effet micro-prosodique pris en

compte est l'effet d'abaissement des valeurs de F0 par endroits. Les seuls cas où les facteurs micro-prosodiques causent une augmentation de la valeur de F0 sont les attaques consonantiques non voisées (présence d'un « t »). L'algorithme modélise F0 en trois étapes.

1. D'abord une étape préliminaire permet d'éliminer les effets d'augmentation de F0 mentionnés ci-dessus. Toutes les valeurs supérieures à un seuil donné par rapport aux valeurs voisines (5%) sont éliminées.
2. Ensuite, on procède à une régression modale asymétrique : il s'agit d'estimer les points cibles optimaux potentiels ou « candidats ».
3. Enfin on réduit les candidats de chaque portion à un point cible.

Les paramètres de l'algorithme ont été optimisés sur un petit corpus (valeurs optimales : la fenêtre d'analyse A = 300 ms, le seuil d'écart D = 5%, la fenêtre de réduction R = 200 ms). L'optimisation de ces paramètres a ensuite été appliquée à la stylisation d'autres corpus de français montrant un pourcentage d'erreur de 5% (cf. figure ci-dessous).

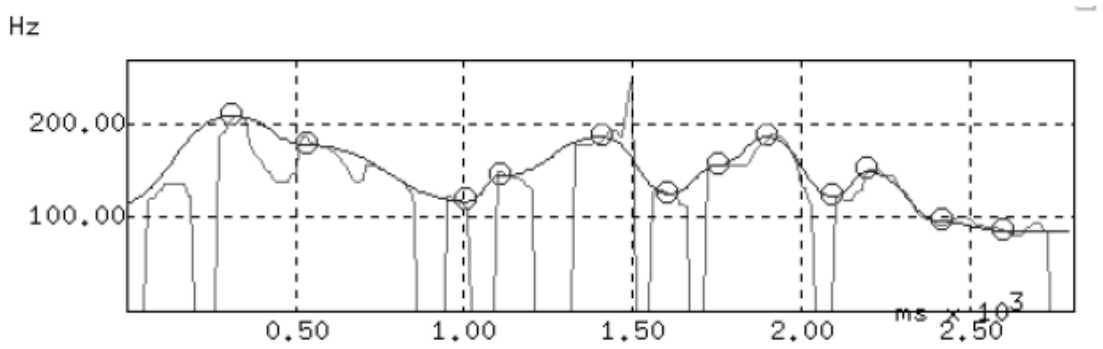


Figure 21 : Exemple de courbe de F0 et courbe spline quadratique obtenue avec MOMEL.

La modélisation automatique de MOMEL ainsi que le codage avec INTSINT dont nous parlerons dans le prochain paragraphe, ont été utilisés pour la modélisation phonétique et le codage symbolique de nombreuses langues telles que l'anglais, le français, le catalan, le portugais brésilien, l'espagnol du Vénézuela, le russe, l'arabe et le isizulu.

Ces travaux ont permis d'évaluer ces systèmes de modélisation et de codage et d'optimiser certains de leurs paramètres. À partir des travaux d'Astésano, Hirst et Espesser dans le cadre d'une évaluation multilingue (Astésano *et al.* 1997) ou une étude de l'intonation du Catalan (Estruch 2000), on peut donner quelques pistes d'évaluation de MOMEL. Ces deux travaux soulignent la qualité de modélisation automatique obtenue avec MOMEL et en présentent aussi les principaux problèmes. Il apparaît que les points cibles sont mal détectés avant ou après les pauses, il s'agit en général de valeurs de F0 soit très basses soit très aiguës. Il est donc nécessaire d'ajouter ou de déplacer des points cibles manuellement dans certains cas. Les contours les plus problématiques sont en général les contours ascendants situés avant ou après une pause. MOMEL ne détecte pas le point final avant la pause quand la fréquence est très élevée. Le même problème se pose dans le cas de certains contours descendants situés avant une

pause et pour lesquels MOMEL ne détecte pas le point cible final du mouvement dans des fréquences très graves. Dans Campione (2001), l'algorithme a été évalué à partir du corpus EUROM1 et pour cinq langues (l'anglais, l'allemand, l'espagnol, le français et l'italien), ce qui représentait cinq heures de parole. L'évaluation a montré un coefficient d'efficacité globale de 95,5% en comparaison avec les points cibles manuellement corrigés.

Plus récemment, dans Hirst (2007), une version révisée de MOMEL a été implémentée sous forme d'un « plugin » dans Praat. Auparavant, l'implémentation de MOMEL était essentiellement accessible sous Unix avec un programme en C, ce qui représentait un obstacle à une utilisation plus large de ce système de modélisation. La modélisation et le codage ont été implémentés sous forme de scripts Praat que l'on peut sélectionner directement dans le menu, ce qui facilite l'utilisation de l'algorithme. Les analyses se déroulent en cinq étapes :

- La spécification des valeurs maximales et minimales de F0 qui peut être faite automatiquement ou manuellement ;
- Le calcul de la valeur des points cibles : la courbe de F0 brute est modélisée grâce à la fonction quadratique ;
- La correction des points cibles : l'utilisateur peut visualiser le signal et les point cibles détectées qui peuvent être interpolés. On peut alors manipuler les points cibles manuellement et comparer le résultat avec la courbe originale en utilisant le système de resynthèse Psola ;
- Les points cibles automatiques ou corrigés sont codés par l'algorithme INTSINT ;
- On peut ensuite comparer les trois représentations (originale, modélisée par MOMEL, en tons INTSINT).

La majorité des erreurs systématiques liées aux pauses sont corrigées dans cette nouvelle version de MOMEL. Le « plug in » contient aussi une nouvelle version d'INTSINT que l'on présentera ci-dessous.

3.3.2.2 Codage de l'intonation avec INTSINT

Le système de codage INTSINT (INternational Transcription System for INTonation) peut être obtenu automatiquement à partir du traitement de la fréquence fondamentale par l'algorithme MOMEL.

Les patrons mélodiques sont codés à l'aide d'un ensemble de symboles mélodiques abstraits : {T, M, B, H, S, L, U, D} (pour Top, Mid, Bottom, Higher, Same, Lower, Upstepped, Downstepped). Chaque symbole caractérise l'un des points cibles de la représentation phonétique de MOMEL. Par opposition aux autres systèmes d'annotation tels que TOBI, l'utilisation d'INTSINT ne requiert pas la connaissance au préalable de l'inventaire des patrons mélodiques de la langue que l'on veut étudier. Hirst et Di Cristo (1998) présentent ainsi INTSINT comme un outil très utile pour rassembler des données et élaborer des descriptions phonologiques de langues peu étudiées auparavant.

Les tons INTSINT représentent deux dimensions de F0 : une dimension locale avec les tons relatifs H et U (plus haut que la cible précédente), L et D (plus bas), S (au même niveau), et une dimension globale avec les tons absolus T, B, M qui représentent les extrémités de l'étendue vocale du locuteur au niveau d'une unité intonative.

Une distinction est aussi faite entre les tons dits itératifs tels que {U, D} et les tons non itératifs {H, L, S}. Cette distinction est en accord avec les descriptions des contours de type « upstep » ou « downstep » qui montrent des paliers ou diminutions de F0 généralement inférieurs aux contours descendants isolés.

On s'appuie ainsi sur des cibles mélodiques alignées avec la transcription orthographique. Tout d'abord, on peut qualifier le registre de voix du locuteur en utilisant trois symboles : T (Top), M (Mid) et B (Bottom). Ensuite, on note les variations de F0 d'un point de la courbe à un autre grâce aux symboles suivants : H (Higher), L (Lower), S (Same). Les symboles D et U s'ajoutent pour décrire les phénomènes de « downstep » et « upstep ». La distinction entre les tons itératifs et non itératifs peut être basée sur deux critères. D'abord un critère configurationnel : H et L représentent les sommets et les fossés (« Peaks and valleys »), alors que D et U sont des plateaux dans des séquences ascendantes ou descendantes. Il y a aussi une différence scalaire puisque H et L correspondent à des intervalles fréquentiels plus larges que U et D (cf. figure ci-dessous).

....		Montant	Neutre	Descendant
ABSOLU		T [⇑]	M [⇒]	B [⇓]
RELATIF	*Non-itératif*	H [↑]	S [→]	L [↓]
	Itératif	U [<]	·	D [>]

Figure 22 : Le système INTSINT.

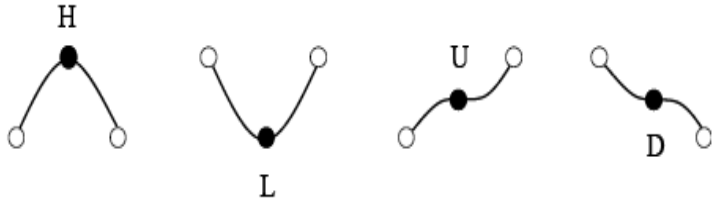


Figure 23 : Configurations correspondant aux symboles relatifs.

Hirst (2001) propose une deuxième version d'INTSINT dans laquelle l'encodage des points cibles dépend de deux paramètres : la clé (key) et l'étendue tonale (« range ») dépendants du locuteur et de l'unité intonative. Le mot-clé (comme une clé musicale) fait référence au point de repère défini par une valeur absolue de F0. C'est un point de référence mélodique en valeur

absolue. Le mot « étendue » détermine l'intervalle (en octaves) entre les valeurs minimales et maximales dans un énoncé. T et B représentent donc les limites de l'étendue de la voix du locuteur et M la valeur centrale ou moyenne et sont calculés de la façon suivante, en octaves :

$$T = key + range/2$$
$$M = key$$
$$B = key – range/2$$

Puis la valeur des tons relatifs est définie en fonction de l'intervalle entre le point cible précédent (P_{i-1}) et les valeurs de T et B selon les équations suivantes :

$$H = P_{i-1} + 0,5 * (T - P_{i-1})$$
$$U = P_{i-1} + 0,25 * (T - P_{i-1})$$
$$S = P_{i-1}$$
$$L = P_{i-1} + 0,5 * (B - P_{i-1})$$
$$D = P_{i-1} + 0,25 * (B - P_{i-1})$$

Une séquence de cibles tonale [M T L H L H D B], en supposant une valeur de clé de 240 HZ et une étendue d'un octave pour une locutrice donnée, serait convertie en points cibles comme il suit :

[240 340 240 286 220 273 242 170]

Ces points cibles peuvent ensuite être utilisés pour générer une courbe quadratique pour cet énoncé.

C'est donc un système d'annotation prosodique réversible. À partir d'une représentation phonologique de surface en tons INTSINT, on peut obtenir une représentation phonétique en points cibles MOMEL. Ainsi il est possible de proposer un modèle formel et de le comparer avec des données modélisées avec MOMEL.

3.3.3 Annotation des fonctions prosodiques : StarBI et IF

Une première tentative de création d'un système de transcription purement fonctionnel peut être dérivée du système TOBI. On parle d'un « Toneless TOBI » ou « StarBi » (cf. Wightman *et al.* 2000). Il s'agit d'extraire les informations purement fonctionnelles du système TOBI qui comprend comme on l'a vu plus haut, des informations formelles et fonctionnelles. D'après Hirst (2005), il est beaucoup plus évident de parvenir à un bon taux d'accord entre différents transcripteurs lorsque l'on se base sur l'aspect fonctionnel de la prosodie.

On retiendrait ainsi de TOBI, la notation des frontières et des proéminences. Cette notation reste toutefois très rudimentaire et demande à être enrichie. Elle a été mise en pratique par Wightman & Rose (1999), qui ont demandé à des sujets non-linguistes de faire un travail d'annotation des

proéminences principales et des frontières et ont pu constater que les taux d'accord entre annotateurs étaient largement supérieurs de cette manière.

Un autre système d'annotation fonctionnel plus complet a été développé dans les travaux de Daniel Hirst en 1974 et publié dans l'ouvrage « *Intonative Features* » (1977). Ce système s'inscrit plutôt dans la continuité avec les théories britanniques de l'intonation.

Ce qui motive la création d'un tel système est la recherche de la grammaire d'une langue définie comme un système de règles internalisées établissant des liens entre les sons et le sens :

> « *[…] a grammar being defined as an internalised system of rules which establishes a certain relation between sound and meaning.* » Hirst (1977, p. 464)

Hirst (1977) cherche alors à recenser et à analyser les cas où un changement au niveau de la forme prosodique peut être mis en relation avec un changement au niveau du sens ou un contraste. Il prend l'exemple de l'usage des paires minimales pour établir un inventaire des phonèmes d'une langue. Au niveau de l'intonation et par opposition à la plupart de ses prédécesseurs (O'Connor & Arnold, Pike, etc.) qui lient les variations de l'intonation à l'expression des attitudes, il se place au niveau de la syntaxe ou plus précisément des ambiguïtés syntaxiques pour développer son système de représentation. Il avance que cinq traits intonatifs sont suffisants pour rendre compte de toutes les ambiguïtés syntaxiques pour lesquelles un changement d'intonation implique un changement de sens. Ces traits sont codifiés sous la forme d'une annotation appelée IF pour « Intonative Features » ou « Intonative Functions ». On compte les traits {accent, frontière, centre, terminal, contraste}. À l'origine, ces traits étaient codifiés de façon identique aux traits distinctifs utilisés dans la phonologie structurelle ([±emphatic], [± terminal]). Dans les publications plus récentes, on utilise une représentation plus compacte (cf. intitulé de la figure 24).

Ce système permet ainsi de distinguer quatre niveaux de proéminence : inaccentué [‘], nucléaire [°] et emphatique [_] et deux types de frontières : frontière terminale [||] et non terminale [+]. Le trait « accentué » permet de marquer les accents mélodiques ou non d'un énoncé, le trait « centre » annote la proéminence nucléaire de l'énoncé qui peut être plus marquée ou contrastive (« _ »). Le trait terminal indique la présence d'une frontière et son effet : l'énoncé est terminé. Le trait non terminal indique une rupture accompagnée de la perception d'un énoncé non achevé (trait continuatif). Ce trait permet notamment dans de nombreux cas, de distinguer les contours mélodiques descendants des contours mélodiques montants.

Dans Hirst (2005), l'exemple suivant est donné *:*

|If you 'can't °lift it + 'ask °Peter to 'help you ||

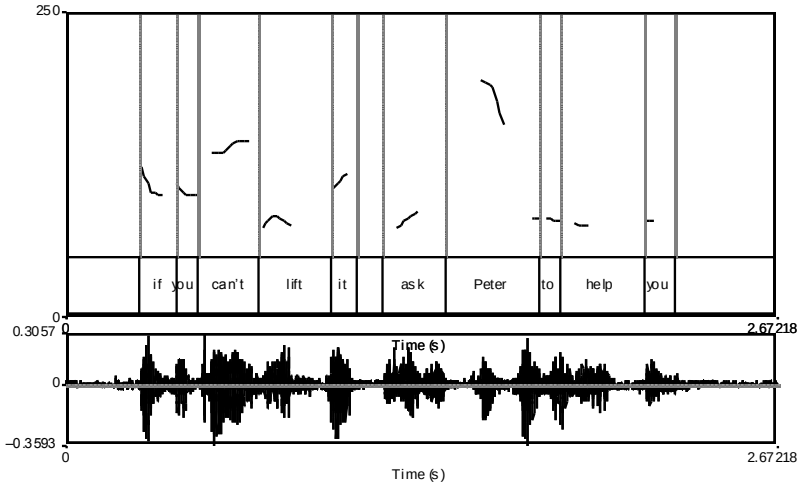


Figure 24 : *Signal et F0 pour « If you can't lift it ask Peter to help you » (extrait de Hirst 2005, p. 339)*

Hirst y ajoute aussi une représentation sous forme d'arbre phonologique pour le même exemple :

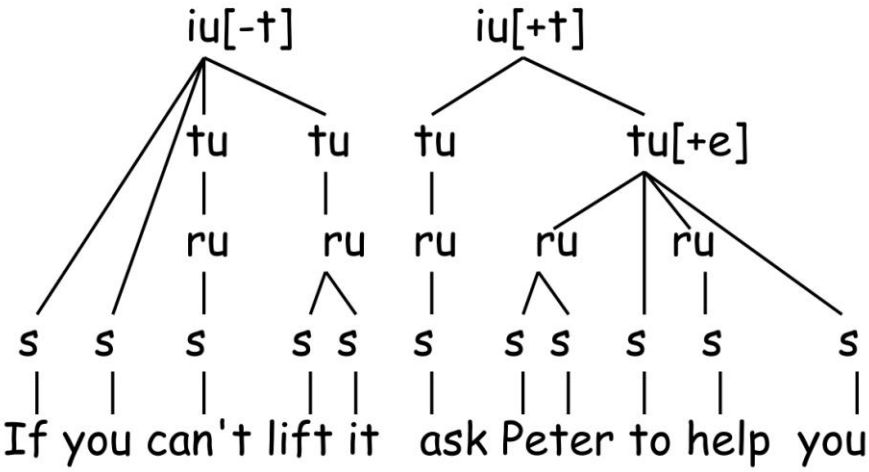


Figure 25 : *Représentation fonctionnelle IF sous forme d'un arbre phonologique (extrait de Hirst 2005, p. 340).*

L'annotation du type de frontière se fait donc au niveau le plus large correspondant à une unité intonative. Le niveau suivant est celui des unités tonales correspondant au domaine de l'accent mélodique qui peut se décomposer en unités rythmiques qui dépendent elles-mêmes de l'accent lexical.

Dans son ouvrage, Hirst montre donc que cette annotation permet de rendre compte de la contribution de la prosodie à l'interprétation syntaxique d'un énoncé.

Il ajoute aussi qu'il manque évidemment une annotation des phénomènes liés aux attitudes et à l'affect, mais que ces éléments pourront se rajouter par la suite aux fonctions syntaxiques révélées dans ce système sans qu'il n'y ait une contradiction mais plutôt un enrichissement du système.

C'est un point de départ très intéressant à plus de recherches qui compléteront une annotation fonctionnelle de ce type et c'est le cadre que nous adoptons dans cette recherche.

3.4 Conclusion

Ce chapitre a été consacré à la question de l'annotation prosodique de l'anglais. Nous avons montré que les premiers systèmes d'annotation étaient essentiellement de nature iconique et représentaient donc d'abord les formes prosodiques. L'évolution de ces systèmes a conduit à introduire beaucoup plus d'abstraction en réalisant une représentation conjointe des formes et fonctions prosodiques (représentation des contours mélodiques et des accents lexicaux par exemple dans la notation en têtards). Les avancées technologiques et la possibilité de travailler sur des corpus de plus en plus grands ont motivé le développement de systèmes informatisés comme le TSM et TOBI.

TOBI est le système d'annotation le plus largement utilisé aujourd'hui, mais nous avons émis des réserves quant à l'efficacité de ce système d'annotation dans l'optique de traiter un grand nombre de données à cause du coût temporel d'une telle annotation et du problème des taux d'accord entre différents annotateurs (cf. Wightman 2002).

Enfin suivant Hirst (2005), nous avons vu que les difficultés rencontrées dans ce type d'annotation pourraient être évitées en séparant la représentation de l'information formelle de la représentation de l'information fonctionnelle. Wightman (2002) a en effet montré que les taux d'accord entre annotateurs et le temps d'annotation étaient considérablement diminués en utilisant une version de TOBI purement fonctionnelle (en annotant les frontières et les proéminences). Un cadre d'analyse qui rend cette distinction possible a ensuite été présenté.

Nous séparerons donc l'annotation des formes et fonctions prosodiques en basant l'annotation de nos corpus sur le système IF du point de vue fonctionnel, car il s'agit du système d'annotation fonctionnelle le plus complet à notre connaissance, et le système MOMEL-

INTSINT du point de vue formel. L'annotation fonctionnelle sera réalisée au niveau des deux unités prosodiques de base définie par Hirst (2004) : l'unité intonative et l'unité tonale[2].

L'utilisation de MOMEL-INTSINT n'est pas obligatoire dans un tel travail, mais nous apparaît comme un point de départ et un cadre d'analyse prometteur. Ce système de modélisation formelle est quasiment dépourvu d'aprioris théoriques ce qui en fait un cadre d'évaluation idéal pour les approches théoriques de l'anglais britannique. Le système d'annotation INTSINT est par conception multilingue et il n'est pas nécessaire de connaître l'inventaire des contours phonologiques d'une langue pour l'utiliser, ce qui laisse place à une libre exploration des données. Enfin le système de modélisation et de codage est réversible ce qui ouvre la voix à une évaluation subjective des modèles de prédiction que l'on choisit de tester en utilisant la resynthèse avec Psola qui est un système de resynthèse automatique implémenté sous Praat.

Après avoir abordé la question des fonctions attribuées à la prosodie, celle des modèles de représentation des formes et les systèmes de transcription qui en découlent, il reste à évoquer les travaux expérimentaux s'intéressant à l'explicitation de l'articulation fonction/forme.

---

[2] Traduits de l'anglais « Intonation Unit » et « Tonal Unit ». Tout au long de notre mémoire nous ferons référence à ces deux unités prosodiques en utilisant les abréviations IU et TU correspondant aux termes anglais.

Chapitre 4

Différentes approches de l’articulation forme/fonction

Si de nombreuses fonctions prosodiques sont quasi-universelles comme la marque d'identité lexicale, de proéminence ou de frontière (Hirst et Di Cristo 1998), nous n'utilisons pas dans toutes les langues les mêmes formes pour exprimer les mêmes fonctions (Hirst 2005). L'articulation forme/fonction doit donc être explicitée au niveau d'une langue spécifique.

> *« First of all, many prosodic functions may be quasi- universal, in nearly all languages prosody contributes in some way to lexical identity (via tone, quantity and accent), expressing prominence, boundaries non-finality etc […] The mapping between form and function, however is certainly not universal. »* Hirst (2005, p.337)

Après avoir présenté différentes fonctions attribuées à la prosodie dans le premier chapitre et exploré la question des représentations formelles dans les deux suivants, nous nous attacherons donc dans ce dernier chapitre à présenter une analyse synthétique des différentes approches de la recherche sur l'articulation forme/fonction. Les questions que nous nous posons sont les suivantes : comment les différents chercheurs expliquent-ils la relation des représentations formelles à l'expression du sens ? quelles sont les méthodes utilisées afin de mener une recherche sur la relation entre les représentations formelles et les représentations fonctionnelles ? comment peut-on s'inspirer des résultats de ces différents travaux de recherche et définir notre propre méthode d'analyse ?

Nous distinguerons deux grands types d'approches : les approches théoriques et les approches expérimentales.

4.1 Les approches théoriques : « Top down approaches »

4.1.1 L'approche syntaxique ou grammaticale

Nous faisons ici référence aux modèles de l'intonation britannique déjà évoqués du point de vue formel dans le second chapitre, c'est-à-dire l'approche globale et l'approche des tons nucléaires. D'après la première approche, le choix de l'intonation d'un énoncé dépend de certaines caractéristiques grammaticales ou syntaxiques telle que la modalité. On attribue à un énoncé en entier une signification précise qui peut être d'ordre syntaxique ou grammatical (cf. l'intonation de la question), ou d'ordre sémantique ou « expressif » (surprise). C'est ce que l'on retrouve dans les travaux de Jones (1909) et Armstrong et Ward (1931).

Pour les partisans de l'approche basée sur les contours nucléaires, la relation au sens paraît plus complexe au premier abord, car les contours sont divisés en différents éléments (« Prehead, Head, Nucleus, Tail ») jouant potentiellement un rôle particulier. Les différents tons mélodiques décrits dans ces modèles sont au cœur de la construction du sens. Dans l'analyse de Faure (1970), on trouve les caractéristiques suivantes :

- La chute inférieure ou « low fall » (médium au grave) : c'est le ton de l'objectivité qui peut exprimer la neutralité et le détachement car sur le plan physique, elle exige un minimum de tension ; on l'utilise dans l'affirmative simple, les questions en WH ;
- La grande chute ou « high fall » (aigu au grave) : elle exprime un caractère achevé d'un énoncé qui se suffit à lui-même ; elle exige une plus grande tension et peut donc marquer plus d'implication de la part du locuteur comme l'expression de l'emphase ou du contraste ;
- La montée inférieure ou « low rise » (grave au médium) : elle porte un caractère d'inachèvement sur le plan de la réalisation concrète : la tension nerveuse du locuteur se maintient durant toute l'émission du message ;
- Le ton creusé « fall rise » est considéré comme une version emphatique de la montée inférieure, car dans cette analyse, c'est le mouvement final qui prime et la chute précédant ce mouvement vient intensifier le mouvement ;
- Le ton circonflexe ou « rise-fall » est aussi une variante de la chute inférieure, rendue plus particulière par la présence d'une montée sur le nucléus.

L'apport sémantique de chaque type de contour semble être mis en relation avec les caractéristiques de sa réalisation concrète. On met notamment en relation un contour qui exige plus d'effort tel que la grande chute avec l'expression d'une plus grande implication de la part du locuteur.

Peter Roach (1991) décrit la signification des tons mélodiques en relation directe avec l'attitude du locuteur en affirmant que :

- La chute exprime la finalité :

  Iam absolutely ↓certain.

- La montée fait référence à ce qui est inachevé (questions générales, listes, encouragements..) :

  Can you ↑help me ? ↑Yellow, ↑brown,↑ red…

- Le contour creusé exprime l'incertitude, le doute :

  You may be ˅right,

- Le contour circonflexe marque la surprise, le fait d'être impressionné par quelque chose :

  You were ^ first !

Toutefois, malgré quelques tentatives de donner un sens précis aux contours nucléaires dans ces approches, la relation à la signification est plus généralement établie de façon globale comme leurs prédécesseurs au niveau d'un énoncé en entier.

> *« Tone group 1 (falling tone) is sometimes assumed to express some sort of finality, but the exceptions are so numerous that it would be unwise to formulate any general rule to this effect. We shall consider the several values of tone group 1 according to the nature of its heads. »* Palmer (1924, p.72)

Ainsi Palmer analyse les différentes utilisations des tons mélodiques en fonction du contour de tête qui précède : une chute précédée d'une tête inférieure sera utilisée dans les déclarations catégoriques ayant un caractère conclusif, ou dans les questions en WH par exemple :

> *[What's a ↓table* ?] > *It's a piece of ↓furniture.*

Les travaux d'O'Connor & Arnold (1961) donnent lieu au même type d'analyse globale de l'apport sémantique de l'intonation.

Halliday (1970) quant à lui, part d'une analyse sémantique des tons mélodiques de base qui sont d'après lui la chute et la montée. Il leur attribue une valeur sémantique générique à partir de laquelle il déduit toutes les différentes variations possibles dans l'utilisation de ces contours. Cette valeur sémantique est liée à l'expression de la certitude et de l'incertitude. Ces valeurs affectent la polarité de l'énoncé en anglais :

> *« Basically, a falling contour means certainty and a rising contour uncertainty. […] In English, it takes this particular form: a falling contour means certainty with regard to yes or no. We go down when we know whether something is positive or negative, and we go up when we do not know. In other words we go down when we know the polarity of what we're saying. »* Halliday (1970, p. 23)

Cela implique que l'on utilise le contour descendant dans les affirmations et dans un type de question : les questions en WH. En effet, on ne s'attend pas dans ces cas précis à une réponse par oui ou non donc la polarité n'est pas mise en doute. Dans les questions appelées « Yes/No question, » on utilisera donc un contour montant car une incertitude existe quant à la polarité de la réponse. Le modèle de Halliday s'inscrit ainsi dans le même mouvement de pensée que ses prédécesseurs mais constitue un niveau d'abstraction plus important.

Cette approche de l'apport sémantique de l'intonation est encore une fois très utilisée dans l'enseignement de l'anglais seconde langue. Sur le plan de la recherche cependant, elle fait l'objet de plus de critiques. Cette approche est considérée comme réductrice. En effet il existe beaucoup plus de façons d'interpréter un message, beaucoup de connotations ou d'implications qu'il n'existe de contours intonatifs. Ainsi un contour intonatif donné ne peut pas être associé à une structure grammaticale unique. On doit pouvoir exprimer quantité de significations

différentes avec un contour identique en fonction du contexte plutôt que de la grammaire comme l'avance T'Hart (1984) :

> *« There are et least ten times as many meanings or rather implications or interpretative possibilities as there are different intonation patterns, so that at least a given pattern must correspond to a set of implication instead of only one. »* T'Hart (1984 , p.198)

Pike (1945) partage ce point de vue et souligne les dangers de ces approches syntaxiques qui d'après lui peuvent être trop générales et ne pas représenter la véritable richesse et complexité de l'apport sémantique de la prosodie dans la communication orale. Il fait référence au fait que des généralisations sont faites à l'issue de recherches menées sur des corpus trop étroits et des contextes trop spécifiques. C'est une critique de la méthode d'investigation utilisée dans ces approches :

> « *An error -which has vitiated much work in the past-, lies in the failure to get the common meaning from a large enough number of contexts. By abstracting a particular contour just from a single context, or from contexts which are all grammatically or physically similar even although that contour occurs elsewhere in grammatically and physically diverse contexts, one tends to assume that the meaning is much more concrete than it actually is[..] »* Pike (1945, p. 23)

Certains théoriciens se tournent alors vers une analyse de l'apport sémantique de l'intonation encore plus abstraite et qui s'inscrit dans le discours et la pragmatique plutôt que dans la grammaire et la syntaxe.

4.1.2 Approches discursives et pragmatiques

On peut retrouver cette approche dans les travaux de Jassem (1952), Brazil (1975) et Gussenhoven (1983 et 1984). Si les auteurs mentionnés précédemment s'intéressent à la signification des tons mélodiques et à l'interprétation des combinaisons possibles entre les différents éléments prosodiques, ceux-ci se dirigent vers une analyse plus abstraite et généralisante. Cruttenden (1986) oppose ainsi les théoriciens étudiant l'apport sémantique au niveau local et ceux qui proposent une analyse plus globale. Il souligne le fait qu'aucun chercheur n'ait tenté de mettre en relation les apports sémantiques au niveau local et les analyses sémantiques plus abstraites :

« Firstly, no analyst has systematically related abstract to local meanings, rather, some have concerned themselves principally with the description of local meanings and some with the characterization of abstract meaning. » Cruttenden (1986)

Ces modèles ou systèmes de l'intonation plus abstraits ne sont pas développés à partir de six à huit contours nucléaires comme les précédents par exemple mais réduisent leur analyse à deux ou trois contours de base.

Ainsi, l'idée centrale développée par Brazil (1975) dans son ouvrage est que la meilleure façon de rendre compte de l'apport sémantique de la prosodie est de se placer au niveau du discours :

> *« Intonation is more satisfactorily accounted for at the level of discourse than at the level of grammar. »* Brazil (1975)

En s'appuyant sur la thèse de Jassem (1952), Brazil définit ce qu'il appelle les « fonctions internes » des contours mélodiques. A la base de cette description se trouvent les contours mélodiques descendants (« falling tones ») et descendants-montants ou creusés (« falling-rising tones »). Un contour neutre est aussi ajouté pouvant prendre la forme d'une petite montée ou d'un ton statique. Chacun de ces tons a une valeur intrinsèque de base à partir de laquelle des variations sémantiques peuvent être développées. La chute a une valeur de « proclamation », ce contour signale souvent la présence d'une information nouvelle alors que le ton creusé fait référence à une information précédemment connue (c'est l'opposition entre : « proclaiming tone *vs* referring tone »).

Exemples d'annotation avec ce système d'analyse :

> */ p* Kate */ r washed the* curtains*/*

(/=frontière du groupe intonatif, p= proclaiming tone, r= referring tone, italique= nucléus)

Dans cet exemple, on peut imaginer que le locuteur répond à la question: « *Who washed the curtains?* ». « *Kate* » porte un ton de proclamation en tant qu'information nouvelle alors que le reste de l'énoncé est déjà contenu dans la question d'origine donc annoté « r ».

Pour chacun de ces tons, il existe une version dite « intensifiée » ou plus marquée. Cela indique un engagement plus important de la part du locuteur. On obtient alors le système tonal suivant :

« Referring tone » → Non marqué : contour creusé (chute-montée)
« Referring tone » → Marqué : contour montant

« Proclaiming tone » → Non marqué : chute
« Proclaiming tone » → Marqué : contour de montée-chute

Contour neutre → petite montée

À ce système s'ajoute la notion de clé ou registre dans tout énoncé dont l'importance est soulignée très tôt par Sweet :

> *« Each sentence or sentence group has a general key or pitch of its own: high, mid or low. »* Sweet (1906) dans Brazil (1975, p. 9)

On peut donc qualifier le registre dans ce type d'annotation en se basant sur une opposition catégorielle entre deux types de registre : le registre neutre et le registre marqué. On obtient les trois niveaux de registre aigu, médium et grave, classés en fonction de leur valeur phonologique. Les fonctions internes du registre de voix se résument ainsi :

Valeur neutre ➔ registre médium

Valeur marquée ➔ contrastive : registre aigu
égale : registre grave

Plus récemment, Gussenhoven (1983) a proposé une analyse de l'intonation basée sur le discours et la pragmatique. Il rappelle que l'étude de l'intonation et du sens reposent sur deux éléments : la segmentation ou la manière dont le chercheur divise le contour intonatif en unités distinctives et le regard porté sur l'apport sémantique de la prosodie dans un message. Son approche est résolument pragmatique comme l'illustre l'affirmation suivante :

> *« Part of what intonation is used for is to mark the status of the information conveyed by the sentence with respect to a background hypothesised by the speaker. »* Extrait de Gussenhoven (1984, p. 200)

Le système prosodique qu'il développe repose sur trois contours nucléaires : la chute, la chute-montée et la montée. Ce qui est mis en valeur dans cette analyse (et par là même marqué par un ton mélodique) est la contribution du locuteur à la conversation. Hors « focus » se trouve le « background », la connaissance partagée par les locuteurs ou contexte. Les éléments mis en relief sont appelés variables alors que les autres sont étiquetés « background » ou contexte. D'après Gussenhoven (1984), le choix d'un contour nucléaire permet donc une manipulation des variables en fonction du contexte.

Le premier ton est associé à l'addition : le locuteur ajoute une nouvelle information dans un contexte précis. La signification de l'annotation « V-addition » peut être paraphrasée de la façon suivante. Le locuteur signale :

> *« Je veux que tu saches qu'à partir de maintenant cet élément fait partie de notre connaissance commune »* Gussenhoven (1984, p. 384).

Le second ton est associé à la « sélection » : le locuteur sélectionne une variable dans l'arrière plan commun ce qui peut être paraphrasé ainsi :

> *« Je veux que tu prennes conscience du fait que cet élément linguistique fait partie de notre connaissance commune »*. Gussenhoven (1984, p. 384).

Enfin, le dernier ton est associé à la pertinence ou au test : le locuteur choisit de ne pas s'impliquer en n'indiquant pas si la variable ou information appartient ou non à l'arrière plan commun :

« Je te laisse libre de décider si nous devrions établir cette variable comme une partie de notre arrière plan commun. »

Gussenhoven illustre ces différentes significations grâce aux exemples suivants :

The \house is on fire (I am informing you)

The \house is on /fire (I am confirming to you that it's the house that's on fire not something else for instance)

The/ house is on fire? (I am asking you if that's true)

Des variations de hauteur, registre et dans la complexité des contours sont aussi prises en compte comme des variations graduelles.

Les tons peuvent être orientés vers le locuteur ou son interlocuteur « speaker serving » ou « hearer-serving » comme le montrent les exemples suivants (Cruttenden 1986) :

Of `course/ he's an `expert (I remember now)

,See/ he's an `expert now (Now you must believe me)

Le premier énoncé peut être interprété comme étant orienté vers le locuteur qui se rappelle qu'un certain individu est un expert alors que le second exemple vise à impliquer l'interlocuteur, ce qui est marqué par le contour montant sur « *see* » (« tu vois ? »), et la chute sur la première syllabe de « *expert* » (qui confirme l'information).

En conclusion, ces deux approches pragmatique et discursive sont très intéressantes, car elles permettent de réduire le système de l'intonation anglaise à trois tons nucléaires seulement à partir desquels on développe toute une grammaire de la relation au sens. Elles représentent une rupture radicale avec les approches grammaticales de l'intonation. Cependant les théoriciens leur reprochent de ne pas permettre d'interpréter tous les contours intonatifs et contextes associés. De plus ces systèmes d'annotation supposent une prise de recul importante par rapport à ce qui est dit pour analyser ce qui est supposé être « proclamé ou déjà connu », il est nécessaire de connaître les intentions du locuteur, ce qui n'est pas chose facile dans un travail d'annotation de corpus.

D'autre part, une autre approche discursive de l'intonation a été développée à partir du modèle autosegmental métrique de Pierrehumbert (1980), il s'agit de l'approche compositionnelle.

4.1.3 L'approche compositionnelle

Cette approche a été développée à partir des avancées apportées à la théorie autosegmentale évoquée dans une partie précédente. Elle est essentiellement représentée par les travaux de J. Pierrehumbert et J. Hirshberg (1990). Celles-ci se distinguent des théoriciens de l'intonation

anglaise présentés précédemment avec lesquels elles sont en désaccord. Ces derniers (tels que O'Connor & Arnold 1961, Liberman 1975…) associent les contours mélodiques à l'expression d'attitudes telles que la politesse, la surprise, la séduction ou à l'expression d'émotions telles que la colère ou la haine ou bien encore aux actes de langage comme les requêtes, les affirmations (Liberman & Sag 1974). Or Pierrehumbert et Hirshberg (1990) pensent qu'il est plus judicieux pour rendre compte de l'apport sémantique de l'intonation de se placer au niveau du discours pour pallier au fait qu'un même contour mélodique peut être associé à une multiplicité de significations différentes selon le contexte :

> *« Though speaker attitude may sometimes be inferred from choice of a particular tune, the many-to-one mapping between attitudes and tune suggests that attitude is better understood as derived from tune meaning interpreted in context than as representing that meaning itself. »* Pierrehumbert et Hirshberg (1990, p. 284)

Une autre critique fondamentale faite aux modèles précédents est celle du domaine utilisé pour étudier l'apport sémantique de la prosodie. C'est une critique envers l'approche holistique qui met en relation des phrases intonatives avec des significations et connotations variées. Ainsi d'après Pierrehumbert (1990), il est plus évident d'analyser la signification d'un énoncé en entier que de tenter d'extraire l'apport spécifique de chacun de ses éléments, mais une telle approche ne permet pas d'atteindre un niveau de généralisation et d'abstraction suffisant pour l'élaboration d'un modèle phonologique de la relation forme/sens.

> *« This approach has probably been more a practical matter than a theoretical conviction: it is difficult to identify the "meaning" of parts of a contour until one has some idea of what the meaning of the whole unit might be. »* Extrait de Pierrumbert et Hirschberg (1990, p. 285)

L'interprétation de la signification de l'intonation dans l'approche de Pierrehumbert et Hirschberg (1990) est résolument compositionnelle. Chaque catégorie accentuelle : « pitch accents, phrase accents, boundary tones », apporte une part de sens à l'énoncé et contribue à faciliter l'interprétation du locuteur. Ce sont donc des morphèmes intonatifs dans les termes de Pierrehumbert (1990).

Les locuteurs utilisent un contour mélodique pour exprimer un lien particulier entre le « contenu propositionnel » réalisé dans la phrase intonative et les croyances mutuelles des différents participants de la conversation.

De façon plus précise, ce qui est mis en valeur est ce que le locuteur pense être mutuellement partagé et connu par lui-même et son interlocuteur : en utilisant certains contours mélodiques, le locuteur a pour intention d'ajouter un élément aux croyances de son interlocuteur, avec d'autres il n'ajoute rien de nouveau.

Les accents mélodiques (H*, L*, L*+H, L+ H*) transmettent des informations relatives aux référents, aux « modifieurs », aux prédicats et aux divers liens existant entre ces derniers qui sont eux-mêmes indiqués dans le lexique.

Les tons de phrase (H- et L-) expriment le degré de dépendance ou le lien existant entre les énoncés intermédiaires alors que les tons de frontière (H% et L%) nous renseignent sur le lien existant entre les phrases intonatives.

Les accents mélodiques sont généralement utilisés pour mettre en relief un élément.

H* marque une information nouvelle, L* marque un élément qui doit être mis en valeur, mais ne fait pas partie de ce que le locuteur avance comme information nouvelle dans l'énoncé.

En combinant les différents types d'accents mélodiques, on peut obtenir les interprétations suivantes :

- H*L-L% : le contour des déclaratives non-marquées :
  He didn't EAT it.
- H*H H% : une information est présentée et mise en question simultanément :
  (to a receptionist) I'm the enginEER ?
- L* H-H% : le contour traditionnel des questions fermées « Yes/No questions ») :
  Has he gone aWAY ?
- L*L-H% : poser une question à propos de quelque chose que l'interlocuteur doit savoir :
  I don't eat PORK.
- L*+HL-H% incertitude par rapport à un élément :
  He's a good BADminton player
  Ward and Hirschberg (1985)

Le ton de phrase H- indique que la phrase intermédiaire fait partie de la même unité d'interprétation que la phrase intermédiaire suivante. Il y a une interdépendance entre les deux phrases intermédiaires. L-, au contraire met en relief la séparation entre les phrases intermédiaires consécutives. Exemples extraits de Cruttenden (1986):

Do you want apple juice or orange juice?

H H-/ H* L-L%*

vs

Do you want apple juice or orange juice?

H L-/ H* L-L%*

Dans le premier exemple, « *apple juice* » et « *orange juice* » font partie d'une même entité alors que dans le second, on a deux entités distinctes.
En ce qui concerne les tons de frontière, H% indique le fait que la phrase intonative doit être interprétée en relation avec la phrase intonative suivante, alors que dans le cas d'un ton L%, la « direction » d'interprétation n'est pas spécifiée.

Exemples :

1. My new car manual is almost unreadable.

It's quite annoying.

*H*L-H%*

I spent two hours figuring out how to use the jack.

2. My new car manual is almost unreadable.

It's quite annoying.

*H*L-L%*

I spent two hours figuring out how to use the jack

Cruttenden (1986).

Dans (1) le contour mélodique indique un regard vers l'avant : « *forward looking* », on attend la suite. Dans (2), on a une valeur de regard vers l'arrière : « *backward looking* », cela donne une valeur de commentaire à l'énoncé « *It's quite annoying* ».

Enfin il faut noter que cette approche est à son début dans les travaux sur lesquels sont basés les exemples et explications donnés dans ce chapitre :

> *« In the following discussion we put forward a first approximation of a compositional theory of tune interpretation […].»* Pierrehumbert et Hirshberg (p. 270)

Elle ne fait pas l'unanimité comme le montre la citation suivante (Cruttenden 1986) :

> *« [..] it appears that, while decomposing contours into sequences may simplify speech synthesis, meanings are probably still more easily assigned to overall contours. »*

Plus récemment encore, cette approche de l'apport sémantique de la prosodie a été contestée dans les travaux de Dainora (2002) qui posait la question suivante :

> *« Does Intonational Meaning come from Tones or Tunes ? ».*

Autrement dit quelle est la méthode la plus pertinente pour étudier l'apport sémantique de la prosodie : la méthode holistique qui se place au niveau du contour en entier ou l'approche compositionnelle qui étudie chaque ton mélodique individuellement. Elle s'interroge alors sur l'affirmation théorique qui dit que chaque ton contribue de façon autonome à l'interprétation d'un énoncé. Elle se concentre dans cette recherche sur la nature de l'interaction entre les accents mélodiques et les accents de frontière. Le choix d'un accent mélodique a-t-il une influence sur le choix des accents de frontière qui lui succèdent ? Dainora fait l'hypothèse suivante : si les tons ne sont pas choisis de façon indépendante les uns des autres alors la signification de l'énoncé n'est pas construite à partir de la somme des significations de chaque ton :

> *« If tones are not chosen independently then we can infer that meaning doesn't arise from the sum of the meaning of individual tones. »*

L'analyse d'un corpus de données par un modèle probabiliste Markov donne des résultats allant à l'encontre de l'approche compositionnelle. En effet, les données montrent que le choix de l'accent nucléaire influence de façon déterminante celui du ton de frontière associé.

Après un accent nucléaire L*, le ton de frontière associé est haut (H%) dans 83% des cas alors que le ton de frontière apparaissant après un ton L*+H n'est haut que dans 17% des cas.

Ces résultats ont aussi été confirmés par des tests d'information mutuelle. Le choix de l'accent mélodique est aussi en relation étroite avec celui des tons de phrase d'après cette étude. La prise en compte de l'accent mélodique et des tons de phrase permet de prédire certains tons de frontière.

En conclusion, Dainora (2002) démontre que les tons qui composent un contour mélodique sont liés d'une manière qu'un modèle qui attribue une signification particulière à chaque élément ne peut pas rendre compte. Les tons se combineraient en contours prédictibles si bien qu'une approche holistique serait plus apte à représenter leur apport sémantique.

> *« The strong interrelations among tones suggest that the meaning of a tune is more than the sum of its tones. »*

> *« Tones behaving not like numbers but like phonemes in that they combine in predictable patterns. […] In sequences of tones, certain combinations are more common than others so that a "tunal approach" should be better. »* Dainora (2002)

4.2 Les approches expérimentales: « Bottom up approaches »

4.2.1 L'approche perceptive

Ce mouvement de recherche a été développé à l'institut de recherche sur la perception à Eindhoven. On se basera pour ce chapitre sur l'ouvrage de TE'Hart, Collier et Cohen publié en 1990, dans un premier temps et sur les travaux d'Alessandro et Mertens (1995) dans une seconde section.

Il s'agit d'une approche résolument expérimentale par opposition aux approches présentées précédemment qui sont ici critiquées pour leurs prérequis théoriques trop contraignants sur les données. En effet d'après ces chercheurs, une bonne théorie de l'intonation devrait comprendre une composante phonétique et une composante linguistique puis l'énonciation de liens entre ces deux niveaux. Le problème qui se pose alors est d'ordre méthodologique, comment doit-on s'y prendre ? On relève un « dilemme » initial à cette tentative, résumé dans la citation suivante :

> *« The student of intonation faces a dilemma: either he chooses the linguistic approach at the risk of overlooking phonetically important features, or he opts for an instrumental-phonetic angle, thus increasing the chance of missing the communicatively relevant essentials. »* T'Hart, Collier et Cohen (1990, p. 4)

Deux approches sont donc confrontées, l'approche «top down » et l'approche «bottom up », toutes deux étant à elles seules incapables d'aboutir à une théorie juste de l'intonation. La solution est donc pour ces chercheurs d'introduire entre l'approche instrumentale ou acoustique et l'approche linguistique, un modèle de l'interlocuteur ou « model of the listener ». Afin de ne pas se noyer dans les détails acoustiques et phonétiques, on introduit un « filtre » basé sur ce que l'interlocuteur perçoit. La question principale devient donc la suivante :

> *« What does the listener make of pitch in speech? »* (p. 4)

Afin de se lancer dans une recherche sur la façon dont la prosodie contribue au sens de la phrase, il faut utiliser un métalangage spécifique à ces phénomènes. Or à l'heure actuelle, il n'y a pas de consensus dictant la meilleure façon de décrire l'intonation. On se basera ainsi dans cette approche sur les limites de la perception de celui qui écoute en définissant des seuils physiologiques au-delà desquels les phénomènes étudiés seront considérés comme non pertinents. Les variations de F0 sont mesurées en demi-tons et non en Hertz afin de mieux représenter ce que l'interlocuteur perçoit. Il ne s'agit pas d'utiliser des valeurs absolues en hertz mais plutôt de représenter des écarts de hauteur perceptibles. À partir de cette représentation, une théorie de l'intonation est développée.

4.2.1.1 Vers une théorie de l'intonation basée sur la perception

Afin de définir un modèle de l'intonation du hollandais, on part d'abord d'une étude acoustique ou phonétique des données. L'objectif premier est de parvenir à classer, étiqueter et décrire les

phénomènes phonétiques pertinents dans la perception. Ce n'est que dans un dernier temps que ces formes mélodiques seront éventuellement mises en relation avec des fonctions abstraites.

La première étape consiste à observer les données du point de vue acoustique et à essayer d'en retirer une représentation simplifiée en fonction des limites de la perception. On considère par exemple deux types de phénomènes prosodiques essentiels : les variations de F0 volontaires et non volontaires. Les premiers sont étiquetés « perceptuellement pertinents » alors que les seconds sont considérés comme des détails non significatifs. L'étape de simplification ou de stylisation comprend donc l'omission des détails non pertinents. La courbe de F0 est ensuite décrite ou représentée par l'intermédiaire de lignes droites. La pertinence et l'efficacité de cette représentation sont testées par resynthèse en utilisant le système LPC (Linear Predictive Coding). Les stylisations ou « close copies » doivent être identiques aux patrons d'origine du point de vue auditif.

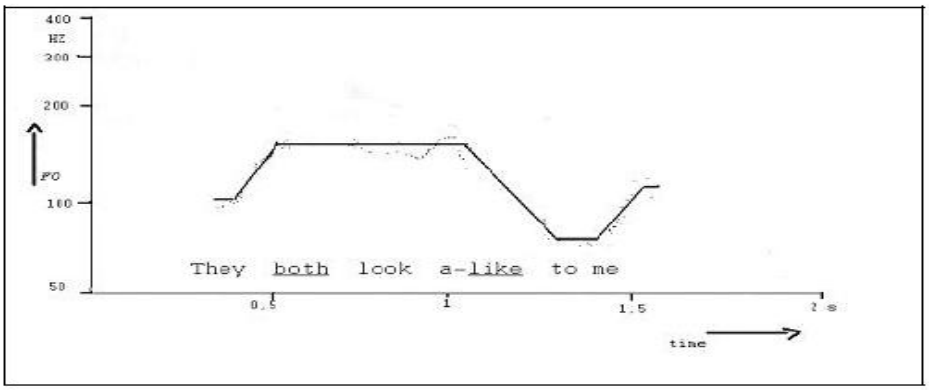


Figure 26 : Contour stylisé selon la méthode IPO.

À partir de ces représentations simplifiées de F0, ces chercheurs ont tenté d'établir une grammaire de l'intonation du hollandais et de l'anglais (Willem, Collier et TE'Hart 1983). Cela consistait à postuler les combinaisons possibles de contours définis à partir d'un premier corpus puis à confronter cette grammaire préliminaire à un autre corpus afin d'en voir les limites et de l'enrichir à nouveau.

À l'issue de différentes expériences menées dans ce cadre de recherche, différentes propositions théoriques ont été avancées, nous n'en relèverons que quelques-unes qui nous paraissent pertinentes pour notre thème de recherche. Les premières se rapportent à la définition de la structure de l'intonation du hollandais et les secondes aux interactions entre l'intonation et l'accentuation, la syntaxe ou plus généralement le sens.

4.2.1.2 Structure de l'intonation du hollandais

Tout d'abord, les variations de F0 sont à interpréter en termes de mouvements ou « glides » et non pas en termes de sauts ou niveaux. Toutes les variations pertinentes de F0 requièrent du temps pour être réalisées et perçues. Ce sont d'après ces expériences, les mouvements

mélodiques qui sont les unités minimales de perception et d'analyse, ce qui correspond aux théories holistiques de l'intonation.

On peut décomposer les mouvements mélodiques en éléments perceptifs selon leur direction, leur alignement par rapport aux frontières syllabiques, la vitesse de changement de F0 et leur taille.

À un second niveau de l'analyse de la mélodie, les mouvements mélodiques se combinent en « configurations ». Celles-ci peuvent être de trois types :

- *« Root configuration » :* la seule configuration essentielle tout comme le nucléus ;
- *« Prefix configuration » :* configurations possibles avant la racine tout comme les contours de tête ;
- *« Suffix configuration » :* configurations possibles après la racine correspondant à la queue.

Au troisième niveau ces configurations forment un contour en fonction des contraintes syntagmatiques. En Hollandais, on retient six patrons intonatifs de base qui correspondent aux six configurations « racines » existantes. Enfin les travaux menés sur le hollandais et l'anglais permettent de formuler quelques hypothèses quant à la façon dont on peut analyser l'articulation forme/fonction.

4.2.1.3 Intonation et accentuation, intonation et syntaxe, intonation et sens

Une des affirmations retenues au regard des expériences menées à l'IPO est la suivante : l'intonation prévaut sur l'accentuation pour déterminer la forme du contour mélodique. Les contours mélodiques ne seraient pas prédictibles en fonction du type de syllabe accentuée ou non, ce qui est en contradiction totale avec les modèles britanniques qui se basent sur le niveau de proéminence pour attribuer des configurations tonales possibles.

> *« Some pitch movements have the capacity to lend prominence to the syllable on which they occur and their position in a contour is flexibly adjusted so as to coincide with words or syllables that require a pitch accent. »* T'Hart, Collier et Cohen (1990, p.97)

La correspondance avec la syntaxe n'est ni unique ni obligatoire. Dans les énoncés contenant deux propositions, la frontière syntaxique peut être marquée au moyen de l'intonation mais le locuteur a toujours le choix de ne pas utiliser l'intonation. Quand la mélodie est utilisée comme marqueur de frontière en hollandais, il existe un choix entre trois configurations tonales. Ces configurations peuvent aussi apparaître dans des contextes syntaxiques différents. Cela confirme donc l'affirmation de Bolinger (1957, vol 8, p. 36) :

> *« Intonation operates in its own sphere, and the uses that grammar makes of it are catch-as-catch-can : the two do not go separate ways, but they do not cling together all the time either. »*

Ainsi dans de nombreux cas étudiés, ce sont les indices temporels comme la durée des syllabes finales qui indiquent la présence d'une frontière syntaxique. En ce qui concerne l'ambiguïté syntaxique, on remarque que c'est principalement l'accentuation et non l'intonation qui permet d'y remédier.

Enfin, ces chercheurs affirment que l'intonation n'a pas de signification intrinsèque par opposition à l'approche compositionnelle de Pierrehumbert & Hirschberg (1990).

> *« The number of basic intonation patterns per language is very limited, whereas the shade of meaning and attitude is nearly infinite. [..] Rather the relation between intonational and attitudinal features must be one to many, possibly even many to many. »* Pierrehumbert & Hirschberg (1990, p. 111)

Les difficultés relevées dans une recherche visant à mettre en relation les mouvements mélodiques et l'expression d'une attitude sont multiples. Quirk *et al*. (1964) et Crystal (1969) se sont heurtés à ces difficultés. Les différents termes relatifs aux attitudes et émotions peuvent renvoyer à des réalités différentes pour chaque locuteur, il est difficile d'isoler le rôle de l'intonation elle-même par rapport à la syntaxe, la sémantique et la pragmatique. La traditionnelle opposition entre les contours montants et descendants transformant une affirmation en question est aussi remise en question.

> *« Evidently, a rising pitch contour can turn a statement into a question, but the extent to which this can be done is determined by the lexico-pragmatic characteristic of the utterance itself, and even under the best pragmatic circumstances the effect of intonational cue remains limited. »* Pierrehumbert & Hirschberg (1990, p. 113)

Dans les travaux d'Alessandro et Mertens (1995) et ceux de Mertens (1987, 1993, 1997, 1999, 2004b), une approche perceptive de la modélisation de l'intonation est aussi développée.

4.2.1.4 Le prosogramme : un modèle perceptif de l'intonation

Le système de Mertens est basé sur une simulation de la perception du locuteur. C'est un système de transcription semi-automatique. Il requiert d'avoir un signal avec une transcription phonétique alignée. Il a été appliqué sur l'étude de plusieurs langues, dont la principale est le français.

Dans ce système, on prend le noyau vocalique ou syllabique comme unité de base. La transcription issue de ce modèle est considérée comme une estimation du contour perçu par l'interlocuteur moyen.

Les auteurs se penchent donc sur la question des limites de la perception de la mélodie. Ils montrent alors qu'une variation de F0 doit avoir une taille minimale pour être perçue. Cette taille varie en fonction de la fréquence de départ et de la durée du mouvement (la taille diminue quand la durée augmente). Un seuil de glissando a donc été défini. La variation de F0 est pour cela exprimée comme un intervalle mélodique en demi-tons (le demi-ton est une échelle musicale dans laquelle une octave est divisée en 12 intervalles égaux sur une échelle logarithmique). Le seuil de glissando est calculé de la manière suivante :

$G = 0{,}16/T^2$ (ST/s)
où G = Glissando, T = durée de la variation, ST = semi-tones.

Cette équation permet de quantifier le glissando pour les variations de F0 linéaires alors que de telles variations sont assez rares dans la parole continue. Les auteurs ajoutent alors le calcul du seuil de glissando différentiel pour rendre compte des contours dynamiques de F0 comme les chutes-montées. La question qui se pose est alors la suivante : quels changements de pente de F0 sont perceptibles ? La formule de calcul du glissando différentiel est la suivante :

$DG = G2-G1$

Où DG = « Differential glissando », G2 = valeur du glissando au point 1, G2 = valeur du glissando au point 2.

Ces formules correspondent au calcul du glissando pour un son isolé tel que la voyelle. Cependant quand un environnement plus complexe est pris en compte, le calcul du seuil de glissando devient plus problématique. En effet selon House (1990), une même variation de F0 peut être perçue différemment selon son alignement par rapport aux frontières syllabiques. Les changements simultanés d'intensité, de spectre et de voisement semblent empêcher la perception des variations mélodiques. C'est pourquoi les auteurs se basent dans leurs travaux sur une segmentation du signal au niveau du noyau syllabique.

Stylisation et transcription de la mélodie

Les contours de F0 sont stylisés suivant l'approche IPO. On obtient une représentation simplifiée de la courbe de F0 qui doit préserver les éléments pertinents du contour. Les paramètres du modèle sont représentés par des seuils psycho-acoustiques (les glissandos) qui sont appliqués aux variations simples de F0 (chutes, montées). Comme nous l'avons déjà mentionné, l'unité de base de la stylisation est le noyau vocalique. On obtient les deux représentations de base suivante :

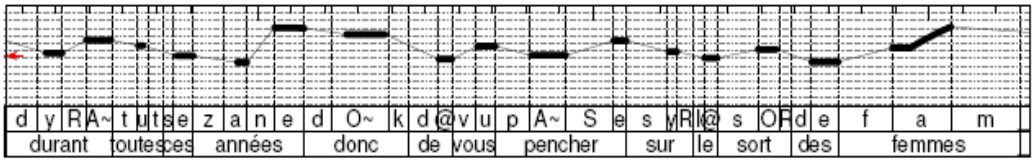


Figure 1. *Compact plain prosogram (using glissando threshold 0.32/T², cf. infra)*

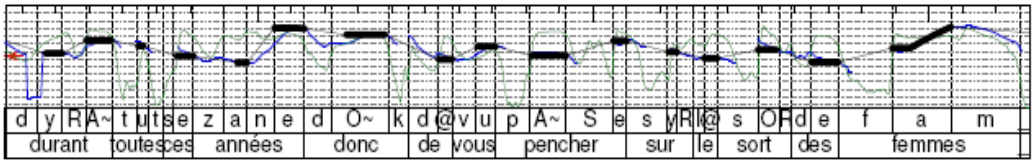


Figure 2. *Compact rich prosogram (using glissando threshold 0.32/T², cf. infra)*

Figure 27 : Transcription simple et riche du prosogramme (extrait de Mertens 2004).

Les lignes en pointilles sont séparées par deux demi-tons comme dans une partition de musique. On les utilise pour interpréter les intervalles mélodiques entre les syllabes ou les voyelles.

On peut paramétrer le registre en fonction du locuteur : une flèche au niveau de la frontière gauche indique la valeur de référence de 150 HZ qui est la valeur probable pour les hommes et les femmes. Dans la seconde représentation (dite riche) on peut visualiser les contours de F0 brute et les comparer avec les segments pertinents retenus par le modèle.

Les conventions de notation issues de ce modèle et développées dans Mertens (1987, 1993, 1997) permettent de distinguer quatre niveaux de hauteur de base :

1. Le ton bas (B) ;
2. Le haut (H) ;
3. L'infra bas (B-) ;
4. Le suraigu (H+).

B- et H+ représentent les extrémités de la tessiture d'un locuteur. Toutefois, le locuteur se sert essentiellement du registre usuel. L'interprétation des niveaux se fait progressivement de gauche à droite et localement en fonction des intervalles mélodiques entre syllabes successives ou entre les parties successives d'une même syllabe.

Les niveaux B et H peuvent être rehaussés ou abaissés suite à la présence d'un intervalle mineur (typiquement trois demi-tons). Ceci est indiqué par les signes « / » et « \ ».

Les niveaux de hauteur ne constituent donc pas des repères absolus, mais plutôt des paliers relatifs et locaux qui se définissent les uns par rapport aux autres.

Enfin, la relation au sens est étudiée de façon secondaire à la modélisation perceptive tout comme dans l'approche IPO. Il s'agit d'abord de décrire et analyser les contours mélodiques pertinents du point de vue de la perception et de les mettre en relation avec des fonctions d'ordre syntaxique ou structurel. Pour obtenir des contours naturels et variés, il est indispensable d'ajouter au texte des marqueurs pour signaler des aspects pragmatiques ou déclencher les formes prosodiques souhaitées. Ces marqueurs sont appelés des balises selon Mertens (2004a), par analogie avec les balises utilisées dans les documents hypertexte.

Le terme balise est emprunté à la terminologie du traitement automatique des langues et ces balises peuvent être de type linguistique, pragmatique, émotif, expressif, phonostylistique etc.

On a bien vu dans la présentation de cette approche, que l'on s'appuie sur des critères expérimentaux (les limites de la perception) pour établir un modèle de l'intonation et donc de l'articulation forme/sens. On se détache donc bien des modèles théoriques de l'intonation présentés précédemment. Dans la section suivante, on présentera une approche qui est encore plus ancrée dans l'expérimental : l'approche physiologique ou articulatoire.

4.2.2 Les approches physiologiques et articulatoires

Cette section est consacrée à la présentation d'une approche résolument originale et novatrice représentée dans les travaux de Fujisaki (1983, 1997, 2006), Gussenhoven (2002) et Xu (2002 et 2005). Ces chercheurs se sont intéressés tout comme leurs prédécesseurs à la relation entre les formes prosodiques et les fonctions. Ce qui fait leur originalité toutefois est la prise en compte dans leur analyse de l'interface forme/sens, de l'aspect physiologique ou articulatoire. On s'intéresse toujours ici à la relation entre la phonétique et la phonologie tout en intégrant à cette analyse bidimensionnelle une troisième dimension qui interagit avec les deux autres : la dimension physiologique.

> *« The understanding of speech melody can be improved by simultaneously taking into account two factors: that speech conveys communicative meanings and that it is produced by human articulators. »* Xu (2005)

4.2.2.1 Le modèle de Fujisaki

Fujisaki (1983) a développé une méthode de recherche et un modèle génératif des éléments prosodiques en travaillant tout d'abord sur la langue japonaise. Ce modèle a été par la suite largement appliqué à la modélisation d'autres langues dont l'anglais. Celui-ci met en valeur d'une part le contenu du message lors d'un acte de langage et d'autre part les canaux permettant de véhiculer ce message. Le rôle de la physiologie, de la physique et des mathématiques, est alors mis en valeur (Fujisaki 2006).

Son modèle est développé par une approche qui combine l'induction et la déduction, donc qui vise à réconcilier les « théoriciens de l'intonation » et les « mesureurs » dans les termes de Ladd (1996). Il avance que la prosodie a des manifestations mesurables, mais répond aussi à des

principes sous-jacents si bien que les approches déductives et inductives devraient être utilisées de concert dans l'étude de la prosodie :

> *« Prosody has both measurable manifestations and underlying principles [...] Thus the two processes- induction and deduction- are not dichotomous; both can be profitably used in the scientific quest for precise and generalizable knowledge. »* Fujisaki (1983, p. 28)

Fujisaki considère que la prosodie permet de transmettre trois types d'information : une information linguistique, paralinguistique et non linguistique.

Par linguistique, on entend l'information symbolique représentée par un ensemble de symboles discrets et de règles régissant leur utilisation. Cette information peut être représentée par le langage écrit ou bien déduite du contexte. Par exemple, le type d'accent d'un mot Japonais est une information discrète, car elle spécifie un type d'accent parmi un nombre fini d'autres types d'accents.

L'information paralinguistique correspond à l'information que l'on ne peut pas déduire du langage écrit, mais qui est ajoutée délibérément par le locuteur afin de modifier ou compléter l'information linguistique. Une phrase écrite peut ainsi être prononcée de différentes façons pour exprimer différentes attitudes, intentions ou styles de paroles contrôlés par le locuteur.

L'information non linguistique correspond à différents aspects tels que l'âge, le sexe, l'état physique et psychique du locuteur, etc. Ce type d'information peut être discret ou continu pour exprimer différentes gradations par exemple.

Ces différents types d'information vont être véhiculés à travers des mécanismes articulatoires tels que la phonation et l'articulation. Il s'agit alors pour ce type de modèle de mettre en relation le contenu linguistique du message et la réalité acoustique. C'est dans cette optique que l'outil mathématique est utilisé. Il permet de fournir une description quantitative des phénomènes physiologiques impliqués dans la production de la prosodie et plus particulièrement les variations de la fréquence fondamentale.

Deux éléments principaux sont mis en valeur dans la production de F0 : les cordes vocales et les mouvements du cartilage thyroïdien dans le larynx.

Dans le cas d'un locuteur normal, les cordes vocales sont une paire de membranes élastiques et symétriques. La fréquence fondamentale correspond à leur mode de vibration normal. La tension des cordes vocales est le premier facteur déterminant de cette fréquence de vibration.

D'autre part le rôle des muscles crico-thyroïdiens est mis en valeur par ce modèle. L'analyse de la structure du larynx suggère que le mouvement du cartilage thyroïdien par rapport au cartilage cricoïde peut se faire de deux façons : sous la forme d'une translation horizontale par le muscle crico-thyroïdien, ou sous la forme d'une rotation autour de l'articulation crico-thyroïdienne. La rotation et la translation de la thyroïde modifient la longueur des cordes vocales. Le mouvement

de translation est caractérisé par une constante temporelle beaucoup plus large que le mouvement de rotation si bien qu'il est utilisé pour indiquer des phénomènes globaux comme l'énonciation alors que le second mouvement indique des phénomènes locaux : l'accent de mot ou la production d'un ton. Cette analyse des mécanismes physiques et physiologiques contrôlant F0 est à la base du modèle « commande-réponse » proposé par Fujisaki. On a donc deux commandes fondamentales appelées « phrase command » et « accent command », les changements de F0 résultant de l'action de ces commandes sont appelés « accent component » et « phrase component ». On peut ainsi expliquer et générer la courbe de F0 au niveau d'une phrase comme un contour global descendant agrémenté de différents pics mélodiques correspondant aux évènements locaux comme les accents mélodiques (cf. illustration ci-dessous).

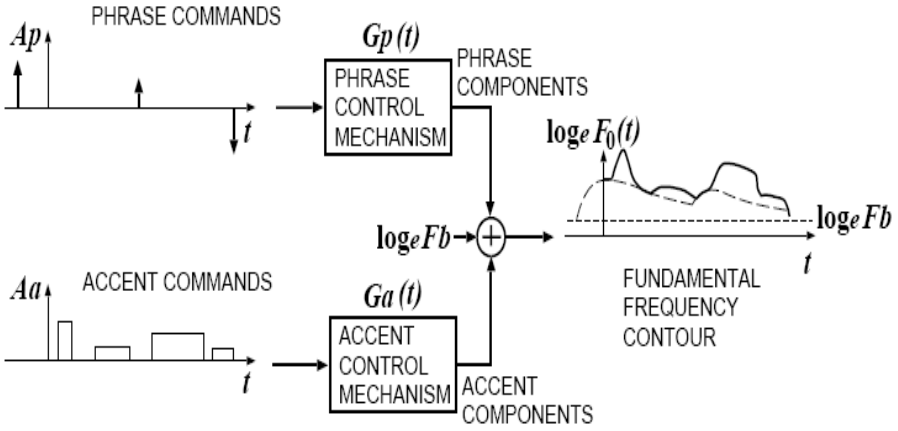


Figure 28 : Illustration du processus de génération des contours de F0 dans le modèle Fujisaki (extrait de Fujisaki 2002).

Ce modèle a été testé sur plusieurs langues dont le Japonais en utilisant l'analyse par synthèse. Cela a permis de quantifier plus précisément le modèle en termes de magnitude et d'alignement. En ce qui concerne le japonais, le modèle permet de générer le contour de F0 en entier avec succès : les deux commandes de phrases positives correspondent au sujet et au prédicat tandis que la commande de phrase négative correspond à la chute finale de F0. Les commandes accentuelles qui sont toujours positives en japonais correspondent aux mots prosodiques.

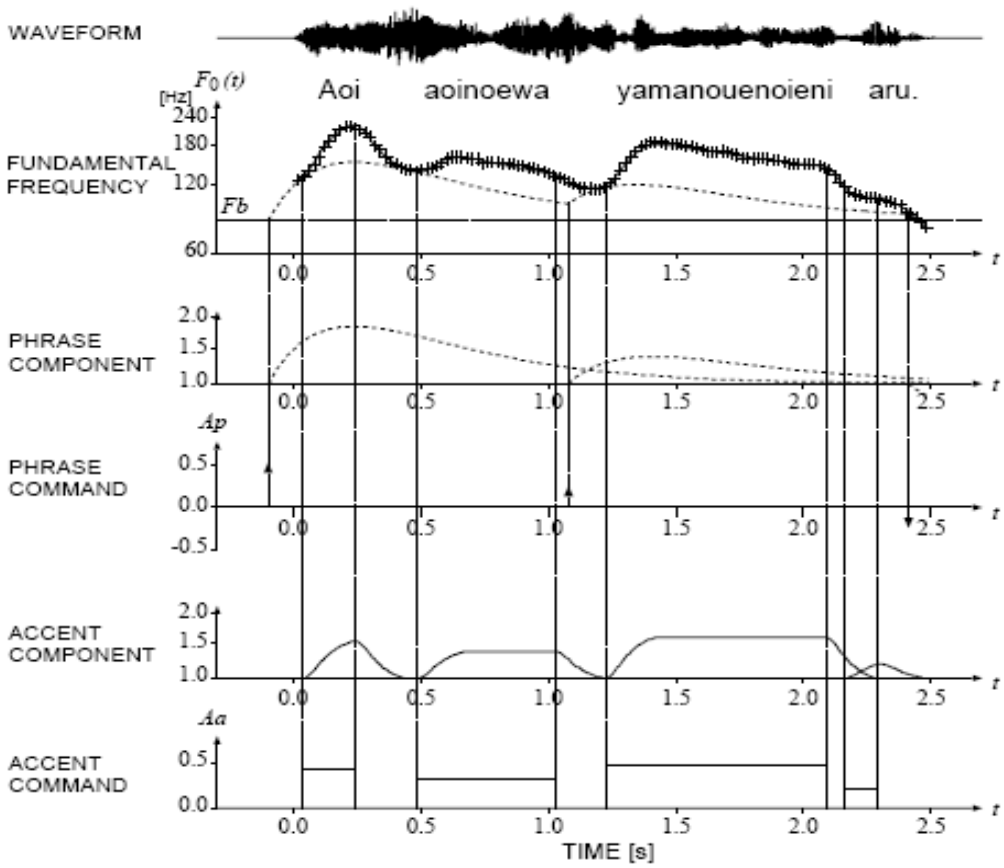


Figure 29 : Application du modèle Fujisaki au japonais (Fujisaki 2006).

D'après les résultats des travaux de Fujisaki (2006), on peut retenir quelques points de repère au niveau de l'alignement des contours de F0 avec les composants linguistiques. Les commandes décrites sont temporellement étroitement liées aux composants linguistiques de la phrase : la commande accentuelle commence 40 à 50 ms avant l'attaque vocalique d'une more aiguë et se termine aussi 40 à 50 ms avant la fin de cette more aiguë. La commande de phrase est localisée 200 ms avant le début d'une phrase et 200 ms avant une frontière syntaxique telle que la séparation entre le sujet et le prédicat. En général elle est beaucoup plus large en début de phrase qu'en milieu de phrase si bien que le contour général de F0 sans compter les montées et descentes dues à la commande accentuelle, décrit une courbe globalement descendante.

Nous décrirons dans la section suivante l'approche des codes biologiques en nous appuyant sur les travaux de Gussenhoven (2002).

4.2.2.2 Les trois codes biologiques

Tout d'abord Gussenhoven (2002) s'intéresse à l'apport sémantique de l'intonation en soulignant l'existence de deux composants principaux : d'une part la réalisation phonétique et d'autre part la grammaire intonative. La réalisation phonétique est utilisée pour véhiculer des significations universelles et se matérialise par les trois codes biologiques suivants : le code de

la fréquence, le code de l'effort et le code de la production. Ces codes permettent de véhiculer des valeurs affectives et informationnelles. L'apport sémantique de l'intonation relevé dans les études essentiellement grammaticales conduit souvent d'après Gussenhoven à former des relations « forme/sens » arbitraires alors que l'étude des codes biologiques ne serait pas en conflit avec les liens naturels existant entre les formes et les fonctions :

> « *Speech communities will vary in the extent to which they employ these meanings [affective and informational] and in the choices they make when they conflict. What they will never do however is change the natural form-function relations that they embody. By contrast, grammaticalised meanings often mimic the natural meaning but linguistic change may create quite arbitrary form-meaning relations when forms are phonologised and the semantics is systematised.* » Gussenhoven (2002)

Le code de la fréquence

On l'appelle aussi « size code » (Ohala 1984), c'est-à-dire le code de la taille. La taille du larynx influence celles des cordes vocales et ainsi la vitesse de vibrations des cordes vocales. Un larynx de petite taille produira souvent des sons plus aigus qu'un larynx large dont les vibrations seront moins rapides produisant des sons plus graves. La corrélation entre la taille du larynx et la vitesse de vibration des cordes vocales est exploitée dans l'expression des rapports dominant/dominé notamment.

Gussenhoven présente dans son article deux grands types d'interprétations sémantiques de chaque code biologique : l'expression de l'affect (en relation avec le locuteur) et l'aspect informationnel ou grammatical (en relation avec le message linguistique).

Les interprétations sémantiques du code de la fréquence sont nombreuses : on lui accorde essentiellement une valeur affective et on peut distinguer l'expression de la soumission, de valeurs féminines et leur opposé de dominance masculine. Par un registre de voix aigu, on peut marquer un caractère amical, la politesse, la vulnérabilité par opposition à nouveau au registre grave marquant la confiance, l'assurance, un côté protecteur ou bien de manière plus négative un tempérament agressif.

Du point de vue informationnel, un registre plus aigu marque l'incertitude par opposition au registre grave marquant la certitude. Ceci revient donc à utiliser la fréquence pour distinguer l'acte de poser une question (registre aigu) de celui d'affirmer (registre grave) comme dans la tradition classique d'interprétation grammaticale du rôle de l'intonation.

Le code de l'effort

La quantité d'énergie déployée pour parler est variable : un effort plus marqué entraîne des mouvements articulatoires plus précis, des contours mélodiques plus canoniques et nombreux.

Les mouvements mélodiques sont aussi plus amples (le registre est plus étendu).

L'interprétation la plus fréquente est la suivante : le locuteur produit un effort particulier, car il estime que le contenu du message est important. La signification la plus évidente est donc l'emphase du point de vue informationnel. Un son plus aigu aura tendance à être plus proéminent, mais les expériences ont montré que l'étendue du registre utilisé et l'ampleur des mouvements mélodiques jouent aussi un rôle majeur. Une application intéressante est l'utilisation en langue Bantu d'un registre de voix compressé pour exprimer la négativité ou la dissimulation de certaines informations.

L'interprétation affective du code de l'effort inclut l'expression de la surprise et de la relation d'aide. L'utilisation de mouvements mélodiques d'une grande ampleur peut être indicative d'une disposition obligeante. Les tests de perception effectués sur le Néerlandais (Thomas 1978) montrent que c'est la distance entre les cibles tonales L et H qui est la plus pertinente et pas seulement l'ampleur de n'importe quel mouvement mélodique.

Le code de la production

La production d'énergie ou d'un effort est liée aux phénomènes respiratoires et à la présence de groupes de souffle. De façon générale ce code met en relation les débuts d'énoncé avec un registre aigu et les valeurs plus graves de F0 avec les fins d'énoncé.

Les locuteurs semblent ainsi dépenser plus d'énergie au début d'un énoncé qu'à la fin, cela est dû à la relation existant entre les énoncés et les groupes de souffle. Une conséquence naturelle de cette perte d'énergie est une diminution graduelle de l'intensité et un abaissement progressif de la fréquence fondamentale appelé effet de déclinaison ou « downstepping effect ».

Du point de vue informationnel, le code de la production est associé à l'initialité et à la finalité. Les débuts d'énoncés aigus signalent l'introduction d'un nouveau thème (« topic »), les débuts d'énoncés dans les graves signalent la continuité d'un même thème de conversation. Les fins d'énoncés fonctionnent de façon contraire : les fins aiguës signalent la continuité alors que les fins graves marquent la finalité ou le changement de tour de parole.

D'après Gussenhoven, ces trois codes permettent d'interpréter ce qui est universel dans les variations de la hauteur de F0.

> *« Universal meaning in intonation derives from three biological codes […]These codes are biological in the sense that they represent aspects of the speech production mechanism that affect rate of vocal cords vibration. »* Gussenhoven (2002)

Enfin, dans la dernière partie dédiée à l'approche physiologique, nous présenterons le modèle PENTA.

4.2.2.3 Le modèle PENTA (Xu 2005)

Le modèle PENTA est issu des travaux de recherches menés par Xu (2001, 2002, 2005). Tout comme les deux auteurs cités précédemment, il inclut dans son modèle la dimension articulatoire ou physiologique, mais comme on le verra dans cette partie, il se distingue des autres théoriciens de diverses façons. À la base du modèle de Xu (2005) se trouve l'affirmation suivante : le langage permet de véhiculer du sens et il est produit par le système articulatoire humain.

La plupart des modèles de l'intonation existants définissent d'abord les composants du système intonatif en termes de forme et secondairement seulement en termes fonctionnels. De plus, ils tentent généralement d'observer un lien direct entre les contours de F0 et les fonctions intonatives. Or Xu (2005) affirme qu'il n'existe pas de relation directe entre les fonctions intonatives et les corrélats acoustiques. Les fonctions mélodiques ne seraient donc pas codées directement sous une forme invariable, mais seraient soumises à des processus de manipulation du système articulatoire définis ici comme des « primitives phonétiques ».

Quatre primitives phonétiques sont reconnues par le modèle : les cibles mélodiques locales, le registre, la force articulatoire et la durée.

Ainsi on ne reconnaît pas de corrélats acoustiques invariants, mais Xu (2005) postule l'existence de cibles mélodiques idéales et sous-jacentes que l'on essaierait toujours d'atteindre dans un acte de communication, mais qui seraient approchées de façon variable en fonction des différentes primitives mélodiques. Ces cibles mélodiques sous-jacentes peuvent être hautes (« High »), basses (« Low »), moyennes (« Mid »), montantes (« Rising ») ou descendantes (« Falling »).

La réalisation mélodique des fonctions communicatives dans ce modèle est donc soumise à des contraintes articulatoires et se fait aussi de façon parallèle (plusieurs fonctions mélodiques peuvent être encodées simultanément).

La vitesse maximale de changement de la valeur de F0 fait partie des facteurs biologiques déterminants dans la génération des contours mélodiques. Les contours mélodiques sont générés par le larynx qui est comparé à un instrument de musique :

> *« If likened to a musical instrument, the larynx belongs to the type that can produce only one note at a time and shift to a new note only after the previous note is over. »* Xu (2005, p. 224)

La vitesse à laquelle deux notes successives peuvent être produites dépend donc des propriétés vibratoires des cordes vocales et de leur tension en particulier. D'après les différentes expériences menées par Xu et Sun (2000), le locuteur moyen a besoin d'un laps de temps de 100 ms pour produire un changement mélodique le plus petit soit-il.

De plus, plusieurs muscles sont impliqués dans la production de F0. On trouve par exemple de nombreux muscles permettant la mise en tension des cordes vocales d'une part et les propriétés des cordes vocales créant un mode de vibration spécifique. La coordination de ces différents articulateurs limite aussi le degré de liberté de mouvement du larynx (pour plus de détails cf. Xu 2005).

L'appellation PENTA : « Parallel Encoding and Target Approximation » est tirée de cinq principes :

- La vitesse de changement de F0 et la synchronisation des mouvements mélodiques imposent des contraintes articulatoires conséquentes sur les contours de F0 ;
- Le mécanisme articulatoire de base générant F0 est très probablement réalisé en synchronie avec la syllabe ;
- L'affectation de cible mélodique au travers de diverses fonctions est spécifique à chaque langue ;
- Les composants mélodiques du langage sont définis et organisés en termes de fonction non de forme ;
- Plusieurs fonctions communicatives peuvent être transmises en parallèle.

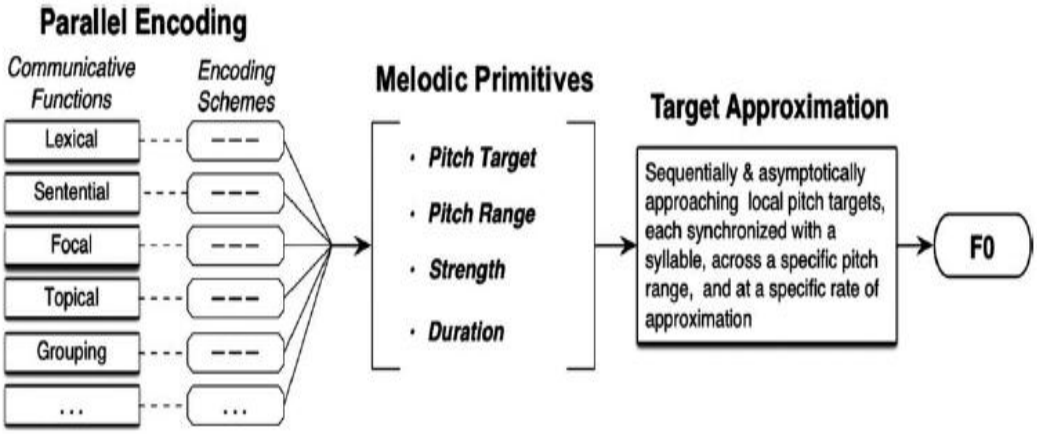


Figure 30 : Génération des contours de F0 dans le modèle PENTA (extrait de Xu 2005).

On part donc bien dans ce modèle des fonctions communicatives comme la focalisation, le groupement de mots etc. Ces fonctions sont transmises grâce à un schéma d'encodage particulier à chacune, celui-ci correspond aux propriétés numériques des primitives mélodiques. Ainsi le registre peut être étendu, étroit ou normal pour exprimer une fonction particulière. Cela résulte ensuite en la production d'une cible mélodique se rapprochant plus ou moins de l'une des cibles mélodiques idéales définies plus haut.

La primitive « registre » est associée notamment à la focalisation : afin de mettre en valeur une partie du message, le registre serait volontairement étendu sur certains segments et réduit sur les autres. La primitive « strength » ou force articulatoire qui spécifie la vitesse à laquelle une cible

tonale peut être approchée est associée aux groupements de mots et à la marque d'information lexicale.

Ce modèle se distingue du modèle de Fujisaki de plusieurs façons : tout d'abord, il contient une contrainte forte de synchronisation de la cible mélodique avec les frontières de la syllabe. Ensuite, par opposition au modèle « command-response », on ne prend pas en compte une forte tendance de retour vers une ligne mélodique de base (« baseline »). Enfin, les composants de base du modèle PENTA sont des patrons mélodiques idéaux plutôt que des commandes musculaires. En effet, Xu (2005) avance que les mouvements des muscles sont constamment ajustés et modifiés lors d'un acte de langage et ne peuvent donc pas être considérés comme des mouvements de base invariants.

Dans ce modèle, il n'y a pas eu non plus de tentative de lier les fonctions communicatives et les codes biologiques de façon directe à l'image des travaux d'Ohala (1983) et Gussenhoven (2002).

Les avancées technologiques, la mise à disposition de nouveaux outils et de plus grands corpus de parole ouvrent la voie à de nouvelles méthodes de recherche remettant en cause ou venant tester les modèles les plus classiques de l'intonation.

Nous présenterons dans une dernière partie, deux méthodes d'approche de l'articulation forme/fonction basées sur l'interaction entre l'étude de l'intonation et le traitement automatique des langues : l'approche probabiliste et l'analyse par synthèse.

4.2.3 L'approche probabiliste

Il s'agit d'établir la grammaire prosodique d'une langue en se basant sur l'étude du comportement des formes prosodiques sur de grands corpus. L'ouvrage récent de J. Coleman (2000), rend compte des différentes approches expérimentales possibles de nos jours en linguistique expérimentale.

Un chapitre entier est ainsi consacré à l'explicitation des « automates à état fini ». Ces automates permettent de générer des grammaires linguistiques. On peut par exemple générer une grammaire finie de la syllabe anglaise. Celle-ci illustre les combinaisons de syllabes possibles en anglais. Ceci est tout à fait particulier à une langue donnée dans laquelle certaines séquences seront permises ou non. Cependant, Coleman (2000) ne présente pas dans son ouvrage d'application de ces automates ou modèles Markov cachés à l'étude de la prosodie. On peut mentionner deux exemples récents de mise en application des méthodes probabilistes dans l'étude de la prosodie : le modèle probabiliste d'Audra Dainora (2001) et les travaux de Nesterenko, Hirst et Rauzy (2007).

Dans une thèse récente (Dainora 2001), un modèle probabiliste de l'anglais américain a été établi. A. Dainora (2001) s'appuie sur un automate à état fini, pour décrire et analyser la distribution des différents morphèmes intonatifs que sont les accents mélodiques, les tons de

phrase et les tons de frontière. Ce modèle se base donc en amont sur la théorie autosegmentale de l'intonation et utilise la notation TOBI. Il s'agit ici d'étudier les transitions possibles entre les différents morphèmes intonatifs en termes probabilistes et à partir d'un grand corpus d'anglais américain. Le modèle final permet de prédire par exemple l'accent de frontière qui apparaîtra de la façon la plus probable après un accent mélodique particulier. Une telle approche permet de remettre en question l'inventaire des tons phonologiques existant dans l'approche autosegmentale, comme les phénomènes d'abaissement (« downstep »), qui d'après Dainora (2001) ne sont pas distinctifs et ne devraient donc pas apparaître dans l'inventaire phonologique de base.

Enfin, dans une thèse franco-russe récente (Nesterenko 2006 et Nesterenko, Hirst et Rauzy 2007), la question de l'explicitation de l'articulation fonction/forme est adressée en proposant une approche probabiliste. On cherche alors à établir une grammaire INTSINT à partir de l'étude probabiliste de la primitive fonctionnelle « accentuée » dans un corpus de parole spontanée en langue russe.

Les auteurs adoptent le point de vue de Pierrehumbert (2001) qui souligne que la tradition générativiste déprécie la valeur des analyses probabilistes en privilégiant l'étude de la compétence linguistique chez l'humain à celle de l'étude de la variabilité observée au niveau de la performance. Dans une approche probabiliste, on n'écarte donc pas les structures prosodiques atypiques. Celles-ci sont seulement considérées comme plus rares et marquées.

L'unité prosodique de base, dans cette étude, est le mot prosodique. Sur la totalité du corpus, 315 mots prosodiques ont été annotés comme portant un accent mélodique.

Deux concepts mathématiques d'intérêt sont retenus dans ces analyses : celui de la probabilité conditionnelle et celui de l'entropie. D'après les auteurs, les probabilités conditionnelles renseignent sur le lien qui existe entre deux événements, en l'occurrence, deux étiquettes de tons d'INTSINT. On interprète les valeurs de la probabilité conditionnelle de la façon suivante :

- si B n'exerce aucune influence sur la présence de A, $p(A|B) = p(A)$;
- si B défavorise la présence de A, $p(A|B) < p(A)$;
- si B favorise la présence de A, $p(A|B) > p(A)$.

Les auteurs utilisent aussi la mesure de l'entropie qui est une mesure de l'organisation de l'information. La notion étant propre à la théorie de l'information, elle quantifie pour une distribution donnée l'écart par rapport à une distribution équiprobable. L'entropie d'une distribution varie entre 0, valeur correspondant à un système déterministe, et lnN pour un système dont les N états sont équiprobables.

Une première ébauche de grammaire probabiliste INTSINT a donc été proposée pour l'encodage de la primitive fonctionnelle « accentuée ». Toutefois, avant d'aller plus loin, les auteurs se posent la question de l'unité optimale de description de la fonction accentuée et se

proposent aussi d'inclure les phénomènes d'alignement des contours mélodiques dans une version plus approfondie de leur grammaire.

Les modèles probabilistes offrent donc une piste de recherche très intéressante dans l'étude de l'articulation forme/fonction appliquée à de grands corpus d'anglais. C'est une approche cohérente avec la démarche visant à déduire la structure prosodique à partir des données observables. Plus le corpus utilisé est grand, plus le potentiel d'enrichissement des grammaires prosodiques est important.

Dans la section suivante, nous présenterons une autre approche de l'étude de l'intonation basée sur les systèmes de traitement automatique des langues : l'analyse par synthèse.

4.2.4 L'analyse par synthèse

> *« Quand on s'interroge sur les échanges entre la linguistique et le traitement automatique des langues (TAL) dans le domaine de la prosodie, on constate d'emblée que la principale application en TAL où intervient la prosodie, c'est la synthèse de la parole à partir du texte »*
>
> affirme Piet Mertens (2004) dans une publication sur *« les aller-retour entre la prosodie et son traitement automatique* ».

Une telle application de l'étude de la prosodie se retrouve dans les systèmes de lecture de documents pour les non voyants, les systèmes de réponse vocale, etc.

Toutefois nous nous intéressons aux systèmes de synthèse vocale dans ce travail de recherche, essentiellement dans la mesure où ils constituent un outil privilégié pour mettre à l'épreuve les théories linguistiques mises en œuvre et pour leur réalisation. On peut alors s'en servir pour simuler et évaluer les modèles de la prosodie : les imperfections au niveau de l'intonation produite par le système de synthèse peuvent constituer des indices d'erreurs ou de lacunes dans le modèle de représentation choisi.

La problématique principale est la suivante : quelles sont les informations pertinentes en tant qu' « input » pour un système de génération des contours de F0 ? De la finesse de l'analyse en amont dépendra la qualité de la parole synthétique crée. Les systèmes de synthèse actuels ne montrent pas encore autant de finesse comme le remarque Halliday (1970) :

> *« When the information provided is to be converted into machine readable instructions [...] the description lacks the precision and explicitness required [...] »*

C'est une approche à double sens de la prosodie, à la fois déductive et inductive, c'est-à-dire que les résultats en synthèse et notamment les manques et défauts constatés permettront d'enrichir le modèle théorique de départ. En s'appuyant sur ce test de la synthèse, de nombreux chercheurs

ont proposé de modéliser l'intonation d'une langue en se basant sur des descriptions formelles et fonctionnelles différentes.

On prendra ici l'exemple des modèles de Pijper (1983), Williams et Alderson (1986) et de Taylor (2000) qui sont des modèles essentiellement phonétiques d'une part et d'autre part, les modèles de Bailly (2002) et Auran et Hirst (2005) qui accordent une place plus importante à la modélisation du niveau phonologique.

4.2.4.1 Le modèle de Pijper (1983) : l'analyse par synthèse à partir de l'approche perceptive de l'intonation

L'ouvrage relatant les travaux de Jan Roelof de Pijper s'intitule : « Modelling British English Intonation, An analysis by resynthesis of British English intonation ».

La resynthèse des contours intonatifs à partir d'une représentation spécifique de l'intonation est utilisée pour évaluer un modèle de l'intonation et l'enrichir à partir des données. Ce chercheur s'inscrit dans le mouvement de recherche de l'institut IPO et en adopte donc le modèle phonétique et phonologique. Les contours de F0 sont représentés par des lignes droites formant des contours stylisés appelés « close copy stylizations » qui sont dérivées de l'observation étroite des données.

La structure du modèle est basée en partie sur le modèle existant en hollandais et en partie sur les résultats obtenus après avoir analysé un grand corpus de contours stylisés.

Elles répondent aussi à l'impératif de pertinence au niveau de la perception c'est-à-dire que l'on réduit l'information prosodique à ce qui est utile et pertinent pour celui qui écoute. Un corpus de base est utilisé pour faire un premier inventaire des contours stylisés à retenir. Ce même inventaire est ensuite utilisé pour annoter un corpus plus large qui contient donc souvent des contours non présents dans l'inventaire de départ. La comparaison entre les contours stylisés resynthétisés et les contours présents dans le corpus permet d'isoler les segments différents dans les nouveaux contours et d'enrichir le modèle de départ.

Les deux éléments de base de ce modèle sont les mouvements mélodiques et les règles régissant leurs combinaisons.

Huit contours mélodiques sont définis et se différencient selon plusieurs paramètres mélodiques : leur direction, le registre, la pente, la durée du mouvement en relation avec la syllabe.

Le registre peut être complet (« full ») ou partiel (« half »), la pente abrupte (« steep ») ou progressive (« gradual »). La valeur d'une pente standard est définie à 7,5 demi-tons et la durée standard d'un mouvement est de 160 ms. Les mouvements mélodiques sont alignés avec l'attaque de la syllabe et la partie voisée de celle-ci. Une montée étiquetée « abrupte » commencera 160 ms avant l'attaque de la syllabe alors qu'une montée partielle commencera 80 ms avant l'attaque syllabique.

Pijper affirme que ce modèle fournit une bonne représentation de l'intonation de l'anglais. Le cadre de développement de ce modèle permet aussi de tester d'autres approches comme celle de Halliday basée sur sept tons primaires.

Cependant les sept contours mélodiques ne suffisent pas à représenter toute la richesse mélodique de l'anglais.

L'inconvénient d'un tel modèle vient du fait que l'on ait besoin d'énormément de données au départ pour fournir les contours stylisés. De plus il n'y a aucune tentative ici de mise en relation de cet inventaire des formes prosodiques avec le niveau phonologique ou fonctionnel.

Ce modèle est purement phonétique : on part de l'observation des données acoustiques pour développer un modèle des formes existantes et des combinaisons possibles. Les fonctions prosodiques ne sont donc pas prises en compte :

« *It's a truly melodic model, ie it doesn't include any knowledge concerning what intonation patterns exist in English, or what functions they are*. » Pijper (1983, p. 43)

Un modèle de synthèse de l'intonation de l'anglais britannique a aussi été développé à partir cette fois-ci des théories britanniques de l'intonation.

4.2.4.2 Un modèle de synthèse de l'anglais britannique

B.J Williams et P.R Alderson (1986), proposent aussi un modèle de synthèse de l'intonation de l'anglais britannique basé sur l'approche des tons nucléaires. Il s'agit de synthétiser l'intonation de l'anglais à partir d'une annotation en TSM. Ce système de TSM est en fait une variation des systèmes utilisés par O'Connor & Arnold (1961) et Crystal (1969). Des règles ont ensuite été formulées pour convertir les TSM en cibles tonales sur une échelle de 1 à 10 similaire à celle utilisée par Pierrehumbert (1980).

La dernière étape consiste en une superposition de la ligne de fréquence de base et de la ligne de fréquence maximale associant une valeur à chaque syllabe. Le contour de F0 résultant est ensuite comparé avec l'original pour une évaluation objective. Rentrons maintenant dans plus détails en explicitant le système de transcription utilisé, les règles de synthèse et la vision de la relation forme/fonction.

Un système britannique « adapté »

À l'image du système britannique traditionnel, les énoncés sont divisés en unités tonales majeures correspondant à des phrases, elles-mêmes composées d'une ou de plusieurs unités tonales mineures (correspondant fréquemment à un constituant syntaxique plus petit). Chaque unité tonale mineure contient au moins une syllabe accentuée. Les syllabes accentuées peuvent être proéminentes de plusieurs façons : du point de vue rythmique quand on constate une augmentation de la durée et/ou de l'intensité, du point de vue mélodique quand une discontinuité est introduite dans la courbe de F0 (c'est l'opposition classique « stressed » *vs*

« accented »). Les catégories « head, prehead, nucleus, tail » sont conservées donnant le système suivant :

- Les frontières d'unités tonales sont notées « | »ou « || » ;
- Les syllabes accentuées (proéminente mélodiquement et ou rythmiquement) peuvent prendre la forme d'une chute (grande ou petite), d'une montée (grande ou petite), d'une chute-montée et d'une montée-chute et d'un ton statique haut ou bas ;
- Les syllabes non accentuées (non proéminentes mélodiquement) peuvent être rythmiquement proéminentes (« stressed »), notifiée d'une flèche montante (« booster ») pour indiquer une augmentation de la fréquence ou d'une flèche descendante pour indiquer une diminution (« drop ») ;
- Les unités tonales mineures se trouvent la plupart du temps au niveau d'une pause courte ou d'une pause pleine avec un allongement final de la dernière syllabe.

Les cinq types d'accents mélodiques s'appliquent ici aussi bien au nucléus qu'à la tête, ce qui simplifie considérablement l'analyse. Toute différence entre ces deux entités est traitée comme des variations graduelles non catégorielles ou discrètes :

> *« For O'Connor and Arnold, as for Crystal, the types of pitch pattern found in the head are phonemically distinct from those found in the nucleus. […] Any differences [in this model] are treated as allophonic. »* Williams et Alderson (1986, p. 4)

Les accents mélodiques peuvent tous être grands ou petits (« high and low »), ce qui fait référence à la hauteur initiale de F0 au niveau de l'attaque de la syllabe accentuée en comparaison avec la hauteur de F0 au niveau de fin de la syllabe précédente. On évite ainsi les problèmes d'asymétrie dans le système d'O'Connor et Arnold qui ne considèrent pas l'opposition « high » et « low » pour les chutes-montées et montées-chutes.

L'utilisation des flèches dénote un registre plus large ou plus étroit que la normale en certains points. La hauteur est déterminée de façon essentiellement relative. La valeur de F0 au niveau d'une syllabe donnée peut être plus haute que les valeurs associées à la précédente ou bien plus basse, au niveau, beaucoup plus basse ou beaucoup plus haute.

Utilisant ce système de transcription comme un « input » pour le modèle de synthèse, on attribue à chaque syllabe une valeur allant de 1 à 10, représentant une échelle abstraite de hauteur mélodique pertinente au niveau linguistique. Par opposition au modèle de Pierrehumbert (1980), chaque syllabe est marquée par une valeur particulière.

Les contours synthétiques issus de ce modèle montrent des différences notables par rapport aux contours de F0 brute. Cependant l'évaluation subjective de ces contours donne un résultat perceptuellement équivalent aux contours originaux ce qui rapproche ce modèle du modèle IPO. Les auteurs soulignent la difficulté de trouver une mesure appropriée et objective afin d'évaluer

la qualité des contours synthétique. Ils proposent alors de s'appuyer sur la notion d'équivalence au niveau de la perception.

> *« It's difficult to find the appropriate measures by which to judge the correctness of synthesized contours. […] In no sense is the precise F0 of an utterance to be taken as canonical. […] It's in this sense that the notion of perceptual equality is particularly useful. »* Williams et Alderson (1986, p. 20)

Mouvement mélodique et sens

Les auteurs de ce papier parlent de leur modèle comme un modèle essentiellement phonétique dans sa première version. Toutefois il est basé sur une analyse fonctionnelle de départ déterminante. Les contours mélodiques utilisés sont déduits d'une analyse linguistique de l'intonation :

> *« Not just any pitch movement is appropriate in a synthesized utterance: the underlying linguistic analysis must be taken into account when assigning pitch patterns. » (*p. 21)

Ils ajoutent aussi que le modèle doit mettre en relation les mouvements de F0 et la structure prosodique de façon transparente comme l'affirme Ladd (1984) :

> *« It makes no sense to model trend lines if the points to which the trend lines are fitted are known to be phonologically equivalent. »* Ladd (1984, p.56)

Il s'agit donc d'un modèle de synthèse basé sur une analyse phonologique de départ correspondant à une adaptation des théories britanniques traditionnelles. On observe une tentative d'isoler seulement les mouvements mélodiques pertinents par opposition aux variations arbitraires produites dans le langage. Les fonctions retenues sont essentiellement structurelles : il s'agit de la fonction accentuelle et de la marque des frontières.

On présentera dans la section suivante une autre approche d'analyse par synthèse de l'intonation basée sur un modèle phonétique unique en son genre : le modèle TILT de Taylor (2000).

4.2.4.3 Le modèle Tilt

Ce modèle a été développé par Paul Taylor (2000) et est une extension du modèle RFC (« Rise Fall Connection », 1995) du même auteur. Il s'agit aussi d'un modèle résolument phonétique décrivant les contours observables de F0. La représentation de l'intonation est totalement distincte du modèle précédent : c'est une séquence linéaire d'évènements tels que les accents mélodiques (« pitch accent ») et les tons de frontière (« boundary tone »). Ce modèle est donc proche des théories autosegmentales de l'intonation.

Chaque évènement mélodique est caractérisé par des paramètres continus : l'amplitude, la durée et le « tilt » c'est-à-dire la mesure de la forme du contour.

Un détecteur d'évènement prosodique a été créé dans cette approche qui rend possible le développement d'un système de reconnaissance de l'intonation.

Des algorithmes de synthèse sont aussi utilisés afin de générer les contours de F0 à partir de la représentation Tilt et de comparer ces contours synthétiques avec les contours originaux de F0 présents dans les données. Ce test de la synthèse fait partie des qualités fondamentales de ce modèle :

> « *The basic aim of intonation models is to provide a system of intonational description that is linguistically meaningful in such a way that representations in this system can be automatically derived from the relevant parts of an utterance's acoustics, and that the acoustics can be automatically synthesized from the representation* » Taylor (2000, p. 3)

Une description significative au niveau linguistique est donc une description qui contient l'information nécessaire à l'interprétation d'un énoncé. Cela exclut donc les phénomènes redondants affectant les contours de F0 comme les perturbations segmentales ou micro-prosodiques :

> « *smooth contours are more meaningful representation to measure against than raw contours.* »

Dans le modèle RFC, chaque évènement est décrit en termes de montée ou chute. Celles-ci sont distinguées en fonction de leur amplitude et de leur durée. Pour une forme montante ou descendante donnée, trois points de repères sont retenus : le début de l'évènement, le point culminant ou « peak » et la fin du mouvement. Les connections représentent les contours de F0 entre les évènements. Dans cette analyse, on se concentre uniquement sur les parties voisées de F0 : les paramètres du modèle Tilt sont dérivés automatiquement à partir de l'analyse RFC. D'après Taylor, si le modèle RFC permet de suivre les contours de F0 de façon efficace, l'extension en modèle Tilt permet par ailleurs une meilleure interprétation fonctionnelle des évènements. En effet, le tilt est une mesure de la forme des évènements et représente les tailles relatives des montées et descentes détectées. L'amplitude, par exemple, est mise en relation avec la proéminence phonétique d'un évènement, plus l'amplitude sera importante, plus l'évènement sera proéminent.

La qualité de la synthèse produite par le modèle est mesurée par l'écart quadratique moyen[3] (RMSE ou Root Mean Square Error) entre le contour original et le contour synthétique et par le calcul de la valeur de la corrélation entre ces deux courbes.

---

[3] Source de la traduction des termes statistiques : http://isi.cbs.nl/glossary/term2872.htm . « International Statistical Institute ». Le terme RMSE peut aussi être traduit comme l'erreur quadratique moyenne.

Le modèle Tilt étant essentiellement présenté comme un modèle phonétique de l'intonation, la question qui se pose est la suivante : « Is the Tilt representation linguistically meaningful ? ». Les modèles RFC et Tilt ont en effet souvent souffert de la critique de n'être que des modèles de codage des contours de F0 non signifiants au niveau linguistique. Si le modèle Tilt n'a pas encore donné lieu à une analyse phonologique de l'intonation, il permet tout de même de remettre en question la représentation des variables phonétiques qui sont à la base de toute théorie sémantique de la prosodie. En effet, le modèle Tilt utilise des paramètres ou variables continues pour décrire les évènements mélodiques alors que la plupart des représentations traditionnelles s'appuient sur des catégories discrètes.

> *« This [the bias against continuous representations] has led to properties of intonation that are clearly continuous, such as pitch range and prominence, being somewhat ignored, and study concentrating on categorical issues only, such as accent type. [..] The bias towards purely discrete representations in linguistics is a hangover from traditional linguistics and has often been justified because such representations are seen as being properly cognitive. »*
> Taylor (2000, p. 20)

D'après Taylor (2000), il n'y a aucune preuve montrant que les frontières strictes entre les unités prosodiques généralement définies signalent un changement sémantique significatif. Sa position est la suivante : l'intonation est continue en termes de son et d'apport sémantique, mais la relation entre ces niveaux est complexe et non linéaire.

Dans les deux derniers modèles que nous présenterons dans ce chapitre, la relation au sens est abordée de façon plus directe et aboutit à un modèle de l'articulation forme/fonction.

4.2.4.4 Le modèle SFC

Plus récemment Holm et Bailly (2002) et Bailly et Holm (2005) s'appuient aussi sur un système de synthèse pour révéler « la structure cachée de l'intonation » : « Learning the Hidden Structure of Intonation : Implementing Various Functions of Prosody » et développent un modèle prosodique complet appelé SFC (« Superposition of functional contours »).

Cette approche a été développée initialement à travers l'étude du français, mais a été tout aussi bien étendue à d'autres langues comme l'anglais et le mandarin. Ce modèle possède trois dimensions : une dimension phonologique, une dimension phonétique (description des contours observables de F0) et un modèle de l'articulation entre ces deux niveaux. Ce modèle est constitué par l'entraînement de générateurs de contours (un par fonction) en utilisant l'analyse par synthèse (« analysis by synthesis loop »). Il s'agit d'un système basé sur des contraintes théoriques et sur l'exploitation de données expérimentales, il permet de générer les formes prosodiques à partir de l'information métalinguistique.

Le modèle phonologique

Le continuum prosodique est considéré comme une superposition de contours à paramètres multiples s'ajoutant les uns aux autres. Les formes phonétiques et les fonctions linguistiques sont très étroitement liées si bien que les fonctions métalinguistiques s'appliquant à des domaines divers, sont directement réalisées sous forme de contours à paramètres multiples.

Par fonctions métalinguistiques, on entend la faculté de la prosodie de démarquer les unités phonologiques et de véhiculer des informations concernant les fonctions propositionnelles et interactionnelles de ces unités dans le discours. Ces fonctions comme, la fonction de segmentation, de hiérarchisation, d'emphase ou d'expression de l'attitude, sont implémentées de façon directe sous la forme d'un contour « prototypé ». La prosodie du message en entier est alors obtenue par l'addition et la superposition de tous les contours impliqués. Ce modèle ne fait donc pas usage d'une représentation intermédiaire de la prosodie (cf. Hirst *et al.* 2000).

Le modèle phonétique

La mélodie et la structure rythmique de l'énoncé sont analysées : chaque syllabe est caractérisée par un mouvement mélodique. Celui-ci est stylisé à l'aide de trois valeurs de F0 sur le noyau vocalique et un facteur d'allongement. Ces quatre valeurs relevées pour toutes les syllabes, constituent ce que les auteurs appellent le « squelette d'un énoncé ». Ce squelette est articulé par les segments prosodiques élémentaires qui sont superposés. Les paramètres prosodiques donnant vie au squelette sont les contours de F0 et la durée des phonèmes. Une première étape de décomposition de la mélodie est effectuée en utilisant une procédure de stylisation similaire à MOMEL. Contrairement à MOMEL cependant, cette stylisation comprend une contrainte d'alignement avec la ligne segmentale.

Le « mapping model »

Chaque fonction métalinguistique est encodée par un contour prototypé spécifique lié à un domaine d'application (la syllabe, la clause..). Puisque la fonction peut être appliquée à différents domaines, elle est caractérisée par une famille de contours ou clichés prosodiques (cf. Fonagy et Bailly 1984). La phase d'apprentissage des générateurs de contours est effectuée à l'aide de réseaux de neurones appelé PMEM (« Prosodic Movement Expansion Model »).

Le cycle d'analyse par synthèse fonctionne de la façon suivante : les générateurs de contours prédisent des contours fonctionnels pour toutes les unités du corpus. Pour chaque énoncé, les contours prosodiques synthétiques sont informatisés en superposant les contours associés aux unités de l'énoncé. Pour chaque énoncé, ces contours prédits sont comparés aux contours observés. L'erreur de prédiction est analysée et distribuée parmi les contours contribuant à l'énoncé. Cette différence de prédiction est ensuite ajoutée aux contours fonctionnels prédits

pour former de nouveaux contours cibles pour l'entraînement des générateurs de contours. Une fois les générateurs entraînés, à nouveau on retourne à l'étape de départ jusqu'à ce que le taux d'erreur soit le plus petit possible.

Cette procédure est à la fois déductive (« bottom up close copy stylisation ») et inductive (fournit une description phonologique). Le plus grand intérêt de ce modèle SFC est qu'il permet à la fois de générer la prosodie et de l'analyser en testant par exemple différentes fonctions métalinguistiques et domaines associés.

4.2.4.5 L'environnement PROZED

Enfin, un autre cadre d'approche d'analyse par synthèse de la prosodie est celui développé par Hirst et Auran (2005) dans l'article suivant : « Analysis by synthesis of speech prosody : the ProZed environment ». On propose ici un système d'analyse de la prosodie par la synthèse multilingue.

L'analyse se base sur la représentation symbolique des formes prosodiques sans faire référence aux fonctions prosodiques par opposition à certains des modèles précédents tels que celui de Williams et Alderson (1986). Les paramètres du modèle se limitent pour l'instant à la fréquence fondamentale et à la durée, mais pourront être étendus par la suite à d'autres paramètres telle que la qualité de voix. Ce cadre théorique est intégré au programme Praat pour l'analyse et Mbrola pour la synthèse.

L'originalité de ce système vient du fait que l'on distingue et sépare bien la représentation des formes prosodiques de celle des fonctions.

La représentation symbolique du système peut être directement et automatiquement dérivée des données acoustiques en spécifiant les domaines et unités pertinentes pour l'analyse. Cette analyse est alors réversible c'est-à-dire que le codage symbolique peut être converti en données prosodiques et donner lieu à une comparaison avec le contour d'origine. Les erreurs ou imprécisions du modèle synthétique permettent de chercher à approfondir l'analyse de départ.

Ce cadre d'analyse a été utilisé pour étudier notamment le rythme et la mélodie.

Pour chacune de ces analyses, on fait une distinction entre les variations locales ou à court terme (les tons distinctifs lexicaux ou non lexicaux, la quantité) pour lesquelles des unités spécifiques sont définies, puis les variations globales ou à long terme (variation de registre, tempo). Ainsi d'après cette analyse, on considère pour un domaine rythmique particulier, que l'on peut définir un tempo constant utilisé comme une référence par défaut pour l'analyse du phénomène à court terme associé. De même la mélodie est décrite par domaines au niveau desquels le niveau de hauteur de F0 de référence du locuteur appelé « clé » et la variabilité possible par rapport à cette valeur appelée « étendue », sont considérés comme constants.

En ce qui concerne la modélisation de la mélodie, on utilise l'algorithme MOMEL dont on a déjà parlé et le système d'annotation INTSINT. Ceci n'est cependant en rien obligatoire. En matière de codage, l'annotation INTSINT est complétée par des algorithmes permettant d'indiquer les informations de durée, d'alignement et de domaine prosodique.

La première étape de modélisation consiste à extraire deux composantes distinctes à partir de la courbe de F0 brute : une composante macroprosodique (le patron mélodique d'un énoncé) et une composante microprosodique (déviations par rapport à ce patron causées par les perturbations au niveau segmental). Ainsi comme l'affirment Hirst et Auran (2005) :

> *« Ideally, our model should separate the components so that when two different texts are pronounced with the same intonation pattern, the macroprosodic component of the two utterances will be the same and when the same text is pronounced with two different tunes, the microprosodic pattern of the two utterances will be the same. »*

On obtient donc la représentation suivante extraite de Hirst et Auran (2005) :

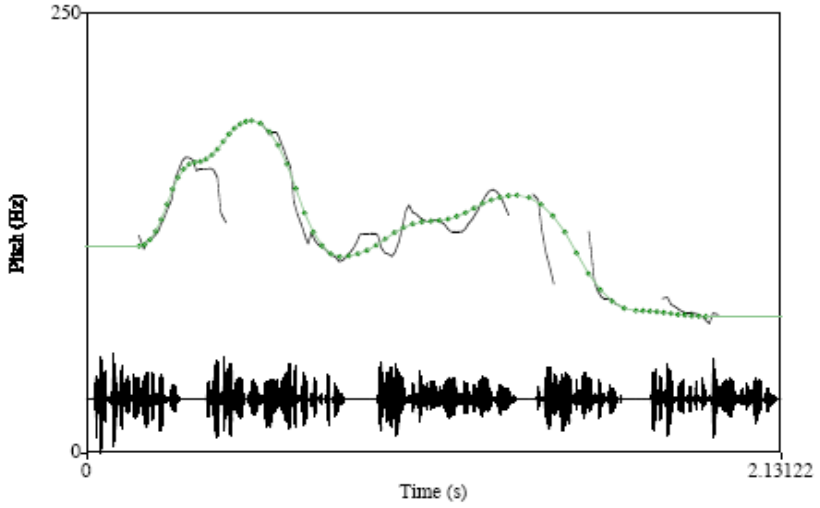


Figure 31 : Courbe de F0 brute (ligne continue) et courbe modélisée (en pointillés) pour le passage « I have a problem with my water softener » du corpus EUROM1

L'alignement des symboles mélodiques est défini par rapport aux frontières de la syllabe la plus proche, les catégories discrètes que l'on obtient sont les suivantes : « initial ([), early (<), middle (:), late (>) and final (]) ». L'unité de temps utilisée pour définir l'alignement des tons est l'unité tonale c'est-à-dire l'unité commençant soit après une frontière intonative soit au début d'une syllabe accentuée et qui s'étend jusqu'à la prochaine frontière intonative ou la prochaine syllabe accentuée. C'est ce que l'on peut voir dans la figure ci-dessous :

{}{M}{T>}{S<L:H]}{U<B>}{}{M<}{T<L>}{}{B[H<L:}{
H:U]}{L<L:U>B]}{}{T>}{S<D>}{H<D>}{S]}{L>}{H<D>
}{U[B:S>H>}{T:}{L<}{H<}{L<}{H<L>}{B>}{}{M<U>}{
H<L:H>}{D[H:}{L<H>}{D<B>}

Figure 32 : Codage INTSINT du passage illustré dans la figure 3.1.

> { } Indiquent les frontières d'unités tonales, l'alignement de chaque symbole est spécifié comme initial ([), tôt (<), médian (:), tardif (>) ou final (]) par rapport aux frontières des unités tonales.

Enfin, les auteurs ajoutent que la représentation phonétique fournie par ce modèle est neutre en ce qui concerne la production et la perception par opposition aux modèles articulatoires par exemple et au modèle IPO. Toutefois, ils ont l'intention avec ce modèle d'avoir accès aux aspects distinctifs issus de la production et de la perception.

4.3 Conclusion

À l'issue de cette revue de lecture de différentes approches de la relation forme/fonction au cœur de l'intonation, on peut procéder à quelques regroupements et généralisations.

Tout d'abord, deux types d'approches semblent se détacher ici : la première présentant une analyse allant des formes vers les fonctions et la seconde beaucoup moins développée allant des fonctions vers les formes. Dans la première catégorie, on peut rassembler le modèle IPO, le modèle des codes biologiques, l'approche compositionnelle, l'approche globale grammaticale, le modèle de Fujisaki et le TILT. Dans toutes ces approches, la relation au sens semble être établie de façon secondaire à l'étude, l'observation et la modélisation des données acoustiques. On peut résumer cette approche grâce à la citation suivante :

> *« Function should be considered only after the phenomenon has been thoroughly explored in its phonetic aspects. »* Collier (1974, p. 21)

C'est à partir des différents patrons intonatifs observés dans les données que l'on va ensuite s'intéresser à leur apport sémantique au niveau de l'énoncé comme dans l'approche compositionnelle de Pierrehumbert qui a défini une grammaire des évènements prosodiques en anglais américain et s'attache ensuite à donner un sens à chacun des éléments de cette grammaire.

Dans la seconde catégorie, on peut regrouper le modèle PENTA de Xu (2005), le modèle de Brazil, le modèle SFC et l'environnement Prozed, qui donnent un rôle primordial à l'analyse des fonctions prosodiques :

> *« […] detailed research on physical features is a secondary investigation to be entered only when we think we have got the meaning right. »* Brazil, Coulthard et Johns (1980, p. 98)

On part ici d'une étude des fonctions véhiculées par la prosodie pour ensuite en générer une grammaire des formes existantes.

Ce qui différencie aussi fondamentalement ces différentes approches de l'apport sémantique de la prosodie, est le « prisme principal » au travers duquel ce processus est explicité. Certains modèles se placent au niveau de la grammaire et de la syntaxe pour expliquer l'apport sémantique de l'intonation, d'autres se placent au niveau du discours. Pour les partisans de l'approche articulatoire, c'est au travers de la connaissance profonde des mécanismes biologiques liés au langage que l'on peut interpréter les fonctions prosodiques, alors que pour les partisans de l'approche IPO, c'est le filtre de la pertinence au niveau perceptif qui permet de donner du sens. Enfin, les chercheurs utilisent les avancées technologiques dans le traitement automatique des langues pour mettre leurs modèles théoriques à l'épreuve.

CONCLUSION GÉNÉRALE

Le développement de cette première partie nous a permis d'explorer et de nous interroger sur les implications, les enjeux, les problématiques liés à un travail de recherche sur la relation entre l'annotation des formes et fonctions prosodiques de l'anglais contemporain.

Nous avons essayé d'apporter des réponses aux questions principales qui découlent de l'étude d'un tel sujet. Qu'entend-on par fonction prosodique et quelles fonctions sont généralement attribuées à l'intonation dans la littérature ?

Qu'entend-on par forme prosodique, et quelles sont les caractéristiques des principaux modèles de représentation des formes prosodiques en anglais britannique ? Comment l'intonation est décrite et analysée dans la littérature ?

Quel système de transcription peut-on utiliser, quelles sont les caractéristiques des principaux systèmes de transcription de l'anglais ?

Quelle approche méthodologique adopter pour explorer l'articulation forme/fonction : pouvons-nous nous inspirer des différentes approches existantes ?

À la question des fonctions attribuées à l'intonation, nous répondons qu'il nous paraît impossible de faire une liste complète de celles-ci. De multiples fonctions ont été évoquées dans la littérature. Au-delà de ce fait si l'on essaie de réduire ces fonctions à une liste réduite qui semble mettre une majorité des chercheurs d'accord telle que la fonction grammaticale, lexicale, etc., on se rend compte que ces termes ne sont pas utilisés de la même façon dans les différentes approches. Ces fonctions ne sont pas non plus totalement discrètes et indépendantes : l'intonation n'a pas un rôle purement grammatical dans certains cas et purement expressif dans d'autres, mais sert plutôt probablement à véhiculer plusieurs types de significations à la fois. De plus elle interagit avec d'autres canaux de la communication comme le lexique et le contexte. Comment extraire alors la part de sens véhiculé par l'intonation seule ? Cela reste une question ouverte et complexe.

Dans un deuxième chapitre, c'est à la façon dont les formes prosodiques ont été décrites dans la littérature que nous nous sommes intéressée. Et à nouveau nous avons affaire à paysage contrasté. En ce qui concerne les formes de l'intonation, deux aspects nous intéressent tout particulièrement : les configurations de F0 et leur alignement. Deux grands types de modèles s'opposent : les modèles britanniques et les modèles autosegmentaux métriques. On constate alors la multiplicité des inventaires des formes prosodiques, la diversité dans la façon d'analyser le contour intonatif : modèles du nucléus/accents mélodiques, accent de phrase et accent de frontière. Au niveau des modèles britanniques, nous avons relevé des problèmes non résolus comme la question de la nature des accents postnucléaires dans les contours de chute + montée,

la nature du nucléus ou l'opposition entre les accents mélodiques et les accents rythmiques. Les modèles autosegmentaux ouvrent la voie à des études expérimentales sur l'alignement tonal. Si celles-ci n'ont pas encore permis à notre connaissance de fournir un modèle invariant explicitant les phénomènes d'alignement, elles permettent de mieux connaître les facteurs influençant l'alignement des points cibles, les contraintes articulatoires et physiologiques ainsi que les principaux points d'ancrage mis en valeur.

La question de l'annotation prosodique nous a aussi permis de mettre en valeur différentes conceptions de l'intonation (on pense à l'opposition entre TOBI et les TSM). Nous avons montré que les limites des plus grands systèmes d'annotation pouvaient être diminuées en séparant l'annotation des formes et des fonctions prosodiques et avons montré l'intérêt de systèmes d'annotation permettant cette séparation (MOMEL-INTSINT et IF).

Pour finir, on s'est interrogé sur les méthodes d'approche de l'articulation forme/fonction. Des approches théoriques comme les approches syntaxiques et pragmatiques ont été opposées à des approches plus expérimentales se plaçant du point de vue de la perception, de la physiologie (avec les modèles articulatoires par exemple), ou des méthodes utilisant le traitement automatique et notamment l'analyse par synthèse. Deux tendances ont été mises en valeur : celle qui donne priorité à la description des formes prosodiques avant de s'attacher à en décrire les fonctions et l'approche concurrente basée sur une analyse préliminaire essentielle des fonctions prosodiques.

Il nous semble évident que les deux partis ont raison. Afin de comprendre, analyser et décrire le système prosodique d'une langue, il faut en connaître les formes c'est-à-dire être en mesure d'évaluer, d'observer et de distinguer les différentes formes existantes au moyen d'oscillogramme, de spectrogrammes ou d'algorithmes. Il faut aussi avoir une idée de ce qui est important, de ce qui est distinctif ou significatif dans un système donné et s'intéresser donc à l'apport fonctionnel de l'intonation dans une langue. À quoi sert l'intonation, à quoi peut-on s'attendre ? Cependant on sait aussi que d'autres facteurs multiples interviennent dans la création du sens si bien que la relation forme/fonction n'est pas directe et transparente.

On peut alors réconcilier Brazil (1975) et Collier (1974) en mettant en place une méthode de recherche basée sur un constant aller-retour entre l'analyse des formes de l'intonation et celle de ses fonctions.

Nous ne disposons pas d'un modèle des formes prosodiques représentatives unique et consensuel. Nous avons besoin d'un outil (un outil aussi neutre que possible au niveau de la représentation des formes et fonctions) qui nous permettrait de tester nos idées. On pourrait alors partir d'un modèle de représentation minimal contenant peu d'aprioris théoriques et le confronter à la réalité des données en essayant de prédire les contours de F0 à partir de celui-ci. Il y aura des manques, des erreurs de prédiction et l'on compte sur cette fenêtre d'évaluation pour revenir sur notre modèle de base et le remettre en question puis l'enrichir jusqu'à toucher à des apports significatifs de plus en plus complexes.

Il nous semble qu'une telle démarche est possible si l'on utilise des systèmes de transcription distincts pour représenter les formes et fonctions de l'intonation et une méthode d'analyse par synthèse. C'est la méthodologie que nous allons tenter de mettre en place dans la seconde partie de ce mémoire.

PARTIE II

TRAVAIL EXPÉRIMENTAL

Dans la première partie de cette thèse, nous avons présenté différentes approches de notre problématique qui est l'explicitation de l'articulation forme/fonction au cœur de l'intonation. Nous nous sommes interrogée sur la nature et définition des formes et fonctions prosodiques à partir d'une revue de la littérature actuelle sur le sujet ainsi que sur les implications et enjeux d'un tel travail de recherche. La deuxième partie de ce mémoire est consacrée aux expériences menées pour développer une nouvelle méthode d'investigation de cette articulation forme/fonction.

Cette seconde partie s'articule autour de quatre axes principaux : premièrement la présentation de nos choix de corpus et de systèmes d'annotation prosodique, deuxièmement, la mise en place d'une méthode de recherche qui est l'analyse par synthèse, troisièmement les différentes étapes d'optimisation des paramètres du modèle d'origine à partir de deux corpus (EUROM1 et MARSEC) et enfin un chapitre analytique traitant des résultats de nos expériences. Notre travail expérimental est donc basé sur deux corpus différents, le premier nous a essentiellement permis de développer une méthode de recherche ainsi que différents outils nécessaires au traitement automatique des données, le second est utilisé afin d'approfondir notre recherche sur l'articulation forme/fonction.

Chapitre 5

Choix des corpus et de l'annotation

Pour ce travail expérimental, nous avons choisi d'utiliser deux corpus de base : le corpus EUROM1 et le corpus MARSEC. Tout d'abord, notre premier objectif était de mettre en place et de tester une méthode d'approche particulière de l'articulation forme/fonction qui est l'analyse par synthèse dont on a déjà parlé dans le chapitre précédent. Cette approche sera présentée en détail dans la section suivante. On peut toutefois avancer qu'une telle approche implique la création de toute une batterie de scripts permettant l'automatisation de la procédure. Dans notre première tentative, nous avons cependant utilisé une annotation manuelle comme point de départ. Il nous fallait donc un corpus de petite taille d'où le choix du corpus EUROM1 dont on a extrait une partie. Ce corpus présente aussi l'intérêt d'être assez simple : il s'agit d'un corpus de lecture de phrases que l'on peut qualifier de parole de laboratoire. De plus ce corpus est disponible gratuitement et comporte différents niveaux d'annotation de base très utiles pour notre travail expérimental.

Après avoir mis en place notre procédure d'analyse par synthèse à partir du corpus EUROM1, nous avons souhaité aller plus loin en utilisant un corpus de base plus riche et un style de parole plus expressif. De cette façon nous pensions pouvoir enrichir notre annotation fonctionnelle de base en la confrontant à des données plus complexes. En effet suivant Auran *et al.* (2004), on peut parler de parole authentique dans le corpus MARSEC par opposition à la parole de laboratoire contenue dans le corpus EUROM1. Le terme authentique fait ici référence à de la parole destinée non pas à réaliser une expérience mais à communiquer et donc à délivrer un véritable message.

De plus nous avons choisi de nous focaliser sur la partie intitulée « fiction » de ce corpus contenant des lectures de nouvelles pour enfants et adultes animées par des acteurs professionnels.

Nous présenterons donc d'abord l'extrait du corpus EUROM1 que nous avons utilisé pour nos expériences puis le corpus MARSEC. Enfin les annotations du corpus EUROM1 seront détaillées dans une dernière section.

5.1 Le corpus EUROM1

EUROM1 est probablement une des premières bases de données de parole produite en Europe. Des corpus équivalents pour chaque langue européenne ont été collectés avec le même nombre de locuteurs sélectionnés de la même manière et enregistrés dans les mêmes conditions et mêmes formats de fichiers. La base de données EUROM1 contient des enregistrements de 60 locuteurs en huit langues européennes. Elle a été explicitement conçue pour faciliter les

comparaisons phonétiques entre langues, en s'appuyant sur des matériaux similaires et des protocoles d'enregistrement identiques dans toutes les langues. Au départ, huit pays d'Europe ont participé aux enregistrements : l'Italie, le Royaume-Uni, l'Allemagne, les Pays-Bas, le Danemark, la Suède, la Norvège et la France. Cette base a ensuite été complétée par les enregistrements réalisés en Grèce, au Portugal et en Espagne (grâce au projet SAM-A du programme européen Esprit (Chan, D & *al.* 1995)).

Le contenu consiste en :

1. De la parole continue :

   40 passages composés de phrases,
   50 phrases d'ajustement (« patching »), constituées dans le but de compenser la répartition des phonèmes irréguliers dans les passages.

2. Des nombres :

   Les nombres ont été divisés en cinq blocs, chacun contenant vingt nombres. Chaque bloc a été enregistré d'un seul trait.

3. Des mots CVC :

   Les listes de mots CVC contiennent cinq types de listes ainsi que des phrases porteuses correspondant au type. Quatre-vingt-deux mots isolés ont été utilisés.

Le corpus utilisé pour ce travail de recherche est dans un premier temps le corpus EUROM1 anglais (Il s'agit essentiellement d'anglais britannique méridional). Nous avons choisi quatre locuteurs parmi l'ensemble des participants. Il s'agit de deux hommes et de deux femmes. Le corpus retenu est composé de 53 fichiers son contenant chacun la lecture de cinq phrases. Il y a donc 265 phrases lues en tout.

Le corpus contient déjà plusieurs niveaux d'annotations manuelles et automatiques :

- Un découpage en phonèmes ;
- Une transcription orthographique en mots ;
- Une annotation manuelle des pauses silencieuses ;
- Une « rangée[4] » MOMEL et INTSINT avec corrections manuelles sur MOMEL.

C'est sur cet étiquetage de base que nous nous sommes appuyée pour développer les niveaux d'annotation fonctionnelle que l'on présentera dans la dernière section de cette partie.

[4]: nous employons le terme « rangée » pour traduire le terme anglais « tier » en suivant l'exemple de H.S Cho, dans sa thèse de doctorat récente (septembre 2009). Nous préférons ce terme à celui de « tire » imité de l'anglais et utilisé par François Dell et Jacques Durand.

Commençons par présenter le second corpus de base utilisé pour nos expériences, le corpus Aix-MARSEC.

5.2 Le corpus Aix-MARSEC

Le corpus Aix-MARSEC a été élaboré à partir du corpus SEC (« Spoken English Corpus ») enregistré à IBM UK dans les années quatre-vingt. Développé par Peter Roach en 1994, il devient le corpus MARSEC (« Machine Readable Corpus ») avant de prendre sa forme la plus complète en tant que corpus Aix-MARSEC. L'annotation, l'alignement des phonèmes et la correction de la transcription orthographique ont été effectués au laboratoire LPL à Aix-en-Provence par Daniel Hirst, Cyril Auran et Caroline Bouzon. C'est cette dernière version complète que nous utilisons pour ce travail.

Le corpus MARSEC représente cinq heures et demie de parole. Il regroupe des styles de parole très variés allant de la lecture de romans aux enregistrements de la BBC ou de sermons, de cours de mathématiques. Nous parlons ici de parole authentique suivant Auran *et al.* (2004), car il s'agit de parole destinée à être écoutée, à délivrer un message ou à communiquer. Il ne s'agit donc pas de parole enregistrée pour la recherche uniquement. Le corpus est découpé en fichiers d'une minute environ à l'intérieur de chaque style de parole.

Nous nous concentrerons ici sur la partie « fiction ». Cinq locuteurs différents proposent la lecture de nouvelles pour enfants et pour adultes. Cette partie contient 54 fichiers sons et représente 47 minutes de parole quasiment continue.

Chaque fichier contient les éléments suivants alignés avec le signal ;

- une transcription phonémique ;
- une transcription syllabique avec la marque de l'accent lexical ;
- un découpage en pieds métriques selon Abercrombie et en unités rythmiques étroites (NRU) et en anacruses (ANA) selon Jassem ;
- une transcription orthographique et enfin un étiquetage prosodique en TSM (cf. figure ci-dessous) ;
- les tons INTSINT ;
- la « rangée » MOMEL représentant les valeurs des points cibles de F0 que nous avons corrigés manuellement.

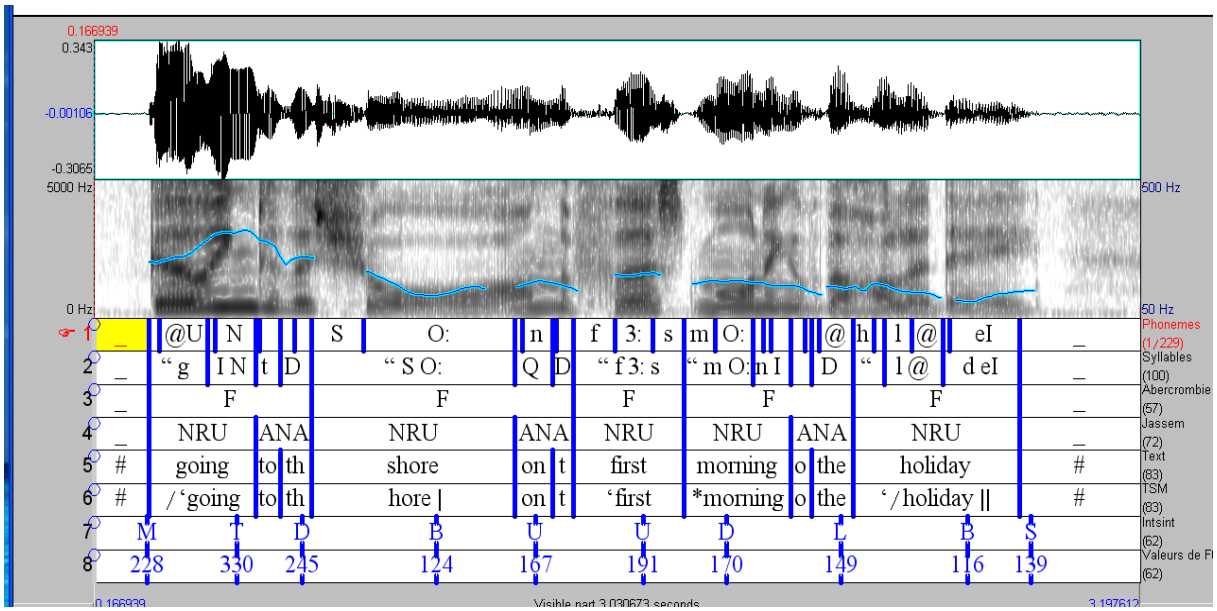


Figure 33 : TextGrid G0101B avec l’étiquetage initial du corpus.

Nous allons surtout nous concentrer sur les trois dernières rangées qui représentent notre information de base pour l’annotation des formes et des fonctions. La sixième rangée représente les TSM, la septième les séquences de tons INTSINT et la dernière les valeurs des points cibles MOMEL. La rangée des points cibles MOMEL corrigés manuellement permettra de générer la courbe de référence à partir de laquelle nous testerons la qualité de nos prédictions tout comme dans le travail effectué sur EUROM1.

La transcription prosodique a été réalisée par Gerry Knowles et Briony Williams. Il s’agit d’une base d’information unique pour un corpus de si grande taille et d’un apport précieux pour notre travail d’annotation.

Elle sera utilisée pour extraire une annotation fonctionnelle de départ. Comme nous l’avons expliqué dans le premier chapitre de cette thèse, les TSM sont un système d’annotation prosodique qui annote des formes tout autant que des fonctions. Nous allons donc tenter de convertir ce type d’annotation en annotation IF en extrayant par règle seulement l’information fonctionnelle.

5.3 Développements de l’annotation du corpus EUROM1

Nous avons d’abord ajouté à l’étiquetage de départ, un découpage en syllabes et la marque de l’accent lexical. Le découpage automatique en phonèmes a aussi fait l’objet de corrections manuelles. Les annotations fonctionnelles ont été réalisées au niveau des unités intonatives et des unités tonales selon une tradition remontant aux travaux de Jassem (1952) (les abréviations anglaises « TU » et « IU » seront utilisées dans le reste du document).

Cette annotation représente un travail d’écoute du corpus en prenant à chaque fois un énoncé entre deux pauses. Les questions qui se posent alors sont les suivantes : cet énoncé contient-il

plusieurs sous-énoncés ? Peut-on introduire une pause sans altérer la cohérence mélodique (cf. Cho et Hirst 2006) ? Si l'on perçoit une frontière, de quel type de frontière s'agit-il ? (l'énoncé paraît-il être terminé ou continuatif ?) À l'intérieur de chaque groupe ainsi défini quelles sont les syllabes qui sont accentuées, quelle est la syllabe la plus proéminente de l'énoncé ? Y a-t-il une emphase ?

Il faut bien préciser que l'on parle d'accent au niveau mélodique et non pas d'accent lexical. Il ne s'agit pas de se contenter de noter les accents lexicaux qui font partie de l'identité de chaque mot même hors contexte mais de relever les mots qui sont perçus comme proéminents dans un contexte donné.

Ces aspects fonctionnels sont codifiés dans le système IF dont on changera la forme de codage pour ce travail. Au lieu d'utiliser les diacritiques « 0 », « ' », « * », « !* » correspondant respectivement à une unité tonale inaccentuée, accentuée, nucléaire et emphatique nucléaire, nous utiliserons une annotation alphabétique. De cette façon il sera plus aisé dans les phases expérimentales suivantes d'enrichir l'annotation fonctionnelle de base. On obtient le codage suivant :

- « U » pour un TU inaccentué (= « unaccented ») ;
- « A » pour un TU accentué (= « accented ») ;
- « N » pour un nucléus non marqué ;
- « !N » pour un nucléus emphatique.

Les étapes d'annotation sont alors les suivantes :

1. Une annotation des frontières intermédiaires entre deux pauses silencieuses[5] ;
2. Une identification du type de frontière auquel on a affaire : terminal ou non terminal « | » ou « + » ;
3. Une annotation des syllabes proéminentes ;
4. Un découpage des unités intonatives en unités tonales allant d'une syllabe accentuée jusqu'à la prochaine syllabe accentuée et leur identification (U, A, N, !N).

[5] Les pauses silencieuses sont étiquetées avec le symbole { #})

Ci-dessous un exemple de passage annoté en IF :

[^{u}Mu^{A}nicipal ^{N}Fire ^{A}Service ^{A}speaking| # [^{u}We're ^{A}trying to lo^{A}cate an e^{A}mergency ^{N}caller+ [^{u}who ^{A}rang ^{N}off+ # [^{u}wi^{A}thout ^{A}giving any ^{A}personal ^{N}details| # [^{u}He ap^{A}peared to ^{A}be on the ^{A}local ^{N}network| # [^{u}He con^{A}nected on our ! ^{N}line ^{A}number+ # [^{A}seven six ^{N}two+ ^{A}five ^{A}eight ^{N}four| # [!^{u}We'd ap^{A}preciate im^{A}mediate at^{A}tempts to ^{N}trace him| [^{u}be^{A}cause he ^{A}sounded ^{N}desperate|

Figure 34 : Passage FAo30072 avec l'annotation fonctionnelle utilisée dans l'analyse.

La première unité intonative de ce passage sous la forme d'un TextGrid est montrée dans la figure ci-dessous.

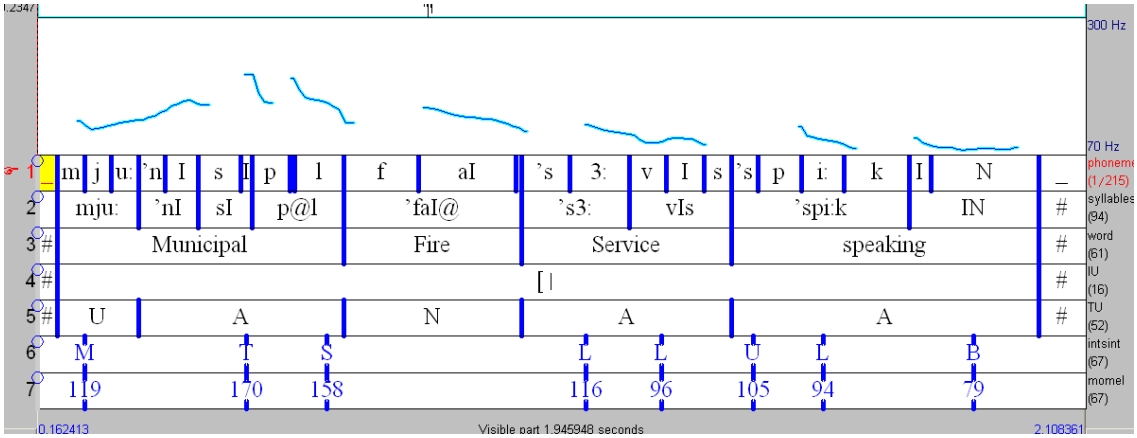


Figure 35 : Un extrait du TextGrid correspondant au premier IU du passage ci-dessus.

> Rangée 1= phonèmes, rangée 2= syllabes, rangée 3= mots, rangée 4= unités intonatives, rangé 5= unités tonales, rangée 6= tons INTSINT, rangée 7= points cibles de MOMEL.

Le découpage en unités tonales correspond au modèle développé par Jassem (1952). Le principe de base est en accord avec le système IF (Hirst 2005). Il s'agit d'une annotation purement fonctionnelle, on se concentre donc sur l'information perçue ou le message comme un point de départ avant d'observer ensuite les formes correspondant à ces fonctions annotées. Ces formes et leur annotation sont présentées dans la section suivante portant sur la description de l'approche de l'analyse par synthèse de l'articulation forme/fonction.

Chapitre 6

Analyse par synthèse des contours intonatifs de l'anglais britannique

6.1 Méthode d'analyse

Nous nous proposons de caractériser la relation entre la représentation des formes et des fonctions prosodiques en utilisant l'analyse par la synthèse. Dans le contexte scientifique de cette thèse, plusieurs exemples d'utilisation de l'analyse par synthèse appliqués à l'étude de la prosodie ont été présentés. Tout comme les auteurs cités, nous utiliserons l'analyse par synthèse comme un outil privilégié pour simuler et évaluer des modèles de l'intonation. Les imperfections au niveau de l'intonation produite par le système seront utilisées comme des indices d'erreurs ou de lacunes dans le modèle de représentation choisi que l'on cherchera à enrichir.

Ce travail s'inscrit dans la continuité du cadre d'analyse par synthèse PROZED. Nous présupposons pour cette analyse, trois niveaux intermédiaires de représentation entre le niveau du signal acoustique et le niveau fonctionnel (Hirst 2000) : la représentation phonétique, la représentation phonologique de surface et la représentation phonologique sous-jacente. Nous nous intéresserons essentiellement dans ce travail à la relation entre les deux premiers niveaux de représentation des formes (phonétique et phonologique de surface) et la représentation fonctionnelle.

La représentation phonétique est fournie par l'algorithme MOMEL qui permet d'obtenir une représentation automatique discrète de la courbe brute de la fréquence fondamentale sous forme d'une séquence de points cibles. On obtient ainsi une courbe lisse dépourvue de perturbations micro-mélodiques.

Les points cibles sont ensuite convertis automatiquement en symboles tonals au moyen de l'alphabet INTSINT (représentation phonologique de surface).

Enfin le système de codage IF (Hirst, 1977, 2005), représente le niveau d'annotation fonctionnelle. Nous prenons en considération deux unités prosodiques : l'unité intonative caractérisant les types de frontières (terminal/non terminal) et l'unité tonale pour annoter les niveaux de proéminence (inaccentué, accentué, nucléaire, emphatique).

Il s'agit alors de générer des représentations formelles à partir de cette représentation minimale des fonctions prosodiques et de comparer le résultat obtenu avec les données observées. Cela permet ensuite d'enrichir la représentation fonctionnelle et de voir si cela permet une meilleure reproduction des données.

Cinq modèles successifs, de plus en plus complexes de l'articulation entre formes et fonctions sont présentés et mis en œuvre sur 53 fichiers son du corpus EUROM1 (chacun contenant cinq phrases lues). Ces modèles sont inspirés par les modèles descriptifs des contours intonatifs développés par les théoriciens de l'école britannique de l'intonation. Les contours décrits par ceux-ci sont traduits en tons INTSINT puis convertis en points cibles MOMEL pour la resynthèse.

Nous évaluons enfin la qualité de la prédiction des modèles en calculant la corrélation linéaire entre la courbe de F0 modélisée et la courbe de MOMEL corrigée manuellement pour chaque unité intonative, mais aussi en écoutant les résultats de la resynthèse des modèles que nous avons prédits. Nous disposons donc de deux types d'évaluation de nos modèles de synthèse : une évaluation objective quantifiée par la corrélation linéaire et une évaluation subjective rendue possible par la resynthèse sous PSOLA.

Il s'agit d'une approche incrémentale comparant à chaque niveau d'approximation le résultat de la prédiction et les patrons intonatifs observés dans les données. À ce niveau d'expérimentation, l'annotation fonctionnelle est faite manuellement.

Les règles de dérivation de l'annotation fonctionnelle vers l'annotation formelle par l'alphabet INTSINT sont implémentées dans un script Praat. Les symboles INTSINT sont ensuite convertis en points cibles que l'on interpole quadratiquement afin d'obtenir un « pitch-tier » utilisable dans PSOLA pour la resynthèse. C'est ce segment ainsi resynthétisé qui peut faire l'objet d'une évaluation subjective.

6.2 Modèles d'analyse par synthèse de l'intonation de l'anglais

Les modèles développés dans cette section s'inspirent des modèles prosodiques développés dans la tradition de recherche britannique d'une part et de la simple observation des données dont nous disposons d'autre part.

6.2.1 Le modèle « nul »

Le modèle « nul » s'appuie sur l'hypothèse que l'intonation n'est dérivée d'aucune information fonctionnelle. Aucun repère fonctionnel au niveau segmental ne serait donc nécessaire dans le but de prédire et décrire les variations de F0.

Prenons donc l'exemple d'un passage du corpus EUROM1 en entier (cf. figure 36). L'observation la plus évidente semble être une diminution progressive de la valeur de F0 tout au long du passage après avoir atteint une hauteur maximale au début de l'énoncé. C'est ce qui a été décrit dans la littérature comme l'effet de déclinaison ou « downstepping effect ». Les cordes vocales lors d'un acte de parole ne sont pas maintenues à la même fréquence de vibration et se relâchent progressivement de façon naturelle. Il s'agit ensuite de traduire cet effet à l'aide

de tons INTSINT. On peut interpréter ce contour globalement descendant en utilisant les tons INTSINT suivants : M (« mid ») pour le point de départ, T (« top ») pour le contour montant initial, et le ton B en fin d'énoncé pour marquer la fin du mouvement descendant. La figure 36 montre la courbe du modèle superposée à la courbe dérivée des points cibles de MOMEL pour un passage.

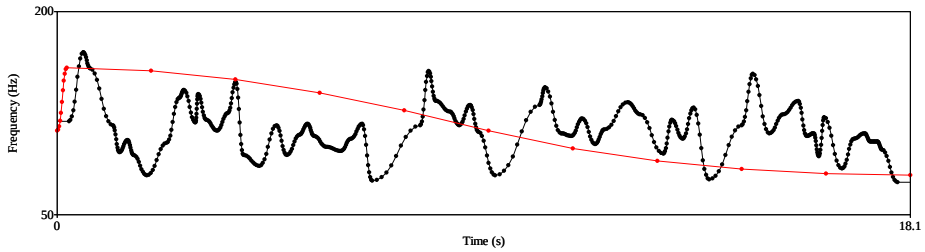


Figure 36 : La courbe de F0 générée par MOMEL superposée au résultat du modèle « nul ».

En rouge la courbe de F0 dérivée de l'annotation INTSINT [MT…..B] superposée à la courbe crée par MOMEL interpolée quadratiquement pour le passage FAo30072 du corpus EUROM1.

Évidemment, on se rend très vite compte que cette interprétation des variations de F0 est très simpliste et réductrice ce qui nous pousse à prendre en compte plus d'informations fonctionnelles comme le montrent les modèles suivants.

6.2.2 Le modèle « IU »

De toute évidence le contour intonatif du passage illustré dans la figure 36, ne consiste pas en un seul contour continu mais plutôt en une séquence de contours distincts constituant plusieurs « blocs successifs » à l'intérieur du passage. Ces différents énoncés correspondent aux unités intonatives ou « sense group » chez O'Connor et Arnold (1961) ou « tone group » chez Palmer (1924). Fujisaki (1983) parle aussi dans ce cas-là de commande de phrase et associe ces commandes avec un contour globalement descendant. On peut alors utiliser la même séquence de tons INTSINT que dans le modèle précédent pour décrire ces contours en se plaçant cette fois-ci au niveau de chaque unité intonative. Le texte est annoté grâce à des symboles de frontière ([ ]) le divisant en unités intonatives. Chaque unité intonative est alors modélisée en une séquence de tons {M-T…..B} où M et B sont alignés avec les frontières gauches et droites de la séquence. T est aligné à une distance constante du début de l'unité qui est fixée de façon empirique à 200 ms. Fujisaki (1983) mentionne aussi ce chiffre de 200 ms pour indiquer le début de la commande de phrase.

La représentation fonctionnelle a été effectuée sous Praat, dans un TextGrid où chaque intervalle correspond à une unité intonative permettant de générer les contours les plus proches des points cibles manuellement corrigés.

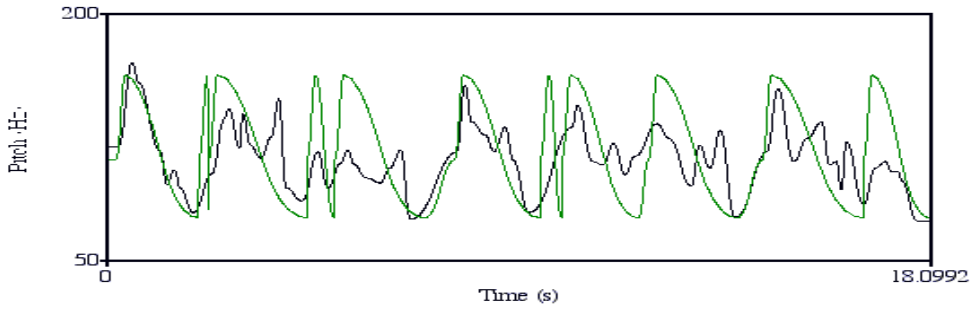


Figure 37 : La courbe de F0 générée par MOMEL superposée au résultat du modèle IU.

On observe déjà dans ce modèle basé sur les unités intonatives, une nette amélioration visuelle en comparaison avec le modèle « nul ».

6.2.3 Le modèle « accent»

À l'image de Fujisaki (1983), après avoir souligné l'importance des variations de F0 globales au niveau de l'énoncé ou de l'unité intonative, on s'intéresse aux variations locales de F0 à l'intérieur de ces unités intonatives. Celles-ci sont traditionnellement associées au niveau fonctionnel, à l'accent de mot ou de façon plus large aux proéminences. Cette question est loin d'être anodine en anglais, car comme on l'a déjà mentionné dans le chapitre théorique, le phénomène d'accentuation existe au niveau lexical et supra-lexical. On parle d'accents de mot et d'accents de phrase ou d'accents lexicaux par opposition à des accents mélodiques.

Tous les accents lexicaux ne portent pas forcément une proéminence supérieure au niveau de la phrase (accent de phrase ou « stressed » *vs* « accented »). Ces accents de phrase peuvent aussi marquer des syllabes non accentuées lexicalement (bien que ce ne soit pas le cas dans ce corpus). Les accents de phrase sont représentés dans Praat sur une seconde rangée nommée « Tonal Unit » (unité tonale) suivant une tradition remontant aux travaux de Jassem (1952). D'après celui-ci les accents de mot participeraient à la création du rythme par opposition aux accents de phrase qui s'inscrivent dans l'intonation. Il décrivait donc des unités rythmiques et des unités tonales.

Les théoriciens britanniques décrivent le contour standard de référence en anglais comme un contour globalement descendant, utilisant les syllabes accentuées comme des paliers successifs amenant progressivement F0 jusqu'au niveau le plus grave de la tessiture d'un locuteur. Il s'agit du « downstepping pattern ».

On peut traduire ce contour en tons INTSINT de la façon suivante : le ton M est aligné avec le début de chaque unité intonative et B avec la fin de celle-ci. Le ton T est aligné avec le début de la première unité tonale de chaque unité intonative qui est annotée « A » et un ton D est aligné avec les unités tonales suivantes. Un ton B est aligné après le dernier T ou D de la dernière unité tonale étiquetée « A ».

Il manque un élément fondamental à cette interprétation du contour de base de l'anglais : il s'agit du nucléus décrit très longuement dans la littérature britannique. D'après ces derniers, le nucléus permet de distinguer les accents dits prénucléaires des accents postnucléaires. Il est donc considéré comme un pivot, un centre à partir duquel la mélodie est construite et change.

6.2.3.1 Le Nucléus

Adoptant l'annotation fonctionnelle IF de Hirst (2005), on introduit ensuite la notion de nucléus ou de centre intonatif. La structure intonative que l'on décrit alors est familière dans la tradition britannique de l'intonation (O'Connor 1961, Cruttenden 1986) :

(prehead) (head body) nucleus (tail)

She's going to the cinema after school.

Les tons associés à cette structure sont alors :

(prehead) (head(body…….)) nucleus(tail)

[	A	A	N	A	|
M	T	D	D/H	-B	B

C'est le « tone group II » chez O'Connor and Arnold (1961) : « stepping head+low fall ».

Ou de manière plus compacte :

[M (^{A}T (^{A}D)) ^{N}H/D ^{A}-BB|]

où le choix entre H/D dépend du fait qu'il y ait ou non une tête ou « head » précédant ce ton. S'il y a une ou des syllabes de tête précédant le nucléus, celui-ci est marqué par le contour {DB} et s'il n'y en a pas par le contour {HD}.

6.2.4 Le modèle « terminal »

Dans les modèles précédents, toutes les unités intonatives sont représentées par un contour descendant. Toutefois, très tôt dans l'étude de l'intonation de l'anglais, deux contours de base ont été décrits : l'un globalement descendant et se terminant par une chute de F0, l'autre identique en partie mais se terminant par une montée de F0. Ce sont les contours I et II décrits par Jones (1909) et Armstrong et Ward (1931).

Au niveau fonctionnel, cela permet de distinguer le caractère final ou terminal du caractère non terminal d'un énoncé (terminal « [| » / « [+ ») (Hirst 1977). L'interprétation en tons INTSINT peut alors être la suivante :

[M (^{A}T (^{A}D)) ^{N}H/D -^{A}BB|] *vs* [M (^{A}T (^{A}D)) ^{N}H/D -^{A}BH+]

Ce dernier contour correspond au « contour 7 » ou «tone group VII» de O'Connor et Arnold (1961) appelé : « Stepping head + low rise ».

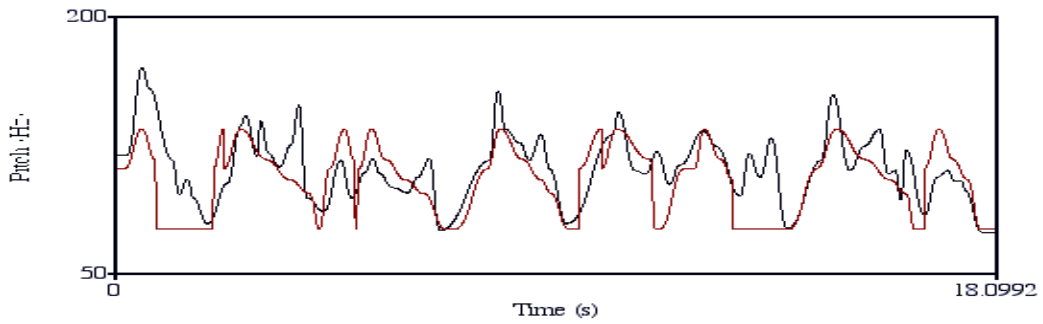


Figure 38 : La courbe de F0 générée par MOMEL superposée au résultat du modèle accent.

6.2.5 Le modèle « emphatique »

Dans la plupart des études portant sur l'intonation de l'anglais britannique, on distingue des contours marqués de contours non marqués (cf. Kingdon 1958, Armstrong et Ward 1931, etc.). Cette emphase particulière peut s'appliquer à un niveau local (celui du centre intonatif) ou à un niveau global («special heads and preheads»).

6.2.5.1 Le nucléus emphatique

La caractéristique de base d'un nucléus emphatique est la réalisation de son point de départ à une valeur de F0 supérieure à la hauteur des accents précédents. Cela correspond à l'opposition entre la petite chute (« low fall ») et la grande chute (« high fall ») chez O'Connor et Arnold (1961) ou la simple chute (« Falling nucleus ») par opposition à la chute intensifiée (« intensified falling nucleus ») chez Palmer (1924).

Cette fonction emphatique est représentée dans le système IF par un point d'exclamation apposé à l'étiquette nucléus : « !N ».

On peut traduire le contour emphatique en tons INTSINT par une séquence {TB} au niveau du nucléus plutôt que {DB} ou {HB} s'il n'y a pas de syllabes de tête. Ainsi la continuité induite par le ton D (« downstep ») est interrompue pour effectuer une nouvelle montée à partir de la syllabe emphatique avec le ton T.

[M (^{A}T (^{A}D)) ^{N}T -^{A}BB/H|/+]

6.2.5.2 La tonique de tête emphatique

Le nucléus emphatique terminal ou non terminal peut être présignalé au niveau de la tête de l'énoncé (Hirst 1998). Il existe donc des contours prénucléaires plus marqués mettant en valeur

le nucléus emphatique. On peut alors noter le caractère emphatique au niveau du TU pour le nucléus ou au niveau de l'IU pour une tête emphatique.

Nous avons pris l'exemple de deux contours emphatiques décrits par O'Connor et Arnold (1961) : la tête montante (« upstepping head ») et la tête descendante montante (« sliding head ») associée à une grande chute dans le premier cas et à une chute-montée dans le second. D'après Hirst (1998), la tête descendante montante a une valeur neutre ou non marquée dans certains dialectes tels que l'anglais américain mais en anglais britannique standard, on lui préfère le contour descendant progressif en paliers (« downstepping Head »). L'utilisation d'un tel contour est ainsi perçue comme emphatique.

On utilise alors deux séquences de tons INTSINT : soit une suite de chutes {T+L, H+L H+L, T+B+U} pour un IU non terminal soit une séquence ascendante {B+U+U + TB} pour un IU terminal.

6.2.5.3 Cas particuliers

Certains phénomènes intonatifs peuvent être annotés avec une utilisation spécifique des frontières intonatives (Hirst 1998). Cela signifie qu'une unité intonative pourrait n'avoir qu'une frontière initiale « [» mais aucune frontière finale (cf. the « interrupted glide down » de Kingdon 1958) ou le cas contraire comme dans l'intonation des parenthèses, les commentaires ou « afterthoughts » et de la chute + montée (« divided fall-rise »).

La dernière modélisation contient les catégories fonctionnelles suivantes :

Unités intonatives : { [, +, |, [|, [+, [!, [!+, [!|}

Unités tonales *:* {U0, A, N, !N }

Cela permet de définir les contours intonatifs suivants :

1) Un contour initial interrompu (« interrupted glide down ») :

   [M (^{A}T (^{A}D$_1$))

4. Un contour global neutre et terminal :

   [M (^{A}T (^{A}D$_1$)) ^{N}H/D-B B|

5. Un contour neutre non terminal :

   [M (^{A}T (^{A}D$_1$)) ^{N}H/D-B H+

6. Un contour terminal emphatique :

   [M (^{A}T (^{A}D$_1$)) ! ^{N}T -B B|

7. Un contour emphatique non terminal :

   [M (^{A}T (^{A}D$_1$)) ! ^{N}T-B H+]

8. Un contour terminal avec une emphase sur la tête et le nucléus :

   [!T (^{A}B (^{A}U$_1$)) ! ^{N}T-B B|

9. Un contour non terminal avec une tête et un nucléus emphatique :

   [!M (^{A}TL (^{A}HL$_1$)) ! ^{N}T-B H+

10. Une parenthèse terminale basse :

    (^{A}B (^{A}B)) B|

11. Une parenthèse basse non terminale :

    (^{A}B (^{A}B)) H+

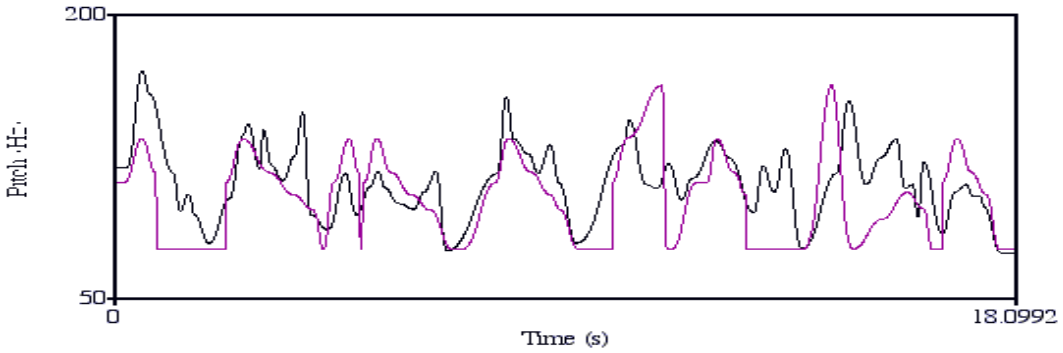


Figure 39 : La courbe issue du modèle emphatique superposée à la courbe issue de MOMEL.

Nous avons ensuite calculé la corrélation linéaire entre les courbes issues de nos modèles et les courbes MOMEL corrigées manuellement pour le corpus EUROM1.

6.3 Évaluation des résultats

La courbe brute de la fréquence fondamentale des enregistrements a été modélisée avec l'algorithme MOMEL. Les paramètres spécifiques au registre de chaque locuteur c'est-à-dire la clé et l'étendue, sont calculés automatiquement par l'algorithme INTSINT. Les contours modélisés en tons INTSINT décrits ci-dessus sont convertis en représentations phonétiques, c'est-à-dire en points cibles MOMEL automatiquement par l'intermédiaire d'un script Praat. C'est cette nouvelle courbe MOMEL issue de nos modèles que l'on compare à la courbe MOMEL d'origine. On a ensuite calculé pour chaque IU, le coefficient de corrélation linéaire entre la courbe MOMEL d'origine et la courbe issue des différents modèles comme on peut le voir dans le tableau ci-dessous.

Tous	correlation	[+	correlation	[|	correlation
iu	0,597	iu	0,613	emphasis	0,658
accent	0,559	accent	0,574	terminal	0,650
emphasis	0,549	terminal	0,510	iu	0,636
terminal	0,542	emphasis	0,491	accent	0,619
none	0,501	none	0,480	none	0,596

Tableau 1 : Corrélation moyenne pour chaque modèle et par type d'unité intonative (tous, [+, [|).

La première colonne de ce tableau rassemble tous les résultats du calcul quelle que soit la nature de l'unité intonative. Dans ce cas-là, la progression vers les modèles les plus complexes n'est pas représentée. C'est le modèle IU qui obtient la meilleure moyenne de corrélation. Contrairement à notre hypothèse de départ, on n'obtient pas d'amélioration de la modélisation en prenant en compte la fonction accentuelle, c'est le modèle global ou la commande de phrase (en terme Fujisakien) qui obtient la meilleure corrélation globale.

Toutefois, si l'on se concentre sur la dernière colonne représentant les unités intonatives terminales seulement, on obtient bien une amélioration de la corrélation de modèle en modèle. Plus on introduit d'informations fonctionnelles, meilleure devient notre prédiction des contours de F0.

Ces résultats paraissent mitigés, mais il faut bien préciser qu'aucune tentative d'optimisation des paramètres de nos modèles n'a été entreprise à ce stade de notre travail expérimental. Quelles sont les meilleures séquences de tons INTSINT à utiliser dans les différents modèles, comment peut-on optimiser l'alignement des points cibles dans nos modèles ? Qu'en est-il du calcul de la corrélation linéaire ? Est-ce une évaluation suffisante de la similitude entre deux courbes de F0 ?

La prochaine étape consistera donc à tenter d'optimiser les paramètres de nos modèles de synthèse afin d'obtenir une meilleure représentation des formes mais aussi le cas échéant de remettre en question l'annotation fonctionnelle de départ.

Chapitre 7

Phases d'optimisation des modèles de synthèse

Ayant estimé les paramètres de clé et d'étendue automatiquement avec l'algorithme INTSINT, il nous reste à optimiser le choix des séquences de tons INTSINT, et l'alignement de ces tons. L'évaluation objective de la qualité de nos modèles fera aussi l'objet d'une optimisation au cours de ces expériences.

Nous avons décidé d'introduire un modèle supplémentaire pour ces expériences d'optimisation. Celui-ci est basé sur l'annotation des pauses silencieuses. On peut en effet supposer que ces pauses silencieuses constituent une information linguistique de base qui nous permettra de dégager une première modélisation des contours de F0.

L'optimisation des représentations formelles sera effectuée dans le modèle des pauses, le modèle des unités intonatives et enfin au niveau du modèle le plus complexe appelé modèle emphatique lui-même basé sur l'annotation des unités tonales. Ceux-ci seront décrits dans les sections suivantes. À l'issue de ces étapes d'optimisation des représentations formelles, nous essaierons d'optimiser cette fois-ci les représentations fonctionnelles c'est-à-dire que l'on cherchera à extraire les informations fonctionnelles directement à partir des formes.

7.1 Premières phases d'optimisation : modèle des pauses

Il s'agit là d'un modèle intermédiaire ajouté entre le modèle « nul » et le modèle des unités intonatives. La seule information fonctionnelle retenue à ce stade sera donc la pause silencieuse annotée manuellement avec le symbole « # ». Pour chaque segment situé entre deux pauses silencieuses, on se propose d'optimiser la représentation formelle de la courbe F0. Il y a donc cinq segments par fichier son, correspondant aux cinq phrases énoncées par passage.

7.1.1 Optimisation de la séquence de tons INTSINT

Tout d'abord, les tons INTSINT sont alignés de façon fixe pour cette étape. La représentation de la mélodie entre deux pauses se fait de façon empirique, au moyen de quatre tons INTSINT [t1-t2 t3-t4] afin de modéliser de façon identique le mouvement mélodique initial et le contour mélodique final. t1 est aligné à une distance de 10 ms de la frontière gauche et t4 à 10 ms de la frontière droite, t2 et t3 sont quant à eux alignés respectivement à 300 ms de la frontière gauche pour le premier et à 300 ms de la frontière droite pour le second. Un alignement fixe est proposé à ce stade pour faire varier dans un premier temps la valeur des tons INTSINT.

Nous avons ensuite réalisé un script permettant de tester toutes les combinaisons de tons INTSINT possibles pour t1, t2, t3, t4. Pour le premier ton cependant on ne choisit que des tons absolus tels que M, T ou B car il n'y a pas de ton précédant ce point. Pour les autres points, il y a sept tons possibles (T, B, H, L, U, D, S). Pour les points t2 et t3, on inclut la possibilité de n'avoir aucun ton (_). On a donc un total de 1344 (= 3*8*8*7) séquences de tons possibles comme on peut le voir dans la figure ci-dessous :

	#	Municipal fire service speaking					#	We're trying to locate an emergency caller					#
Aligne-ment des tons en ms	0 à 300ms	t1à +10 ➔ 210	t2 à +300 ➔ 500	… ….	t3 à -300 ➔ 1720	t4 à -10 ➔ 2010	2020 à 2240	t1à +10 ➔ 2250	t2 à +300 ➔ 2540		t3 à -300 ➔ 4100	t4 à -10 ➔ 4390	4400
Tons possibles		M T B	T B H L U D S _		T B H L U D S _	T B H L U D S		M T B	T B H L U D S _		T B H L U D S _	T B H L U D S	
Combinai sons obtenues		M M M M M M M M M Etc … B	T T T T T T T T T Etc … _		T T T T T T T B B Etc … _	T B H L U D S T B Etc … S		M M M M M M M M M Etc … B	T T T T T T T T T … … _		T T T T T T T T T … … _	T B H L U D S T B … … S	

Tableau 2 : Optimisation des séquences de tons INTSINT entre deux pauses.

Pour évaluer la qualité de chaque séquence de ton testée, on génère la courbe MOMEL correspondante (les tons INTSINT sont convertis en points cibles MOMEL) et on calcule la corrélation linéaire entre la courbe issue du modèle et la courbe de MOMEL issue des données. Les résultats obtenus pour chaque combinaison de tons au niveau de tous les énoncés entre deux pauses sont stockés dans un fichier de format « .txt » dont on peut voir deux extraits ci-dessous.

Locuteur	Fichier	Début	Fin	Type_IU	Modèle	Forme	Corrélation
fa	fao20071	12	242	[|	pause	MTTB	0.422
fa	fao20071	12	242	[|	pause	MTTL	0.288
fa	fao20071	12	242	[|	pause	MTTD	0.13
fa	fao20071	12	242	[|	pause	MTBT	0.624
fa	fao20071	12	242	[|	pause	MTBB	0.873
fa	fao20071	12	242	[|	pause	MTBH	0.799
fa	fao20071	12	242	[|	pause	MTBU	0.848
fa	fao20071	12	242	[|	pause	MTLT	0.574
fa	fao20071	12	242	[|	pause	MTLB	0.83
fa	fao20071	12	242	[|	pause	MTLH	0.753
fa	fao20071	12	242	[|	pause	MTLL	0.857
fa	fao20071	12	242	[|	pause	MTLU	0.813

Tableau 3 : Extrait du fichier récapitulatif des séquences de tons testées entre deux pauses.

> La première colonne indique le locuteur traité ici « FA », la seconde le fichier traité, la troisième le temps du début de l'énoncé suivi par le temps à la fin de l'énoncé (« début » et « fin »). Le type de frontière apparaissant avant la pause silencieuse est noté dans la cinquième colonne et les trois dernières colonnes indiquent le modèle, la combinaison de tons testés et la valeur de la corrélation.

À partir de l'ensemble des résultats de corrélation, il s'agit ensuite de sélectionner les meilleures séquences de quatre tons. Nous cherchons à extraire les séquences de quatre tons qui obtiennent une bonne note de corrélation dans le plus de cas possible. On fait un premier tri des résultats en éliminant les séquences qui obtiennent une corrélation inférieure à 0,5. Ce chiffre correspond à un seuil minimal placé de façon empirique. L'observation d'une série de courbes générées à partir de l'optimisation des tons nous a conduits à utiliser cette valeur. Cette valeur minimale correspond aussi aux résultats obtenus par les modèles de synthèse (cf. tableau 1 section 6.3).

Afin de faire ressortir les séquences qui obtiennent de bons résultats, nous avons cumulé pour chaque combinaison de tons les corrélations obtenues sur l'ensemble du corpus d'un locuteur donné. La forme qui obtient le cumul le plus élevé est retenue.

Tous types d'IU	Corrélation tous	[+	Corrélation [+	[|	Corrélation [|
MHLL	50.67	BTBB	14.307	MHLB	29.895
MHLB	50.352	BHBB	14.304	MUDL	29.545
MUDD	50.153	BUBB	14.298	MUDD	29.458
MULL	50.137	MULD	14.003	MHLL	29.337
MULB	49.837	MU_L	13.856	MTLB	29.107

Tableau 4 : Résultat pour FA : meilleures formes par type de frontière

> La première colonne indique les résultats tous types d'IU confondus, la colonne corrélation correspond au cumul des corrélations > 0,5. Les colonnes qui suivent montrent les résultats pour les IU de type non terminal « [+ » et terminal « [| ».

De façon globale, un énoncé entre deux pauses est modélisé au mieux par une séquence montante descendante globale codée {MHLL}.

D'autre part la séquence qui obtient la meilleure corrélation au niveau d'un segment est enregistrée dans le fichier TextGrid dans une rangée nommée « inttone1 », comme le résume le fichier Praat suivant :

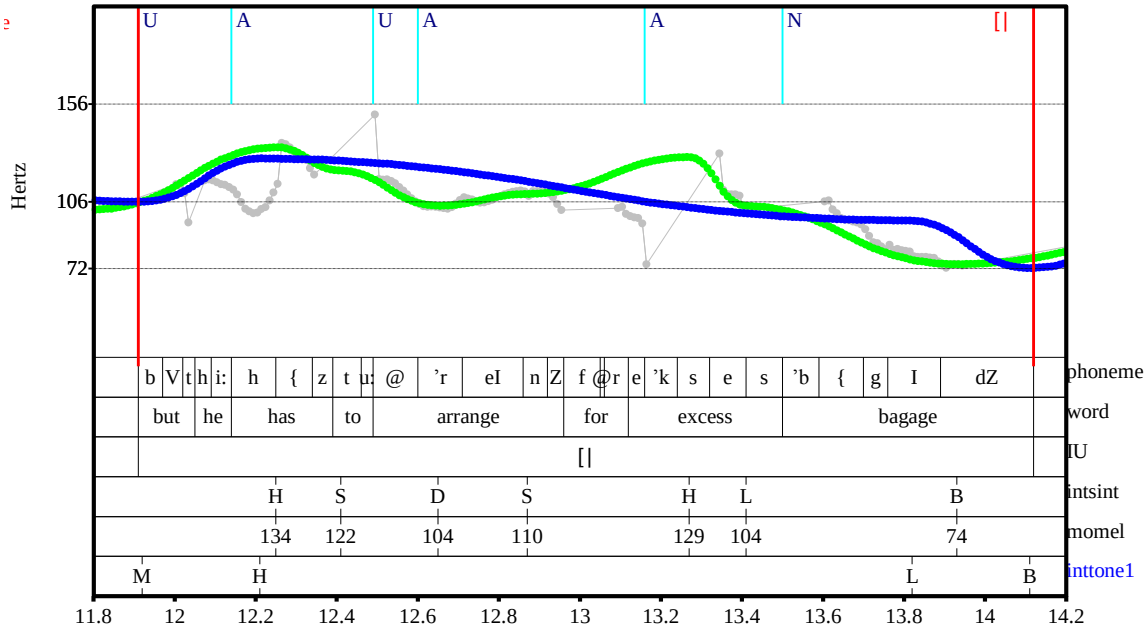


Figure 40 : Courbe issue du processus d'optimisation (en bleu) comparée à MOMEL d'origine (en vert) et TextGrid représentant les niveaux d'annotation et d'analyse pour le fichier FAO40073.

> Rangée 1= phonèmes, rangée 2= mots, rangée 3= annotation de l'unité intonative, rangée 4= tons INTSINT, rangée 5= MOMEL, rangée 6= optimisation de la représentation formelle en 4 tons INTSINT (« inttone1 »).

Au niveau de dernière rangée nommée « inttone1 », on peut observer le résultat de l'optimisation de la séquence de quatre tons INTSINT {MHLB}.

La même procédure d'optimisation est reproduite pour chacun des quatre locuteurs de notre corpus. Après avoir sélectionné la meilleure séquence pour chaque locuteur : FA {MHLL}, FE {MULL}, FF {MULL}, FG {MSLL}, nous faisons varier l'alignement des points cibles.

7.1.2 Optimisation de l'alignement des points cibles

Les tons t1 et t4 sont maintenus à une distance fixe des frontières gauches et droites à ce stade, seuls t2 et t3 font l'objet de cette optimisation de la manière suivante : prenons l'exemple du locuteur FE :

M fixé à une distance de 10 ms de la frontière gauche ;

U varie de 100 à 800 ms de la frontière gauche par pas de 50 ms (15 itérations) ;

L varie de 100 à 800 ms de la frontière droite par pas de 50 ms (15 itérations) ;

L fixé à une distance de 10 ms de la frontière droite.

De manière identique à l'étape d'optimisation des tons, on conserve la paire t2-t3 qui donne la meilleure corrélation avec la courbe de MOMEL issue des données (cf. tableau de résumé des résultats ci-dessous).

Locuteur	Forme	t1 Alignement à gauche	t2 Alignement à gauche	t3 Alignement à droite	t4 Alignement à droite	Corrélation moyenne
FA	M H L L	10	200	350	10	0.749
FE	M U L L	10	300	400	10	0.775
FF	M U L L	10	250	250	10	0.756
FG	M S L L	10	200	300	10	0.722

Tableau 5 : Meilleures séquences de tons et alignements pour chaque locuteur.

Pour le locuteur FA, on observe que le meilleur alignement pour t2 est à 200 ms du début de l'IU, t3 lui, s'aligne de façon optimale à 350 ms de la fin de l'IU.

Avant de nous attacher à montrer comment on peut optimiser les représentations fonctionnelles à partir de ces formes optimales, nous décrirons l'optimisation des représentations formelles à partir des deux modèles suivants qui correspondent à l'utilisation de deux autres unités prosodiques : l'unité intonative et l'unité tonale.

7.2 Optimisation des représentations formelles au niveau de l'unité intonative

Dans cette étape, on décide de se placer au niveau de l'unité intonative (IU) déjà annotée manuellement pour optimiser les séquences de tons et leurs alignements. De la même façon que dans l'expérience précédente, les séquences de tons et leurs alignements sont optimisés pour chaque locuteur comme le montre l'extrait du fichier récapitulatif des meilleures séquences trouvées :

Tous types d'IU	Corrélation tous	[+	Corrélation [+	[|	Corrélation [|
MT_B	73.173	MT_B	36.051	MULL	27.546
MH_B	73.044	MU_L	35.487	MULB	27.282
MU_L	72.912	MH_B	35.411	MSLL	27.113
MU_B	71.448	MH_L	35.381	MHLB	27.092
MHLD	71.394	MHLD	34.665	MH_B	27.041

Tableau 6 : Meilleures séquences de tons INTSINT pour le locuteur FA avec modélisation au niveau de l'IU.

> Il s'agit là d'un extrait du fichier R montrant les meilleures séquences de tons trouvées pour le locuteur Fa. Les différentes colonnes représentent les différents types d'unités intonatives : la première regroupe tous les cas (terminal/non terminal) ensuite les unités intonatives non terminales et terminales séparément.

L'étape suivante consiste à optimiser l'alignement des tons. Nous garderons cette fois-ci plusieurs séquences de tons et à la différence de l'expérience précédente, nous ne garderons pas un alignement fixe mais une tranche optimale pour chaque cible.

Tous types d'IU	Correlation tous	[+	Correlation [+	[|	Correlation [|
MT_B, 300,-100	87.959	MTBU , 250,-150	41.575	MULL , 300,-250	36.956
MT_B ,300,-150	87.959	MTBU ,200,-100	41.353	MULL , 300,-300	36.904
MT_B, 300,-200	87.959	MTBU 350,-100	41.281	MULL , 300,-200	36.79
MT_B , 300,-250	87.959	MTBU 250,-100	41.179	MULL , 250,-250	36.779
MT_B , ,300,-300	87.959	MTBU 300,-100	41.16	MULL , 250,-300	36.711
MT_B , 200,-100	87.854	MTBU 300,-150	41.125	MH_B ,300,-100	36.691
MT_B , 200,-150	87.854	MTBU , 200,-150	41.084	MH_B , 300,-150	36.691
MT_B, 200,-200	87.854	MTBU ,350,-150	40.785	MH_B, 300,-200	36.691

Tableau 7 : Meilleures séquences de tons et alignements pour le locuteur FA avec modélisation au niveau de l'IU.

On peut observer dans ce tableau les meilleures formes retenues pour chaque type d'IU ainsi que les meilleurs alignements qui leur sont associés. Pour un IU terminal par exemple, on voit dans la première cellule de la dernière colonne que la meilleure forme est représentée par la séquence {MULL} pour laquelle t2 est aligné à 300 ms du début de l'IU et t3 à 250 ms de la fin de l'IU. Ainsi pour le locuteur FA par exemple on garde les paramètres optimisés suivants : les meilleures séquences retenues sont {MH_B} pour les IU terminaux et {MTBU} pour les IU non terminaux et on peut réduire les variations de l'alignement en fonction des résultats obtenus.

L'alignement optimal choisi est alors le suivant :

- t1 est aligné à 10 ms du début de l'IU ;
- t2 varie de 200 à 350 millisecondes du début de l'IU en 4 itérations de 50 ms ;
- t3 varie de 300 à 100 ms de la fin de l'IU en 5 itérations de 50 ms ;
- t4 est aligné à 10 ms de la fin de l'IU.

Enfin, l'optimisation des représentations formelles est effectuée au niveau des unités tonales.

7.3 Optimisation des représentations formelles au niveau des unités tonales

On se place ici au niveau du dernier modèle de synthèse présenté dans la première partie : le modèle emphatique. Tous les niveaux d'annotation fonctionnelle représentés dans le système IF seront donc pris en compte : le niveau de l'unité intonative (IU), le caractère terminal/non terminal, le niveau des unités tonales (TU) et leurs différents niveaux de proéminence (accentué : A / non accentué : U / nucléus : N / emphatique : !N). On procède de la même façon que dans les premiers modèles à partir de l'annotation des unités tonales cette fois-ci. La prise en compte de ce domaine plus étroit nous conduit à développer l'annotation fonctionnelle et remettre en question l'évaluation objective par le calcul de la corrélation linéaire uniquement.

7.3.1 Optimisation de l'évaluation objective des modèles

Comme on l'a vu dans le premier chapitre de cette thèse, la corrélation linéaire est la principale mesure utilisée pour quantifier la similitude entre deux courbes de F0. Hermes (1998) affirme qu'il s'agit de l'indice le plus fiable de l'équivalence perceptive de deux courbes de F0. De nombreux chercheurs utilisent la corrélation linéaire de concert avec le calcul de l'écart quadratique moyen entre les courbes afin d'obtenir une mesure plus complète de la qualité d'un modèle de synthèse. La corrélation permet de juger de l'équivalence perceptuelle d'un modèle par rapport à un enregistrement original. L'écart quadratique moyen permet de quantifier plus précisément les écarts entre la courbe originale et la courbe modélisée (évaluation auditive *vs* évaluation visuelle). Dans les deux premiers modèles optimisés (modèle des pauses et modèles des IU), nous n'avons utilisé que le calcul de la corrélation linéaire avec un résultat qui paraissait tout à fait satisfaisant. Toutefois en s'attaquant au modèle des unités tonales, qui implique la prise en compte d'un domaine plus étroit que l'unité intonative, nous avons été conduite à remettre en question notre façon d'évaluer objectivement la qualité des modèles de synthèse. Il faut noter que nous utilisons la corrélation linéaire à la fois pour sélectionner les meilleures séquences au niveau d'une unité prosodique et pour évaluer la qualité de nos modèles dans l'ensemble.

Nous avons donc mis en place une série de tests permettant de comparer différentes façons d'évaluer l'optimisation des représentations formelles (séquences de tons et alignement). Tout d'abord, nous avons essayé d'optimiser les représentations formelles au niveau des TU en utilisant la corrélation linéaire seule, puis l'écart quadratique moyen seul et enfin une combinaison des deux calculs.

Avant d'étudier quelques exemples de résultats de ces tests, nous présenterons les formules de calcul de la corrélation linéaire et de l'écart quadratique moyen.

7.3.1.1 Formule du calcul du coefficient de corrélation linéaire et de l'écart quadratique moyen

Afin de calculer le coefficient de corrélation entre deux séries de même longueur, on peut procéder de la manière suivante : on suppose que l'on a les tableaux de valeurs suivants : $X(x_1, \ldots, x_n)$ et $Y(y_1, \ldots, y_n)$ pour chacune des deux séries, alors, pour connaître le coefficient de corrélation liant ces deux séries, on applique la formule suivante :

$$r_p = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Si r vaut 0, les deux courbes ne sont pas corrélées. Les deux courbes sont d'autant mieux corrélées que r est loin de 0 (proche de -1 ou 1). Cependant une corrélation négative indique que les courbes sont inversées donc seules les valeurs de corrélations positives sont intéressantes pour nos expériences.

Le calcul de l'écart quadratique moyen correspond à la racine carrée de la somme du carré des écarts entre les points de la courbe MOMEL $X(x_1,\ldots,x_n)$ et les points de la courbe modélisée $Y(y_1,\ldots,y_n)$. Cette valeur est ensuite divisée par le nombre de points N :

$$\sqrt{\left(\sum (x_i - y_i)^2\right) / N}$$

7.3.1.2 Optimisation des représentations formelles en utilisant différents types d'évaluation objective

Il s'agit de se servir de différents critères mathématiques pour choisir les meilleures séquences de tons INTSINT et leur alignement, au niveau de chaque TU à l'intérieur d'un IU pour tous les fichiers d'un locuteur de MARSEC (G01). Pour chaque modèle, les séquences de tons INTSINT seront converties en points cibles MOMEL et resynthétisées afin de pouvoir comparer cette modélisation avec le résultat de MOMEL d'origine. Le modèle permettant d'obtenir la courbe la plus proche de MOMEL d'origine sera donc le modèle utilisant le meilleur critère mathématique : la corrélation seule, l'écart quadratique moyen ou les deux combinés. Nous utiliserons donc l'analyse par synthèse pour tester la validité de nos modes de calcul.

Observons quelques exemples de courbes modélisées à partir des différents types de calcul.

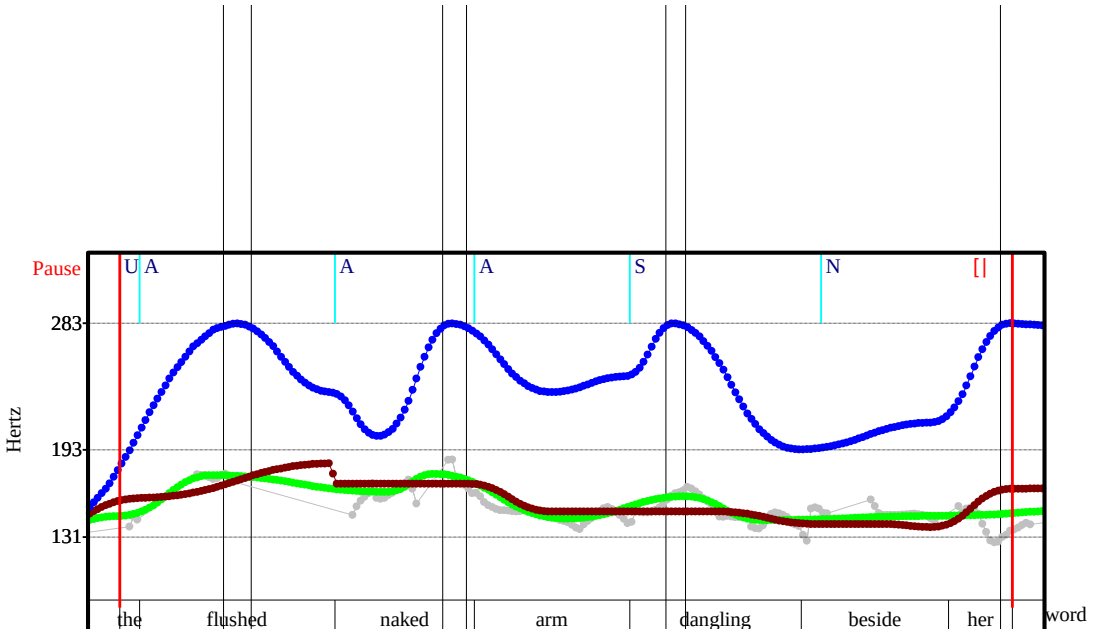


Figure 41 : *Courbes optimisées à partir de la corrélation seule ou de l'écart quadratique moyen pour le fichier G0110G, comparées à MOMEL.*

Dans la figure ci-dessus, nous montrons les résultats de l'optimisation des séquences de tons pour un passage à partir de la corrélation seule (en bleu) et avec l'écart quadratique moyen (en marron). Les courbes ainsi générées sont comparées avec MOMEL (en vert). La première rangée montre les mots, la seconde les tons INTSINT correspondant au passage, la troisième les meilleurs tons INTSINT sélectionnés avec l'écart quadratique moyen (« inttone 81 ») et enfin la dernière les tons sélectionnés avec la corrélation linéaire (« inttone80). Les frontières des TU sont délimitées par les traits bleu clair au-dessus de la figure.

On peut voir dans l'exemple ci-dessus que la meilleure optimisation des séquences de tons est effectuée à partir de l'écart quadratique moyen. La courbe marron suit de près la courbe verte.

La corrélation seule dans ce cas donne une courbe respectant la forme de F0, mais qui est très écartée de la courbe d'origine.

Au niveau de la courbe MOMEL d'origine, on observe une suite de légers contours descendants-montants qui sont reproduits de façon très amplifiée lorsqu'on utilise le calcul de la corrélation qui génère une courbe éloignée de la courbe de MOMEL. Au niveau de la courbe générée en utilisant l'écart quadratique moyen comme métrique, on constate l'effet inverse. Les mouvements mélodiques sont atténués. C'est une tendance que nous avons pu remarquer dans de nombreux exemples.

On trouve aussi d'autres exemples typiques pour lesquels la courbe optimisée par la corrélation seule est encore plus problématique.

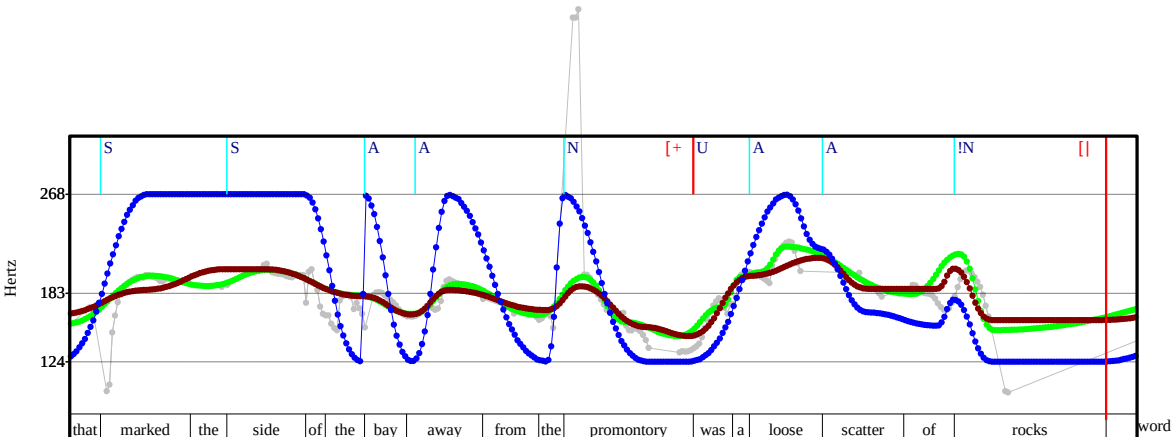


Figure 42 : *Courbes optimisées à partir de la corrélation seule (en bleu) ou de l'écart quadratique moyen (en marron) comparées à MOMEL (en vert) pour le fichier G0105G.*

> Rangée 1= mots, rangée 2= tons INTSINT d'origne, rangée 3= meilleures séquences de tons INTSINT sélectionnées avec l'écart quadratique moyen, rangée 4= meilleures séquences de tons INTSINT sélectionnées avec la corrélation seule.

Au niveau de plusieurs TU, on voit se dessiner un contour en dents-de-scie avec la corrélation comme seul critère. Le calcul de l'écart quadratique moyen permet de choisir une courbe optimale bien plus proche de MOMEL d'origine.

C'est donc une métrique qui semble permettre de générer une courbe très proche de la courbe d'origine, mais les courbes optimisées peuvent montrer des contours généralement simplifiés par rapport à la courbe d'origine, ce qui diminue la qualité de la resynthèse et l'authenticité au niveau de la perception.

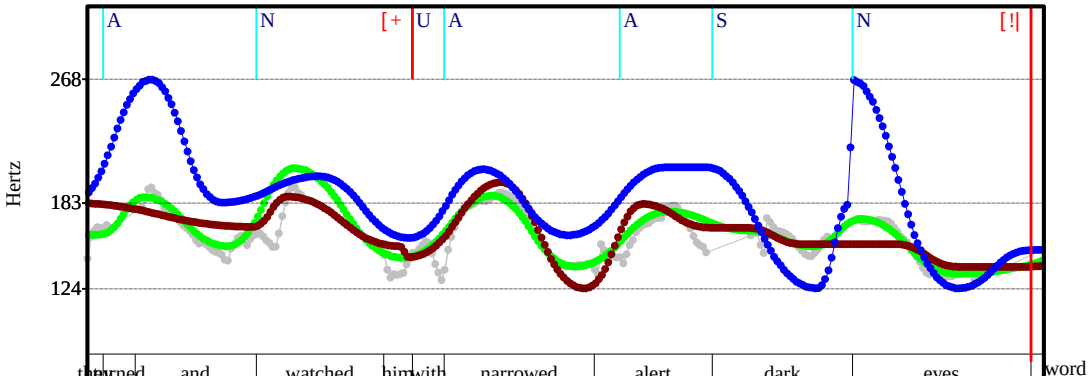


Figure 43 : *Courbes optimisées à partir de la corrélation seule (en bleu), de l'écart quadratique moyen (en marron) pour le fichier G0105B comparées à MOMEL (en vert).*

Au niveau de la courbe marron qui représente le résultat de l'optimisation des tons à partir de l'écart quadratique moyen, on voit que les contours de F0 sont à nouveau escamotés en deux endroits : au niveau de la première unité tonale « A » (« they turned and ») et de la dernière qui est un nucléus non marqué. La séquence INTSINT correspondant au premier contour est {BHL} (voir première rangée), alors que l'écart quadratique moyen optimise le contour avec une séquence {–D} (voir seconde rangée). Le calcul de la corrélation pour cet extrait permet une meilleure modélisation des contours de F0.

L'observation des courbes générées à partir de chaque modèle et pour l'ensemble des fichiers du locuteur G01 permet de faire les remarques suivantes : tout d'abord on trouve toute une série d'exemples pour lesquels une bonne corrélation linéaire ne correspond pas forcément à une courbe satisfaisante. L'écart quadratique moyen seul permet de suivre précisément la courbe dans la plupart des cas, mais on compte aussi des exemples pour lesquels la courbe générée est simplifiée et donne donc un résultat nettement inférieur à la corrélation linéaire.

Nous avons ensuite observé les variations de la valeur de l'écart quadratique moyen en fonction de la corrélation sur plusieurs extraits du corpus de G01. Pour chaque unité tonale, nous avons testé toutes les séquences de deux tons INTSINT et calculé les valeurs de la corrélation et de l'écart quadratique moyen comme le montre l'extrait du tableau de résultat suivant :

Locuteur	Fichier	Début	Fin	Forme	Corrélation	RMSE
G01	G0101B	330	800	TD	0.907	2.411
G01	G0101B	330	800	TD	0.867	2.721
G01	G0101B	330	800	HD	0.857	4.512
G01	G0101B	330	800	TT	0.848	1.725
G01	G0101B	330	800	HS	0.848	3.717
G01	G0101B	330	800	US	0.848	5.435
G01	G0101B	330	800	HD	0.751	4.775
G01	G0101B	330	800	HU	0.738	3.391
G01	G0101B	330	800	TT	0.725	1.33
G01	G0101B	330	800	HS	0.725	3.363

Tableau 8 : Extrait du tableau montrant les séquences de ton INTSINT testées au niveau d'un TU et les valeurs de corrélation et de l'écart quadratique moyen obtenues.

La première colonne indique le nom du locuteur, la seconde le fichier traité, la troisième et la quatrième le début et la fin du TU traité en ms. La cinquième colonne montre la séquence de tons testée et les deux colonnes suivantes la valeur de la corrélation puis de l'écart quadratique moyen. Quand une séquence de tons apparaît plusieurs fois, cela signifie que l'alignement testé est différent.

Les résultats sont ensuite représentés sous forme de graphique. Dans l'extrait ci-dessous, seules les séquences obtenant une corrélation comprise entre 0,8 et 1 pour un TU ont été retenues. Pour chacune de ces combinaisons de tons testés, on affiche alors la valeur de la corrélation (en bleu) et les valeurs de l'écart quadratique moyen (en rouge). Les traits en abscisse représentent les combinaisons de tons testées :

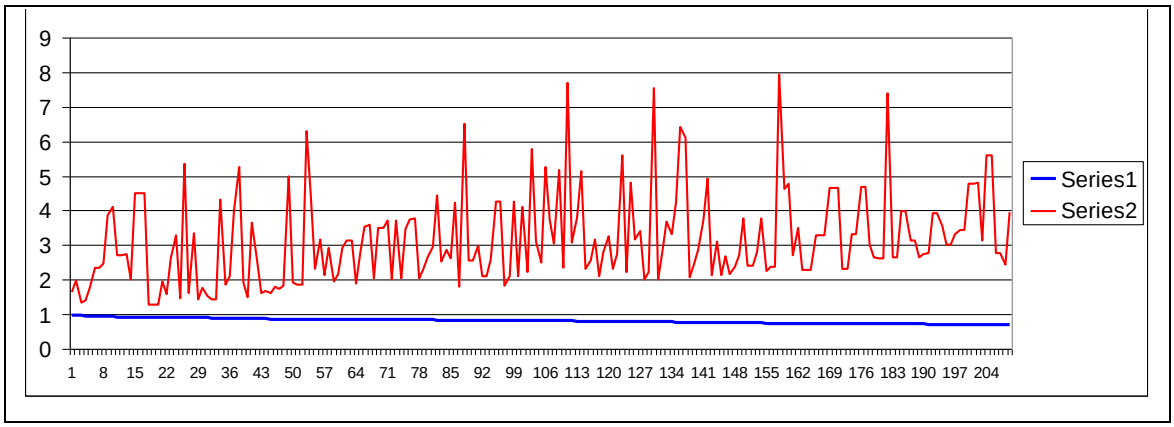


Figure 44 : Comparaison des valeurs de corrélation et de l'écart quadratique moyen pour l'ensemble des séquences de tons testées au niveau d'un TU sur G0101B.

> Série 1 : variation des valeurs de corrélation linéaire, série 2 : variation des valeurs de l'écart quadratique moyen, en ordonnées : valeurs de l'écart quadratique moyen en demi-tons (de 0 à 9), en abscisses : les traits représentent chaque combinaison de tons INTSINT testée.

On peut observer de façon globale que pour une bonne valeur de corrélation qui ne varie que de 0,8 à 1, on peut obtenir de multiples valeurs de l'écart quadratique moyen. On peut voir sur l'axe des ordonnées que la valeur de celui-ci varie de 1,5 à environ 7,5 demi-tons.

Observons les résultats obtenus pour un même passage en gardant toutes les valeurs de corrélation (de -1 à 1) :

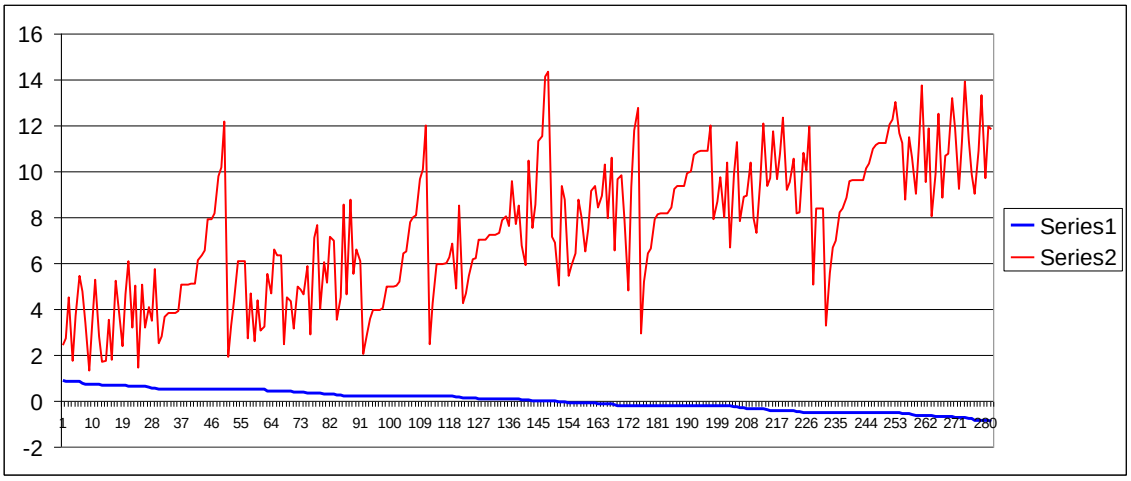


Figure 45 : Comparaison des valeurs de corrélation et de l'écart quadratique moyen pour l'ensemble des séquences de tons testées au niveau d'un TU sur G0101B toutes corrélations.

On peut voir dans ce graphique que, lorsque les valeurs de la corrélation diminuent, les valeurs de l'écart quadratique moyen augmentent. Cependant la variabilité des valeurs de l'écart quadratique moyen reste importante en tout point.

L'observation des variations de l'écart quadratique moyen en fonction de la corrélation sur plusieurs échantillons semble suggérer une relation complexe entre les deux métriques.

Afin d'améliorer le choix automatique des meilleures séquences et en même temps la qualité de nos modèles, nous avons décidé de tester une métrique combinant la corrélation et l'écart quadratique moyen. Nous avons fait plusieurs essais sur nos corpus. Il était essentiel pour garder un choix automatique des meilleures séquences, de pouvoir utiliser une métrique unique représentée par un chiffre unique que nous appellerons « score » dont nous détaillons le calcul ci-après. Nous avons gardé la corrélation comme indice principal de la direction de la courbe et utilisé l'écart quadratique moyen pour pondérer le résultat et pour affiner notre modélisation. Au niveau de la corrélation, on retient une seule décimale sur laquelle on applique une règle d'arrondie : quand l'écart quadratique moyen est faible (inférieur ou égal à un demi-ton) on arrondit par excès, quand l'écart quadratique moyen est plus élevé (supérieur à deux demi-tons) on arrondit par défaut sinon on applique la règle d'arrondi standard. Ce chiffre est alors multiplié par dix et devient la partie entière, il sera ensuite complété par l'écart quadratique moyen pour sa partie décimale. Pour avoir une bonne modélisation, il faut que la corrélation soit élevée et l'écart quadratique moyen le plus faible possible, aussi pour combiner ces deux indices, on fait le complément à 9999 (on a retenu quatre décimales).

Prenons un exemple :

- Si pour une séquence donnée, on obtient une corrélation égale à 0,9062 et un écart quadratique moyen égal à 0,875 alors le score sera 10,9124 (1*10 concaténé à 9999-875) ;
- Si la corrélation est égale à 0,9862 et l'écart quadratique moyen est égal à 3,713 alors le score sera 9,6286 (0,9*10 concaténé à 9999-3713).

C'est une façon empirique de combiner les deux métriques principales dont les résultats nous ont paru satisfaisants sur nos données comme on peut le voir dans l'exemple ci-dessous où la courbe rose représente le résultat avec le score.

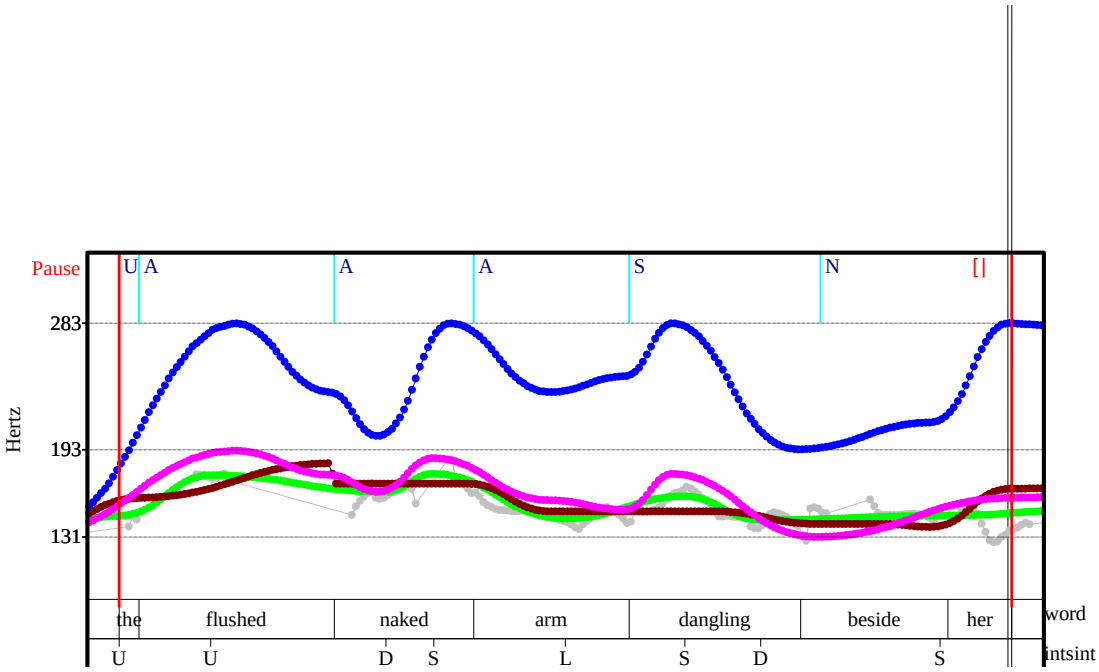


Figure 46 : *Courbes optimisée à partir de la corrélation seule (en bleu), de l'écart quadratique moyen (en marron) et du « score » (en rose) pour le fichier G0105B comparées à MOMEL (en vert).*

Rangée 1= mots, rangée 2= tons INTSINT d'origine, rangée 3= meilleures séquences de tons INTSINT sélectionnées avec le score, rangée 4= meilleures séquences de tons INTSINT sélectionnées avec l'écart quadratique moyen, rangée 5= meilleures séquences de tons INTSINT sélectionnées avec la corrélation seule.

Dans la suite du travail expérimental, nous utiliserons donc essentiellement le calcul du score pour choisir les meilleures séquences et alignement automatiquement et continuerons à utiliser chaque métrique séparément pour comparer le résultat des différents modèles.

7.3.2 Optimisation de la séquence de tons en fonction de l'unité tonale

7.3.2.1 Développement de l'annotation des fonctions

Les trois types de TU principaux sont classés en fonction de leur position dans l'IU en utilisant les signes {0, 1, M, F, 1F} :

- 0 représente la première unité tonale de l'unité intonative lorsqu'elle est inaccentuée ;
- 1 est le premier TU accentué de l'IU ;
- M représente les TU en position médiane ;
- F est utilisé pour distinguer le TU final ;
- 1F représente une unité tonale en position initiale et non suivie par d'autres unités tonales.

On cherche à caractériser les différentes configurations tonales en fonction des niveaux de proéminence et de la position du TU dans l'IU. Si ces éléments s'avéraient être vraiment

significatifs dans notre analyse, l'optimisation de leur représentation formelle devrait permettre de les distinguer.

Deux symboles diacritiques « + » et « | » sont aussi ajoutés avant ou après les premiers TU initiaux (0 ou 1) et les TU finaux (F). Si le premier TU est précédé d'un IU non terminal, on note « +0 » ou « +1 », si le dernier TU se trouve dans un IU non terminal, il est noté « F+ ».

On obtient par exemple les séquences suivantes au niveau d'un IU :

|U0 A1 AM NM AF+ +A1 AM NF| | U0 !N1 AM AF|

Le tableau ci-dessous résume l'ensemble des fonctions prises en compte pour les TU en position initiale.

	Position 0	Position 1 en début d'IU	Position 1 précédé d'un U0
IU précédent de type terminal	|U0	|A1 |N1 |!N1	A1 N1 !N1
IU précédent de type non terminal	+U0	+A1 +N1 +!N1	A1 N1 !N1

Tableau 9 : Différents types de TU annotés en position initiale

Le tableau ci-dessous résume l'ensemble des fonctions prises en compte pour les TU en position médiane ou finale.

	Position M	Position F	Position 1F (1 seul TU / IU)
IU de type terminal	AM NM !NM	AF| NF| !NF|	N1F|
IU de type non terminal	AM NM !NM	AF+ NF+ !NF+	N1F+

Tableau 10 : Différents types de TU annotés en position médiane et finale

Nous décrirons ensuite les paramètres d'optimisation des représentations formelles dans ce nouveau modèle.

7.3.2.2 Paramètres d'optimisation dans les modèles des TU

Les TUS en position initiale et médiane sont modélisés à l'aide de trois cibles tonales[6] afin de pouvoir représenter à la fois des contours simples comme les chutes et montées mais aussi des contours plus complexes comme les chutes montées :

- T1 est aligné à une distance fixe de 10 ms du début du TU, tons INTSINT possibles : {M,T,B} en début d'IU sinon { _ } ;
- T2 varie de 100 à 200 ms du début du TU en deux itérations, tons possibles : {T, B, H, L, U, D, S, _} ;
- T3 est aligné à 50 ms de la fin du TU, tons possibles : {T, B, H, L, U, D, S}.

Les TU finaux sont codés quant à eux au moyen de quatre cibles tonales ce qui peut nous permettre le cas échéant de représenter les contours de chute + montée :

- T1 est aligné à une distance fixe de 10 ms du début du TU, tons INTSINT possibles : {_};
- T2 varie de 100 à 200 ms du début du TU en deux itérations, tons possibles : {T, B, H, L, U, D, S, _} ;
- T3 est aligné à 100 ms de la fin du TU, tons possibles : {T, B, H, L, U, D, S, _} ;
- T4 est aligné à 10 ms de la fin du TU, tons possibles : {T, B, H, L, U, D, S}.

Au niveau de chaque TU composant un IU, toutes les combinaisons de trois ou quatre tons autorisées sont testées. Les séquences de tons INTSINT ainsi obtenues sont converties en points cibles MOMEL et extrapolées de façon linéaire (un point toutes les dix millisecondes) afin de pouvoir comparer la courbe ainsi obtenue avec la courbe de MOMEL présente dans les données.

7.3.2.3 Résultats de la recherche des meilleures séquences de tons

Il faut noter que l'optimisation des séquences de tons pour ce modèle n'est pas spécifique à chaque locuteur. Les paramètres de clé et d'étendue sont conservés, mais l'optimisation au niveau des TU est globale pour tous les locuteurs. En effet nous ne cherchons pas à extraire un modèle particulier pour chaque locuteur mais plutôt un modèle généralisant pour chaque type de fonction annotée au niveau de l'unité tonale. En d'autres termes, on cherche les meilleures formes associées aux unités tonales accentuées par opposition aux unités tonales nucléaires, etc.

Au niveau de chaque type de TU, toutes les combinaisons de séquences de tons INTSINT sont testées selon les paramètres décrits plus haut. Un script R permet ensuite de résumer les

[6] Dans un TU en position médiane, t1 a la valeur « no tone » notée « _ ».

résultats pour chaque étiquette fonctionnelle comme on peut le voir dans le tableau ci-dessous qui est un extrait du fichier de résultat complet :

NF|	score NF|	!NF|	score !NF|	NF+	score NF+	!NF+	score !NF+	N1F|
__LS	3998.65	_SBU	886.565	__LS	4346.163	_UBU	209.4098	_UBB
__DS	3891.4226	_DLS	878.023	_DDS	4279.4686	_HBU	209.365	_SLS
__BB	3843.15	_DBB	876.119	__DS	4205.8584	_ULU	208.4954	_SDS
___B	3838.5158	__LS	873.875	_D_D	4183.8954	_U_B	206.4306	_HBB
_DLS	3815.27	_DBU	865.718	_LLU	4173.3766	_ULS	203.4804	_SLD
__LD	3754.5752	_DLD	862.8938	__BU	4119.4822	_ULD	201.549	_S_L

Tableau 11 : Meilleures séquences de tons INTSINT au niveau des TU.

Le tableau ci-dessus montre les six meilleures séquences retenues pour les fonctions NF| (nucléus final dans une unité intonative terminale), !NF| (la même chose pour un nucléus emphatique), etc. Pour chaque forme retenue, le cumul des scores obtenus est aussi indiqué dans la colonne qui suit.

Les cinq meilleures séquences de tons par fonction sont ensuite sélectionnées automatiquement. C'est à partir de cet ensemble de meilleures séquences que l'on procède à l'optimisation de l'alignement tonal.

7.3.3 Optimisation de l'alignement des points cibles

Tout comme dans le premier modèle d'optimisation, les alignements du second et du troisième point sont optimisés laissant le premier point cible et le dernier s'il y en a un, alignés à une distance fixe de la frontière gauche ou droite du TU. Deux modes d'optimisation de l'alignement des points cibles sont testés : le premier étant basé sur l'expression des distances en valeur absolue et le second exprimé en pourcentage de la durée du segment pris en compte.

Dans le premier modèle (alignement en valeur absolue), les paramètres d'optimisation de l'alignement des tons sont les suivants :

Pour un TU en position initiale ou médiane (1 ou M) :

- T1 est aligné à 10 ms du début de l'unité tonale ;
- T2 varie de 50 à 300 ms du début du TU en 6 itérations ;
- T3 varie de 300 à 50 ms de la fin du TU.

Pour un TU en position finale :

- T1 est aligné à 10 ms du début de l'unité tonale ;
- T2 varie de 50 à 300 ms du début du TU en 6 itérations ;
- T3 varie de 300 à 50 ms de la fin du TU en 6 itérations ;
- T4 est aligné à 10 ms de la fin du TU.

Les paramètres du second type d'expérience d'optimisation (alignement en fonction de la durée du TU) sont les suivants :

Pour un TU en position initiale ou médiane (1 ou M) :

- T1 est aligné avec le début du TU ;
- T2 varie de 5 à 40% du début du TU en 8 itérations ;
- T3 varie de 85 à 95% en 3 itérations.

Pour un TU en position finale :

- T1 est aligné avec le début du TU ;
- T2 varie de 5 à 40% du début du TU en 8 itérations ;
- T3 varie de 60 à 75% du début du TU en 4 itérations ;
- T4 est aligné à 10 ms de la fin du TU.

À l'issue du traitement de toutes les combinaisons possibles de tons et paires de points, on gardera une tranche d'alignement optimale plutôt que deux points uniques et fixes afin de permettre plus de flexibilité à notre modèle. Ci-dessous un extrait du fichier récapitulatif des meilleures positions d'alignement en pourcentage.

|A1	score |A1	+A1	score +A1	A1	score A1	AM
MUT ,0,20,85	688.3166	_UU ,0,30,85	342.9031	_HS ,0,35,85	2152.5451	__D ,0,20,85
MUT ,0,10,85	687.3913	_UU ,0,25,85	341.8114	_HS ,0,40,85	2151.1078	__D ,0,25,85
MUT ,0,15,85	678.4944	__H ,0,15,85	341.2607	_HU ,0,40,85	2148.0349	__D ,0,30,85
M_T ,0,10,85	675.751	__H ,0,20,85	341.2607	_HU ,0,30,85	2139.9795	__D ,0,35,85
M_T ,0,15,85	675.751	__H ,0,25,85	341.2607	_HU ,0,25,85	2133.3456	__D ,0,40,85
M_T ,0,20,85	675.751	__H ,0,30,85	341.2607	_HU ,0,35,85	2128.8469	_DD ,0,35,85
M_T ,0,25,85	675.751	__H ,0,35,85	341.2607	_HU ,0,20,85	2103.019	__D ,0,15,85
M_T ,0,30,85	675.751	__H ,0,40,85	341.2607	_HS ,0,30,85	2102.8503	__D ,0,10,85

Tableau 12 : Les huit meilleurs alignements pour les séquences de tons optimales dans les TU A1 et AM.

> Au niveau de chaque colonne, on observe pour un type de TU donné, les meilleures séquences de tons retenues ainsi que les huit meilleurs alignements pour T2 et T3. Pour le TU |A1 la meilleure forme est la séquence MUT avec T2 aligné à 20% du TU et T3 à 85%.

L'alignement du second point cible (T2) semble varier de 20 à 40% de l'unité tonale alors que le troisième point (T3) s'aligne presque toujours à 85% de l'unité tonale.

Après avoir optimisé les représentations formelles des trois types de modèles ci-dessus, nous tenterons d'optimiser les représentations fonctionnelles.

7.4 Optimisation de la représentation des fonctions

Le principal objectif de ce travail de recherche est de mettre en relation la représentation des formes et fonctions prosodiques. Ayant optimisé les représentations formelles pour les modèles des pauses, des unités intonatives et des unités tonales, la prochaine étape inclura une optimisation de la représentation des fonctions dans chacun de ces modèles. Nous chercherons donc à extraire l'information fonctionnelle à partir de l'information formelle contenue dans les données.

À partir des deux premiers modèles, une tentative de détection automatique des frontières intermédiaires sera menée. À partir de l'optimisation du dernier modèle, nous chercherons à détecter les types de TU c'est-à-dire les niveaux de proéminence représentés par les étiquettes « A, N, !N » ainsi que les frontières des unités intonatives.

7.4.1 Détections des frontières d'IU dans les modèles des pauses et des IU

Les informations fonctionnelles connues pour cette étape sont les fins de mots, les pauses silencieuses et la place de l'accent lexical. Chaque segment entre deux pauses peut être découpé en quatre sous-segments, ce qui correspond à la possibilité d'insérer un maximum de trois frontières intermédiaires[7]. Un script a été élaboré afin de tester toutes les positions possibles pour une, deux ou trois frontières dans chaque segment et de sélectionner la meilleure solution c'est-à-dire celle qui permet d'obtenir la courbe la mieux corrélée avec la courbe MOMEL d'origine.

Afin de tester les meilleurs positionnements pour une frontière, la procédure est la suivante :

- Tout d'abord, on essaie de placer une frontière après le premier mot, le second mot etc. ;
- Au niveau des nouvelles unités intonatives, les formes optimales retenues sont testées en utilisant la corrélation et l'écart quadratique moyen ;
- Le placement de frontières qui permet d'obtenir le meilleur score global au niveau du segment est conservé ;
- Cette procédure est répétée en essayant de placer deux frontières puis trois jusqu'à ce que l'on trouve la meilleure solution.

Le positionnement des frontières est détaillé de façon plus approfondie dans la section suivante.

Prenons un exemple de détection de frontières dans le modèle le plus simple (modèle des pauses ou modèle 1). Tout d'abord on considère un segment entre deux pauses pour lequel la séquence optimale est appliquée comme on peut le voir dans la figure ci-dessous :

[7] Une étude statistique des données montre qu'il y a rarement plus de quatre frontières intermédiaires entre deux pauses. Dans la plupart des cas, on a deux segments.

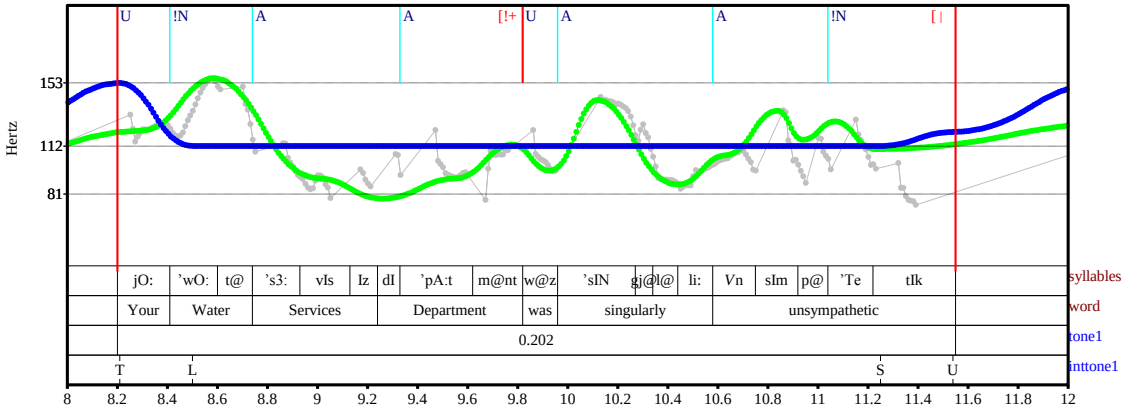


Figure 47 : *Courbe modélisée au niveau de la pause (en bleu) comparée à la courbe de MOMEL originale (en vert) pour un passage du fichier FAo30072.*

> En haut de la figure, on peut voir l'annotation manuelle des frontières d'IU en rouge et des TU en bleu clair. Rangée 1= les syllabes, rangée 2= les mots, rangée 3 = valeur de la corrélation entre les deux courbes, la rangée 4= la forme optimisée dans le modèle 1.

On voit bien que la modélisation en quatre points cibles et ne prenant en compte aucune information fonctionnelle mise à part les pauses silencieuses, est tout à fait médiocre. La corrélation linéaire pour ce passage est d'ailleurs de 0,202.

Après détection des frontières intermédiaires selon la procédure décrite ci-dessus, on obtient la modélisation suivante pour le même passage :

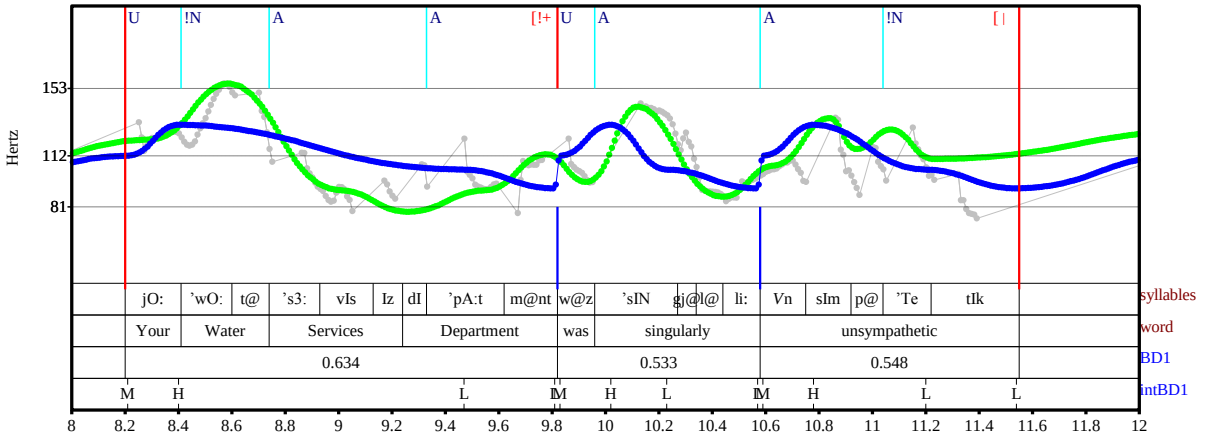


Figure 48 : *Courbe MOMEL modélisée après la détection des frontières (en bleu) comparée à la courbe de MOMEL d'origine.*

> Frontières détectées en bas en bleu à comparer avec les frontières annotées en rouge. Rangée 1= les syllabes, rangée 2= les mots, rangée 3= corrélation entre la courbe de MOMEL d'origine et la courbe issue du modèle 1 après détection des frontières (BD1), rangée 4= séquences de tons optimales au niveau de chaque nouveau segment.

On voit bien dans cette figure que l'ajout de frontières a nettement amélioré la modélisation de ce passage. Bien qu'une frontière supplémentaire ait été détectée après « *singularly* ». On

obtient des corrélations de 0,634, 0,533 et 0,548 pour les trois sous-segments respectifs par rapport à une corrélation globale de 0, 202 sans frontières.

La détection des frontières a été effectuée dans les deux premiers modèles optimisés (le modèle des pauses et celui des unités intonatives). Nous présenterons les résultats de détection des frontières obtenus à partir de ces deux modèles dans les sections suivantes. Mais nous décrirons tout d'abord le fonctionnement du script.

7.4.1.1 Positionnement des frontières

On considère l'ensemble des cas de figure suivant :

Pas de frontières (0,0,0)

Positionnement de la 1ère frontière après chaque mot et ce jusqu'à la fin du segment

- ➔ Essai avec une seule frontière (1,0,0)
  - Positionnement de la 2ème frontière après chaque mot à partir de la première frontière et ce jusqu'à la fin
  - Et pour chaque 2ème frontière possible :
    - ➔ Essai avec 2 frontières (1,2,0)
      - Positionnement de la 3ème frontière (si possible)
      - Puis on essaie de positionner la 3ème frontière après chaque mot à partir de la 2ème frontière et ce jusqu'à la fin du segment

Dans tous les cas on estime qu'on ne peut pas ajouter une frontière à moins de 300 ms de la précédente.

Exemples de toutes les combinaisons successives essayées :

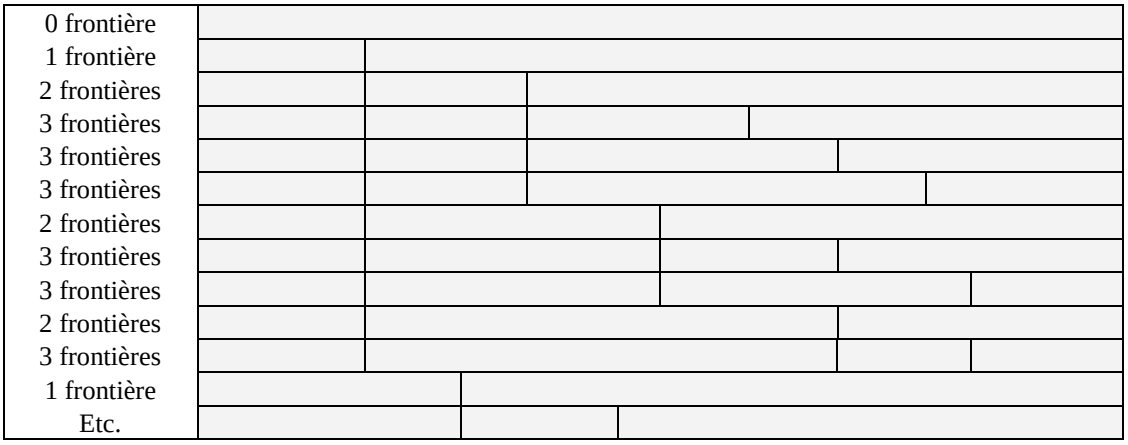


Tableau 13 : Exemple des combinaisons de frontières successivement essayées lors de la détection des frontières d'IU.

7.4.1.2 Évaluation des résultats

Les résultats de la détection automatique des frontières sont comparés avec les frontières annotées manuellement. On classe les résultats en fonction du type d'unité intonative qui avait été annoté manuellement (terminal/non terminal etc.) et du type de modèle (BD1, BD3 pour « Boundary Detection 1 » et « Boundary Detection 3 » qui correspond à la détection des frontières à partir du modèle 1 puis du modèle 3). Dans ce tableau, les frontières déterminées par une pause silencieuse n'apparaissent pas.

Type d'unité intonative	Nombre d'IU annotés	Nombre de frontières détectées avec le modèle BD1	Nombre de frontières détectées avec le modèle BD3
#	351	348	348
[	25	2	6
[!+	1	1	1
[!+_NW			
[_NW		3	3
[|	18	3	3
[|_NW		5	9
[+	292	57	110
[+_NW		59	63
+	1	1	1
(blank)		72	158
Total	**688**	**551**	**702**

Tableau 14 : Résultats de la détection des frontières dans le modèle des pauses (BD1) et dans le modèle des IU (BD3).

> Dans la première colonne, on indique les différents types de frontières annotées ([, [!+ etc.].
>
> On distingue un cas particulier celui des frontières détectées un mot après la frontière d'origine : [|_NW = frontière placée un mot après un IU terminal.

Tout d'abord, 3 frontières terminales sur 18 sont détectées exactement. Ensuite le modèle 1 permet de détecter précisément 57/292 frontières non terminales ce qui correspond à un taux de 19,5 %. Dans le modèle 3, 110/292 frontières non terminales sont détectées ce qui revient à un taux de 38%. Ces résultats peuvent être améliorés si l'on prend en considération les cas où la frontière est détectée un mot inaccentué après la frontière annotée manuellement (c'est la ligne « NW » pour « next word » dans le tableau ci-dessus), on a alors 8/18 frontières terminales détectées dans le premier modèle contre 12/18 dans le second. Au niveau des frontières non terminales, on a 116/292 frontières détectées dans le modèle 1 (40%) contre 173/292 dans le second (59,2%).

Trois types d'erreur dans la détection automatique de frontières ont pu être observés.

- Tout d'abord, les frontières sont fréquemment identifiées un ou deux mots après la frontière annotée manuellement. Ces mots sont généralement des mots grammaticaux, inaccentués que l'on a noté « NW » dans le tableau ci-dessus. Il s'agit surtout de pronoms,

articles ou prépositions. Il semble qu'aucun changement majeur ne se produise dans la direction de F0 à leur niveau si bien que la corrélation est meilleure en les incluant dans la première unité intonative.

- Ensuite des frontières sont parfois omises lorsque la valeur de la corrélation avec l'introduction d'une frontière est légèrement inférieure à la corrélation obtenue au niveau d'un énoncé sans frontières intermédiaires.
- Enfin des frontières supplémentaires sont fréquemment détectées au niveau de l'attaque d'une proéminence.

Ci-dessous quelques exemples tirés du corpus EUROM1.

Exemples pour le fichier Feo40707:

Please take a request | for an early morning taxi (Auteurs)

Please take a request for | an early morning taxi (BD1)

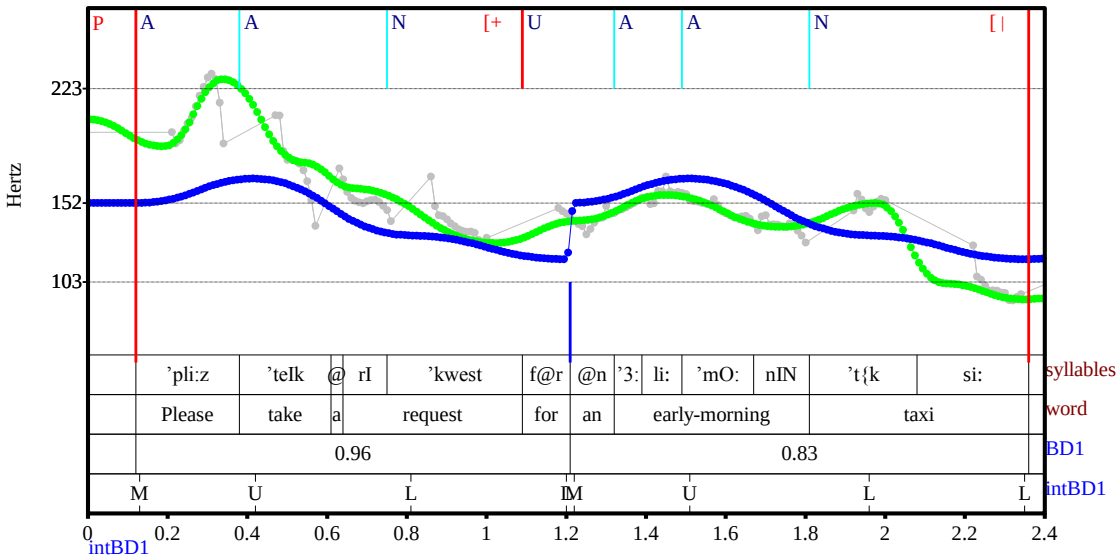


Figure 49 : Courbe de MOMEL originale (en vert) comparée à la courbe modélisée à partir de la détection des frontières dans le modèle BD1 (en bleu) et TextGrid correspondant.

Les annotations manuelles des TU et frontières d'IU sont représentées en haut de la figure. Les frontières détectées sont représentées en bas de la figure en bleu.

Rangée 1= syllabes, rangée 2= mots, rangée 3= corrélation entre MOMEL et la courbe issue du modèle de détection (BD1), rangée 4= les formes optimisées au niveau de chaque segment.

Exemples pour le fichier Fao40073 :

Mark it as top priority | and ensure punctuality (annoté)

Mark it as top priority and| ensure punctuality (BD1)

Exemples pour le fichier Feo1074 :

I have a problem | with my water softener (annoté)

I have a problem with my | water softener (BD1)

The water level | is too high (annoté)

The water level is | too high (BD1)

On compte aussi des cas pour lesquels aucune frontière intermédiaire n'est rajoutée, car la valeur de la corrélation en créant deux groupes est très légèrement inférieure à celle que l'on obtient en un seul groupe.

Exemples pour le fichier Fer1094 :

Maybe a salad | would be more suitable (annoté)

Maybe a salad would be more suitable (BD1)

La dernière erreur récurrente observée concerne les proéminences locales : le début d'une proéminence est assimilé à un début d'unité intonative.

Exemple pour le fichier Fao30072 :

We'd appreciate immediate attempts to trace him|because he sounded desperate| (annoté)

We'd appreciate| immediate attempts to |trace him|because he sounded desperate| (BD3)

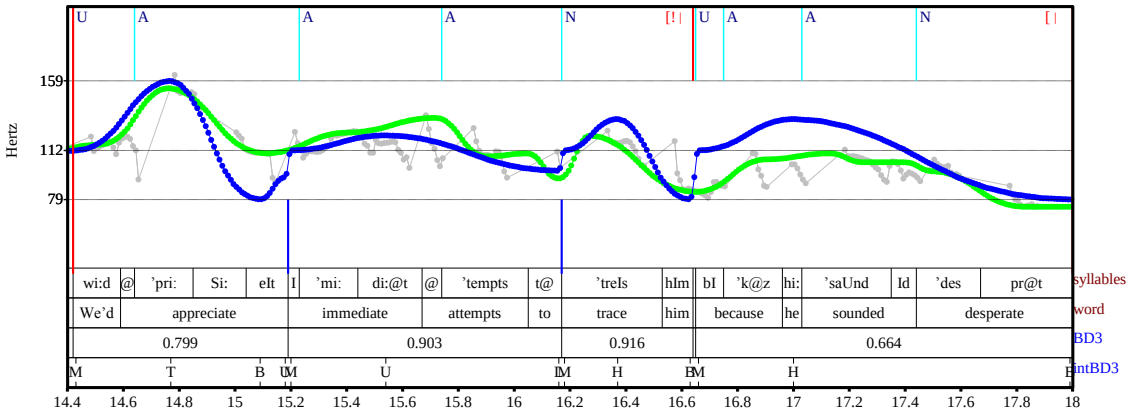


Figure 50 : *Courbe modélisée après détection des frontières dans le modèle 3 (en bleu) comparée à MOMEL (en vert).*

> Rangée 1= syllabes, rangée 2= mots, rangée 3= corrélation entre MOMEL et la courbe issue du modèle de détection 3 (BD3), rangée 4= les formes optimisées au niveau de chaque segment (intBD3).

Exemple pour le fichier Fao60075 :

They moved approximately three months ago (annoté)

They moved approximately three | months ago (BD1)

Dans ces deux derniers exemples une frontière est détectée avant ce que l'on avait annoté comme un nucléus. La valeur de F0 augmente au niveau du nucléus, ce qui est interprété comme le début d'une nouvelle unité intonative. À nouveau ce genre d'erreur suggère le besoin de travailler dans un modèle plus complexe prenant en compte les niveaux de proéminence.

Nous présenterons ensuite dans la section suivante, les étapes de détection de fonctions effectuées dans les modèles des unités tonales.

7.4.2 Détection des types de TU

La deuxième expérience d'extraction de l'information fonctionnelle à partir des représentations formelles consiste à détecter automatiquement le type de TU. Les informations fonctionnelles connues pour cette étape sont les frontières et types d'unités intonatives et les accents lexicaux. Dans ce corpus, tous les TU commencent au niveau d'un accent lexical donc cela revient à connaître aussi les frontières d'unités tonales. Au niveau de chaque TU annoté, on cherche à connaître le niveau de proéminence auquel on a affaire.

On prend en compte deux types de détection des types de TU. La première est dite libre ou sans contrainte. Au niveau de chaque TU, tous les types de fonctions possibles sont pris en compte, c'est la fonction associée à la meilleure séquence de ton pour un TU donné qui est retenue quelle que soit sa nature. Cela implique qu'il est possible de détecter plusieurs nucléus dans une unité intonative. Il s'agit des modèles 5 et 7.

La seconde expérience de détection de TU se fait dans les modèle 6 et 8 qui comprennent la contrainte du nucléus. On considère qu'il ne peut y avoir qu'un nucléus par unité intonative tout comme dans les modèles britanniques. Par conséquent, afin de déterminer les fonctions possibles pour chaque TU, on fait d'abord varier la position du nucléus.

La procédure de détection des types de TU pour les modèles sans contrainte comprend les étapes suivantes :

- Détermination de la position du TU dans l'unité intonative IU ;
- Détermination des fonctions possibles pour chaque TU ;
- Recherche de la fonction optimale pour chaque TU.

La procédure de détection des types de TU pour les modèles avec la règle du nucléus se déroule de la façon suivante :

- Détermination de la position de chaque TU dans l'unité intonative ;
- Pour chaque position du nucléus possible :
  - Détermination des fonctions possibles pour chaque TU,
  - Recherche de la fonction optimale pour chaque TU,
  - Calcul du score global obtenu au niveau de l'unité intonative ;
- La position du nucléus qui donne le meilleur score global au niveau de l'IU est retenue.

7.4.2.1 Détermination de la position du TU dans l'unité intonative

La position de chaque TU dans un IU est déterminée selon les critères suivants :

- Si le TU est inaccentué (U) sa position est « 0 » ;
- S'il s'agit du premier TU accentué ou s'il suit un U0, il a la position 1 ;
- Si c'est le dernier TU d'un IU, il a la position F ;
- Les TU intermédiaires ont la position M ;
- Si le TU de position 1 est aussi le dernier, il aura la position 1F.

7.4.2.2 Détermination des fonctions possibles pour chaque TU dans les modèles sans contrainte

Les fonctions possibles sont déterminées par la position du TU dans l'IU et aussi, pour les TU en position initiale par le type de l'IU précédent, enfin pour les TU en position finale par le type de l'IU :

- Si le TU est de type 0, il peut être caractérisé de deux façons :
  - Si l'IU précédent est de type terminal, il devient : « |U0 »,
  - Si l'IU précédent est de type non terminal, il devient : « +U0 » ;
- Si le TU est en position 1, il peut être caractérisé de trois façons :
  - S'il est en position 1 après un TU U0, les fonctions possibles sont {A1, N1, !N1},
  - Si l'IU précédent est de type terminal et s'il n'y a pas de U0, les fonctions possibles sont {|A1, |N1, | !N1},
  - Si l'IU précédent est de type non terminal sans U0, les fonctions possibles sont {+A1, +N1, + !N1} ;
- Si le TU est en position M, on a les fonctions suivantes : {AM, NM, !NM} ;

- Si le TU est en position F :
  - Si l'IU est de type terminal, les fonctions possibles sont { AF|, NF|, !NF|},
  - Si l'IU est de type non terminal, les fonctions possibles sont {AF+, NF+, !NF+} ;
- Si le TU est de position 1F (le seul TU de l'IU) on a les fonctions suivantes : { N1F|, N1F+}, selon que l'IU est respectivement de type terminal ou non terminal.

Le tableau ci-dessous résume les fonctions possibles :

TU en Position 0	Position 1 en début d'IU	Position 1 précédé d'un TU 0	Position M	Position F	Position 1F
|U0 +U0	|A1 +A1 |N1 +N1 |!N1 +!N1	A1 N1 !N1	AM NM !NM	AF| AF+ NF| NF+ !NF| !NF+	N1F| N1F+

Tableau 15 : Tous les types de TU possibles dans les modèles sans contrainte en fonction de la position du TU dans l'IU.

7.4.2.3 Détermination des fonctions possibles pour chaque TU dans les modèles avec règle du nucléus

Comme pour les modèles sans contrainte, les fonctions possibles dépendent de la position du TU dans l'IU, du type de l'IU précédent, et du type de l'IU.

Pour les modèles qui comprennent la contrainte du nucléus, on cherche d'abord la meilleure position pour le nucléus, ce qui va déterminer la nature des autres TU. Pour chaque position du nucléus possible, une nouvelle série d'étiquettes fonctionnelles pour chaque TU est alors définie comme le montre le tableau récapitulatif ci-dessous :

Position du nucléus	TU en position 0	TU en position 1	TU en position M	TU en position M	TU en position F
1	|U0 +U0	N1 |N1 +N1 !N1 | !N1 + !N1	ATM	ATM	ATF+ ATF|
2	|U0 +U0	A1 |A1 +A1	NM !NM	ATM	ATF+ ATF|
3	|U0 +U0	A1 |A1 +A1	AM	NM !NM	ATF+ ATF|
4	|U0 +U0	A1 |A1 +A1	AM	AM	NF| NF+ !NF| !NF+

Tableau 16 : Tous les types de TU possibles avec la règle du nucléus.

On rajoute donc essentiellement la lettre T pour « tail » qui fait référence aux unités tonales se trouvant après le nucléus. On peut donc avoir des TU de type « ATM » et » ATF ».

7.4.2.4 Détection des types de TU

Afin de déterminer la fonction de chaque TU, on s'appuie sur toutes les séquences de tons et alignements retenus pour chaque fonction initialement. Les séquences de tons optimales sont converties en points cibles MOMEL. La courbe modélisée qui se rapproche le plus de la courbe d'origine est choisie et les fonctions correspondantes retenues.

7.4.2.5 Détection des types de TU dans les modèles avec règle du nucléus

On place le nucléus dans toutes les positions possibles, et pour chaque position du nucléus, on détermine la fonction de chaque TU comme expliqué ci-dessus. Puis on calcule le score global au niveau de l'IU. La position du nucléus qui donne le meilleur score global au niveau de l'IU est retenue et par là-même les fonctions déterminées avec cette position du nucléus.

Dans la section suivante, nous décrirons la procédure de détection des frontières d'IU adaptée au modèle des TU.

7.4.3 Détection des frontières d'IU dans le modèle des TU

La procédure de détection des frontières d'IU dans le modèle des TU est identique à la précédente à l'exception du fait que l'on retient les meilleures séquences obtenues au niveau des TU cette fois-ci afin de tester et de trouver la meilleure frontière intermédiaire possible.

Afin de tester les meilleurs positionnements pour une frontière, la procédure est donc la suivante :

- Tout d'abord, comme dans la détection des frontières dans les modèles des pauses et des IU (cf. 7.4.1), on essaie de placer une frontière après le premier mot, le second mot etc. ;
- Pour chacune de ces nouvelles unités intonatives, les nouvelles unités tonales sont étiquetées automatiquement en se basant sur les accents lexicaux. Un TU va d'une syllabe accentuée à la prochaine syllabe accentuée :
    - Si la première syllabe n'est pas accentuée, on obtient un TU de type « U » jusqu'à la prochaine syllabe accentuée,
    - Si la première syllabe est accentuée on aura un TU de type « A » ou « N » ou « !N » jusqu'à la prochaine syllabe accentuée ;
- Au niveau des nouvelles unités tonales, les formes optimales retenues pour chaque type de TU sont testées en utilisant la corrélation et l'écart quadratique moyen ;
- Le placement de frontière qui permet d'obtenir le meilleur score global au niveau de l'IU est conservé ;
- Cette procédure est répétée en essayant de placer deux frontières puis trois jusqu'à ce que l'on trouve la meilleure solution.

La procédure a déjà été décrite en détail dans les sections précédentes, nous ajouterons donc seulement ici un exemple détaillé :

Ci-dessous un exemple des traitements effectués afin de placer une frontière intermédiaire sur un passage du TextGrid FEO10704, dans un modèle sans contrainte :

D'abord pas de frontière

	I	have a	problem with my	water	softener
Type de TU	U	A	A	N	A
Position	0	1	M	M	F
Fonctions possibles	+U0 |U0	A1 N1 !N1	AM NM !NM	AM NM !NM	AF+ AF| (!)NF+ (!)NF|
Retenu niveau TU + score	U0 M _ U -8.0123	A1 _ H U 9.99906	AM _ D L 9.98679	NM _ H L 9.9863	NF+ _ D L 9.99028
score global	9.98866				

Une frontière après « *problem* »

	I	'have a	'problem	with my	'water	'softener
Type de TU	U	A	A	U	A	A
Position	0	1	F	0	1	F
Fonctions possibles	+U0 |U0	A1 N1 !N1	AF+ AF| (!)NF+ (!)NF|	U0	A1 N1 !N1	AF+ AF| (!)NF+ (!)NF|
Retenu niveau TU + note	U0 M _ U -8.0123	A1 _ H U 9.99906	!NF+ _ H L 9.99469	U0 _ U U 7.98354	!N1 _ H D 9.98533	NF| _ L L 9.99456
Note niveau ss-segment	9.99454			9.98735		
Note globale	9.99094					

Une frontière après « *problem* » et une 2ème après « *water* »

	I	'have a	'problem	with my	'water	'softener
Type de TU	U0	A	A	U	A	A
Position	0	1	F	0	1F	1F
Fonctions possibles	+U0 |U0	A1 N1 !N1	AF+ AF| (!)NF+ (!)NF|	U0	A1F+ A1F| (!)N1F+ (!)N1F|	A1F+ A1F| (!)N1F+ (!)N1F|
Retenu niveau TU + note	U0 M _ U -8.0123	A1 _ H U 9.99906	!NF+ _ H L 9.99469	U0 _ U U 7.98354	N1F _ H L 8.98683	N1F _ H B 6.97006
Note niv ss-seg	9.99454			6.98506		6.97006
Note globale	7.98322					

Une frontière après « *with* »

	I	‘have a	‘problem with	my	‘water	‘softener
Type de TU	U	A	A	U	A	A
Position	0	1	F	0	1	F
Fonctions possibles	+U0 |U0	A1 N1 !N1	AF+ AF| (!)NF+ (!)NF|	+U0 |U0	A1 N1 !N1	AF+ AF| (!)NF+ (!)NF|
Retenu niveau TU + note	U0 M _ U -8.0123	A1 _ H U 9.99906	!NF+ _ U L 9.99367	U0 _ _ U 9.99219	!N1 _ H D 9.99587	NF _ L L 9.9986
Note niveau sous-segment	9.99392			9.99564		
Note globale	**9.99477** **C’est cette combinaison qui sera retenue**					

Une frontière après « *with* » et après « *water* »

	I	‘have a	‘problem with	my	‘water	‘softener
Note niveau sous-segment	M _ U	_ H U 9.99392	_ U L	_ _ U	_ H L 9.99323	_ H B 6.9714
Note globale	8.98618					

Une frontière après « *my* »

	I	‘have a	‘problem with my	‘water	‘softener
Note niveau sous-segment	M _ U	_ H U 9.98884	_ D L	_ H S 9.9944	_ L L
Note globale	9.99162				

Tableau 17 : Un exemple des traitements effectués afin de placer une frontière intermédiaire sur un passage du TextGrid FEO10704.

7.4.4 Résultats des détections des fonctions

7.4.4.1 Détection des frontières d'unités intonatives

Les résultats de détections des frontières comparés aux frontières annotées manuellement sont regroupés dans le tableau ci-dessous :

Type de frontière	NB frontières annotées	Nombre de frontières détectées selon le modèle					
		BD1	BD3	BD5	BD6	BD7	BD8
#	351	348	348	347	348	348	348
[	25	2	6	6	7	3	5
[!+	1	1	1				1
[!+_NW				1			
[_NW		3	3	7	8	4	1
[|	18	3	3	5	4	5	6
[|_NW		5	9	6	7	4	4
[+	292	57	110	80	98	89	95
[+_NW		59	63	83	78	63	66
+	1	1	1	1		1	
(blank)		72	158	253	204	128	142
Total	**688**	**551**	**702**	**789**	**754**	**645**	**668**

Tableau 18 : Résultats des détections de frontières dans les modèles 1 à 8.

Dans ce tableau, les frontières déterminées par une pause silencieuse n'apparaissent pas. En prenant en compte les frontières décalées d'un mot inaccentué, les résultats sont les suivants.

Pour les frontières terminales, on obtient les résultats suivants :

- Modèles 5 et 6 (BD5 et BD6) = 11/18 ou 61%
- Modèle 7 (BD7) = 9/18 ou 50%
- Modèle 8 (BD8) = 10/18 ou 55%

Pour les frontières non terminales, on obtient les résultats suivants :

- BD5 = 163/292 ou 56%
- BD6 = 176/292 ou 60%
- BD7 = 152/292 ou 52%
- BD8 = 161/292 ou 55%

C'est donc le modèle 6 appliquant la règle du nucléus et un alignement en valeurs absolues, qui permet la meilleure détection automatique de frontières intermédiaires.

Observons quelques exemples d'erreurs au niveau du placement des frontières :

La frontière est décalée d'un mot inaccentué

She's going to the Far East| for a holiday. (auteurs)

She's going to the Far East for| a holiday. (BD5)

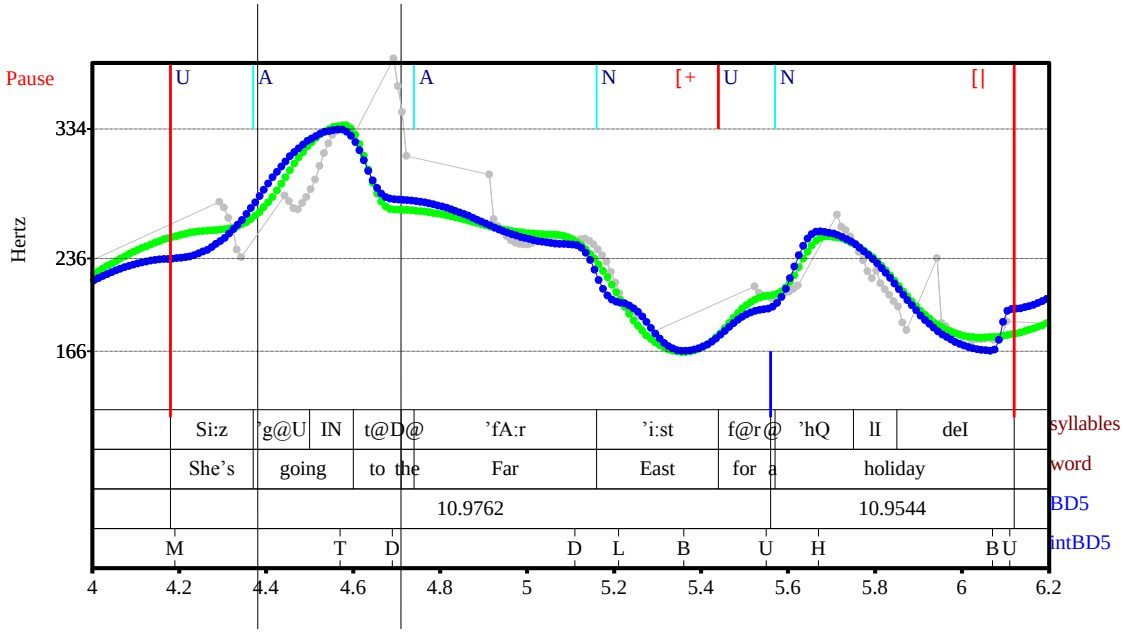


Figure 51 : *Courbe modélisée après détection des frontières dans le modèle 5 (en bleu) comparée à MOMEL (en vert) pour le fichier* FFO00792.

> Rangée 1= syllabes, rangée 2= mots, rangée 3= valeur du score (combinaison de la corrélation et de l'écart quadratique moyen) (BD5), rangée 4= meilleures formes optimisées au niveau de chaque segment (intBD5).

Ci-dessus un cas classique dans les modèles précédents : la frontière est placée après un mot court de type grammatical, ici « for ».

On trouve aussi des exemples de décalage de frontière par un mot accentué.

Un début de proéminence est confondu avec un début d'unité intonative

They moved approximately three months ago| (auteurs)

They moved approximately |*three months ago*| (BD5)

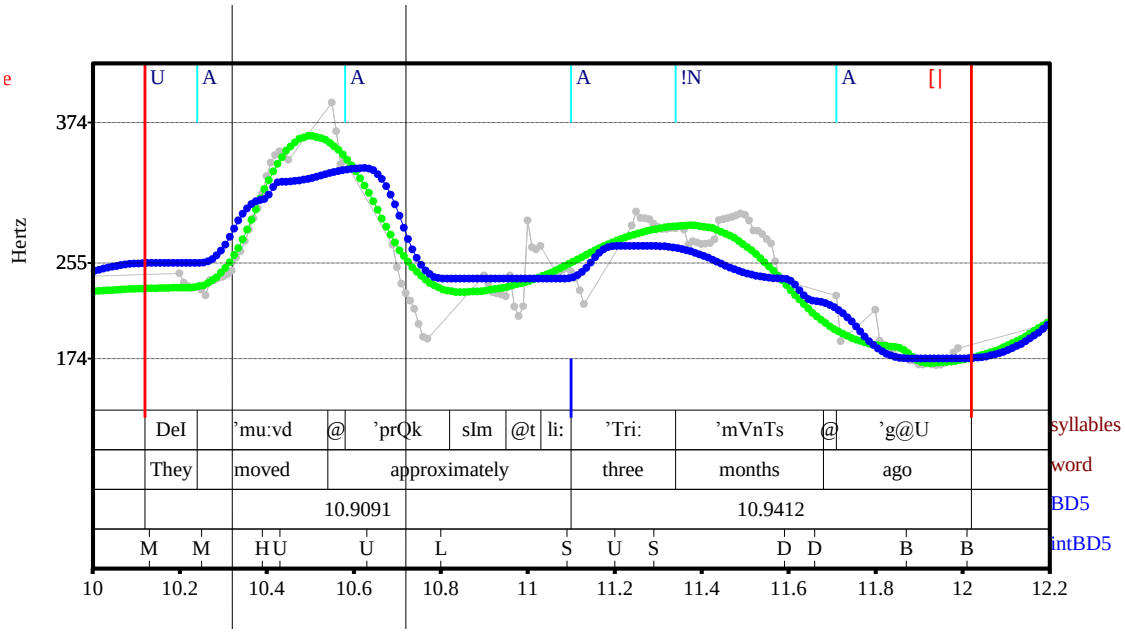


Figure 52 : *Courbe modélisée après détection des frontières dans le modèle 5 (en bleu) comparée à MOMEL (en vert) pour le fichier FFO60788.*

Il s'agit pour le script de choisir la combinaison de frontières et de séquences de tons qui permet de suivre la courbe au mieux. L'optimisation des tons au départ peut ainsi limiter les choix possibles. A chaque fois que la courbe entame une franche montée, ou un contour global montant-descendant, on détecte automatiquement un début d'IU qui ne coïncide pas forcément avec la pause syntaxiquement correcte mais qui marque le début d'un mouvement prosodique cohérent. On trouvait déjà ce type d'erreur dans les modèles précédents qui ne prenaient pas les proéminences en compte.

Omissions de frontières

Could you arrange|to send an engineer| on Tuesday morning please ? (Auteurs)

Could you arrange to send an engineer on Tuesday morning please ? (BD7)

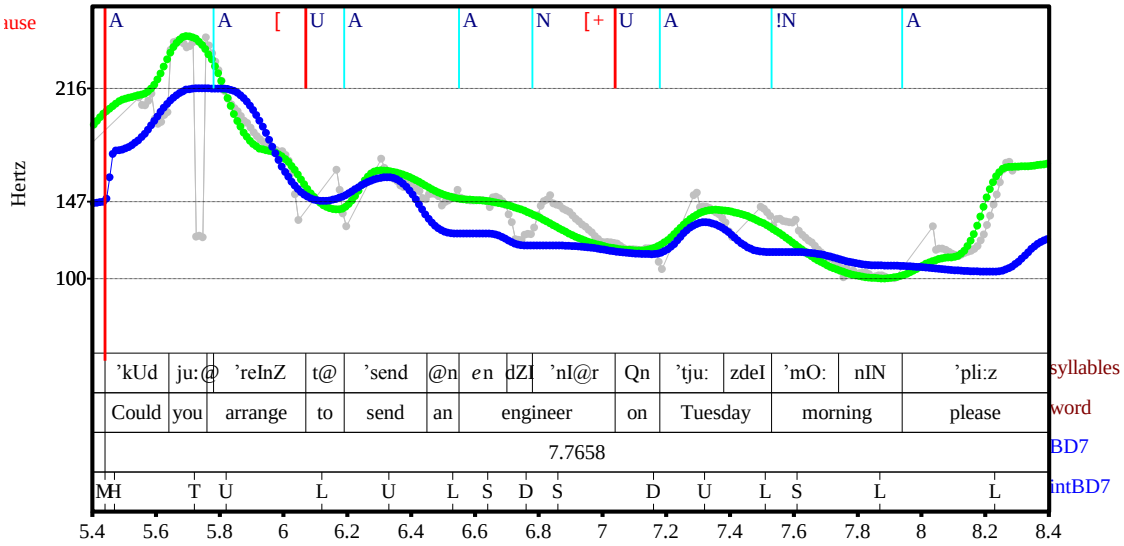


Figure 53 : Courbe modélisée après détection des frontières dans le modèle 7 (en bleu) comparée à MOMEL (en vert) pour le fichier FEo2075.

7.4.4.2 Détection des types de TU

Afin d'observer et d'évaluer les résultats de détection des types de TU, on se place dans le modèle où les frontières intonatives sont déjà fixées. Les unités intonatives sont ensuite découpées en TU et il ne reste plus qu'à trouver pour chaque TU, la fonction qui permettra de suivre la courbe au mieux.

Les résultats sont présentés sous la forme de matrices de confusion comparant les types de TU annotés manuellement et les types de TU détectés automatiquement. On distingue quatre modèles de détection : le modèle 5 dont l'alignement est exprimé en valeur absolue et qui n'obéit pas à la règle du nucléus, le modèle 6 lui est identique, mais répond à la règle du nucléus, les modèles 7 et 8 qui sont les modèles correspondants avec un alignement exprimé en pourcentage cette fois-ci (cf. section 7.3.3 pour les paramètres).

Les résultats de détection tous modèles confondus sont les suivants :

TU annotés	TU détectés				
	!N	A	N	U	Total
!N	136	90	121		347
A	338	2284	744		3366
N	436	675	449		1560
U	2	7	3	1825	1837
Total	912	3056	1317	1825	7110

Tableau 19 : Matrice de confusion des types de TU annotés et des types de TU détectés tous modèles confondus.

De façon globale, 136/347 nucléus emphatiques sont bien détectés ce qui correspond à un taux de réussite de 39,2 %. 2284/3366 TU accentués sont bien identifiés ce qui revient à 68% et enfin 449/1560 nucléus non marqués sont bien identifiés ce qui revient à 29%. Dans l'ensemble, il y a un nombre important de nucléus confondus avec des TU accentués simples. On peut analyser les résultats de manière plus approfondie en faisant apparaître les résultats par modèle.

TU annotés	Modèle	TU détectés					Total
		!N	A	N	U	(blank)	
!N	5	41	21	58			120
	6	38	42	42			122
	7	48	22	52			122
	8	37	52	33			122
A	5	159	651	420		10	1240
	6	89	975	167		10	1241
	7	189	628	424			1241
	8	91	951	199			1241
N	5	209	154	163		1	527
	6	156	222	152		1	531
	7	123	237	171			531
	8	88	247	196			531
U	5	1	1	1	457		460
	6		2	1	457		460
	7	1	3		456		460
	8	1	2	1	456		460
Total		1271	4210	2080	1826	22	9409

Tableau 20 : Matrice de confusion des types de TU annotés et des types de TU détectés pour les modèles 5, 6, 7 et 8.

On cherche surtout à travers cette étape à affiner la définition de chaque fonction. Peut-on distinguer les TU de type « A », « N » et « !N » ? On observe que dans le modèle 5, 651 TU «A « sont bien détectés alors que 420 d'entre eux sont pris pour un nucléus et 159 pour un nucléus emphatique. Dans le modèle 6, on trouve beaucoup plus de TU « A », en effet on a ajouté une

contrainte ne permettant qu'un seul nucléus par unité intonative. Il y a donc moins de confusions dans ce sens-là, mais cela signifie-t-il que le nucléus identifié est bien le même que celui qui a été annoté au départ ? Pas forcément de toute évidence, si l'on observe la ligne des résultats pour les TU « N », 222 TU « A » sont pris pour de TU « N » et l'on détecte moins de nucléus exacts que dans le modèle 5 (152 contre 163). En ce qui concerne les modèles alignés en pourcentage, on retrouve un schéma similaire aux précédents. Le modèle 8 incluant la règle du nucléus a des résultats légèrement supérieurs au modèle 7.

L'observation de ces résultats dans plus de détails pourrait nous permettre de choisir la meilleure représentation fonctionnelle pour ce corpus, est-il plus judicieux de n'avoir qu'un nucléus par unité intonative où un modèle permettant la présence de plusieurs nucléus ou accents mélodiques ne serait-il pas plus représentatif des formes dont nous disposons ?

Quelques exemples :

Cas où le modèle 7 permettant la présence de plusieurs nucléus donne une meilleure représentation formelle

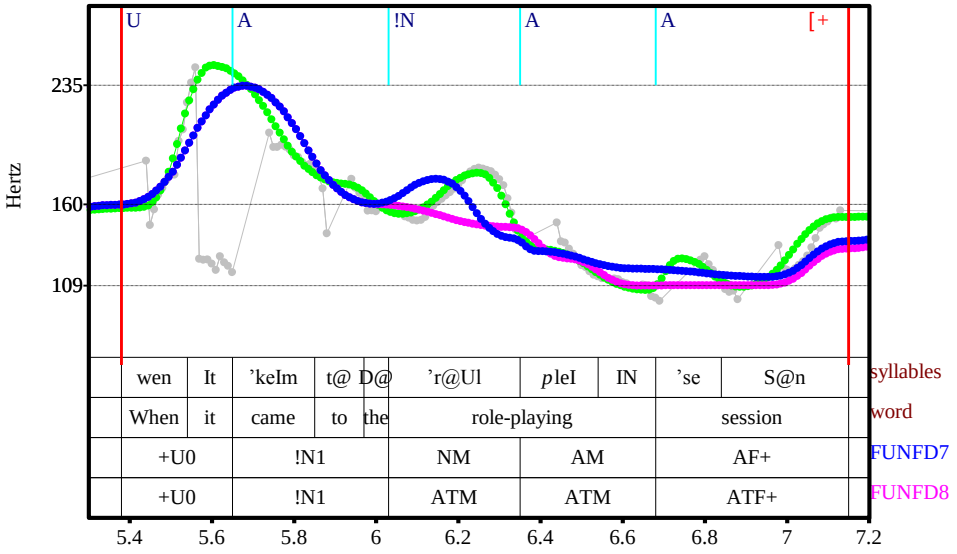


Figure 54 : Comparaison entre deux courbes modélisées après détections des fonctions dans les modèles 7 (en bleu) et 8 (en rose)

> En haut de la figure = annotation manuelle des TU.
>
> Rangée 1= syllabes, rangée 2= mots, rangée 3= TU détectés automatiquement avec le modèle 7 (FUNFD7), rangée 4= TU détectés automatiquement avec le modèle 8 (FUNFD8).

Dans le modèle 7, deux nucléus sont détectés, l'un commençant sur « came » l'autre sur « role ». Cette modélisation permet de suivre la courbe de MOMEL d'origine de façon plus précise que le second modèle qui met un nucléus sur « came » et des accents postnucléaires sur les TU suivants.

Cas où le modèle du nucléus donne une meilleure représentation

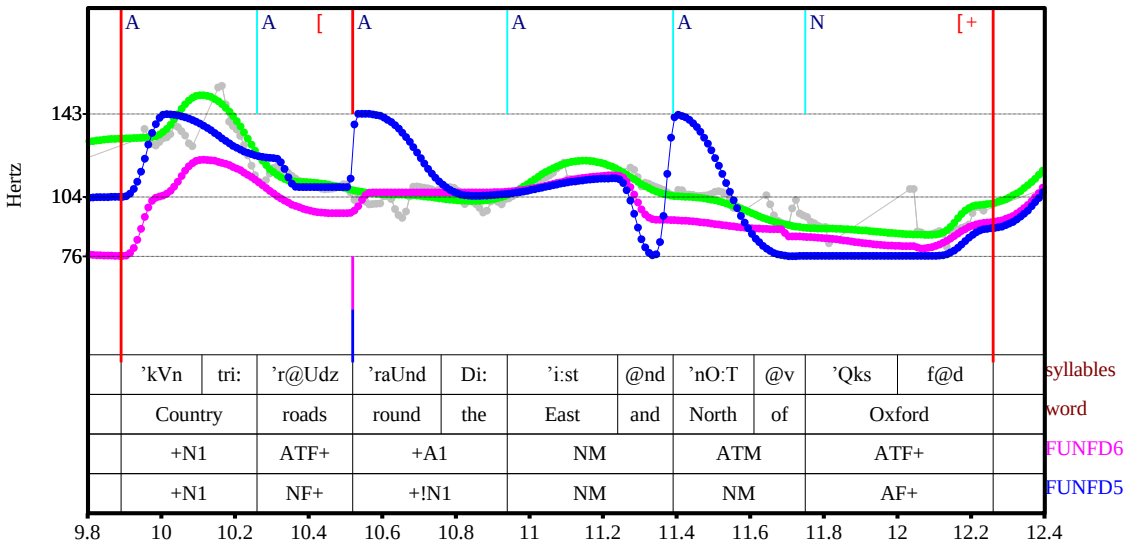


Figure 55 : Courbes issues de la détection des fonctions dans les modèles 5 et 6.

Dans cet exemple, c'est le modèle 6 qui permet de suivre la courbe de MOMEL le plus précisément en ne détectant qu'un TU nucléus sur « *East* » plutôt que trois nucléus dans la modèle 5 sur « *Round, East et North* ».

7.5 Conclusions

7.5.1 Evaluation des modèles

Nous présenterons dans cette section, une évaluation des modèles que nous avons optimisés en fonction des trois critères que sont la corrélation moyenne, le score moyen et l'écart quadratique moyen.

On rassemble ici les résultats par type d'optimisation (optimisation des séquences de tons « inttone », optimisation de l'alignement « intalign », détection des frontières « intBD », détection des autres fonctions « intFD »), par modèle (5, 6, 7, 8) et par type d'IU (tous, « [+ », « [| ».

Résultats pour les modèles d'optimisation des tons :

tous		Non terminal [+		Terminal [|	
Moyennes des corrélations					
inttone5	0.885234	inttone5	0.909098	inttone5	0.853341
inttone6	0.885234	inttone6	0.909098	inttone6	0.853341
inttone7	0.865016	inttone7	0.882738	inttone7	0.847801
inttone8	0.865016	inttone8	0.882738	inttone8	0.847801
Scores moyens					
tous	**score**	**[+**	**score**	**[|**	**score**
inttone5	9.866994	inttone5	10.16472	inttone5	9.453733
inttone6	9.866994	inttone6	10.16472	inttone6	9.453733
inttone7	9.660523	inttone7	9.912521	inttone7	9.388712
inttone8	9.660523	inttone8	9.912521	inttone8	9.388712
Ecart quadratique moyen					
tous	**RMSE**	**[+**	**RMSE**	**[|**	**RMSE**
inttone6	1.578243	inttone6	1.10127	inttone6	2.15311
inttone5	1.578243	inttone5	1.10127	inttone5	2.15311
inttone7	1.753628	inttone7	1.371645	inttone7	2.155028
inttone8	1.753628	inttone8	1.371645	inttone8	2.155028

Tableau 21 : Classement des modèles d'optimisation des tons en fonction de la corrélation, du score et de l'écart quadratique moyen.

Au niveau de cette première étape d'optimisation, les meilleurs modèles sont le modèle 5 et le modèle 6.

Résultats pour les modèles d'optimisation de l'alignement :

tous		Non terminal [+		Terminal [|	
Moyennes de corrélation					
intalign6	0.79439	intalign6	0.818572	intalign8	0.768472
intalign5	0.781535	intalign5	0.800912	intalign7	0.764679
intalign7	0.779128	intalign7	0.793547	intalign6	0.760809
intalign8	0.777311	intalign8	0.790229	intalign5	0.752
Scores moyens					
intalign6	8.723972	intalign6	9.008939	intalign8	8.417813
intalign5	8.575703	intalign5	8.80364	intalign7	8.358052
intalign7	8.523114	intalign7	8.686789	intalign6	8.339692
intalign8	8.518246	intalign8	8.663276	intalign5	8.242708
Ecart quadratique moyen					
intalign6	2.218496	intalign6	1.736537	intalign8	2.661919
intalign7	2.443555	intalign5	2.045204	intalign7	2.691081
intalign8	2.404246	intalign8	2.009537	intalign6	2.668984
intalign5	2.47111	intalign7	2.051474	intalign5	2.739293

Tableau 22 : Classement des modèles d'optimisation de l'alignement en fonction de la corrélation, du score et de l'écart quadratique moyen.

Le meilleur modèle d'alignement est le modèle 6 qui obtient les meilleures moyennes à partir des trois critères testés. L'alignement en valeur absolue donne donc ici de meilleurs résultats.

Résultats pour les modèles de détections des frontières

tous		Non terminal [+		Terminal[|	
Moyennes de corrélation					
intBD5	0.836865	intBD6	0.851448	intBD5	0.820589
intBD6	0.835894	intBD5	0.851065	intBD6	0.817163
intBD7	0.820815	intBD7	0.835544	intBD7	0.808488
intBD8	0.81307	intBD8	0.826441	intBD8	0.801159
intBD3	0.749472	intBD3	0.768572	intBD3	0.724224
intBD1	0.605773	intBD1	0.604678	intBD1	0.623089
Scores moyens					
intBD6	9.254108	intBD6	9.439287	intBD5	9.054155
intBD5	9.233873	intBD5	9.401483	intBD6	9.032409
intBD7	9.043112	intBD7	9.227915	intBD7	8.883367
intBD8	8.98083	intBD8	9.155656	intBD8	8.811367
intBD3	7.86412	intBD3	8.146481	intBD3	7.457263
intBD1	6.406869	intBD1	6.397989	intBD1	6.586686
Ecart quadratique moyen					
intBD6	1.905349	intBD6	1.479665	intBD6	2.375122
intBD5	1.96056	intBD5	1.554048	intBD5	2.398785
intBD8	2.032712	intBD8	1.622045	intBD8	2.509728
intBD7	2.062429	intBD7	1.652287	intBD7	2.521593
intBD3	3.203426	intBD3	2.66738	intBD1	3.708134
intBD1	3.537314	intBD1	3.278411	intBD3	3.863224

Tableau 23 : Classement des modèles de détection des frontières en fonction de la corrélation, du score et de l'écart quadratique moyen.

La meilleure modélisation est à nouveau obtenue à partir du modèle 6 suivi de près par le modèle 5.

Résultats pour les modèles de détection des proéminences :

Tous		Non terminal [+		Terminal [|	
Moyennes de corrélation					
intFD6	0.794805	intFD6	0.827801	intFD6	0.779915
intFD8	0.772565	intFD8	0.802695	intFD8	0.77052
intFD5	0.76607	intFD7	0.799169	intFD7	0.766963
intFD7	0.764443	intFD5	0.798219	intFD5	0.743024
Scores moyens					
intFD6	8.716495	intFD6	9.118006	intFD6	8.4694
intFD8	8.455191	intFD8	8.837056	intFD8	8.324581
intFD7	8.351543	intFD7	8.770397	intFD7	8.256573
intFD5	8.346161	intFD5	8.746911	intFD5	8.016499
Ecart quadratique moyen					
intFD5	2.714542	intFD5	2.161655	intFD5	3.139646
intFD7	2.435615	intFD7	1.925275	intFD7	3.053183
intFD8	2.406439	intFD8	1.887907	intFD8	2.946533
intFD6	2.263972	intFD6	1.743592	intFD6	2.7965

Tableau 24 : Classement des modèles de détection des proéminences en fonction de la corrélation, du score et de l'écart quadratique moyen.

C’est à nouveau le modèle 6 qui se détache pour cette étape.

C’est le modèle 6, avec un alignement en valeurs absolues et la règle du nucléus qui permet d’obtenir les meilleurs résultats de façon générale. On remarque aussi que l’optimisation des modèles de synthèse nous a permis d’améliorer la modélisation des contours de F0 de façon significative. On passe de moyennes de corrélation avoisinant la valeur de 0,5 à des corrélations moyennes allant jusqu’à environ 0,8.

7.5.2 Modèles des formes prosodiques

Les contours intonatifs « canoniques » développés à partir du corpus EUROM1 restent très proches des contours prédits dans les modèles de synthèse de départ. En se basant sur l’optimisation des modèles 5, 6, 7 et 8, on peut en déduire les contours généraux trouvés dans les données.

Pour un IU non terminal ([+) non marqué :

[MHT D DD ULU] « downstepping head+ fall-rise »

[A A A !N +

[MUU LD DL BU] « downstepping head+ low rise »

[A A A N +

Pour un IU non terminal marqué :

[MHU HBU] « high head+ fall rise »

[A !N +

[MHD UL DU BD LU] « sliding head + fall-rise »

![A A A N A +

Un énoncé incomplet (« interrupted glide down ») :

[M H D D S[

[A A A A[

Pour un IU terminal ([|) :

[MHTDLDD/HBB] « downstepping head+low/high fall »

[A A A !N |

[BUTDULHBB] « sliding head+high fall »

![A A A !N |

On remarque que les contours non terminaux sont mieux représentés à partir de ce corpus en utilisant le ton {U} plutôt que le ton {H}. La montée finale est généralement plus réduite et peut même être optimisée par la séquence {DS} ou {LD} qui semble correspondre à ce qui a été décrit dans la littérature comme une chute courte ou interrompue (« interrupted glide down ») donnant une impression d'inachèvement même s'il n'y a pas de contour final montant évident. Liberman (1975) représente ce contour nucléaire à l'aide de la séquence {HM} (où M = «mid ») plutôt que {HL}, alors que Faure (1962) parle de chute supérieure allant des aigus vers les médiums.

Les contours prénucléaires restent quasiment identiques aux contours de départ à l'exception du cas de la tête haute « high head ». La pré-signalisation de l'emphase au niveau des accents prénucléaires est très peu représentée dans ce corpus si bien que cette étiquette fonctionnelle n'a pas pu être enrichie à partir de ces données.

7.5.3 Détection automatique et réinterprétation des fonctions prosodiques

En ce qui concerne la détection des frontières, on remarque que les modèles qui donnent les meilleurs résultats semblent être les modèles 3 et 6 (cf. tableau 18). Le modèle simple en quatre tons et à partir de l'annotation manuelle des frontières intonatives est aussi performant qu'un modèle plus complexe prenant en compte les unités tonales. Ce sont surtout les contours initiaux et finaux qui nous intéressent quand il s'agit de démarquer des frontières. Il reste à trouver le moyen de distinguer le début d'une proéminence et le début d'un énoncé qui sont trop souvent confondus dans cette étape de détection et conduisent à un découpage erroné des énoncés. Comment peut-on distinguer l'attaque d'une proéminence au début d'un énoncé de celle d'une proéminence marquée (nucléus ou nucléus emphatique) à l'intérieur d'un énoncé ? La prise en compte des variations de la durée permettrait certainement d'améliorer la distinction entre ces deux cas de figure comme Campbell (1993) l'a montré.

La détection des types d'unité tonale est plus problématique dans l'ensemble et conduit à se poser la question suivante : combien de niveaux de proéminence devraient être pris en compte afin de reproduire correctement les contours de F0 ? On a bien vu que le modèle 6 appliquant la règle du nucléus, donne de meilleures corrélations que le modèle 5 par exemple. Cependant il reste beaucoup de marge de confusion entre l'étiquette « A » et l'étiquette « N ». Si l'optimisation à partir de l'étiquette « A » donne généralement un contour du type {DL} ou {MU} c'est-à-dire un mouvement mélodique réduit presque statique, on trouve aussi des cas d'accents plus marqués ne ressortant pas dans l'étape d'optimisation globale. Ainsi lorsque l'on se place dans une étape de détection des fonctions du type 5 (sans contrainte de n'avoir qu'un seul nucléus), on place volontiers plusieurs nucléus dans un même énoncé. Il nous semble qu'en considérant tous les accents prénucléaires avec une seule étiquette « A », la spécificité de certains accents plus marqués mélodiquement noyés dans la masse d'accents plus neutres, est perdue. Ainsi, il serait peut-être judicieux d'ajouter un niveau de proéminence supplémentaire pour représenter ces cas précis. On aurait alors cinq niveaux de proéminence : un accent faible, un accent marqué, un accent nucléaire et un accent nucléaire emphatique.

Ce système correspond à celui développé à partir des théories britanniques de l'intonation et plus particulièrement dans les TSM. L'accent faible sera ainsi noté « S » (pour stressed) et correspondra à un accent majoritairement rythmique, la lettre « A » signifiera « accented » et désignera une proéminence principalement mélodique.

Le travail mené sur le corpus EUROM1 a permis de tester notre méthode de recherche par analyse par synthèse. Il en ressort des résultats prometteurs, mais il semble aussi que le corpus de base utilisé présente des limites quant à son contenu sémantique ou plus particulièrement quant au style de parole utilisé. Il s'agit en effet de parole de laboratoire et dans l'ensemble d'un style plutôt neutre ou peu expressif d'où le manque de représentation des phénomènes d'emphase par exemple. Il nous paraît donc important de tester notre méthode de recherche sur un corpus plus expressif, sur de la parole plus authentique afin d'enrichir en particulier notre système d'annotation fonctionnelle. Le corpus Aix-MARSEC sera utilisé à cet effet et décrit dans une seconde section de cette partie expérimentale.

Chapitre 8

Optimisation de la relation forme/fonction à partir du corpus Aix-MARSEC

8.1 Introduction

À partir des enregistrements de quatre locuteurs d'EUROM1, nous avons développé et optimisé la procédure d'analyse par synthèse de l'articulation forme/fonction. Toute une série de scripts a été élaborée afin d'automatiser l'optimisation des représentations formelles à partir des représentations fonctionnelles et inversement. Le mode d'évaluation des résultats des différentes étapes ainsi que leur représentation ont aussi été optimisés.

Nous utiliserons cette base méthodologique comme un point de départ pour approfondir notre travail de recherche sur le corpus MARSEC. L'annotation fonctionnelle en IF ne sera pas effectuée manuellement dans ces expériences mais déduite automatiquement d'une conversion des TSM en IF. Il sera donc possible de traiter un nombre beaucoup plus important de données dans le cadre de l'analyse par synthèse. En effet, on dispose de 47 minutes de parole continue dans ce corpus.

À travers l'étude de ce corpus, nous approfondirons le travail d'optimisation des représentations formelles et fonctionnelles. Nous présenterons d'abord dans la section suivante, les différentes questions expérimentales auxquelles nous essaierons d'apporter des réponses en utilisant l'analyse par synthèse sur le corpus MARSEC. Le développement de l'annotation fonctionnelle ainsi que les différentes étapes d'optimisation des représentations formelles seront ensuite décrits avant de s'étendre sur les différents résultats. Enfin dans une dernière partie, de nouvelles tentatives d'extraction des informations fonctionnelles à partir des représentations formelles seront présentées.

8.2 Hypothèses et questions de départ

Les questions expérimentales sont de deux sortes : les questions concernant la façon d'annoter les fonctions prosodiques et celles qui concernent la représentation ou annotation des formes prosodiques.

8.2.1 Les questions liées aux fonctions prosodiques

Comme nous l'avons déjà vu dans les sections précédentes, l'annotation fonctionnelle de base sera obtenue automatiquement par une conversion des TSM en IF. Celle-ci sera présentée en détail dans les sections suivantes. Nous chercherons à évaluer plusieurs types d'annotation fonctionnelle à partir de l'annotation en TSM. Le développement de ces différents types

d'annotation fonctionnelle est essentiellement motivé par la recherche du type d'annotation qui permettra de prédire au mieux les variations de F0. On peut alors se poser les questions suivantes :

- Combien de niveaux de proéminence peut-on distinguer ou sont importants pour prédire les variations essentielles des contours de F0 ?
- Quelle unité prosodique de base devrait-on utiliser pour modéliser les variations de F0 ?
- Le nucléus est-il nécessaire ? Comment le définir : un maximum de proéminence au niveau de l'intonation ? le dernier accent mélodique ?

Trois types d'annotation fonctionnelle sont utilisés pour tenter d'apporter des réponses à ces questions.

Tout d'abord en se basant sur une interprétation stricte de l'école britannique de l'intonation, on peut faire les hypothèses suivantes :

- On présuppose qu'il existe au moins cinq niveaux de proéminence principaux : un niveau non proéminent ou non accentué noté « U », un accent de nature essentiellement rythmique noté « S » pour « stressed », un accent de nature mélodique noté « A » pour « accented », un nucléus non marqué noté « N » et enfin nucléus emphatique noté « !N » ;
- L'unité prosodique de base au niveau de laquelle les fonctions sont annotées est l'unité tonale ;
- Le nucléus se définit comme le dernier accent mélodique d'un énoncé ;
- Les accents postnucléaires se distinguent des accents prénucléaires.

Si ce modèle est satisfaisant (ou correspond bien à des variations significatives de F0), on pourra distinguer ces différents types de proéminence au niveau de la modélisation des formes. On s'attend à trouver des contours dynamiques dans le cas des TU « A » et « N » et des contours quasiment statiques pour les TU « S ». Les contours associés aux accents prénucléaires seront typiquement distincts des contours associés aux accents postnucléaires.

Nous avons aussi mis en place deux autres types d'annotation adaptés à partir du premier modèle et afin de proposer d'autres réponses alternatives à nos questions initiales.

Dans les deux types d'annotation suivants, on réduit le nombre de niveaux de proéminence à quatre : « U, A, N, !N ». On présuppose que la distinction « A » et « S » est ambiguë et pas nécessaire, on note seulement les proéminences mélodiques et les unités rythmiques sont intégrées aux précédentes.

L'unité prosodique de base est l'unité tonale élargie (que l'on a appelé TTU[8]). On s'attend à pouvoir prédire les variations de F0 à partir des accents mélodiques essentiellement.

Dans le premier modèle, le nucléus est à nouveau considéré comme le dernier accent mélodique et se réalise sur une unité nucléaire élargie comprenant tous les accents rythmiques postnucléaires (c'est l'unité prosodique appelée TTU1 inspirée des travaux de Grice 1995).

Dans le second modèle, on considère qu'il peut y avoir un accent mélodique dans les énoncés non terminaux au moins. Le nucléus peut alors prendre une définition plus large et représenter un maximum de proéminence dans le cas des chutes-montées divisées.

8.2.2 Les questions liées aux représentations formelles

Tout d'abord nous nous interrogerons sur le codage et le calcul des tons INTSINT. Peut-on proposer des variantes ou alternatives au codage des tons INTSINT à l'image des travaux d'Estelle Campione (2000) ? Ces variantes permettront-elles d'améliorer la modélisation des contours de F0 ? Deux variantes du codage INTSINT seront introduites et testées.

Ensuite, nous nous étendrons tout particulièrement sur la question de l'alignement tonal.

La question de l'alignement tonal a été abordée dans le premier chapitre de cette thèse. Les travaux de recherche entrepris depuis les années 80 ont permis de rendre compte de la nature des phénomènes d'alignement, des contraintes et facteurs influençant l'alignement des points cibles mais aussi de mettre en avant des points d'ancrage ou points de repère principaux. Nous utiliserons à nouveau l'analyse par synthèse pour tester certaines des hypothèses utilisées pour expliquer l'alignement des points cibles en définissant plusieurs modèles d'alignement. Nous nous poserons les questions suivantes :

- Quels sont les points d'ancrage principaux à prendre en compte pour modéliser les phénomènes d'alignement tonal ?
- Comment les cibles tonales s'alignent-elles avec le matériau segmental ? L'alignement doit-il être exprimé en fonction de la durée de l'unité prise en compte (en pourcentage) ou en valeur absolue c'est-à-dire en terme de distance fixe par rapport à un point de repère (Pierrehumbert et Steele 1989) ?
- Peut-on distinguer l'alignement tonal au niveau des accents prénucléaires et postnucléaires ? Ou de façon plus générale, la position d'une unité tonale dans une unité intonative a-t-elle un effet sur la façon dont s'alignent les tons qui lui sont associés ?

Dans la tradition la plus ancienne, on associe les tons aux syllabes accentuées (au début de celles-ci).

[8] L'abréviation TTU signifie « Total Tonal Unit » et nous est inspirée de la description des unités rythmiques étroites et larges (« RU » et « TRU ») dans Hirst (1998).

Les expériences mettent aussi en valeur le rôle du pied métrique puisqu'on observe que certains points et surtout le pic mélodique peuvent être alignés en dehors de la syllabe accentuée : « le pic est une propriété du pied métrique attiré par la syllabe forte » (cf. Pierrehumbert et Beckman 1988).

Dans plusieurs travaux, « H* » est aligné proportionnellement à la durée de la syllabe accentuée ou à celle du pied. Les travaux de Pierrehumbert montrent que l'on a le même mode d'alignement dans les accents prénucléaires et postnucléaires. L'attaque de la rime est aussi considérée comme un bon point de repère pour l'alignement du pic.

Dans le cas d'un contour montant de type « LH », c'est l'attaque de la syllabe accentuée qui est définie comme un point d'ancrage.

Enfin, Bruce (1990) parle de l'alignement du contour en entier : pour un contour du type « HLH », L est aligné par rapport à H1 et H2 par rapport à la fin du pied. Quand on a affaire à une chute + montée divisée certains travaux montrent que L2 est aligné en pourcentage de la durée totale du pied et par rapport à la montée finale.

Les points d'ancrage testés seront donc les suivants : la syllabe accentuée, l'unité tonale, la voyelle accentuée, la distance par rapport à la cible tonale précédente, les frontières droites et gauches de l'unité tonale.

8.3 Annotations fonctionnelles

8.3.1 Conversion des TSM en IF : vers un IF « British School »

Comme nous l'avons déjà vu dans la première partie de cette thèse, les TSM représentent à la fois des formes prosodiques (les configurations de F0) et des fonctions prosodiques (proéminences et frontières). Il s'agit alors d'extraire l'information fonctionnelle des TSM et de trouver des règles de correspondance pour coder cette information à l'aide du système IF.

Le système IF permet de distinguer deux types de frontières : les frontières terminales ([|) et les frontières non terminales ([+). Il permet aussi de coder quatre niveaux de proéminence : inaccentué, accentué, nucléaire et nucléaire emphatique.

Les TSM permettent aussi de distinguer deux types de frontières : les frontières intonatives majeures (||) et les frontières intonatives mineures (|).

Peter Roach (1994) résume la signification de ces termes en notant que les frontières majeures sont souvent accompagnées d'une pause remarquable alors que dans le cas des frontières mineures, il peut s'agir uniquement d'une rupture de type rythmique ou mélodique sans pause silencieuse. Dans Williams et Alderson (1986), on définit une frontière majeure comme correspondant le plus souvent à une phrase complète alors que les pauses mineures peuvent se

trouver à d'autres points syntaxiques inférieurs, c'est-à-dire entre deux syntagmes par exemple. Dans une première approximation automatique, nous noterons donc les frontières mineures comme des frontières non terminales et les frontières majeures comme des frontières terminales.

	TSM	IF
Minor intonation unit	|	[+
Major intonation unit.	||	[|

Tableau 25 : Annotation des frontières en TSM et en IF.

En ce qui concerne les niveaux de proéminence, trois questions principales se posent : combien de niveaux de proéminence doivent être pris en compte au niveau prénucléaire ? comment identifier le nucléus à partir des TSM ? qu'en est-il de la marque de l'emphase ?

Une particularité de l'approche britannique de l'intonation, qui est présente dans les TSM, est le fait de distinguer les accents rythmiques (rendus proéminents par des phénomènes d'ordre rythmique) des accents mélodiques (rendus proéminents par la mélodie essentiellement). Les TSM annotent donc des syllabes inaccentuées, accentuées au niveau rythmique (« * ») et accentuées au niveau mélodique (les chutes, montées, tons statiques hauts, etc.).

Nous testerons donc dans un premier temps, une adaptation du système IF à l'approche britannique en introduisant cette distinction entre les accents rythmiques notés « S » pour « stressed » [9] et les accents mélodiques notés « A » pour « accent ».

Le nucléus, présent dans le système IF, sera identifié à partir du dernier accent mélodique d'une unité intonative dans les TSM. En effet dans la notation TSM, la distinction entre les accents nucléaires et prénucléaires n'est pas codifiée, mais on peut supposer que dans la plupart des cas, le dernier accent mélodique annoté correspondra au nucléus dans l'approche IF (cf. Alderson 1986).

En ce qui concerne la fonction emphatique, on distingue les contours mélodiques dits « grands » des « petits » tels que la grande chute et la petite chute dans les TSM. Phonétiquement, cette distinction est marquée par l'attaque du contour à une hauteur supérieure à la hauteur du contour associé à la syllabe précédente. C'est cet écart mélodique qui fait percevoir une emphase. Ainsi quand le dernier accent mélodique est de type « grand », on le traduira comme un nucléus emphatique « !N ». Nous marquerons aussi un contour « !N » quand le dernier accent mélodique est précédé du signe « > » qui indique un élargissement conséquent du registre en ce point.

[9] S stressed but unaccented (accentué rythmiquement pas mélodiquement)

Les règles de correspondance entre les TSM et IF sont résumées dans le tableau ci-dessous.

TSM	IF
* (stressed but unaccented)	S
Low fall (‘), low rise (,), low fall-rise (\’), low level (_), high level (~)	A ou N (si c’est le dernier accent mélodique)
High fall(\), high rise (/), high fall-rise (‘/), >+ pitch accent	A ou !N (si c’est le dernier accent mélodique)
Minor boundary (|)	[+
Major boundary (||)	[|

Tableau 26 : Équivalences entre les TSM et IF.

Exemple tiré du fichier G0101B :

/`going to the \,shore | on the `first *morning of the `/holiday || # TSM

[{A }{ N } +] [{ U }{A }{ S }{ !N } |] IF

Où {} représentent les frontières d’unités tonales. La première unité intonative comporte donc deux syllabes accentuées mélodiquement (en TSM on a un « rise-fall » et un « fall-rise ») qui se traduisent par un TU du type « A » suivi d’un nucléus « N ». Dans la seconde unité intonative les syllabes appartenant au « prehead » prennent l’étiquette « U » et après le premier accent mélodique on trouve un accent rythmique noté « S » suivi d’un « high fall-rise » en TSM ce qui correspond fonctionnellement à un nucléus emphatique.

Toutes ces règles sont ensuite résumées dans un script Perl permettant de créer une nouvelle rangée « IF » pour chaque fichier son du corpus et qui correspond donc à la conversion à partir de la rangée « TSM » comme le montre la figure ci-dessous :

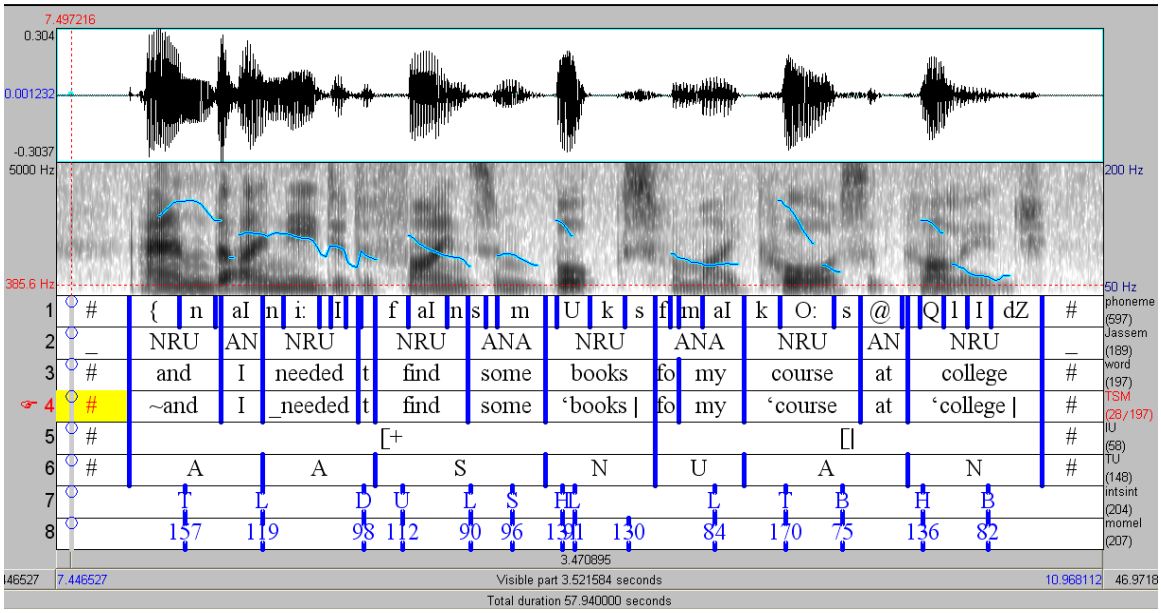


Figure 56 : TextGrid G0301B montrant les différentes rangées d'annotation .

> Rangée 1= phonèmes, rangée 2= découpage en unités ryhtmiques selon Jassem, rangée 3= mots, rangée 4= transcription de la prosodie avec les TSM, rangée 5= annotation des types d'IU obtenue à partir de la conversion des TSM en IF, rangée 6= annotation des types de TU déduite des TSM, rangée 7= tons INTSINT, rangée 8= points cibles MOMEL.

Les unités tonales sont aussi classées en fonction de leur position dans l'unité intonative et en fonction de la position du nucléus. On distingue les positions 0 (un début inaccentué), 1 (premier TU de l'IU), M (les TU intermédiaires), F (TU en position finale). On distingue aussi les TU prénucléaires des TU postnucléaires en ajoutant la lettre T (pour « tail ») pour ces derniers. Un accent postnucléaire est noté « STF ». Dans ce premier modèle fonctionnel, on adopte l'approche britannique traditionnelle, si bien que seuls des accents rythmiques ne peuvent apparaître en position postnucléaire. L'ensemble des TU pris en compte est résumé dans le tableau suivant :

Position nucléus/position du TU dans l'IU	0	1	M	M	F
1	U0	N1 !N1	STM	STM	STF
2	U0	A1 S1	NM !NM	STM	STF
3	U0	A1 S1	AM SM	NM !NM	STF
4	U0	A1 S1	AM SM	AM SM	NF !NF

Tableau 27 : Type de TU en fonction de la position dans l'IU et de la position du nucléus.

L'annotation fonctionnelle complète d'une unité tonale comprend aussi le type d'unité intonative dans laquelle ce TU apparaît : on distingue donc par exemple, les nucléus finaux dans un IU non

terminal annoté « [+NF » des nucléus finaux dans un IU terminal annoté « [|NF ». Il en est de même pour tous les types de proéminence.

Position du TU/ type d'IU	Initial (1)	Median (M)	Final (F)	Initial et final (1F)
[+	[+(!)N1	[+(!)NM	[+(!)NF	[+(!)N1F
[|	[|(!)N1	[|(!)NM	[|(!)NF	[|(!)N1F

Tableau 28 : Exemples de l'annotation fonctionnelle complète pour un nucléus.

8.3.1.1 Développement des types d'unités intonatives

Nous avons aussi enrichi l'annotation fonctionnelle au niveau des IU à partir de ce corpus.

Le modèle de synthèse de départ permet de distinguer les unités intonatives suivantes :

{ [, +, |, +, [|, [+, [!+, [!|}.

À partir du corpus EUROM1, nous avons essentiellement utilisé les IU {[, [| et [+}.

On trouve dans ce second corpus beaucoup de contours emphatiques du type « [!| » ou « [!+ » s'accompagnant souvent des contours suivants après une première observation sur la rangée INTSINT d'origine :

Pour « [!| »**:**

- Falling head + falling N = {BUBUB} ou {TLHL}
- Upstepping head + high fall = {BUHL} / high preH + same = {UHBTB} (G0503G, temps 49,26)
- High head + low fall = {MHUDL}

Ou pour « [!+ » :

- Falling head + fall rise = {MTLHLUBU}
- High head + fall rise = {MHUHBU}
- Divided fall rise : {HBHT/TLT} (G0503B, temps 47,84)

  {MDDHBBU} (G0503G, temps 25,33)
- High head + divided fall rise = {TTTTDST}

Le corpus comporte de nombreux cas de parenthèses basses du type « *said the little girl* » qui permettront de mieux optimiser les IU « + » et « | ».

Les énoncés interrogatifs aussi sont ajoutés avec des étiquettes distinctes : les questions terminales « [?| » et non terminales « [?+ », qui correspondent au niveau formel aux questions se terminant par un contour descendant par opposition aux questions se terminant par un contour montant.

Exemples :

G0501B : *Why are you making bread?* *vs* G0501G : *What is it*?

TLT D L U B S B T

[?| [?+

8.3.1.2 Corrections manuelles et problèmes de conversion

Le résultat de la conversion automatique a été corrigé manuellement par la suite. Nous avons donc réécouté le corpus en entier tout en observant l'annotation automatique en IF dérivée des TSM. Les règles de correspondance permettent en effet une première approximation, mais ne sont pas toujours suffisantes. Nous avons donc effectué un ensemble de corrections au niveau des IU et des TU que l'on peut observer dans les tableaux récapitulatifs suivants :

Annotation IF après correction	Annotation automatique à partir de la TSM			
	[|	[+	(blank)	Total
[|	441	362	31	834
[+	51	493	15	559
[!+	25	71	3	99
[!|	57	32	6	95
|	28	17		45
+	28	13	2	43
[	6	32	1	39
[?+	14	11	1	26
[?|	5	8		13
(blank)				
Total	655	1039	59	1753

Tableau 29 : Matrice de confusion mettant en valeur les corrections manuelles effectuées au niveau des unités intonatives.

Le tableau ci-dessous est une matrice de confusion comparant les types d'IU automatiquement extraits des TSM et les types d'IU finaux qui ont fait l'objet d'une correction. On a donc complété l'étiquette fonctionnelle de certains IU en rajoutant une pré-signalisation de l'emphase ou en précisant le fait qu'il s'agit d'une question ou d'une parenthèse basse etc.

Nous avons surtout effectué beaucoup de corrections au niveau de la conversion des frontières mineures et majeures en frontières terminales et non terminales. On voit par exemple dans la

matrice ci-dessus que pour 362 IU, nous avons noté une frontière terminale au lieu d'une frontière non terminale à partir de la conversion automatique.

Le premier problème se pose ainsi avec l'annotation des types de frontières dans chaque système. La plupart des frontières majeures correspondent à une frontière de type terminal dans le système IF, mais nous avons pu constater beaucoup plus de différences quand il s'agissait de faire correspondre les frontières mineures et non terminales.

Les fichiers G (« fiction ») sont des lectures par des acteurs confirmés marquant des pauses claires et remarquables ce qui est la définition d'une pause majeure d'après Roach (1994). Or toutes ces pauses ne sont pas forcément majeures dans la transcription. Il semble que les transcripteurs indiquent par ces deux signes le lien plus ou moins fort existant entre plusieurs énoncés et que cela correspond en beaucoup d'endroits à une distinction terminale/non terminale. Les TSM marquant les frontières paraissent être utilisées soit pour marquer l'appartenance d'un énoncé à une unité syntaxique précédente soit pour marquer la perception d'une continuité ou d'un aspect terminal.

Prenons un exemple dans le fichier G0102B (temps 15,69 à 17,73) :

TSM : | <*She *looked im'patient* | <*Then smiled* ||

La conversion automatique TSM-IF donne une frontière « [+ » après « impatient », or à l'écoute, nous pensons qu'il s'agit plutôt d'une frontière terminale.

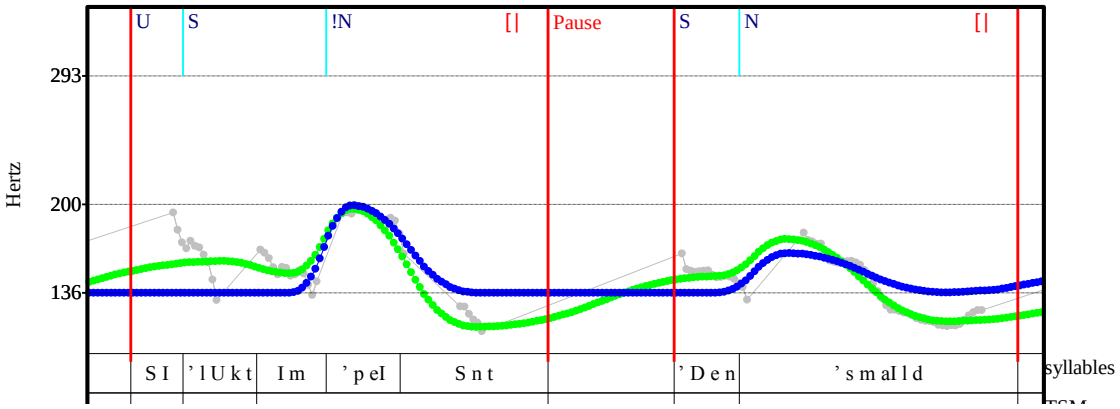


Figure 57 : Courbe de MOMEL et annotation fonctionnelle pour un passage du fichier G0102B.

Rangée 1= syllabes, rangée 2= TSM, rangée 3= annotation des IU, rangée 4= annotation des TU, rangée 5= tons INTSINT

Il semble que l'on puisse percevoir ici une intonation terminale au niveau du premier énoncé. Le second énoncé est comme rajouté après une légère pause.

Au niveau des TSM, on note une frontière mineure puis une frontière majeure : en effet, les deux énoncés sont liés syntaxiquement, on aurait probablement pu les produire en un seul énoncé sans marquer de pause. La TSM reflète peut-être l'interprétation syntaxique de l'énoncé et non pas sa réalisation prosodique dans ce cas. La conversion des frontières mineures en frontières de type non terminal demande donc à être réévaluée systématiquement à ce stade.

Au niveau de l'interprétation des niveaux de proéminence, nous avons aussi effectué un certain nombre de corrections que l'on peut estimer dans le tableau suivant :

Annotation IF après correction	Annotation automatique à partir de la TSM					
	S	A	N	!N	U	Total
S	**1349**	54	64	27		1494
A	55	**1129**	12	23		1219
N	24	34	**568**	206		832
!N	7	53	282	**458**		800
U					**1054**	1054
Total	1435	1270	926	714	1054	5399

Tableau 30 : Matrice de confusion comparant la conversion automatique des TU et la version corrigée.

Ici, le plus grand nombre de changements a été effectué au niveau de l'annotation des nucléus emphatiques. La conversion automatique des grands contours comme nucléus emphatique a fait l'objet d'une vérification systématique. On remarque surtout que les deux transcripteurs adoptent des stratégies différentes dans ces cas : Williams a une approche plus phonétique et note systématiquement un contour de type « grand » dès lors que l'on observe un écart entre l'attaque de l'accent noté et la syllabe précédente même si l'emphase est à peine perceptible dans certains cas. Knowles semble adopter une approche plus fonctionnelle qui correspond exactement à l'interprétation de l'emphase dans le système IF.

L'identification systématique du nucléus à partir du dernier accent mélodique a aussi créé des ambiguïtés en particulier dans le cas des chutes+montées.

Dans un premier cas de figure, la conversion automatique TSM-IF donne une interprétation qui nous paraît satisfaisante, mais pose la question du statut de la dernière unité tonale comme dans l'exemple suivant issu du fichier G0201B :

« Once upon a time there was a little girl, <u>who didn't like going to school. »</u>

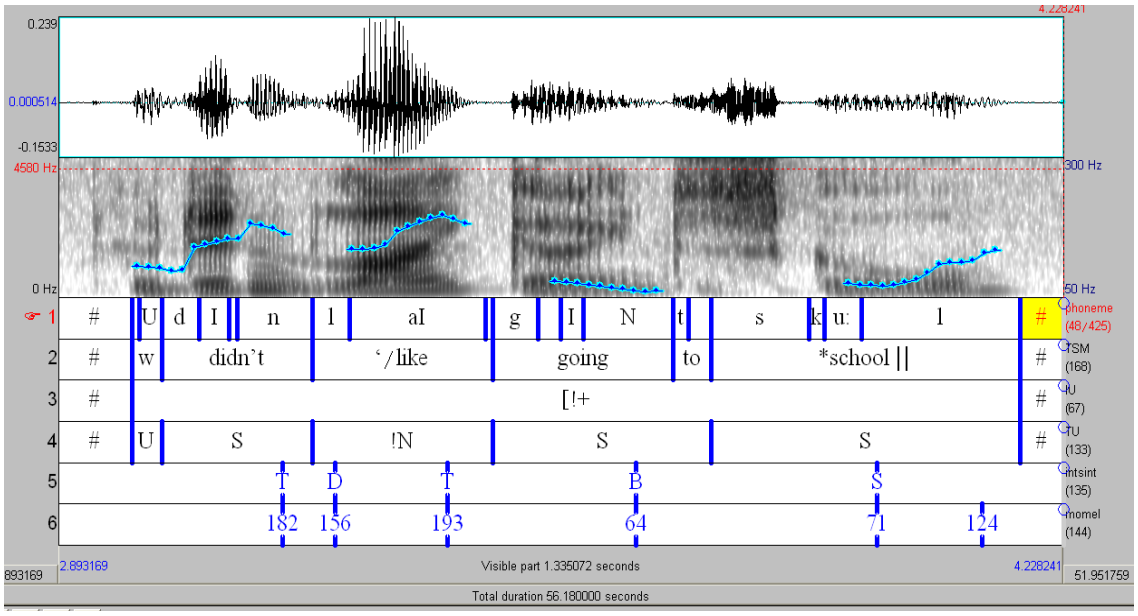


Figure 58 : Exemple de conversion automatique des TSM en IF pour un passage extrait du fichier G0201B.

> Rangée 1= phonèmes, rangée 2= TSM, rangée 3= IU, rangée 4= TU, rangée 5= INTSINT, rangée 6= MOMEL.

Dans l'exemple ci-dessus, on peut observer un contour de chute+montée. Le dernier accent mélodique de l'énoncé est noté au niveau du mot « like » qui est alors interprété comme le nucléus de l'énoncé. À l'écoute, on perçoit bien une proéminence nettement supérieure au niveau du mot « like » qui prend une valeur contrastive. On peut tout de même s'interroger sur la nature de la dernière unité tonale marquée par un contour montant.

Dans un second cas de figure plus problématique, on peut analyser un énoncé marqué par un contour de chute+ montée de plusieurs façons.

Exemple tiré du fichier G0504G : (temps 17,90 à 19,38) :

*And *drying his 'hands I sup,pose||*

Une conversion automatique placerait un accent mélodique au niveau du mot « *hands* » et un nucléus commençant à partir de la seconde syllabe du mot « *suppose* » alors que l'on perçoit nettement une proéminence plus importance au niveau de « *hands* » et que le syntagme « *I suppose* » a ici plutôt une valeur de commentaire, ou de parenthèse. C'est un cas complexe

analysé de façons multiples dans la littérature. On peut proposer diverses annotations fonctionnelles pour ce cas :

« And drying his hands I suppose »

[{U} {S }{ !N }{ S ou A} +

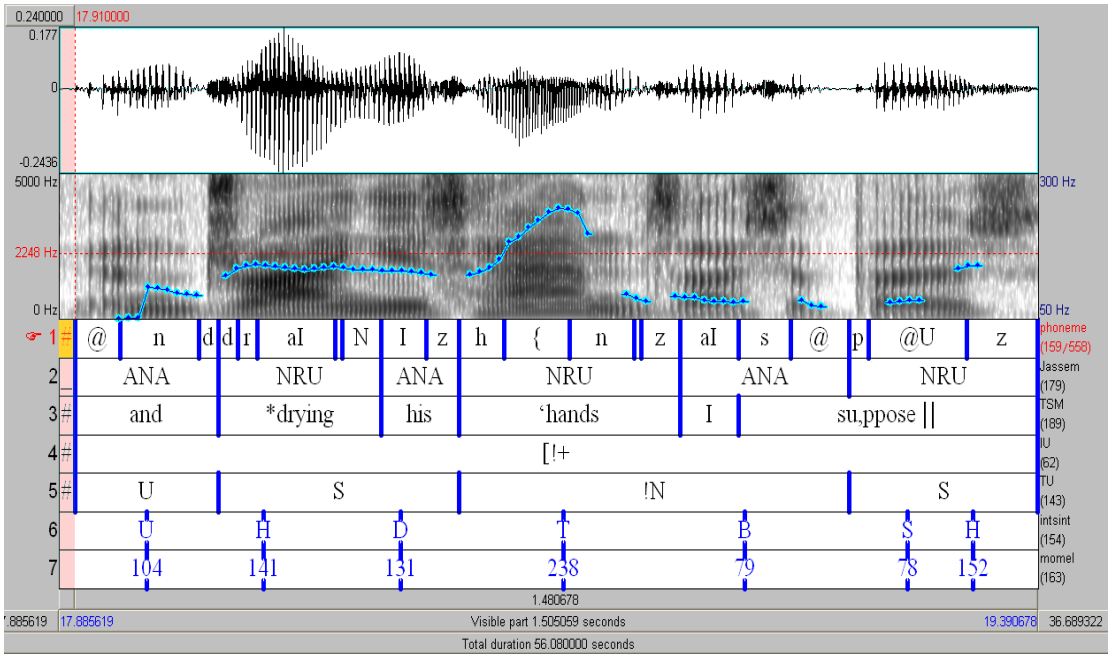


Figure 59 : Exemple de conversion automatique des TSM en IF pour un passage extrait du fichier G0504G.

> Rangée 1= phonèmes, rangée 2= unités ryhtmiques selon Jassem, rangée 3= TSM, rangée 4= IU, rangée 5= TU, rangée 6= INTSINT, rangée 7= MOMEL.

On peut aussi considérer que l'on a affaire à deux unités intonatives distinctes fortement liées (voir structures récursives) :

« And drying his hands I suppose »

[{U} {S }{ !N}|{U}{A} +

Enfin, dans l'ensemble, l'annotation fonctionnelle de départ pourrait faire l'objet de discussions entre plusieurs annotateurs afin d'essayer de choisir l'interprétation la plus proche du message transmis dans les données et de redéfinir des règles de conversion automatique plus précises à l'issue de l'étude de plusieurs cas ambigus.

De façon à comparer notre propre façon d'annoter les différents types de proéminence avec l'annotation fournie par les TSM, nous avons ainsi réalisé une annotation manuelle des fichiers du locuteur G04 dont nous présenterons brièvement les résultats dans la section suivante.

8.3.1.3 Test d'annotation sur le locuteur G04

Les fichiers lus par le locuteur G04 ne sont dans un premier temps pas convertis en IF. Ils sont utilisés pour un test d'annotation manuelle, afin de comparer le résultat de la conversion des TSM en IF avec une annotation manuelle directe. On prend cette occasion pour mettre en application la méthode de Wightman (2002) : « label what you hear ». Une rangée est automatiquement créée correspondant au découpage en syllabes seulement. À partir de ce découpage, on notera les niveaux de proéminence en ayant accès à la ligne de transcription orthographique uniquement sans aucune indication formelle. À partir de cette annotation brute, les TU sont reconstitués puis comparés aux résultats de la conversion des TSM en IF. Les résultats de cette comparaison sont résumés dans la matrice de confusion suivante :

Annotation manuelle	Annotation automatique à partir de la TSM						
	!N	**A**	**N**	**S**	**U**	**(vide)**	**Total**
!N	**30**	10	39	1		1	81
A	26	**56**	8	25	1	8	124
N	56	3	**29**	7		3	98
S	3	21	8	**137**	9	16	194
U		1		4	**113**	38	156
(vide)							
Total	116	91	84	174	123	66	654

Tableau 31 : Matrice de confusion pour les TU annotés manuellement et la conversion directe des TSM.

Tout d'abord on observe que l'étiquette nucléus est identifiée avec un bon taux d'accord entre les deux types d'annotation si l'on ne distingue pas les nucléus dits neutres des nucléus dits emphatiques (154/179). Le nucléus emphatique n'est pas identifié de la même manière chez les différents annotateurs. L'étiquette « S » représentant un accent essentiellement rythmique obtient un bon taux d'accord (137/194). C'est l'identification d'un accent mélodique prénucléaire qui pose ici le plus de problèmes. Dans la moitié des cas considérés, ce que l'on a noté comme un « A » est noté comme un « S » ou un « N » chez l'autre annotateur. On en revient à la question de savoir combien de niveaux de proéminence sont significatifs au niveau fonctionnel et c'est ce à quoi nous tenterons d'apporter des réponses dans les prochaines étapes expérimentales.

8.3.2 Unité nucléaire et accents mélodiques : un IF « britannique adapté »

Nous avons vu dans le chapitre théorique de ce mémoire et notamment à partir des travaux de Grice (1995), que les modèles britanniques faisaient l'objet de critiques diverses. Nous prendrons en compte deux critiques principales : l'une portant sur la distinction entre les accents rythmiques et mélodiques et l'autre portant sur le domaine d'analyse du nucléus.

La distinction entre les accents rythmiques et mélodiques n'est pas partagée par tous et est considérée comme vague et imprécise. On a vu à partir du test d'annotation mentionné plus haut

que la distinction entre ces deux types d'accent posait des difficultés non négligeables du point de vue de l'annotateur. Nous testerons donc un modèle simplifié inspiré de la tradition britannique de l'intonation, mais qui ne prendra en compte que les proéminences mélodiques. Ce système d'annotation pourra être dérivé automatiquement à partir du premier système fonctionnel en IF présenté dans les parties précédentes.

Seulement les unités tonales codées « A » seront prises en compte. Quand un accent mélodique initial « A1 » est suivi d'un accent rythmique « SM », on ne retient qu'une unité tonale globale commençant au début du TU « A1 » et se terminant à la fin du TU « SM ». Si on trouve un ou plusieurs accents rythmiques en début d'énoncé, ceux-ci sont codés « U0 » où « U » signifie « Unaccented » dans le sens britannique du terme et non pas seulement « Unstressed » comme dans le modèle précédent. Les accents rythmiques initiaux sont donc notés « U0 ». Cette nouvelle unité tonale élargie sera appelée TTU : « Total Tonal Unit ».

De plus, l'étendue de l'unité tonale nucléaire sera développée en adoptant l'idée de Grice (1995) tout d'abord. Celle-ci avance que l'on n'accorde pas assez d'importance au contour mélodique apparaissant au niveau des syllabes postnucléaires. L'interdépendance entre le nucléus et les syllabes de queue lui donne l'idée de définir une unité mélodique cohérente plus large : l'unité nucléaire. Cette unité nucléaire commence à partir de la syllabe nucléaire et englobe toutes les syllabes de queue. À nouveau, cette codification peut être déduite automatiquement à partir du système d'annotation précédent. On aura donc les règles de correspondances suivantes :

TTU « A » = TU « A » ou TU « A » + TU « S »

TTU « N » = TU « N » + TU « STM » et « STF »

TTU « U0 » = TU « U0 » ou TU « U0 » + TU « S1 » + « SM »

Type de TTU1 en fonction de la position dans l'IU	**U**	**A**	**N**
0	U0		
1		A1	(!) N1
M		AM	(!) NM
F			(!) NF

Tableau 32 : Types de TTU1 en fonction de la position dans l'IU.

Ce premier modèle britannique adapté sera noté « TTU1 ».

Enfin, un deuxième modèle britannique adapté sera testé. Celui-ci est identique au premier, à l'exception du traitement des syllabes postnucléaires. On intègre cette fois-ci la possibilité d'avoir un accent mélodique après le nucléus dans un cas précis : le cas de la chute+montée. La fonction ajoutée est notée « ATF » ou plus précisément : « [+ ATF ». Dans un premier temps seuls les énoncés non terminaux comportent cette étiquette supplémentaire.

On l'extrait automatiquement de l'annotation du premier modèle, car « ATF » = « [+STF ». Dans ce dernier modèle, la définition du nucléus est donc modifiée : on fait l'hypothèse qu'il s'agit de l'accent le plus proéminent de l'unité intonative plutôt que le dernier accent mélodique de l'unité. C'est le modèle « TTU2 ».

Type de TTU2 en fonction de la position dans l'IU	**U**	**A**	**N**
0	U0		
1		A1	(!) N1
M		AM	(!) NM
F		[+ATF	(!) NF

Tableau 33 : Types de TTU2 en fonction de la position dans l'IU.

Nous testerons donc trois unités prosodiques codées de façon abrégée dans les tableaux de résultats : « TU, TTU1, TTU2 ». Nous avons donc fait un script permettant de générer deux rangées d'annotations fonctionnelles à partir de la première, en appliquant les règles mentionnées plus haut et ceci sur l'ensemble du corpus.

Ci-dessous un exemple d'énoncé non terminal dont le nucléus est marqué par un contour de chute + montée pour la phrase « *There was nothing to see, from the windows* ».

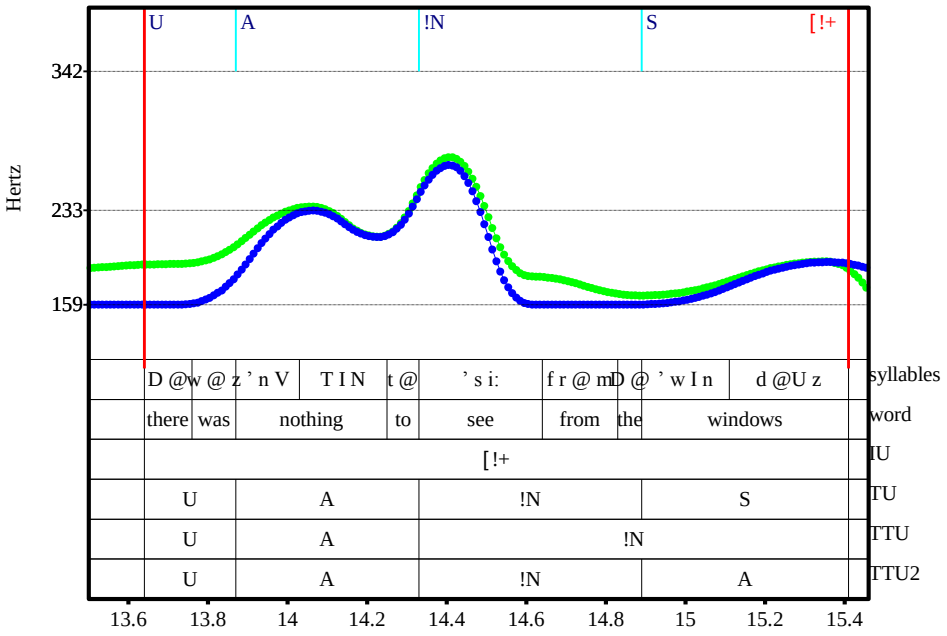


Figure 60 : TextGrid extrait du ficher G0401g montrant les différents modèles fonctionnels et la courbe de MOMEL d'origine (en vert) comparée à la courbe issue du premier modèle fonctionnel (en bleu).

> Rangée 1= syllabes, rangée 2= mots, rangée 3= IU, rangée 4= TU, rangée 5= annotation en TTU1, rangée 6= annotation an TTU2.

Dans l'exemple ci-dessus on a affaire à une chute + montée divisée à partir du nucléus qui est annotée de trois façons différentes. Dans le modèle « TU » le nucléus commence sur « *see* » et son domaine s'étend sur les syllabes inaccentuées suivantes (« *from the* »). Le dernier accent est un accent rythmique du type « STF » au niveau de la première syllabe du mot « *windows* ». Cet accent rythmique s'accompagne d'une montée finale que l'on peut observer sur la courbe de MOMEL (en vert).

Dans le modèle TTU1, l'unité nucléaire commence sur « see » et se prolonge jusqu'à la fin de l'IU. Dans le modèle TTU2, on distingue le dernier accent comme un accent mélodique noté « ATF ».

Dans la section suivante, nous présenterons les différents codages testés au niveau des représentations formelles.

8.4 Annotation des formes

8.4.1 Optimisation du codage INTSINT

Nous nous intéresserons à partir du corpus MARSEC, à la façon de calculer la valeur des tons relatifs dans le système de codage INTSINT. C'est en effet le système d'annotation des formes prosodiques que nous avons décidé d'utiliser pour ce travail de recherche afin d'optimiser et d'étudier les configurations de F0 en fonction du codage fonctionnel. Deux modèles de représentation du codage INTSINT présentant quelques différences par rapport au système d'origine seront présentés. Le premier appelé « INTSINT2 » correspond à un développement du codage INTSINT d'origine en ajoutant de nouveaux tons, le second au contraire est une proposition de simplification de l'utilisation des tons INTSINT sous la forme d'un « INTSINT3 ».

Tout d'abord rappelons que le système de codage INTSINT comprend quatre tons relatifs :

- Le ton {H} qui représente un pic de F0 et une proportion de 50% de l'étendue à partir du point considéré jusqu'au « top » ;
- Le ton {U} qui représente un plateau dans une séquence montante et une proportion de 25% l'étendue à partir du point considéré jusqu'au top ;
- Le ton {L} qui représente un fossé de F0 et une proportion de 50% de l'étendue à partir du point considéré jusqu'au « bottom » ;
- Le ton {D} qui représente un plateau dans une séquence descendante et 25% de l'étendue à partir du point considéré jusqu'au « bottom ».

Les tons {H} et {L} représentent donc en principe des mouvements mélodiques plus conséquents que les tons {U} et {D}. On peut les mettre en relation avec l'opposition entre les grandes et

petites montées et les grandes et petites chutes traditionnellement décrites dans la littérature. Or l'observation du codage INTSINT sur le corpus MARSEC ou à partir des fichiers récapitulatifs des séquences optimales sur le corpus EUROM1, montre que cette opposition scalaire n'est pas toujours respectée.

Les tons {H} et {L} peuvent en fait représenter un mouvement mélodique allant de un à huit demi-tons. Quand il ne reste plus que 30% de l'étendue le placement d'un ton {H} ou {U} correspond en fait à une montée quasiment identique et inversement pour {L} et {D} dans les graves si bien que dans ces cas-là des séquences telles que {MULL} ou {MULD} correspondent en fait au même type de contour. L'opposition entre les tons {L/D} ou {H/U} semble donc être redondante dans ces cas précis. C'est ce qui nous a poussés à tester deux autres façons de calculer la valeur des tons INTSINT.

Dans le premier modèle proposé, on ajoute deux tons supplémentaires {l, h} au codage INTSINT si bien que les tons relatifs sont définis de la façon suivante :

- Le ton {H} représente une proportion de 75% de l'étendue à partir du point considéré jusqu'au «top» ;
- Le ton {h} représente une proportion de 50% de l'étendue à partir du point considéré jusqu'au «top» ;
- Le ton {u} représente une proportion de 25% de l'étendue à partir du point considéré jusqu'au «top» ;
- Le ton {L} représente une proportion de 75% de l'étendue à partir du point considéré jusqu'au «bottom» ;
- Le ton {l} représente une proportion de 50% de l'étendue à partir du point considéré jusqu'au «bottom» ;
- Le ton {d} représente une proportion de 25% de l'étendue à partir du point considéré jusqu'au «bottom».

On a donc six tons relatifs en tout ce qui permet de décrire trois types de montées et parallèlement trois types de chutes (petites, moyennes et grandes). L'utilisation des six tons est toutefois limitée. Prenons l'exemple des contours montants : jusqu'à 60% de l'étendue, les trois tons relatifs {H}, {h} et {u} sont utilisés ; jusqu'à 30% de l'étendue, seuls les tons {h} et {u} peuvent apparaître ; à moins de 30% de l'étendue, seul {u} est utilisé si bien que le ton {H} représente toujours un mouvement mélodique conséquent et plus important que les tons {h} et {u}. La même règle est appliquée de façon symétrique pour les tons {L}, {l} et {d}.

Dans l'optique d'apporter une proposition de réponse au même problème rencontré avec INTSINT, un deuxième modèle, plus simple cette fois-ci, sera testé.

On n'ajoute pas de tons supplémentaires cette fois-ci, mais le calcul des tons relatifs est modifié et limité par des règles :

- Le ton {H} représente une proportion de 60% de l'étendue à partir du point considéré jusqu'au «top» ;
- Le ton {U} représente une proportion de 30% de l'étendue à partir du point considéré jusqu'au «top» ;
- Le ton {L} représente une proportion de 60% de l'étendue à partir du point considéré jusqu'au «bottom» ;
- Le ton {D} représente une proportion de 30% de l'étendue à partir du point considéré jusqu'au «bottom».

Afin d'éviter d'obtenir à nouveau des séquences équivalentes au niveau des extrémités de l'étendue, on ajoute les conditions suivantes : les tons {H} et {U} sont utilisés jusqu'à 30% de l'étendue selon les proportions définies ci-dessus. Cependant à partir de cette limite seul le ton {U} est autorisé et celui-ci prend une valeur de 50% de l'étendue. La même condition est appliquée pour l'opposition entre le ton {L} et le ton {D}.

On peut résumer les règles de codage des trois INTSINT comme il suit :

	INTSINT1	INTSINT3	INTSINT2
<30%	H (50%), U (25%), T, S	U (50%), T, S	U (50%), T, S
>=30%	H (50%), U (25%), T, S	H (60%), U (30%), T, S	H (2/3), u (1/3), T, S
50%	H (50%), U (25%), T, S	H (60%), U (30%), T, S	H (2/3), u (1/3), T, S
>= 60%	H (50%), U (25%), T, S	H (60%), U (30%), T, S	H (75%), h (50%), u(25%), T, S
100%	H (50%), U (25%), T, S	H (60%), U (30%), T, S	H (75%), h (50%), u(25%), T, S

Tableau 34 : Tons possibles en fonction de la proportion de l'étendue et des différents codages d'INTSINT.

La valeur des tons {H}, {U}, {h} et {u} est distincte en fonction du type de codage et de la partie de l'étendue où l'on se trouve. On peut expliciter la comparaison entre les tons {H} et {U} dans les codages INTSINT1 et 2 de la manière suivante:

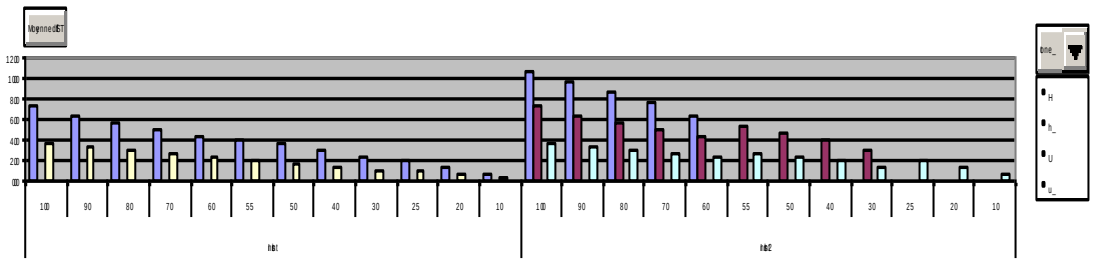


Figure 61 : Histogramme de la valeur des tons H, U, h et en fonction de l'étendue pour INTSINT1 et 2.

Les tons H sont représentés en bleu, les tons U en jaune clair, h en bordeaux et u en bleu clair. Codage avec INTSINT1 à gauche et avec INTSINT2 à droite.

On voit bien que quand il ne reste plus que 30% de l'étendue jusqu'au «top», la différence entre un ton {H} et un ton {U} dans INTSINT est minimale alors que le codage INTSINT2 permet toujours de représenter avec ces deux tons des montées de tailles distinctes.

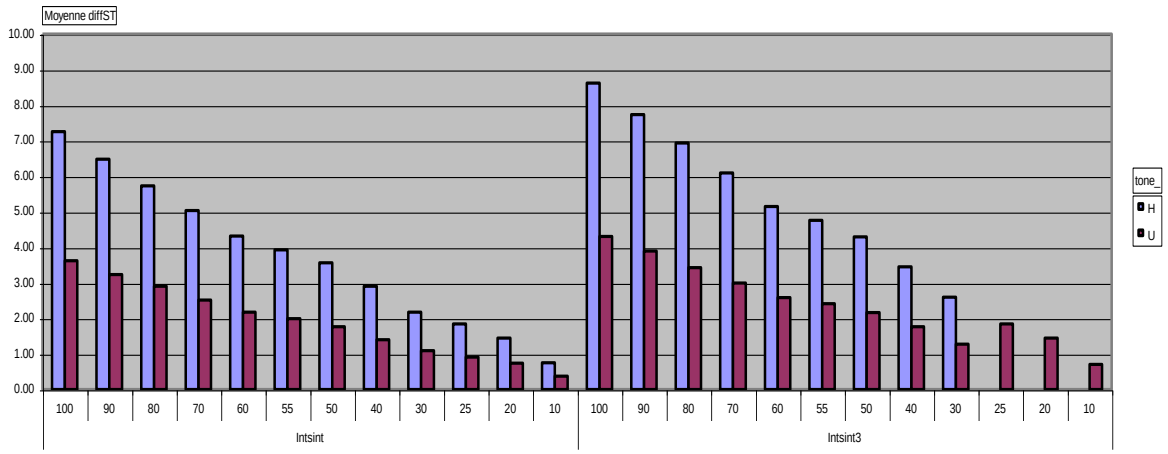


Figure 62 : Histogramme de la valeur des tons H et U en fonction de l'étendue pour INTSINT1 et 3.

Les tons H sont représentés en bleu, les tons U en bordeaux.

A partir de 30% de l'étendue, on ne dispose plus que d'un seul ton {U} dans le système de codage INTSINT3.

Les différents types de montées codés {H}, {U}, {h}, {u} sont représentés de la manière suivante:

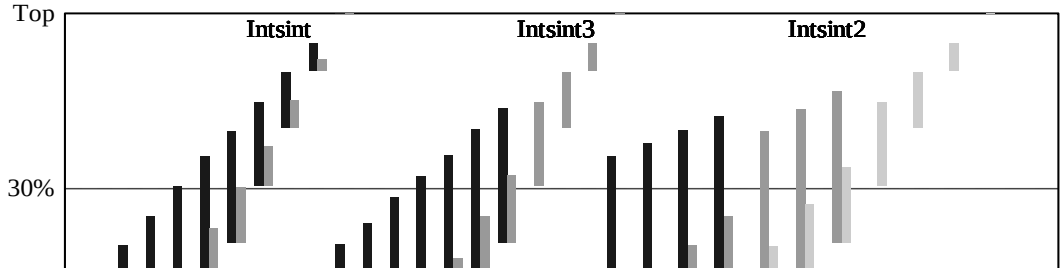


Figure 63 : Représentation des contours montants en fonction du type de codage INTSINT et de la position dans l'étendue.

Nous avons effectué un test préliminaire des deux nouveaux codages INTSINT à partir des enregistrements du locuteur G01. Ceux-ci ont été codés automatiquement avec INTSINT2 et INTSINT3. Les séquences de tons INTSINT2 et 3 ont été converties en points cibles MOMEL pour comparaison avec INTSINT d'origine. L'écart quadratique moyen avec MOMEL d'origine pour ces trois codages a ensuite été calculé comme on peut le voir dans le graphique ci-dessous:

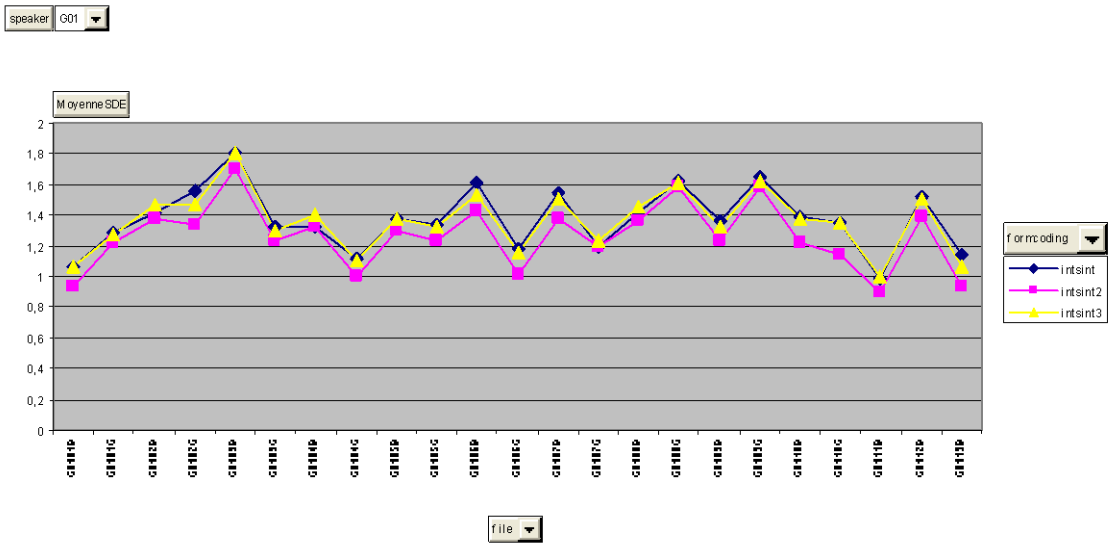


Figure 64 : Écart quadratique moyen entre les courbes d'INTSINT, INTSINT2 et INTSINT 3 et la courbe de MOMEL d'origine.

> Les courbes d'INTSINT sont en noir, INTSINT2 en rose et INTSINT 3 en jaune ..

Dans l'ensemble, l'écart quadratique moyen varie entre 0,9 et 1,8.

Le codage qui permet d'obtenir les valeurs d'écart quadratique moyen les moins élevées, est le codage INTSINT2. Les courbes représentant le résultat des codages INTSINT et INTSINT3 sont très proches.

Les deux nouveaux systèmes de codage INTSINT ont été appliqués à l'extrait du corpus MARSEC « fiction » en entier. Les différents modèles fonctionnels et phases d'optimisation seront donc développés parallèlement à partir du système INTSINT original et d'INTSINT2 et 3 afin de les évaluer par analyse par synthèse.

Afin de tester et mettre en oeuvre les différentes hypothèses et questions dont nous avons discuté dans les premières sections de ce chapitre, toute une série de modèles a été développée. C'est ce qui fera l'objet de la section suivante.

8.5 Description des modèles à tester

Toute une série de modèles a été élaborée en fonction des hypothèses d'alignement à tester mais aussi en fonction du modèle fonctionnel pris en compte (Pauses, IU, TU, TTU1, TTU2) et du codage INTSINT testé (INTSINT, INTSINT1, 2 et 3).

Nous n'avons pas cherché à optimiser les variations globales de F0 à proprement dit dans ce travail. Nous rappelons en effet que l'objet principal de ce travail est d'observer et d'analyser les variations locales de F0. Toutefois nous avons limité dans un premier temps les variations du calcul automatique de l'étendue en testant un plafonnement à 1,1 octave qui correspond à l'étendue standard de la voix parlée puis nous avons effectué une série d'essais avec une étendue maximale de 1,4 octave afin de prendre en compte les variations possibles dans un langage plus expressif[10]. Les modèles codés avec INTSINT d'origine comportent ici une étendue maximale de 1,1 et les modèles codés avec INTSINT2 et 3 ont une étendue maximale de 1,4. Nous avons donc ajouté une série de modèles avec le codage INTSINT d'origine et un plafonnement de l'étendue à 1,4 afin de pouvoir comparer les différents codages. Ce dernier codage est noté « INTSINT1 ».

Nous obtenons alors le tableau récapitulatif suivant :

Unité Prosodique		Codage	Modèles
Pause	Entre 2 pauses	Intsint	1
IU	Unité intonative	Intsint	3
TU	Unité tonale	Intsint	6 à 14
		Intsint 1	50, 52
		Intsint2	40, 42
		Intsint3	62
TTU1	Unité tonale élargie	Intsint	15 à 16
TTU2	Unité tonale élargie avec accent mélodique post nucléaire	Intsint	20 à 24
		Intsint1	56
		Intsint2	46
		Intsint3	66

Tableau 35 : Tableau récapitulatif de tous les modèles utilisés.

[10] Le codage INTSINT d'origine fait varier l'étendue de 0,5 à 2,5 octaves, or c'est une variation beaucoup trop large pour notre corpus. Avec une telle variabilité, nous perdons de la précision dans la représentation des variations locales de petit calibre telles que les petites montées finales ou les chutes+ montées.

Remarque sur la numérotation des modèles : nous ne présentons qu'une partie des modèles que nous avions développés à l'origine dans ce travail, si bien que la numérotation présente des vides (comme le modèle 2, ou les modèles 17, 18, et 19). La série des modèles 50 correspond au codage INTSINT1, la série des 40 à INTSINT 2 et la série des 60 à INTSINT3.

Le modèle 1

C'est le modèle le plus simple, car la seule information fonctionnelle prise en considération est la pause silencieuse.

Les points d'ancrage utilisés sont les frontières droites et gauches de l'unité intonative. Les segments situés entre deux pauses silencieuses sont modélisés au moyen de quatre points cibles :

- Le premier point P1 est aligné à 10 ms du début de l'énoncé entre deux pauses ;
- P2 est aligné à une distance x du début de l'énoncé entre deux pauses ;
- P3 est aligné à une distance x de la fin de l'énoncé (donc de la pause silencieuse droite) ;
- P4 est aligné 10 ms avant la fin de l'énoncé ;

La valeur de x varie selon les paramètres décrits dans le tableau 48 (p. 263) et est spécifique à chaque étape (optimisation des tons, optimisation de l'alignement, création du modèle).

Le modèle 3

Les informations fonctionnelles de base pour ce modèle sont les frontières de l'IU et le type d'IU. Les points d'ancrage utilisés sont les frontières droites et gauches de l'unité intonative. Chaque IU est modélisée au moyen de quatre points cibles selon les paramètres suivants :

- P1 est aligné à 10 ms du début de l'IU ;
- P2 est aligné à une distance x du début de L'IU ;
- P3 est aligné à une distance x de la fin de l'IU ;
- P4 est aligné 10 ms avant la fin de l'IU.

Pour tous les modèles qui suivent, l'annotation fonctionnelle la plus complète est prise en compte. Pour les modèles 6 à 14, on se place dans le modèle fonctionnel des TU, pour les modèles 15 et 16 dans le modèle des TTU1 et pour les modèles allant de 20 à 24 on aborde l'annotation en TTU2. Enfin les derniers modèles testés sont des équivalents de ces derniers avec un codage INTSINT différent (INTSINT1, INTSINT2 et 3).

Remarque : Dans les modèles allant de 6 à 10, le premier point de l'unité intonative fait partie de la première unité tonale que celle-ci soit accentuée ou pas (TU1 ou TU0). Ce premier point peut

prendre les valeurs {M, T, B, _}. (M, T, B après une pause et aucun ton s'il n'y a pas de pause précédente).

Dans tous les autres modèles, le premier point n'est plus inclus dans la première unité tonale et est systématiquement aligné à -10 ms du début de l'IU telle une « note de départ » ou une indication de la « tonalité ».

Le modèle 6

Nous nous plaçons ici dans le premier modèle fonctionnel (TU). Les variations d'alignement sont exprimées en valeurs absolues.

On teste les points d'ancrage suivants : les frontières droites et gauches des TU, la distance par rapport à la cible tonale précédente, le début et la fin de L'IU. On utilise ces points d'ancrage non pas pour tous les TU mais en fonction de la position du TU dans l'IU :

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
0	1	Aligné au début de l'IU			
1	3	Aligné au début du TU	Distance en ms / début du TU	Distance en ms / fin du TU	
M	2	Distance en ms / début du TU	Distance en ms / fin du TU		
F	3	Distance en ms / début du TU	Distance en ms / P1	Aligné à 10 ms de la fin de l'IU	
1F	4	Distance en ms / début de l'IU	Distance en ms / début du TU	Distance en ms / P2	Aligné à 10 ms de la fin de l'IU

Tableau 36 : Mode d'alignement / Point d'ancrage pour le modèle 6.

Le Modèle 7

Par opposition au modèle précédent, les variations d'alignement sont ici exprimées en pourcentage de la durée de l'unité étudiée.

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
0	1	Aligné au début de l'IU			
1	3	Aligné au début du TU	% de la durée du TU	% de la durée du TU	
M	2	% de la durée du TU	% de la durée du TU		
F	3	% de la durée du TU	% de la durée du TU	Aligné à 10 ms de la fin de l'IU	
1F	4	Aligné au début de l'IU	% de la durée du TU	% de la durée du TU	Aligné à 10 ms de la fin de l'IU

Tableau 37 : Mode d'alignement / Point d'ancrage pour le modèle 7.

Le modèle 8

Les variations d'alignement sont exprimées en valeurs absolues.

Les points d'ancrage sont les suivants : le début de l'IU, la fin de l'IU, le début du TU, la fin de la dernière syllabe du TU, l'attaque ou le début de la dernière syllabe du TU. On resserre les points d'ancrage au niveau de la syllabe en prenant la dernière syllabe d'un TU, on s'attend à pouvoir cadrer le dernier mouvement mélodique surtout dans les montées finales ou les chutes+montées.

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
0	1	Aligné au début de l'IU			
1	3	Aligné au début du TU	Distance en ms / début du TU	En nombre de syllabes / fin de la dernière syllabe	
M	2	Distance en ms / début du TU	En nombre de syllabes / fin de la dernière syllabe		
F	3	Distance en ms / début du TU	En nombre de syllabes / début de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	
1F		Aligné au début de l'IU	Distance en ms / début du TU	En nombre de syllabes / début de la dernière syllabe	Aligné à 10 ms de la fin de l'IU

Tableau 38 : Mode d'alignement / Point d'ancrage pour le modèle 8.

Le modèle 9

Ce modèle mélange l'utilisation d'alignements en pourcentages et en valeurs absolues. Les points d'ancrage sont les suivants : le début et la fin de l'IU, la première syllabe du TU, la fin de la dernière syllabe du TU, le début de la dernière syllabe du TU.

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
0	1	Aligné au début de l'IU			
1	3	Aligné au début du TU	% de la durée de la 1ère syllabe	En nombre de syllabes / fin de la dernière syllabe	
M	2	% de la durée de la 1ère syllabe	En nombre de syllabes / fin de la dernière syllabe		
F	3	% de la durée de la 1ère syllabe	En nombre de syllabes / début de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	
1F		Aligné au début de l'IU	% de la durée de la 1ère syllabe	En nombre de syllabes / début de la dernière syllabe	Aligné à 10 ms de la fin de l'IU

Tableau 39 : Mode d'alignement / Point d'ancrage pour le modèle 9.

Le modèle 10

Ce modèle est identique au modèle 9 sauf que l'on cherche à modéliser la pré-tête. Un TU0 est modélisé à l'aide de deux points cibles : le premier est aligné à 10 ms du début de l'IU et le second à une distance x de la fin du TU.

On cherche à reconnaître d'éventuelles pré-têtes spéciales dont on ne pourrait pas prédire le contour à partir du premier TU accentué.

Le modèle 11

En pourcentages et valeurs absolues avec la tonalité positionnée en début de l'IU. Les points d'ancrage testés sont les suivants : le début et la fin de l'IU, la fin du TU, la première syllabe du TU (en %), la dernière syllabe du TU (en %).

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	2	% de la durée de la 1ère syllabe	Distance en ms / fin du TU		
M	2	% de la durée de la 1ère syllabe	Distance en ms / fin du TU		
F	3	% de la durée de la 1ère syllabe	% de la durée de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	
1F	3	% de la durée de la 1ère syllabe	% de la durée de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	

Tableau 40 : Mode d'alignement / Point d'ancrage pour le modèle 11.

Le modèle 12

En pourcentages et valeurs absolues avec la tonalité.

Les points d'ancrage testés sont les suivants : le début et la fin de l'IU, le début et la fin du TU, la dernière syllabe du TU (en%).

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	2	Distance en ms / début du TU	Distance en ms / fin du TU		
M	2	Distance en ms / début du TU	Distance en ms / fin du TU		
F	3	Distance en ms / début du TU	% de la durée de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	
1F	3	Distance en ms / début du TU	% de la durée de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	

Tableau 41 : Mode d'alignement / Point d'ancrage pour le modèle 12.

Le modèle 13

En pourcentages et valeurs absolues avec la tonalité. Ce modèle est équivalent au précédent mais introduit cette fois-ci la voyelle comme point d'ancrage.

Les points d'ancrage testés sont les suivants : la première voyelle du TU, la fin du TU, le début et la fin de l'IU, la dernière syllabe du TU (en %).

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	2	Distance en ms / 1ère voyelle du TU	Distance en ms / fin du TU		
M	2	Distance en ms / 1ère voyelle du TU	Distance en ms / fin du TU		
F	3	Distance en ms / 1ère voyelle du TU	% de la durée de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	
1F	3	Distance en ms / 1ère voyelle du TU	% de la durée de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	

Tableau 42 : Mode d'alignement / Point d'ancrage pour le modèle 13.

Le modèle 14

En valeurs absolues avec la tonalité.

Les points d'ancrage testés sont les suivants : le début et la fin de l'IU, la première voyelle du TU, la fin du TU, l'attaque de la dernière voyelle du TU.

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	2	Distance en ms / 1ère voyelle du TU	Distance en ms / fin du TU		
M	2	Distance en ms / 1ère voyelle du TU	Distance en ms / fin du TU		
F	3	Distance en ms / 1ère voyelle du TU	Distance en ms / attaque dernière voyelle du TU	Aligné à 10 ms de la fin de l'IU	
1F	3	Distance en ms / 1ère voyelle du TU	Distance en ms / attaque dernière voyelle du TU	Aligné à 10 ms de la fin de l'IU	

Tableau 43 : Mode d'alignement / Point d'ancrage pour le modèle 14.

Les résultats de ces optimisations de l'alignement ont été utilisés pour réduire le nombre de configurations dans les modèles suivants. C'est-à-dire pour ceux qui utilisent INTSINT2 et 3 puis les modèles au niveau du TTU1 et TTU2. Nous avons retenu les meilleurs modèles pour les appliquer à ces autres types de modèles.

Le modèle 15

C'est l'équivalent du modèle 6 mais au niveau des TTU1 et avec la tonalité. Les points d'ancrage sont exactement les mêmes que dans le modèle 6. Les unités considérées sont en majorité plus longues, car on regroupe le TU « A » et « S » donc on rajoute un point de plus dans certains TTU.

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	3	Distance en ms / début du TU	Distance en ms / cible tonale précédente P1	Distance en ms / fin du TU	
M	3	Distance en ms / début du TU	Distance en ms / P1	Distance en ms / fin du TU	
F	3	Distance en ms / début du TU	Distance en ms / P1	Aligné à 10 ms de la fin de l'IU	
1F	4	Distance en ms / début du TU	Distance en ms / P1	Distance en ms / fin de l'IU	Aligné à 10 ms de la fin de l'IU

Tableau 44 : Mode d'alignement / Point d'ancrage pour le modèle 15.

Le modèle 16

Au niveau des TTU1, avec la tonalité. Les points d'ancrage testés sont les suivants : le début et la fin de l'IU, la cible tonale précédente, la première syllabe du TTU (en %), la fin de la dernière syllabe du TTU, le début de la dernière syllabe du TTU.

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	3	x% de la durée de la première syllabe du TTU	Distance en ms / cible tonale précédente P1	En nombre de syllabes / fin de la dernière syllabe	
M	2	x% de la durée de la première syllabe du TTU	Distance en ms / P1	En nombre de syllabes / fin de la dernière syllabe	
F	3	x% de la durée de la première syllabe du TTU	En nombre de syllabes / attaque de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	
1F	4	x% de la durée de la première syllabe du TTU	Distance en ms / P1	En nombre de syllabes / attaque de la dernière syllabe	Aligné à 10 ms de la fin de l'IU

Tableau 45 : Mode d'alignement / Point d'ancrage pour le modèle 16.

Le modèle 20

Le niveau d'analyse est le TTU2, c'est l'équivalent parfait du modèle 15.

Le modèle 21

C'est l'équivalent parfait du modèle 16.

Le modèle 22

Au niveau des TTU2, avec tonalité. En pourcentages et valeurs absolues. Les points d'ancrage testés sont les suivants : le début et la fin de l'IU, la première voyelle du TU, la cible tonale précédente, la fin de la dernière syllabe, la dernière syllabe (en %).

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	3	Distance en ms / 1ère voyelle du TU	Distance en ms / cible tonale précédente P1	En nombre de syllabes / fin de la dernière syllabe	
M	3	Distance en ms / 1ère voyelle du TU	Distance en ms / P1	En nombre de syllabes / fin de la dernière syllabe	
F	3	Distance en ms / 1ère voyelle du TU	En % de la dernière syllabe	Aligné à 10 ms de la fin de l'IU	
1F	4	Distance en ms / 1ère voyelle du TU	Distance en ms / P1	En % de la dernière syllabe	Aligné à 10 ms de la fin de l'IU

Tableau 46 : Mode d'alignement / Point d'ancrage pour le modèle 22.

Le modèle 23

Variation à partir des modèles 11 et du 15, niveau TTU2 avec tonalité, pourcentages et valeurs absolues. Les points d'ancrage testés sont les suivants : le début et la fin de l'IU, la cible tonale précédente, la fin du TU, la dernière syllabe du TU en %.

Position du TU	Nb points	Mode d'alignement / Point d'ancrage			
		P1	P2	P3	P4
Tonalité	1	Au début de l'IU			
0	0				
1	3	x% de la durée de la première syllabe du TTU	Distance en ms / cible tonale précédente P1	Distance en ms / fin du TU	
M	2	x% de la première syllabe du TTU	Distance en ms / P1	Distance en ms / fin du TU	
F	3	x% de la première syllabe du TTU	x% de la dernière syllabe du TTU	Aligné à 10 ms de la fin de l'IU	
1F	4	x% de la première syllabe du TTU	Distance en ms / P1	x% de la dernière syllabe du TTU	Aligné à 10 ms de la fin de l'IU

Tableau 47 : Mode d'alignement / Point d'ancrage pour le modèle 23.

Le modèle 24

Le même que le 23 sauf que le TTU « U0 » est modélisé en deux points : P1 étant positionné au début de l'IU et P2 à une distance x de la fin du TTU.

Afin de tester les codages INTSINT1, INTSINT2 et 3, une série de modèles équivalents aux précédents a aussi été créée.

Les modèles codés avec INTSINT1

- Modèle 50 équivalent au modèle 6 avec la tonalité ;
- Modèle 52 équivalent au modèle 11 ;
- Modèle 56 équivalent au modèle 23.

Les modèles codés avec INTSINT2

- Modèle 40 équivalent au modèle 6 mais avec la tonalité ;
- Modèle 42 équivalent au modèle 11 ;
- Modèle 45 équivalent au modèle 20 ;
- Modèle 46 équivalent au modèle 23.

Les modèles utilisant INTSINT3

- Modèle 62 équivalent au modèle 11 ;
- Modèle 66 équivalent au modèle 23.

Après avoir décrit l'ensemble des modèles du point de vue des points d'ancrage et des modes d'alignement, nous présenterons les paramètres de variation de l'alignement pour chaque type de point d'ancrage et en fonction de l'étape d'optimisation.

Les variations de distance sont les mêmes dans tous les modèles, elles sont spécifiques au point d'ancrage considéré.

Les paramètres sont toutefois aussi adaptés à chaque étape d'optimisation. Celles-ci sont au nombre de trois :

1. Optimisation des séquences de tons INTSINT (pour cette étape l'alignement varie de façon très réduite) ;

2. Optimisation large de l'alignement des tons avec les 30 meilleures séquences retenues dans l'étape précédente ;

3. Optimisation étroite ou élaboration du modèle en ne gardant que les dix meilleures séquences de tons et les tranches d'alignement optimisées.

Paramètres d'optimisation de l'alignement en fonction du point d'ancrage et du type d'étape	*Etape 1*	*Etape 2*
Début de l'IU	point aligné exactement au début.	point aligné exactement au début.
Fin de l'IU	un point à 10 ms de la fin.	un point à 10 ms de la fin.
Début du TU	de 0 à 200 ms en 3 itérations.	de 0 à 300 ms en 7 itérations.
Fin du TU	de -100 à 10 ms en 2 itérations.	de -160 à 10 ms en 4 itérations.
Durée du TU à partir du début	de 10 à 40% en 3 itérations.	de 10 à 40% en 7 itérations.
Durée du TU avec tranche médiane	de 10 à 40% en 3 itérations.	de 60 à 75% en 4 itérations.
Durée du TU avec la fin comme point de repère	de 80 à 95 % en 3 itérations.	de 80 à 95 % en 4 itérations.
Cible tonale précédente	de 100 à 200 ms en 2 itérations.	de 50 à 300 ms en 6 itérations.
Durée de la première syllabe du TU	de 0 à 100% en 3 itérations.	de 0 à 150% en 7 itérations.
Durée de la dernière syllabe du TU	de -25 à -75¨% en 3 itérations.	de -40 à -80¨% en 7 itérations.
Début de la dernière syllabe	deux itérations : point aligné soit avec le début de la dernière syllabe, soit avec le début de la syllabe précédente	deux itérations : soit un point aligné avec le début de la dernière syllabe, soit avec le début de la syllabe précédente.
Fin de la dernière syllabe	3 itérations : point soit à la fin de l'avant avant dernière syllabe, soit à la fin de l'avant dernière, soit à la fin de la dernière.	3 itérations : point soit à la fin de l'avant avant dernière syllabe, soit à la fin de l'avant dernière, soit à la fin de la dernière.
Première voyelle du TU	de 0 à 200 en 3 itérations.	de -100 à 300 ms en 9 itérations.
Début de la dernière voyelle du TU	de 0 à 200 ms en 3 itérations	de -100 à 300 ms en 9 itérations

Tableau 48 : Paramètres d'optimisation de l'alignement tonal en fonction du type de point d'ancrage et de l'étape.

Dans les trois sections suivantes, nous décrirons et analyserons les résultats des trois phases d'optimisation des représentations formelles menées à partir des paramètres décrits ci-dessus : d'abord les séquences de tons INTSINT seront optimisées, puis leur alignement et enfin les modèles finaux seront générés.

8.6 Résultats de l'optimisation des représentations formelles modèle par modèle

La première étape d'optimisation des représentations formelles consiste à tester toutes les séquences de tons INTSINT possibles au niveau de chaque modèle basé sur une unité prosodique particulière (la pause, l'unité intonative, l'unité tonale, les unités tonales élargies).

8.6.1 Le modèle des pauses

Les segments situés entre deux pauses silencieuses sont modélisés au moyen de quatre points cibles dont toutes les combinaisons possibles sont testées selon les paramètres suivants (« _ » indique la possibilité de ne pas affecter de ton sur le point considéré).

Points cibles	Tons INTSTINT testés							
P1	M	T	B					
P2	T	B	H	L	U	D	S	_
P3	T	B	H	L	U	D	S	_
P4	T	B	H	L	U	D	S	

Tableau 49 : Paramètres d'optimisation des séquences de tons entre deux pauses.

Les séquences qui obtiennent les meilleures corrélations moyennes sont retenues pour la phase d'optimisation de l'alignement dont les paramètres sont décrits dans la section précédente. Ci-dessous un tableau récapitulatif des 10 meilleures séquences de 4 tons :

Séquences optimisées au niveau du modèle 1	
séquences	score
MULL	3705.9418
MH_B	3695.7981
MU_B	3646.9778
MU_L	3641.9422
MULB	3632.4748
MHLL	3627.9417
MHLB	3589.4657
MT_B	3564.5262
MUDD	3551.7159
MS_L	3544.8696

Tableau 50 : Meilleures séquences de tons INTSINT pour le modèle 1.

La modélisation entre deux pauses et tous types d'énoncé confondus, donne à nouveau un contour commençant par une montée initiale suivie d'une descente progressive à l'image de la composante globale dans le modèle de Fujisaki.

8.6.2 Le modèle des IU

À partir du découpage des différents passages en unités intonatives, les représentations formelles sont optimisées au moyen de quatre points cibles et selon les mêmes paramètres tonals que dans le modèle 1 (voir tableau 49).

L'optimisation se fait en fonction du type d'unité intonative annotée ([+, [|, [!|, [!+, [?+, [?|, [, +, |).

Pour chaque type d'unité intonative, plusieurs séquences optimales sont retenues afin de procéder à l'optimisation de l'alignement des points cibles comme le montrent les résultats suivants :

all	[+	[|	[!+	[!|	[?+	[?|	[	+	|
MT_B,	MT_B,	MHDB,	MTBH,	BHLB,	B_HT,	MHDB,	BHHS,	BBUH,	B__B,
MH_B,	MU_L,	MUDB,	MULU,	MSDB,	BUHT,	MHDL,	BHHD,	B_UH,	B_BB,
MHLB,	MHLL,	BH_B,	MHBH,	M_DB,	B__T,	MUDL,	BHUD,	B_BH,	BB_B,
MUDL,	MHLD,	MUDL,	MHLU,	MSDL,	BH_T,	MUDB,	BHUU,	BBBH,	BBBB,
MTLB,	MH_B,	MHLB,	MUBH,	T_DB,	BUSH,	BTLB,	B_HS,	B_BU,	__BB,
MU_B,	MUDD,	BT_B,	MUBU,	M_DL,	BU_T,	BTTB,	B_TD,	BBBU,	__LS,
MULB,	MTLL,	MULB,	MHBU,	T_LB,	BUHH,	BHLB,	B_HD,	BBUT,	__DS,
MU_L,	MTLB,	MTLB,	MTBU,	BULB,	BUUT,	MH_L,	B_TT,	B_UT,	_LDS,
MHLL,	MH_L,	MHDL,	MSLU,	MSLB,	BHHT,	MTDB,	BHUS,	B_BT,	_L_D,
BH_B,	MHLB,	BHLB,	MSBU,	BTLB,	BHUT,	BTDB,	BUUD,	BBBT,	_L_B,

Tableau 51 : Meilleures séquences de tons retenues par type d'IU pour optimiser l'alignement.

À première vue, les meilleures séquences obtenues correspondent aux prédictions des modèles de synthèse de départ. Seuls les énoncés de type « [+ », c'est-à-dire non terminaux et non emphatiques, n'obtiennent pas la modélisation attendue. Les énoncés terminaux et non terminaux sont modélisés de manière quasiment identique. Quand il s'agit d'énoncés de type emphatique, de questions, ou de parenthèses basses, l'opposition entre les deux types ressort nettement. ({MTBH} vs {BHLB}, {B_HT} vs {MHDB}, {BBUH} vs {B__B}). On se trouve dans une modélisation globale et on s'attend à ce que les montées finales soient mieux représentées dans les modèles plus complexes basés sur les proéminences locales.

8.6.3 Les modèles basés sur les unités tonales

Toutes les combinaisons de tons INTSINT comprises dans les paramètres sont testées au niveau des unités tonales dans les différents modèles existants. On calcule ensuite la corrélation entre la courbe issue de chaque modèle et la courbe MOMEL d'origine et retient les résultats dans un fichier texte (« .txt »). On lance alors un script R afin d'extraire les meilleures séquences de tons. Un nouveau tri de l'ensemble des séquences testées est donc effectué comme décrit dans la section 7.1.1 du chapitre précédent. Seules les séquences qui obtiennent une corrélation minimale

de 0,8[11] sont conservées cette fois-ci. Il en ressort une trentaine de séquences pour chaque type d'unité tonale à l'intérieur d'une unité intonative et cela correspond à une base à utiliser pour passer ensuite à l'étape d'optimisation de l'alignement. Nous nous appuierons dans cette partie, sur quelques modèles, pour illustrer le processus d'optimisation des représentations formelles au niveau des unités tonales et esquisser quelques premières observations. Le résultat de l'optimisation des tons pour les modèles non mentionnés pourra être consulté à partir des annexes.

8.6.3.1 Exemples de séquences optimisées dans une sélection de modèles

Prenons d'abord le modèle 11 pour exemple et la modélisation des accents rythmiques :

[+_S1	score [+_S1	[+_SM	score [+_SM	[+_STM	score [+_STM	[+_STF	score [+_STF
_H	510.4877	_D	1148.1937	_L	120.6811	_BU	263.8044
_T	498.3634	_L	1104.2861	_D	120.4741	_LU	260.1763
UH	493.1585	SD	1063.6678	DL	109.0277	L_U	259.1268
UU	489.6038	SL	1020.6415	_B	106.7799	DLU	253.6374
_U	485.9256	DL	1002.3264	DD	105.9843	LBU	248.1782
DT	477.5049	DD	992.6934	SL	91.5103	B_H	236.604
HT	476.4669	UL	839.3013	SD	91.3108	DDU	236.1388
UT	469.7395	UD	657.8731	UL	87.1329	_LS	233.4482
LT	461.417	HL	652.0975	DB	86.2663	LDU	233.2773
DH	451.843	_B	638.4614	SB	78.6178	D_H	233.0418

Tableau 52 : Extrait du tableau R montrant les résultats de l'optimisation des séquences par TU et IU dans le modèle 11.

Ce tableau montre les meilleures séquences retenues au niveau d'une unité intonative non terminale et par type d'unité tonale. Seuls les différents types d'accents rythmiques apparaissent dans cet extrait. L'étiquette « [+S1 » représente par exemple une unité tonale accentuée rythmiquement en première position dans un IU de type non terminal. Ce type de TU est modélisé par un contour montant du type {_H} par opposition à un accent rythmique en position médiane typiquement représenté par une séquence descendante {_D}.

L'observation des fichiers récapitulatifs R pour chaque modèle permet d'émettre des hypothèses préliminaires quant à la validité des étiquettes fonctionnelles prises en compte. Il s'agit de chercher à voir si l'on peut distinguer les différentes étiquettes fonctionnelles par les meilleures formes avec lesquelles elles sont modélisées. Les TU « A » et « S » ont-ils une forme optimale distincte ? Peut-on distinguer la forme d'un nucléus emphatique de celle d'un nucléus non marqué, etc.?

[11] Avec deux ou trois points au niveau d'une unité tonale, on peut obtenir des courbes modélisées proches de la courbe MOMEL d'origine. 0,8 est un seuil minimum placé de façon empirique qui permet de ne garder que les bonnes séquences.

Si on prend l'exemple des TU « A » et « S » dans le modèle 11, la distinction des représentations varie selon le type d'IU. On observe des formes quasiment identiques quand il s'agit d'une unité intonative terminale comme on peut le voir dans le tableau ci-dessous :

[|_S1	score [|_S1	[|_SM	score [|_SM	[|_A1	score [|_A1	[|_AM
_H	1210.1095	_D	1250.657	_H	1510.2219	_D
_T	1161.6949	_L	1183.7376	HU	1507.9196	_L
UU	1158.1017	SD	1095.0206	_T	1482.4373	UL
UH	1155.7786	DD	1082.1032	HD	1382.6782	SD
_U	1107.3578	SL	1042.7126	_U	1366.9955	HL
HT	1087.9881	DL	1032.0991	UU	1358.5476	DL
UT	1070.4066	_U	898.0115	TD	1333.1093	SL
SH	1048.4107	SU	897.2696	UD	1330.2729	UD
DH	1033.0779	SH	892.8946	HT	1302.3377	DD

Tableau 53 : Les 10 meilleurs séquences par type de TU dans le modèle 11 pour un IU [|.

Pour les TU « S1 » et « A1 », les trois meilleures séquences sont des contours montants similaires de type {_H}. Toutefois la représentation du TU « A1 » se distingue du « S1 » dans le reste des dix meilleures séquences avec une représentation en forme de chute {HD}. On peut faire une observation similaire pour les TU « SM » et « AM » partageant la forme {_D}, {_L} mais se distinguant par un contour de chute {UL} pour les TU « AM » et un contour montant {_U} pour les TU « SM ».

Au niveau des autres types d'IU, la distinction entre les deux types de TU est bien plus évidente. Prenons l'exemple des énoncés terminaux et non terminaux emphatiques :

[!|_S1	score [!|_S1	[!|_SM	score [!|_SM	[!|_A1	score [!|_A1	[!|_AM	score [!|_AM
DT	134.5202	SH	280.8776	UL	415.7472	UL	288.3232
LT	125.3293	_H	277.2762	HL	411.2231	HL	275.3047
ST	123.0642	DT	275.6459	TL	385.9979	UD	257.1237
UH	120.8268	_T	275.0493	UD	363.2569	HB	250.3134
SU	117.8127	ST	268.2516	HD	336.682	UB	235.0274
SH	116.6647	UH	264.233	HB	331.928	TL	233.6707
_T	116.5012	UT	260.6101	TB	323.5736	HD	232.5865
UT	115.4359	SU	259.7273	UB	319.724	TB	227.8193
HT	114.7331	DH	259.6876	TD	266.9949	_L	201.2867
_H	113.4288	_U	254.9773	SL	238.5138	TD	191.383

Tableau 54 : Meilleures séquences pour les TU A et S dans un IU [!| (modèle 11).

Dans les énoncés « [!| » les « S1 » s'optimisent par des montées de type {DT} alors que les accents « A1 » sont typiquement des chutes de type {HL}. La même opposition se trouve en

position médiane. Ces séquences semblent décrire le contour général de suites de chutes + montées suivies par une grande chute finale (falling-rising head + high fall) où les accents rythmiques serviraient de transitions, de montées préparatoires avant les chutes qui ressortent au niveau perceptif. On retrouve aussi des formes distinctes dans les questions terminales par exemple où les « AM » sont systématiquement montants par opposition aux « SM » descendants.

La distinction entre les nucléus emphatiques et non emphatiques se fait comme on s'y attend, par une attaque plus haute pour les premiers tels que les « NM » dans un énoncé non terminal ({DL} *vs* {HL}) :

[+_N1	score [+_N1	[+_!N1	score [+_!N1	[+_NM	score [+_NM	[+_!NM	score [+_!NM
HD	47.9161	UL	173.3078	DL	70.8853	HL	137.8129
DL	42.5845	HL	168.228	_L	70.7208	UL	132.7727
_L	41.6784	TL	164.7645	SL	70.5247	UD	118.7993
_D	41.427	HD	158.6668	_D	70.4221	TL	115.9544
DD	40.5997	UD	150.9102	SD	70.3174	HD	115.6886
SL	39.4944	TD	134.6174	UL	69.289	SD	99.2697
SD	39.2821	HB	110.9126	UD	68.0019	SL	98.5264
TL	38.5958	UB	107.4955	HL	67.9543	_L	89.9581
UD	38.1579	TB	106.7581	DD	67.7176	_D	89.6892

Tableau 55 : Meilleures séquences pour les TU N et !N dans un IU [+ (modèle 11).

Les mêmes observations générales peuvent être faites à partir des autres modèles d'annotation fonctionnelle (TTU1 et TTU2). Nous prendrons les exemples des modèles 16, 23 et 24.

Dans le modèle 16 et au niveau d'un IU terminal emphatique, on retrouve le même type de contour de tête que dans le modèle 11 compris cette fois-ci dans une unité mélodique globale puisque le TTU « A » comprend aussi le TU « S ». On observe alors des chutes montées au niveau des TTU « A1 » et « AM » rappelant la répartition des chutes sur les TU « A » et des montées sur les TU « S » :

[!|_A1	score [!|_A1	[!|_AM	score [!|_AM
TBH	344.7346	SDS	170.2884
HLS	335.0548	ULD	162.9394
HLD	323.3334	SLS	162.7763
HBU	314.9047	ULS	161.7848
ULS	314.2668	HBU	157.9673
TLS	308.9486	UBU	157.0188
UDD	308.386	HLD	150.0192
TBU	305.5655	HLS	145.6561
ULD	295.2107	SLD	145.3927
TLD	290.5226	SBH	143.2018
HLU	287.96	DLS	142.3409

Tableau 56 : Meilleures séquences de tons au niveau des TTU dans le modèle 16.

Au niveau d'un IU emphatique non terminal, le contour de tête est différent : les accents prénucléaires sont marqués par des chutes globales et une montée finale sur le nucléus ou le dernier accent mélodique postnucléaire. Il semble que l'on ait affaire à un contour de tête tombant suivi d'une chute-montée (« falling head+fall-rise ») comme on peut le voir dans le tableau ci-dessous :

[!+_A1	score [!+_A1	[!+_AM	score [!+_AM	[!+_ATF
H_L	177.4162	U_L	86.2821	__T
U_L	168.1609	USL	79.7626	__H
T_L	167.803	_UL	71.9565	U_T
U_D	166.5856	HSL	69.7712	L_T
S_D	158.3472	H_L	67.6028	S_T
D_L	152.9516	UDL	66.7588	LHT
DDL	152.8613	HDL	66.7411	DHT
__L	152.3621	HUL	66.4653	BHT
__D	151.941	S_D	66.4636	D_T
S_L	151.579	_SD	65.7918	B_T
D_D	150.6806	_SL	65.7077	_HT

Tableau 57 : Meilleures séquences de tons au niveau des TTU2 dans le modèle 23 pour un IU [!+.

Au niveau de la représentation des unités nucléaires, l'opposition entre la fonction emphatique et non emphatique est aussi modélisée par une attaque mélodique plus haute ({DDL} vs {U_B}) :

[!|_NF	score [!|_NF	[!|_!NF	score [!|_!NF	[!|_!N1F	score [!|_!N1F
DDL	172.4499	U_B	547.9736	_TDL	18.3809
DLB	171.3526	ULB	538.3843	_TDB	17.4944
D_B	170.3609	HLB	521.0767	UHDL	17.2444
__B	168.3578	SDD	520.3581	_H_D	17.1439
ULB	167.9962	ULD	520.0986	SUDL	17.0176
_LB	167.9654	SDL	519.8287	SU_L	17.0096
SDL	167.4131	H_B	517.9989	TTLD	10.8734
__L	165.6802	SLB	510.2394	_TLD	10.8419
__D	164.8813	S_D	503.9721	TTLS	10.8388
SLB	163.7152	SLD	502.6365	T_LD	10.8259
S_B	163.667	S_B	502.4848	_TLS	10.8131

Tableau 58 : Meilleures séquences de tons au niveau des TTU N dans le modèle 16.

Dans le modèle 24, on ajoute un accent mélodique postnucléaire si bien que l'ensemble des TTU « A » dans un IU non terminal sont représentés de la façon suivante :

[+_A1	score [+_A1	[+_AM	score [+_AM	[+_ATF
_HD	1144.6436	U_L	501.0809	_BU
_TD	1113.8869	S_L	498.5186	LBU
U_D	1059.9255	S_D	497.9228	DLU
UHD	1046.6229	__D	477.3224	_LU
_TL	1043.0987	__L	474.5787	DBU
_UD	1042.77	SDD	448.5674	_LS
H_L	1008.5678	ULD	441.8792	_BH
UUD	993.4127	SDL	433.8109	L_U
H_D	992.4834	_DD	432.2466	_DS

Tableau 59 : Meilleures séquences de tons au niveau des TTU A dans le modèle 24.

On observe une représentation formelle distincte pour les accents prénucléaires (représentés par des chutes : {HD}) et les accents postnucléaires (représentés par une montée finale : {_BU}).

Il est possible d'évaluer et de comparer la qualité de tous les modèles testés au niveau de l'optimisation des tons en rassemblant leurs résultats en termes de corrélation, d'écart quadratique moyen et de façon générale en comparant leur score. Ce sera l'objet de la prochaine section.

8.6.4 Évaluation des modèles d'optimisation des séquences INTSINT

Afin de faire une première évaluation de nos modèles après l'étape d'optimisation des séquences de tons, nous avons calculé leur corrélation moyenne sur l'ensemble du corpus ainsi que la moyenne de l'écart quadratique moyen et le score moyen. On présente les résultats tous types d'IU confondus puis séparément pour les IU terminaux et non terminaux pour chaque métrique comme le montrent les tableaux suivants :

Corrélations moyennes					
tous	correlation	[+	correlation	[|	correlation
int2tone46	0.90737	inttone24	0.92365	int2tone46	0.90652
int1tone56	0.90509	inttone23	0.92248	int1tone56	0.90129
int3tone66	0.90110	int2tone46	0.92206	int3tone66	0.89999
inttone23	0.89892	int1tone56	0.92114	inttone23	0.89230
inttone24	0.89712	inttone10	0.91903	inttone24	0.89100
int1tone50	0.89575	inttone21	0.91831	int1tone50	0.88941
int2tone42	0.89341	int1tone50	0.91603	int2tone42	0.88610
inttone21	0.89101	int3tone66	0.91480	int2tone40	0.88479
inttone10	0.89038	inttone16	0.91462	inttone10	0.88435
int3tone62	0.88923	inttone11	0.91159	int3tone62	0.88296
inttone7	0.88834	int2tone42	0.91139	inttone7	0.88235
int2tone40	0.88807	inttone6	0.91083	inttone21	0.88207
inttone16	0.88783	inttone14	0.91055	inttone16	0.88027
inttone11	0.88664	inttone7	0.90947	int1tone52	0.87586
inttone8	0.88149	inttone13	0.90858	inttone11	0.87584
inttone6	0.87906	int3tone62	0.90734	inttone8	0.87574
inttone14	0.87880	int2tone40	0.90729	inttone6	0.87386
inttone13	0.87763	inttone8	0.90571	inttone14	0.87139
int1tone52	0.87239	inttone12	0.89982	inttone12	0.86523
inttone9	0.87218	int1tone52	0.89748	inttone13	0.86462
inttone12	0.86874	inttone9	0.89657	inttone9	0.86371
inttone22	0.85975	inttone22	0.88985	inttone15	0.85870
inttone20	0.85495	inttone20	0.87586	inttone20	0.85830
inttone15	0.85420	inttone15	0.87450	inttone22	0.84861
inttone1	0.56602	inttone1	0.51319	inttone1	0.58737

Tableau 60 : Classement de tous les modèles en fonction de la corrélation et du type d'IU.

Scores moyens					
tous	score	[+	score	[|	score
int2tone46	10.07204	int2tone46	10.26919	int2tone46	10.04135
int1tone56	10.03333	int1tone56	10.26162	int1tone56	9.97281
int3tone66	9.98282	inttone24	10.20805	int3tone66	9.95979
int1tone50	9.88841	inttone23	10.18538	int1tone50	9.81936
int2tone42	9.86921	int3tone66	10.16893	int2tone42	9.79208
inttone23	9.85519	int1tone50	10.13500	inttone23	9.74859
inttone24	9.84792	int2tone42	10.08933	inttone24	9.74688
int3tone62	9.79754	inttone21	10.08545	int3tone62	9.73843
inttone21	9.75318	inttone10	10.07228	inttone10	9.65207
inttone10	9.72592	inttone16	10.04127	inttone21	9.64903
inttone16	9.71363	int3tone62	10.01460	inttone16	9.62563
inttone7	9.68137	inttone11	9.99823	inttone7	9.61364
inttone11	9.67464	inttone14	9.99372	int2tone40	9.59105
int2tone40	9.64822	inttone6	9.96224	inttone11	9.54762
inttone8	9.62403	inttone13	9.96074	inttone8	9.54417
inttone14	9.60760	inttone7	9.94575	inttone6	9.53235
inttone6	9.59476	inttone8	9.93698	inttone14	9.51143
inttone13	9.57671	int2tone40	9.88710	inttone9	9.42710
inttone9	9.50520	inttone12	9.83765	inttone13	9.41878
inttone12	9.46695	inttone22	9.80263	inttone12	9.41782
inttone22	9.41467	inttone9	9.79198	inttone15	9.39846
inttone20	9.38294	int1tone52	9.68102	inttone20	9.39278
int1tone52	9.38242	inttone20	9.63793	int1tone52	9.38998
inttone15	9.37088	inttone15	9.61766	inttone22	9.26124
inttone1	6.22545	inttone1	5.71810	inttone1	6.43114

Tableau 61 : Classement de tous les modèles en fonction du score et du type d'IU.

Moyennes de l'écart quadratique moyen					
tous	RMSE	[+	RMSE	[|	RMSE
int2tone46	1.3998	int2tone46	1.2862	int2tone46	1.4566
int1tone56	1.4629	int1tone56	1.3458	int1tone56	1.5199
int3tone66	1.4832	int3tone66	1.3582	int3tone66	1.5427
int2tone42	1.5530	int1tone50	1.4499	int2tone42	1.5885
int1tone50	1.5782	int2tone42	1.4580	int1tone50	1.6277
int3tone62	1.6316	int3tone62	1.5272	int3tone62	1.6689
inttone24	1.7446	inttone24	1.5290	inttone24	1.8609
inttone23	1.7672	inttone23	1.5715	inttone23	1.8800
inttone21	1.8221	inttone21	1.6464	inttone21	1.9183
inttone16	1.8410	inttone16	1.6588	inttone16	1.9237
inttone11	1.8753	inttone10	1.6882	inttone11	1.9614
inttone10	1.8898	inttone11	1.7098	inttone10	1.9619
inttone20	1.8974	inttone14	1.7211	inttone15	1.9896
inttone15	1.8984	inttone6	1.7283	inttone20	1.9985
inttone14	1.9151	inttone13	1.7289	inttone7	2.0078
inttone22	1.9298	inttone22	1.7480	inttone14	2.0126
inttone7	1.9341	inttone7	1.7497	inttone6	2.0217
inttone13	1.9388	inttone20	1.7562	inttone8	2.0244
inttone6	1.9425	inttone15	1.7622	inttone12	2.0494
inttone8	1.9628	inttone8	1.7645	inttone22	2.0517
inttone12	1.9666	inttone12	1.7804	inttone13	2.0530
inttone9	2.0194	inttone9	1.8354	inttone9	2.0767
int2tone40	2.0277	int2tone40	1.9424	int2tone40	2.0974
int1tone52	2.1869	int1tone52	2.0570	int1tone52	2.2905
inttone1	3.1414	inttone1	3.1229	inttone1	3.1773

Tableau 62 : Classement de tous les modèles en fonction de l'écart quadratique moyen et du type d'IU.

Le modèle qui obtient les meilleurs résultats quelle que soit la métrique prise en compte est le modèle 46 codé avec INTSINT2, suivi de près par le modèle 56 codé avec INTSINT d'origine et un registre élargi.et le modèle 66 codé avec INTSINT3. Ces trois modèles sont basés sur une annotation au niveau des unités tonales élargies. Au niveau des modèles basés sur les TU, on obtient la meilleure modélisation avec les modèles 50 et 42 codés respectivement avec INTSINT1 et INTSINT2. Le modèle le moins performant au niveau de tous les IU est le modèle 1 comme on pouvait aisément s'y attendre.

L'ensemble des modèles seront à nouveau comparés à partir de l'optimisation de l'alignement de leurs 30 meilleures séquences dans la section suivante.

8.7 Analyse et évaluation des résultats de l'optimisation de l'alignement

L'alignement des 30 meilleures séquences par type de TU est optimisé à partir des paramètres de variation définis dans le tableau 48. Un script Praat permet de tester toutes les distances prévues dans les paramètres et de choisir le meilleur alignement pour chaque TU en utilisant les résultats conjoints de la corrélation et de l'écart quadratique moyen.

La meilleure séquence de tons et le meilleur alignement par unité sont conservés et utilisés comme base de l'analyse statistique menée par la suite. La distance entre les points cibles et les points d'ancrage testés dans les modèles est conservée afin d'évaluer la qualité de chaque point d'ancrage. Les questions qui se posent sont les suivantes : à la lumière des résultats de l'étape d'optimisation de l'alignement des points cibles, quels sont les points de repère qui permettent de rendre compte le plus efficacement des variations d'alignement des tons ? À partir de quels points d'ancrage peut-on prédire au mieux les contours de F0 ? La position du TU a-t-elle un effet sur la façon dont s'alignent les points cibles ? Peut-on distinguer l'alignement dans les TU prénucléaires et postnucléaires ?

Nous observerons les variations d'alignement des points 1 et 2 essentiellement, car dans la plupart des cas le troisième correspond à un alignement fixe par rapport à la fin de l'unité concernée. Les résultats seront présentés en deux parties, la première correspondant aux modèles basés sur les TU, la seconde sur les modèles basés sur les TTU.

8.7.1 Alignement de P1 dans les modèles basés sur l'unité tonale

On peut rassembler les résultats d'alignement en deux parties : les variations d'alignement de P1 en pourcentages et en valeurs absolues. Plusieurs points d'alignement ont été testés et l'on cherchera à trouver la distance qui rassemble le plus de meilleures séquences dans chaque modèle.

Les points d'ancrage testés pour P1 sont le TU (durée et frontière gauche), la première syllabe du TU (durée) et la première voyelle du TU (frontière gauche).

P1 est optimisé en fonction de la durée d'une unité donnée dans les modèles 7, 9, 10 et 11.

8.7.1.1 Alignement du point 1 en pourcentages (modèles 7 et 11)

Le tableau ci-dessous montre les résultats de l'alignement de P1 en pourcentage de la durée de la première syllabe ou de la durée du TU, quelle que soit la position du TU dans l'IU.

Point d'ancrage + mode d'alignement	Variation de l'alignement	Nombre de points	En %	En % cumulé
Pourcentage / durée	0	1220	28.98%	28.98%
1ère syllabe	25	873	20.74%	49.71%
	50	829	19.69%	69.41%
	75	617	14.66%	84.06%
	100	272	6.46%	90.52%
	125	175	4.16%	94.68%
	150	224	5.32%	100.00%
Total		4210		
Pourcentage / durée TU	10	1077	27.58%	27.58%
	15	321	8.22%	35.80%
	20	345	8.83%	44.64%
	25	331	8.48%	53.11%
	30	338	8.66%	61.77%
	35	366	9.37%	71.14%
	40	1127	28.86%	100.00%
Total		3905		

Tableau 63 : Répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la syllabe accentuée ou de la durée de l'unité tonale.

En ce qui concerne l'alignement de P1 en fonction de la durée de la première syllabe du TU dans le modèle 11, on remarque que la plupart des meilleures séquences apparaissent à l'intérieur de la syllabe. 84 % des points apparaissent entre 0 et 75 % de la première syllabe.

La situation est bien moins définie quand il s'agit de prendre la durée du TU en entier comme point de repère. La plupart des points se concentrent à 10 % et à 40% du TU comme on peut le voir dans le graphique ci-dessous. 40% représentant la valeur extrême autorisée, on peut penser que 20% des points s'alignent à plus de 40% du TU.

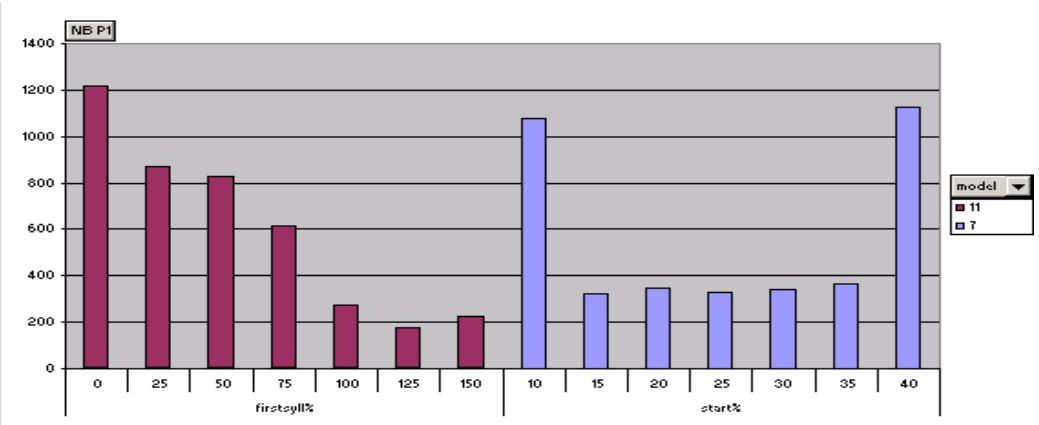


Figure 65 : Histogramme de la répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la première syllabe ou de la durée du TU.

> La répartition des meilleurs points d'alignement en % de la durée syllabe en bordeaux, en % de la durée du TU en bleu.
>
> En ordonnées = nombre de points cibles, en abscisses= points d'alignement testés pour la durée de la syllabe et la durée du TU.

Au niveau de la durée du TU en entier les résultats suggèrent qu'une fourchette de variation plus large aurait permis la présence de points beaucoup plus tard dans l'unité tonale.

8.7.1.2 Alignement de P1 en valeurs absolues (modèles 6 et 14)

Alignement par rapport au début du TU et par rapport à l'attaque de la première voyelle :

Alignement par rapport à	Variation de l'alignement	Nombre de points	En %	En % cumulé
Début du TU	0	685	16.94%	16.94%
	50	970	23.99%	40.92%
	100	777	19.21%	60.14%
	150	652	16.12%	76.26%
	200	457	11.30%	87.56%
	250	260	6.43%	93.99%
	300	243	6.01%	100.00%
Total		4044	100.00%	
Attaque de la	-100	330	7.92%	7.92%
Première	-50	695	16.67%	24.59%
voyelle	0	1137	27.28%	51.87%
	50	757	18.16%	70.03%
	100	576	13.82%	83.85%
	150	340	8.16%	92.01%
	200	163	3.91%	95.92%
	250	96	2.30%	98.22%
	300	74	1.78%	100.00%
Total		4168	100.00%	

Tableau 64 : Répartition des meilleurs points d'alignement : alignement par rapport au début de l'unité tonale et par rapport à l'attaque de la première voyelle.

Dans le modèle 6 (alignement par rapport au début du TU), on observe que 87,56% des points se concentrent dans une fourchette allant de 0 à 200 ms de la frontière gauche du TU.

Dans le modèle 14 (alignement à l'attaque de la première voyelle) 83,85% des points sont alignés entre -100 ms de la première voyelle et 100 ms de celle-ci comme on peut le voir dans les histogrammes ci-dessous :

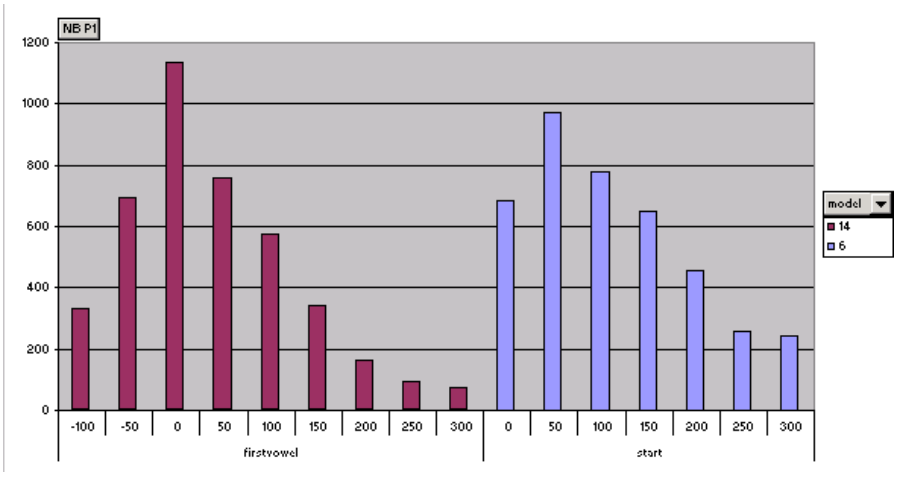


Figure 66 : Histogramme des meilleurs points d'alignement dans les modèles en valeur.

> La répartition des meilleurs points d'alignement à x de la 1ère voyelle en bordeaux, à x du début du TU en bleu.

En ce qui concerne l'alignement de P1, le meilleur point de repère, c'est-à-dire celui qui semble offrir le cadre le plus précis, est la durée de la première syllabe du TU.

Avant de passer à l'observation des résultats d'optimisation de l'alignement de P2, observons l'effet de la position du TU sur l'alignement de P1.

8.7.2 Effet de la position du TU sur l'alignement de P1

8.7.2.1 Effet de la position du TU sur l'alignement de P1 par rapport à la durée de la syllabe accentuée

Dans les modèles 11, 12 et 13, l'alignement de P1 varie en fonction de la durée de la syllabe de 0 à 150% en 7 itérations. Les premiers résultats généraux ont montré que la majorité des meilleurs points d'alignement se trouvait à l'intérieur de la syllabe (de 0 à 75%).

Le fait qu'un TU se trouve en position initiale, médiane ou finale, a-t-il un impact sur la façon dont s'alignent les points cibles qui lui sont associés et ici P1 ?

Variation de l'alignement	1	M	F	Total
0	29,53%	20,97%	36,29%	28,85%
25	14,88%	18,43%	29,56%	20,76%
50	19,71%	20,17%	17,82%	19,26%
75	16,09%	18,12%	9,28%	14,58%
100	7,48%	9,10%	2,79%	6,52%
125	5,29%	5,54%	2,22%	4,39%
150	7,02%	7,67%	2,05%	5,65%
Total	100,00%	100,00%	100,00%	100,00%

Tableau 65 : Répartition des meilleurs points d'alignement pour P1 en fonction de la position du TU.

On observe de façon globale que la majorité des meilleurs points d'alignement se trouve dans la syllabe. On peut remarquer aussi des différences d'alignement en fonction de la position prise en compte :

- Au niveau d'un TU initial : 87,69 % des points se concentrent dans la syllabe (entre 0 et 100%) ;
- Au niveau d'un TU en position médiane : 86,79% des points se concentrent dans la syllabe ;
- Au niveau d'un TU final : 83, 67 % des points se concentrent entre 0 et 50% de la syllabe et 92, 74 % des points sont représentés dans la syllabe en entier.

L'alignement de P1 en position finale se fait plus tôt dans la syllabe. On a un effet de décalage vers la gauche qui a déjà été décrit dans la littérature. La figure ci-dessous confirme cet effet :

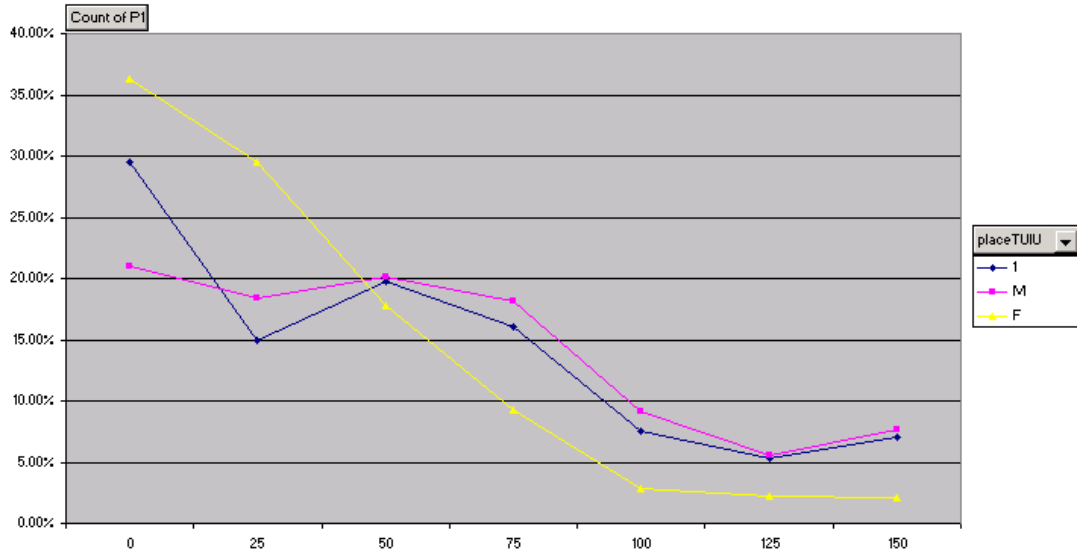


Figure 67 : Variation de l'alignement de P1 en % de la durée de la première syllabe et en fonction de la position du TU.

> En ordonnées = le nombre de points en %, en abscisses = les points d'alignement testés (25%, 50% etc.).
>
> La courbe bleue représente les résultats d'alignement de P1 dans un TU1, la courbe rose dans un TUM et la courbe jaune dans un TUF.

On voit que dans un TU final le nombre de P1 optimaux diminue très rapidement à partir de 50 % de la syllabe alors que la diminution est bien plus progressive au niveau des TU « 1 » et « M ».

On s'intéressera ensuite à l'effet de la position du TU sur l'alignement de P1 par rapport à la durée du TU en entier.

8.7.2.2 Effet de la position du TU sur l'alignement de P1 en % de la durée du TU

Variation de l'alignement	1	M	F	Total
10	22,82%	25,10%	32,02%	26,61%
15	7,40%	8,31%	10,21%	8,64%
20	8,02%	9,51%	9,25%	8,95%
25	8,11%	9,35%	8,81%	8,78%
30	8,99%	8,07%	9,86%	8,95%
35	10,31%	9,43%	8,12%	9,29%
40	34,36%	30,22%	21,73%	28,79%
Total	100,00%	100,00%	100,00%	100,00%

Tableau 66 : Répartition des meilleurs points d'alignement pour P1 en % de la durée du TU et en fonction de la position du TU.

La position du TU dans l'IU ne semble pas avoir un effet significatif sur l'alignement de P1 en fonction de la durée du TU. Quelle que soit la position choisie, on retrouve deux points de repère principaux rassemblant la plupart des meilleures séquences : à 10 % de la durée du TU et à 40 % d'où la forme du graphique ci-dessous :

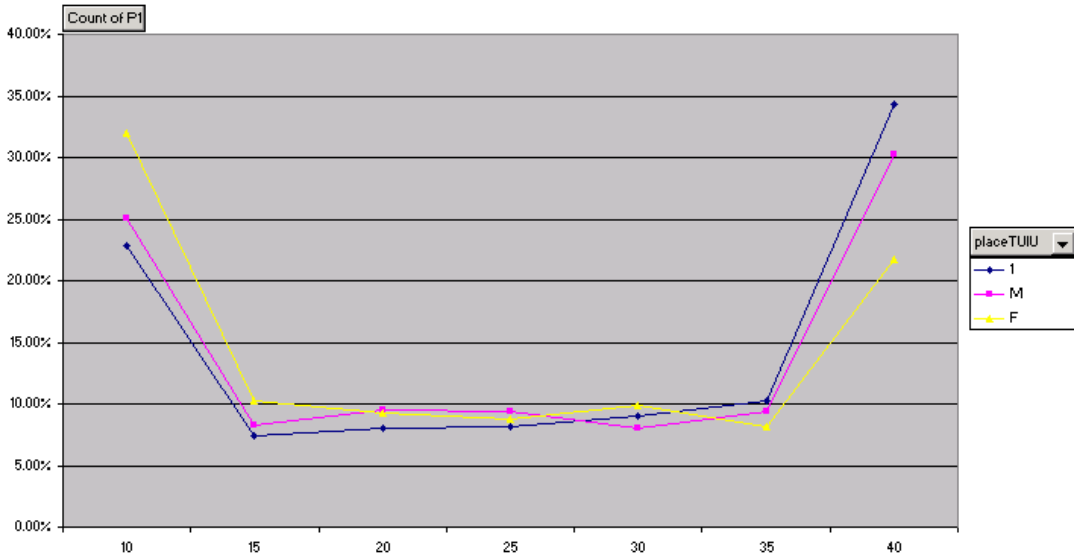


Figure 68 : Variation de l'alignement de P1 en % de la durée du TU en fonction de la position du TU.

On voit bien que les trois courbes représentant les différentes positions du TU ont la même forme incluant une augmentation du nombre de meilleurs alignements sur les valeurs extrêmes à 10 et 40% du TU.

8.7.2.3 Effet de la position du TU sur l'alignement de P1 par rapport au début du TU

Variation de l'alignement	1	M	F	Total
0	0.00%	23.20%	32.24%	18.67%
50	31.85%	16.55%	21.05%	23.00%
100	23.11%	17.58%	18.67%	19.73%
150	17.39%	15.60%	14.31%	15.75%
200	13.36%	12.19%	7.07%	10.87%
250	6.89%	6.89%	4.36%	6.05%
300	7.39%	8.00%	2.30%	5.91%
Total	100,00%	100,00%	100,00%	100,00%

Tableau 67 : Répartition des meilleurs points d'alignement pour P1 par rapport au début du TU et en fonction de la position du TU.

De façon générale, l'alignement de P1 dans un TU en position finale se distingue de son alignement dans les TU « 1 » et « M ». P1 s'aligne plus tôt dans le TU s'il est en position finale. Pour un TU « 1 », 85,71% des points se trouvent entre 0 et 200 ms du début du TU, dans un TU « M » ils sont au nombre de 85,11% dans ce même intervalle. Dans un TU « F » 86,27% des points sont déjà représentés entre 0 et 150 ms du début du TU et 93,34% dans l'intervalle de 0 à 200 ms.

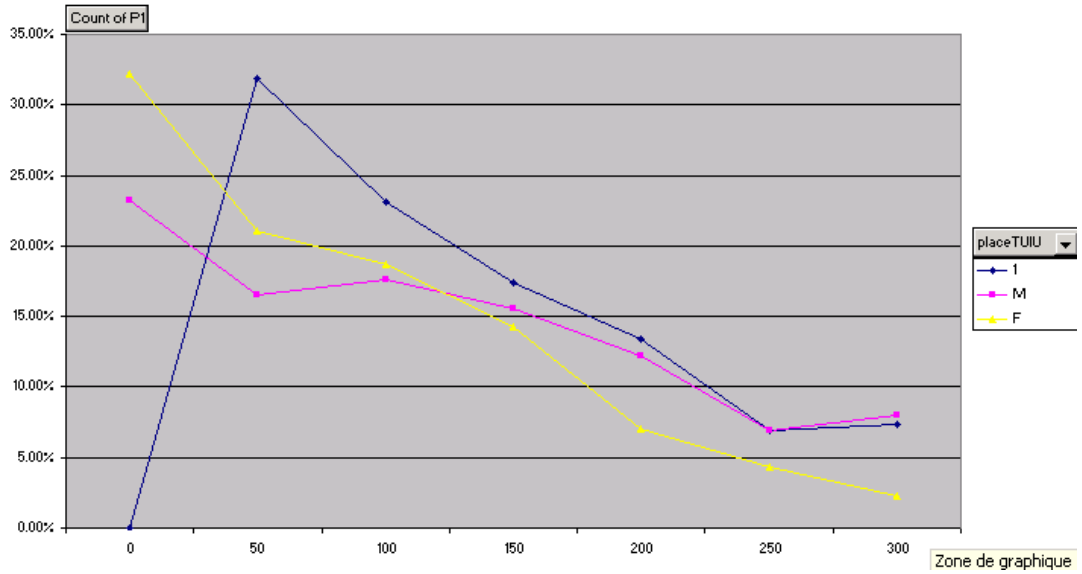


Figure 69 : Variation de l'alignement de P1 par rapport au début du TU et en fonction de la position du TU

8.7.2.4 Effet de la position du TU sur l'alignement de P1 avec la première voyelle du TU

Variation de l'alignement	1	M	F	Total
-100	5.98%	6.00%	12.26%	7.99%
-50	14.91%	14.39%	22.29%	17.09%
0	28.69%	21.98%	31.40%	27.34%
50	18.77%	16.71%	17.32%	17.62%
100	15.44%	16.39%	9.28%	13.79%
150	9.08%	11.19%	3.65%	8.04%
200	3.79%	5.44%	2.32%	3.86%
250	1.89%	4.48%	0.91%	2.43%
300	1.44%	3.44%	0.58%	1.83%
Total	100.00%	100.00%	100.00%	100.00%

Tableau 68 : Répartition des meilleurs points d'alignement pour P1 par rapport au début de la voyelle accentuée et en fonction de la position du TU.

C'est l'alignement de P1 dans les TU « M » qui varie le plus largement puisque la plupart des points se répartissent entre -100 ms et 150 ms de la voyelle. L'alignement de P1 dans les TU « 1 » est plus variable que dans les TU « F » où les points s'alignent légèrement plus tôt : entre -100 et 100 ms 83,80% des points sont représentés dans un TU « 1 », l'intervalle est réduit de -100 ms à 50 ms pour 83,26% des points dans les TU « F ».

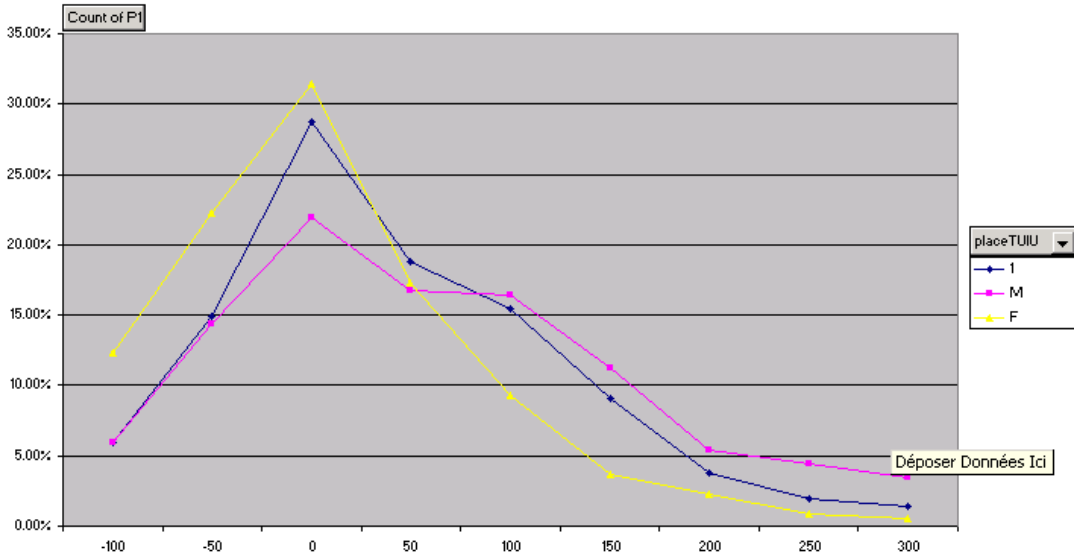


Figure 70 : Variation de l'alignement de P1 par rapport au début du de la voyelle et en fonction de la position du TU.

On voit que les courbes représentant les TU « 1 » et « M » se suivent d'assez près alors que la courbe des TU finaux atteint des valeurs minimales plus tôt.

De façon globale, on a pu voir que l'alignement de P1 était influencé par l'effet de l'initialité et de la finalité. P1 s'aligne plus tôt dans les TU en position finale que dans les TU en position initiale. La position médiane semble être le lieu de la variation la plus grande. Nous observerons ensuite l'effet de la position du nucléus sur l'alignement de P1.

Nous chercherons donc à savoir si les accents prénucléaires et postnucléaires s'alignent de façon distincte à partir de l'étude des résultats de l'optimisation de l'alignement dans les différents modèles.

8.7.3 Effet de la position du nucléus sur l'alignement de P1

8.7.3.1 Effet de la position du nucléus sur l'alignement de P1 en % de la durée de la syllabe

Exprimé en nombre de points

Variation de l'alignement	Avant le nucléus	Au niveau du nucléus	Après le nucléus	Total
0	636	445	139	1220
50	397	418	58	873
100	456	331	42	829
150	381	201	35	617
200	193	71	8	272
250	128	40	7	175
300	175	38	11	224
Total	2366	1544	300	4210

Tableau 69 : Répartition des meilleurs points d'alignement de P1 en % de la durée de la syllabe en fonction de la position du nucléus.

On peut voir de façon générale que l'alignement de P1 varie dans une fourchette plus large dans les TU prénucléaires que dans les TU nucléaires et postnucléaires. Pour être plus précis, observons les résultats exprimés en pourcentages cumulés :

Variation de l'alignement	Avant le nucléus	Au niveau du nucléus	Après le nucléus
0	26,88%	28,82%	46,33%
25	43,66%	55,89%	65,67%
50	62,93%	77,33%	79,67%
75	79,04%	90,35%	91,33%
100	87,19%	94,95%	94,00%
125	92,60%	97,54%	96,33%
150	100,00%	100,00%	100,00%

Tableau 70 : Répartition des meilleurs points d'alignement en fonction de la position du nucléus et en pourcentages cumulés.

On voit bien que l'alignement de P1 est de plus en plus tôt dans la syllabe alors que l'on se déplace vers la droite. Pour les accents prénucléaires, 92,60% des points se trouvent entre 0 et 125% de la syllabe, ce qui constitue une fourchette assez large de variation. Au niveau du nucléus, 90,35% des points apparaissent entre 0 et 75% de la syllabe et au niveau des TU postnucléaires, 91,33% des points sont représentés aux trois quarts de la syllabe.

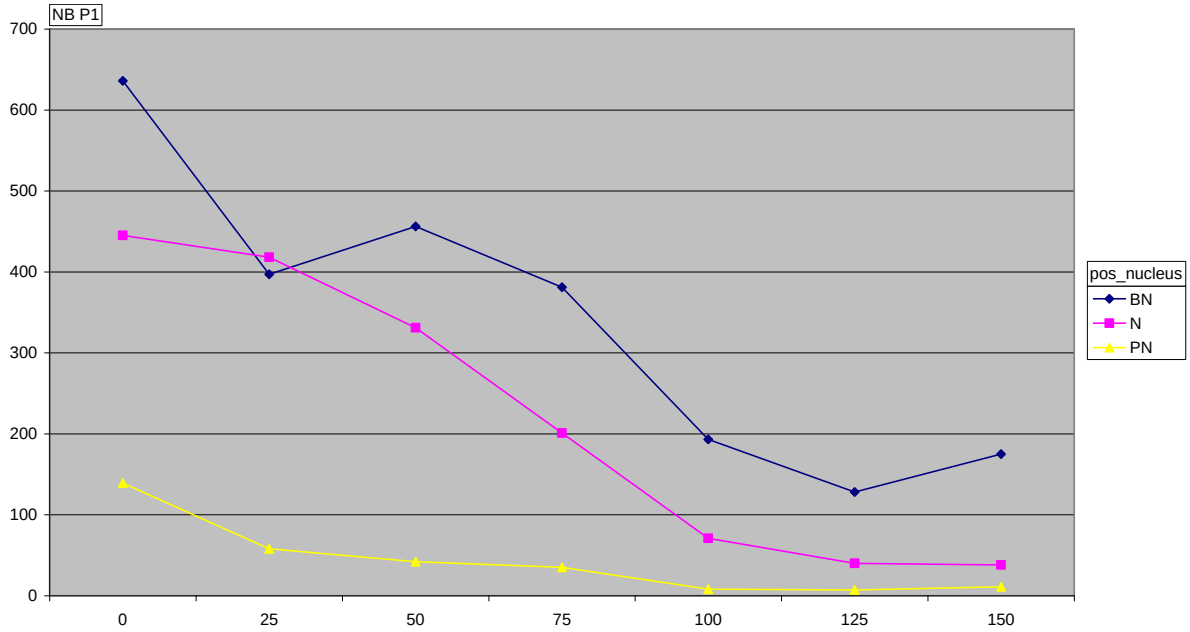


Figure 71 : Variation de l'alignement de P1 en % de la durée de la syllabe et en fonction de la position du nucléus.

> BN = « Before nucléus » en bleu, N = « at the level of the Nucléus » en rose,
> PN = « post-nuclear » en jaune

Les courbes représentant les TU prénucléaires et nucléaires atteignent leur valeurs minimales à 125% de la durée de la syllabe, alors que la courbe représentant les TU postnucléaires atteint son minimum à 100% de la syllabe.

8.7.3.2 Effet de la position du nucléus sur l'alignement de P1 en valeur (début du TU)

Exprimé en pourcentage cumulé.

Variation de l'alignement	Avant le nucléus	Au niveau du nucléus	Après le nucléus
0	12.47%	17.58%	45.95%
50	37.04%	41.04%	68.28%
100	57.61%	59.95%	79.29%
150	73.81%	77.53%	87.70%
200	85.46%	89.62%	92.56%
250	92.44%	95.57%	97.41%
300	100.00%	100.00%	100.00%

Tableau 71 : Répartition des meilleurs points d'alignement pour P1 par rapport au début du TU et en fonction de la position du nucléus.

On observe le même effet de la position du nucléus sur l'alignement de P1 par rapport au début du TU que dans l'expérience précédente. La plupart des points se trouvent entre 0 et 250 ms du début du TU pour les TU BN et N alors qu'ils ne varient essentiellement qu'entre 0 et 150 ms dans le cas des TU PN. C'est ce que l'on peut observer dans le graphique ci-dessous :

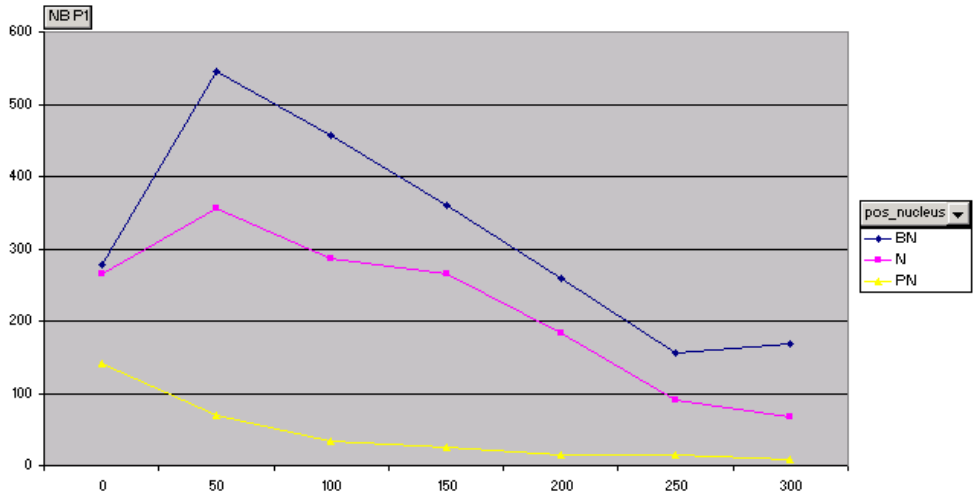


Figure 72 : Variation de l'alignement de P1 par rapport au début du TU et en fonction de la position du nucléus.

8.7.3.3 Effet de la position du nucléus sur l'alignement de P1 par rapport à la voyelle accentuée

Variation de l'alignement	Avant le nucléus	Au niveau du nucléus	Après le nucléus
-100	5.72%	9.98%	14.73%
-50	20.74%	27.98%	37.67%
0	46.73%	56.69%	67.81%
50	64.87%	75.93%	80.48%
100	80.07%	88.85%	88.01%
150	89.50%	95.76%	92.47%
200	94.41%	98.11%	96.58%
250	97.70%	98.89%	98.97%
300	100.00%	100.00%	100.00%

Tableau 72 : Répartition des meilleurs points d'alignement pour P1 par rapport au début de la voyelle et en fonction de la position du nucléus.

L'alignement de P1 est à nouveau plus tardif dans les TU prénucléaires et nucléaires cependant on observe aussi un nombre de points maximal au niveau du début de la voyelle exactement, reconnaissable à un pic caractéristique dans le graphique ci-dessous :

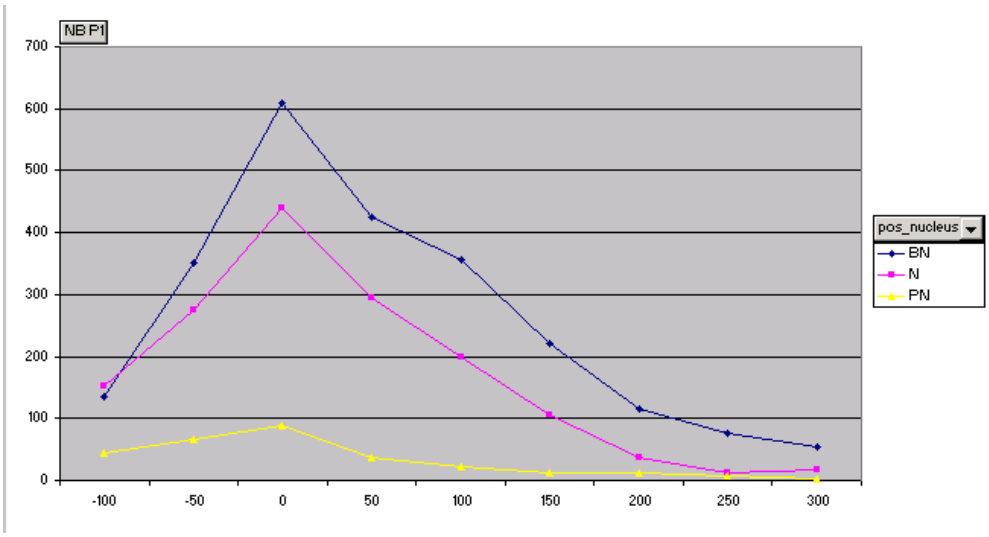


Figure 73 : Variation de l'alignement de P1 par rapport au début de la voyelle et en fonction de la position du nucléus.

On a vu à travers l'observation de l'alignement de P1 par rapport aux différents points d'ancrage, que la position du nucléus semble avoir un effet sur l'alignement de P1. Il est plus tardif en position prénucléaire qu'en position postnucléaire. La position nucléaire semble être un intermédiaire entre ces deux extrêmes. On peut aussi analyser cette distinction comme un simple effet de l'initialité et de la finalité puisque la plupart des TU postnucléaires sont en position finale.

8.7.4 Alignement de P1 en fonction du contour

Après avoir étudié les variations de P1 de façon globale, nous rentrerons dans plus de détails en prenant en compte la configuration de F0. Nous prendrons les exemples des contours du type « HL », « LH » et « HLH ». Les contours de type « HL » correspondent dans nos modèles à des séquences du type {HD}, {TD}, {HB}, {UL}, etc. Les contours de type « LH » correspondent aux séquences INTSINT {LU}, {BU}, {LH}, {LT}, etc. et les contours « HLH » aux séquences de tons INTSINT telles que {UBU}, {HBH}, etc.

Il s'agira, par exemple, de voir si l'alignement d'un pic « H » faisant partie d'un contour descendant se distingue de l'alignement d'un ton « L » débutant un contour montant.

On comparera d'abord l'alignement de P1 dans les contours « HL », « LH » et « HLH », en fonction de la durée de la première syllabe, puis en fonction du début de l'unité tonale et enfin en fonction de la voyelle accentuée.

8.7.4.1 Alignement de P1 dans les contours de type « HL », « LH » et « HLH » par rapport à la durée de la syllabe accentuée

	Position 1		Position M		Position F		
P1	HL	LH	HL	LH	HL	HLH	LH
0	19.40%	43.70%	17.80%	20.63%	19.83%	21.13%	53.47%
25	15.72%	14.07%	18.85%	9.52%	32.47%	42.25%	20.82%
50	25.08%	13.33%	23.82%	13.76%	26.44%	12.68%	12.24%
75	17.06%	11.85%	22.77%	17.99%	11.49%	16.90%	9.39%
100	8.86%	4.44%	7.07%	13.76%	3.45%	1.41%	0.82%
125	5.02%	5.93%	3.66%	8.99%	2.87%	4.23%	1.63%
150	8.86%	6.67%	6.02%	15.34%	3.45%	1.41%	1.63%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tableau 73 : . Variations de l'alignement de P1 dans les contours HL, LH, HLH par rapport à la syllabe accentuée et en fonction de la position du TU.

Au niveau du contour « HL », le pic « H » s'aligne de façon régulière entre 0 et 75 % de la syllabe accentuée. On observe un alignement plus concentré dans les TU en position finale dans lesquels la plupart des meilleurs points d'alignement apparaissent entre 0 et 50% de la syllabe accentuée.

Au niveau du contour « LH », on a une préférence nette pour un alignement avec le début de la syllabe accentuée (à 0%) dans les TU en position initiale et finale (avec 43,70 et 53,47% des points alignés exactement au début). L'alignement du début de la montée est beaucoup plus variable en position médiane.

L'alignement du ton « H » dans un contour du type « HLH » varie essentiellement de 0 à 75% de la syllabe accentuée. C'est ce que l'on peut voir plus précisément dans les figures ci-dessous :

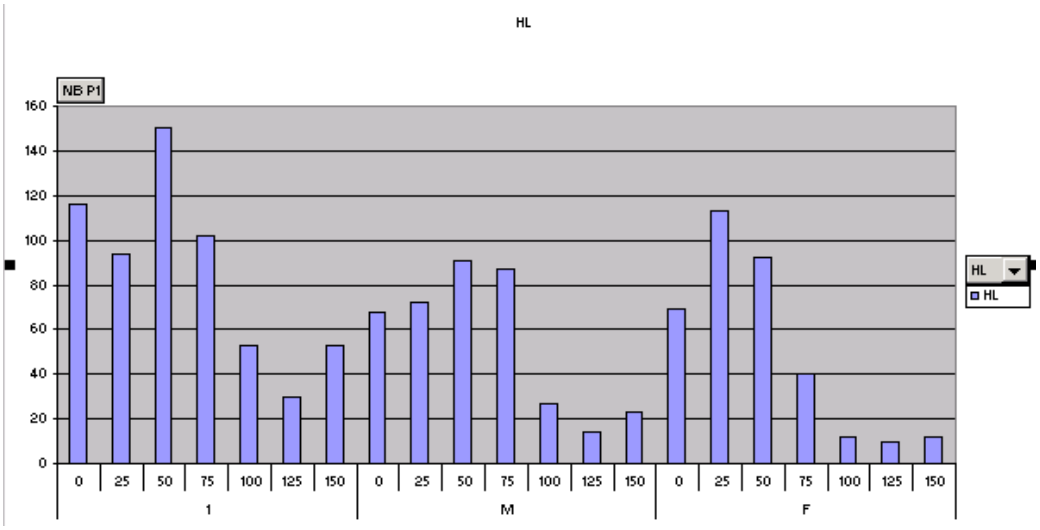


Figure 74 : Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour HL et en fonction de durée de la syllabe accentuée.

On voit aussi ici que dans les TU en position initiale, l'alignement du pic « H » varie assez largement, le nombre le plus important de points est cependant aligné à 50% de la syllabe accentuée.

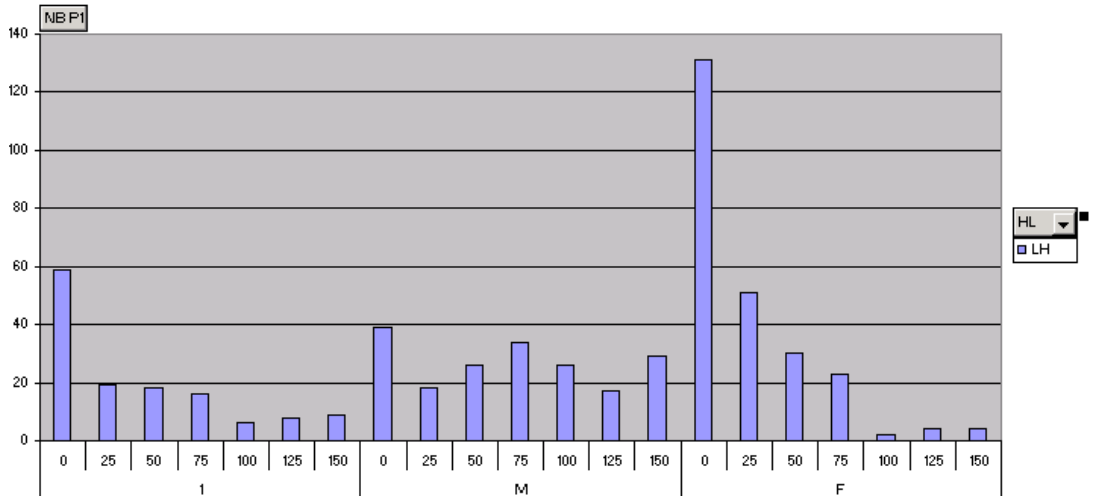


Figure 75 : .Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour LH et en fonction de la durée de la syllabe accentuée.

En position finale, la majorité des points sont alignés au début de la syllabe accentuée pour le contour « LH ».

8.7.4.2 Alignement de P1 dans les contours HL, LH, HLH par rapport au début du TU en ms

	Position 1		M		F	
P1	HL	LH	HL	LH	HL	LH
0	0.00%	0.00%	17.52%	23.83%	18.97%	49.76%
50	35.94%	26.22%	14.56%	13.55%	18.43%	22.44%
100	19.35%	22.85%	18.87%	13.08%	23.31%	11.22%
150	18.43%	11.99%	20.22%	12.62%	18.70%	6.83%
200	15.67%	17.98%	16.17%	13.55%	10.03%	3.41%
250	5.30%	10.11%	7.82%	7.94%	7.05%	4.39%
300	5.30%	10.86%	4.85%	15.42%	3.52%	1.95%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tableau 74 : Variations de l'alignement de P1 par rapport au début du TU en fonction de la position du TU.

L'alignement du pic « H » dans un TU en position initiale varie essentiellement de 50 à 200 ms du début du TU alors qu'il varie de 0 à 150 ms en position finale. On obtient les mêmes résultats pour les contours « HLH » en position finale. Dans les TU médians la variation est plus large et va de 0 à 200 ms.

En position initiale, le premier ton du contour de type « LH » s'aligne de façon quasiment identique au pic « H » soit de 50 à 200 ms du début du TU. Dans les TU en position finale, on observe une variation nettement plus réduite comprise entre 0 et 100 ms du début avec une préférence pour un alignement exactement au début du TU. La position médiane montre à nouveau le plus de variation dans l'alignement de P1 :

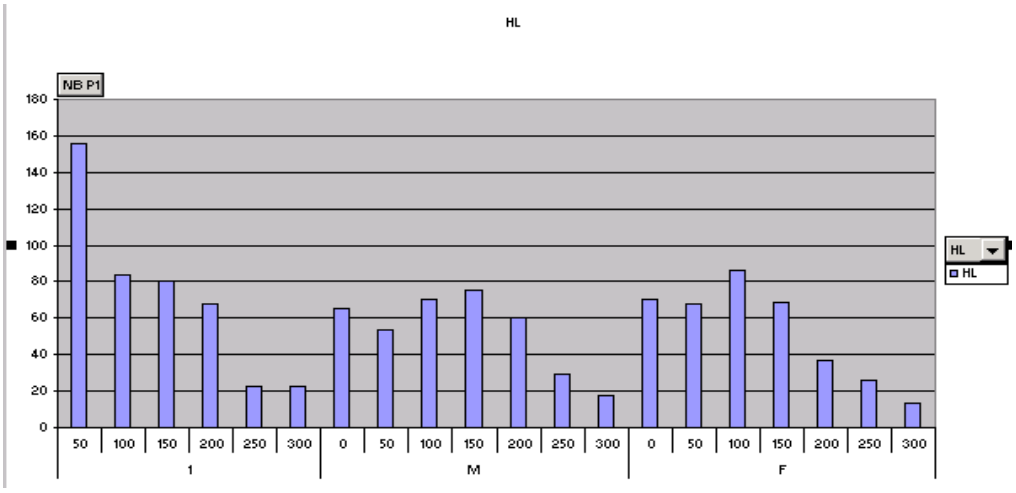


Figure 76 : Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour HL et en fonction de la position du TU.

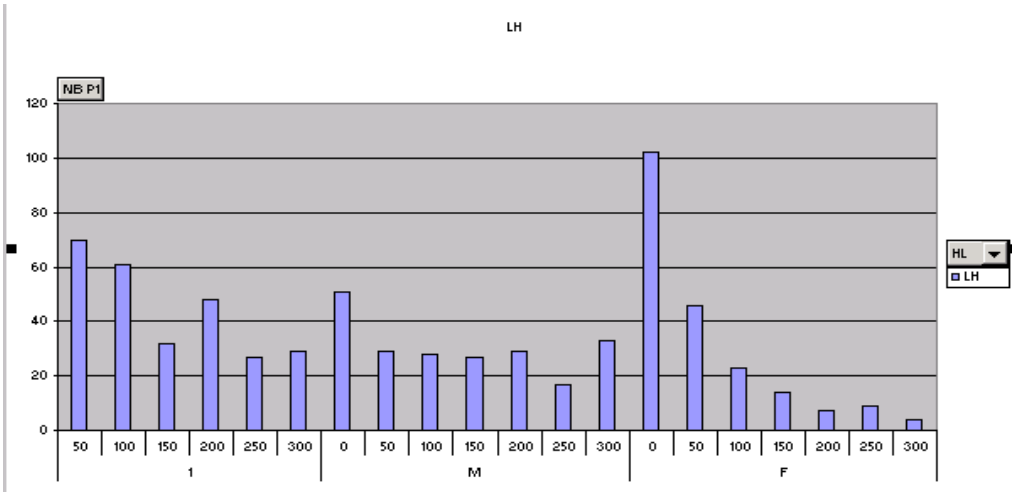


Figure 77 : .Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour LH et en fonction de la position du TU.

8.7.4.3 Alignement de HL, LH et HLH par rapport à l'attaque de la voyelle accentuée

	Position 1		Position M		Position F		
P1	HL	LH	HL	LH	HL	HLH	LH
-100	4.88%	6.90%	5.31%	4.79%	6.07%	6.15%	19.79%
-50	14.15%	20.69%	17.88%	8.51%	20.23%	21.54%	30.48%
0	25.04%	29.31%	21.23%	16.49%	32.95%	27.69%	30.48%
50	20.49%	16.38%	23.18%	10.11%	20.23%	24.62%	10.16%
100	17.24%	10.34%	14.53%	14.36%	12.14%	10.77%	4.28%
150	10.24%	8.62%	9.50%	20.74%	4.05%	6.15%	2.14%
200	4.39%	4.31%	3.91%	10.64%	2.31%	1.54%	2.14%
250	2.44%	1.72%	2.23%	8.51%	0.87%	0.00%	0.53%
300	1.14%	1.72%	2.23%	5.85%	1.16%	1.54%	0.00%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tableau 75 : Variations de l'alignement de P1 par rapport au début de la voyelle en fonction de la position du TU.

Les meilleurs points d'alignement pour le premier point du contour « HL » se répartissent entre -50 et 150 ms du début de la voyelle en position initiale et médiane. Dans un TU en position finale, l'alignement varie plutôt de -50 à 100 ms de la voyelle.

Pour le contour « LH » dans un TU initial, la plupart des points s'alignent entre -50 et 0 ms de l'attaque de la voyelle. Dans les TU en position finale, cette fourchette s'élargit quelque peu : de -100 ms à 0 ms de la voyelle. Dans les TU en position médiane, l'alignement varie de -50 à 150 ms de la voyelle.

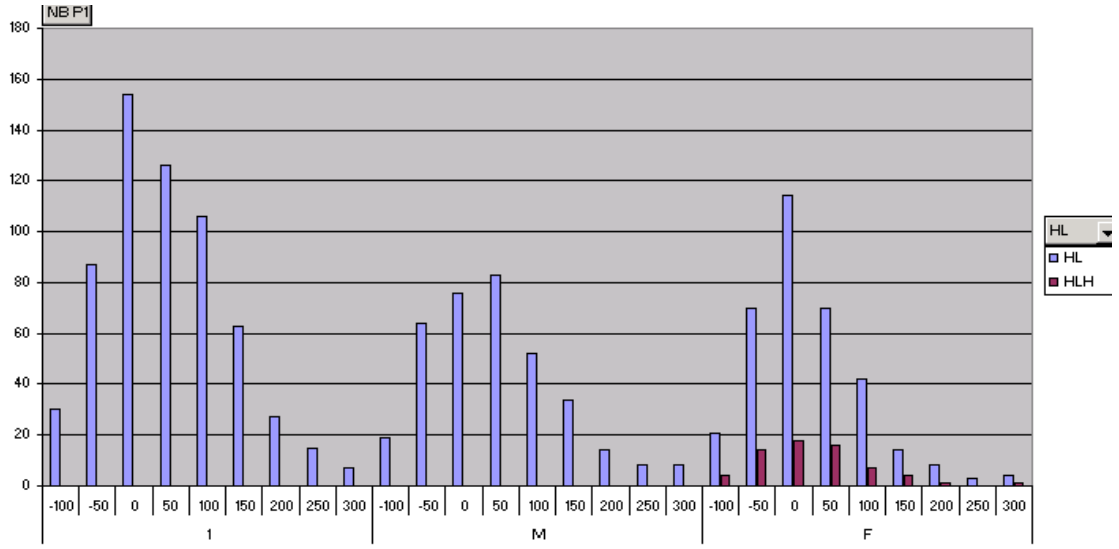


Figure 78 : Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour HL et en fonction de la position du TU.

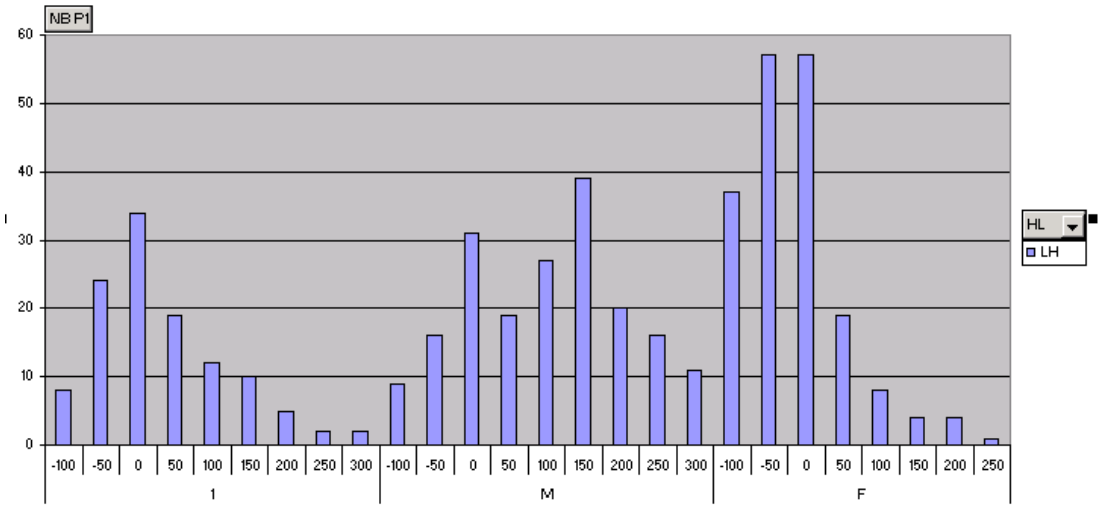


Figure 79 : .Histogrammes de la répartition des meilleurs points d'alignement de P1 pour un contour LH et en fonction de la position du TU.

Dans l'ensemble, on remarque que les pics « H » ou « L » s'alignent différemment par rapport aux différents points d'ancrage testés. Les variations de l'alignement du pic « H » reflètent assez clairement les résultats obtenus de façon globale dans les sections précédentes. La durée de la syllabe accentuée est un point de repère pour son alignement qui varie essentiellement à l'intérieur de celle-ci. Les variations de l'alignement en position initiale ou médiane sont généralement plus importantes qu'en position finale.

Au niveau du contour de type « LH », on obtient des résultats plus contrastés que les résultats globaux. Les deux points d'ancrage qui ressortent ici sont le début de la syllabe accentuée et l'attaque de la voyelle. Si l'on prend en compte essentiellement les TU initiaux et finaux, on observe un alignement massif exactement au début de ces unités prosodiques.

Nous nous pencherons dans les sections suivantes sur l'étude des variations de P2 en fonction des différents points d'ancrage et modes d'alignement testés.

8.7.5 Alignement de P2 dans les modèles basés sur l'unité tonale

L'optimisation de l'alignement du point P2 dépend étroitement de la position de l'unité tonale en question. En effet les TU en position 1 ou M sont modélisés à l'aide de deux points seulement si bien que P2 représente le dernier point de l'unité tonale. Les TU en position F ou 1F sont modélisés par trois points cibles si bien que P2 devient le point médian.

Nous analyserons donc séparément l'optimisation de l'alignement de P2 : tout d'abord P2 situé dans un TU en position 1 ou M, puis P2 situé dans un TU en position F ou 1F.

8.7.5.1 Alignement de P2 dans les TU en position 1 ou M

L'alignement de P2 est ici optimisé en fonction de la frontière droite du TU :

- en distance en millisecondes de la fin du TU (ce point de repère apparaît sous l'abréviation « end » dans les graphique de résultats) ;
- ou en pourcentage de la durée du TU (« end% ») ;
- en nombre de syllabes depuis la fin de dernière syllabe du TU (« LASTSOFF » pour « Last Syllable Offset ») : P2 peut être aligné soit au niveau de la fin de la dernière syllabe, soit à la fin de l'avant dernière syllabe, soit à la fin de l'avant avant dernière syllabe.

Nous retiendrons comme exemple respectivement les modèles 11, 7, 8 pour les modes d'alignement et points d'ancrage cités ci-dessus.

Point d'ancrage + mode d'alignement	Variation de l'alignement	Nombre de points en position 1	Nombre de points en position M	Total position 1 et M	Total position 1 et M En %
Fin du TU	-160	38	27	65	2.51%
	-110	50	30	80	3.09%
	-60	93	62	155	5.99%
	-10	1143	1145	2288	88.41%
Total		1324	1264	2588	
Pourcentage de la durée du TU	80	90	49	139	5.83%
	85	69	36	105	4.40%
	90	55	78	133	5.57%
	95	921	1088	2009	84.20%
Total		1135	1251	2386	
Fin de la dernière syllabe	-2	53	19	72	2.96%
	-1	146	87	233	9.58%
	0	976	1150	2126	87.45%
Total		1175	1256	2431	

Tableau 76 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1 et M.

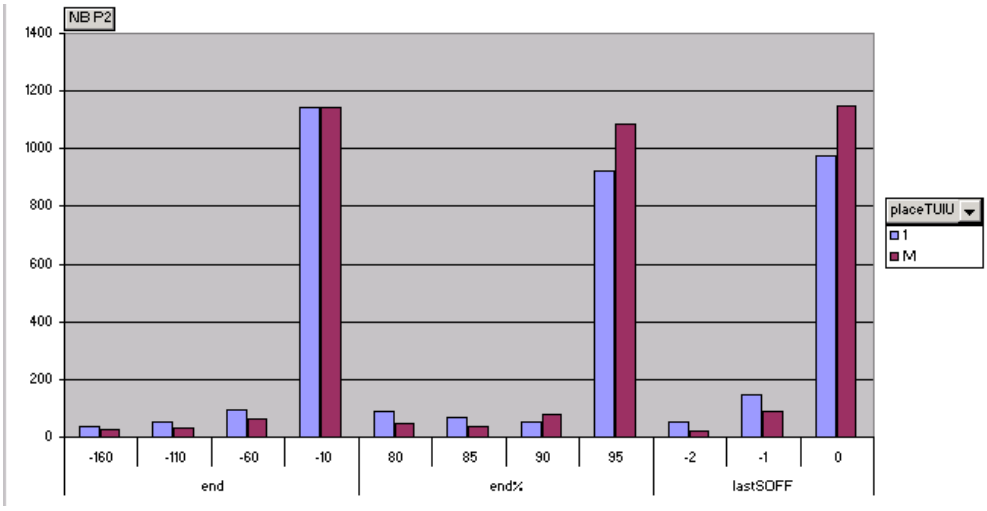


Figure 80 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1 ou M.

La fin du TU apparaît comme un excellent point de repère pour l'alignement de P2 puisque 88% des points sont alignés de façon fixe à 10 ms de celle-ci. On voit aussi que lorsque l'on teste trois points d'alignements (la fin de la dernière syllabe, la fin de l'avant dernière syllabe, etc.), P2 s'aligne majoritairement avec la fin de la dernière syllabe qui coïncide avec la fin du TU. Quand on mesure les variations de l'alignement de P2 par rapport à la durée du TU en entier, la plupart des points se concentrent aussi à 95% de l'unité tonale.

8.7.5.2 Alignement de P2 dans les TU en position finale F

En position finale, nous avons trois points cibles. Le point P2 est alors un point médian.

L'alignement de P2 est optimisé en fonction de plusieurs points d'ancrage :

- en distance en millisecondes du point précédent (« prec ») ;
- en pourcentage de la durée du TU (« end% ») ;
- en distance en millisecondes de l'attaque de la dernière voyelle (« lastVON ») ;
- en nombre de syllabes depuis l'attaque de la dernière syllabe du TU (« LASTSON » pour « Last Syllable Onset ») ;
- en pourcentage de la durée de la dernière syllabe (« lastsyll% »).

Alignement de P2 par rapport à la cible précédente ou en pourcentage de l'unité tonale

Nous retiendrons ici comme exemple respectivement les modèles 6 et 7 .

Alignement par rapport à	Variations de l'alignement	Nombre de points en position F	En %
Cible précédente	50	224	18.42%
	100	261	21.46%
	150	236	19.41%
	200	205	16.86%
	250	150	12.34%
	300	140	11.51%
Total		1216	
Pourcentage de la durée du TU	80	739	64.49%
	85	103	8.99%
	90	89	7.77%
	95	215	18.76%
Total		1146	

Tableau 77 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position F avec point d'ancrage la cible précédente ou la durée du TU.

Lorsque le point de repère est la cible tonale précédente, nous avons une répartition régulière des points d'alignement de 50 ms à 300 ms de celle-ci.

Pour un alignement par rapport à la durée du TU, 64 % des points semblent s'aligner à 80%% de la durée du TU, mais ce chiffre n'est pas significatif, car c'est la limite inférieure que nous avions fixée dans les paramètres. La répartition laisse supposer que si nous avions élargi la fourchette, les points se seraient répartis sur cette nouvelle fourchette : ce repère n'est pas précis.

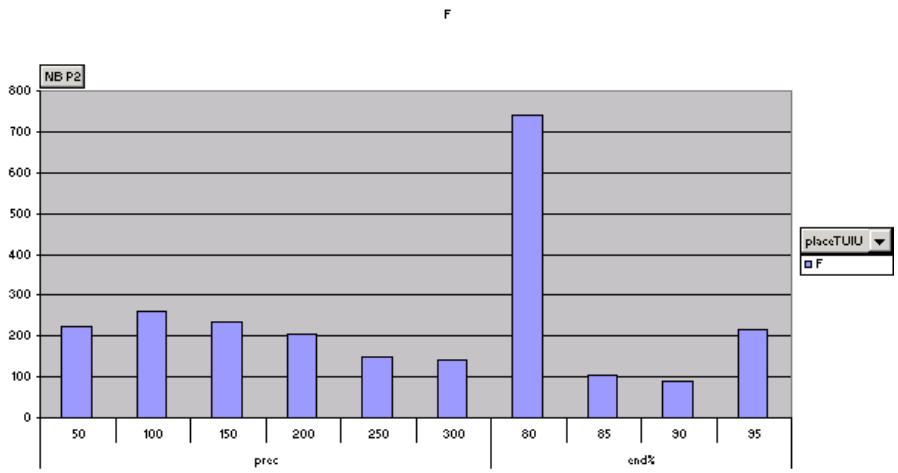


Figure 81 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position F avec point d'ancrage la cible précédente ou un pourcentage de la durée du TU.

Alignement de P2 par rapport à la dernière voyelle ou à la dernière syllabe

Nous retiendrons comme exemple respectivement les modèles 14, 8, 11 où P2 s'aligne en :

- distance en millisecondes de l'attaque de la dernière voyelle ;
- distance en nombre de syllabes de l'attaque de la dernière syllabe ;
- pourcentage de la durée de la dernière syllabe.

Point d'ancrage + mode d'alignement	Variation de l'alignement	Nombre de points en position F	En %	En % cumulé avant et à partir du point d'ancrage
Attaque de la dernière voyelle	-100	100	8.29%	
	-50	114	9.44%	17.73%
	0	161	13.34%	82.27%
	50	190	15.74%	
	100	216	17.90%	
	150	188	15.58%	
	200	113	9.36%	
	250	70	5.80%	
	300	55	4.56%	
Total		1207		

Attaque de la dernière syllabe	-1	23	1.92%	1.92%
	0	1174	98.08%	98.08%
Total		1197		
Pourcentage de la durée de la dernière syllabe	-40	72	5.91%	
	-20	38	3.12%	9.03%
	0	47	3.86%	90.97%
	20	153	12.56%	
	40	207	17.00%	
	60	341	28.00%	
	80	360	29.56%	
Total		1218		

Tableau 78 : Répartition des meilleurs points d'alignement pour P2 en fonction de la dernière syllabe ou de la dernière voyelle.

L'attaque de la dernière voyelle ne semble pas représenter un point de repère très informatif puisque les points sont éparpillés entre 100 ms avant l'attaque de la voyelle et 250 ms après celle-ci.

Lorsque l'on prend comme point d'ancrage l'attaque de la dernière syllabe, le script a la possibilité d'aligner P2 avec l'attaque de la dernière syllabe ou alors l'attaque de la syllabe précédente. Il choisit dans 98% des cas l'attaque de la dernière syllabe. Ce résultat prometteur nous a amené à tester un mode d'alignement en % de la dernière syllabe et là près de 91% des points sont alignés dans la dernière syllabe comme on peut le voir dans le graphique ci-dessous :

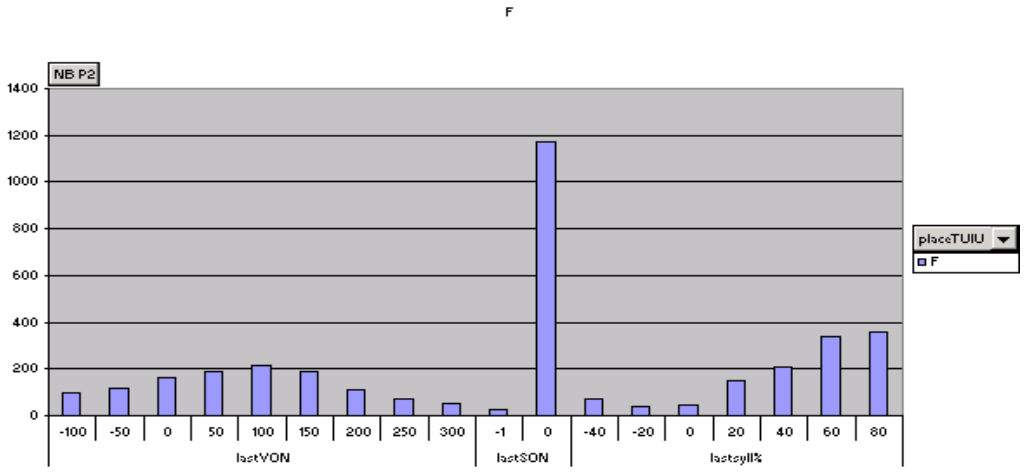


Figure 82 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position F avec point d'ancrage la dernière syllabe ou la dernière voyelle.

8.7.5.3 Alignement de P2 dans les TU en position « 1F »

Pour le cas où il y a un seul TU dans l'IU, la modélisation est identique à un TU en position finale.

Alignement de P2 par rapport à la cible précédente ou en pourcentage de l'unité tonale

Alignement par rapport à	Variations de l'alignement	Nombre de points en position 1F	En %
La cible précédente	50	87	23.20%
	100	99	26.40%
	150	75	20.00%
	200	46	12.27%
	250	38	10.13%
	300	30	8.00%
Total		375	
Pourcentage de la durée du TU	80	206	55.23%
	85	27	7.24%
	90	24	6.43%
	95	116	31.10%
Total		373	

Tableau 79 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la cible précédente ou la durée du TU.

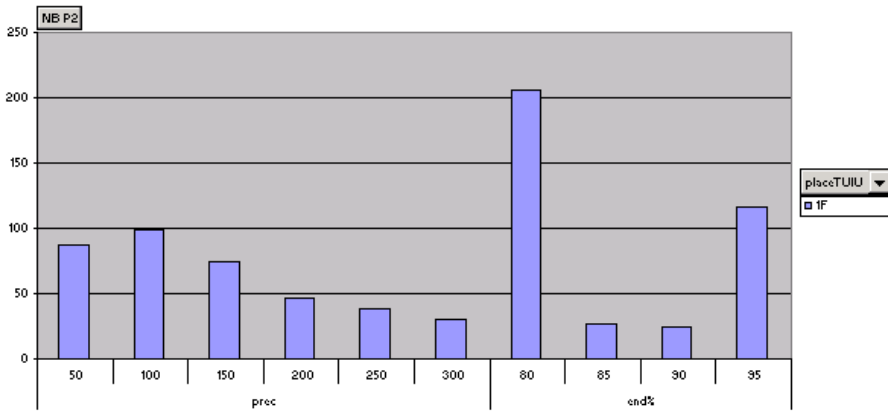


Figure 83 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la cible précédente ou un pourcentage de la durée du TU.

On voit à nouveau que les meilleurs points s'alignent à une distance allant de 50 à 200 ms de la cible précédente. En observant les variations de l'alignement en pourcentage de la tranche finale du TU, on voit ressortir deux points de repère à 80 et à 95% du TU.

Alignement de P2 par rapport à la dernière voyelle ou à la dernière syllabe

Point d'ancrage + mode d'alignement	Variation de l'alignement	Nombre de points en position 1F	En %	En % cumulé avant et à partir du point d'ancrage
Attaque de la dernière voyelle	-100	30	7.71%	
	-50	40	10.28%	17.99%
	0	34	8.74%	82.01%
	50	75	19.28%	
	100	60	15.42%	
	150	54	13.88%	
	200	46	11.83%	
	250	34	8.74%	
	300	16	4.11%	
Total		389		
Attaque de la dernière syllabe	-1	9	2.43%	2.43%
	0	362	97.57%	97.57%
Total		371		
Pourcentage de la durée de la dernière syllabe	-40	30	7.43%	
	-20	12	2.97%	10.40%
	0	17	4.21%	89.60%
	20	42	10.40%	
	40	73	18.07%	
	60	105	25.99%	
	80	125	30.94%	
Total		404		

Tableau 80 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la dernière syllabe ou la dernière voyelle.

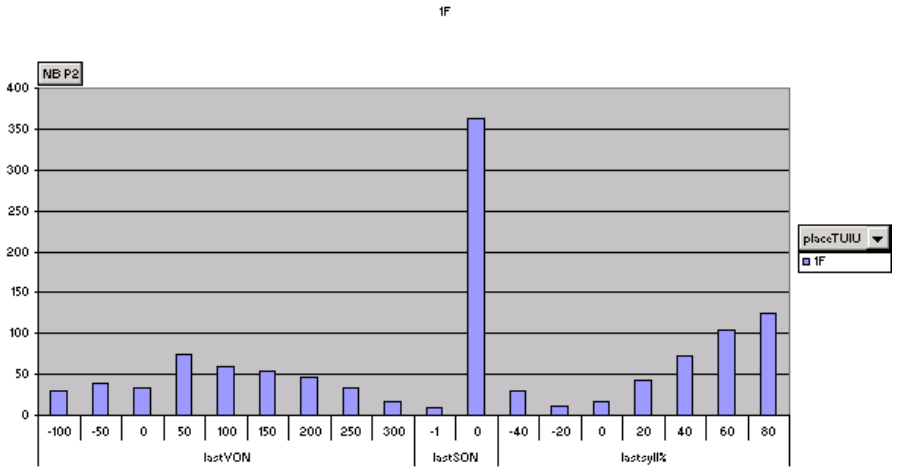


Figure 84 : Répartition des meilleurs points d'alignement pour P2 dans les TU en position 1F avec point d'ancrage la cible précédente ou un pourcentage de la durée du TU.

Les deux points de repère qui semblent les plus informatifs pour P2 sont la fin du TU ou plus précisément un alignement fixe à 10 ms de la fin du pour les TU en position 1 ou M et la dernière syllabe du TU pour les TU en position finale où P2 est un point médian.

8.7.6 Alignement de P2 en position finale en fonction du contour

Après avoir observé les résultats d'alignement de façon globale pour P2, nous observerons l'effet du type de contour pris en compte. Au niveau d'un TUF on peut avoir des contours de type « HLH », « LH », « HL » et « _L » (simplement descendants). On a vu que P2 s'aligne régulièrement en fonction de la durée de la dernière syllabe de façon globale. Voyons alors dans cette section si la configuration de F0 permet de nuancer ces résultats d'alignement.

Alignement de P2 dans les contours HL et _L par rapport à la durée de la syllabe accentuée

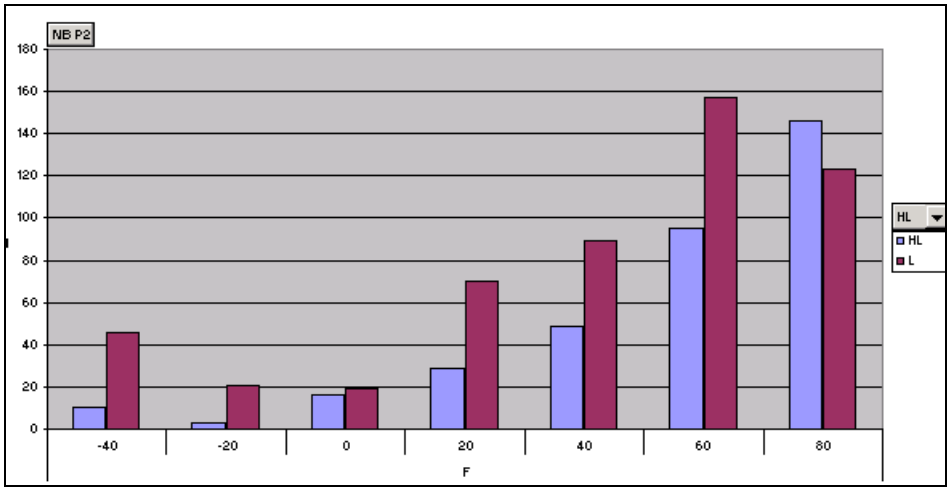


Figure 85 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage la syllabe accentuée.

Dans les contours de type « HL » et « _L», P2 s'aligne de façon identique avec la plupart des meilleurs points situés à l'intérieur de la dernière syllabe accentuée (de 20 à 80%).

Alignement de P2 dans les contours HL et _L par rapport au début de la voyelle accentuée

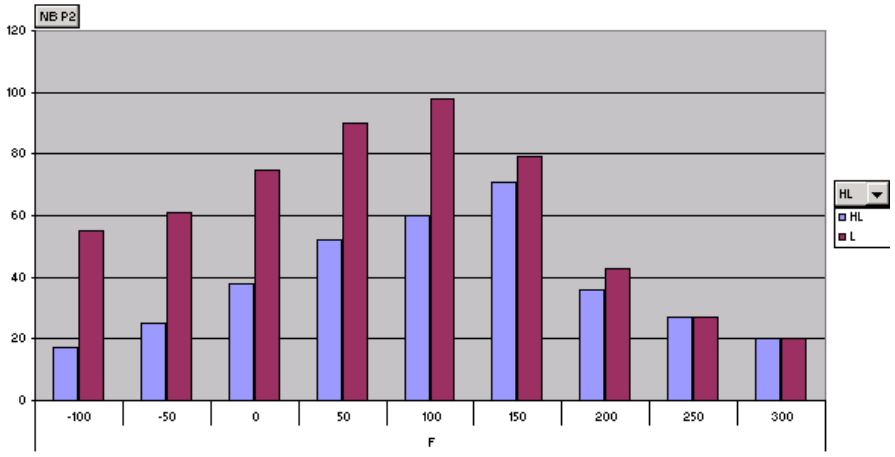


Figure 86 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage la voyelle accentuée.

Pour la plupart des contours de type « _L » et « HL », l'alignement de P2 varie de -100 à 150 ms de l'attaque de la voyelle.

Alignement de P2 dans les contours HL et _L par rapport au point cible précédent

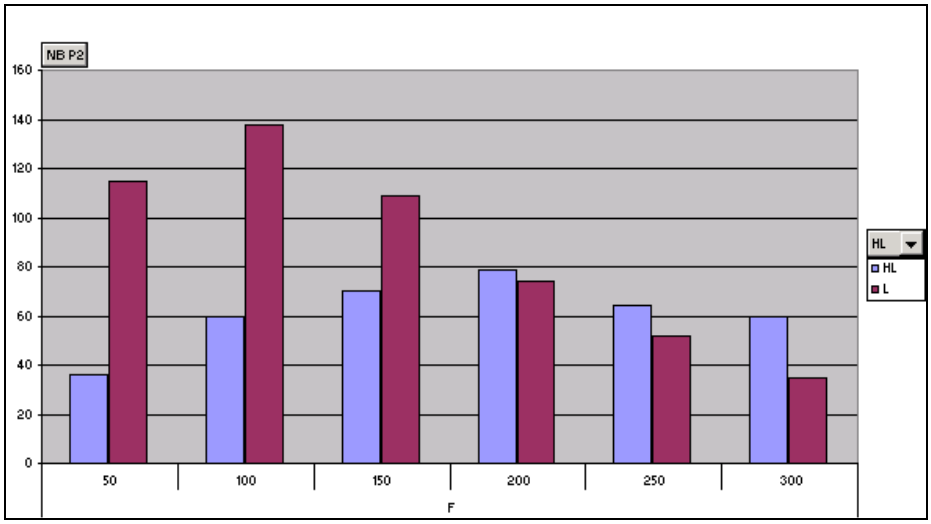


Figure 87 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage le point cible précédent.

Le point de repère du point cible précédent n'est pas précis dans le cas du contour « HL » : P2 s'aligne essentiellement de 50 à 300 ms de la cible précédente. La variation de l'alignement est légèrement plus réduite dans le cas du contour « _L », variant de 50 à 200 ms du point cible précédent.

Observons maintenant les résultats d'optimisation de l'alignement pour les contours « HLH » et « LH ».

Alignement de P2 en position F pour les configurations « HLH » et « LH » et en fonction de la durée de la syllabe accentuée

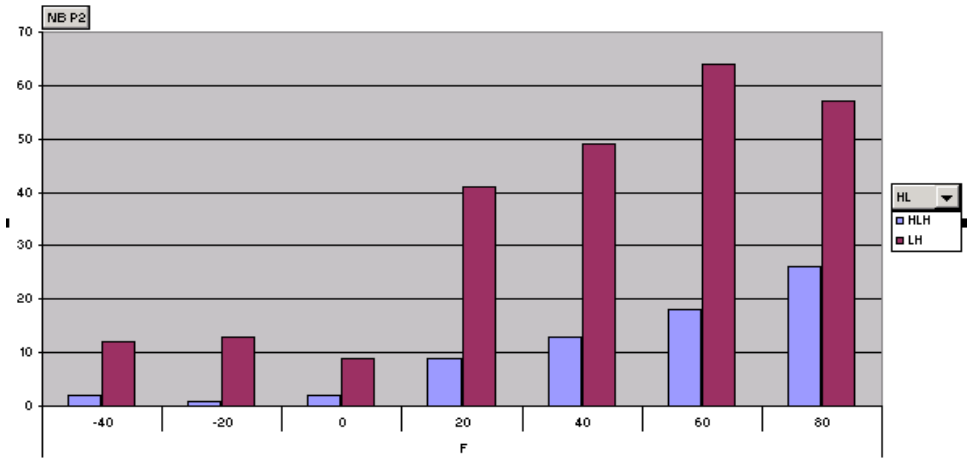


Figure 88 : Histogramme des meilleurs points d'alignement de P2 en position F en fonction du contour avec comme point d'ancrage la syllabe accentuée.

À partir du point d'ancrage de la syllabe accentuée, P2 s'aligne à nouveau le plus souvent entre 20 et 80% de la durée de la syllabe quel que soit le contour pris en compte.

Alignement de P2 en position F pour les configurations « HLH » et « LH » et en fonction de la voyelle accentuée

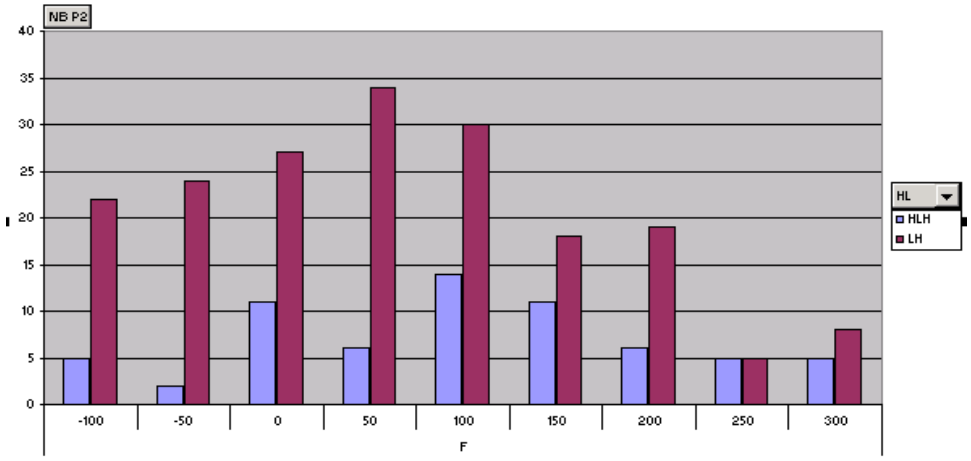


Figure 89 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour avec comme point d'ancrage la voyelle accentuée.

Pour un contour montant du type « LH », l'alignement de P2 varie de -100 à 200 ms. Dans le cas des contours « HLH », les variations sont très irrégulières et il semble difficile de dégager une tranche optimale d'alignement.

Alignement de P2 en position F pour les configurations « HLH » et « LH » et en fonction du point cible précédent

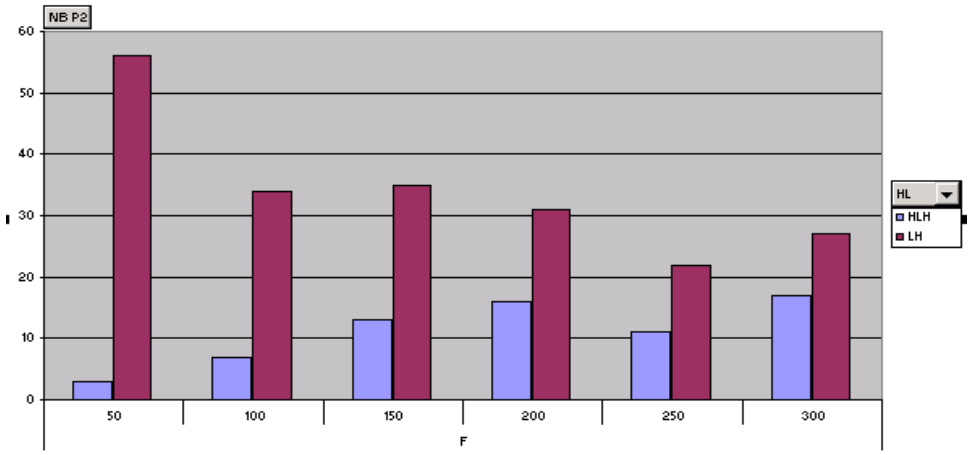


Figure 90 : Histogramme des meilleurs points d'alignement de P2 en fonction du contour et du point cible précédent.

Pour le contour « LH », l'alignement de P2 varie de 50 à 300 ms avec un nombre important de points alignés à 50 ms du point précédent. Dans les contours « HLH », P2 s'aligne un peu plus tard : de 150 à 300 ms du point cible précédent.

Dans les sections suivantes, nous présenterons l'étude des résultats d'optimisation de l'alignement dans les modèles basés sur les TTU.

8.7.7 Alignement des points cibles dans les modèles basés sur l'unité tonale élargie (TTU)

Au niveau des modèles des TTU, on peut observer les variations d'alignement de trois points cibles principaux (P1, P2 et P3) en fonction des points d'ancrage testés et des modes d'alignement.

8.7.7.1 Alignement de P1 dans les modèles des TTU

L'alignement de P1 est optimisé en pourcentage de la durée de la première syllabe accentuée, en distance en millisecondes de l'attaque de la première voyelle et distance en millisecondes du début du TTU.

Points d'ancrage	Variation de l'alignement	Nombre de points	En %	En % cumulé
Pourcentage de la durée de la première syllabe	0	1106	36.38%	36.38%
	25	667	21.94%	58.32%
	50	552	18.16%	76.48%
	75	333	10.95%	87.43%
	100	163	5.36%	92.80%
	125	99	3.26%	96.05%
	150	120	3.95%	100.00%
Total		3040		
Distance en millisecondes de l'attaque de la première voyelle	-100	329	11.00%	11.00%
	-50	581	19.43%	30.43%
	0	808	27.02%	57.46%
	50	547	18.29%	75.75%
	100	340	11.37%	87.12%
	150	195	6.52%	93.65%
	200	92	3.08%	96.72%
	250	47	1.57%	98.29%
	300	51	1.71%	100.00%
Total		2990		
Distance en millisecondes du début du TTU	10	1154	38.42%	38.42%
	60	469	15.61%	54.03%
	110	493	16.41%	70.44%
	160	424	14.11%	84.55%
	210	232	7.72%	92.28%
	260	109	3.63%	95.91%
	310	123	4.09%	100.00%
Total		3004		

Tableau 81 : Répartition des meilleurs points d'alignement pour P1 dans les TTU2 avec comme points d'ancrage la première syllabe, première voyelle ou début du TTU.

87,43% des meilleurs points se trouvent entre 0 et 75% de la syllabe accentuée

87,12 % des meilleurs points apparaissent entre -100 ms et 100 ms du début de la voyelle.

84,55% des points se trouvent entre 10 et 160 ms du début du TTU, mais on observe aussi une concentration de points très importante à 10 ms du début du TU.

Comme on peut le voir dans le graphique ci-dessous :

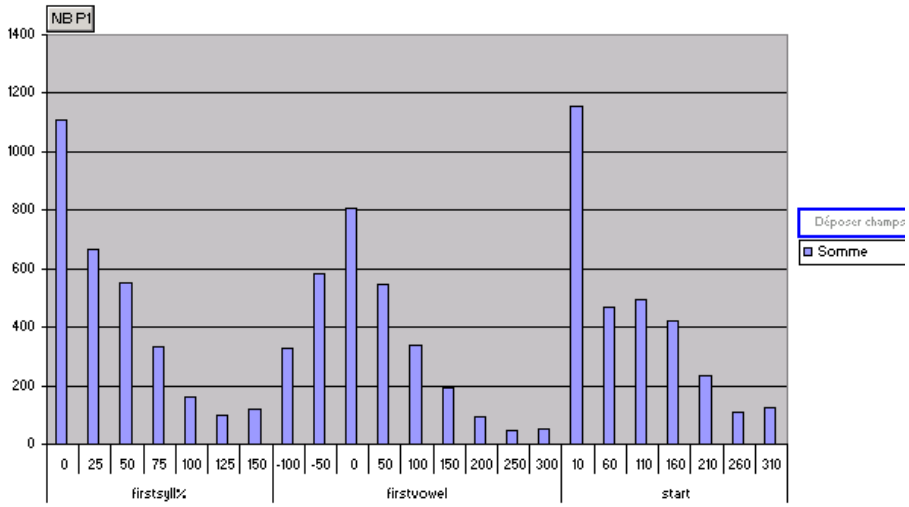


Figure 91 : Répartition des meilleurs points d'alignement pour P1 dans les TTU2 avec point d'ancrage la première syllabe, première voyelle ou début du TTU.

En observant ce graphique et dans la perspective de réduire de façon significative la variation de l'alignement des points cibles, on pourrait réduire notre intervalle de variation de 0 à 50% de la syllabe accentuée.

Le point de repère le plus précis pour l'alignement de P1 au niveau des unités tonales élargies semble à nouveau être la durée de la syllabe accentuée. Qu'en est-il de l'effet de la position du TTU sur l'alignement de P1 ?

8.7.8 Effet de la position du TTU sur l'alignement de P1 dans les modèles basés sur l'unité tonale élargie

8.7.8.1 Effet de la position du TTU sur l'alignement de P1 en % de la durée de la syllabe accentuée

Variation de l'alignement en %	1	M	F	Total
0	37.18%	32.68%	35.34%	35.58%
25	18.19%	14.60%	27.50%	21.65%
50	16.70%	18.52%	18.54%	17.83%
75	10.93%	15.03%	10.27%	11.36%
100	7.06%	7.41%	3.66%	5.62%
125	5.27%	3.70%	2.00%	3.56%
150	4.67%	8.06%	2.70%	4.40%
Total	100.00%	100.00%	100.00%	100.00%

Tableau 82 : Répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la première syllabe et de la position du TTU.

L'alignement de P1 en position finale et initiale se distingue à nouveau. Dans un TTU final, P1 est aligné généralement plus tôt : 91,64 % des points se trouvent entre 0 et 75% de la syllabe contre 83% pour un TTU « 1 » et 80,83% pour un TTU « M ».

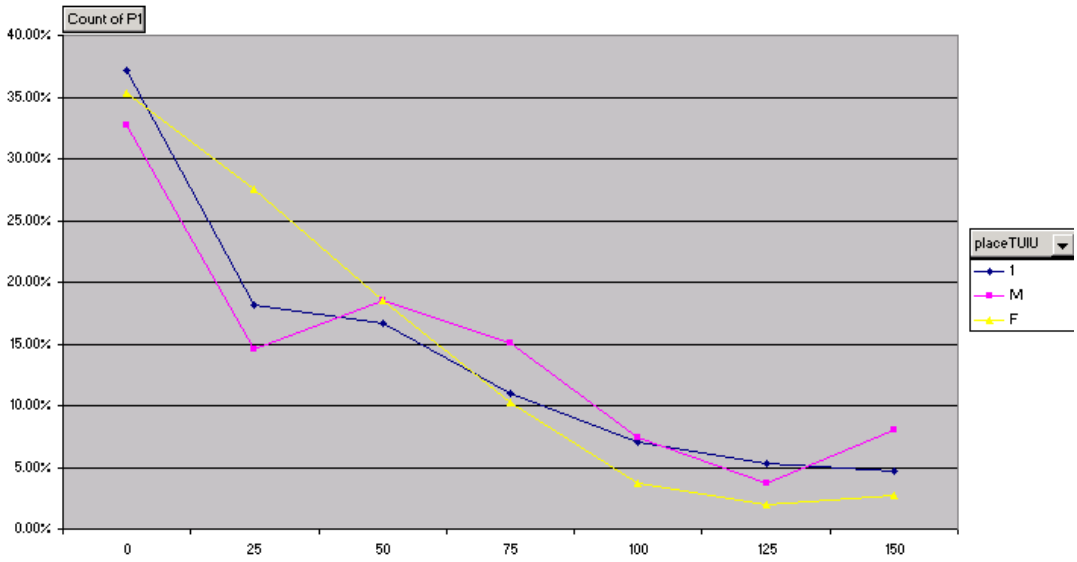


Figure 92 : Variation de l'alignement de P1 en fonction de la durée de la première syllabe et de la position du TTU.

8.7.8.2 Effet de la position du TTU sur l'alignement de P1 par rapport au début du TTU

Variation de l'alignement en ms	1	M	F	Total
10	55.20%	25.00%	29.54%	39.24%
60	10.22%	14.91%	21.04%	15.61%
110	11.34%	15.83%	21.04%	16.22%
160	10.40%	19.72%	14.41%	13.66%
210	7.22%	9.86%	7.25%	7.67%
260	2.06%	6.19%	4.12%	3.63%
310	3.56%	8.49%	2.60%	3.97%
Total	100.00%	100.00%	100.00%	100.00%

Tableau 83 : Répartition des meilleurs points d'alignement pour P1 par rapport au début du TTU et en fonction de la position du TTU.

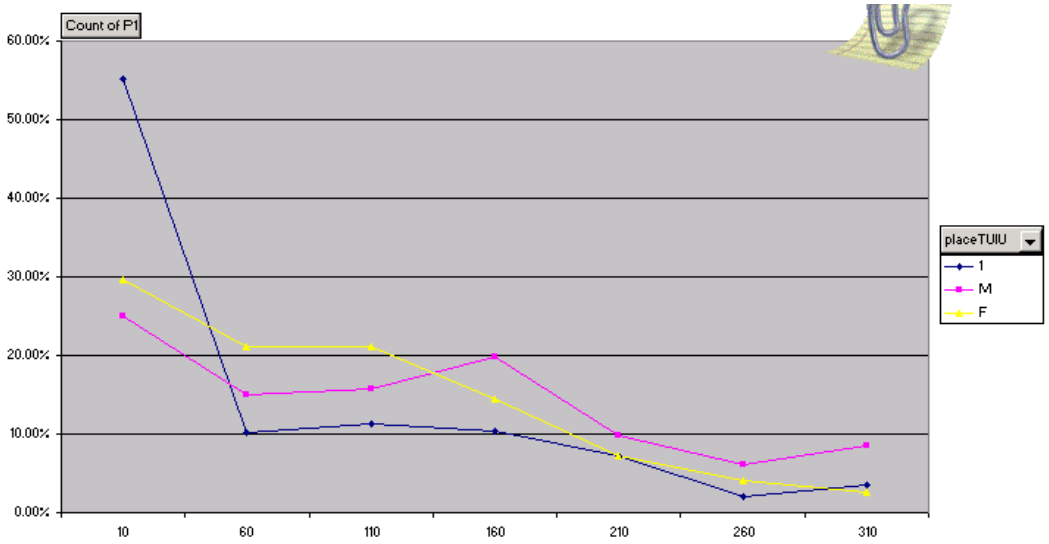


Figure 93 : Variation de l'alignement de P1 en fonction du début du TTU et en fonction de la position du TTU.

8.7.8.3 Effet de la position du TTU sur l'alignement de P1 par rapport à la voyelle

Variation de l'alignement en ms	1	M	F	Total
-100	10.86%	8.31%	12.08%	10.95%
-50	18.91%	19.78%	20.13%	19.58%
0	26.67%	25.84%	27.46%	26.87%
50	16.29%	16.40%	19.59%	17.73%
100	10.86%	12.36%	10.64%	11.03%
150	7.66%	7.19%	6.17%	6.94%
200	4.27%	3.82%	2.15%	3.28%
250	1.84%	3.37%	1.07%	1.77%
300	2.62%	2.92%	0.72%	1.85%
Total	100.00%	100.00%	100.00%	100.00%

Tableau 84 : Répartition des meilleurs points d'alignement pour P1 par rapport au début de la voyelle et en fonction de la position du TTU.

La place du TTU ne semble pas avoir d'effet sur l'alignement de P1 en fonction de la voyelle. Dans tous les cas prévus, la plupart des points s'alignent entre -100 et 100 ms de la voyelle accentuée.

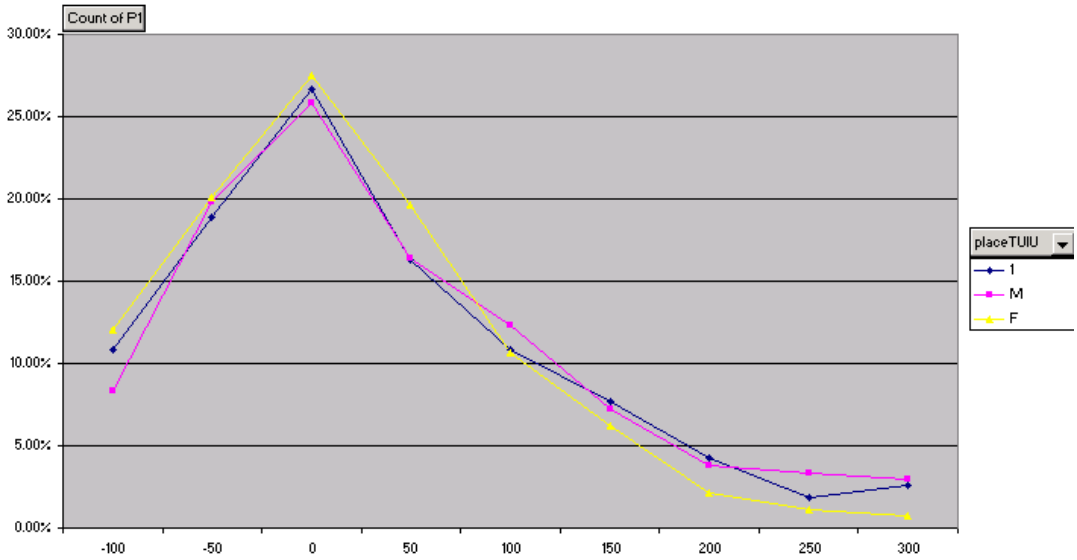


Figure 94 : Variation de l'alignement de P1 par rapport au début de la voyelle et en fonction de la position du TTU.

8.7.9 *Effet de la position du nucléus sur l'alignement de P1 dans les modèles basés sur l'unité tonale élargie*

8.7.9.1 Effet de la position du nucléus sur l'alignement de P1 en % de la durée de la syllabe accentuée

Alignement de P1	Avant le nucléus	Nucléus	Après le nucléus	Total
0	530	516	60	1106
25	251	392	24	667
50	234	301	17	552
75	157	163	13	333
100	87	71	5	163
125	61	35	3	99
150	79	34	7	120
Total	1399	1512	129	3040

Tableau 85 : Répartition des meilleurs points d'alignement pour P1 en fonction de la durée de la syllabe et de la position du nucléus.

L'alignement de P1 par rapport à la durée de la première syllabe du TTU est de plus en plus tôt alors que l'on passe d'un TTU prénucléaire à un TTU postnucléaire. Pour les TTU prénucléaires, 84% des points se trouvent entre 0 et 75 % de la syllabe, pour les TTU nucléaires et postnucléaires, 90% des points appartiennent à la même tranche.

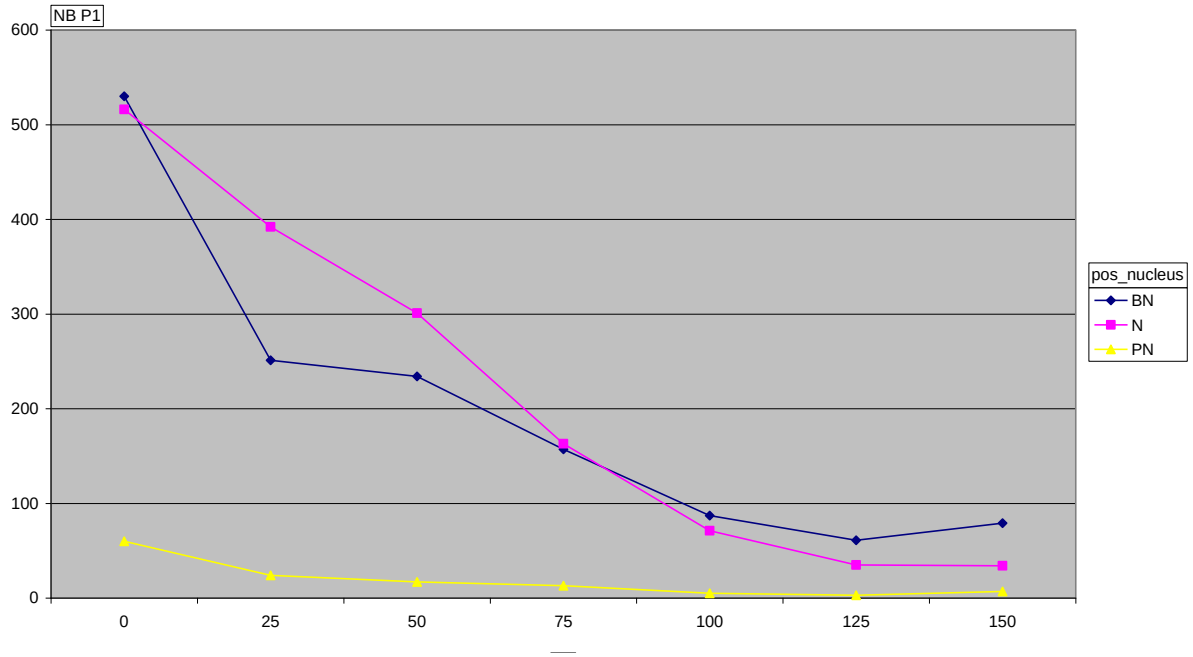


Figure 95 : Variation de l'alignement de P1 en fonction de la durée de la syllabe et de la position du nucléus.

8.7.9.2 Effet de la position du nucléus sur l'alignement de P1 par rapport à la première voyelle

Alignement de P1	Avant le nucléus	Nucléus	Après le nucléus	Total
-100	149	166	14	329
-50	284	270	27	581
0	375	401	32	808
50	234	295	18	547
100	141	180	19	340
150	95	90	10	195
200	59	29	4	92
250	32	13	2	47
300	36	12	3	51
Total	1405	1456	129	2990

Tableau 86 : Répartition des meilleurs points d'alignement pour P1 en fonction de la première voyelle et de la position du nucléus.

La plupart des meilleurs points se trouvent entre -100 ms et 100 ms de la voyelle accentuée quelle que soit la position du nucléus. C'est ce que l'on peut observer dans le graphique ci-dessous où les courbes représentant les accents prénucléaires et nucléaires sont identiques :

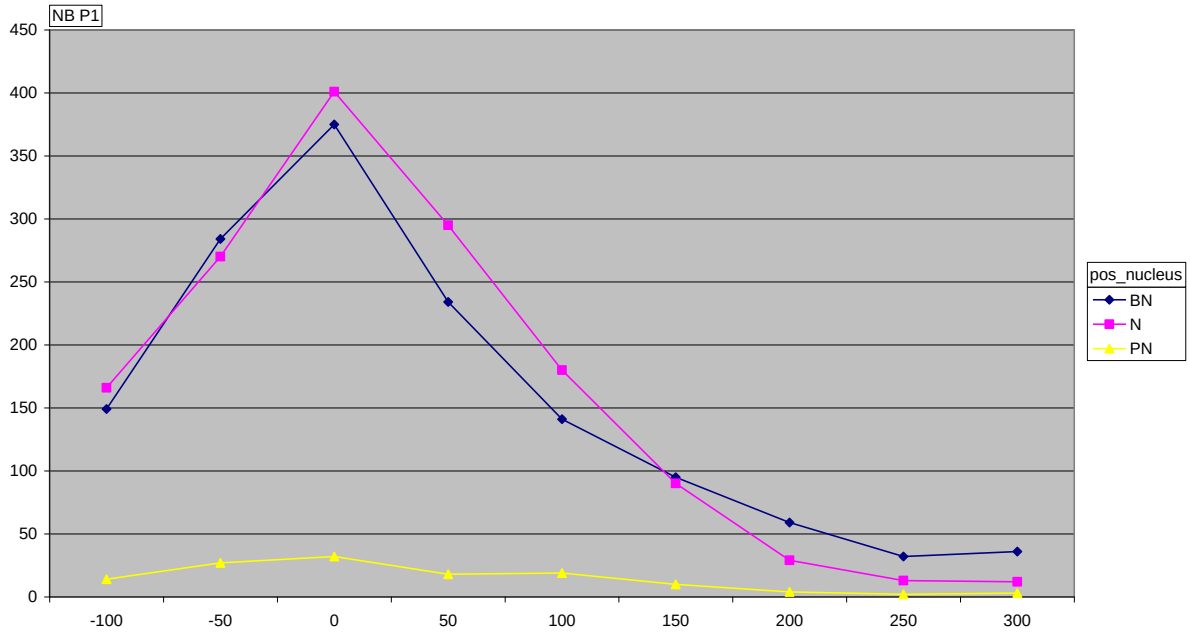


Figure 96 : Variation de l'alignement de P1 en fonction de la première voyelle et de la position du nucléus.

De façon générale, on retrouve au niveau des modèles des TTU un effet de la position du nucléus sur l'alignement de P1 (l'alignement dans les accents prénucléaires est plus tardif). Seul l'alignement par rapport à l'attaque de la voyelle ne donne pas de résultats concluants : les résultats sont identiques au niveau de l'effet de la position du TTU dans l'IU.

8.7.10 Alignement de P2 dans un TTU

8.7.10.1 Alignement de P2 en position initiale ou médiane

En position « 1 » ou « M », nous avons testé un seul point d'ancrage : la distance en millisecondes du point cible précédent.

Variation de l'alignement en ms	1	M	Total
50	26.84%	18.74%	24.30%
100	23.46%	18.74%	21.98%
150	19.28%	19.61%	19.39%
200	12.13%	15.03%	13.04%
250	8.65%	12.85%	9.97%
300	9.64%	15.03%	11.33%
Total	100.00%	100.00%	100.00%

Tableau 87 : Répartition des meilleurs points d'alignement pour P2 en fonction de la cible tonale précédente.

La plupart des points (78,70%) se trouvent à une distance variant de 50 à 200 ms de la cible précédente.

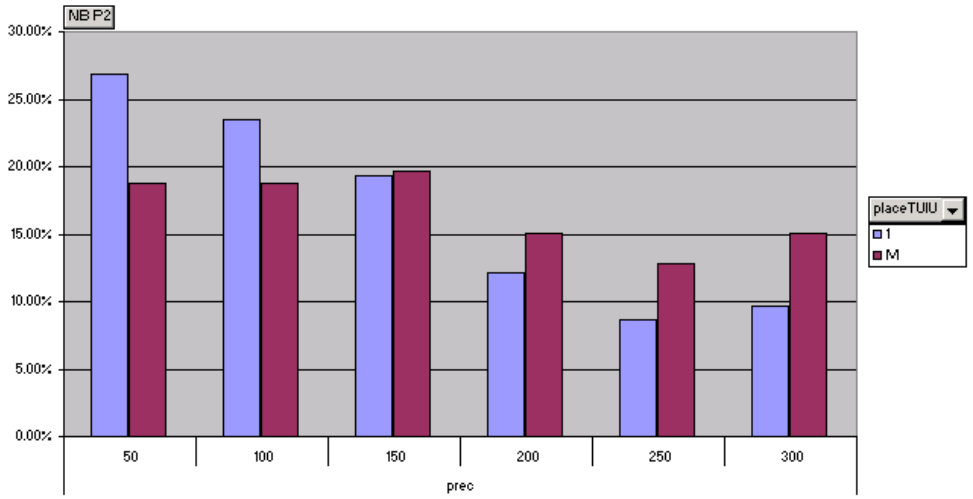


Figure 97 : Répartition des meilleurs points d'alignement pour P2 en fonction de la cible tonale précédente.

8.7.10.2 Alignement de P2 dans un TTU en position finale

L'alignement de P2 dans un TTU en position finale est optimisé en fonction de :

- la distance en nombre de syllabes de l'attaque de la dernière syllabe ;
- la durée de la dernière syllabe.

Point d'ancrage	Variation de l'alignement en Nb syll et %	Nombre de points	En %
Attaque de la dernière syllabe	-2	9	0.83%
	-1	32	2.94%
	0	1047	96.23%
Total		1088	100.00%
Durée de la dernière syllabe	-40	86	7.48%
	-20	36	3.13%
	0	50	4.35%
	20	149	12.97%
	40	216	18.80%
	60	299	26.02%
	80	313	27.24%
Total		1149	100.00%

Tableau 88 : Répartition des meilleurs points d'alignement pour P2 en fonction de la durée de la dernière syllabe du TTU.

89,38 % des meilleurs points se trouvent entre 0 et 80% de la dernière syllabe du TTU. Quand on a le choix entre un alignement avec l'attaque de la dernière syllabe, de l'avant dernière ou de l'avant avant dernière, c'est presque toujours la dernière syllabe qui est choisie comme on peut le voir dans le graphique ci-dessous :

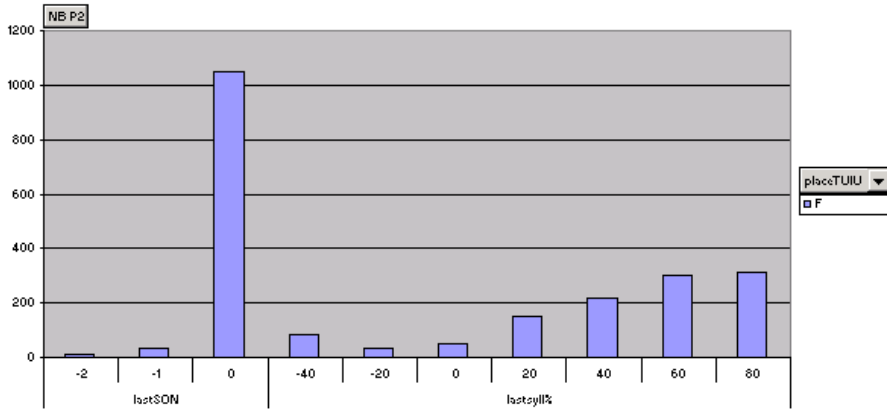


Figure 98 : Variation de l'alignement de P2 en fonction de l'attaque et de la durée de la dernière syllabe du TTU.

La durée de la dernière syllabe du TTU apparaît à nouveau comme un point de repère très informatif.

8.7.11 Alignement de P3 dans un TTU

8.7.11.1 Alignement de P3 dans un TTU en position initiale ou médiane

Dans les TTU en position « 1 » ou « M », on a cherché le point le plus près possible de la frontière qui donne un bon résultat. Deux positions possibles étaient prises en considération :

- soit la fin du TTU (variation de l'alignement exprimé en millisecondes) ;
- soit la fin de la dernière syllabe du TTU, soit la fin de l'avant dernière syllabe (variation de l'alignement exprimé en nombre de syllabes).

Point d'ancrage	Variation de l'alignement en ms et Nb syll	Nombre de points	En %
Fin du TTU	-160	99	6.76%
	-110	74	5.05%
	-60	101	6.89%
	-10	1191	81.30%
Total		1465	100.00%
Fin de la dernière syllabe	-2	78	5.36%
	-1	140	9.62%
	0	1237	85.02%
Total		1455	100.00%

Tableau 89 : Répartition des meilleurs points d'alignement pour P3 en fonction de la fin du TTU et de la fin de la dernière syllabe du TTU.

On a à nouveau une préférence nette pour un alignement très proche de la fin du TTU (à 10 ms ou fin de la dernière syllabe).

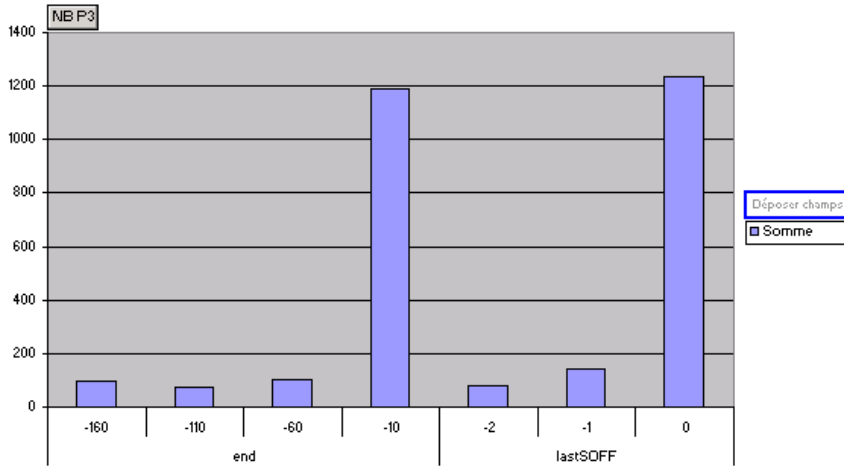


Figure 99 : Variation de l'alignement de P3 en fonction de la fin du TTU et de la fin de la dernière syllabe du TTU.

Trois points d'ancrage principaux sont mis en valeur à l'issue de l'observation des variations de l'alignement de P1, P2 et P3 dans les TTU. Tout d'abord comme dans les modèles précédents, l'alignement de P1 peut être cadré assez précisément en fonction de la durée de la syllabe accentuée. P2 s'aligne en général entre 20 et 80 % de la dernière syllabe du TTU et P3 à 10 ms de la fin du TTU.

8.7.12 Conclusion provisoire

On peut résumer les résultats des expériences ci-dessus en dégageant quelques tendances d'alignement pour les deux points cibles principaux**.**

P1 s'aligne la plupart du temps de façon régulière en fonction de la durée de la première syllabe, en pourcentage. La durée totale de l'unité tonale ne donne pas un point de repère précis d'après nos expériences. Les points de repère fixes testés sont la voyelle accentuée et le début de l'unité tonale, on obtient alors des variations de location des meilleurs points d'alignement nettement supérieures.

Pour P2, on a testé l'alignement par rapport à la durée du TU, à la durée de la dernière syllabe. Puis en valeur absolue : la distance par rapport à la frontière droite, l'attaque de la dernière voyelle et la distance par rapport au point cible précédent. On en retient plusieurs points de repère précis : la distance par rapport à la fin de l'unité tonale (avec une préférence pour un alignement à 10 ms de celle-ci) dans les TU en position initiale et médiane et la durée de la dernière syllabe (entre 20 et 80% de celle-ci) pour les TU en position finale.

L'effet de la position du TU semble visible dans la plupart des expériences et avec tous les points d'ancrage testés : on a une différence entre les TU initiaux et médians et les TU en position finale. L'alignement se fait plus tardivement dans les deux premiers cas. Cet effet est moins visible quand le point d'ancrage choisi est la voyelle.

L'opposition entre les accents prénucléaires et postnucléaires semble montrer le même effet. On a un alignement très variable en position prénucléaire et un alignement plus restreint et plus anticipé en position postnucléaire.

L'observation des résultats de l'optimisation de l'alignement en fonction de la configuration permet d'approfondir nos résultats. Au niveau de P1 et pour un contour de type « HL » les variations de l'alignement sont identiques aux résultats globaux, son alignement varie essentiellement de 0 à 75% de la syllabe accentuée et l'on observe un alignement plus tôt dans la syllabe en position finale. Pour les contours de type « LH » on trouve un alignement plus précis au début de la syllabe accentuée surtout pour les TU en position finale.

L'alignement de P2 dans les contours de type « HLH » est aussi mieux analysé en fonction de la durée de la dernière syllabe d'un TU qu'en fonction du point cible précédent.

Les modèles basés sur une annotation en unité tonale globale montrent les mêmes résultats.

Ces résultats pointent vers le modèle 11 qui contient les points d'ancrage suivants : la durée de la syllabe accentuée, la distance par rapport à la fin du TU et la durée de la dernière syllabe du TU.

Voyons ensuite les résultats de ces modèles en termes de score, corrélation et écart quadratique moyen.

8.7.13 Évaluation objective des modèles d'optimisation de l'alignement

Après avoir évalué les différents modes d'alignement et points d'ancrage du point de vue des variations chiffrées, nous évaluerons les résultats de l'optimisation de l'alignement en fonction de trois critères : la corrélation, l'écart quadratique moyen et le score. Cela nous permettra de trouver le modèle qui reproduit au mieux les contours de F0. Nous nous concentrerons sur les modèles qui regroupent tous les types d'alignement testés pour un même codage INTSINT en comparant donc les modèles allant de 6 à 14, 15 et 16 puis de 20 à 24.

Observons d'abord les corrélations moyennes obtenues par chaque modèle de façon globale et en fonction du type d'IU.

Correlations moyennes par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
intalign11	0.87127	intalign11	0.87884	intalign11	0.87528
intalign14	0.86605	intalign14	0.87736	intalign14	0.87419
intalign13	0.86182	intalign13	0.87263	intalign12	0.87200
intalign12	0.85908	intalign12	0.87128	intalign13	0.86954
intalign6	0.84018	intalign6	0.85620	intalign10	0.84720
intalign24	0.83830	intalign24	0.85225	intalign24	0.84181
intalign10	0.83697	intalign9	0.84484	intalign6	0.83869
intalign9	0.83670	intalign8	0.84215	intalign8	0.83110
intalign8	0.83361	intalign23	0.84008	intalign9	0.83013
intalign7	0.82726	intalign7	0.82493	intalign7	0.82967
intalign23	0.81767	intalign22	0.82212	intalign23	0.81782
intalign21	0.81018	intalign21	0.82193	intalign15	0.81561
intalign16	0.80765	intalign20	0.81961	intalign20	0.81372
intalign20	0.80764	intalign16	0.81858	intalign22	0.80989
intalign22	0.80695	intalign10	0.81240	intalign16	0.80742
intalign15	0.80507	intalign15	0.80837	intalign21	0.80663

Tableau 90 : Classement des modèles en fonction de la corrélation moyenne et du type d'IU.

Le modèle 11 obtient pour chaque type d'IU, la meilleure moyenne de corrélation.

Ecart quadratique moyen par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
intalign14	1.97467	intalign14	1.84866	intalign11	2.02416
intalign11	1.97964	intalign13	1.88051	intalign14	2.02989
intalign12	1.99660	intalign11	1.88365	intalign12	2.05717
intalign13	1.99929	intalign12	1.90644	intalign13	2.08165
intalign24	2.07577	intalign24	1.90994	intalign24	2.13982
intalign23	2.14898	intalign23	1.99587	intalign10	2.18186
intalign10	2.17417	intalign20	2.02623	intalign23	2.24922
intalign6	2.18725	intalign21	2.02801	intalign6	2.28559
intalign15	2.21664	intalign22	2.03901	intalign15	2.32708
intalign20	2.21808	intalign15	2.04175	intalign20	2.34939
intalign21	2.25301	intalign6	2.06273	intalign8	2.35354
intalign8	2.27257	intalign16	2.07412	intalign7	2.35986
intalign22	2.27406	intalign10	2.07923	intalign9	2.39876
intalign16	2.27480	intalign9	2.22454	intalign21	2.42014
intalign9	2.33451	intalign8	2.22608	intalign16	2.42961
intalign7	2.34219	intalign7	2.24356	intalign22	2.46778

Tableau 91 : Classement des modèles en fonction de l'écart quadratique moyen et du type d'IU.

L'écart quadratique moyen le plus faible est obtenu par le modèle 14 suivi par le modèle 11.

Scores moyens par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
intalign11	9.50480	intalign11	9.62377	intalign11	9.54234
intalign14	9.43799	intalign14	9.58904	intalign14	9.51654
intalign13	9.40284	intalign13	9.55893	intalign12	9.50065
intalign12	9.37967	intalign12	9.52745	intalign13	9.46464
intalign6	9.12091	intalign6	9.32587	intalign10	9.21900
intalign24	9.12036	intalign24	9.30469	intalign24	9.13683
intalign10	9.10900	intalign23	9.19215	intalign6	9.07811
intalign9	9.03620	intalign9	9.18063	intalign8	8.96844
intalign8	9.02869	intalign8	9.15575	intalign7	8.94950
intalign7	8.93498	intalign21	9.00247	intalign9	8.94474
intalign23	8.90959	intalign22	8.98746	intalign23	8.88884
intalign21	8.80733	intalign20	8.96011	intalign15	8.82952
intalign20	8.78295	intalign16	8.95672	intalign20	8.80325
intalign22	8.78077	intalign7	8.94772	intalign22	8.77941
intalign16	8.77270	intalign10	8.89619	intalign21	8.72926
intalign15	8.75692	intalign15	8.85123	intalign16	8.72477

Tableau 92 : Classement des modèles en fonction du score moyen du type d'IU.

On peut remarquer que les meilleurs résultats quel que soit le critère objectif pris en compte, sont obtenus avec le modèle 11. De façon générale les résultats sont serrés, mais on a une tendance générale à avoir de meilleurs résultats avec une optimisation de l'alignement au niveau de l'unité tonale par rapport à l'unité tonale élargie, les TTU2 étant meilleurs que les TTU1. Ce sont donc les paramètres d'alignement du modèle 11 qui sont à nouveau mis en valeur c'est-à-dire l'alignement en fonction de la durée de la syllabe accentuée, la durée de la dernière syllabe de l'unité tonale et le point d'ancrage fixe de la frontière gauche du TU.

À partir de l'observation des résultats de cette étape d'optimisation large de l'alignement des trente meilleures séquences de tons, les paramètres des modèles finaux sont générés. Tout d'abord, les dix meilleures séquences retenues par TU ou TTU à l'issue de cette seconde étape, sont automatiquement conservées pour les prochaines étapes. Au niveau de l'alignement, on réduit les paramètres de variation en fonction des résultats observés. Au niveau du point d'ancrage « durée de la première syllabe », P1 peut maintenant varier de 0 à 70% de la durée de la syllabe en trois itérations. P2 est systématiquement aligné à 10 ms de la fin du TU etc. On obtient donc un modèle simplifié en termes de séquences de tons et de variations de l'alignement (cf. annexes 3, p.429 à 433 pour les paramètres en détails). Cette simplification est le résultat de l'optimisation des représentations formelles par analyse par synthèse.

Nous avons aussi représenté les résultats de ces modèles finaux sous forme de grammaires que l'on présentera dans la prochaine sous-section.

C'est à partir de ces modèles finaux que l'on procédera à l'optimisation des représentations fonctionnelles de la même façon qu'à partir du corpus EUROM1.

8.8 Évaluation des modèles finaux

Les modèles finaux sont donc basés sur les dix meilleures séquences sélectionnées à partir des deux premières étapes d'optimisation et une variation réduite de l'alignement. On peut à nouveau comparer ces différents modèles en calculant pour chacun la valeur moyenne de la corrélation, de l'écart quadratique moyen et du score. Nous présenterons d'abord les résultats pour tous les modèles codés avec INTSINT et ensuite les modèles codés avec les variantes d'INTSINT :

Correlations moyennes par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
intmodel13	0.81499	intmodel11	0.82516	intmodel12	0.82509
intmodel12	0.81304	intmodel13	0.82364	intmodel13	0.82476
intmodel14	0.81295	intmodel14	0.82106	intmodel14	0.82391
intmodel11	0.81052	intmodel12	0.81781	intmodel11	0.81838
intmodel24	0.79463	intmodel24	0.80404	intmodel24	0.80379
intmodel10	0.78153	intmodel22	0.79145	intmodel10	0.79305
intmodel6	0.77513	intmodel23	0.78023	intmodel6	0.78491
intmodel22	0.77364	intmodel21	0.77218	intmodel7	0.77733
intmodel23	0.77069	intmodel6	0.76971	intmodel23	0.77610
intmodel7	0.76963	intmodel16	0.76750	intmodel22	0.77587
intmodel8	0.76867	intmodel8	0.76691	intmodel21	0.77439
intmodel9	0.76578	intmodel7	0.76549	intmodel16	0.77373
intmodel21	0.76410	intmodel9	0.76500	intmodel8	0.77273
intmodel16	0.76106	intmodel20	0.76449	intmodel9	0.77009
intmodel20	0.75529	intmodel15	0.76374	intmodel20	0.76846
intmodel15	0.75405	intmodel10	0.75804	intmodel15	0.76796

Tableau 93 : Classement des modèles en fonction de la corrélation moyenne.

Ecart quadratique moyen par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
intmodel14	2.25347	intmodel14	2.11662	intmodel24	2.32240
intmodel13	2.25894	intmodel11	2.13787	intmodel13	2.32753
intmodel12	2.27467	intmodel24	2.15681	intmodel12	2.33290
intmodel11	2.29886	intmodel13	2.16198	intmodel14	2.33406
intmodel24	2.30343	intmodel12	2.19572	intmodel11	2.37735
intmodel22	2.37545	intmodel22	2.19592	intmodel22	2.50769
intmodel23	2.43777	intmodel23	2.27974	intmodel23	2.54420
intmodel20	2.51266	intmodel6	2.34063	intmodel10	2.56090
intmodel15	2.51534	intmodel21	2.34694	intmodel20	2.58429
intmodel21	2.51915	intmodel20	2.35286	intmodel15	2.58581
intmodel6	2.53105	intmodel16	2.36561	intmodel21	2.62668
intmodel10	2.54645	intmodel15	2.36705	intmodel6	2.66021
intmodel16	2.55278	intmodel7	2.39793	intmodel7	2.67099
intmodel7	2.55795	intmodel10	2.40970	intmodel16	2.67847
intmodel8	2.75896	intmodel8	2.69029	intmodel9	2.85841
intmodel9	2.76973	intmodel9	2.74888	intmodel8	2.89533

Tableau 94 : Classement des modèles en fonction de l'écart quadratique moyen.

Scores moyens par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
intmodel13	8.82494	intmodel11	8.98819	intmodel14	8.90920
intmodel14	8.81706	intmodel13	8.96082	intmodel12	8.90290
intmodel12	8.80636	intmodel14	8.94013	intmodel13	8.89511
intmodel11	8.78181	intmodel12	8.89877	intmodel11	8.84385
intmodel24	8.59378	intmodel24	8.75585	intmodel24	8.66927
intmodel10	8.43346	intmodel22	8.60454	intmodel10	8.53697
intmodel22	8.37834	intmodel23	8.48897	intmodel6	8.42370
intmodel6	8.35910	intmodel6	8.38318	intmodel22	8.37403
intmodel23	8.31247	intmodel21	8.37847	intmodel7	8.34585
intmodel7	8.29465	intmodel16	8.33012	intmodel21	8.33492
intmodel21	8.25477	intmodel7	8.32361	intmodel23	8.32275
intmodel8	8.23594	intmodel20	8.28305	intmodel20	8.29494
intmodel16	8.20970	intmodel15	8.27986	intmodel16	8.29356
intmodel9	8.19932	intmodel10	8.25959	intmodel15	8.29127
intmodel20	8.16430	intmodel8	8.25834	intmodel8	8.24100
intmodel15	8.15781	intmodel9	8.23325	intmodel9	8.20815

Tableau 95 : Classement des modèles en fonction du score moyen.

Dans l'ensemble, on observe une meilleure modélisation au niveau des modèles basés sur les unités tonales simples, mais les résultats sont plus serrés qu'à partir de l'optimisation de l'alignement. Au niveau des modèles basés sur les unités tonales, les modèles 13, 14, 12 et 11 se détachent. Cependant les modèles 13 et 14 basés sur un alignement par rapport à la voyelle, contiennent des paramètres de variations nettement plus larges que les deux autres modèles basés

sur l'unité tonale, la durée de la syllabe accentuée et celle de la dernière syllabe accentuée. De façon plus globale et en matière de précision, les modèles 11 et 12 semblent être les plus intéressants.

Les modèles basés sur les unités tonales élargies viennent en tête dans un cas précis la modélisation des IU emphatiques terminaux.

Le meilleur modèle au niveau des TTU est le modèle 24 qui réunit les mêmes paramètres d'alignement que le modèle 11. Les modèles basés sur le TTU simple (modèles 15 et 16) sont à nouveau moins performants dans l'ensemble.

Observons ensuite les résultats de modélisation obtenus à partir des variantes d'INTSINT en comparant pour une même étendue maximale (maximum 1.4) les modèles codés avec INTSINT1, 2 et 3.

Correlations moyennes par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
int2model42	0.81384	int2model42	0.82444	int3model62	0.82136
int3model62	0.81062	int3model62	0.81594	int2model42	0.81990
int1model52	0.80230	int1model52	0.80983	int1model52	0.81916
int3model66	0.77724	int1model56	0.77826	int2model46	0.78759
int2model46	0.77396	int3model66	0.77676	int3model66	0.78722
int1model56	0.77117	int2model46	0.77660	int1model56	0.77930

Tableau 96 : Classement des modèles basés sur les variantes d'INTSINT en fonction de la corrélation moyenne.

Ecart quadratique moyen par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
int3model62	2.23803	int3model62	2.13393	int3model62	2.28610
int2model42	2.27138	int1model56	2.19719	int2model42	2.33010
int3model66	2.31966	int2model42	2.20387	int3model66	2.36047
int2model46	2.32188	int3model66	2.21920	int2model46	2.38402
int1model56	2.33621	int2model46	2.23047	int1model56	2.40766
int1model52	2.56363	int1model52	2.47865	int1model52	2.69006

Tableau 97 : Classement des modèles basés sur les variantes d'INTSINT en fonction de l'écart quadratique moyen .

Scores moyens par modèle					
Tous type d'IU		Non terminal [+		Terminal [|	
int2model42	8.82087	int2model42	8.97266	int3model62	8.91368
int3model62	8.80480	int3model62	8.88852	int2model42	8.85985
int1model52	8.57632	int1model52	8.69036	int1model52	8.69573
int3model66	8.44010	int1model56	8.52508	int2model46	8.53857
int2model46	8.41128	int3model66	8.46568	int3model66	8.53007
int1model56	8.38209	int2model46	8.45159	int1model56	8.42971

Tableau 98 : Classement des modèles basés sur les variantes d'INTSINT en fonction du score moyen.

Le modèle qui donne les meilleurs résultats est le modèle 42 basé sur un codage avec INTSINT2 et une annotation au niveau des TU. Ce modèle est suivi de très près par le modèle 62 codé avec INTSINT3 et basé sur les TU. Les codages INTSINT2 et 3 permettent dans l'ensemble une meilleure modélisation que leurs équivalents exacts codés avec INTSINT d'origine (modèles 52 et 56).

Prenons quelques exemples de passages modélisés au niveau des TU, TTU1 et TTU2 pour illustrer ces résultats.

La reproduction des contours de F0 est plus précise dans les modèles basés sur les unités tonales comme on peut le voir dans l'exemple suivant :

Exemple tiré du fichier G0301B (temps 13,01 à 14,64) :

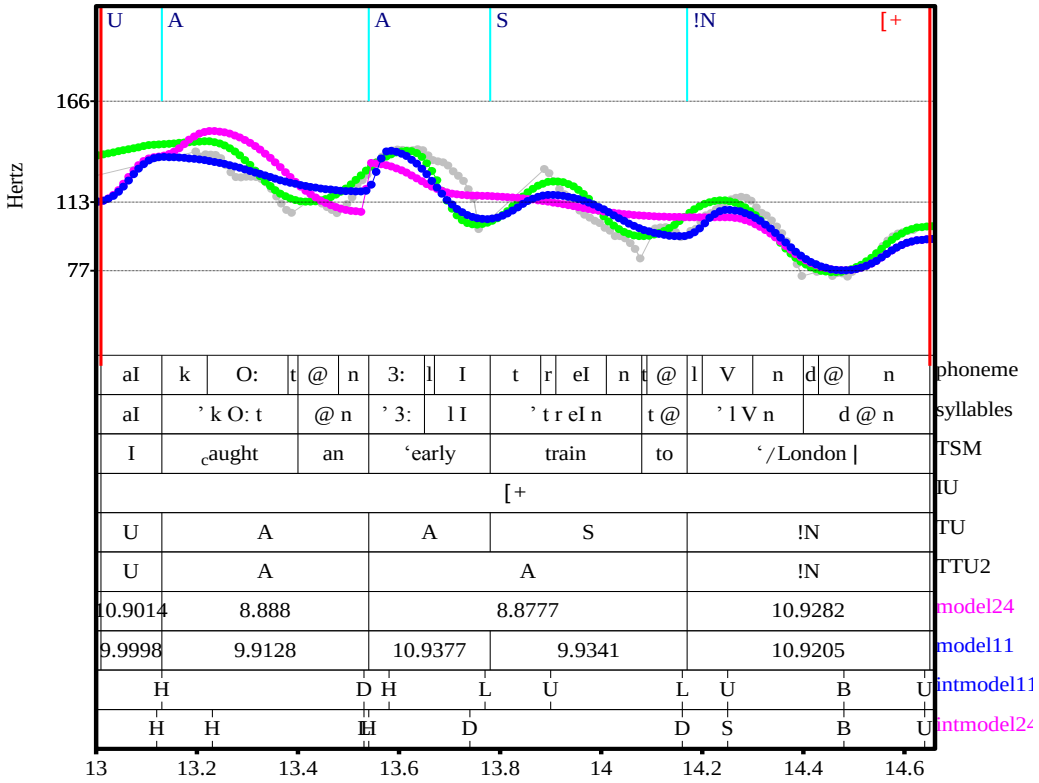


Figure 100 : Courbe générée à partir des modèles 11 et 24 pour un passage de MARSEC.

> Courbe verte= MOMEL d'origine, courbe bleue= courbe générée à partir du modèle 11, courbe rose= générée à partir du modèle 24.
>
> Rangée 1= phonèmes, rangée 2= syllabes, rangée 3= TSM, rangée 4= IU, rangée 5= TU, rangée 6= TTU2, rangée 7= valeurs du score pour le modèle 24, rangée 8= valeurs du score pour le modèle 11, rangée 9= séquences de tons INTSINT optimisées dans le modèle 11, rangée 10= séquences de tons INTSINT optimisées dans le modèle 24.

La courbe bleue représente le modèle 11 et suit de près la courbe verte qui est la courbe de MOMEL d'origine présente dans les données. Quand on se base sur les unités tonales élargies, on ne distingue pas les accents de type « S » ou « A ». A partir du modèle 11 les TU « *early* » et

« *train* » sont modélisés séparément par deux chutes de F0 : {HL} et {UL}. Le contour obtenu à partir de l'unité élargie est généralement descendant et la seconde montée + chute est escamotée, on a un contour {HDD}. On voit au niveau des rangées 7 et 8 que le score est meilleur pour le modèle 11. C'est ce type d'exemples qui explique la meilleure performance des modèles des TU.

Les modèles basés sur les TTU1 et les TTU2 se distinguent essentiellement par la façon de modéliser le nucléus dans les IU non terminaux. On obtient une meilleure modélisation quand le dernier accent mélodique le cas échéant est modélisé séparément.

Exemple extrait du fichier G0302B de 16,2 à 18 :

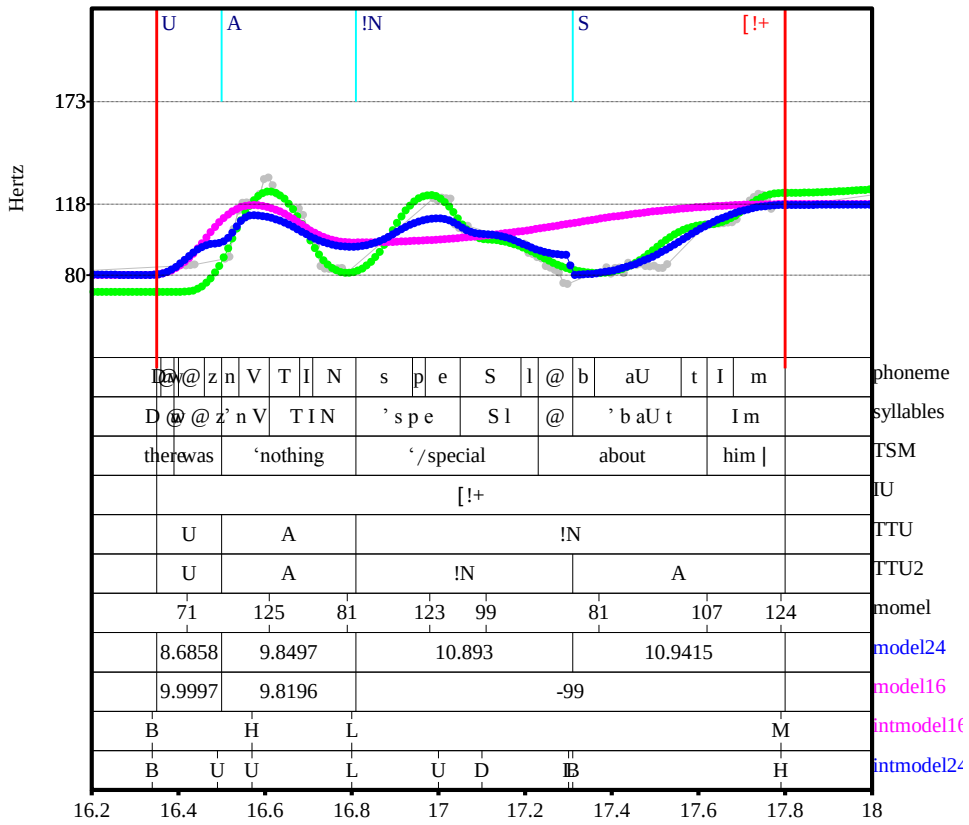


Figure 101 : Modélisation d'un contour de chute+montée à partir des TT1 (modèle 16) et TTU2 (modèle 24).

Courbe verte= MOMEL d'origine, courbe bleue= courbe générée à partir du modèle 24, courbe rose= générée à partir du modèle 16.

Rangée 1= phonèmes, rangée 2= syllabes, rangée 3= TSM, rangée 4= IU, rangée 5= TTU, rangée 6= TTU2, rangée 7= MOMEL, rangée 8= valeurs du score pour le modèle 24, rangée 9= valeurs du score pour le modèle 16, rangée 10= séquences de tons INTSINT optimisées dans le modèle 16, rangée 11= séquences de tons INTSINT optimisées dans le modèle 24.

On peut voir que la meilleure modélisation est obtenue à partir du modèle 24 représenté par la courbe bleue et la séquence INTSINT {UDBH} visible sur la dernière rangée. Le nombre –99 au niveau de la rangée, montrant le score pour le modèle 16, indique que l'on n'a pas pu proposer une modélisation correcte avec les paramètres réduits.

La section suivante porte sur les grammaires représentatives développées à partir des modèles.

8.9 Grammaires formes/fonctions

À partir des résultats de l'optimisation des formes en matière de codage INTSINT et de l'alignement des points cibles, plusieurs grammaires ont été élaborées. En retenant donc pour les modèles testés, cinq meilleures séquences par fonction et une variation d'alignement réduite, on établit une représentation visuelle des résultats modèle par modèle. Au niveau d'un modèle donné, il s'agit de considérer chaque type d'unité intonative et à l'intérieur de celle-ci, de représenter l'agencement des fonctions ou types de TU possibles ainsi que les formes optimisées qui leur sont associées à partir des données du corpus. On peut ainsi visualiser l'ensemble des contours mélodiques prédits par nos modèles ainsi que la structure fonctionnelle des énoncés de façon générale. C'est une autre façon d'observer les résultats d'optimisation des représentations formelles par analyse par synthèse. Nous prendrons les exemples du modèle 11 (basé sur l'annotation en TU) dans les IU « [| » et « [+ », « [!| » et « [!+ », « [?+ » et « [?| » et le modèle 24 (basé sur une annotation en TTU). Les grammaires générées à partir des autres modèles sont à consulter dans les annexes.

8.9.1 Grammaires à partir du modèle 11

Rappelons que ce modèle est basé sur une annotation en unités tonales et un alignement basé sur la durée de la première syllabe et de la dernière syllabe du TU ainsi que les frontières droites et gauches de celui-ci. C'est le modèle qui obtient les meilleurs scores dans la plupart des cas.

Observons d'abord les grammaires des IU « [+ » et « [| ».

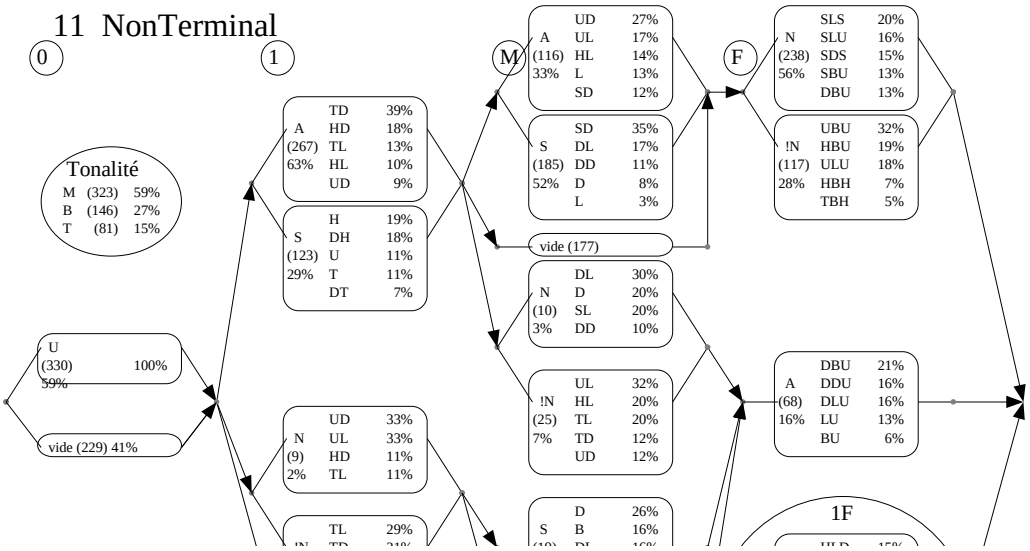


Figure 102 : Grammaire des IU de type non terminal pour le modèle 11.

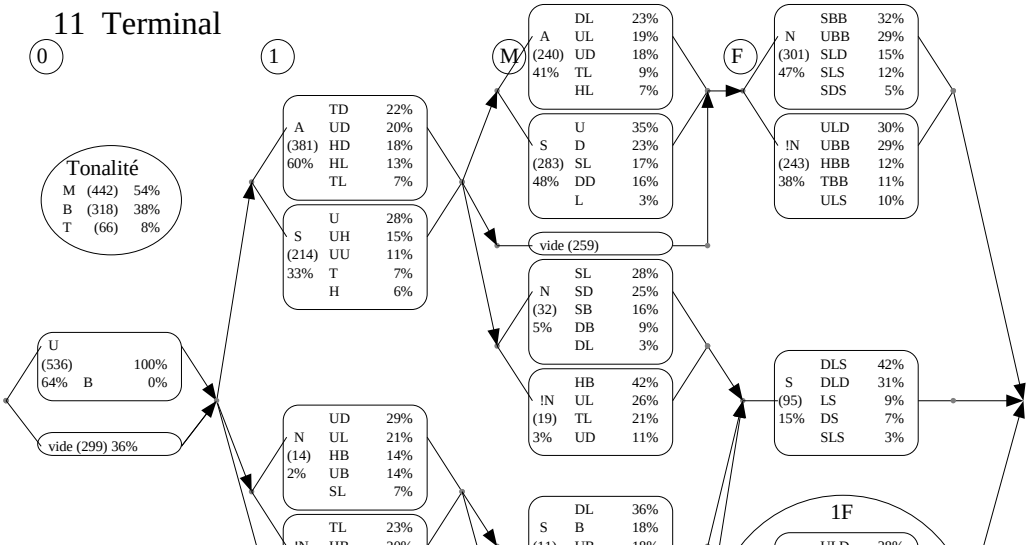


Figure 103 : Grammaire des IU de type terminal pour le modèle 11.

Les rectangles représentent des types de TU (A, S, N, !N). Ceux-ci sont représentés dans l'espace en fonction de leur position qui peut être initiale, médiane, finale ou finale et initiale (1F). À l'intérieur de chaque rectangle, on peut voir le type de TU ainsi que les séquences de tons optimales retenues. La fréquence d'utilisation de ces séquences est indiquée en pourcentages. On a aussi une indication du pourcentage de TU « A », « S » ou « N » apparaissant dans chaque position. On peut donc observer la façon dont les fonctions s'enchaînent.

Pour un énoncé « [+ » la séquence de fonctions la plus courante est la suivante :

> [U0 A1 SM NF] et le contour optimal associé est le suivant : {M_ TD SD SLS/U}. Le contour le plus courant peut être assimilé au contour classique « downstepping head + low rise ».

Pour un IU « [| » la configuration fonctionnelle/formelle la plus courante et la suivante :

> [U0 A1 SM NF] {M_TD SL SBB} = « downstepping head + low fall ».

A partir de ces deux grammaires, on peut observer les résultats d'optimisation des TU « A » et « S » d'une part et des TU « N » et « !N » d'autre part.

En position « 1 » et dans les deux types d'IU, on voit que les accents rythmiques et mélodiques se distinguent : les TU « A » sont modélisés par une chute de F0 du type {TD} par opposition à un contour montant sur les TU « S » {DH}.

En position médiane, la distinction entre les deux types de proéminence semble plus problématique, car les contours optimaux sont très proches, en particulier dans les contours terminaux « [| » : TU « A » = {DL} et TU « S » = {_D}, {_U}, etc. Dans les IU non terminaux « [+ » les deux types de proéminence décrivent le même mouvement descendant, mais les accents mélodiques sont toutefois plus marqués comme on s'y attend : {UL} *vs* {SD}.

En position finale seules les proéminences rythmiques peuvent apparaître dans ce modèle. On peut alors les comparer aux nucléus. Dans un IU « [| » on observe deux contours descendants l'un représenté par les séquences {DLD} et {DLS} pour les accents rythmiques et l'autre par les séquences {SBB} ou {UBB} pour les nucléus. Le nucléus est associé à un mouvement plus ample allant jusqu'au « plancher » de la tessiture du locuteur et pouvant aussi être marqué par une montée au départ du mouvement {UBB}. C'est ce qui le rend probablement plus proéminent par rapport à un accent rythmique postnucléaire. Dans un IU « [+ », on observe des contours finaux montants. On peut avoir une petite montée au niveau d'un nucléus non marqué {SLU} et une chute + montée au niveau d'un nucléus final emphatique {UBU}. Le contour associé à un accent rythmique postnucléaire est très proche des précédents : {DBU}, on n'observe toutefois pas de remontée initiale avant le début de la chute. Ce type de proéminence sera potentiellement une source de confusion avec le nucléus non marqué. On peut s'interroger sur la nature de cette proéminence postnucléaire qui semble avoir les caractéristiques d'un accent mélodique.

En ce qui concerne la modélisation de l'emphase ou plus précisément l'opposition entre les TU « N » et « !N », on peut faire les observations suivantes : en position initiale dans les IU « [| », la forme optimale pour les deux types de fonctions peut être la même : le contour {HB} est le plus fréquent, cependant un TU « !N » aura aussi tendance à être représenté par un contour {TB} par opposition à un contour {UD} dans le cas d'un nucléus non marqué. Le point de départ à un niveau de fréquence supérieur dans le cas d'un TU « !N » est bien une marque d'emphase. Dans les IU « [+ », la distinction entre ces deux fonctions en position initiale est plus évidente, on a {TD} vs {UL}.

En position médiane, les nucléus non marqués sont caractérisés par un contour descendant sans remontée initiale de F0 par opposition aux nucléus emphatiques : on a {UL} vs {SD}. Toutefois on remarque que les nucléus non marqués ont alors la même représentation que les accents rythmiques postnucléaires.

En position finale, l'opposition entre les deux types de fonction est assez claire. Pour les IU « [+ » on a une opposition entre un contour montant simple pour les nucléus non marqués et un contour descendant-montant pour les nucléus emphatiques : {SLU} vs {HBU}. La chute + montée serait donc au niveau fonctionnel, une version emphatique de la petite montée comme cela a déjà été dit dans la littérature (cf. Armstrong et Ward 1931).

Dans les IU « [| », la distinction entre les TU « !N » et « N » est moins précise, de façon globale on a deux contours descendants, l'un marqué par une remontée initiale de F0 et l'autre par une chute dans la continuité des accents précédents : {UBB} vs {SBB}. Quand un énoncé n'est composé que d'un nucléus, on a une opposition entre les contours {ULD} (N) et {TBU} (!N) pour un IU « [+ » et {ULD} (N) et {HBB} (!N) pour un IU « [| ».

Observons ensuite les grammaires générées à partir des IU emphatiques : « [!+ » et « [!| ».

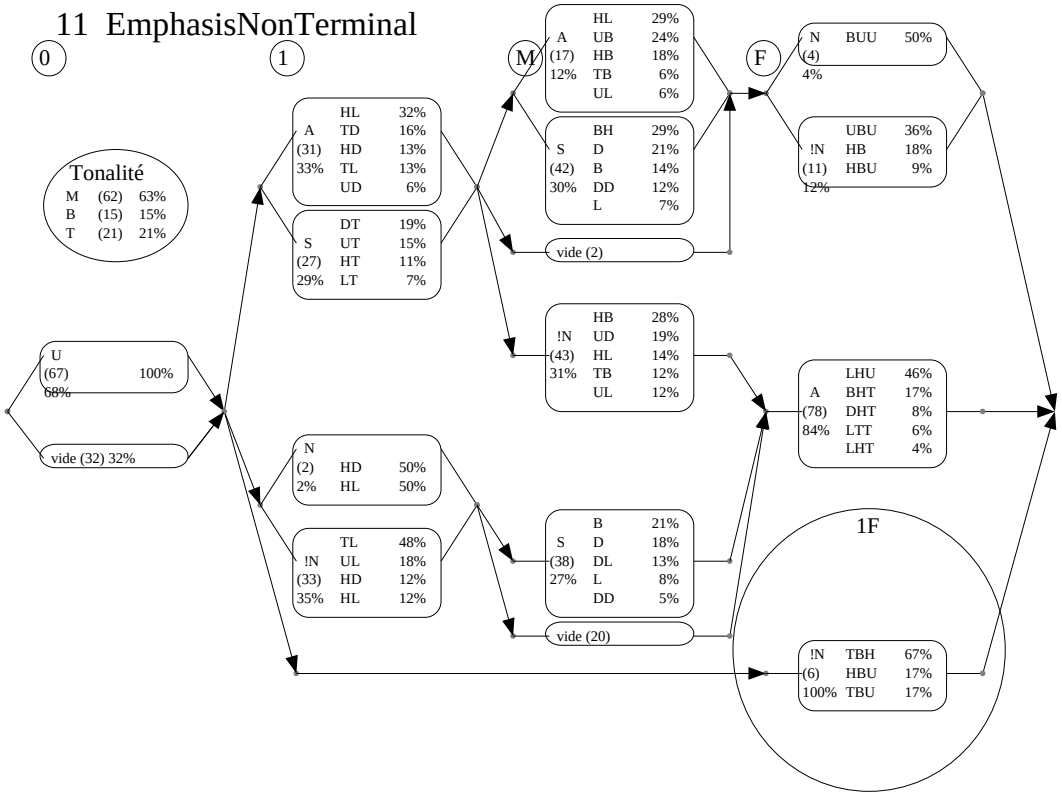


Figure 104 : Grammaire des IU de type emphatique non terminal pour le modèle 11.

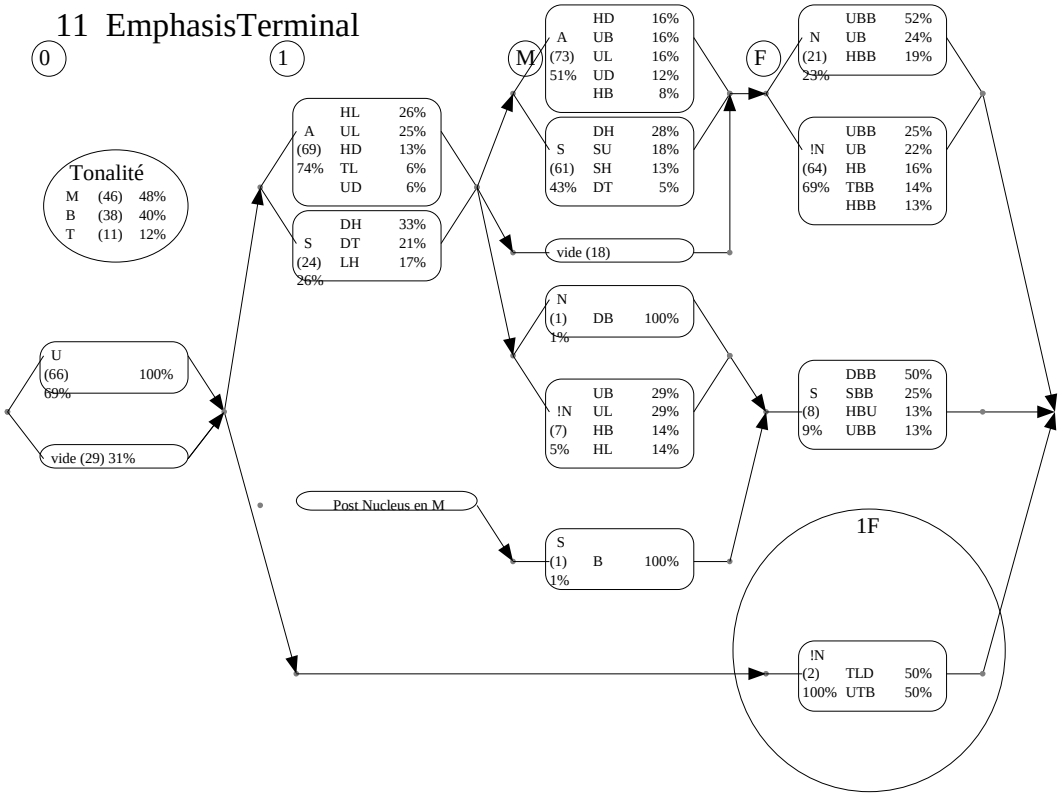


Figure 105 : Grammaire des IU de type emphatique terminal pour le modèle 11.

Les configurations formes/fonctions les plus courantes pour les énoncés de type emphatique non terminal « [!+ » et terminal « [!| » sont les suivantes :

[!U0 A1 SM !NF+]	*vs*	[!U0 A1 AM !NF|]
{ M_HL BH UBU}		{M/B_HL HD UBB}

On a donc en termes britanniques une tête descendante-montante suivie d'une chute + montée pour les IU « [!+ » (« falling-rising head + fall-rise ») et une tête en suite de chutes suivie d'une grande chute (« sliding head + high fall).

Les IU « [!+ » sont aussi bien représentés par un contour global de chute + montée divisée quand le nucléus est en position initiale. On a alors l'association suivante fonctions/formes :

[!U0 !N1 STM ATF+ {M_TL SB LHU}

Les TU « A » et « S » se distinguent généralement beaucoup mieux que dans les IU non marqués : pour un IU terminal « [| » , les TU « A » qu'ils soient en position initiale ou médiane sont toujours modélisés par une chute du type {HL} alors que les TU « S1 » et « SM » sont modélisés par un contour montant du type {DH}. Dans les IU emphatiques « [!+ », les TU « A » sont aussi modélisés par des chutes du type {HL} quelle que soit leur position, les TU « S » peuvent cependant arborer deux types de contours en position médiane : {BH} ou {_D}.

Dans un IU emphatique terminal « [!| », les accents rythmiques en position finale se distinguent des nucléus non marqués avec un contour de chute simple {DBB} par opposition à un contour {UBB} pour les nucléus. Dans un IU emphatique non terminal « [!+ » on a un contour montant plutôt qu'une chute + montée dans le cas des nucléus emphatiques mais à nouveau, le contour du nucléus non marqué est très proche de celui de l'accent rythmique postnucléaire. Au niveau des TU « N » et « !N », les observations sont les mêmes que dans les IU précédents.

Observons maintenant les grammaires générées à partir de l'opposition entre les questions terminales et non terminales.

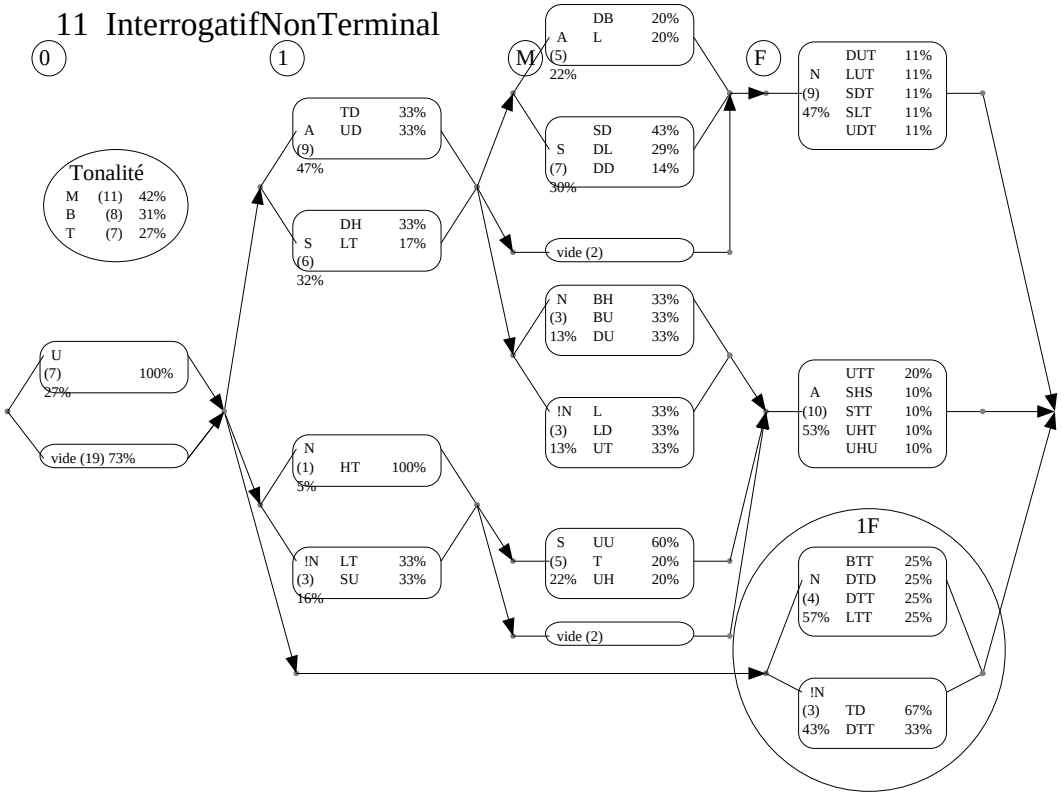


Figure 106 : Grammaire des IU de type interrogatif non terminal pour le modèle 11.

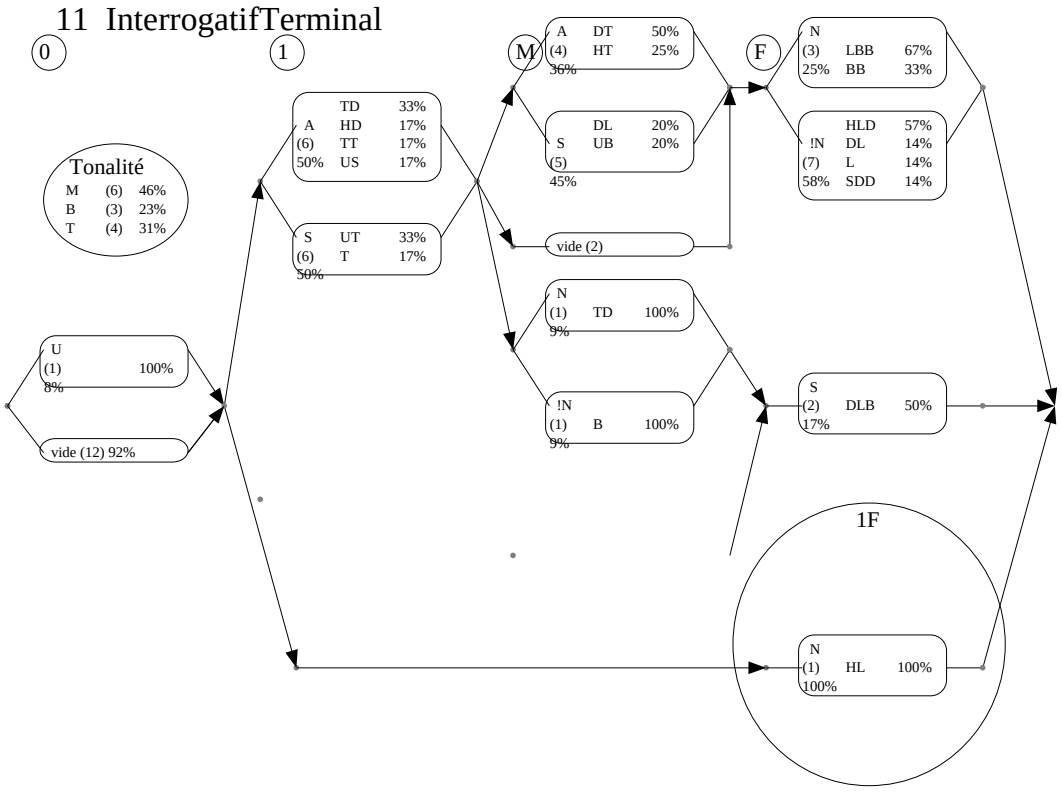


Figure 107 : Grammaire des IU de type interrogatif terminal pour le modèle 11.

La structure fonctionnelle la plus courante pour une question non terminale est la suivante :

[A1 SM NF ?+ associée au contour {TD SD DUT} c'est-à-dire un contour globalement descendant ponctué par une grande montée au niveau du nucléus.

Quand le nucléus est en position médiane, la montée commence plus tôt et l'on a un contour global descendant-montant :

[A1 NM AF ?+ {TD BH UTT}

On peut comparer ces contours de base, à ceux que l'on trouve à partir des questions terminales.

La structure la plus courante : [A1 SM !NF ?| donne un contour généralement descendant fini par une grande chute sur le nucléus : {TD DL HLD}

Si le nucléus est en position médiane, la grande chute est anticipée et l'on a le contour suivant :

[A1 NM SF ?| {TD TD DLB}

L'opposition entre les accents mélodiques et rythmiques est visible en position initiale dans les questions non terminales : on a les contours {TD} *vs* {DT}. Tout comme dans les énoncés de type emphatique, les TU « A » sont modélisés par des chutes par opposition aux TU « S » modélisés par des montées. En position médiane, les deux types de proéminence sont marqués par un contour descendant, le contour est cependant plus ample dans le cas d'un accent mélodique : on a {DB} vs {DL}.

On observera ensuite deux exemples des grammaires issues de la modélisation en TTU à partir du modèle 24. Cette fois-ci seules les formes associées aux unités mélodiques globales sont prises en compte. Les accents rythmiques ne sont pas modélisés séparément et font partie d'une unité mélodique globale. Observons donc les représentations formelles au niveau des IU de type non terminal « [+ » puis des IU de type terminal « [| ».

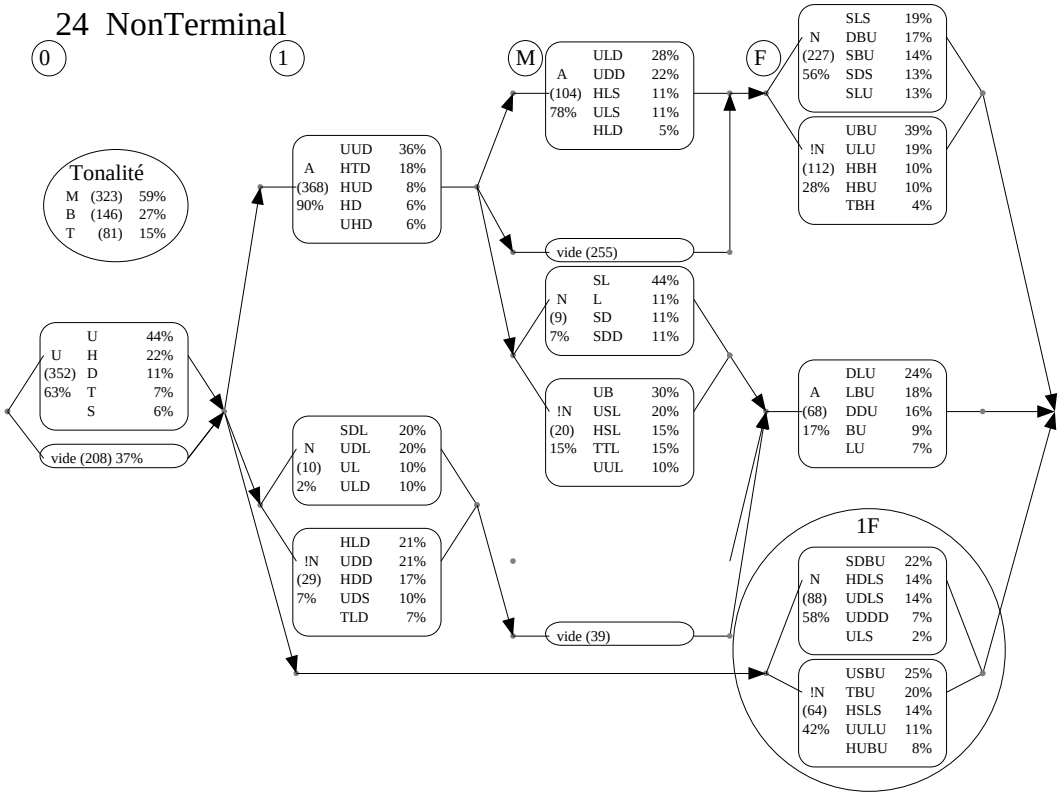


Figure 108 : Grammaire des IU de type non terminal pour le modèle 24.

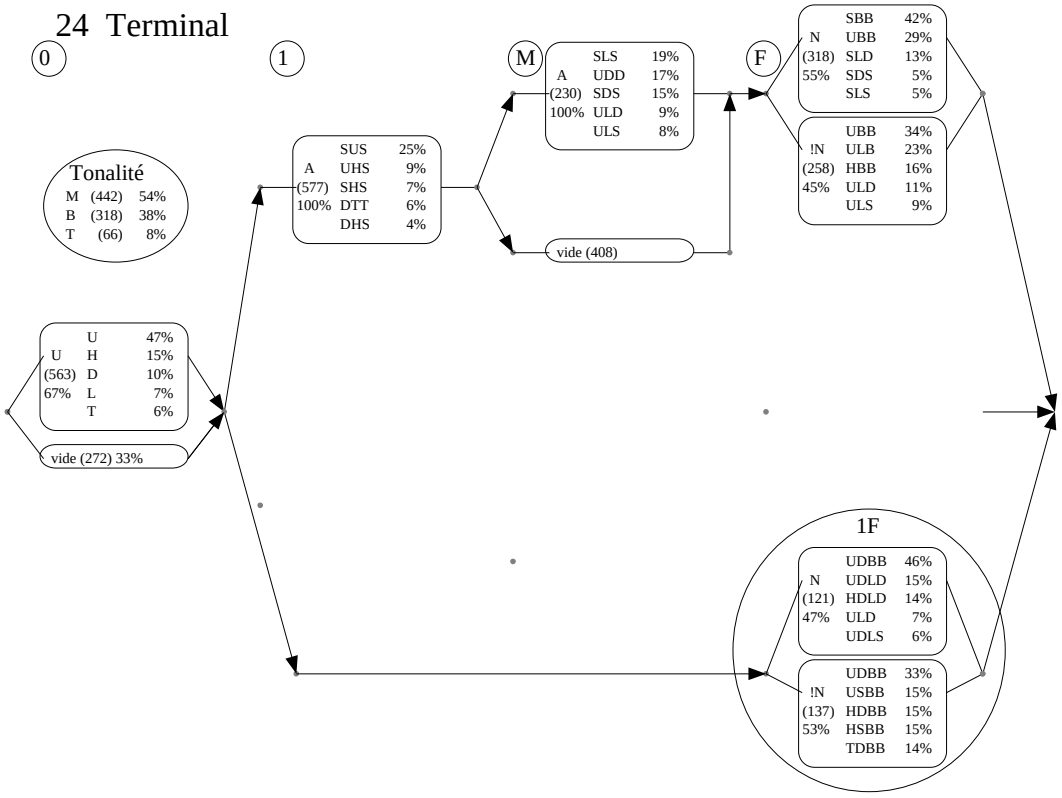


Figure 109 : Grammaire des IU de type terminal pour le modèle 24.

Pour un énoncé de type non terminal la structure fonctionnelle la plus courante est la suivante :

[U0 A1 AM NF + associée au contour {MU UUD ULD SLS/SBU}

L'opposition entre les nucléus marqués et non marqués est représentée par l'opposition entre les contours : {SDL} vs {HLD} en position initiale et {SL} vs {UB} en position médiane. Un nucléus en position finale a la même forme qu'un nucléus final non marqué, celle d'une petite montée du type {LBU} par opposition au contour de chute + montée trouvé au niveau d'un nucléus final emphatique {UBU}, {HBH}, etc.).

Pour un énoncé de type terminal, la structure fonctionnelle la plus courante est la suivante :

[U0 A1 AM NF| associée au contour {MU SUS SLS SBB}

On remarque que les nucléus en position « 1F » c'est-à-dire les énoncés ne contenant qu'une unité nucléaire sans accent mélodique postnucléaire, sont modélisés de la même façon avec une séquence du type {UDBB} qu'ils soient emphatiques ou non.

Les paramètres de ces modèles sont ensuite utilisés afin de tenter d'extraire automatiquement les informations fonctionnelles.

8.10 Optimisation des représentations fonctionnelles dans les différents modèles

Par opposition à l'optimisation des représentations formelles, et afin d'optimiser les représentations fonctionnelles, nous utiliserons d'abord des modèles comportant un nombre important d'informations fonctionnelles pour aller progressivement vers des modèles de plus en plus simples permettant une extraction des informations fonctionnelles basée essentiellement sur les formes.

Trois types de détection de fonctions seront pris en compte : tout d'abord on cherchera à détecter automatiquement le type d'IU en utilisant nos modèles formels optimisés, puis les niveaux de proéminence et enfin on cherchera à détecter les frontières d'unités intonatives. À partir des modèles basés sur les TTU, on essaiera aussi de détecter les frontières de TTU c'est-à-dire les proéminences les plus importantes. L'analyse des résultats de détection automatique de fonctions dans les différents modèles nous permettra ensuite de revenir sur les composants fonctionnels des modèles testés pour les évaluer et éventuellement les remettre en question.

8.10.1 Détection des types d'IU

Tout d'abord, une détection simple des types d'unité intonative est effectuée dans tous les modèles présentés précédemment. Dans une unité intonative donnée, on cherche à savoir à quel type d'énoncé on a affaire : s'agit-il d'une question terminale, d'une affirmation emphatique etc. ?

Les informations fonctionnelles connues pour cette étape sont les frontières des unités intonatives, les frontières des unités tonales et leur nature (U, A, S, N, !N). Il s'agit alors de tester au niveau de chaque unité tonale les séquences optimales retenues pour chaque type de fonction possible. Au niveau d'un TU « N » connu, on teste toutes les meilleures séquences de tons associées aux fonctions « [+N », « [|N », « [!|N », etc.

La suite de fonctions dont les séquences optimales permettent de générer la courbe de MOMEL la plus proche de la courbe d'origine (mesurée par le score), est retenue et le type d'IU associé est alors attribué à l'unité intonative considérée. Si la suite « [+N1 + [+SM + [+STF » permet de générer la meilleure courbe de MOMEL pour l'IU en question, alors celui-ci est étiqueté non terminal.

Les résultats sont résumés sous forme de matrices de confusion entre les types d'IU annotés et les types d'IU détectés automatiquement dans les modèles des TU et des TTU.

8.10.1.1 Résultats pour les modèles des TU (6 à 14)

Types	Types d'IU détectés									
IU annotés	[|	[+	[!|	[!+	[?|	[?+	+	[	|	Total
[|	**37,46%**	17,08%	22,80%	9,24%	10,43%	1,39%	0,88%	0,68%	0,05%	100,00%
[+	20,36%	**31,16%**	11,37%	15,48%	10,46%	6,66%	1,93%	2,36%	0,22%	100,00%
[!+	9,97%	24,95%	7,19%	**47,57%**	3,60%	5,98%	0,06%	0,14%	0,53%	100,00%
[!|	19,67%	12,26%	**45,03%**	12,91%	9,64%	0,00%	0,28%	0,00%	0,23%	100,00%
|	0,00%	0,78%	0,00%	0,00%	0,00%	0,00%	20,14%	33,33%	**45,75%**	100,00%
+	0,00%	0,65%	0,00%	0,98%	0,65%	0,65%	**65,22%**	10,94%	20,91%	100,00%
[?+	6,28%	6,28%	1,20%	10,31%	4,93%	**68,91%**	0,00%	1,79%	0,30%	100,00%
[	0,00%	0,49%	0,00%	0,00%	0,00%	2,59%	21,36%	**61,17%**	14,40%	100,00%
[?|	25,53%	12,61%	22,22%	4,50%	**35,14%**	0,00%	0,00%	0,00%	0,00%	100,00%

Tableau 99 : Matrice de confusion des types d'IU annotés et détectés pour les modèles des TU.

Les IU terminaux sont bien détectés dans 37,46% des cas. Ils sont principalement confondus avec les IU terminaux emphatiques, les IU non terminaux et les questions terminales. Pour tous ces types d'IU le contour global optimisé est très proche en effet.

31,16 % des IU non terminaux sont détectés correctement et l'on a la plus importante marge de confusion avec les IU terminaux.

Les IU emphatiques sont mieux identifiés avec un taux de réussite de 47,57% pour les IU « [!+ » et 45,03% pour les IU « [!| ».

Le meilleur taux de réussite est obtenu avec les IU « [?+ » (68,91%) suivis de parenthèses basses non terminales « + » avec un taux de 65,22%.

8.10.1.2 Résultats pour les modèles basés sur les TTU2

Types	Types d'IU détectés									
IU annotés	[|	[+	[!|	[!+	[?|	[?+	+	|	[	Total
[|	**38,38%**	15,84%	22,91%	8,51%	9,92%	3,26%	1,05%	0,00%	0,12%	100,00%
[+	19,73%	**33,42%**	9,48%	14,98%	9,20%	10,22%	1,86%	0,00%	1,12%	100,00%
[!|	15,23%	10,98%	**48,71%**	14,69%	6,13%	4,26%	0,00%	0,00%	0,00%	100,00%
[!+	3,57%	37,45%	6,47%	**48,74%**	1,18%	2,60%	0,00%	0,00%	0,00%	100,00%
|	0,00%	0,56%	0,00%	0,00%	0,00%	0,00%	28,00%	**60,78%**	10,67%	100,00%
+	0,00%	2,93%	0,00%	0,00%	0,53%	0,00%	**77,76%**	14,91%	3,86%	100,00%
[?+	2,46%	9,34%	0,98%	7,86%	8,11%	**67,32%**	0,98%	0,00%	2,95%	100,00%
[?|	28,23%	8,06%	9,27%	15,73%	**35,08%**	2,42%	1,21%	0,00%	0,00%	100,00%
[	0,00%	56,32%	0,00%	0,00%	0,00%	13,79%	11,49%	0,00%	**18,39%**	100,00%

Tableau 100 : Matrice de confusion des types d'IU annoté et détectés pour les modèles en TTU2.

C'est dans ces derniers modèles que les IU emphatiques sont le mieux identifiés avec des taux de 48,71% pour les IU « [!| » et 48,74% pour les IU « [!+ ». Les meilleurs taux de détection sont donc obtenus, quel que soit le modèle, avec les IU : « [?+ », « + », « [», « [!+ », « [!| ». Les IU les plus problématiques sont les IU « [| », « [+ » et « [?| », dont les formes optimales globales sont moins marquées (cf. tableau 51, p. 265). Le contour global de l'IU « [?| » se confond avec un IU « [| ». Celui d'un IU « [| » se confond avec un IU « [+ ».

8.10.2 Détection des types de TU

Pour cette étape de détection des types de proéminence, les informations fonctionnelles connues sont les frontières et types d'IU ainsi que les frontières des TU. Sachant que l'on se trouve dans une IU de type « [| » contenant trois TU, on cherche à connaître la nature de ceux-ci : le premier TU est-il un accent rythmique, un nucléus ou un accent mélodique etc.?

Afin de déterminer la nature des proéminences, on s'appuiera à nouveau sur les représentations formelles optimisées. La séquence de fonctions dont les représentations formelles permettent de générer la courbe la mieux corrélée avec MOMEL d'origine sera retenue.

Tout d'abord nous exposerons les résultats de détection sous forme de matrices de confusion globales : la première représentant les modèles des TU et la seconde celle des TTU.

8.10.2.1 Résultats globaux

Dans les modèles des TU (6-14)

Types	Types de TU détectés						
TU annotés	!N	A	N	S	U	(vide)	Total
!N	**2505**	478	2816	1393		17	7209
A	818	**4615**	1395	4141		20	10989
N	1464	72	**3531**	2398		23	7488
S	638	1909	2131	**8234**		696	13608
U					**9861**	12	9873
Total	5425	7074	9873	16166	9861	768	49167

Tableau 101 : Matrice de confusion des types de TU annotés et détectés dans les modèles 6 à 14.

Au niveau des TU « !N », 35% sont bien détectés. 39% d'entre eux sont confondus avec un nucléus non marqué et 19% avec un accent rythmique. Les TU « A » sont principalement confondus avec des accents rythmiques ou des nucléus non marqués. Les proéminences les mieux détectées sont les proéminences rythmiques.

Observons les résultats de détection des proéminences à partir des modèles des TTU.

Résultats pour les modèles TTU2

Pour les modèles basés sur les unités tonales élargies, on distingue seulement trois types de proéminence : les proéminences nucléaires (N), nucléaires emphatiques (!N) et les accents mélodiques (A).

Types	Types de TTU détectés				
TTU annotés	A	U	N	!N	Total
A	**11342**		912	564	12818
U	1	**9117**	7		9125
N	669		**3937**	1653	6259
!N	735		2536	**2861**	6132
Total	12747	9117	7660	5078	34912

Tableau 102 : Matrice de confusion des types de TTU annotés et détectés dans les modèles TTU2.

Les proéminences de type « A » obtiennent les meilleurs résultats de détection puisque 88% sont bien identifiés. Les nucléus marqués et non marqués sont souvent confondus comme dans les modèles précédents.

Après avoir observé les résultats de façon globale, nous rentrerons dans plus de détails en prenant les résultats des types de proéminence un par un et en les détaillant en fonction de la position du TU (NM, N1, NF) et le type d'IU ([|, [+, etc.). Ceci nous permettra de préciser les cas qui posent problème en détection. Dans quels cas les nucléus sont-ils confondus avec des accents rythmiques, etc. ?

8.10.2.2 Résultats détaillés des détections de proéminences dans les modèles basés sur les TU

Les TU N (nucléus)

Tout d'abord les résultats de détection du nucléus sont présentés par modèle, tous types d'IU confondus et quelle que soit la position du nucléus :

	Types de TU détectés					
modèle	!N	A	N	S	(blank)	Total
6	11,06%	0,72%	**52,16%**	35,22%	0,84%	100,00%
7	16,35%	0,60%	**49,28%**	33,65%	0,12%	100,00%
8	18,75%	0,60%	**45,19%**	35,22%	0,24%	100,00%
9	18,63%	0,96%	**47,12%**	33,17%	0,12%	100,00%
10	26,56%	0,84%	**37,62%**	34,74%	0,24%	100,00%
11	16,59%	1,32%	**54,45%**	27,40%	0,24%	100,00%
12	23,08%	1,32%	**45,55%**	29,69%	0,36%	100,00%
13	26,20%	1,08%	**42,79%**	29,57%	0,36%	100,00%
14	18,75%	1,20%	**50,24%**	29,57%	0,24%	100,00%
Total	19,55%	0,96%	**47,16%**	32,02%	0,31%	100,00%

Tableau 103 : Résultats de la détection du nucléus par modèle et en pourcentage.

Le modèle qui permet de détecter le plus de nucléus exactement est le modèle 11 qui a un taux de réussite de 54,45%. Dans l'ensemble les nucléus sont surtout confondus avec des accents rythmiques. Il reste à observer le rôle de la position du TU et le type d'IU sur ces résultats : dans quels types d'IU obtient-on la meilleure détection du nucléus ? Qu'en est-il de la détection d'un nucléus final, initial etc. ?

Détection du nucléus en position initiale

Fonctions annotées	Fonctions détectées					
	IU	N1	A1	!N1	S1	Total
N1	[!+	**72,22%**	16,67%	11,11%	0,00%	100,00%
	[?+	**55,56%**	22,22%	0,00%	22,22%	100,00%
	[|	**63,49%**	16,67%	16,67%	3,17%	100,00%
	[+	**61,73%**	27,16%	11,11%	0,00%	100,00%
Total		**63,25%**	20,51%	13,68%	2,56%	100,00%

Tableau 104 : Résultats de détection du nucléus en position initiale.

63,25% des nucléus en position initiale sont bien identifiés. Le meilleur résultat est obtenu dans les IU « [!+ ». En position initiale c'est la confusion avec les TU « A » qui ressort le plus même si le taux reste faible (20,51%).

Détection du nucléus en position médiane

Fonctions annotées	IU	Fonctions détectées					
		STM	NM	SM	!NM	AM	Total
NM	[!|	88,89%	**11,11%**	0,00%	0,00%	0,00%	100,00%
	[?|	0,00%	**77,78%**	22,22%	0,00%	0,00%	100,00%
	[?+	25,93%	**44,44%**	18,52%	11,11%	0,00%	100,00%
	[|	33,68%	**27,78%**	23,96%	8,68%	5,90%	100,00%
	[+	33,33%	**23,33%**	16,67%	18,89%	7,78%	100,00%
Total		33,57%	**28,61%**	21,51%	10,64%	5,67%	100,00%

Tableau 105 : Résultats de détection du nucléus en position médiane.

La détection du nucléus en position médiane est beaucoup plus problématique, seuls 28,61% des nucléus sont bien identifiés. Dans l'ensemble, le nucléus est confondu avec un accent rythmique postnucléaire. Le nucléus en position médiane est toutefois bien identifié dans les IU « [?+ » (44,44%) et les IU « [?| » (77,78%).

Détection du nucléus en position finale

Fonctions annotées	IU	Fonctions détectées			
		NF	STF	!NF	Total
NF	[!|	**64,02%**	31,22%	4,76%	100,00%
	[!+	**66,67%**	5,56%	27,78%	100,00%
	[?|	**74,07%**	25,93%	0,00%	100,00%
	[?+	**45,68%**	54,32%	0,00%	100,00%
	[|	**42,74%**	40,64%	16,62%	100,00%
	[+	**37,91%**	44,38%	17,71%	100,00%
Total		**41,90%**	41,73%	16,37%	100,00%

Tableau 106 : Résultats de détection du nucléus en position finale.

41,90% des nucléus finaux sont bien identifiés. 41,73% sont confondus avec des accents rythmiques postnucléaires. Le taux de confusion le plus important apparaît dans les IU « [?+ » et « [+ ».

Les TU « !N » (nucléus emphatiques)

Tout d'abord les résultats de détection du nucléus emphatique sont présentés par modèle, tous types d'IU confondus et quelle que soit la position du nucléus :

	Types de TU détectés					
Modèles	!N	A	N	S	(blank)	Total
6	**25,47%**	6,24%	42,82%	24,97%	0,50%	100,00%
7	**31,59%**	5,99%	43,20%	19,10%	0,12%	100,00%
8	**32,33%**	6,37%	39,20%	21,72%	0,37%	100,00%
9	**32,46%**	7,24%	38,83%	21,35%	0,12%	100,00%
10	**44,44%**	6,74%	27,34%	21,35%	0,12%	100,00%
11	**31,59%**	7,12%	44,69%	16,48%	0,12%	100,00%
12	**39,20%**	6,99%	36,45%	17,23%	0,12%	100,00%
13	**43,32%**	5,87%	34,71%	15,86%	0,25%	100,00%
14	**32,33%**	7,12%	44,32%	15,86%	0,37%	100,00%
Total	**34,75%**	6,63%	39,06%	19,32%	0,24%	100,00%

Tableau 107 : Résultats de détection du nucléus emphatique par modèle.

Dans l'ensemble le nucléus emphatique est confondu avec le nucléus non emphatique ce qui laisse à penser que la modélisation des représentations formelles dans le cas de la fonction emphatique est insuffisante. C'est dans le modèle 10 que l'on obtient les meilleurs résultats de détection.

Détection du nucléus emphatique en position initiale

Fonctions annotées	IU	Fonctions détectées				
		N1	A1	!N1	S1	Total
!N1	[!+	27,61%	50,17%	**10,77%**	11,45%	100,00%
	[?+	18,52%	11,11%	**55,56%**	14,81%	100,00%
	[|	57,04%	15,56%	**14,07%**	13,33%	100,00%
	[+	26,85%	37,50%	**33,33%**	2,31%	100,00%
Total		36,91%	33,95%	**19,38%**	9,75%	100,00%

Tableau 108 : Résultats de détection du nucléus emphatique en position initiale.

Le nucléus emphatique en position initiale est confondu avec un nucléus non marqué dans 36,91% des cas et avec un accent « A » dans 33,95% des cas. Le meilleur taux d'identification est obtenu dans les questions non terminales (55,56%).

Détection du nucléus emphatique en position 1F

Fonctions annotées		Fonctions détectées		
	IU	!N1F	N1F	Total
!N1F	[!|	**100,00%**	0,00%	100,00%
	[!+	**100,00%**	0,00%	100,00%
	[?+	**66,67%**	33,33%	100,00%
	[|	**44,55%**	55,45%	100,00%
	[+	**60,78%**	39,22%	100,00%
Total		**53,05%**	46,95%	100,00%

Tableau 108b. Résultats de détection du nucléus emphatique en position 1F.

C'est la position où la détection du TU « !N » est la meilleure (53,05%). Dans les IU emphatiques la détection est parfaite (100%).

Détection du nucléus emphatique en position médiane

Fonctions annotées		Fonctions détectées					
	IU	!NM	AM	SM	NM	STM	Total
!NM	[!|	**25,40%**	6,35%	14,29%	41,27%	12,70%	100,00%
	[!+	**33,07%**	38,50%	17,83%	0,00%	10,59%	100,00%
	[?|	**77,78%**	0,00%	22,22%	0,00%	0,00%	100,00%
	[?+	**74,07%**	0,00%	14,81%	11,11%	0,00%	100,00%
	[|	**12,87%**	6,43%	23,39%	36,26%	21,05%	100,00%
	[+	**36,00%**	17,33%	15,11%	15,56%	16,00%	100,00%
Total		**31,07%**	23,02%	17,91%	14,29%	13,72%	100,00%

Tableau 109 : Résultats de détection du nucléus emphatique en position médiane.

Le nucléus emphatique en position médiane est mieux identifié que le nucléus non emphatique (31,07% *vs* 28,61). La plus grande marge de confusion existe avec les TU « AM » (23,02 %).

Détection du nucléus emphatique en position finale

Fonctions annotées		Fonctions détectées			
	IU	NF	!NF	STF	Total
!NF	[!|	65,28%	**18,40%**	16,32%	100,00%
	[!+	5,05%	**76,77%**	18,18%	100,00%
	[?|	60,32%	**30,16%**	9,52%	100,00%
	[|	43,60%	**33,58%**	22,83%	100,00%
	[+	28,87%	**31,43%**	39,70%	100,00%
Total		42,14%	**31,83%**	26,02%	100,00%

Tableau 110 : Résultats de détection du nucléus emphatique en position finale.

42,14% des TU « !N » sont confondus avec des nucléus finaux non emphatiques. Les meilleurs résultats sont obtenus dans les IU « [!+ » (76,77%).

Les TU « A »

Dans les modèles des TU, les accents mélodiques n'apparaissent qu'en position initiale et médiane. On se pose alors les questions suivantes : nos modèles d'optimisation permettent-ils de distinguer un accent rythmique d'un accent mélodique ou un accent mélodique prénucléaire d'un nucléus ?

	Types de TU détectés					
Modèles	!N	A	N	S	(blank)	Total
6	8,19%	**38,74%**	14,58%	38,08%	0,41%	100,00%
7	9,01%	**37,67%**	12,45%	40,70%	0,16%	100,00%
8	7,53%	**40,29%**	14,58%	37,18%	0,41%	100,00%
9	9,91%	**45,13%**	11,30%	33,50%	0,16%	100,00%
10	11,47%	**41,28%**	11,71%	35,38%	0,16%	100,00%
11	5,65%	**44,47%**	11,22%	38,57%	0,08%	100,00%
12	3,93%	**40,87%**	14,00%	41,11%	0,08%	100,00%
13	5,90%	**43,82%**	12,53%	37,67%	0,08%	100,00%
14	5,41%	**45,70%**	11,88%	36,94%	0,08%	100,00%
Total	7,44%	**42,00%**	12,69%	37,68%	0,18%	100,00%

Tableau 111 : Résultats de détection du TU A par modèle.

42% des TU « A » sont bien identifiés tous modèles confondus. Le meilleur modèle de détection est le modèle 14 basé sur un alignement par rapport à la voyelle. Comme on pouvait s'y attendre, la plus grande marge de confusion est trouvée avec les accents rythmiques.

Détection des TU A en position initiale

Fonctions annotées		Fonctions détectées				
	IU	A1	S1	N1	!N1	Total
A1	[!|	**78,58%**	21,42%	0,00%	0,00%	100,00%
	[!+	**70,76%**	16,25%	10,47%	2,53%	100,00%
	[?|	**72,22%**	27,78%	0,00%	0,00%	100,00%
	[?+	**49,38%**	40,74%	8,64%	1,23%	100,00%
	[|	**42,25%**	31,18%	18,39%	8,18%	100,00%
	[+	**53,21%**	21,35%	17,76%	7,67%	100,00%
Total		**50,85%**	26,34%	15,93%	6,88%	100,00%

Tableau 112 : Résultats de détection du TU A en position initiale.

Un accent mélodique en position initiale est bien identifié dans la plupart des types d'IU puisque le taux de réussite moyen est de 50,85%. Les meilleurs résultats de détection sont obtenus dans les IU emphatiques (78,58% et 70,76%). C'est en effet dans ces IU emphatiques que les formes optimisées apparaissaient les plus distinctes.

Détection des TU A en position médiane

Fonctions annotées		Fonctions détectées					
	IU	SM	AM	STM	!NM	NM	Total
AM	[!|	26,33%	**49,47%**	5,94%	6,39%	11,87%	100,00%
	[!+	20,26%	**75,82%**	1,96%	1,96%	0,00%	100,00%
	[?|	33,33%	**63,89%**	0,00%	0,00%	2,78%	100,00%
	[?+	13,33%	**11,11%**	31,11%	28,89%	15,56%	100,00%
	[|	46,53%	**19,44%**	22,87%	5,09%	6,06%	100,00%
	[+	37,93%	**22,03%**	14,75%	17,05%	8,24%	100,00%
Total		39,63%	**27,33%**	17,19%	8,45%	7,40%	100,00%

Tableau 113 : Résultats de détection du TU A en position médiane.

Dans 39,63% des cas, les TU A sont confondus avec des accents rythmiques. En effet, la modélisation de ces types de TU en position médiane donne des résultats identiques à la modélisation des accents rythmiques. Les résultats sont tout de même très satisfaisants quand on se concentre sur les IU emphatiques et les questions terminales (49,47%, 75,82% et 63,89%).

Les TU « S »

Les accents rythmiques peuvent apparaître en position initiale, médiane et finale ou même postnucléaire. En position prénucléaire, on s'attend à les confondre avec les accents mélodiques, en position finale, on cherche à les distinguer des accents nucléaires.

		Fonctions détectées				
Modèle	!N	A	N	S	(blank)	Total
6	5,95%	13,29%	16,34%	**59,06%**	5,36%	100,00%
7	3,90%	13,49%	15,67%	**61,97%**	4,96%	100,00%
8	5,29%	12,96%	15,94%	**60,78%**	5,03%	100,00%
9	5,36%	16,07%	16,34%	**57,41%**	4,83%	100,00%
10	5,62%	14,22%	14,55%	**60,12%**	5,49%	100,00%
11	3,64%	15,87%	16,01%	**59,59%**	4,89%	100,00%
12	4,23%	12,63%	15,74%	**61,90%**	5,49%	100,00%
13	4,17%	13,89%	14,42%	**62,57%**	4,96%	100,00%
14	4,03%	13,82%	15,94%	**61,18%**	5,03%	100,00%
Total	4,69%	14,03%	15,66%	**60,51%**	5,11%	100,00%

Tableau 114 : Résultats de détection du TU S par modèle.

Les TU « S » sont les types de proéminence les mieux identifiés en général avec un taux de réussite de 60,51%. C'est à partir du modèle 13 basé sur un alignement par rapport aux frontières du TU et à la durée de la dernière syllabe, que le taux de réussite est le plus élevé (62,57%).

Détection des TU « S » en position initiale

Fonctions annotées		Fonctions détectées				
	IU	S1	A1	N1	!N1	Total
S1	[!|	**72,43%**	27,57%	0,00%	0,00%	100,00%
	[!+	**55,14%**	42,80%	1,23%	0,82%	100,00%
	[?|	**74,07%**	25,93%	0,00%	0,00%	100,00%
	[?+	**40,74%**	37,04%	7,41%	14,81%	100,00%
	[|	**52,86%**	29,70%	12,25%	5,19%	100,00%
	[+	**45,15%**	35,99%	13,78%	5,08%	100,00%
Total		**51,95%**	32,44%	10,99%	4,62%	100,00%

Tableau 115 : Résultats de détection du TU S en position initiale.

51,95% des accents rythmiques sont bien détectés en position initiale. 32,44% des TU « S » sont toutefois confondus avec des accents mélodiques comme prévu.

Détection des TU « S » en position médiane

Fonctions annotées		Fonctions détectées					
	IU	SM	STM	AM	NM	!NM	Total
SM	[!|	**56,99%**	9,50%	25,09%	5,91%	2,51%	100,00%
	[!+	**49,86%**	23,89%	15,42%	0,00%	10,83%	100,00%
	[?|	**55,56%**	0,00%	17,78%	11,11%	15,56%	100,00%
	[?+	**43,52%**	34,26%	3,70%	9,26%	9,26%	100,00%
	[|	**52,50%**	28,67%	8,74%	7,90%	2,19%	100,00%
	[+	**41,07%**	26,03%	12,85%	12,31%	7,73%	100,00%
Total		**48,91%**	25,34%	12,35%	8,17%	5,23%	100,00%

Tableau 116 : Résultats de détection du TU S en position médiane.

Détection des TU « S » en position finale

Fonctions annotées		Fonctions détectées			
	IU	STF	NF	!NF	Total
SF	[!|	**81,94%**	18,06%	0,00%	100,00%
	[!+	**51,42%**	35,47%	13,11%	100,00%
	[?|	**77,78%**	11,11%	11,11%	100,00%
	[?+	**87,78%**	12,22%	0,00%	100,00%
	[|	**74,62%**	21,87%	3,51%	100,00%
	[+	**70,05%**	23,57%	6,38%	100,00%
Total		**67,25%**	25,81%	6,94%	100,00%

Tableau 117 : Résultats de détection du TU S en position finale.

62,25% des TU « S » sont bien identifiés en position finale, 25,81% sont confondus avec des nucléus non marqués. Les types d'IU pour lesquels cette confusion est la plus importante sont les IU « [!+ » et « [+ ».

8.10.2.3 Conclusion

Les nucléus non marqués sont bien détectés en position initiale et « 1F » (initiale et finale). En position médiane, ce type de proéminence pose un problème important, car il est très souvent confondu avec un accent rythmique. En effet, l'observation des séquences optimales associées à ces deux types de proéminence permet de prévoir cette difficulté. Ces deux types de TU obtiennent la représentation formelle typique : {SD} ou {SL}.

En position finale, on remarque aussi que le nucléus peut être confondu avec un accent rythmique final tout particulièrement dans les énoncés non terminaux. Ceci laisse à penser que dans certains cas, on a affaire non pas seulement à un accent rythmique final après le nucléus mais à un accent mélodique. Dans les énoncés de type « [+ », « [!+ », « [?+ » nous ajouterons donc la possibilité d'avoir un accent mélodique postnucléaire étiqueté « ATF ». Ceci permet de représenter les cas de chutes montées divisées que l'on s'attend à trouver dans ces types d'énoncés.

La détection du nucléus emphatique permet de mettre en valeur d'autres problématiques. En position médiane, il est bien identifié et se distingue de l'accent rythmique. Les formes optimales associées à ces deux fonctions sont {HL} *vs* {SD}. En position « 1F » les résultats de détection sont généralement très bons. Les deux positions qui sont source de difficultés sont les positions initiales et finales pour lesquelles le nucléus emphatique est confondu avec un nucléus non marqué. En effet, les formes optimales retenues pour ces deux types de fonction en position initiale et finale sont très proches. Dans ces cas précis, le facteur mélodique ne semble pas suffire à distinguer la fonction emphatique de la fonction non marquée. Notre modélisation doit être enrichie.

En ce qui concerne la distinction entre les accents rythmiques et mélodiques et comme on pouvait s'y attendre à l'issue des essais d'annotation et de l'observation des grammaires, quelques problèmes se posent aussi. En position initiale, l'accent mélodique se distingue généralement de l'accent rythmique, mais en position médiane la distinction est beaucoup plus difficile.

8.10.2.4 Résultats détaillés des détections de proéminences dans les modèles basés sur les TTU

Dans les modèles britanniques adaptés, on se base sur des unités mélodiques globales pouvant avoir les fonctions suivantes « A », « N » ou « !N » dans les positions initiales, médianes et finales. La possibilité d'avoir un accent mélodique après le nucléus a été ajoutée dans les IU de type non terminaux. On cherche essentiellement à distinguer les accents nucléaires des accents mélodiques simples.

Les TTU « N »

	Types de TTU détectés			
Modèles	N	!N	A	Total
20	**60,64%**	36,92%	2,44%	100,00%
21	**51,86%**	34,79%	13,35%	100,00%
22	**57,50%**	30,12%	12,39%	100,00%
23	**63,25%**	27,01%	9,74%	100,00%
24	**65,73%**	23,91%	10,36%	100,00%
Total	**62,90%**	26,41%	10,69%	100,00%

Tableau 118 : Résultats de détection du TTU N.

62,90 % des TTU « N » sont bien identifiés. Les meilleurs résultats sont obtenus avec le modèle 24. La plus grande source d'erreur est la confusion avec des TTU « !N » (26,41%).

Détection des TTU « N » en position initiale

Types de TTU annotés		Types de TTU détectés				
	IU	!N1	A1	N1	(blank)	Total
N1	[!+	12,50%	25,00%	**62,50%**	0,00%	100,00%
	[+	22,50%	28,75%	**37,50%**	11,25%	100,00%
Total		20,83%	28,13%	**41,67%**	9,38%	100,00%

Tableau 119 : Résultats de détection du TTU N en position initiale.

41,67% des TTU « N » sont bien identifiés en position initiale, il y a cependant une plus grande marge de confusion avec les TTU « A » (28,13%).

Détection des TTU « N » en position médiane

Types de TTU annotés		Types de TTU détectés			
	IU	!NM	AM	**NM**	Total
NM	[+	19,44%	52,78%	**27,78%**	100,00%
Total		19,44%	52,78%	**27,78%**	100,00%

Tableau 120 : Résultats de détection du TTUN en position médiane.

De nouveau, la position médiane est problématique puisque 52,78% des TTU « N » sont confondus avec des TTU « A ».

Détection des TTU « N » en position 1F

Types de TTU annotés		Types de TTU détectés			
	IU	!N1F	N1F	(blank)	Total
N1F	[?|	0,00%	**87,50%**	12,50%	100,00%
	[?+	23,21%	**76,79%**	0,00%	100,00%
	[|	41,43%	**58,06%**	0,52%	100,00%
	[+	40,34%	**57,39%**	2,27%	100,00%
Total		40,21%	**58,53%**	1,27%	100,00%

Tableau 121 : Résultats de détection du TTU N en position 1F.

58,53% des TTU « N » en position « 1F » sont bien identifiés. Les meilleurs taux de réussite sont obtenus dans les questions terminales (87,50%).

Détection des TTU « N » en position finale

Types de TTU annotés		Types de TTU détectés				
	IU	!NF	ATF	NF	(blank)	Total
NF	[!|	25,57%	0,00%	**71,02%**	3,41%	100,00%
	[!+	15,63%	28,13%	**56,25%**	0,00%	100,00%
	[?|	21,88%	0,00%	**78,13%**	0,00%	100,00%
	[?+	28,75%	0,00%	**61,25%**	10,00%	100,00%
	[|	18,16%	0,00%	**70,09%**	11,75%	100,00%
	[+	20,87%	32,76%	**45,87%**	0,50%	100,00%
Total		19,68%	12,91%	**60,53%**	6,88%	100,00%

Tableau 122 : Résultats de détection du TTU N en position finale.

60,53% des TTU « N » en position finale sont bien identifiés. Dans les types d'IU où l'on a annoté des accents mélodiques postnucléaires, on retrouve une marge de confusion de 30%.

Les TTU « !N »

	Types de TTU détectés			
Modèle	!N	N	A	Total
20	**50,75%**	33,71%	15,54%	100,00%
21	**55,18%**	32,37%	12,45%	100,00%
22	**52,92%**	36,74%	10,34%	100,00%
23	**51,32%**	38,13%	10,55%	100,00%
24	**44,36%**	42,67%	12,97%	100,00%
Total	**46,66%**	41,36%	11,99%	100,00%

Tableau 123 : Résultats de détection du TTU !N.

46,66% des TTU « !N » sont bien identifiés. La plus grande marge de confusion est trouvée avec les TTU « N » (41,36%).

Détection des TTU « !N » en position initiale

Types de TTU annotés		Types de TTU détectés			
	IU	!N1	A1	N1	Total
!N1	[!+	**18,29%**	53,35%	28,35%	100,00%
	[+	**33,62%**	41,81%	24,57%	100,00%
Total		**24,64%**	48,57%	26,79%	100,00%

Tableau 123b : Résultats de détection du TTU !N en position initiale.

48,57% des TTU « !N » sont confondus avec des TTU « A ».

Détection des TTU « !N » en position « 1F »

Types de TTU annotés		Types de TTU détectés			
	IU	!N1F	N1F	(blank)	Total
!N1F	[!|	**100,00%**	0,00%	0,00%	100,00%
	[!+	**100,00%**	0,00%	0,00%	100,00%
	[?+	**51,79%**	48,21%	0,00%	100,00%
	[|	**51,46%**	48,27%	0,27%	100,00%
	[+	**63,48%**	31,64%	4,88%	100,00%
Total		**56,83%**	41,55%	1,62%	100,00%

Tableau 124 : Résultats de détection du TTU !N en position 1F.

Les TTU « !N » en positon « 1F » sont tous identifiés dans le cas des IU emphatiques et obtiennent un taux global de réussite de 56,83%. 41,55% des TTU « !N » sont toutefois confondus avec des nucléus non-marqués.

Détection des TTU « !N » en position médiane

Types de TTU annotés		Types de TTU détectés				
	IU	!NM	AM	NM	(blank)	Total
!NM	[!+	**53,57%**	43,93%	0,00%	2,50%	100,00%
	[+	**43,75%**	38,13%	13,13%	5,00%	100,00%
Total		**50,00%**	41,82%	4,77%	3,41%	100,00%

Tableau 125 : Résultats de détection du TTU !N en position médiane.

Contrairement aux TTU « N » les TTU « !N » en position médiane sont bien identifiés dans 50% des cas. 41,82% des TTU « !N » sont aussi confondus avec les TTU « A ».

Détection des TTU « !N » en position finale

Types de TTU annotés		Types de TTU détectés				
	IU	!NF	ATF	NF	(blank)	Total
!NF	[!|	**57,39%**	0,00%	34,68%	7,92%	100,00%
	[!+	**75,00%**	13,64%	11,36%	0,00%	100,00%
	[?|	**40,63%**	0,00%	59,38%	0,00%	100,00%
	[?+	**56,25%**	0,00%	0,00%	43,75%	100,00%
	[|	**37,11%**	0,00%	53,34%	9,54%	100,00%
	[+	**36,61%**	29,80%	33,59%	0,00%	100,00%
Total		**41,15%**	7,55%	44,56%	6,74%	100,00%

Tableau 126 : Résultats de détection du TTU !N en position finale.

En position finale, 44,56% des TTU « !N » sont confondus avec des TTU « N » et 41,15% sont bien identifiés.

Les TTU « A »

	Types de TTU détectés			
Modèle	A	N	!N	Total
20	**91,30%**	7,32%	1,38%	100,00%
21	**86,73%**	6,95%	6,32%	100,00%
22	**89,38%**	4,93%	5,69%	100,00%
23	**89,40%**	5,36%	5,24%	100,00%
24	**87,94%**	7,40%	4,66%	100,00%
Total	**88,48%**	7,11%	4,40%	100,00%

Tableau 127 : Résultats de détection du TTU A.

C'est le type de TTU le mieux identifié quel que soit le type d'IU et la position dans l'IU avec un taux de réussite de 88,48%. Les résultats sont excellents pour les positions initiales et médianes (cf. tableaux ci-dessous). Seule la position finale pose quelques difficultés.

Détection des TTU « A » en position initiale

Types de TTU annotés		Types de TTU détectés					
	IU	!N1	A1	AM	N1	(blank)	Total
A1	[!|	0,00%	**94,09%**	0,00%	0,00%	5,91%	100,00%
	[!+	1,50%	**95,25%**	0,00%	1,25%	2,00%	100,00%
	[?|	0,00%	**100,00%**	0,00%	0,00%	0,00%	100,00%
	[?+	0,00%	**91,67%**	0,00%	0,00%	8,33%	100,00%
	[|	0,00%	**89,56%**	0,00%	0,00%	10,44%	100,00%
	[+	9,95%	**76,83%**	0,03%	11,45%	1,73%	100,00%
Total		3,36%	**86,12%**	0,01%	3,84%	6,67%	100,00%

Tableau 128 : Résultats de détection du TTU A en position initiale.

Détection des TTU « A » en position médiane

Types de TTU annotés		Types de TTU détectés				
	IU	!NM	AM	NM	(blank)	Total
AM	[!|	0,00%	**96,83%**	0,00%	3,17%	100,00%
	[!+	11,81%	**88,19%**	0,00%	0,00%	100,00%
	[?|	0,00%	**100,00%**	0,00%	0,00%	100,00%
	[?+	0,00%	**80,00%**	0,00%	20,00%	100,00%
	[|	0,00%	**94,57%**	0,00%	5,43%	100,00%
	[+	15,63%	**74,04%**	8,41%	1,92%	100,00%
Total		4,33%	**89,48%**	2,06%	4,13%	100,00%

Tableau 129 : Résultats de détection du TTU A en position médiane.

Détection des TTU « A » en position finale

Types de TTU annotés		Types de TTU détectés				
	IU	!NF	ATF	NF	(blank)	Total
AF	[!+	13,30%	**51,60%**	35,10%	0,00%	100,00%
	[+	6,43%	**59,38%**	34,01%	0,18%	100,00%
Total		10,10%	**55,22%**	34,59%	0,09%	100,00%

Tableau 130 : Résultats de détection du TTU A en position finale.

34, 59% des TTU A en position finale sont confondus avec des nucléus.

8.10.2.5 Conclusion

Dans l'ensemble, les détections de proéminences sont meilleures dans les modèles britanniques adaptés.

Le nucléus non marqué est bien détecté en position initiale, finale et 1F. Toutefois la position médiane reste problématique, car la plupart des nucléus sont confondus avec des accents mélodiques « AM ».

Le nucléus emphatique est à nouveau mieux détecté en position médiane, mais tend à se confondre avec un nucléus non marqué en position initiale et finale.

Les accents mélodiques sont très bien identifiés à l'exception des accents mélodiques postnucléaires souvent confondus avec des nucléus non marqués.

Nous présenterons dans la section suivante, une autre façon de détecter à la fois les types d'IU et de TU.

8.10.3 Détection jointe des types de TU-IU

Cette étape permet de détecter à la fois les types d'unités intonatives et les types d'unités tonales. Les informations fonctionnelles connues sont alors les frontières des IU et celles des TU. Connaissant donc les frontières d'une unité intonative et des unités tonales qui la composent, on peut faire les hypothèses suivantes : si l'on suppose que cette IU est de type terminal « [| », alors toutes les séquences optimales associées aux TU « [|N1 », « [|SM », « [|A1 » etc., seront testées au niveau de chaque TU jusqu'à que l'on obtienne le meilleur score possible au niveau de l'unité intonative. La même procédure est ensuite répétée pour chaque type d'unité intonative hypothétique (« [+ », « [!+ », « [?| », etc.). Enfin, le type d'IU et la séquence de TU qui permettent d'obtenir le meilleur score sont retenus.

Les deux modèles ayant obtenu les meilleurs scores seront retenus pour cette étape supplémentaire : le modèle 11 et le modèle 24. Ces modèles sont basés sur la modélisation réduite de l'alignement et cinq meilleures séquences optimales au lieu de dix comme dans les modèles

précédents. À partir de ces modèles simplifiés et de l'étape de détection jointe des types d'IU et de TU, nous présenterons les résultats globaux de détection des fonctions dans les modèles des TU et les modèles des TTU.

8.10.3.1 Détection des types d'IU

Résultats dans un modèle basé sur les TU : le modèle 11

Types d'IU annotés	Types d'IU détectés								
	[|	[+	[!|	[!+	[?|	[?+	[	+	Total
[|	**43.56%**	10.29%	22.86%	6.03%	13.97%	1.82%	0.31%	1.16%	100.00%
[+	22.31%	**26.78%**	7.50%	14.81%	14.03%	9.92%	2.24%	2.42%	100.00%
[!|	33.42%	4.30%	**43.54%**	7.59%	6.58%	3.54%	0.25%	0.76%	100.00%
[!+	16.79%	12.28%	3.01%	**45.86%**	6.77%	13.53%	0.50%	1.25%	100.00%
[?|	27.03%	0.00%	35.14%	0.00%	**37.84%**	0.00%	0.00%	0.00%	100.00%
[?+	4.00%	0.00%	0.00%	13.33%	0.00%	**72.00%**	4.00%	6.67%	100.00%
[	14.49%	10.14%	7.25%	4.35%	8.70%	33.33%	**11.59**%	10.14%	100.00%
|	37.17%	8.85%	14.16%	1.77%	8.85%	6.19%	11.50%	11.50%	100.00%
+	6.80%	2.91%	0.00%	8.74%	5.83%	30.10%	1.94%	**43.69%**	100.00%
Total	32.52%	14.64%	17.18%	11.75%	12.56%	7.26%	1.36%	2.73%	100.00%

Tableau 131 : Matrice de confusion des types d'IU annotés et détectés dans le modèle 11.

De façon générale les résultats de détection sont légèrement inférieurs aux résultats obtenus dans les expériences précédentes pour lesquelles nous disposions au départ de plus d'informations fonctionnelles (le degré de proéminence des TU appartenant à l'IU considéré était déjà connu). Les IU « [?+ » et « [!+ » obtiennent les taux de détection les plus élevés (respectivement 72% et 45,86%).

Résultats dans un modèle basé sur les TTU2 : le modèle 24

Types d'IU annotés	Types d'IU détectés									
	[|	[+	[!|	[!+	[?|	[?+	[	|	+	Total
[|	**43.60%**	13.78%	18.45%	9.99%	8.37%	3.93%	0.23%	0.18%	1.48%	100.00%
[+	19.86%	**28.83%**	8.06%	19.79%	6.86%	10.81%	1.34%	0.21%	4.24%	100.00%
[!|	11.60%	11.29%	**49.84%**	14.11%	5.33%	6.90%	0.00%	0.00%	0.94%	100.00%
[!+	6.03%	15.56%	8.57%	**40.63%**	6.35%	16.51%	0.00%	0.00%	6.35%	100.00%
[?|	32.26%	9.68%	6.45%	19.35%	**25.81%**	6.45%	0.00%	0.00%	0.00%	100.00%
[?+	6.00%	4.00%	0.00%	22.00%	8.00%	**54.00%**	6.00%	0.00%	0.00%	100.00%
[	11.11%	0.00%	14.81%	0.00%	33.33%	14.81%	**14.81%**	0.00%	11.11%	100.00%
|	27.78%	10.19%	10.19%	4.63%	23.15%	13.89%	3.70%	**0.00%**	6.48%	100.00%
+	4.55%	11.36%	3.41%	3.41%	5.68%	18.18%	9.09%	0.00%	**44.32%**	100.00%
Total	29.45%	18.09%	15.92%	15.37%	8.10%	8.33%	0.95%	0.16%	3.63%	100.00%

Tableau 132 : Matrice de confusion des types d'IU annotés et détectés dans le modèle 24.

Dans l'ensemble, on obtient des taux de détection inférieurs à ceux qui sont trouvés à partir du modèle 11. Seul l'IU emphatique terminal « [!| » est mieux détecté dans le modèle des TTU. Il s'agit en effet d'un type d'énoncé mieux modélisé au niveau de l'unité tonale élargie comme on l'a vu plus tôt.

8.10.3.2 Détection des types de proéminence

Types de TTU annotés	Types de TTU détectés						
	!N	A	N	S	U	(blank)	Total
!N	**38,17%**	11,38%	39,90%	9,92%	0,00%	0,64%	100,00%
A	4,90%	**49,81%**	11,61%	32,93%	0,00%	0,75%	100,00%
N	27,34%	11,42%	**46,75%**	13,10%	0,00%	1,39%	100,00%
S	7,07%	27,06%	12,65%	**52,40%**	0,00%	0,83%	100,00%
U	0,00%	0,00%	0,00%	0,00%	**99,91%**	0,09%	100,00%

Tableau 133a : Résultats de détection des types de proéminence dans les modèles britanniques.

La détection des types de proéminence est sensiblement meilleure dans cette étape de détection pour laquelle les séquences optimales sont simplifiées (on retient cinq séquences par type de TU). Les meilleurs taux de détection sont obtenus pour les accents rythmiques et mélodiques (respectivement 52,40% et 49,81%). On observe une même amélioration dans les modèles basés sur les TTU.

Types de TTU annotés	Types de TTU détectés					
	!N	A	N	U	(blank)	Total
!N	**44,03%**	12,09%	40,69%	0,00%	3,19%	100,00%
A	4,57%	**80,76%**	9,61%	0,29%	4,77%	100,00%
N	32,38%	9,23%	**55,86%**	0,00%	2,53%	100,00%
U	0,00%	0,27%	0,00%	**98,02%**	1,71%	100,00%

Tableau 133 : Résultats de détection des types de proéminence dans les modèles britanniques adaptés.

Les TTU de type « A » obtiennent le meilleur taux de détection qui est de 80,76% suivi par l'unité nucléaire détectée dans 55,86% des cas.

Dans les sections suivantes, nous décrirons les tentatives de détection des frontières de TTU et d'IU.

8.10.4 Détection des frontières de TTU

On cherche à détecter automatiquement les frontières des unités mélodiques globales dans chaque IU. Les informations fonctionnelles connues sont les frontières d'IU et les accents lexicaux (ce qui nous permet de déduire les frontières des unités tonales). La méthode de détection des frontières de TTU est identique à la détection des frontières d'IU à l'exception du fait que l'on s'appuie sur les débuts et fins de TU comme points de repère plutôt que sur les frontières de mots. Un début de TTU coïncide toujours avec un début de TU donc c'est à ce niveau que l'on testera

les meilleures séquences de tons possibles. Le découpage en TTU dont les séquences optimales correspondantes permettront d'obtenir la meilleure corrélation avec MOMEL d'origine, sera retenu.

On teste cette détection automatique à partir du modèle 24 simplifié comportant cinq séquences optimales par fonction. Pour ne pas fausser les résultats, les TTU dont la frontière gauche coïncide avec le début d'une unité intonative tels que les TTU « U0 », les TTU « A1 », « N1 » et « !N1 » qui ne sont pas précédés d'un TTU « U0 », ne sont pas pris en compte.

Les résultats sont d'abord présentés de façon globale et ensuite en fonction de la position du TTU annoté :

TTU annotés	Nombre de TTU annotés	Nombre de frontières de TTU détectées	Pourcentages
A	1298	734	56.55%
N	740	480	64.86%
!N	722	573	79.36%
Total	2760	1787	64.75%

Tableau 134 : Résultats de détection des frontières de TTU dans le modèle 24.

Le meilleur taux de détection est obtenu pour les frontières de TTU « !N » (79,36%), suivi des TTU « N » (64,86%), des TTU « A » dont les frontières sont détectées dans 56,55% des cas. On peut détailler ces résultats en fonction de la position du TTU dans l'IU.

TTU annotés	Nombre de TTU annotés	Nombre de frontières de TTU détectées	Pourcentages
A1	722	457	63.30%
AM	424	207	48.82%
AF	146	66	45.21%
A1F	6	4	66.67%
N1	8	8	100.00%
NM	9	5	55.56%
NF	587	342	58.26%
N1F	136	125	91.91%
!N1	48	42	87.50%
!NM	55	40	72.73%
!NF	463	344	74.30%
!N1F	156	147	94.23%
Total	2760	1787	64.75%

Tableau 135 : Résultats de détection des frontières de TTU en fonction du type de TTU et de la position dans l'IU.

Les meilleures détections de frontières de TTU sont obtenues pour les TTU « N1 » suivies par les TTU « N1F » et « !N1F » (exactes dans 94,23% des cas).

Le taux de détection le moins élevé est obtenu avec les TTU « A » en position médiane et finale (48,82% et 45,21%).

Exemple tiré du TextGrid G0101G (4,6-8,8) :

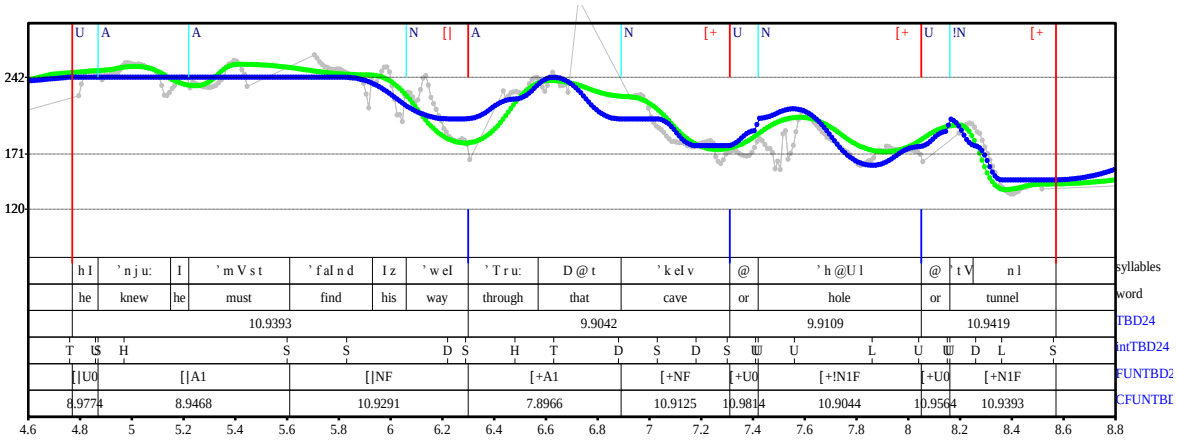


Figure 110 : Courbe issue du modèle de détection des frontières de TTU comparée à MOMEL.

> Courbe issue du modèle de détection des frontières de TTU (en bleu), courbe MOMEL (en vert), frontières annotées en haut (bleu), frontières détectées en bas (bleu) et TextGrid correspondant
>
> Rangée 1= phonèmes, rangée 2= mots, rangée 3= valeur du score pour la courbe issue du modèle de détection et MOMEL d'origine, rangée 4= séquences de tons INTSINT optimisées dans le modèle 24, rangée 5= fonctions détectées, rangée 6= score obtenu à partir des fonctions détectées et par TTU.

Dans l'exemple ci-dessous, la majeure partie des frontières de TTU sont bien détectées. Seul un TTU n'est pas détecté au début de l'énoncé, il s'agit d'un TTU « A » en position médiane commençant au niveau du mot « *must* ».

8.10.5 Détection des frontières d'unités intonatives (IU)

La méthode de détection des frontières est identique à celle qui a été développée à partir du corpus EUROM1 (cf. section 7.4.3). Les informations fonctionnelles connues sont les pauses silencieuses, les frontières de mots et les accents lexicaux. Trois modèles de détection des frontières sont testés sur la totalité du corpus : le modèle 1 basé sur une optimisation des formes au niveau des énoncés entre deux pauses, le modèle 3 basé sur l'optimisation au niveau de l'IU et le modèle 11 basé sur IF britannique (annoté en TU).

8.10.5.1 Détection des frontières d'IU dans les modèles 1 et 3

Les résultats de détection automatique sont comparés avec l'annotation originale du corpus :

Type d'unité intonative	Nombre d'IU annotés	Nombre de frontières détectées avec le modèle 1	Nombre de frontières détectées avec le modèle 3	% frontières détectées avec le modèle 3	% frontières détectées avec modèle 3 incluant le mot suivant
[+	262	86	104	40%	61%
[+_NW		68	56		
[|	116	28	43	37%	56%
[|_NW		24	22		
[!+	18	10	8	44%	67%
[!+_NW		2	4		
[!|	7	2	3	43%	57%
[!|_NW			1		
[?+	1		1	100%	100%
[?+_NW					
[?|	1		1	100%	100%
[?|_NW					
+	1			0%	0%
[	16	2	2	13%	31%
[_NW		2	3		
Total	**422**	**224**	**248**	**38%**	**59%**
Frontières ajoutées		634	567	134%	

Tableau 136 : Résultats de détection des frontières d'IU en fonction du type d'IU dans les modèles 1 et 3.

Le modèle 3, optimisé à partir des unités intonatives annotées manuellement, permet une meilleure détection des frontières que le modèle 1 optimisé à partir des énoncés entre deux pauses. Pour les IU terminaux, on obtient (en prenant en compte les frontières décalées d'un mot en général inaccentué) un taux de détection de 37% pour le modèle 1 et 56% pour le modèle 3. Pour les IU non terminaux ce taux est de 40% pour le modèle 1 et de 61% pour le modèle 3.

Un peu plus d'une frontière sur deux est détectée correctement ou un mot après la frontière manuelle mais beaucoup de frontières sont aussi rajoutées de manière intempestive : 634 frontières sont ajoutées avec le modèle 1 et 567 avec le modèle 3.

L'observation des résultats de prédiction des frontières dans les fichiers « TextGrid » montre qu'il s'agit du même type d'erreur que dans les expériences menées sur le corpus EUROM1. Des frontières sont ajoutées au début d'une proéminence forte telle qu'accent mélodique « A » ou un nucléus dans la plupart des cas. Le corpus MARSEC utilisé est d'un style plus expressif et marqué si bien que ces proéminences sont plus nombreuses et expliquent probablement le grand nombre de frontières rajoutées.

Exemple tiré du fichier G0402G :

Frontières annotées : *« A lot of the passengers +were travelling home |to spend christmas with their families.| »*

Frontières détectées : *« A lot of the passengers + were travelling home to |spend Christmas |with their families.| »*

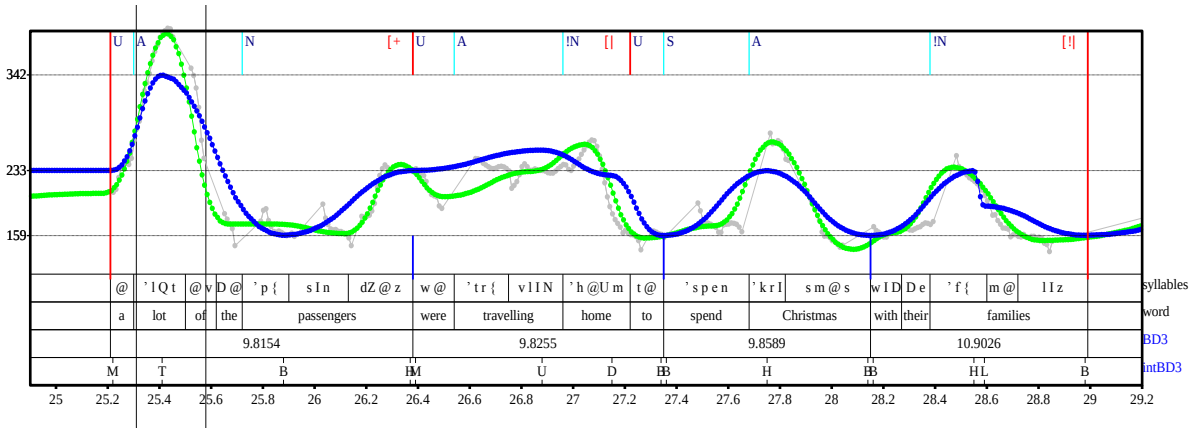


Figure 111 : Exemple de frontières détectées dans le modèle 3.

> On distingue en bleu la courbe issue du modèle 3 et en bas de la figure les frontières détectées, en vert la courbe MOMEL d'origine et en rouge les frontières annotées manuellement.
>
> Rangée 1= syllabes, rangées 2= mots, rangée 3= valeurs du score, rangée 4= séquences de tons INTSINT optimisées avec le modèle 3 et les nouvelles frontières détectées.

Dans la figure ci-dessus, on observe les cas de figure les plus fréquents : tout d'abord, la première frontière est détectée correctement après « *passengers* », ensuite la frontière du deuxième énoncé est détectée un mot inaccentué après la frontière manuelle (après « *to* »), enfin on détecte une frontière supplémentaire avant le nucléus emphatique porté par « *families* ». On ne parvient toujours pas dans ce modèle à distinguer le début d'un énoncé de celui d'une proéminence.

Les résultats de détection des frontières intonatives à partir du modèle 11 sont présentés dans la section suivante.

8.10.5.2 Détection des frontières d'IU dans le modèle 11

De façon globale, on détecte plus de frontières intermédiaires à partir du modèle 11 :

Type d'unité intonative	Nombre d'IU annotés	Nombre de frontières détectées avec le modèle 11	% frontières détectées avec le modèle 11	% frontières détectées incluant le mot suivant
[+	260	122	47%	68%
[+_NW		54		
[|	116	53	46%	73%
[|_NW		32		
[!+	20	7	35%	50%
[!+_NW		3		
[!|	7	3	43%	86%
[!|_NW		3		
[?+_NW				
[?+	1		0%	0%
[?|	1		0%	0%
[?|_NW				
+	1		0%	0%
[	16	4	25%	44%
[_NW		3		
Total	**422**	**284**	**45%**	**67%**
Frontières ajoutées		**549**	**130%**	**130%**

Tableau 137 : Résultats de détection des frontières d'IU en fonction du type d'IU pour modèle 11.

85/116 frontières terminales sont détectées contre 52 dans le modèle 1 et 65 dans le modèle 3, ce qui équivaut à un taux de réussite de 73% si l'on prend en compte les frontières décalées d'un mot, la plupart du temps inaccentué. 176/260 frontières non terminales sont détectées contre 154 et 160 dans les modèles précédents ce qui donne un taux de réussite de 68%%. On trouve le même type d'erreur que dans les modèles précédents (déplacement d'un mot inaccentué et confusion avec le début d'une proéminence).

Observons les résultats de détection de frontières pour le même exemple que dans la section précédente :

Exemple tiré du fichier G0401G de 24,9 à 29,2 :

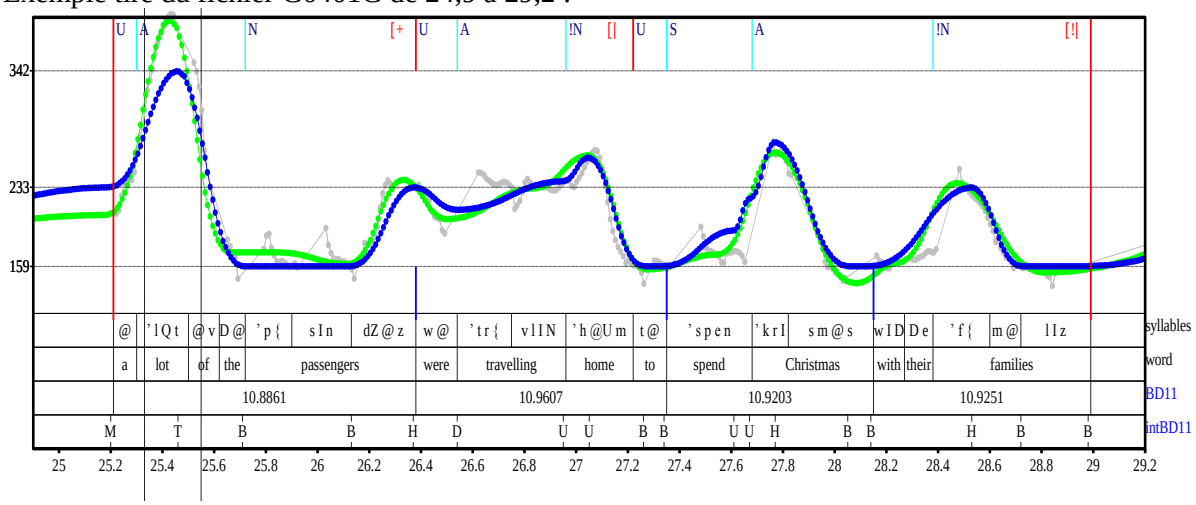


Figure 112 : Exemple de frontières détectées dans le modèle 11.

> On distingue en bleu la courbe issue du modèle 5 et en bas de la figure les frontières détectées, en vert la courbe MOMEL d'origine et en rouge les frontières annotées manuellement.
>
> Rangée 1= syllabes, rangée 2= mots, rangée 3= valeurs du score, rangée 4= séquences de tons INTSINT optimisées avec le modèle 11 et les nouvelles frontières détectées.

Les mêmes frontières sont détectées (la première exactement, la deuxième après « *to* », la dernière rajoutée avant « *with* »). La modélisation de la courbe MOMEL issue du modèle 11 est toutefois bien meilleure qu'à partir du modèle 3.

Bien que les proéminences soient prises en compte dans ce modèle, cela ne suffit pas à permettre de distinguer les débuts d'énoncés des débuts de proéminences. À nouveau, le paramètre mélodique s'avère insuffisant dans ce cas-là.

8.10.6 Conclusion provisoire sur les détections de fonctions

À partir des modèles optimaux développés par analyse par synthèse, nous avons mis en place trois types de détection de fonctions : une détection des types d'énoncés ou types d'unités intonatives, une détection des niveaux de proéminence et la détection des frontières (frontières d'IU et de TTU).

La reconnaissance des différentes fonctions montre une marge de confusion importante dans certains cas, mais il faut se rappeler que l'on ne s'occupe que de l'intonation et que l'on ne retient pour chaque étiquette fonctionnelle que les formes les plus courantes, les plus utilisées.

En ce qui concerne la reconnaissance des types d'IU, on détecte bien les énoncés spéciaux comme les parenthèses basses, les questions non terminales, les énoncés emphatiques par rapport aux énoncés les plus courants et les moins marqués « [| » et « [+ ».

Les modèles des TTU2 permettent une meilleure reconnaissance des contours emphatiques.

C'est lorsque le modèle est basé sur une moyenne de nombreux cas que la détection est la moins bonne. Il faudrait affiner les étiquettes « [+ » et « [| ».

Les meilleurs résultats de détection des proéminences sont obtenus avec les paramètres du modèle 11. De façon générale, on remarque que la position médiane est plus problématique quel que soit le type de TU. Le cas du nucléus en position médiane confondu avec un accent rythmique en est un bon exemple. Le même problème de détection se pose notamment avec les accents mélodiques en position médiane. Les niveaux de proéminence étiquetés en position initiale et finale donnent de meilleurs résultats.

Les meilleurs résultats de détection des fonctions sont obtenus à partir des modèles en TTU2. Les frontières de TTU sont très bien détectées.

Les résultats de détection des frontières d'unités intonatives sont similaires aux résultats obtenus à partir du corpus EUROM1. On détecte environ une frontière sur deux.

Ces résultats permettent d'avoir une vision globale de la performance de nos modèles en matière de reconnaissance des fonctions. Afin de mieux comprendre et expliquer les erreurs et confusions constatées, il faut se pencher sur les données du corpus et plus particulièrement sur les prédictions de nos modèles. Il s'agira de comparer ce que nos modèles peuvent prédire avec les enregistrements du corpus et essayer de relever des manques systématiques dans le but d'enrichir la modélisation de départ. Nous nous baserons sur une sélection réduite de modèles ayant donné les meilleurs résultats en matière de score et de détection de fonction : les modèles 11 pour le niveau des TU et le modèle 24 pour les TTU2.

8.11 Retour sur les données et évaluation des modèles par analyse par synthèse

Après avoir développé une procédure d'optimisation de la relation entre l'annotation des formes et des fonctions et par là-même mis en place plusieurs modèles de l'articulation forme/fonction, nous allons approfondir l'analyse des résultats en nous appuyant sur des exemples précis. Les différents tableaux de résultats chiffrés présentés précédemment permettent de se faire une idée de la performance des modèles de détection. Nous allons maintenant entrer dans les détails pour mieux comprendre et analyser les erreurs et manques de nos modèles et essayer d'atteindre notre objectif : se nourrir des erreurs pour enrichir le modèle de départ de façon incrémentale.

Nos modèles seront évalués du point de vue de leur pouvoir de prédiction : prédiction des contours de F0 à partir de l'annotation fonctionnelle de base et reconnaissance des fonctions à partir du modèle des formes prosodiques. Nous avons déjà présenté les résultats d'un point de vue objectif en utilisant les critères de corrélation linéaire et l'écart quadratique moyen. Reste maintenant à illustrer le processus d'évaluation subjective de nos modèles.

Il s'agira d'identifier les passages mal modélisés et d'en écouter la resynthèse par comparaison aux enregistrements d'origine pour s'intéresser aux raisons. La question qui se pose est la suivante : que pouvons-nous apprendre des « ratés » de notre modèle ?

Du point de vue de la détection des fonctions, nous chercherons à mieux comprendre et illustrer les cas où les fonctions sont mal attribuées dans la même optique d'en apprendre quelque chose et d'améliorer notre modèle de départ.

8.11.1 Hypothèses expliquant les échecs de modélisation

On peut supposer que les échecs de modélisation peuvent provenir de deux problèmes différents : un problème de modélisation et d'analyse des formes prosodiques ou bien un problème de modélisation et d'analyse des fonctions prosodiques.

Si en un point donné, nous ne parvenons pas à modéliser un contour prosodique, ceci peut s'expliquer de plusieurs façons.

Il s'agit d'un problème de forme :

- Nous avons affaire au niveau d'une unité tonale par exemple à un contour mélodique non commun, celui-ci n'apparaît pas dans les formes optimales retenues pour les modèles ;
- Nous avons affaire à un problème de codage INTSINT, en particulier un problème de registre ;

- Nous avons affaire à un problème d'alignement : le modèle que nous avons retenu a des limites et ne permet pas de couvrir tous les cas, c'est simplement celui qui obtient le meilleur score moyen ;
- Le calcul de la corrélation et de l'écart quadratique moyen conduit à faire le mauvais choix : c'est l'évaluation objective utilisée dans notre modèle qui est remise en cause.

Il s'agit d'un problème fonctionnel :

- Il y a un problème d'annotation. La conversion automatique TSM-IF donne un résultat erroné. L'annotation subjective est discutable (la solution dans un travail futur : diminuer les erreurs et les facteurs de subjectivité en utilisant plusieurs experts annotateurs). Peut-être que ce type de fonction pose un problème récurrent (difficulté à percevoir un niveau de proéminence) ? On pense à remettre en question l'étiquette fonctionnelle ambiguë.
- Nous avons affaire à une fonction non prise en compte dans le modèle de départ. Nous observons une forme particulière récurrente associée à cette fonction et nous devons l'étiqueter.
- La configuration de F0 paraît tout à fait correcte, mais ce n'est pas l'élément le plus important dans la perception de cette fonction : dans ce cas, ce sont la durée ou l'intensité ou autre qui dominent. Nous arrivons aux limites de ce que l'on peut prédire à partir de l'intonation seulement.

Nous allons donc nous pencher sur plusieurs exemples tirés de l'observation des modèles 11 et 24.

8.11.2 Relevé de problèmes de modélisation à partir des modèles 11 et 24

8.11.2.1 Méthode

Les résultats de la modélisation sont regroupés dans un fichier EXCEL. Les scores et meilleures formes trouvées sont notés TU par TU à partir du corpus en entier. Une colonne est aussi ajoutée à partir du score du modèle pour chaque TU. On peut avoir la mention TB (très bien) pour identifier un score supérieur ou égal à 9, B (bien) pour un score supérieur ou égal à 7 ou PB (problème) pour cas où les scores sont moins élevés. Un simple tableau croisé montrant par exemple les types de TU et le nombre de « bonnes » ou « mauvaises » notes permet alors d'obtenir une liste des cas problématiques pour chaque fonction. On prend l'exemple de la modélisation des types de proéminence à partir du modèle 11.

	Notes obtenues			
TU annotés	TB	B	PB	Total
!N	647	71	83	801
A	803	210	208	1221
N	647	94	91	832
S	922	233	357	1512
U	1096		1	1097
Total	4115	608	740	5463

Tableau 138 : Répartition des notes obtenues en fonction du type de TU modélisé dans le modèle 11.

Le tableau récapitulatif permet de voir par exemple, que la modélisation du nucléus emphatique donne une note de type TB dans 647 TU, mais une mauvaise note PB dans 83 cas. Il suffit ensuite de cliquer sur le chiffre qui nous intéresse pour accéder à une liste des TU concernés contenant l'adresse détaillée (locuteur, fichier, temps, etc.) pour visualiser l'ensemble des cas.

Nous développerons cette procédure en prenant deux exemples : celui de la modélisation des accents mélodiques puis celui des nucléus emphatiques ou de l'emphase.

8.11.2.2 Le cas des accents mélodiques « A »

La plupart des TU mal modélisés sont du type « [|AM » c'est-à-dire que l'on a affaire à des accents mélodiques en position médiane dans des IU terminaux. On remarque aussi à partir du tableau EXCEL que ces TU sont associés à un codage INTSINT particulier. Trois types de contours ressortent : un contour montant du type {BU}, un contour complexe du type {HLH} et un contour descendant du type {HD}.

Dans le premier cas, la forme associée aux TU « A » est non commune. À partir de la modélisation et de l'optimisation des formes, on obtient en effet essentiellement deux types de contours pour un accent mélodique : un contour descendant {HL} ou plus rarement une montée {_U}. En position médiane et quel que soit le type d'IU concerné, l'optimisation des formes ne fait ressortir qu'un contour descendant unique du type {DL}. Or à partir de la liste retraçant les problèmes de modélisation, on s'aperçoit qu'un ensemble de TU « A » est associé à une forme descendante-montante du type {BU}. À partir du contour moyen extrait du corpus, on ne parvient donc pas à modéliser ces TU. Ils correspondent quasiment systématiquement au niveau des TSM à un ton statique bas noté « _ ».

Prenons un exemple tiré du fichier G0505G :

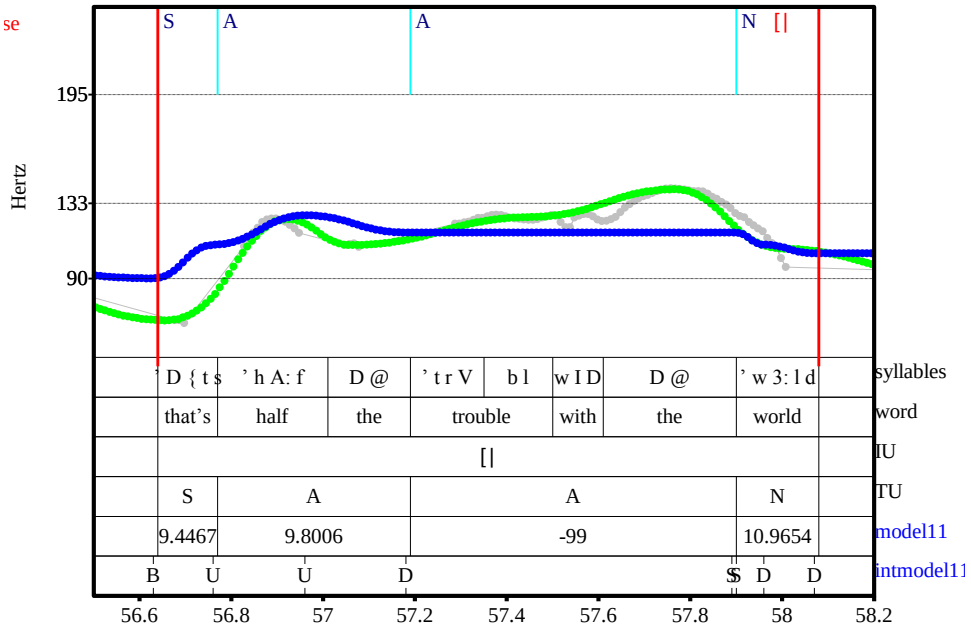


Figure 113 : Exemple de modélisation problématique d'un TU A

Courbe de MOMEL d'origine en vert comparée à la courbe issue du modèle 11 en bleu.

Rangée 1= syllabes, rangée 2= mots, rangée 3= IU, rangée 4= TU, rangée 5= valeurs du score, rangée 6= séquences de tons INTSINT optimales pour le modèle 11.

Il s'agit d'un IU terminal contenant un accent rythmique initial suivi de deux accents mélodiques et d'un nucléus en position finale. Le problème de modélisation est signalé au niveau de la rangée du score par « -99 ». Cela signifie que dans ce cas précis aucune des meilleures formes optimales retenues pour un TU « [|AM » ne permet de suivre la courbe de MOMEL. En effet on a un abaissement de F0 suivi d'un contour montant marqué dans les TSM par un ton statique « _ » et interprété comme un accent mélodique dans le système IF. Il ne s'agit pas ici d'un problème d'annotation erronée, on perçoit bien une proéminence au niveau du TU commençant par « *trouble* ». C'est un type de proéminence particulier qui peut correspondre à ce que Faure appelle un accent démarcatif. La seule étiquette fonctionnelle « A » ne permet pas de faire ressortir ces cas particuliers. Il faudrait donc rajouter une étiquette fonctionnelle spécifique pour annoter ce type de proéminence. Ce cas montre que l'on peut affiner l'annotation fonctionnelle de départ pour l'ensemble des TU « A » associés à un ton statique grave dans les TSM.

On observe le même type de cas problématiques dans les IU emphatiques terminaux.

Prenons un exemple tiré du fichier G0302B :

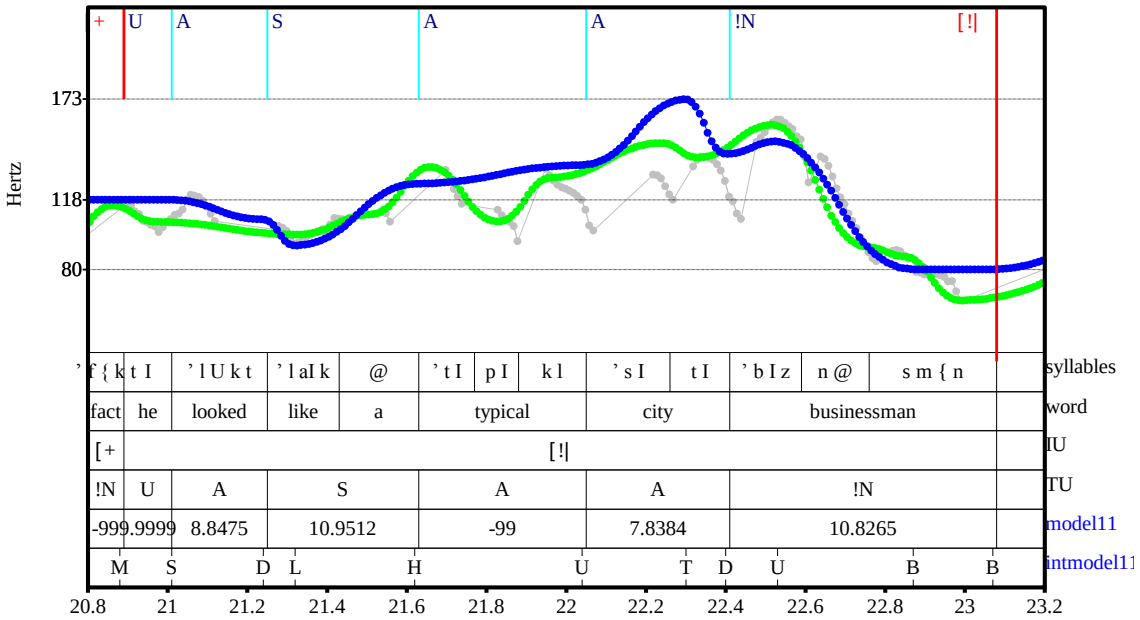


Figure 114 : Exemple de modélisation problématique d'un TU A dans un IU emphatique

Voir figure précédente pour la légende détaillée des rangées.

Au niveau du TU commençant par « *typical*», on observe un contour mélodique complexe du type {HBH}. L'ensemble des TU « [!|A » sont modélisés par des simples chutes du type {HL} dans nos modèles. Ici il n'est pas possible d'obtenir un bon score d'où l'annotation « -99 ». En écoutant l'enregistrement original pour ce cas précis, on se rend compte que l'effet de proéminence est équivalent à celui du ton statique simple. On a un effet accentuel démarcatif. De plus cet accent fait partie d'un contour emphatique général sur les accents mélodiques prénucléaires qui décrivent une montée progressive jusqu'à l'attaque du nucléus (« upstepping head + high fall »).

Le dernier cas correspond à des TU « A » ayant une forme typique mais qui ne sont tout de même pas correctement modélisés. La plupart de ces cas correspondent à des problèmes d'alignement. Les paramètres de variations ont été réduits et simplifiés pour nos modèles si bien qu'ils ne permettent pas de représenter tous les cas possibles. Le second point d'un TU dans le modèle 11 est aligné à 10 ms de la fin du TU, ce qui a permis d'obtenir une bonne corrélation dans la plupart des cas, mais on trouve des contre-exemples.

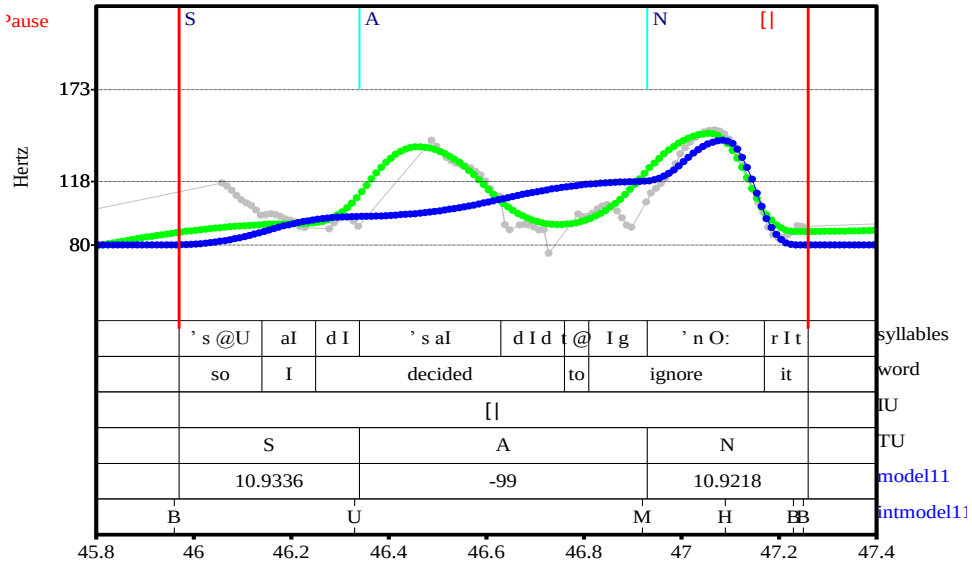


Figure 115 : Exemple de problème de modélisation du TU A dû aux paramètres d'alignement sur G0302B.

Dans l'exemple ci-dessus, l'accent mélodique porté par « *decided* » a une forme typique descendante, mais n'est pas modélisé. Le ton L devrait ici être aligné à une distance largement supérieure à 10 ms de la frontière droite du TU pour parvenir à suivre le mouvement. Avec le modèle 24 basé sur une annotation fonctionnelle en TTU, le problème d'alignement est résolu, car on utilise trois points cibles.

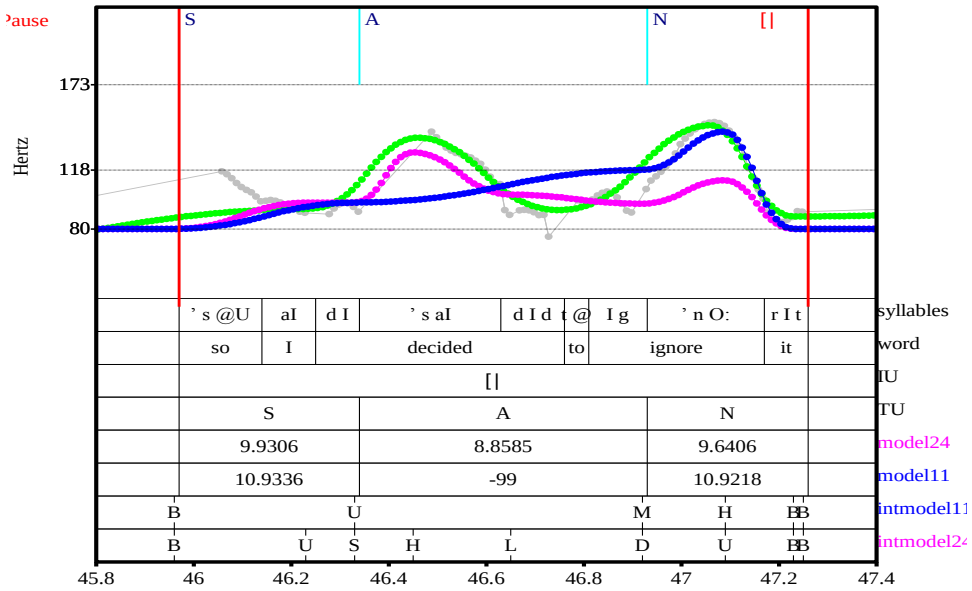


Figure 116 : Comparaison des modèles 11 et 24 pour le même exemple.

Courbe verte= MOMEL d'origine, courbe bleue= générée à partir du modèle 11, courbe rose= générée à partir du modèle 24.

Rangée 1= syllabes, rangée 2= mots, rangée 3= IU, rangée 4= TU, rangée 5= valeurs du score du modèle 24, rangée 6= valeurs du score pour le modèle 11, rangée 7= séquences de tons optimisées dans le modèle 11, rangée 8= séquences de tons optimisées dans le modèle 24.

À partir de la modélisation en TTU, le problème d'alignement est résolu, car on a trois points cibles {HLD}. La chute peut être mieux alignée avec MOMEL d'origine avec ces paramètres.

L'observation des problèmes de modélisation des TU « A » permet de mettre en valeur deux domaines à enrichir. Tout d'abord, l'étiquette « A » est trop générale et l'on peut annoter au moins un type de proéminence supplémentaire, l'accent mélodique démarcatif. Ceci permettra de représenter tous les cas d'accents statiques graves du corpus. Du point de vue de la modélisation des formes, les restrictions de variation de l'alignement peuvent poser problème. Il faudrait élargir la variation de l'alignement du second point à plus de 10 ms de la fin du TU. D'autre part certains TU plus longs que la moyenne nécessiteraient une modélisation en trois points cibles comme dans les modèles basés sur les TTU.

8.11.2.3 Le cas des nucléus emphatiques

Au niveau des nucléus emphatiques l'observation de certains exemples non modélisés par nos modèles montre que l'on peut affiner les étiquettes. Tout d'abord de nombreux cas non modélisés correspondent à un problème de registre. Le registre est typiquement beaucoup plus étendu que la moyenne dans certains cas et le codage INTSINT ne permet pas de les modéliser. Du point de vue de l'annotation fonctionnelle, on pourrait étiqueter ces cas précis comme des cas d'emphase extrêmes associés à un état émotif intense en particulier chez le locuteur G04 ou à des contours vocatifs/chantant chez le locuteur G05. À nouveau l'étiquette générale d'emphase peut être développée pour prendre en compte ces cas précis.

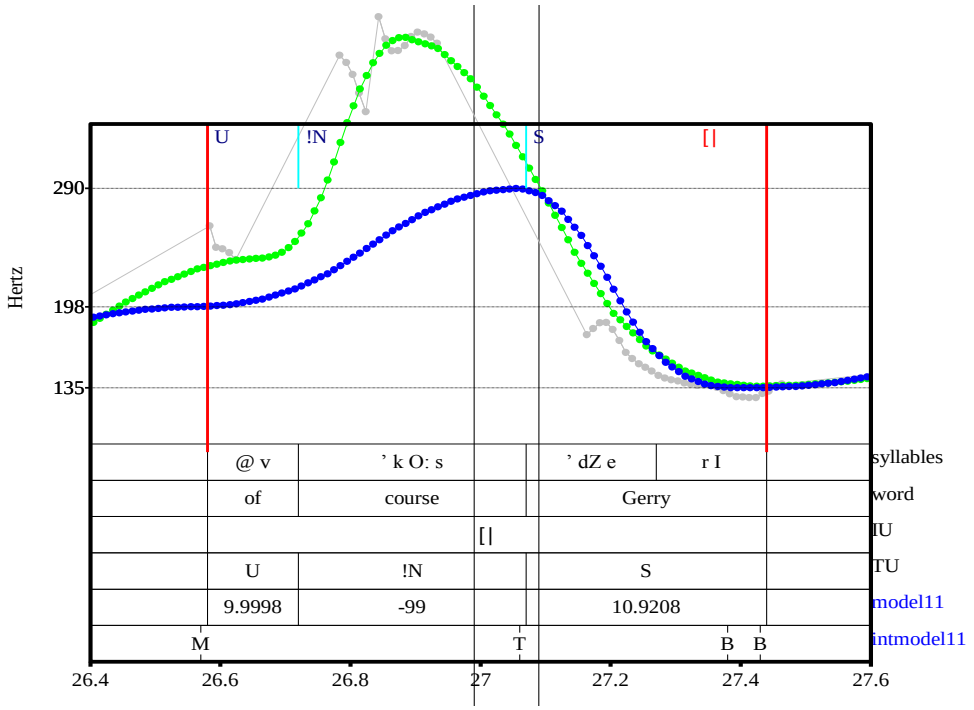


Figure 117 : Exemple de registre étendu localement non modélisé sur le fichier G0103B.

Courbe verte= MOMEL, courbe bleue= générée à partir du modèle 11

Rangée 1= syllabes, rangée 2= mots, rangée 3= IU, rangée 4= TU, rangée 5= valeurs du score, rangée 6= séquences de tons optimisées dans le modèle 11.

Dans l'exemple ci-dessus, le registre est localement supérieur à la moyenne calculée sur l'ensemble du fichier pour le TU « !N ». On perçoit un registre infra-aigu qui rend l'énoncé beaucoup plus marqué. Les formes optimales du TU ! »N » ne permettent pas de suivre la courbe correctement. Dans ce cas, on a à la fois un problème de registre et d'alignement.

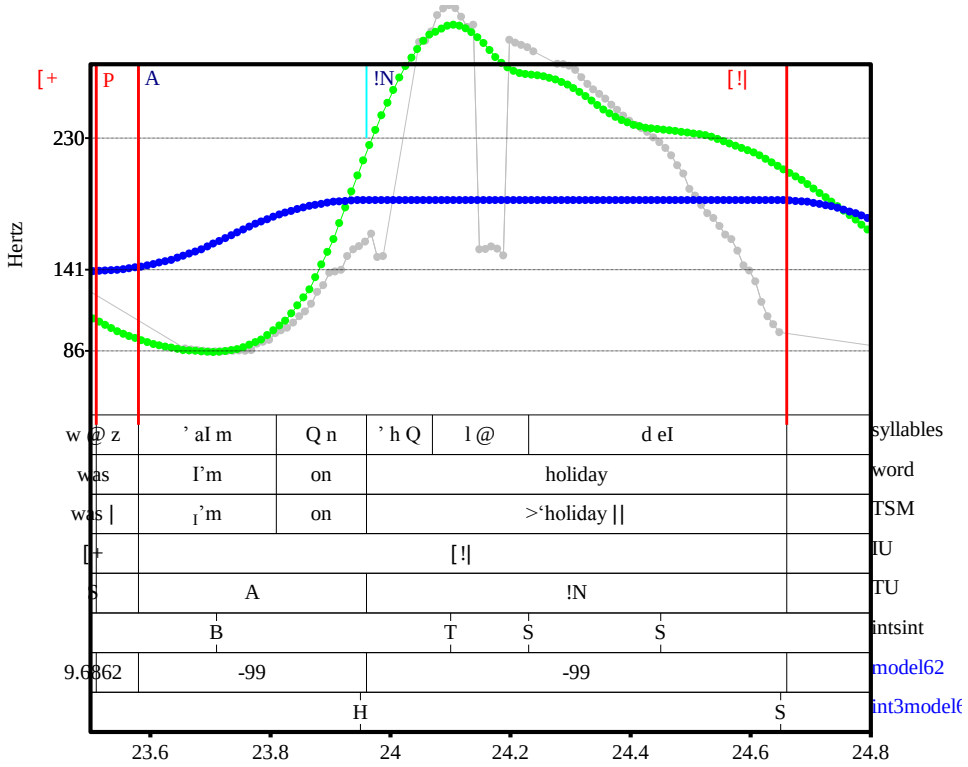


Figure 118 : Exemple de contour vocatif non modélisé sur le fichier G0501G.

On observe toute une série de contours emphatiques non modélisés chez le locuteur G05. Ceux-ci sont annotés comme des nucléus emphatiques et on trouve au niveau de la TSM une indication de registre étendu. Ce cas précis d'emphase pourrait aussi être rajouté dans l'annotation fonctionnelle de départ.

Ce cas d'emphase intensifiée s'accompagne aussi souvent d'un contour de tête particulier qui ne ressort pas dans les modèles optimaux. Il s'agit d'une montée progressive suivie d'une très grande chute (« upstepping head + high fall »).

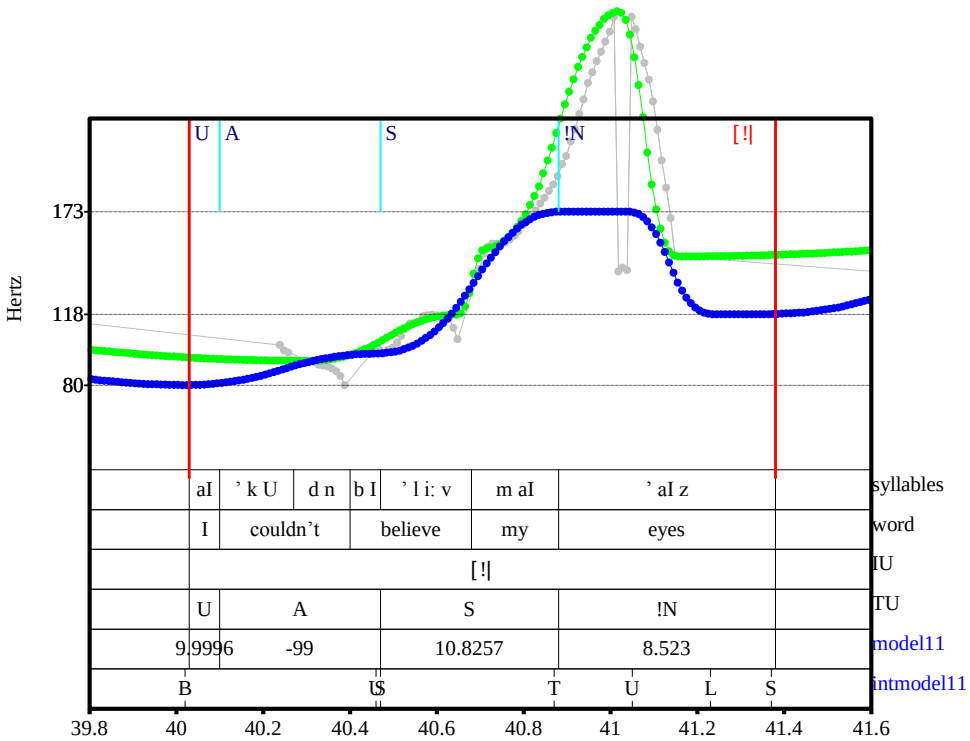


Figure 119 : Exemple de contour montant + grand chute associée à un IU [!| sur G0302B.

Comme on peut l'observer dans l'exemple ci-dessus, la grande chute sur le nucléus emphatique est précédée d'une montée préparatoire sur les accents prénucléaires. C'est un contour intonatif classique qui n'est pas représenté dans nos modèles. En effet, l'optimisation des formes au niveau des énoncés emphatiques terminaux donne en priorité un contour com posé de suites de chutes et d'une grande chute sur le nucléus (« falling Head +high Fall »). L'annotation de l'emphase au niveau de l'IU peut donc être affinée en incluant l'ensemble de ces cas.

Nous présenterons le même travail d'analyse des résultats dans la prochaine section à partir d'une série d'exemples de problèmes de détection de fonctions.

8.11.3 Exemple de détections de fonctions problématiques

8.11.3.1 Le cas de la détection du nucléus en position médiane

On a déjà vu dans la section précédente que le taux de détection des nucléus en position médiane était le plus bas. En observant quelques exemples, on remarque deux types de cas problématiques. Le nucléus est confondu avec un accent rythmique ou mélodique quand les formes optimales sont très proches ou bien on observe un cas d'annotation fonctionnelle ambiguë ou discutable.

Tout d'abord, le nucléus est très souvent confondu avec un accent rythmique postnucléaire. Dans ces cas, on remarque que la forme associée au nucléus est identique à celle d'un accent postnucléaire. Prenons un exemple sur le fichier G0106G :

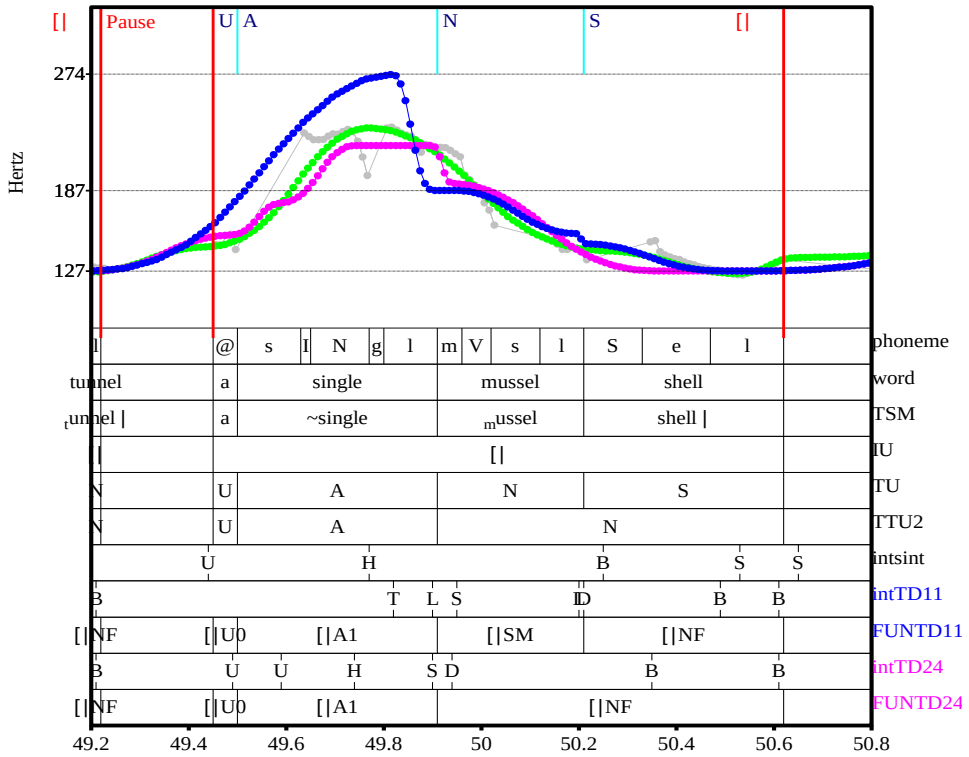


Figure 120 : Exemple de nucléus confondu avec un accent rythmique postnucléaire

> Courbe rose= issue de la détection des fonctions dans le modèle 24, courbe bleue= issue de la détection des fonctions dans le modèle 11
>
> Rangée 1= phonèmes, rangée 2= mots, rangée 3= TSM, rangée 4= IU, rangée 5= TU, rangée 6= TTU2, rangée 7= tons INTSINT, rangée 8= séquences de tons INTSINT optimales pour le modèle 11, rangée 9= fonctions détectées dans le modèle 11, rangée 10= séquences de tons INTSINT optimales pour le modèle 24, rangée 11= fonctions détectées dans le modèle 24.

On compare dans cet exemple la détection de fonction à partir du modèle 11 (en TU dont les résultats apparaissent dans la neuvième rangée « FUNTD11 ») et du modèle 24 (en TTU dont les résultats apparaissent dans la dernière rangée « FUNTD24 »). Les rangées « intTD11 » et « intTD24 » montrent les séquences de tons INTSINT associées aux fonctions détectées.

On perçoit un nucléus sur « *mussel* ». Le contour mélodique associé à celui-ci n'est cependant pas saillant. Il s'inscrit dans la continuité de la courbe avec une séquence INTSINT du type {SD}. La proéminence supérieure perçue sur « *Mussel* » ne se traduit pas par un contour mélodique particulier et peut très bien provenir de l'utilisation d'autres facteurs prosodiques comme la durée ou l'intensité. Le TU qui suit s'accompagne d'une forme typique associée au nucléus final : {DBB}, donc le script interprète la présence d'un TU « SM » suivi d'un TU « NF » ce qui permet

d'obtenir une bonne corrélation avec la courbe de MOMEL. À partir du modèle 24 basé sur une annotation en unités nucléaires globales, on parvient cependant à détecter le nucléus correctement.

On a le même type de problème dans l'exemple suivant pour le fichier G0108B :

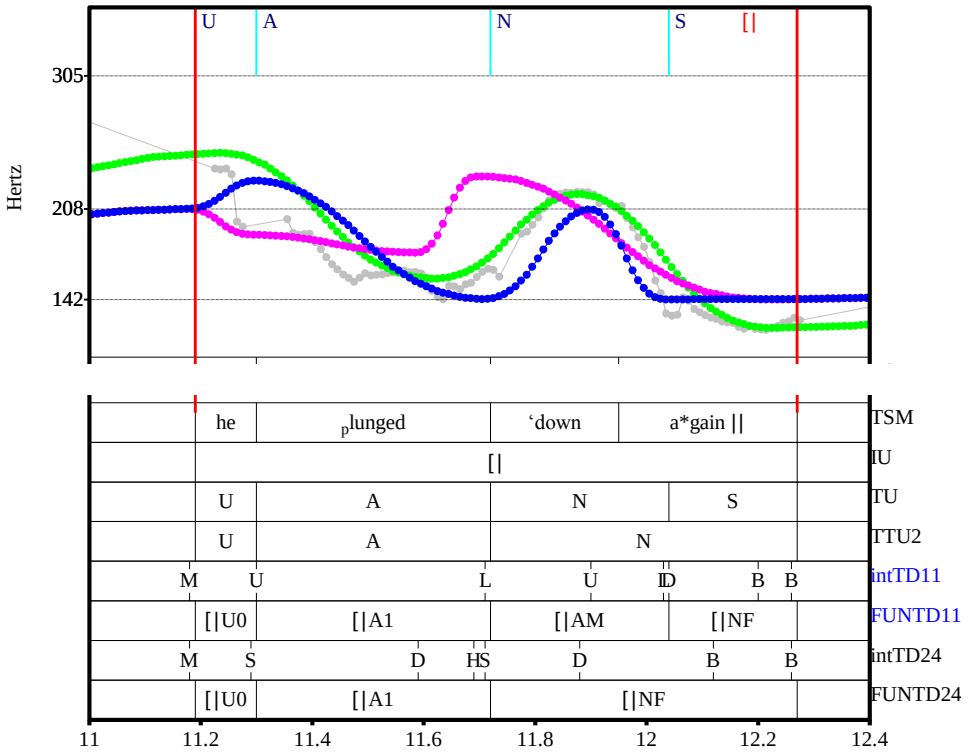


Figure 121 : Exemple de nucléus confondu avec un accent rythmique postnucléaire dans les modèles 11 et 24

> Courbe rose= issue de la détection des fonctions dans le modèle 24, courbe bleue= issue de la détection des fonctions dans le modèle 11
>
> Rangée 1= TSM, rangée 2= IU, rangée 3= TU, rangée 4= TTU2, rangée 5= séquences de tons INTSINT optimales pour le modèle 11, rangée 6= fonctions détectées dans le modèle 11, rangée 7= séquences de tons INTSINT optimales pour le modèle 24, rangée 8= fonctions détectées dans le modèle 24.

Dans le modèle 11, on détecte un TU « AM » sur « *down* » associé à une forme typique {UD} et un TU nucléaire sur « *gain* » associé à la forme typique {DBB}. À nouveau, une analyse en unités nucléaires globales permet d'éviter cette confusion et de reconnaître le mouvement nucléaire correctement mais la modélisation est moins bonne.

Un deuxième cas de figure courant provient d'une annotation fonctionnelle discutable. Cela revient souvent à la possibilité d'analyser un énoncé comme deux unités intonatives ou bien une seule.

Un exemple sur le fichier G0502B :

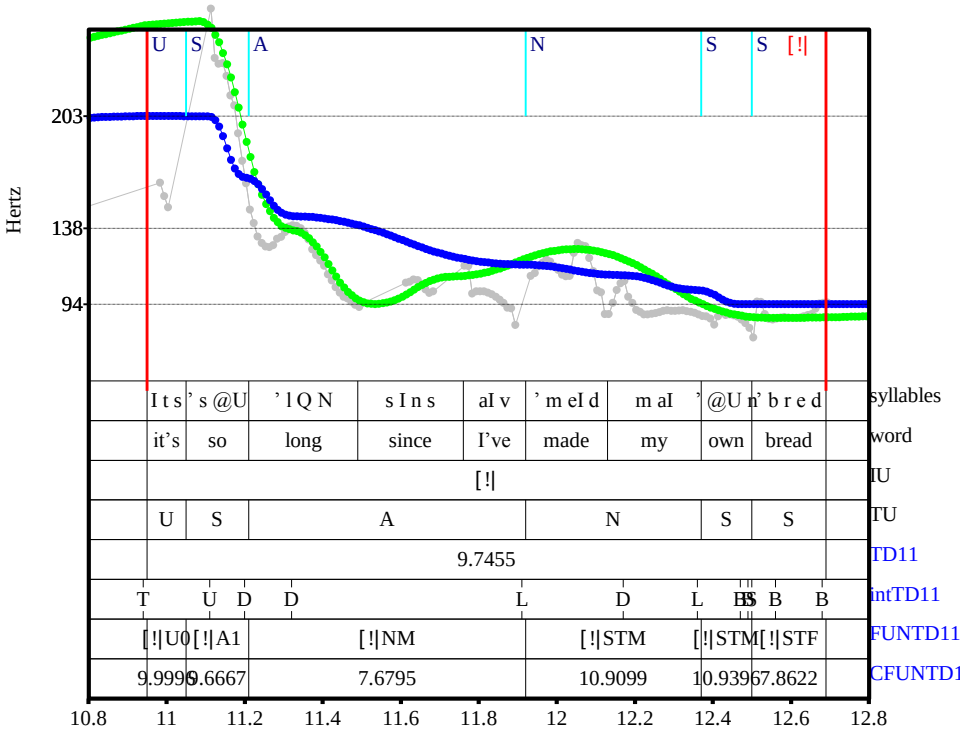


Figure 122 : Exemples de détection de fonction pour un énoncé ambigu

La conversion automatique des TSM en IF donne une unité intonative avec un nucléus en position médiane sur « *made* ». En fait, on aurait pu annoter une frontière après « *long* » ce qui aurait probablement permis une meilleure modélisation de la mélodie.

Chapitre 9

Conclusion et synthèse des résultats

Nous avons développé et testé une méthode d'analyse et d'investigation de la relation entre l'annotation des formes et fonctions prosodiques en anglais britannique. Cette méthode de travail étant basée sur l'analyse par synthèse de l'intonation et le traitement automatique des données mais aussi sur une annotation distincte des formes et fonctions prosodiques en utilisant MOMEL-INTSINT et IF. Nous cherchions d'une part à développer un modèle des formes prosodiques constant à partir de l'annotation en IF et d'autre part à enrichir notre annotation fonctionnelle de base en étudiant les manques de notre modèle et les écarts entre les contours prédits et les enregistrements originaux.

La partie expérimentale de ce mémoire de thèse s'articule autour de trois mouvements principaux. Tout d'abord nous avons proposé une série de modèles explicites de la représentation de l'articulation forme/fonction. On allait donc d'un modèle dit « nul » pour lequel aucune information fonctionnelle n'était prise en compte à un modèle plus complet où les formes prosodiques étaient associées à différentes unités prosodiques comme l'unité intonative et l'unité tonale. Ces modèles étaient de plus en plus complexes et nous cherchions à montrer qu'en ajoutant de plus en plus d'informations fonctionnelles, la prédiction des contours de F0 deviendrait de plus en plus précise. Les différents contours intonatifs étaient fixés manuellement en s'inspirant des contours décrits dans la littérature portant sur l'anglais britannique standard. En utilisant une évaluation objective avec la corrélation linéaire et une évaluation subjective en écoutant les résultats de la resynthèse de nos modèles, nous avons pu montrer que plus on ajoute d'information fonctionnelle, plus notre modélisation des contours de F0 est efficace.

Nous avons ensuite optimisé les paramètres de nos modèles à partir d'un échantillon du corpus EUROM1. Nous avons annoté cet extrait du corpus EUROM1 manuellement en utilisant le système d'annotation fonctionnelle IF. Il s'agissait alors ensuite de dégager un nouveau modèle de l'articulation forme/fonction à partir de cette annotation de base. Ce modèle ne serait pas fait manuellement cette fois-ci mais développé à partir des données en utilisant des méthodes de traitement automatique.

Nous avons donc développé une série de scripts afin d'optimiser les représentations formelles à partir de nos représentations fonctionnelles. C'est-à-dire que l'on a testé des métriques et outils afin de trouver les meilleurs contours représentatifs pour une fonction donnée. Nous avons optimisé le codage des tons INTSINT ainsi que leur alignement. Trois métriques se sont avérées utiles pour extraire les meilleurs contours mais aussi pour évaluer la qualité de nos modèles : la corrélation linéaire que nous avions déjà utilisée dans le premier essai mais aussi le calcul de l'écart quadratique moyen et enfin une combinaison des deux métriques utilisée surtout dans les scripts pour choisir les meilleures séquences.

Les contours canoniques obtenus pour ce corpus sont proches des contours décrits dans les premiers modèles. Nous avons retrouvé notamment des contours similaires au « downstepping head + low rise » ou « downstepping head + low fall » en relation avec les unités intonatives non terminales et terminales. Nous avons obtenu des contours en général moins marqués que ceux que nous avions décrits dans le premier essai. Au niveau d'un énoncé entre deux pauses silencieuses, le contour optimal est une séquence du type {MULL} plutôt que {MTB}. L'opposition entre les étiquettes terminales/non terminales est aussi mieux représentée par de faibles montées du type {LU} voir de simples chutes du type {LD}. Nous avons associé ce contour spécifique à la chute courte ou chute interrompue qui aurait un caractère inachevé correspondant à une annotation de type non terminal. À partir de l'annotation des accents prénucléaires sur le corpus EUROM1, nous avons pu observer la présence de deux types de formes distinctes qui correspondent à nos yeux à deux niveaux de proéminence, l'un étant assez faible et marqué par des contours quasi-statiques et l'autre étant dynamique. Ceci nous a inspiré l'idée d'introduire dans les essais suivants une distinction au niveau prénucléaire entre les accents rythmiques et les accents mélodiques.

Un modèle constant ayant été dégagé, nous avons utilisé ces résultats pour extraire l'information fonctionnelle automatiquement, c'est-à-dire que nous avons essayé de détecter les frontières des unités intonatives et les différents types d'unité tonale. À partir des meilleurs contours extraits du corpus au niveau de l'unité intonative comme au niveau des unités tonales, nous parvenons à identifier à peu près une frontière sur deux. L'utilisation de l'intonation seulement ne nous permet cependant pas de distinguer le début d'un énoncé de celui d'une proéminence majeure. L'information apportée par la mesure de la durée permettant d'identifier un allongement final notamment serait très utile dans ces cas.

Enfin, après avoir développé et testé un ensemble d'outils sur le corpus EUROM1 afin de modéliser les représentations formelles à partir des informations fonctionnelles et inversement, nous nous sommes tournée vers un second corpus dans le but d'enrichir et approfondir notre travail. En effet nous sommes passée d'un corpus de parole de laboratoire de petit calibre à un grand corpus de parole authentique dans l'espoir de pouvoir enrichir l'annotation des formes et des fonctions. L'annotation fonctionnelle a été fournie automatiquement à partir d'une conversion des TSM en IF dont nous avons défini les règles. Cette première conversion automatique a fait l'objet d'une série de corrections manuelles. L'équivalence simple entre les frontières mineures et les frontières non terminales a posé quelques difficultés. Il en a été de même en ce qui concerne l'identification systématique du nucléus à partir du dernier accent mélodique et l'interprétation de l'emphase. Nous avons testé plusieurs versions ou interprétations adaptées du système IF en introduisant notamment dans le premier modèle une distinction entre les accents de type rythmique et les accents de type mélodique. Nous avons aussi testé une modélisation des formes au niveau d'une unité tonale dite élargie correspondant à la prise en compte des accents mélodiques seulement. Nous nous sommes donc posé un ensemble de questions sur les niveaux de proéminence à prendre en compte pour modéliser les contours de F0, sur les unités prosodiques possibles ou sur le rôle du nucléus.

Au niveau des représentations formelles, nous avons testé différentes façons de coder les tons INTSINT et nous nous sommes attachée à approfondir le travail d'optimisation de l'alignement tonal. À l'issue du développement de nos nouveaux modèles de l'articulation forme/fonction, nous avons aussi fait de nouvelles tentatives de détection de fonctions. Nous allons donc, dans les prochains paragraphes, présenter une synthèse des résultats obtenus à partir du travail expérimental effectué sur le corpus MARSEC.

9.1 Annotations fonctionnelles

Tout d'abord, en ce qui concerne les représentations fonctionnelles, nous nous sommes interrogée sur l'unité prosodique à prendre en compte pour prédire et modéliser au mieux les variations de F0. Au moyen de multiples modèles, nous avons testé successivement les énoncés entre deux pauses silencieuses, l'unité intonative, l'unité tonale et deux sortes d'unités tonales élargies. C'est à partir du découpage en unités tonales que l'on parvient à reproduire au mieux les contours de F0. En effet, l'ensemble des modèles qui sont basés sur cette annotation, obtiennent les meilleures moyennes sur tout le corpus que cela soit au niveau de la corrélation ou de l'écart quadratique moyen.

Les modèles basés sur l'unité tonale élargie et l'unité nucléaire que l'on a appelée « TTU1 » sont de loin les moins performants. En modélisant le nucléus comme une unité globale commençant à partir de la syllabe nucléaire et s'étendant jusqu'à la fin de l'énoncé, on perd de toute évidence beaucoup de précision. Les résultats sont significativement améliorés lorsque l'on dégage au moins les accents mélodiques apparaissant après le nucléus dans les énoncés non terminaux (modèles des TTU2).

En regardant les résultats de plus près on s'aperçoit aussi que la modélisation obtenue à partir de l'unité élargie « TTU2 » est en fait la meilleure quand il s'agit des énoncés terminaux emphatiques. Ceci s'explique par la présence de formes complexes de type descendant-montant au niveau des accents prénucléaires. Ces contours sont mieux modélisés avec les paramètres utilisés au niveau des TTU qui comportent trois points cibles.

À la question des niveaux de proéminence qui semblent permettre de reproduire les variations de F0, on peut apporter des éléments de réponse en s'appuyant sur les formes optimales obtenues à partir des meilleurs modèles « British school ». Il semble que l'on puisse distinguer au moins trois niveaux de proéminence significatifs. Tout d'abord, un niveau de proéminence faible représenté par les formes associées aux accents rythmiques de façon générale, puis un niveau de proéminence forte représenté par les contours associés aux accents mélodiques et aux nucléus non marqués et enfin un niveau de proéminence très fort porté par les nucléus emphatiques. Comme on a pu l'observer à partir des grammaires ou bien des résultats de nos expériences de détection de fonctions, il existe une marge considérable de confusion entre les différents niveaux de proéminence, car les formes qui leur sont associées ne sont pas toujours distinctes. En ce qui concerne l'opposition entre les accents « A » et « S », nos résultats varient en fonction de la

position de l'unité tonale en question et du type d'énoncé dans lequel on se trouve. Ce sont les énoncés marqués par une emphase au niveau prénucléaire qui permettent de faire ressortir la distinction entre les accents rythmiques et mélodiques le plus clairement. On a alors une tendance générale qui oppose des contours montants dans le cas des accents rythmiques et des contours descendants dans le cas des accents mélodiques. On observe la même opposition entre les accents mélodiques et les accents rythmiques dans les questions non terminales et en position initiale.

Au niveau de la position de l'unité tonale, c'est la position médiane qui parait la plus complexe. Les contours mélodiques associés aux unités tonales en position médiane quelle que soit leur étiquette fonctionnelle « A » ou « S », sont simplement représentés par un abaissement léger de F0 du type {SD}. Dans ces cas-là, on peut supposer que d'autres facteurs prosodiques interviennent et rendent un accent « A » plus proéminent au niveau perceptif qu'un accent noté « S » même si le contour intonatif reste le même.

L'opposition entre les nucléus emphatiques et non emphatiques se traduit dans nos modèles de façon attendue par une différence de hauteur au niveau de l'attaque du contour : pour les nucléus en position médiane, on a {SL} *vs* {UL}, en position finale {HL} *vs* {TL}. Lorsque l'on se trouve dans un IU non terminal, on observe essentiellement une opposition entre un contour montant simple de type {BU} et un contour descendant-montant de type {HBU}.

L'observation de quelques exemples d'échec de modélisation laisse entrevoir la possibilité de développer et approfondir notre annotation fonctionnelle en exploitant la variabilité au niveau des formes associées aux accents mélodiques par exemple ou à l'étiquette fonctionnelle emphatique.

On pourrait notamment annoter un accent mélodique de type « démarcatif » qui serait associé à une forme de type {BU} plutôt qu'à la forme canonique de la petite chute {UL} extraite de l'analyse par synthèse. On aurait alors au moins deux étiquettes pour un accent mélodique : « A1 » = {UL} et « A2 » = {BU}.

Au niveau de la fonction emphatique, on pourrait ajouter de nombreux degrés ou niveaux d'emphase, ce qui nous conduirait à définir des contours associés à une emphase de type expressif par exemple. La modélisation de la pré-signalisation de l'emphase à partir de l'étiquette « [!| », donne essentiellement un contour en suite de chutes, suivi d'une grande chute. Cependant on a bien vu en observant certains échecs de modélisation que l'on pouvait aussi trouver un contour généralement montant suivi d'une grande chute finale correspondant au contour classique « upstepping head + high fall ». Il faudrait ajouter une étiquette fonctionnelle particulière pour ce cas précis qui fait partie du même ensemble que les IU emphatiques terminaux mais porte une particularité au niveau de la forme et du sens.

Au niveau d'un nucléus emphatique en position finale, on pourrait enrichir l'annotation en prenant en compte les différents types de contours que l'on peut trouver dans le corpus comme les contours vocatifs ou les contours accompagnés d'un élargissement important du registre.

On aurait alors différentes sous-catégories fonctionnelles pour une même étiquette de base :

[| !NF : 1 = HL

2 = TTS (vocatif)

3 = TB (emphatique ++)

À la question du statut du nucléus, on peut aussi apporter quelques éléments de réponse. L'étiquette « nucléus » ne paraît pas distinctive en elle-même quand on observe les formes qui lui sont associées. Le nucléus se comporte comme un accent mélodique surtout quand on le trouve en position initiale ou médiane. En s'appuyant sur l'annotation de l'intonation seulement, on ne le distingue pas d'un accent mélodique non nucléaire s'il est non emphatique. Il peut alors apparaître comme un élément redondant dans la plupart des cas, comme on l'a déjà fait remarquer auparavant, car il représente presque toujours le dernier accent mélodique d'un énoncé. Il reste cependant un élément distinctif dans les cas où il a une valeur contrastive. Dans les cas de contours de chute+montée, l'annotation du nucléus permet de distinguer deux accents mélodiques, l'un étant plus proéminent et permettant d'exprimer le contraste désiré. On observe aussi des nucléus contrastifs apparaissant dans des unités intonatives terminales et en position médiane. Le nucléus se définit alors par la distinction entre les contours qui le précèdent et les contours postnucléaires comme cela a été décrit dans la littérature. On observe dans ces cas précis un effet de compression du registre repéré dans nos modèles par des contours de type {SBB} associés aux accents postnucléaires. Ce sont les contextes de focalisation étroite par opposition aux contextes de focalisation large qui donnent tout son sens à l'étiquette nucléaire du moins du point de vue de l'observation des contours intonatifs. Il resterait tout de même à observer et à mesurer les variations de la durée ou de l'intensité dans les différents contextes pour voir si la proéminence nucléaire n'est pas manifestée par d'autres canaux dans le cas d'une focalisation large.

9.2 Codage des formes et alignement tonal

Du point de vue des représentations formelles, nous avons aussi effectué un ensemble de tests afin de répondre à quelques questions initiales que l'on va repasser en revue ici.

Au niveau du codage des tons INTSINT, nous avons testé deux variantes, l'une développant le nombre de tons utilisés appelée INTSINT2 et la seconde proposant une autre façon de calculer les tons appelée INTSINT3. Avec un codage plus descriptif tel qu'INTSINT2, on a pu améliorer la reproduction des contours de F0. Toutefois un tel codage surgénère des contours et il reste à démontrer que ces nouveaux contours sont justifiés au niveau linguistique. Il serait intéressant de voir si les différents types de montées codées par exemple permettent de distinguer différentes fonctions. C'est un travail qui reste à approfondir à ce stade.

L'ensemble des tests menés sur l'alignement des points cibles nous permettent de mettre en valeur des points d'ancrage particuliers : la durée de la syllabe accentuée et de la dernière syllabe d'une unité tonale, les frontières droites et gauches de l'unité tonale.

Le mode d'alignement qui apparaît comme le plus régulier à partir de nos expériences est un alignement en pourcentage de la durée de l'unité prosodique concernée et ici essentiellement de la syllabe accentuée.

En ce qui concerne le premier point cible d'une unité tonale, nous avons vu qu'il s'alignait de façon régulière en fonction de la durée de la syllabe accentuée. En effet, les variations de l'alignement se concentrent souvent entre 0 et 75% de cette syllabe, ce qui fait de ce point d'ancrage un point de repère intéressant quand on cherche à modéliser et réduire les variations de l'alignement. Les tests menés à partir de la voyelle accentuée ou de l'attaque de l'unité tonale comme points de repère fixe ont montré beaucoup plus de variabilité.

Le second point montre un alignement régulier en fonction de deux points de repère : la durée de la dernière syllabe d'une unité tonale d'une part et la fin de l'unité tonale. Dans les unités tonales en position initiale et médiane contenant chacune deux points cibles seulement, on a obtenu un bon résultat de modélisation avec un point cible aligné à une distance de 10 ms de la fin de l'unité tonale.

Les unités tonales en position finale ont été modélisées à l'aide de trois points cibles afin de prendre en compte les contours complexes comme les chutes + montées. Nous avons trouvé dans ces cas-là que le second point s'alignait de façon régulière en fonction de la durée de la dernière syllabe plutôt qu'à une distance fixe du point cible précédent ou de la fin de l'unité tonale. Quant au troisième point cible, il a été maintenu à nouveau à une distance fixe de 10 ms de la fin de l'unité tonale.

Nos résultats, du point de vue de l'alignement du premier point cible d'une unité tonale, confirment les résultats trouvés par Pierrehumbert et Silverman (1990) qui en se concentrant sur l'alignement du pic H*, ont montré que celui-ci s'alignait proportionnellement à la durée de la syllabe accentuée plutôt qu'à une distance fixe dans la voyelle. Nous avons aussi pu observer tout comme dans les travaux de House et Wichmann (1996), que le premier point pouvait s'aligner plus tard dans le pied et donner des valeurs supérieures à 100% ce qui souligne l'importance de l'unité tonale. En ce qui concerne l'alignement des contours complexes du type « HLH », nos résultats ne semblent pas confirmer le fait que les cibles mélodiques s'aligneraient de façon régulière les unes par rapport aux autres. En effet dans la plupart des cas, nous avons trouvé que le second point cible s'alignait proportionnellement à la durée de la dernière syllabe de l'unité tonale plutôt qu'à une distance variable mais régulière du premier point cible.

Enfin nous avons observé l'effet de la position de l'unité tonale sur la façon dont s'alignent les points cibles. Dans l'ensemble, nos résultats rejoignent à nouveau ceux de House et Wichmann (1996). Le premier point s'aligne en général plus tard dans la syllabe accentuée quand il s'agit d'un accent prénucléaire et plus tôt au contraire pour un accent postnucléaire. On observe aussi le

même effet quand on compare l'alignement du premier point dans les unités tonales en position initiale et finale.

L'observation des résultats en fonction de la configuration de F0 prise en compte permet aussi de faire des remarques supplémentaires : l'alignement du premier point dans les contours de type « HL » ou « HLH » reflète exactement les tendances générales observées précédemment (la durée de la syllabe accentuée est un bon point de repère et l'on a une distinction entre l'alignement en position initiale et finale). L'alignement du premier point dans les contours de type « LH » ou « _L » en position finale essentiellement est essentiellement concentré au niveau du début de la syllabe accentuée voir exactement au niveau de l'attaque de la voyelle. Les différences d'alignement constatées pour P1 dans un TU prénucléaire et postnucléaire peuvent donc aussi être expliquées par le type de contour apparaissant à ces niveaux. En effet, les contours apparaissant après le nucléus sont de type simplement descendant « _L » ou montants « LH » dans la plupart des cas.

9.3 Tentatives de reconnaissance automatique des fonctions prosodiques

Nous avons achevé notre travail expérimental sur le corpus MARSEC par une série de nouvelles tentatives de détection automatique des informations fonctionnelles à partir des informations formelles. En nous appuyant essentiellement sur les contours intonatifs canoniques qui composent les modèles que nous avons développés par analyse par synthèse, nous avons essayé d'identifier automatiquement les types d'énoncés ou types d'IU, les niveaux de proéminence et enfin les frontières d'unité tonale élargie et des unités intonatives.

Dans l'ensemble les résultats restent modestes et les marges de confusion entre les différentes catégories restent conséquentes. Toutefois il faut se rappeler ici que nous ne nous appuyons que sur l'information mélodique dans nos expériences. L'utilisation d'autres facteurs prosodiques comme la durée ou l'intensité ferait certainement la différence dans une telle entreprise.

Au niveau des unités intonatives, on parvient à identifier de façon assez satisfaisante les énoncés spéciaux comme les questions non terminales, les parenthèses basses ou les énoncés interrompus. Les contours qui leur sont associés sont en effet plus marqués et se distinguent plus aisément. Les unités intonatives terminales se confondent très largement avec les unités intonatives non terminales. Si les taux de détection sont légèrement plus élevés pour les énoncés emphatiques terminaux et non terminaux, ceux-ci restent source de confusion.

Au niveau des détections de proéminences, on parvient à identifier les nucléus non marqués en position initiale et « 1F » (initiale et finale). En position médiane toutefois ce type de proéminence est confondu avec un accent rythmique simple. Le nucléus emphatique est mieux identifié en position médiane mais régulièrement confondu avec un nucléus non marqué en position finale et initiale.

Une frontière intonative sur deux est correctement identifiée à partir du corpus MARSEC. On détecte en revanche avec un bon taux de réussite les frontières des unités tonales élargies.

Ces résultats permettent d'évaluer les modèles développés en mettant l'accent sur les catégories fonctionnelles à enrichir et sur les ambiguïtés de nos modèles.

9.4 Perspectives et limites

Le résultat principal de ce travail de thèse est l'élaboration d'une méthode et d'un outil pour analyser et étudier la relation entre l'annotation des formes et fonctions prosodiques. La mise en place de cette méthode d'analyse par synthèse des contours intonatifs de l'anglais comprend des apprentissages tels que l'utilisation de différentes métriques pour évaluer la qualité de nos modèles, un système d'optimisation des représentations formelles, un enrichissement des étiquettes fonctionnelles de base et des tentatives de détection automatique des fonctions prosodiques. On a aussi pu proposer un modèle de l'articulation forme/fonction que l'on peut évaluer subjectivement en écoutant les résultats de la resynthèse. Ceux-ci sont disponibles dans le CD qui accompagne ce mémoire en annexe.

Ce travail reste cependant seulement un début et ouvre la voie à beaucoup d'autres démarches expérimentales pour aller plus loin.

En continuant le travail d'observation des écarts entre les prédictions du modèle de synthèse et les enregistrements d'origine d'abord, on pourrait certainement continuer à enrichir l'annotation fonctionnelle comme on l'a vu précédemment.

Du point de vue de l'annotation fonctionnelle de départ et des cas ambigus de conversion entre les TSM et IF, il serait très productif de faire intervenir plusieurs annotateurs experts et de comparer leurs productions.

On pourrait aussi faire un travail similaire sur un corpus de style différent afin d'extraire de nouveaux contours canoniques et un nouveau modèle de synthèse. On pourrait alors imaginer un système de synthèse qui permettrait d'interpréter un même texte avec plusieurs styles. Ce type de système pourrait être utilisé de façon ludique dans l'enseignement de l'intonation aux apprenants de l'anglais. On peut d'ailleurs déjà développer à partir de nos résultats un outil permettant d'interpréter un même énoncé de plusieurs façons en faisant varier son annotation avec IF. Pour un même énoncé, on peut faire varier la position du nucléus, la nature des accents prénucléaires, ajouter une pré-signalisation de l'emphase, etc.

Comme nous l'avons mentionné de nombreuses fois dans ce mémoire, nous pourrions appliquer cette méthode d'analyse par synthèse à d'autres facteurs prosodiques comme la durée ou l'intensité. De plus il serait très intéressant d'analyser la relation forme/fonction au niveau des variations globales de F0, en observant le registre. C'est d'ailleurs un travail déjà engagé dans le laboratoire d'Aix-en-Provence dans le cadre d'une thèse en cours (cf. Delooze 2009).

Du point de vue des résultats de resynthèse de nos modèles, il serait intéressant de mener une série de tests de perception. Tout d'abord des tests simples du type MOS (Mean Opinion Score) pourraient venir compléter l'évaluation objective donnée par les métriques. Ensuite on pourrait imaginer de réaliser différentes tâches de « shadowing » avec l'aide de locuteurs natifs. La comparaison entre les contours synthétiques et les contours reproduits par ces locuteurs permettrait certainement de mettre en valeur les éléments manquant dans nos modèles.

RÉFÉRENCES BIBLIOGRAPHIQUES

Abercrombie, D. (1964). Syllable quantity and enclitics in English. In Abercrombie, D., Fry., Mac Carthy, N. and Trim, J. (eds): *In Honor of Daniel Jones.* London: Longman, 216-222.

Ali, S. (2004). *Alignement des points cibles dans les contours mélodiques creusés en anglais britannique.* Mémoire présenté pour l'obtention du diplôme de DEA en sciences du langage sous la direction de Daniel Hirst. Université de Provence.

Ali, S. & Hirst, D. (2007). Analysis by synthesis of English intonation patterns: generalising from form to function. In *Proceedings of the International Conference of Phonetic Sciences*. Saarbrücken: University of Saarland, 1205-1208.

Ali, S. & Hirst, D. (2009). Developing an Automatic Functional Annotation System for British English intonation. In *Proceedings of the INTERSPEECH-2009.* (10: September : 6-10). Brighton, 2207-2210.

Anderson, A.H., Bader, M., Bard, E.G. & al. (1991). The HCRH Map Task Corpus. *Language and Speech 34,* vol.4, 351–366.

Armstrong, L.E. & Ward, I.C. (1931). *A Handbook of English Intonation*. Second edition. Heffer. Cambridge.

Astésano, C., Espesser, R., Hirst, D. & Llisterri, J. (1997). Stylisation automatique de la fréquence fondamentale : une évaluation multilingue. *Actes du Quatrième Congrès Français d'Acoustique*, Marseille, 441-443.

Auran C., Bouzon C. & Hirst D. (2004). The Aix-MARSEC project: an evolutionary database of spoken British English and automatic tools. In *Proceedings of Speech Prosody 2004*, March Nara. : Japan. 23-26

Bailly, G. & Benoit, C. (Eds). (1992). *Talking Machines: Theories, Models and Designs.* Institut de la Communication Parlée, Unité de Recherche Associée au CNRS n°368, INPG/Université Stendhal. Grenoble, France.

Bailly, G. (1997). No Future for Comprehensive Models of Intonation? In *Computing Prosody : Computational Models for Processing Spontaneous Speech.* Sagisaka, Y., Campbell, N. and Higuchi, N. (Eds).

Bailly, G. & Aubergé, V. (1997). Phonetic Representation for intonation. In *Progress in Speech Synthesis.* Van Santen, J., Sproat, R., Olive, J.P., Hirschberg, J. (Eds).

Bailly, G. & Holm, H. (2005). SFC a Trainable Prosodic Model. In *Speech Communication.* vol 46, 348-364.

Beckman, M.E & Ayers, E.G. (1997). *Guidelines for ToBI labelling.* Retrieved 2002-01-11, http://www.ling.ohiostate.

Beckman, M.E., Hirschberg, J. & Shattuck-Hufnagel, S. (2005). The original ToBI framework. In S.-A. Jun (Ed.) *Prosodic Typology-The phonology of Intonation and Phrasing.*

Black, W.A. & Hunt, A. (1996). Generating F0 contours from ToBI labels using linear regression. *In Proceedings of ICSLP-1996,* Philadelphia, Penn., vol.3, 1385-1388.

Black, W.A. (1997). Predicting the Intonation of Discourse Segments from examples in Dialogue Speech. In *Computing Prosody: Computational Models for Processing Spontaneous Speech.* Sagisaka, Y., Campbell, N. and Higuchi, N. (Eds).

Boersma, P. & Weenink, D. (2005). *Praat: a system for doing phonetics by computer,* (version 4.3.23). Retrieved September 24 (2005) from http// www.praat.org/.

Bolinger, D. (1945). The minimising downskip. *American Speech* 20, 40-4.

Bolinger, D. (1951). Intonation: Levels Versus Configurations. In *Words,* vol.7. 199-210. Reprinted in Bolinger, *Forms of English.*

Bolinger, D. (1957). Intonation and Grammar, *Language Learning,*, vol.8 (1957-58), 31-38.

Bolinger, D. (1958). A Theory of Pitch Accent in English, *Word*, vol.14, 109-149.

Bolinger, D. (1965). *Forms of English: Accent, Morpheme, Order.* Isamu Abe et Tetsuya Kanekiyo Eds. Cambridge, Harvard University Press, Tokyo, Hokuou.

Bolinger, D. (1972). *Intonation.* Penguin Modern Linguistics Readings.

Bolinger, D. (1986). *Intonation and its Parts*: *Melody in spoken English*. Edward Arnold (publishers).

Bolinger, D. (1989). *Intonation and its Uses: Melody in Grammar end Discourse.* Edward Arnold (publishers).

Brazil, D. (1975). *Discourse Intonation.* Discourse analysis monographs no1. English Research. Birmingham University.

Brazil, D., Coulthard, M. & Johns, C. (1980). *Discourse Intonation and Language Teaching.* London: Longman.

Brown, G., Currie, K.L. & Kenworthy, J. (1980). *Questions of intonation.* Croom Helm, Londres.

Brend, R. (1972). Male-female intonation patterns in American English. *Proceedings of the 7th International Congress of Phonetic Sciences*, Mouton, 866-869.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.

Bruce, G. (1990). Alignment and composition of Tonal Accents: Comments on Silverman and Pierrehumbert's paper. *Between the grammar and physics of Speech.* Kingston, J., Beckman, M.E. (Eds).

Calhoun, S. (2004). Overloaded ToBI and what to do about it: an Argument for Function-Based Intonation Categories. Presented at the *Linguistics Postgraduate Conference,* University of Edinburgh, May 2004 (Retrieved from http://homepages.inf.ed.ac.uk/s0199920/research.html).

Campbell, N. (1993). Automatic detection of prosodic boundaries in speech. *Speech Communication,* vol.13, 343-354.

Campbell, N. (1997). Synthesizing Spontaneous Speech. In *Computing Prosody: Computational Models for Processing Spontaneous Speech.* Sagisaka, Y., Campbell, N. and Higuchi, N. (Eds).

Campione, E., Hirst, D. & Véronis, J. (2000). Automatic stylisation and symbolic coding of F0: implementations of the INTSINT system. In Botinis A. (ed.) *Intonation: Research and Applications*. Dordrecht: Kluwer. 2000, 185-208.

Campione, E. (2001). *Etiquetage prosodique semi-automatique de corpus oraux: algorithmes et méthodologie.* Thèse de doctorat: sciences du langage. Aix-en-Provence: Université de Provence.

Carr, Philip. (1999). *English phonetics and phonology.* Coll. Modern Linguistics Series. Londres, Grande Bretagne: Macmillan Press Ltd.

Caspers, J. (2000). Experiments on the Meaning of Four Types of Single-Accent Intonation Patterns in Dutch. *Language and Speech.* vol .43 , part 2, 127- 161.

Chan, D., Fourcin, A., Gibbon, B & al. (1995). EUROM-A Spoken Language Ressource for the EU. In *Proceedings of the fourth European Conference on Speech Communication and Speech Technology.* (18-21 September). Madrid, vol.1, 867-870.

Cho, H.S. (2009). Etude des propriétés acoustiques de la structure prosodique du coréen. Thèse nouveau régime : Université Aix-Marseille 1. Sous la direction de Albert Di Cristo. Septembre 2009.

Cho, H.S. & Hirst, D. (2006). The contribution of silent pauses of prosodic boundaries in Korean read speech. In *Proceedings of Speech Prosody 2006*. Dresden: Germany.

Coleman, J. (2000). *Introducing Speech and Language Processing.* Cambridge University Press.

Collier, R. (1974). Intonation from a structural linguistic point of view: A criticism. *Linguistics,* vol.129, 5-28.

Couper Kuhlen, E. (1986). *An Introduction to English Prosody.* London, Arnold.

Couper Kuhlen, E. & Selting, M. (1996). *Prosody in conversation: Interactional Studies.* Cambridge University Press.

Cruttenden, Alan. (1962). *Gimson's pronunciation of English*. Oxford University Press.

Cruttenden, Alan. (1981). Falls and rises: meanings and universals. *Journal of Linguistics,* vol.17, 77-91.

Cruttenden, Alan. (1986). *Intonation*. Second edition. Cambridge University Press.

Cruttenden, Alan. (1995). Rises in English. Published in J.Windsor Lewis (eds). In *Studies in General and English Phonetics.* London: Routledge.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge University press.

Crystal, D. (1975). *The English tone of voice: essays in intonation, prosody and paralanguage.* Edward Arnold, London.

Crystal, D. (1980). The Analysis of Nuclear Tones. In *The Melody of Language: Intonation and Prosody.* Linda R., Waugh and C.H. van Schoonveld (Eds). University Park Press, Baltimore.

D'Alessandro, C. & Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. In *Computer Speech and Language* 9 (vol.3), 257-288.

D'Imperio, M. (2002a). Language-specific and universal constraints on tonal alignment: The nature of targets and anchors. *In Proceedings of Speech Prosody* 2002. Aix-en-Provence.

D'Imperio, M. (2002b). Tonal Alignment, scaling and slope in Italian question and statement tunes. *Travaux interdisciplinaires du laboratoire parole et langage d'Aix-en-provence.* (*TIPA*), vol.21, 25-44.

Dainora A. (2001). *An Empirically Based Probabilistic Model of intonation in American English.* University of Chicago, MIT dissertation, 101-106.

Dainora, A. (2002). Does Intonational Meaning come from Tones or Tunes? Evidence against a compositional approach. *In Proceedings of the first international conference on Speech Prosody 2002,* April. Aix-en-Provence.

Delais-Roussarie, E., Rialland, A., Doetjes, J. & Marandin, J.M. (2002). The Prosody of post-focus sequences in French. In *Proceedings of SPEECH PROSODY 2002*, Bel, B., et Marlien, I. (eds). Avril 2002, Aix en Provence.

Delais-Roussarie, E. & Fougeron, C. (2004). Les notions de Domaine et de constituant en phonétique et en Phonologie. *Journées d'Études Linguistiques* (*JEL 2004*), Mai 2004, Nantes.

Delais-Roussarie, E., Post, B. & Portes, C (2006), Annotation prosodique et typologie, *Travaux Interdisciplinaires du Laboratoire Parole et Langage,* vol. 25, 61-95.

Delooze, C. (2009). Analyse et interprétation de l'empan temporel des variations prosodiques en anglais et en français. Thèse de doctorat en cours. Université de Provence : Aix-en-provence.

Di Cristo, A. (1985). De la microprosodie à l'intonosyntaxe. Aix-en-provence, France : Publications de l'Université de Provence.

Di Cristo, A., Hirst, D. (1986). Modelling French micromelody: analysis and synthesis. *Phonetica*, vol. 43, 11-30.

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 language d'Aix-en-Provence* (*TIPA*), vol.23. 2004, 67-211.

Estruch, M. (2000). Evaluation de l'algorithme de modélisation mélodique MOMEL et du système de codage symbolique INTSINT avec un corpus de passages en catalan. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA),* vol.19, 45-62.

F0 Target alignment project. [en ligne] (www.ling.ed.ac.uk/~astrid/project.html. 1998)

ESRC Award Reference No. R000 23 7447. Alignment of fundamental frequency targets in English and Dutch.

Faure, G. (1962). *Recherche sur les critères et le rôle des éléments musicaux dans la prononciation anglaise. (essai de description phonologique).* Etudes anglaises 10. Didier. 4 et 6, rue de la Sorbonne Paris.

Faure, G. (1970). Contribution à l'étude du statut phonologique des structures prosodématiques. In *Prosodic Feature Analysis, Analyse des faits prosodiques,* Léon, P. Faure, G. Rigault, A, (eds), Didier, Ottawa, 93-108.

Fonagy, I. & Bailly, G. (1984). Clichés mélodiques. *Folia Linguistica,* vol.17, 153-185.

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In *The production of Speech,* Heidelberg: Springer-Verlag, 39-55.

Fusijaki, H. (1997). Prosody, Models, and Spontaneous Speech. In *Computing Prosody: Computational Models for Processing Spontaneous Speech.* Sagisaka, Y., Campbell, N. and Higuchi, N. (Eds).

Fujisaki, H. (2006). The role of Physiology, Physics and Mathematics in Modeling Prosodic Features of Speech. (invited paper). In *Proceedings of Speech prosody 2006,* Dresden.

Gee, J. & Grosjean, F. (1983). Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive psychology,* vol.15, 411-458.

Gimson, A.C. (2001). *Gimson's Pronunciation of English*. 6th edition. Revised by Alan Cruttenden. London: Arnold.

Ginésy, M. (1995). Mémento de phonétique anglaise. Edition Nathan.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis,* vol.2, 23-68.

Goldsmith, J, A. (1979). The aims of autosegmental phonology. In D. A. Dinnsen (ed). *Current approaches to phonological theory*. Bloomington: Indiana University Press, 202-222.

Goldsmith, J, A. (1995). *The Handbook of Phonological Theory*. Blackwell Handbooks in linguistics. Blackwell Publishers.

Grabe, E. (1998). Pitch accent realisation in English and German. *Journal of Phonetics,* vol.26, 129-144.

Grabe, E., Kochanski, G. & Coleman, J. (2003). Quantitative modelling of intonational Variation. In *Proceedings of Speech Analysis and Recognition in Technology, linguistics and Medicine 2003* [en ligne] (http://www.phon.ox.ac.uk/~esther/Grabe_Kochanski_Coleman.pdf)

Grice, M. (1995). The intonation of interrogation in Palermo Italian: implication for intonation theory (Linguistische Arbeiten Nr. 334). T.bingen: Niemeyer.

Grice, M., Leech, G., Weisser, M & Wilson, A. (2000) Representation and annotation of dialogue. In Gibbon Dafydd, Inge Mertins and Roger K. Moore (Eds) : *Handbook of Multi-modal and Spoken Dialogue Systems. Resources, Terminology and Product Evaluation.* Boston, Dordrecht, London : Kluwer Academic Publishers, 1-101.

Gussenhoven, C. (1983). A three-dimensional scaling of nine English tones. *Journal of semantics,* vol.2, 186-204.

Gussenhoven, C. (1984). A semantic analysis of the nuclear tones of English. In *On the Grammar and Semantics of Sentence Accents.* Publications in Language sciences 16. Foris Publications.

Gussenhoven, C., Bolinger, D. & Keijsper, C. (1987). *On Accent.* In the Indiana University Linguistics Club.

Gussenhoven, C. (2000). The Behavior of H* and L* Under Variations in Pitch Range in Dutch Rising Contours. *Language and Speech,* vol .43 , part 2,.183- 203

Gussenhoven, C. (2002). Intonation and Interpretation : Phonetics and Phonology. In *Proceedings of The 1st International Conference on Speech Prosody*, Aix-en-Provence, France, 47-57.

Halliday, M.A.K. (1967). *Intonation and Grammar in British English.* The Hague: Mouton.

Halliday, M.A.K. (1970). *A course in Spoken English Intonation.* Oxford: Oxford University Press.

Herman, R. & Tevis McGory, J. (2002). The conceptual Similarity of Intonational Tones and its Effects on Intertranscriber Reliability. *Language and Speech* ,vol.45, part I.

Hermes, D, J. (1998). *Auditory and Visual Similarity of Pitch Contours.* In *JSLHR,* vol. 41, 63-72.

Hermes, D, J. (1998). Measuring the Perceptual Similarity of Pitch Contours.In *JSLHR,* vol. 41, 73-82

Hirshberg, J. & Prieto, P. (1996). Training intonational phrasing rules automatically for English and Spanish text-to-speech. In *Speech Communication,* vol.18, 283-293.

Hirshberg, J. (2002). Communication and Prosody: Functional aspects of prosody. In *Speech Communication,* vol.36, 31-45.

Hirst, D. (1977). *Intonative Features: A syntactic approach to English Intonation.* (Janua Linguarum. Series minor; 139). The Hague: Mouton, 135 pages.

Hirst, D. (1979). The Transcription of English Intonation. In *Problèmes de prosodie*. (vol.1): *Approches théoriques*. Léon et M.Rossi. Institut de phonétique d'Aix-en-Provence.

Hirst, D. (1983). Structures and Categories in Prosodic representations. *Models and Measurements*. In Cutler, A and Ladd, R (eds). Berlin. Springer, 93-109.

Hirst, D. (1987). *La représentation linguistique des systèmes prosodiques. Une approche cognitive*. Thèse d'Etat. Doctorat de Lettres et Sciences Humaines: Université de Provence. Sous la direction de Mario Rossi.

Hirst, D. & Espesser, R. (1993). Automatic modelling of fundamental frequency curves using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, vol.15, 71-85.

Hirst, D., Di Cristo, A. & Espesser, R. (1994). Constituants prosodiques et macro-segmentation du signal: de l'acoustique à la phonologie. Actes, séminaire *Reconnaissance automatique de la parole*. (CRIN/INRIA) 19-11 Mars : Nancy, France.

Hirst, D. (1998). Intonation in British English. In D.J Hirst and A. DiCristo: *Intonation Systems: a survey of twenty languages.*

Hirst, D., Di Cristo, A. & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (ed), 51-87.

Hirst, D. (2001). Optimising the INTSINT coding of F0 targets for multi-lingual speech synthesis. In *Proceedings of Workshop: Prosody* (ESCA/ISCA) (2000 mai 2-5: Krakow, POLAND), non paginé.

Hirst, D. & Auran, C. (2005). Analysis by synthesis of speech prosody: the Prozed environment. In *Proceedings of Interspeech/Eurospeech. European conference on Speech Communication and Technology*. (9 : 2005, septembre 5-8: Lisbon, Portugal).29-32.

Hirst, D. (2005). Form and function in the representation of speech prosody. In *Speech Communication*, vol .46, 334-347.

Hirst, D. (2006). Phonetics: the prosody of speech and language. In K. Brown (ed.) *Encyclopaedia of language and linguistics, 2nd edition*. Oxford: Oxford university press. 2006, vol.10, 167-178.

Hirst, D. (2007). A Praat Plugin for MOMEL and INTSINT with improved algorithms for modelling and coding intonation. In *Proceedings of International Conference of Phonetic Sciences*. Saarbrücken: University of Saarland. 2007, 1233-1236.

Hirst, D., Ali, S., Chentir, A., Cho, H., Nesterenko, I. & Rauzy, S. (2007).The AANVIS project: towards the automatic multilingual analysis of non-verbal information in speech. In *Proceedings of ICPhS 2007 Satellite Meeting, Workshop on Intonational Phonology: Understudied or Fieldwork Languages*. Saarbrücken: Germany, 1-2.

Hobbs, J.R. (1990). The Pierrehumbert-Hirschberg Theory of Intonational Meaning Made Simple: Comments on Pierrehumbert et Hirschberg. In *Intentions in Communication.* Cohen, R, Jerry, M & Pollack, M.E. (Eds). Cambridge Massachusetts: MIT Press.

Holm, B. & Bailly, G. (2002). Learning the Hidden Structure of Intonation: Implementing Various Functions of Prosody. In *Proceedings of the first international conference on SPEECH PROSODY*, Aix-en-Provence, 399-402.

House, D. (1990). *Tonal Perception in Speech.* Lund: Lund University Press.

House, J. (1989). Syllable structure constraints on f0 timing. poster presentation, *LabPhon II*, Edinburgh.

House, J. & Wichmann, A. (1996). Investigating Peak timing in naturally occurring speech: from segmental constraints to discourse structure. *UCL Working Papers in Linguistics.*

Jassem, W. (1952). *Intonation of Conversational English. (Educated southern English).* Travaux de la société des sciences et des lettres de Wroclaw seria A. No 45.

Jones, D. (1909). *Intonation curves.* B. G. Teubner, Leipzig & Berlin.

Jones, D. (1972). *An Outline of English Phonetics.* Cambridge University Press, Cambridge.

Jun, S.A (eds). (2005). *Prosodic Typology: the phonology of intonation and phrasing.* Oxford University Press.

Kingdon, R. (1958). *The Groundwork of English Intonation*. London, Longman

Klinghart, H. & Klemm, G. (1920). *Übungen im Englischen Tonfall für Lehrer und Studierende.* Cöthen, Otto Schulze Verlag.

Knowles, G. (1994). Annotating large speech corpora: building on the experience of MARSEC. *Journal of Linguistics,* vol.13, 87-98.

Knowles, G., Taylor, L., & Williams, B. (1995). *A corpus of formal British English Speech.* London: Longman.

Knowles, G. (1996). From text structure to prosodic structure. In *Working with Speech* .Knowles, G., Wichmann, A. and Alderson, P. (eds).

Ladd, D.R. (1980). *The structure of Intonational Meaning: Evidence from English.* Indiana University Press. Bloomington & London.

Ladd, D.R. (1983). Levels vs. configurations, revisited. In: *Essays in Honor of Charles F. Hockett*. Agard, F.B., Kelley, G.B., Makkai, A. and Makkai, V.B. (eds). Brill, E.J: Leiden, 93-131.

Ladd, D. R., Scherer, K. R. & Silverman, K. (1984). An integrated approach to studying intonation and attitude. In: *Studies in Intonation and discourse,* Catherine John-Lewis, ed., Croom Helm, London.

Ladd, D.R. (1996). *Intonational Phonology*. Coll. Cambridge studies in linguistics 79. Cambridge university press.

Laver, J. (1994). *Principles of phonetics*, Cambridge: CUP.

Leben, W, R. (1976). The tones in English Intonation. *Linguistic analysis,* vol.2, 69-107.

Leech, G., Weisser, M, Wilson, A. & Grice, M. (1998). Le-EAGLES-WP4-4 : Integrated Resources Working Group Survey and guidelines for the representation and annotation of dialogue.

Liberman, M. & Sag, I. (1974). Prosodic form and discourse function. *Papers from the 10th Regional Meeting, Chicago Linguistic Society*, 416-427.

Liberman, M, Y. (1975). The Intonational System of English. Doctoral dissertation, Cambridge, MA: MIT.

Liberman, M. & Prince, A. S. (1977). On Stress and Linguistic Rhythm. In Goldsmith, J. A. (ed): *Phonological Theory. The Essential Readings.* USA: Blackwell publ., 392-404.

Liberman, M, Y. (1978). *The Intonation System of English.* Bloomington, Ind: Indiana university Linguistics Club.

Lieberman, P. (1967). *Intonation, Perception and Language.* Research monograph n°38. MIT Press, Cambridge, Massachusetts.

Liu, F. (2009). *Intonation systems of Mandarin and English: a functional approach*. Phd dissertation submitted to the Faculty of the division of the Humanities. University of Chicago, Illinois. 179p.

Mertens, P. (1993). Accentuation, intonation et morphosyntaxe. *Travaux de linguistique*, vol. 26, 21-69.

Mertens, P. (1997). De la chaîne linéaire à la séquence de tons. *Traitement automatique des langues* 38, vol.1, 27-51.

Mertens, P. (1999). Un algorithme pour la génération de l'intonation dans la parole de synthèse. *Actes Conférence TALN99*, Cargèse, 12-17 juillet, 233-242.

Mertens, P. (2004a). Quelques allers retour entre la prosodie et son traitement automatique. *Le français moderne* 72 , vol.1, 39-57.

Mertens, P. (2004b). The Prosogram: Semi Automatic Transcription of Prosody based on a Tonal Perception Model. In *Proceedings of Speech Prosody 2004*. Nara: Japan.

Möbius, B**.** (1995). Components of a quantitative model of German intonation. In *Proceedings of the 13th International Congress of Phonetic Sciences* (Stockholm), vol. 2, 108-115.

Nesterenko, I. (2006). *Analyse formelle et implémentation phonétique de l'intonation du parler russe spontané en vue d'une application à la synthèse vocale.* Thèse de doctorat en sciences du langage. Université de Provence: Aix-en-provence. 475p.

Nesterenko, I., Hirst, D. & Rauzy, S. (2007). On the probabilistic modelling of the form~function articulation for prosodic phenomena. *Mathématiques et sciences humaines*, n° 180, hiver 2007, spécial Mathématiques et phonologie, [En ligne], mis en ligne le 21 février 2008. URL : http://msh.revues.org/document8093.html.

Nicolas, P. & Hirst, D, J. (1995). Symbolic coding of higher-level characteristics of fundamental frequency curves. In *Proceedings of Eurospeech 95*, 989-992.

O'Connor, J.D. & Arnold, G.F. (1961). *Intonation of Colloquial English*. London: Longman (2nd edition, 1973).

Ohala, J. J. & Ewan, W. G. (1973). Speed of pitch change. *Journal of Acoustic Society of America.* 53, 345(A).

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41, 1-16.

Pablo, D.A., Wimmer, K. & Bonafonte, A. (2004). Automatic Analysis and Synthesis of Fujisaki's Intonation Model for TTS. In *Proceedings of Speech Prosody 2004*. Nara : Japan.

Palmer, D.D., Ostendorf, M. & Burger, J.D. (2000). Robust information extraction from automatically generated speech transcriptions. In *Speech Communication.* Vol.32, 79-95.

Palmer, H, E. (1924). *English intonation with systematic exercises*. Cambridge, W.Heffer & Sons LTD.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English.* Thèse de doctorat. Massachussets Institute of Technology.

Pierrehumbert, J. B & Beckman, M.E. (1988). Japanese Tone Structure. *Linguistic Inquiry Monograph 15*. Cambridge, MA: MIT Press.

Pierrehumbert, J.B & Steele, S.A. (1989). Categories of tonal alignment in English. *Phonetica 46*: 181-196.

Pierrehumbert, J. & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In *Intentions in Communication.* Cohen, R, Jerry, M & Pollack, M.E. (eds). Cambridge Massachusetts: MIT Press.

Pierrehumbert, J. (2001). Stochastic Phonology. *GLOT*, 6, 1-13.

Pijper, J.R. de. (1983). *Modelling British English intonation.* Foris Publications, Dordrecht-Holland.

Pike, K.L. (1945). *The Intonation of American English.* Ann Arbor, MI : University of Michigan Press.

Post, B., D'imperio, M. & Gussenhoven, C. (2007). Fine Phonetic Detail and Intonational Meaning. In *Proceedings of ICPhs XVI*. Saarbrucken, 6-10 August, 191-196.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel. & Fong, C. (1991). The Use of Prosody in Syntactic Disambiguation. In *the Journal of the Acoustical Society of America*. Vol.90, 2956-2970.

Prieto, P., Van Santen, J. & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal Of Phonetics*. Vol.23, 429-451.

Quirk, R., Duckworth, A. P., Rusiecki, J. P. L., Colin, A. J. T. (1964). Studies in the correspondence of prosodic to grammatical features in English. In H. G. Lunt (ed.), *Proceedings of the Ninth International Congress of Linguists*. The Hague: Mouton.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1972). *A Grammar of Contemporary English.* Longman, London.

Rietveld, T & Gussenhoven, C. (1995). Aligning pitch targets in speech synthesis: effects of syllable structure. In *Journal of Phonetics,* vol.23, 375-385.

Roach, P. (1991). *English Phonetics and phonology: a practical course*. Second edition. Cambridge university press.

Roach, P. (1994). Conversion between Prosodic Transcription Systems*:* Standard British and ToBi. *Speech Communication,* vol.15, 91-99.

Sanders, M, J. (1996). *Intonation contour choice in English.* Phd dissertation. Utrecht University.

Scherer, K, R. (1979). Nonlinguistic vocal indicators of emotion and psychopathology. In *Emotions in Personality and Psychopathology,* Caroll E. Izard, (eds). Plenum, New York, 495-529.

Schubiger, M. (1958). *English Intonation: its form and function.* Max Niemeyer Verlag/Tubigen.

Shriberg, E., Stolke, A., Hakkani-Tur & G.Tur. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *In Speech Communication.* vol.33, 125-155.

Silverman, K. (1987). *The structure and processing of fundamental frequency contours.* Phd dissertation, University of Cambridge.

Silverman, K.E.A. & Pierrehumbert, J. (1990).The Timing of Prenuclear High Accents in English. In Kingston, J. & Beckman, M.E. (eds), *Paper in laboratory phonology I*, Cambridge: CUP, 72-106.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992). ToBI: a Standard for Labelling English Prosody. In *Proceedings ICSLP92*, 2, 867- 870, Banff, Canada.

Steele, J. (1775). *An Essay Towards Establishing the Melody and Measure of Speech.* London: Bowyer and Nichols.

Stirling, L., Fletcher, J., Mushin, I. & Wales, R. (2001). Representational issues in annotation: using the Australian map task corpus to relate prosody and discourse structure. *Speech Communication*. Vol.33, 113-134.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7, 71-79.

Svartvik, J. & Quirk, R. (1980). A Corpus of English Conversation. *Lund studies in English* 56. Lund: Lund University Press.

Sweet, H. (1892). *A Primer of Phonetics.* Clarendon Press, Oxford. Second edition (1906).

Syrdal, A.K.J, Mc Gory, J. & Beckman, M. (2001). Automatic ToBI prediction and alignment to speed manual labelling of prosody. *Speech Communication*. Vol.33, 135-351.

T' Hart, J. (1984). A phonetic approach to intonation: from pitch contours to intonation patterns. In Gibbon, D. & Richter, H. (eds), *Intonation, Accent and Rhythm: Studies in discourse phonology*, Berlin, New-York: de Gruyter, 193-202.

T' Hart, J., Collier, R. & Cohen, A. (1990). *A perceptual study of intonation.* Cambridge University Press.

Taylor, P. (1995). The rise/fall/connection model of intonation. *Speech Communication*, vol.15, 169-186.

Taylor, P. (1998). The Tilt Intonation Model. In *Proceedings of the International Conference on Spoken Language Processing 98*. Sydney : Australia.

Taylor, P. (2000). Analysis and Synthesis of Intonation using the Tilt model. *Journal of Acoustical Society of America*. (107) 3, 16 94-1714.

Tench, P. (1996). *The Intonation Systems of English.* Cassell, Wellington House. 125 Strand. London WC2R OBB.

TobI 1999. http://www.ling.ohio-state.edu/~tobi/

Trager, G. L and Smith, H, L. (1951). *An Outline of English structure*. Norman. OK: Battenburg Press.

Trim, J.L.M. (1959). Major and minor tone groups in English. *Le Maître Phonétique*. 112, 26-9.

Uldall, E. (1960). Attitudinal Meanings conveyed by intonation contours. In *Language and Speech.* Vol 3, 223-234.

Vainio, M., Hirst, D.J., Suni, A. & De Looze, C. (2009). Using functional prosodic annotation for high quality multilingual, multidialectal and multistyle speech synthesis. 1*3th International Conference on Speech and Computer SPECOM 2009*, Saint-Petersburg, Russia, June 2009.

Van Santen, J. & Hirschberg, J. (1994). Segmental effects on the timing and height of pitch contours. In *Proceedings of ICSLP-1994*, 719-722.

Van Santen, J. & Möbius, B. (2000). A Quantitative Model of F0 Generation and Alignment. In Botinis, A. (ed), *Intonation, Analysis, Modelling and Technology.* Coll. Text, Speech and language Technology, 15. Dordrecht, Pay-Bas : Kluwer Akademic Publishers, 269-288.

Van Santen, J., Kain, A., Klabbers, E. & Mishra, T. (2005). Synthesis of Prosody using Multi-level Unit Sequences. In *Speech Communication*. Vol 46, 365-375.

W. Stannard Allen. (1977). *Living English Speech*. London. Longman.

Ward, G. & Hirschberg, J. (1985). Implicating uncertainty: the pragmatics of fall-rise intonation. *Language*, vol. 61, 747-76.

Wells, R, S. (1945). The pitch phonemes of English. *Lg*. vol.21, 27-39.

Wichmann, A., House, J. & Rietveld, T. (2000). Discourse constraints on F0 peak timing in English. In Botinis, Antonis(ed), *Intonation: Analysis, modelling and technology*.

Wightman, C. & Rose, R. (1999). Evaluation of an Efficient Prosody Labeling System for Spontaneous Speech. In: *Proceedings of Automatic Speech Recognition and Understanding Workshop*, Keystons, CO, USA. Vol.1, 333-336.

Wightman, C.W., Syrdal, A.K., Stemmer G.& al. (2000). Perceptually based Automatic Prosody Labelling and Prosodically Enriched Unit Selection improve Concatenative text-to-speech Synthesis. In *Proceedings of ICSLP-2000*, vol.2, 71-74.

Wightman, C. (2002). ToBI or not ToBI? In *Proceedings of the First International Conference on Speech Prosody*, Aix en Provence: April 2002, 25-29.

Willem, N., Collier, R. & T'Hart, J. (1988). A synthesis scheme for Brititish English intonation. *The Journal of the Acoustical Society of America*, vol.84, 1250-1262.

Williams, B, J. & Alderson, P, R. (1986) Synthesising British English Intonation using a Nuclear Tone Model. UKSC 154. IBM UK Scientific Centre.

Woo, J, S. (1944). *English Speech-tones*. New York: William-Frederick Press. (ré -édité en 1959 sous le titre: *The Seven English speech tones, analyzed and identified with musical tones and Chinese speech tones*. New York: William-Frederick Press.

Xu, Y. & Sun, X. (2000) How fast can we really change pitch? Maximum speed of pitch change revisited. In *Proceedings of The 6th International Conference on Spoken Language Processing*, Beijing. III-666-669.

Xu Y. & Wang, Q.E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech communication*, vol.33, 319-337.

Xu, Y. (2002). Articulatory constraints and tonal alignment. In *Proceedings of The 1st International Conference on Speech Prosody*, Aix-en-Provence, France, 91-100.

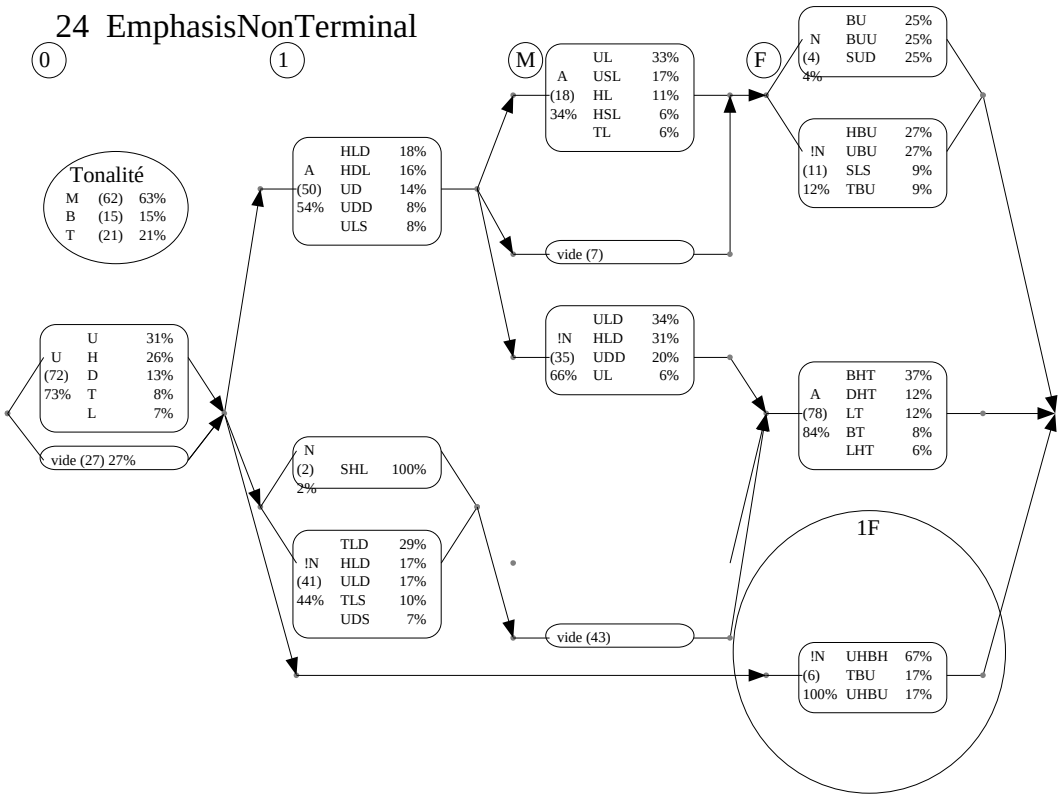
Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *In Speech communication*, vol.46, 220-251.

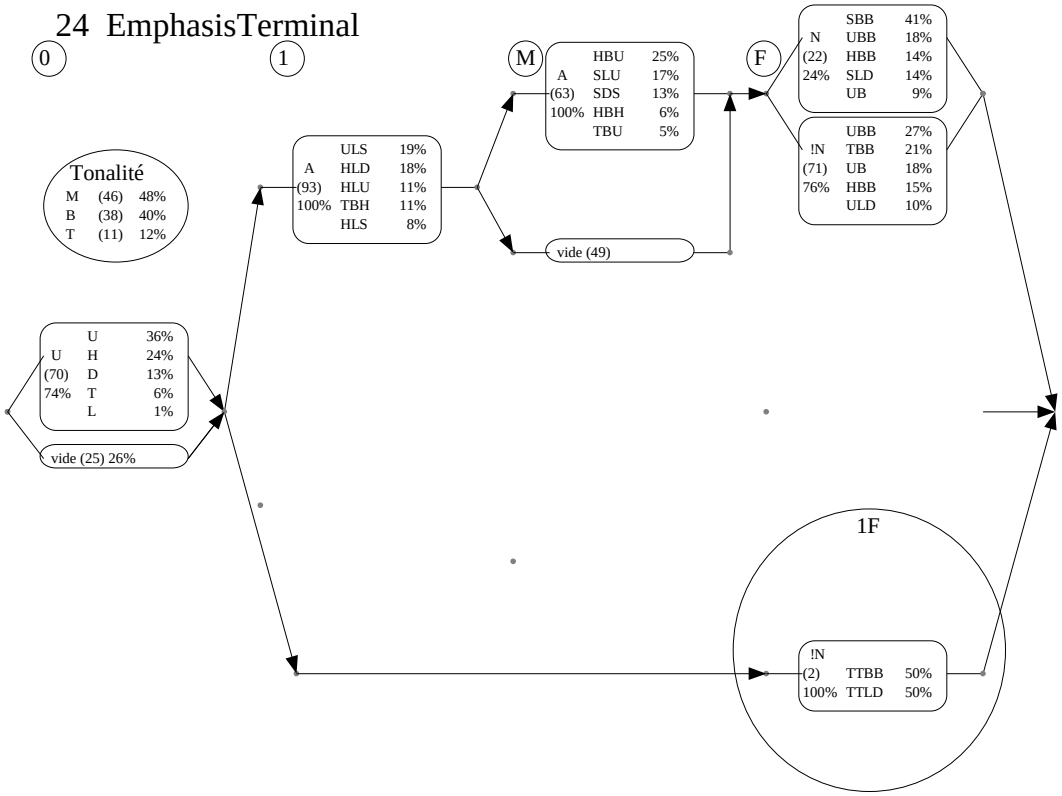
Xu, Y. & Fang, Liu. (2006).Tonal alignment, syllable structure and coarticulation: Towards an integrated model. *Italian Journal of Linguistics*, vol.18, 125-159.

ANNEXES

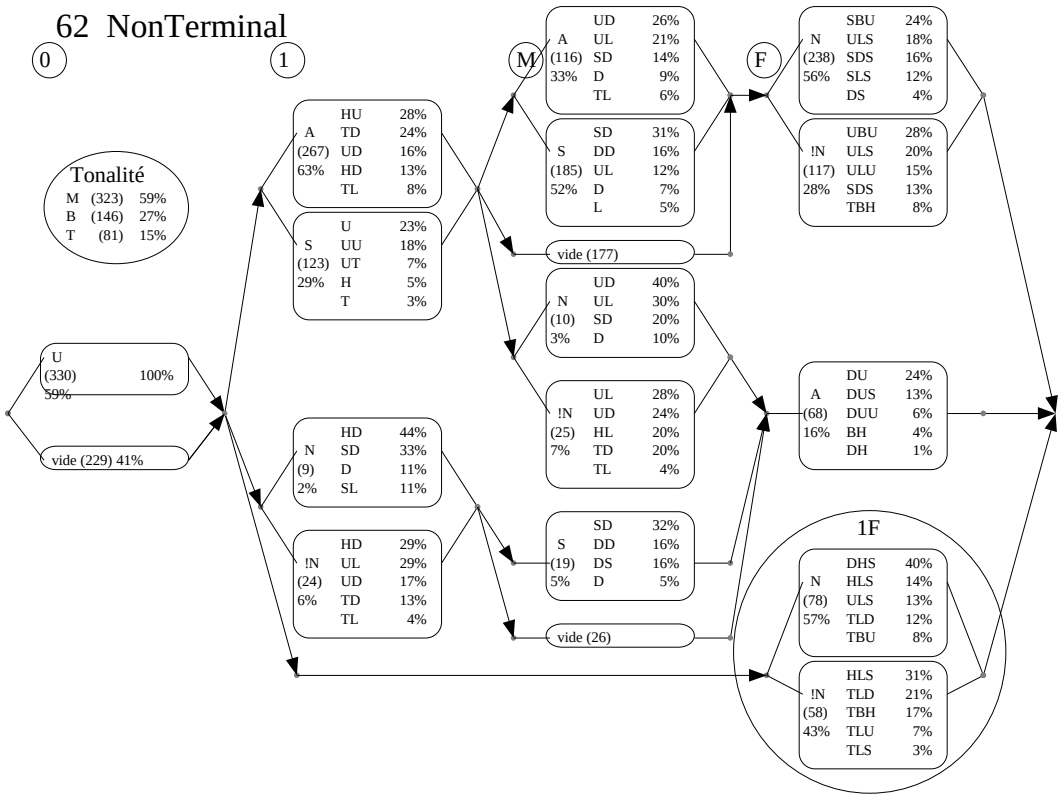
Annexe 1. ***Grammaires de l'articulation forme/fonction dans les différents modèles***

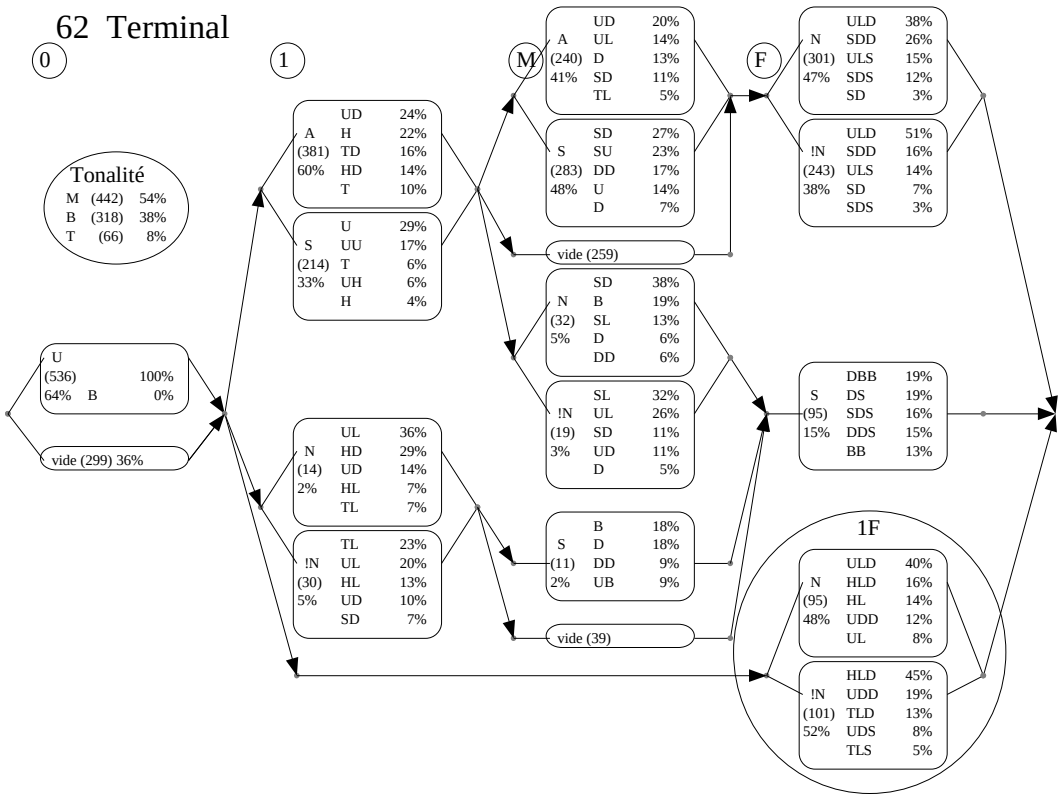
1.1. Grammaires du modèle 24 (TTU et INTSINT)

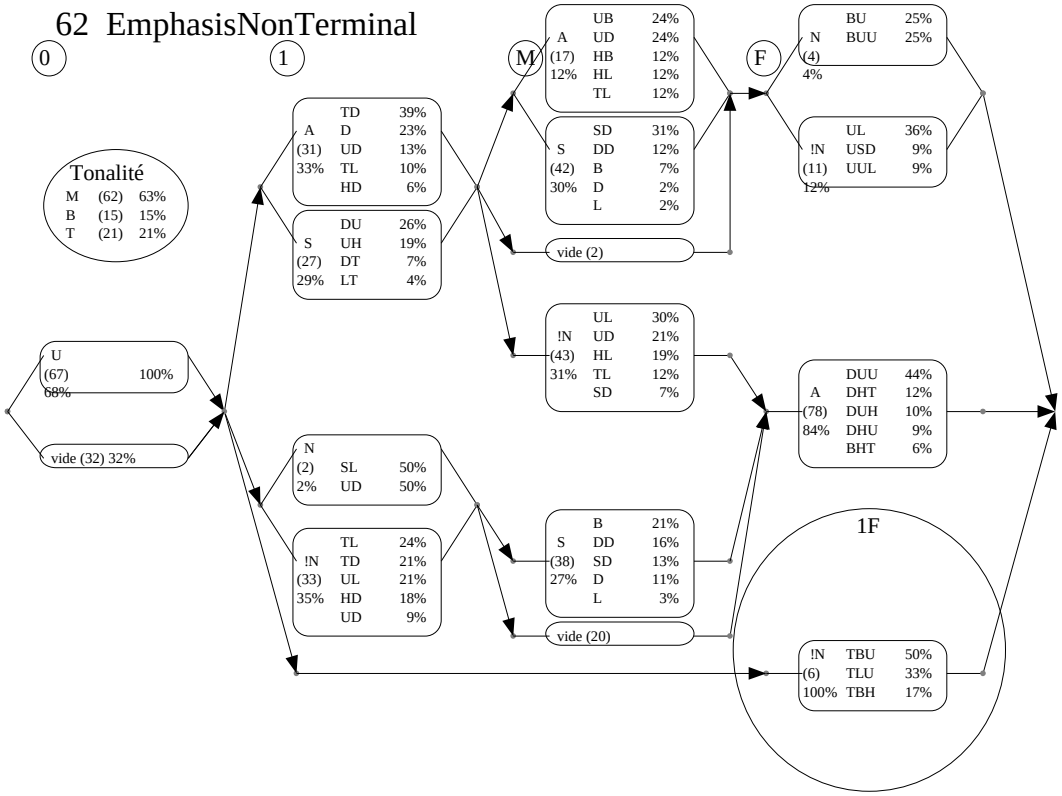


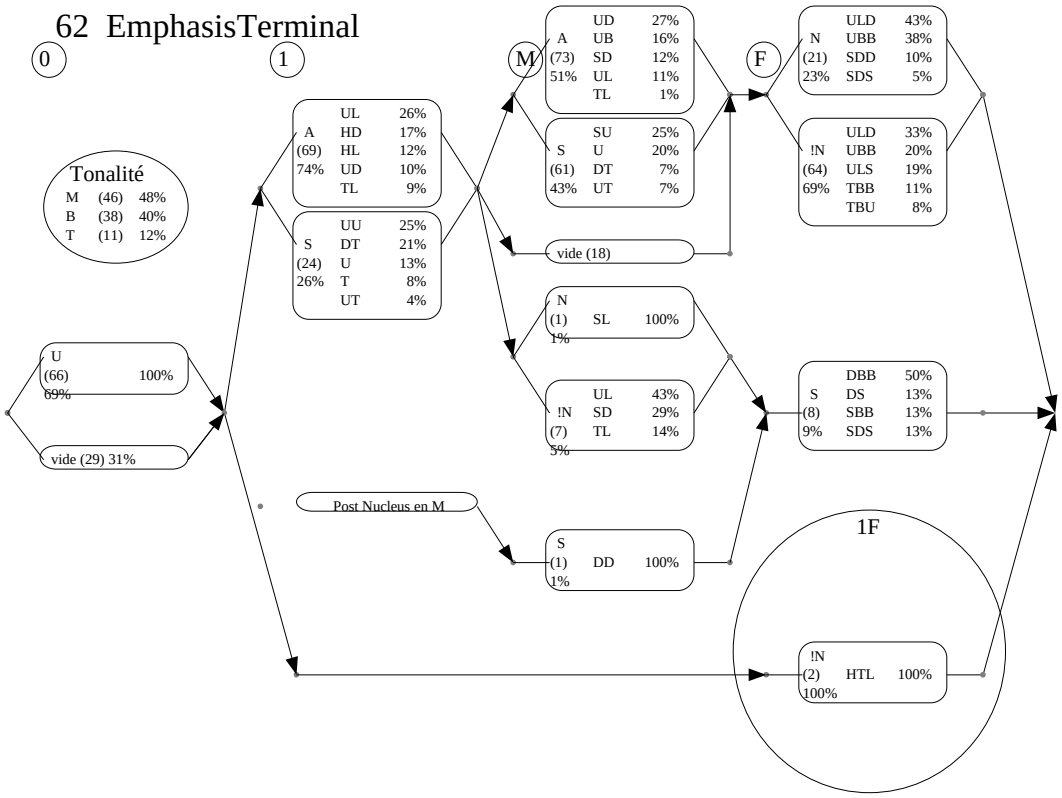


1.2. Grammaires du modèle 62 (TU et INTSINT3)

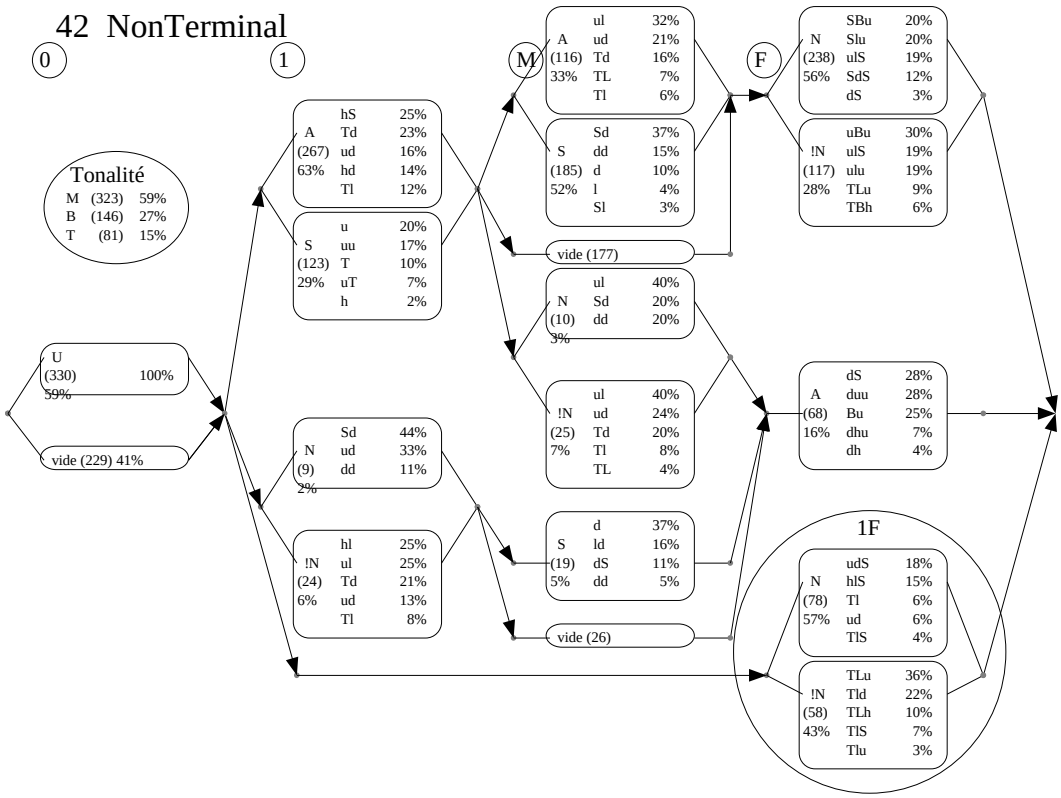


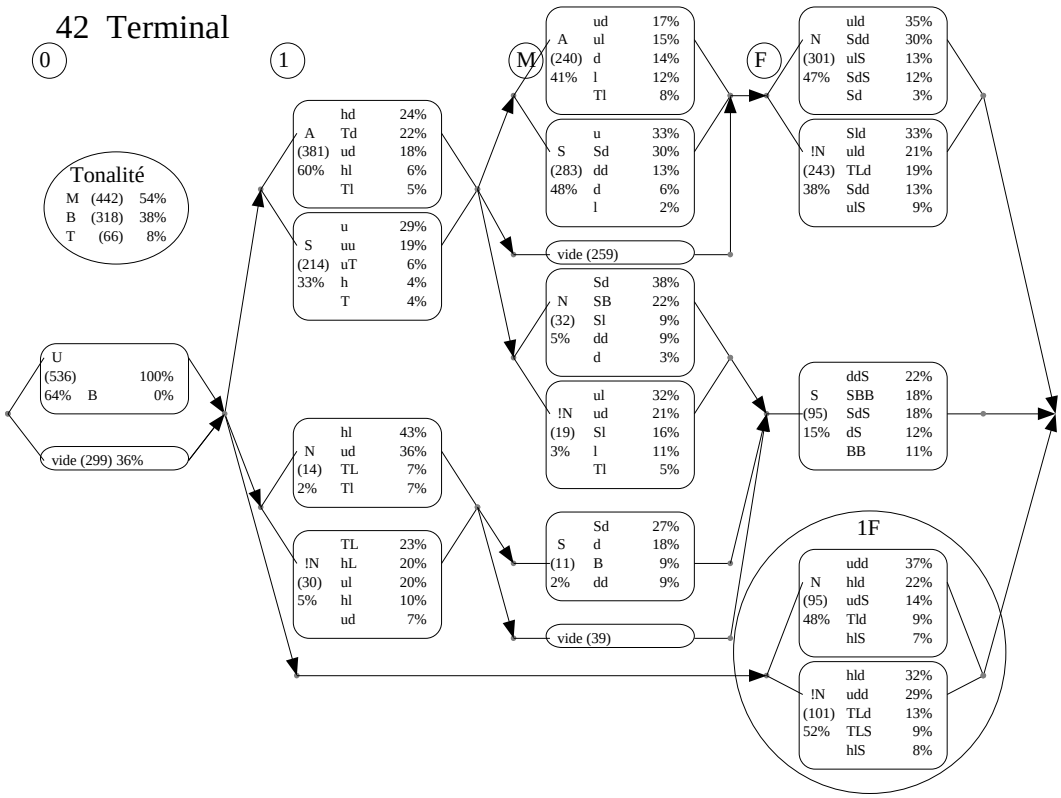


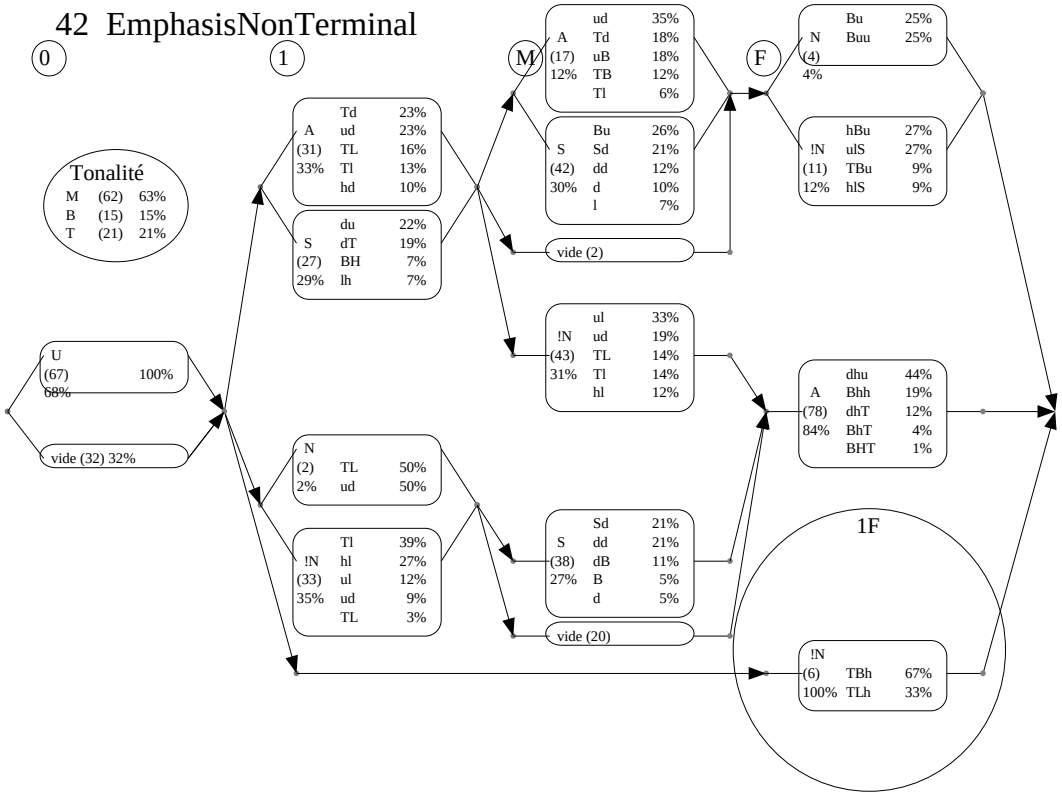

62 EmphasisNonTerminal
0
1
M
F
Tonalité
M (62) 63%
B (15) 15%
T (21) 21%
U
(67) 100%
68%
vide (32) 32%
A (31) 33%
TD 39%
D 23%
UD 13%
TL 10%
HD 6%
S (27) 29%
DU 26%
UH 19%
DT 7%
LT 4%
N (2) 2%
SL 50%
UD 50%
!N (33) 35%
TL 24%
TD 21%
UL 21%
HD 18%
UD 9%
A (17) 12%
UB 24%
UD 24%
HB 12%
HL 12%
TL 12%
S (42) 30%
SD 31%
DD 12%
B 7%
D 2%
L 2%
vide (2)
!N (43) 31%
UL 30%
UD 21%
HL 19%
TL 12%
SD 7%
S (38) 27%
B 21%
DD 16%
SD 13%
D 11%
L 3%
vide (20)
N (4) 4%
BU 25%
BUU 25%
!N (11) 12%
UL 36%
USD 9%
UUL 9%
A (78) 84%
DUU 44%
DHT 12%
DUH 10%
DHU 9%
BHT 6%
1F
!N (6) 100%
TBU 50%
TLU 33%
TBH 17%

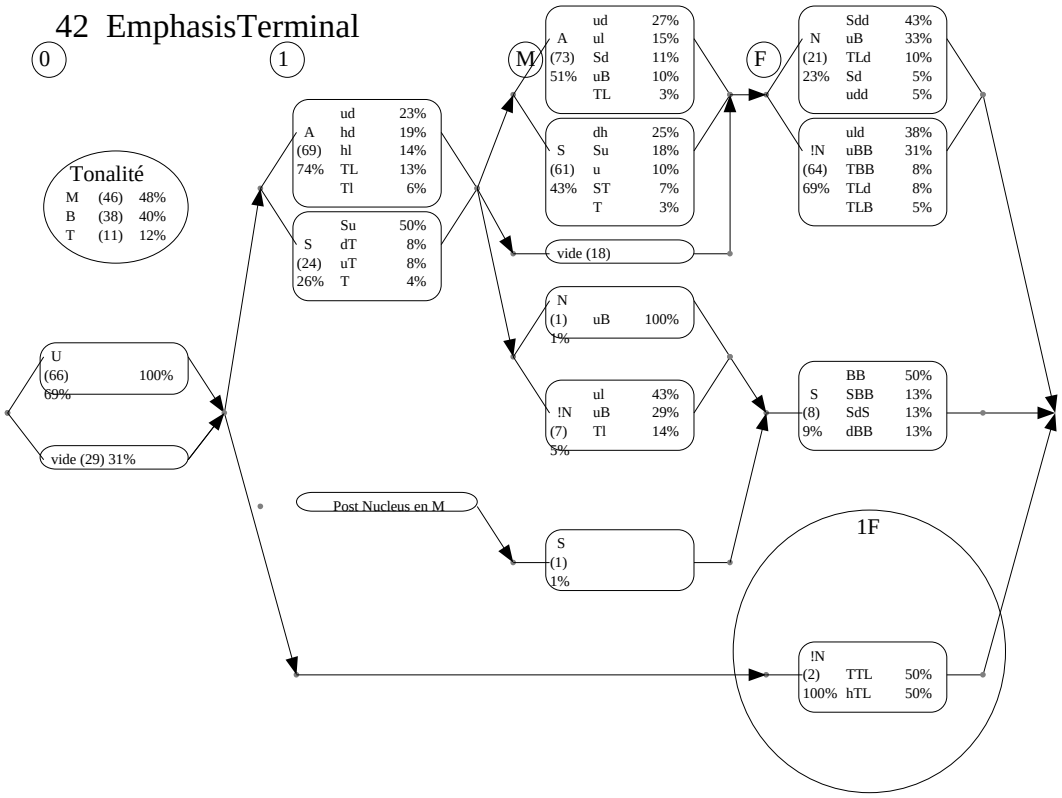

62 EmphasisTerminal
0
1
M
F
Tonalité
M (46) 48%
B (38) 40%
T (11) 12%
U
(66) 100%
69%
vide (29) 31%
A (69) 74%
UL 26%
HD 17%
HL 12%
UD 10%
TL 9%
S (24) 26%
UU 25%
DT 21%
U 13%
T 8%
UT 4%
A (73) 51%
UD 27%
UB 16%
SD 12%
UL 11%
TL 1%
S (61) 43%
SU 25%
U 20%
DT 7%
UT 7%
vide (18)
N (1) 1%
SL 100%
!N (7) 5%
UL 43%
SD 29%
TL 14%
Post Nucleus en M
S (1) 1%
DD 100%
N (21) 23%
ULD 43%
UBB 38%
SDD 10%
SDS 5%
!N (64) 69%
ULD 33%
UBB 20%
ULS 19%
TBB 11%
TBU 8%
S (8) 9%
DBB 50%
DS 13%
SBB 13%
SDS 13%
1F
!N (2) 100%
HTL 100%

1.3. Grammaires du modèle 42 (TU et INTSINT2)

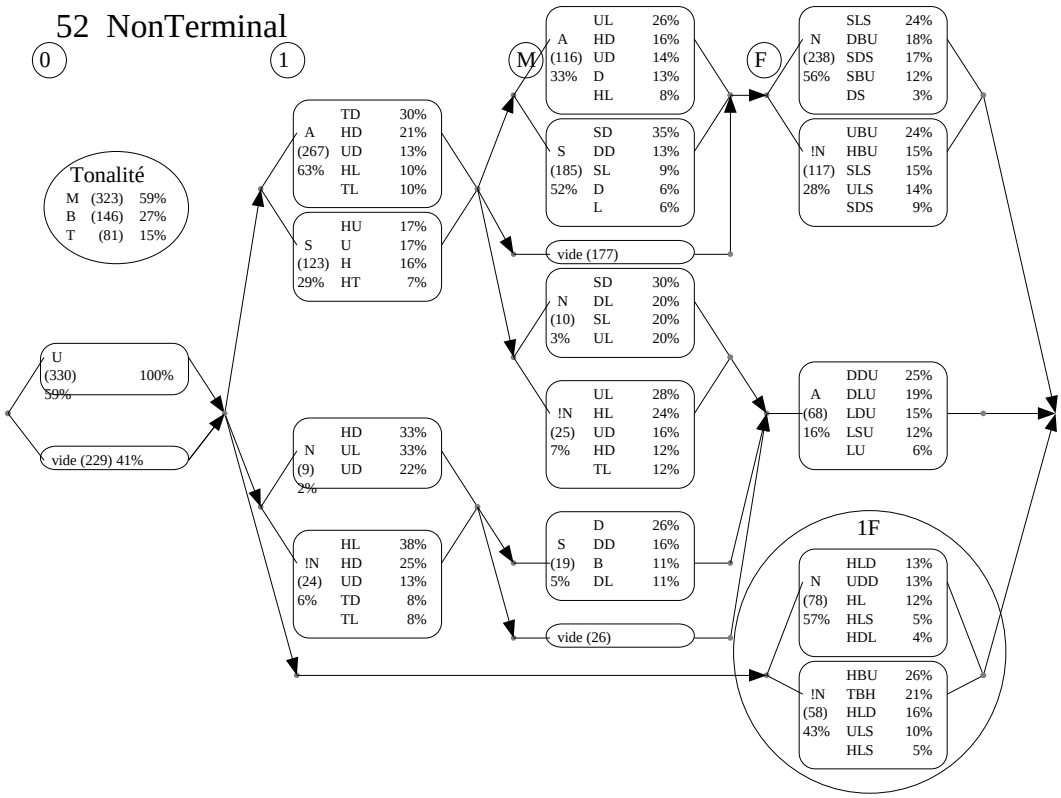


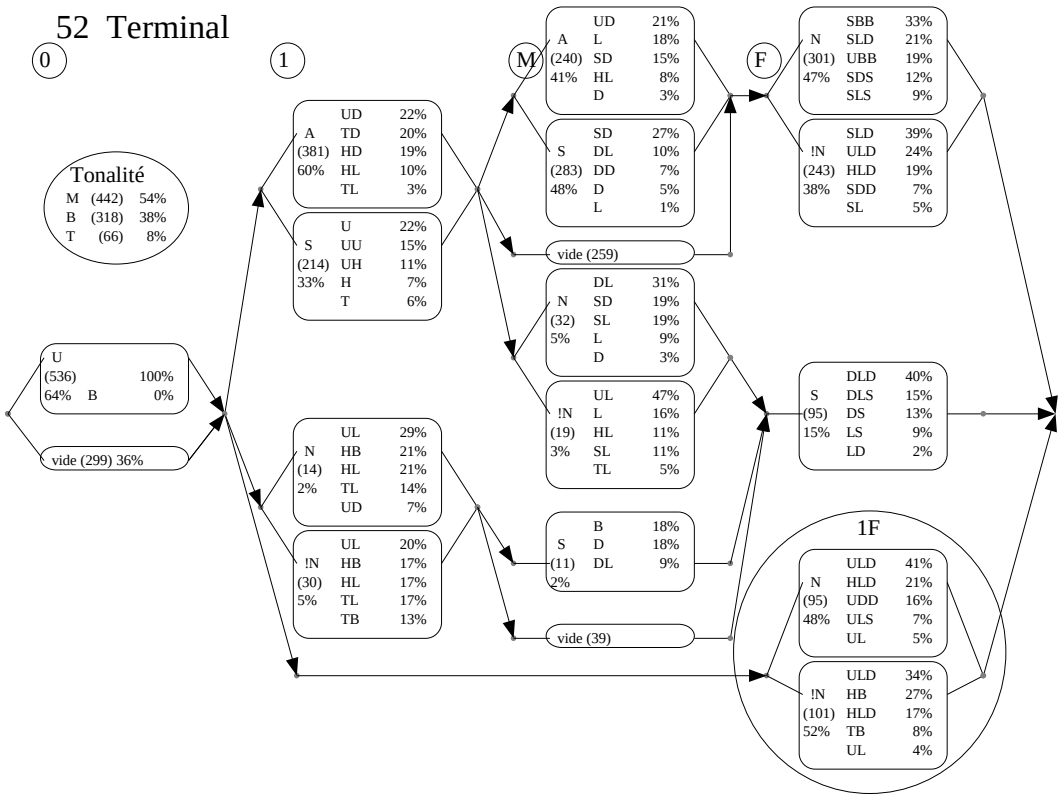


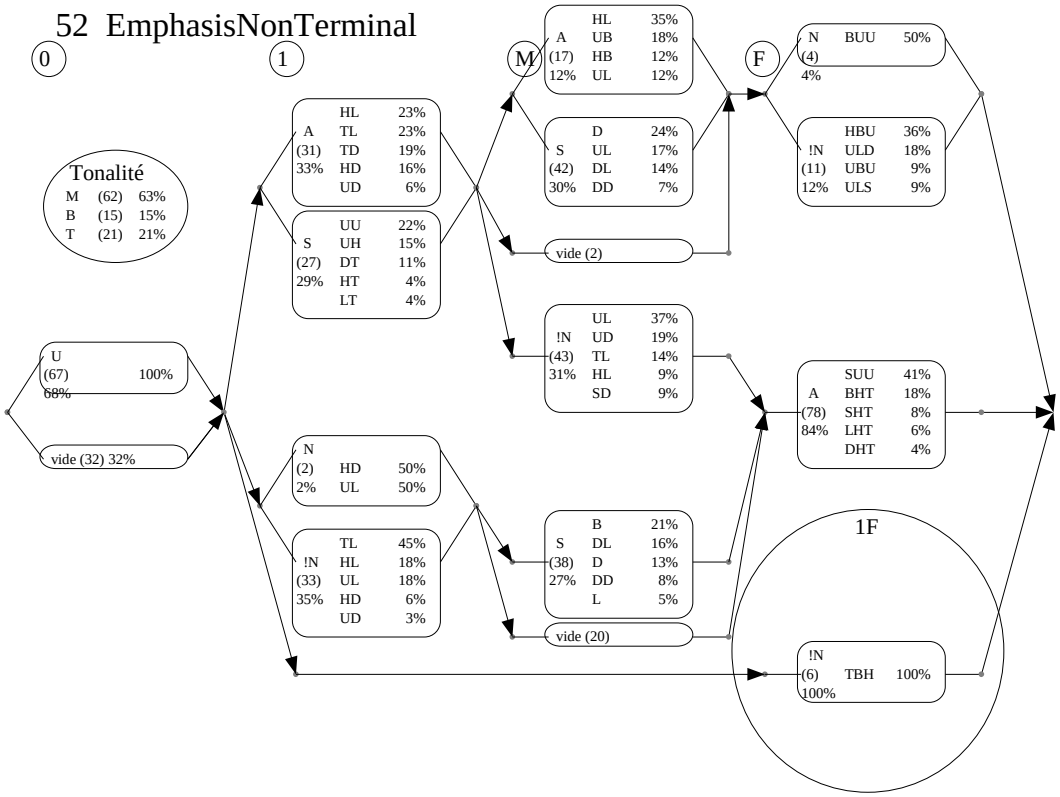

42 EmphasisNonTerminal
0
1
M
F
Tonalité
M (62) 63%
B (15) 15%
T (21) 21%
U
(67) 100%
68%
vide (32) 32%
A (31) 33% Td 23% ud 23% TL 16% Tl 13% hd 10%
S (27) 29% du 22% dT 19% BH 7% lh 7%
N (2) 2% TL 50% ud 50%
!N (33) 35% Tl 39% hl 27% ul 12% ud 9% TL 3%
A (17) 12% ud 35% Td 18% uB 18% TB 12% Tl 6%
S (42) 30% Bu 26% Sd 21% dd 12% d 10% l 7%
vide (2)
!N (43) 31% ul 33% ud 19% TL 14% Tl 14% hl 12%
S (38) 27% Sd 21% dd 21% dB 11% B 5% d 5%
vide (20)
N (4) 4% Bu 25% Buu 25%
!N (11) 12% hBu 27% ulS 27% TBu 9% hlS 9%
A (78) 84% dhu 44% Bhh 19% dhT 12% BhT 4% BHT 1%
1F
!N (6) 100% TBh 67% TLh 33%

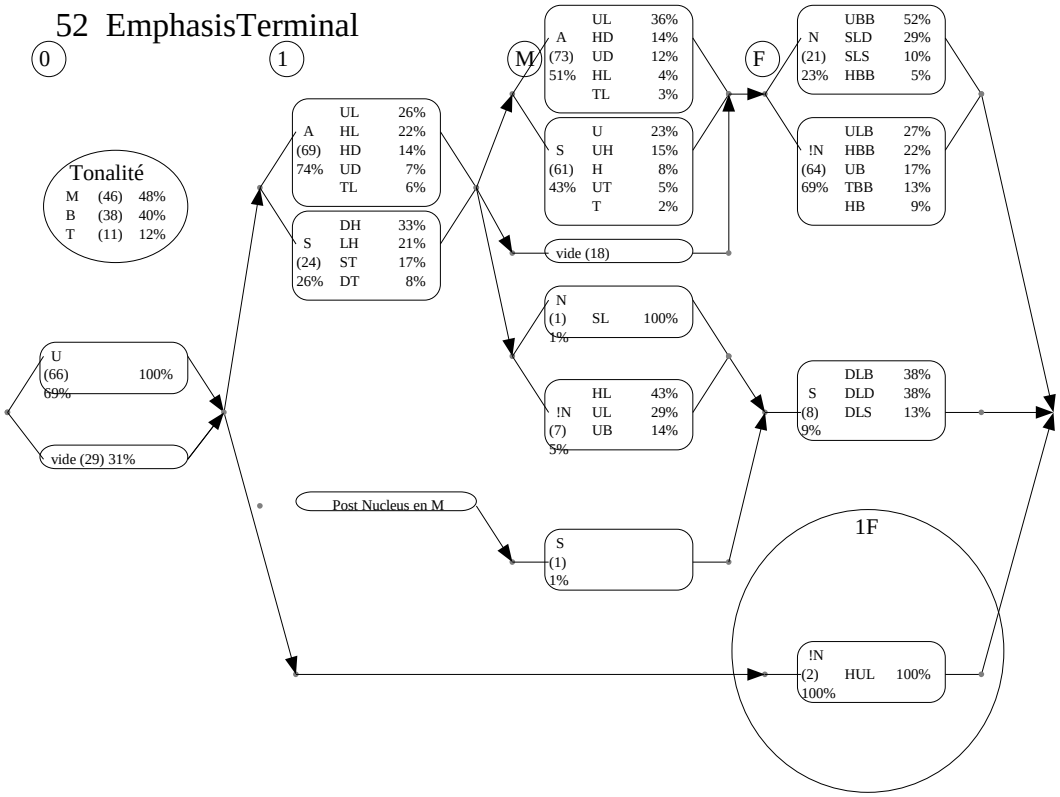

42 EmphasisTerminal
0
1
M
F
Tonalité
M (46) 48%
B (38) 40%
T (11) 12%
U
(66) 100%
69%
vide (29) 31%
A (69) 74% ud 23% hd 19% hl 14% TL 13% Tl 6%
S (24) 26% Su 50% dT 8% uT 8% T 4%
Post Nucleus en M
A (73) 51% ud 27% ul 15% Sd 11% uB 10% TL 3%
S (61) 43% dh 25% Su 18% u 10% ST 7% T 3%
vide (18)
N (1) 1% uB 100%
!N (7) 5% ul 43% uB 29% Tl 14%
S (1) 1%
N (21) 23% Sdd 43% uB 33% TLd 10% Sd 5% udd 5%
!N (64) 69% uld 38% uBB 31% TBB 8% TLd 8% TLB 5%
S (8) 9% BB 50% SBB 13% SdS 13% dBB 13%
1F
!N (2) 100% TTL 50% hTL 50%

1.4. Grammaires du modèle 52 (TU et INTSINT1)

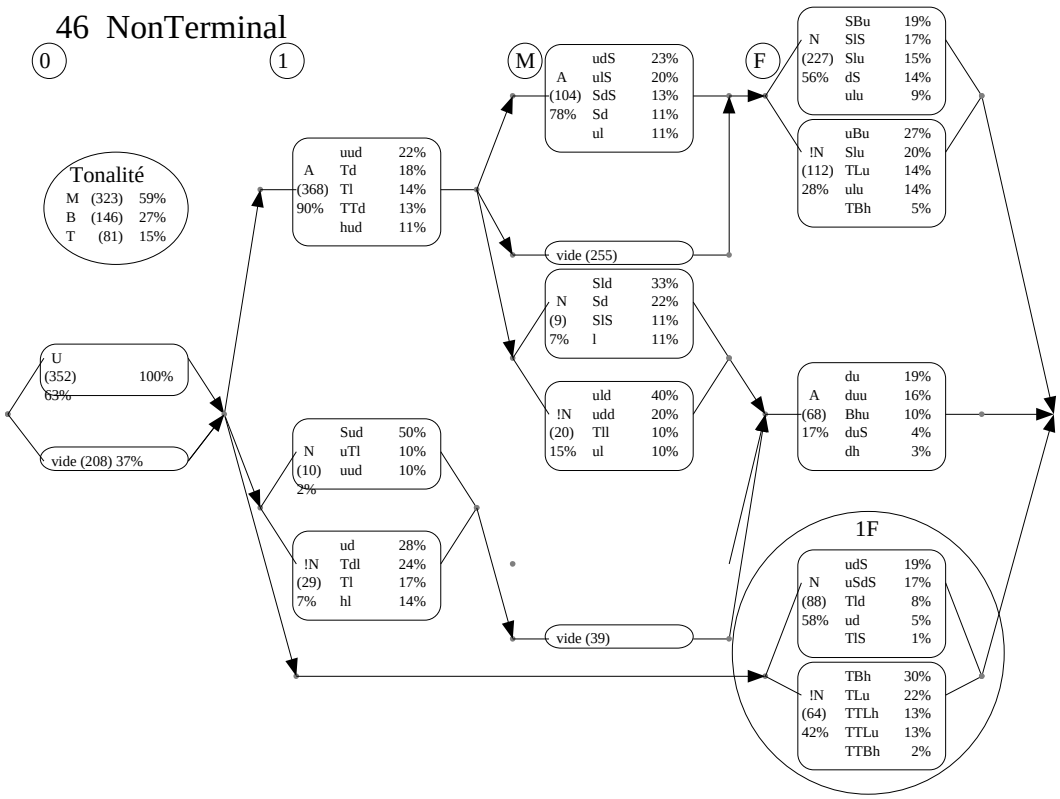


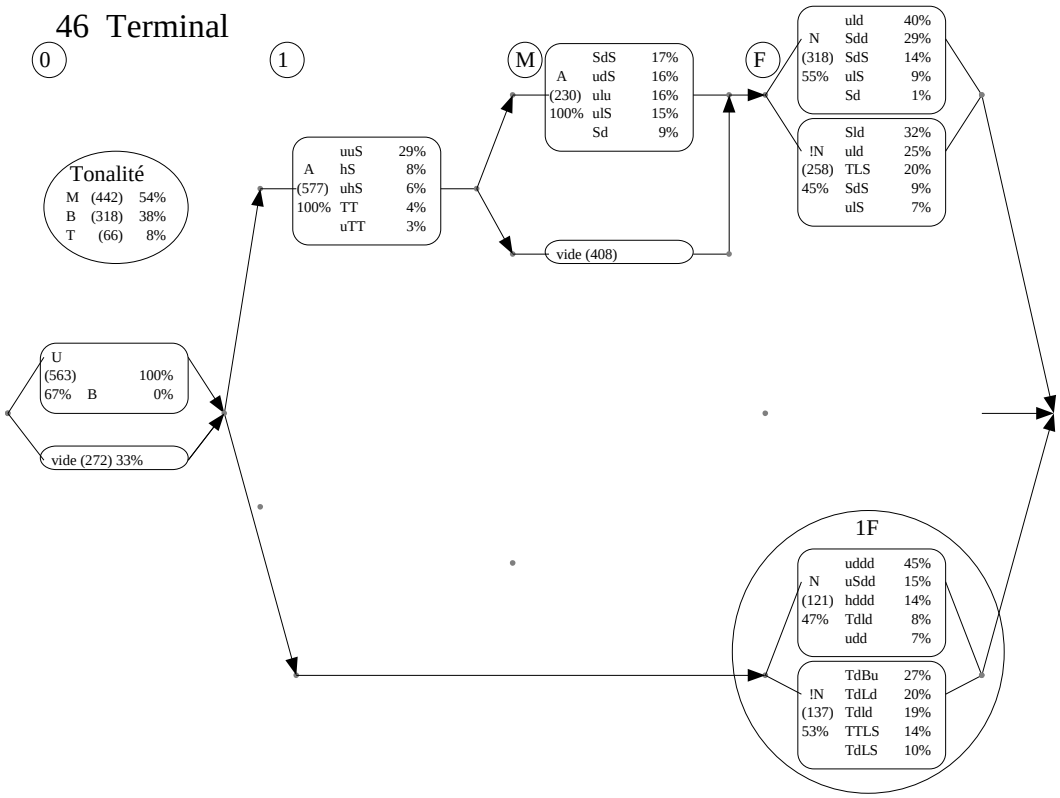


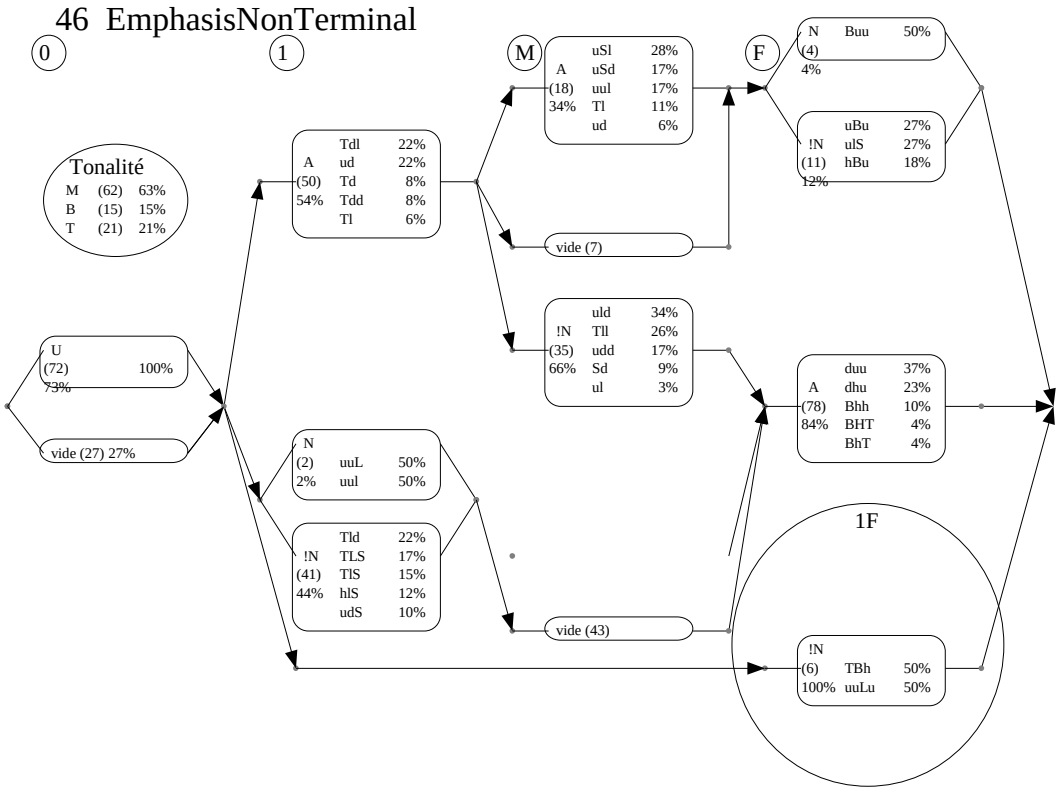

52 EmphasisNonTerminal
0
1
M
F
Tonalité
M (62) 63%
B (15) 15%
T (21) 21%
U
(67) 100%
68%
vide (32) 32%
A (31) 33%
HL 23%
TL 23%
TD 19%
HD 16%
UD 6%
S (27) 29%
UU 22%
UH 15%
DT 11%
HT 4%
LT 4%
N (2) 2%
HD 50%
UL 50%
!N (33) 35%
TL 45%
HL 18%
UL 18%
HD 6%
UD 3%
A (17) 12%
HL 35%
UB 18%
HB 12%
UL 12%
S (42) 30%
D 24%
UL 17%
DL 14%
DD 7%
vide (2)
!N (43) 31%
UL 37%
UD 19%
TL 14%
HL 9%
SD 9%
S (38) 27%
B 21%
DL 16%
D 13%
DD 8%
L 5%
vide (20)
N (4) 4%
BUU 50%
!N (11) 12%
HBU 36%
ULD 18%
UBU 9%
ULS 9%
A (78) 84%
SUU 41%
BHT 18%
SHT 8%
LHT 6%
DHT 4%
1F
!N (6) 100%
TBH 100%

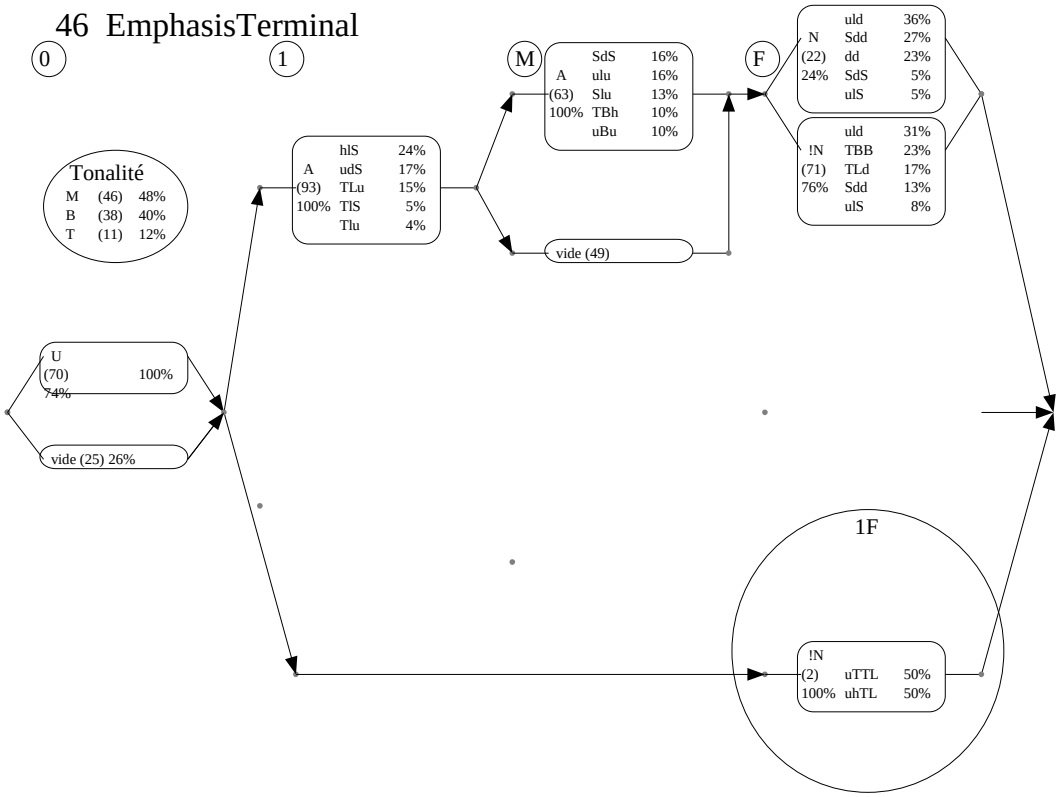

52 EmphasisTerminal
0
1
M
F
Tonalité
M (46) 48%
B (38) 40%
T (11) 12%
U
(66) 100%
69%
vide (29) 31%
A (69) 74%
UL 26%
HL 22%
HD 14%
UD 7%
TL 6%
S (24) 26%
DH 33%
LH 21%
ST 17%
DT 8%
A (73) 51%
UL 36%
HD 14%
UD 12%
HL 4%
TL 3%
S (61) 43%
U 23%
UH 15%
H 8%
UT 5%
T 2%
vide (18)
N (1) 1%
SL 100%
!N (7) 5%
HL 43%
UL 29%
UB 14%
Post Nucleus en M
S (1) 1%
N (21) 23%
UBB 52%
SLD 29%
SLS 10%
HBB 5%
!N (64) 69%
ULB 27%
HBB 22%
UB 17%
TBB 13%
HB 9%
S (8) 9%
DLB 38%
DLD 38%
DLS 13%
1F
!N (2) 100%
HUL 100%

1.5. Grammaires du modèle 46 (TTU et INTSINT2)

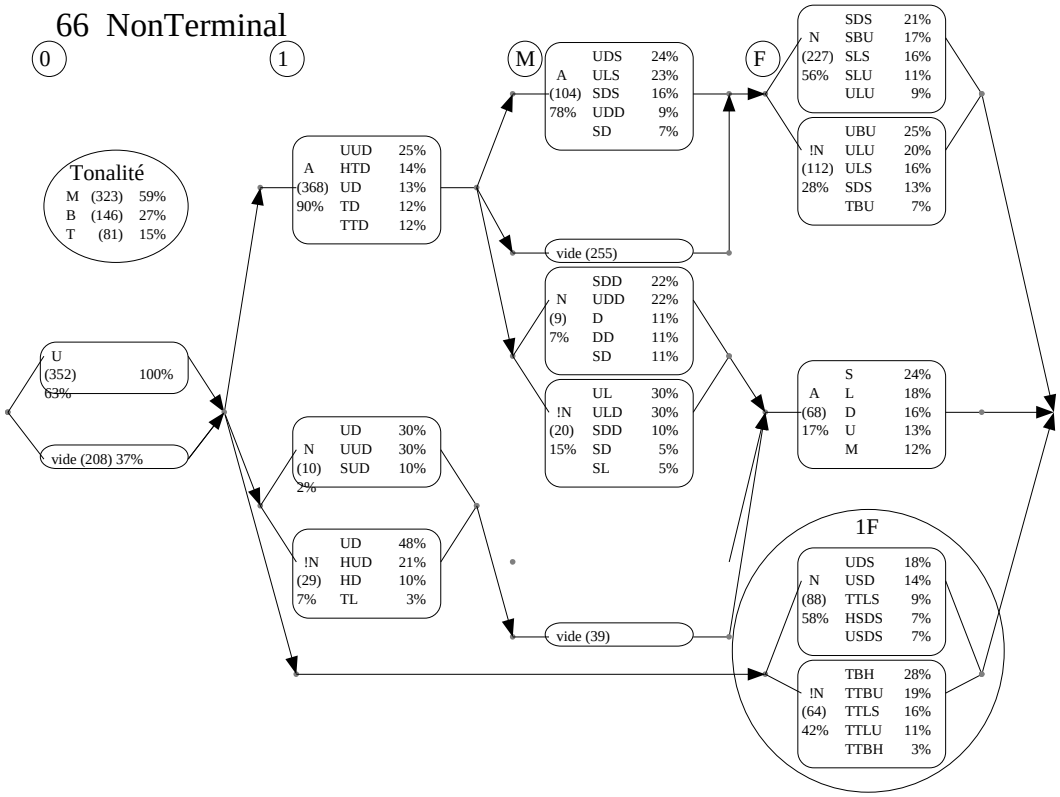


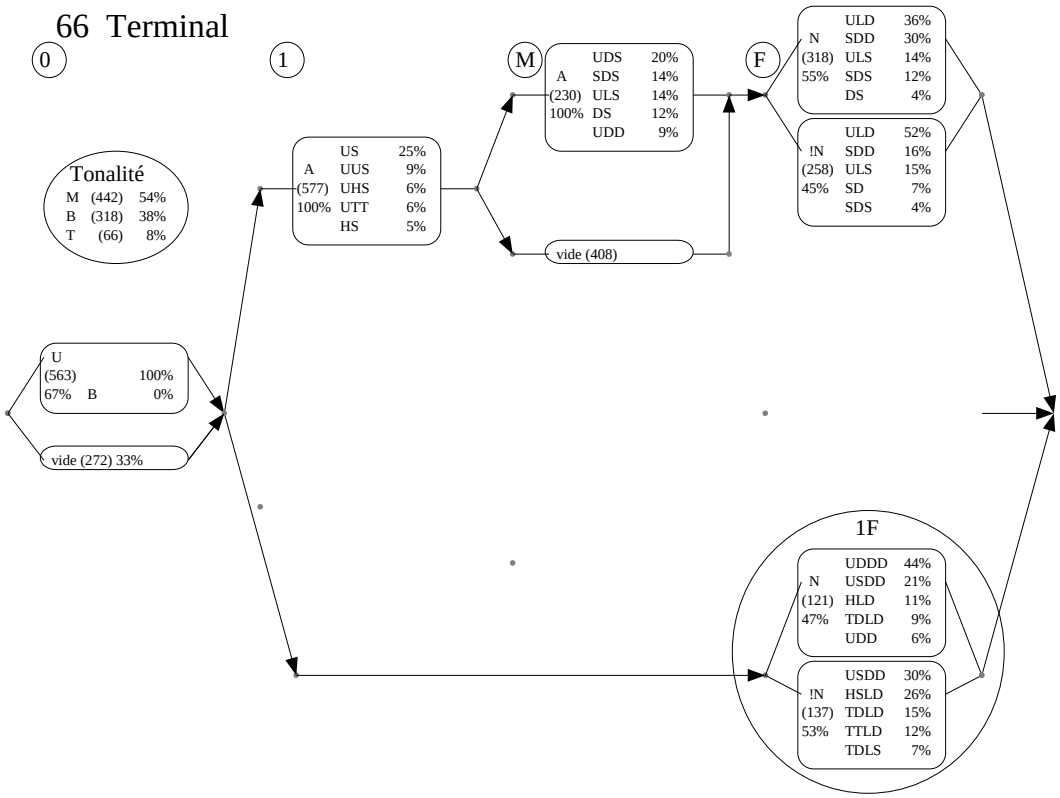


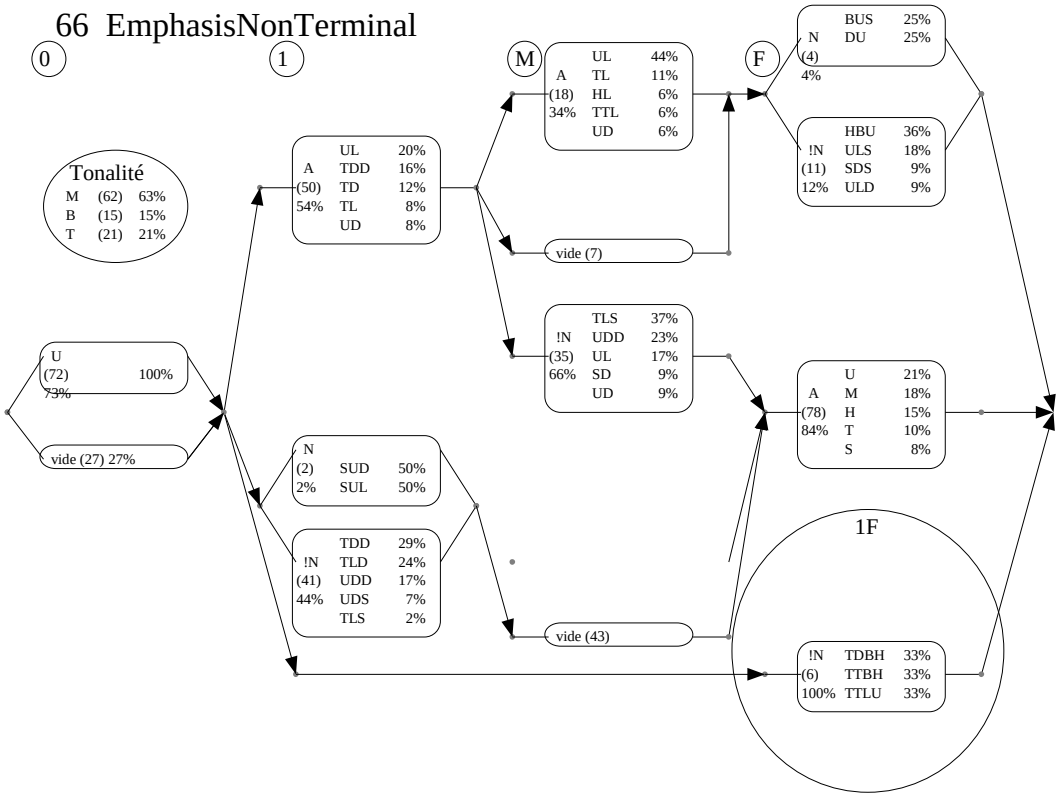

46 EmphasisNonTerminal
0
1
M
F
Tonalité
M (62) 63%
B (15) 15%
T (21) 21%
U
(72) 100%
73%
vide (27) 27%
A (50) 54%
Tdl 22%
ud 22%
Td 8%
Tdd 8%
Tl 6%
N
(2) 2%
uuL 50%
uul 50%
!N (41) 44%
Tld 22%
TLS 17%
TlS 15%
hlS 12%
udS 10%
A (18) 34%
uSl 28%
uSd 17%
uul 17%
Tl 11%
ud 6%
vide (7)
!N (35) 66%
uld 34%
Tll 26%
udd 17%
Sd 9%
ul 3%
vide (43)
N (4) 4%
Buu 50%
!N (11) 12%
uBu 27%
ulS 27%
hBu 18%
A (78) 84%
duu 37%
dhu 23%
Bhh 10%
BHT 4%
BhT 4%
1F
!N (6) 100%
TBh 50%
uuLu 50%

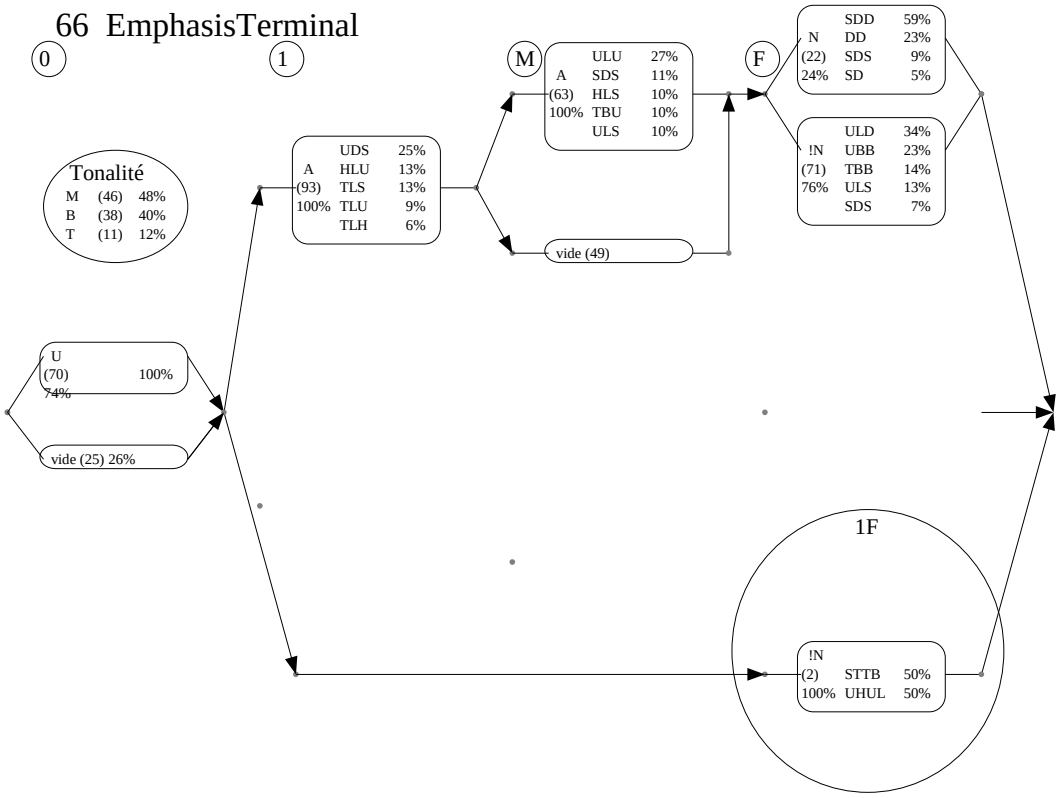

46 EmphasisTerminal
0
1
M
F
Tonalité
M (46) 48%
B (38) 40%
T (11) 12%
U
(70) 100%
74%
vide (25) 26%
A (93) 100%
hlS 24%
udS 17%
TLu 15%
TlS 5%
Tlu 4%
A (63) 100%
SdS 16%
ulu 16%
Slu 13%
TBh 10%
uBu 10%
vide (49)
N (22) 24%
uld 36%
Sdd 27%
dd 23%
SdS 5%
ulS 5%
!N (71) 76%
uld 31%
TBB 23%
TLd 17%
Sdd 13%
ulS 8%
1F
!N (2) 100%
uTTL 50%
uhTL 50%

1.6. Grammaires du modèle 66 (TTU et INTSINT3)






66 EmphasisNonTerminal
0
1
M
F
Tonalité
M (62) 63%
B (15) 15%
T (21) 21%
U
(72) 100%
73%
vide (27) 27%
A (50) 54%
UL 20%
TDD 16%
TD 12%
TL 8%
UD 8%
N
(2) 2%
SUD 50%
SUL 50%
!N (41) 44%
TDD 29%
TLD 24%
UDD 17%
UDS 7%
TLS 2%
A (18) 34%
UL 44%
TL 11%
HL 6%
TTL 6%
UD 6%
vide (7)
!N (35) 66%
TLS 37%
UDD 23%
UL 17%
SD 9%
UD 9%
vide (43)
N (4) 4%
BUS 25%
DU 25%
!N (11) 12%
HBU 36%
ULS 18%
SDS 9%
ULD 9%
A (78) 84%
U 21%
M 18%
H 15%
T 10%
S 8%
1F
!N (6) 100%
TDBH 33%
TTBH 33%
TTLU 33%


66 EmphasisTerminal
0
1
M
F
Tonalité
M (46) 48%
B (38) 40%
T (11) 12%
U
(70) 100%
74%
vide (25) 26%
A (93) 100%
UDS 25%
HLU 13%
TLS 13%
TLU 9%
TLH 6%
vide (49)
A (63) 100%
ULU 27%
SDS 11%
HLS 10%
TBU 10%
ULS 10%
N (22) 24%
SDD 59%
DD 23%
SDS 9%
SD 5%
!N (71) 76%
ULD 34%
UBB 23%
TBB 14%
ULS 13%
SDS 7%
1F
!N (2) 100%
STTB 50%
UHUL 50%

Annexe 2. ***Exemples de fichiers résultats de l'optimisation des tons***

2.1.Résultats optimisation des tons pour modèle 11, IU non terminal, TU S et A

[+_S1	score [+_S1	[+_SM	score [+_SM	[+_STM	score [+_STM	[+_STF	score [+_STF	[+_A1	score [+_A1	[+_AM	score [+_AM
_H	510.4877	_D	1148.1937	_L	120.6811	_BU	263.8044	TD	1269.2765	SD	527.8074
_T	498.3634	_L	1104.2861	_D	120.4741	_LU	260.1763	HD	1266.5718	_D	526.1483
UH	493.1585	SD	1063.6678	DL	109.0277	L_U	259.1268	TL	1080.0552	_L	511.2014
UU	489.6038	SL	1020.6415	_B	106.7799	DLU	253.6374	UD	1060.569	UL	508.3576
_U	485.9256	DL	1002.3264	DD	105.9843	LBU	248.1782	HL	1001.9612	UD	496.1057
DT	477.5049	DD	992.6934	SL	91.5103	B_H	236.604	HU	941.4633	SL	495.2898
HT	476.4669	UL	839.3013	SD	91.3108	DDU	236.1388	_T	923.7286	HL	493.9392
UT	469.7395	UD	657.8731	UL	87.1329	_LS	233.4482	TT	888.2684	HD	421.1758
LT	461.417	HL	652.0975	DB	86.2663	LDU	233.2773	UL	841.6273	TL	419.0872
DH	451.843	_B	638.4614	SB	78.6178	D_H	233.0418	HT	822.067	DL	360.4921
SH	447.9578	LD	576.4394	UB	77.4413	_DS	232.7275	_H	803.7508	DD	338.08
ST	440.8226	HD	547.0477	LB	75.3563	B_U	228.5597	HS	788.3169	UB	296.2392
LH	434.1858	DB	540.4452	HB	75.3428	_BH	228.4808	SD	696.9077	_B	283.0669
HU	430.2763	LB	514.5107	HL	65.3778	D_U	225.1898	UU	687.7085	SB	281.6821
BT	427.7883	TL	512.951	LD	64.4678	L_H	220.3747	UH	675.5352	TD	274.4581
SU	423.5078	LS	508.0168	TB	63.2598	_LD	217.8837	UT	651.9059	DB	273.5481
UD	401.8088	DS	498.7303	LS	56.5434	DLS	217.4321	_U	618.557	HB	257.3159
TL	379.9854	SB	492.1123	DS	56.2407	DBU	216.2947	US	611.01	TB	252.5701
UL	379.3391	UB	463.2685	UD	53.3467	S_H	213.3146	SL	579.005	_H	239.4417
DU	372.2155	TD	415.9538	TL	43.8258	S_U	208.567	_D	538.9335	_T	235.7811
HL	368.0764	_S	401.8461	LU	41.069	__L	208.4442	ST	521.0152	HT	232.1347
HD	367.1374	HB	393.592	DU	39.8755	__D	205.1741	DT	520.4972	_U	231.8063
BH	305.6789	SU	323.0633	BB	39.4428	_BB	202.9513	LT	505.7383	UH	221.3606
TD	303.2295	LU	319.4374	BU	39.2566	B_T	202.6802	SH	487.3064	UT	208.7236
US	293.0123	BB	305.6771	BH	30.3195	SBU	202.2879	DD	472.7843	UU	207.6442

2.2.Résultats optimisation des tons pour modèle 11, IU non terminal, TU N et !NM

[+_N1	score [+_N1	[+_!N1	score [+_!N1	[+_NM	score [+_NM	[+_!NM	score [+_!NM	[+_NF	score [+_NF	[+_!NF	score [+_!NF	[+_N1F	score [+_N1F	[+_!N1F	score [+_!N1F
HD	47.9161	UL	173.308	DL	70.8853	HL	137.813	__D	1005.63	_LS	514.868	HLD	229.713	TBU	323.443
DL	42.5845	HL	168.228	_L	70.7208	UL	132.773	_DS	1001.76	_DS	512.075	_HS	226.485	HLD	299.501
_L	41.6784	TL	164.765	SL	70.5247	UD	118.799	_LS	1000.98	HBU	509.895	ULD	222.572	U_L	296.675
_D	41.427	HD	158.667	_D	70.4221	TL	115.954	__L	985.916	SDS	507.888	ULS	216.92	ULD	296.215
DD	40.5997	UD	150.91	SD	70.3174	HD	115.689	SDS	964.876	SLS	492.74	_HU	215.978	TBH	292.916
SL	39.4944	TD	134.617	UL	69.289	SD	99.2697	_LU	950.485	HBH	490.826	TBU	211.877	HBU	290.75
SD	39.2821	HB	110.913	UD	68.0019	SL	98.5264	_LD	948.023	UBU	490.378	HBU	211.462	HLS	290.231
TL	38.5958	UB	107.496	HL	67.9543	_L	89.9581	SLS	944.672	__L	489.395	_UD	210.744	ULS	289.929
UD	38.1579	TB	106.758	DD	67.7176	_D	89.6892	_BU	921.536	TBU	487.505	SBU	209.191	H_L	285.27
TD	35.3093	SL	88.1388	TL	56.1953	TD	86.1336	SLD	904.297	_BU	485.778	UUU	208.184	U_B	283.031
DB	31.5227	SD	88.0962	DB	53.5624	HB	85.0299	DLS	901.97	ULU	484.893	DTT	208.139	TLU	281.626
LD	30.6368	_D	76.4986	_B	53.3219	UB	82.6187	DLD	895.126	SBU	483.722	HUS	206.962	U_D	280.485
UL	30.6104	_L	76.4193	SB	52.2801	DL	77.9513	D_L	892.716	TBH	478.031	UUS	206.888	H_B	277.48
SB	30.5601	DL	75.3231	LB	52.2587	SB	75.772	SLU	867.057	__D	477.433	LHS	206.595	TLD	273.962
_B	30.3743	SB	69.0037	HD	46.0028	TB	72.8219	SBU	849.731	ULS	458.879	_TT	205.838	T_B	271.945
HL	29.456	_B	67.481	UB	40.2581	_B	69.4156	_DD	842.079	S_D	456.225	DHU	205.395	TLS	268.992
LB	29.3448	DB	66.6873	HB	39.1375	DB	67.1741	S_D	839.776	HLU	455.883	LHU	204.902	TBB	265.264
UB	21.7106	DD	63.581	TB	38.6389	DD	66.2005	D_D	828.502	_LU	449.897	_US	204.682	HLU	264.63
DS	20.7523	LB	56.6935	LD	30.6911	LB	50.0428	ULS	820.992	_LD	448.412	U_L	204.623	S_D	262.116
HB	20.6205	US	46.9247	LS	28.6461	LD	38.0798	S_L	816.589	UBH	446.793	HLS	203.385	HBB	261.567
LS	19.8017	HU	36.7882	TD	27.755	US	38.0317	ULD	815.935	DBU	443.964	TLD	202.68	UBB	261.137
TB	18.3395	TT	36.7571	_S	26.7811	HT	37.2416	DBU	806.733	S_L	442.399	_TD	201.178	T_L	256.521
BB	9.758	HS	29.2947	DS	19.8014	HS	37.1935	__B	804.406	SLD	441.039	UHS	200.018	UBU	255.542
DT	9.3997	_T	28.8916	BB	17.4646	TT	36.5259	U_L	792.932	DLS	430.247	SHS	199.282	S_L	255.158
LT	9.2101	HT	28.0601	_U	10.6928	BB	29.5682	_BB	792.676	SLU	429.187	LTT	199.274	TLH	249.984
DH	8.3266	LD	27.8797	_H	10.5699	LS	28.2053	SDD	792.659	ULD	420.934	H_U	198.789	SLS	240.184

2.3.Résultats optimisation des tons pour modèle 11, IU terminal, TU S et A

[|_U0	score [|_U0	[|_S1	score [|_S1	[|_SM	score [|_SM	[|_STM	score [|_STM	[|_STF	score [|_STF	[|_A1	score [|_A1	[|_AM	score [|_AM
_	5359.89	_H	1210.11	_D	1250.66	UB	49.6818	_LS	584.122	_H	1510.22	_D	887.705
NA	NA	_T	1161.69	_L	1183.74	HB	47.2142	_DS	578.522	HU	1507.92	_L	856.46
NA	NA	UU	1158.1	SD	1095.02	_B	41.1346	_BB	556.98	_T	1482.44	UL	790.182
NA	NA	UH	1155.78	DD	1082.1	_L	40.8254	_LD	549.136	HD	1382.68	SD	757.331
NA	NA	_U	1107.36	SL	1042.71	_D	40.5575	DDS	540.245	_U	1367	HL	755.176
NA	NA	HT	1087.99	DL	1032.1	LB	39.1618	DLS	534.176	UU	1358.55	DL	744.624
NA	NA	UT	1070.41	_U	898.012	UL	39.0997	LBB	533.437	TD	1333.11	SL	732.834
NA	NA	SH	1048.41	SU	897.27	DL	38.9238	__L	528.498	UD	1330.27	UD	722.28
NA	NA	DH	1033.08	SH	892.895	DD	38.8229	L_B	525.145	HT	1302.34	DD	713.528
NA	NA	HU	1032.66	_H	888.207	HL	38.6893	__B	524.911	UH	1279.73	HD	618.393
NA	NA	ST	1015.83	ST	845.044	TB	36.839	DLD	523.416	TL	1218.83	TL	583.165
NA	NA	LT	999.049	DU	842.253	BU	31.6741	D_D	515.817	HL	1213.67	_B	553.54
NA	NA	DT	993.275	DH	831.863	LU	31.5339	D_L	507.16	UT	1178.44	LU	541.334
NA	NA	BT	960.895	_T	830.845	SL	31.276	DDD	504.035	UL	1094.75	DU	505.732
NA	NA	SU	960.871	UH	829.26	SD	31.1879	__D	492.652	HS	1083.7	LS	501.217
NA	NA	LH	930.25	UT	821.759	BH	30.5306	_DD	480.594	DT	1078.12	TD	496.019
NA	NA	DU	892.184	LH	804.671	DB	30.3668	DBB	479.246	LT	1056	LD	495.687
NA	NA	BH	807.161	UL	801.522	SB	30.3283	DDL	456.935	SH	1051.29	UB	494.789
NA	NA	LU	770.255	DT	794.49	DU	30.236	LDL	455.242	ST	1017.24	_H	491.332
NA	NA	BU	591.539	UU	739.936	LH	29.2271	SLD	453.444	DH	1016	_U	488.824
NA	NA	UD	544.713	LT	722.125	UD	29.0901	DLB	452.46	US	1015.12	DS	488.346
NA	NA	HD	535.748	LU	713.981	DH	27.9612	SDS	450.25	TT	1002.02	HB	483.387
NA	NA	UL	528.513	_B	711.057	_U	20.9064	SLS	449.443	SU	978.22	DB	482.251
NA	NA	US	518.547	HT	703.598	UU	19.5622	D_B	446.919	BT	955.401	SB	482.169
NA	NA	HS	511.519	DB	685.237	HD	18.6351	_LB	445.697	LH	923.53	DH	472.04
NA	NA	HL	499.314	BT	682.831	_H	18.5519	SBB	426.775	DU	861.461	UU	462.926

2.4.Résultats optimisation des tons pour modèle 11, IU terminal, TU N et !NM

[|_N1	score [|_N1	[|_!N1	score [|_!N1	[|_NM	score [|_NM	[|_!NM	score [|_!NM	[|_NF	score [|_NF	[|_!NF	score [|_!NF	[|_N1F	score [|_N1F	[|_!N1F	score [|_!N1F
UL	109.77	UL	195.829	SL	220.415	UL	117.859	__L	2038.41	S_L	1759.97	U_L	609.446	U_B	625.073
HB	105.175	HL	194.916	SD	220.398	SL	115.611	__D	1990.06	S_D	1729.11	U_B	600.85	H_B	608.97
HL	103.37	UD	173.802	UL	217.459	HL	115.53	__B	1985.92	U_B	1726.64	H_B	570.315	T_B	599.614
UD	98.7698	HB	156.106	SB	206.832	UD	114.165	_LD	1927.39	S_B	1681.64	ULD	555.248	HBB	568.785
UB	95.9608	TL	154.291	HB	206.047	SD	114.07	D_L	1906.83	H_B	1660.29	H_L	551.946	UBB	565.539
TL	94.3866	TB	153.136	UB	205.339	UB	107.025	D_B	1890.48	SDD	1642.77	T_B	547.281	TBB	552.618
TB	86.7581	UB	151.146	DB	199.945	TL	102.936	S_L	1874.2	U_L	1630.07	HLD	520.037	ULB	537.381
HD	78.3297	HD	146.824	_D	196.842	HB	101.22	S_D	1836.93	D_B	1592.78	ULB	507.823	U_L	534.94
SL	60.0193	TD	124.908	DL	193.688	TB	97.4891	_DD	1831.01	__L	1579.86	U_D	506.388	ULD	527.224
SD	59.7282	SL	99.3242	_L	192.189	_B	87.7822	_LS	1829.6	__D	1569.6	UDD	503.159	TBU	525.511
SB	59.1434	SD	98.2334	TB	183.86	_L	87.2268	SLD	1822.33	SLD	1567.43	ULS	492.721	HLD	493.928
TD	56.2553	_D	85.5556	_B	182.589	_D	86.4853	S_B	1822.08	UBB	1548.51	UDL	475.279	ULS	488.253
HS	54.8082	_L	85.3377	DD	182.205	SB	86.115	DLD	1818.49	SLB	1537.54	UBB	473.162	HLB	482.602
_L	49.7565	DD	77.6013	LB	167.045	DL	85.2414	UBB	1793.69	ULB	1532.56	HLB	471.667	H_L	478.557
DB	49.6981	SB	75.2406	HL	157.656	DB	84.5711	_DS	1791.28	SDL	1527.39	T_L	467.333	HLS	470.64
_D	48.3197	DL	68.613	TL	130.124	HD	75.5397	SDD	1786.58	SLS	1512.11	HLS	464.382	UDD	456.276
TT	43.8844	DB	65.1232	UD	125.412	DD	72.4956	SLS	1777.72	T_B	1508.03	HDD	464.084	TLB	455.659
_B	41.3911	_T	58.7606	HD	87.0737	TD	67.8602	U_B	1775.6	D_L	1507.63	H_D	453.732	U_D	446.399
DL	40.1804	_B	57.3977	LD	68.0688	LB	66.1401	DLS	1766.93	_DD	1507.07	HBB	447.074	HBU	444.21
US	38.249	HT	49.9997	DS	46.3658	BB	19.6599	DBB	1759.19	ULD	1496.73	TLD	441.297	TLD	426.278
LB	31.5112	HU	49.2585	TD	39.1004	LD	19.2952	_BB	1758.58	__B	1495.77	HDL	440.526	T_L	410.465
DD	31.2004	_H	48.736	LS	38.5955	LS	18.2175	SDS	1743.22	SDS	1492.21	TBU	437.147	HDD	406.051
LD	30.5352	LD	47.8029	BB	37.2071	DS	17.8108	SBB	1718.37	HBB	1481.13	TLB	433.127	TLS	403.13
HU	27.3133	LB	47.4796	_S	35.6919	BH	17.1837	DDD	1707.44	ULS	1465.21	S_L	432.817	TBH	396.056
HT	19.656	UU	45.5286	HS	30.6989	LH	16.824	DLB	1706.49	HLB	1447.89	S_D	428.055	UDL	393.539
_T	19.6074	UT	42.265	US	29.6123	DH	16.6243	_LB	1664.81	DLB	1446.35	TBB	413.249	S_D	392.745

2.5.Résultats optimisation des tons pour modèle 11, IU emphatique non terminal, TU S et A

[!+_U0	score [!+_U0	[!+_S1	score [!+_S1	[!+_SM	score [!+_SM	[!+_STM	score [!+_STM	[!+_STF	score [!+_STF	[!+_A1	score [!+_A1	[!+_AM	score [!+_AM	[!+_ATF	score [!+_ATF
_	669.984	HT	108.258	_L	255.875	_D	232.197	__T	464.902	HD	177.563	HL	106.261		
NA	NA	UT	107.96	_D	245.352	_L	226.446	L_T	453.256	TL	168.462	UL	102.768		
NA	NA	UH	103.111	DL	236.884	DL	225.971	D_T	453.109	TD	162.514	HB	93.5111		
NA	NA	HD	102.664	DD	233.948	_B	210.399	__H	449.454	HL	158.913	TB	90.6807		
NA	NA	UU	102.284	SL	212.793	DD	207.424	U_T	445.843	UD	154.686	UB	86.9497		
NA	NA	_H	102.208	_B	205.646	SL	206.375	D_H	437.703	_D	136.051	UD	86.1138		
NA	NA	TL	102.075	SD	202.645	SD	202.719	DHT	430.569	_L	129.793	TL	83.7356		
NA	NA	_T	99.6774	LB	186.654	LB	201.545	S_T	429.803	DD	128.819	HD	67.7389		
NA	NA	HU	98.3714	UL	177.028	DB	199.632	L_H	428.686	UL	127.734	TD	44.5181		
NA	NA	HL	98.3539	DB	176.651	SB	195.275	B_T	426.296	SD	126.21	HS	41.0152		
NA	NA	SH	98.0281	UB	175.438	UL	145.187	LHT	425.605	DL	119.938	HU	38.9598		
NA	NA	DT	98.0196	SB	167.174	UB	145.132	DHU	416.73	LD	116.877	US	38.5801		
NA	NA	ST	97.4777	HB	131.216	HB	141.414	S_H	413.46	SL	112.375	TT	38.1512		
NA	NA	UL	96.5241	LU	116.518	TB	136.885	_HU	408.303	UB	102.757	SD	37.1327		
NA	NA	LT	95.545	UD	115.124	HL	132.974	_HT	406.964	HB	95.2327	DB	28.9962		
NA	NA	BT	94.5871	LD	115.109	UD	131.891	UHU	406.22	TB	94.5222	SB	28.9127		
NA	NA	_U	92.5935	_S	109.304	LD	111.659	UHT	405.794	SB	92.5806	SL	28.5759		
NA	NA	UD	89.6592	TB	108.122	TL	103.826	__U	401.356	_B	89.7233	DL	18.5007		
NA	NA	LH	87.4668	DU	106.312	HD	93.635	BHT	400.432	DB	81.6574	DD	18.3621		
NA	NA	DH	81.2856	HL	104.808	DH	92.4204	SHT	399.927	LB	81.5555	_H	18.2326		
NA	NA	_D	81.0821	BU	103.91	LH	89.4673	U_H	398.196	_T	45.4635	_T	18.1981		
NA	NA	SD	80.9522	BH	102.346	LT	82.5261	LHU	397.118	LS	40.3222	_B	18.0514		
NA	NA	DD	80.1507	LS	96.0657	BT	82.3828	DUH	393.124	DS	39.203	_L	17.7885		
NA	NA	SL	80.0585	DS	94.458	LS	82.0645	_TT	391.196	HU	35.3712	_D	17.6092		
NA	NA	_L	79.9659	LH	93.545	DS	81.3093	LTT	385.133	TT	29.2067	HT	17.2613		
NA	NA	SU	79.628	DH	83.8876	SH	79.6575	_UH	383.731	LU	28.3647	LB	10.9116		

2.6.Résultats optimisation des tons pour modèle 11, IU emphatique non terminal, TU N et !N

[!+_N1	score [!+_N1	[!+_!N1	score [!+_!N1	[!+_NM	score [!+_NM	[!+_!NM	score [!+_!NM	[!+_NF	score [!+_NF	[!+_!NF	score [!+_!NF	[!+_N1F	score [!+_N1F	[!+_!N1F	score [!+_!N1F
HL	19.6378	TL	234.354			UL	231.483	BUU	20.6255	H_B	46.1617			HBU	60.3906
HD	19.4843	HL	203.781			HL	214.422	B_H	20.6079	U_B	37.1704			TBH	59.6249
UD	19.367	HD	202.51			UD	192.078	BUH	20.5188	UBU	32.6682			TBU	59.5348
TL	18.2761	UL	191.307			UB	190.204	L_U	20.4043	_HB	32.668			UBU	59.1326
HB	17.6092	TD	187.407			HB	185.64	L_H	20.3818	HBU	32.6601			ULS	57.832
UL	17.4623	UD	173.625			SD	176.767	D_U	20.2751	SBU	31.6			HLS	56.9399
TD	17.0996	HB	135.834			TL	172.894	B_U	19.6356	UHB	30.6279			HLU	56.8134
DL	10.834	TB	134.151			HD	165.454	BUS	19.6173	UUB	30.6244			ULU	56.7297
_L	10.7747	SD	125.325			TB	162.825	LUU	19.398	UBB	30.5409			ULD	55.8607
_D	10.6425	SL	124.721			SL	157.034	LUS	19.3931	_BU	30.5321			SBU	55.8348
_B	10.5314	UB	123.1			TD	141.816	LUH	19.3017	SBB	30.4692			HLD	54.9853
DB	9.8303	DL	82.8122			_D	128.499	BBH	18.622	TBU	30.3457			SLS	54.6022
DD	9.7527	_D	74.077			_L	110.625	LSH	18.3839	HSB	29.6002			SDS	54.1874
LD	9.7174	SB	66.3243			SB	102.939	DSU	18.2456	UBH	29.5545			HBH	50.857
LB	9.6731	_L	65.8615			_B	76.2459	BUT	18.2447	HUB	29.5532			TLU	48.1861
SB	8.7767	DD	56.3328			DL	74.4109	B_T	18.0937	HBB	29.551			TLH	48.0048
TB	8.7767	_T	56.246			DB	67.4856	LDH	17.4139	HBH	29.5423			TLS	47.3224
UB	8.7767	DB	47.6702			DD	67.1406	LSU	17.3793	DBU	29.5281			HDS	46.6137
LS	8.6993	TT	47.467			LD	54.7638	BBT	17.3476	_LS	29.426			UDS	46.5768
SL	8.6802	HS	45.6234			HU	54.4704	_LH	17.3298	SLS	29.419			TLD	46.3941
DS	8.6607	_H	45.3967			LB	48.0827	DDH	17.2992	_BB	29.4019			DLS	45.7706
SD	8.5664	LD	36.5706			HS	47.6216	_BT	17.2659	DBB	29.392			HDU	44.5083
BB	8.2546	HU	35.7546			HT	46.8535	DSH	17.2528	_UB	29.3756			SLD	44.286
NA	NA	DT	30.3761			TT	45.9244	LDT	17.1897	TBH	29.2828			UBH	41.0177
NA	NA	LT	29.0109			US	45.9129	LST	17.1652	ULU	29.2101			SLU	37.215
NA	NA	_B	28.9469			_T	39.1863	DUU	10.9672	USL	28.5421			DLD	36.5018

2.7.Résultats optimisation des tons pour modèle 11, IU emphatique terminal, TU S et A

[!|_U0	score [!|_U0	[!|_S1	score [!|_S1	[!|_SM	score [!|_SM	[!|_STM	score [!|_STM	[!|_STF	score [!|_STF	[!|_A1	score [!|_A1	[!|_AM	score [!|_AM	[!|_ATF	score [!|_ATF
_	659.987	DT	134.52	SH	280.878	_B	8.9456	_BB	40.1704	UL	415.747	UL	288.323		
NA	NA	LT	125.329	_H	277.276	_D	8.9456	LBB	38.1887	HL	411.223	HL	275.305		
NA	NA	ST	123.064	DT	275.646	_L	8.9456	_LS	37.2174	TL	385.998	UD	257.124		
NA	NA	UH	120.827	_T	275.049	_S	8.9456	_LD	36.2815	UD	363.257	HB	250.313		
NA	NA	SU	117.813	ST	268.252	BB	8.9456	L_B	35.9666	HD	336.682	UB	235.027		
NA	NA	SH	116.665	UH	264.233	DB	8.9456	LDL	35.7355	HB	331.928	TL	233.671		
NA	NA	_T	116.501	UT	260.61	DD	8.9456	DDS	35.125	TB	323.574	HD	232.587		
NA	NA	UT	115.436	SU	259.727	DL	8.9456	LDB	34.8076	UB	319.724	TB	227.819		
NA	NA	HT	114.733	DH	259.688	DS	8.9456	_DS	28.6711	TD	266.995	_L	201.287		
NA	NA	_H	113.429	_U	254.977	LB	8.9456	_BU	28.1611	SL	238.514	TD	191.383		
NA	NA	UU	112.449	HT	245.533	LD	8.9456	LDD	27.9594	SD	235.848	SD	185.426		
NA	NA	_U	111.317	LT	240.681	LS	8.9456	LDS	27.9288	DL	195.694	SL	182.358		
NA	NA	DU	109.424	UU	239.369	SB	8.9456	DLS	27.729	SB	190.893	_D	178.925		
NA	NA	HU	107.347	BT	237.908	SD	8.9456	DDD	27.51	_D	187.149	HU	173.369		
NA	NA	BT	105.686	LH	221.863	SL	8.9456	HBU	27.0004	_L	178.972	UU	172.427		
NA	NA	DH	101.766	BH	216.253	NA	NA	__B	26.9572	DD	166.897	HS	168.108		
NA	NA	LH	99.1664	DU	206.922	NA	NA	__L	26.3797	_B	157.536	US	165.886		
NA	NA	LU	98.8764	LU	187.521	NA	NA	D_D	25.4084	DB	150.73	SB	162.477		
NA	NA	BU	96.1357	_D	182.357	NA	NA	DBB	21.8283	LB	121.895	TT	156.525		
NA	NA	BH	70.1358	SD	172.786	NA	NA	SBB	21.7532	_T	111.845	DL	154.735		
NA	NA	US	67.1563	_L	170.177	NA	NA	UBB	20.6545	TT	111.404	DB	153.576		
NA	NA	HS	57.3557	SL	160.75	NA	NA	B_B	20.6267	UU	104.49	HT	149.356		
NA	NA	TT	52.3718	_S	154.056	NA	NA	BBB	20.6267	LD	101.925	_H	147.858		
NA	NA	HD	48.3057	BU	135.088	NA	NA	HBB	20.5428	_H	97.2172	UH	137.779		
NA	NA	UD	39.2444	HU	134.982	NA	NA	DLB	19.6016	HS	95.1469	_T	137.628		
NA	NA	HL	38.2727	UL	130.199	NA	NA	DLD	19.5	HT	91.7276	_U	136.484		

2.8.Résultats optimisation des tons pour modèle 11, IU emphatique terminal, TU N et !N

[!|_N1	score [!|_N1	[!|_!N1	score [!|_!N1	[!|_NM	score [!|_NM	[!|_!NM	score [!|_!NM	[!|_NF	score [!|_NF	[!|_!NF	score [!|_!NF	[!|_N1F	score [!|_N1F	[!|_!N1F	score [!|_!N1F
				UB	10.9499	UB	48.6784	UBB	161.041	H_B	431.326			TLD	9.8065
				SL	10.8948	HB	47.9166	HBB	157.443	U_B	426.791			TLS	9.7858
				SD	10.89	TB	46.4266	U_B	150.357	T_B	404.189			STB	9.7639
				SB	10.8573	SD	37.9831	TBB	147.311	UBB	391.315			UTB	9.7603
				HB	10.8415	SL	37.9693	H_B	134.819	HBB	383.981			TBH	9.7535
				_D	9.8821	HL	37.9283	SBB	132.923	S_B	379.115			_TB	9.749
				UL	9.876	SB	36.704	S_B	130.223	S_L	365.462			DTB	9.7466
				_L	9.8743	UL	30.3647	SLD	130.009	TBB	362.343			TBU	9.7429
				DB	9.8151	TL	29.2202	SLS	129.914	HLB	361.13			HTB	9.7358
				_B	9.8129	UD	27.9267	__L	128.756	S_D	359.965			LTB	9.7121
				TB	9.5651	DB	25.7697	D_B	128.747	ULB	355.789			HLD	9.6767
				DL	8.8464	_D	19.5367	__B	127.967	U_L	345.371			HLS	9.6741
				NA	NA	_L	19.4839	_LD	127.894	SDL	335.796			UHB	9.6638
				NA	NA	DL	18.289	S_L	127.714	D_B	333.225			HLU	9.6614
				NA	NA	HD	18.279	_BB	127.454	SDD	331.203			SHB	9.6136
				NA	NA	_B	18.2498	DBB	127.442	TLB	325.54			HBU	9.6055
				NA	NA	UT	9.6969	_LS	126.803	SLD	323.592			U_L	9.5909
				NA	NA	ST	9.6827	DLD	126.721	ULD	322.803			UDD	9.5894
				NA	NA	TD	8.7752	SDS	126.208	SLB	322.214			UTL	9.5864
				NA	NA	_T	8.7041	TBU	126.2	TBU	320.99			ULS	9.5849
				NA	NA	HT	8.7038	__D	126.101	__D	318.296			ULD	9.5843
				NA	NA	UH	8.6498	S_D	125.044	__B	318.249			UDS	9.5819
				NA	NA	SH	8.6182	_DD	124.903	H_L	313.343			UHL	9.5801
				NA	NA	DD	8.5379	SLB	124.589	ULS	313.193			_TL	9.5797
				NA	NA	NA	NA	_DS	124.207	__L	312.882			HTL	9.5718

Annexe 3. ***: Description de l’application***

3.1.Le menu

```
C:\Perl\bin\perl.exe
*================================================================*
                MENU  Applications : Saisie du corpus            *
                     1=MARSEC, 2=EUROM1                          *
*================================================================*
Saisir le Corpus xxxxxx (exemple MARSEC) (X=STOP) ==> _
```

```
C:\Perl\bin\perl.exe
*================================================================*
                MENU  Applications : Saisie du corpus            *
                     1=MARSEC, 2=EUROM1                          *
*================================================================*
Saisir le Corpus xxxxxx (exemple MARSEC) (X=STOP) ==> 1
Corpus=MARSEC
*================================================================*
*  MENU  Applications :    Saisie du sous-dossier                *
*  1 = Corrige, 2 = Original, 3 = Test                           *
*  5 = NewIU ,          6 = NewIUTU ,       7 = Final_models     *
*  8 = version2      ,    9=synthese        X= Arret             *
*  ou taper directement le nom du sous-dossier                   *
*================================================================*
Saisir le numero du sous-dossier (exemple 1) (X=STOP) ==>
```

```
C:\Perl\bin\perl.exe
*................ optimisation detection frontiÞres fonctions..........*
6  = Etape 1 Optimisation sequences de tons entre 2 Pauses silencieuses (tone)
7a  = Etape 2 Optimisation de l'alignement des tons (align)(optmize_alignment)
7m  = Etape 3 Application du modele retenu     (model)(ultimate_model)
8  = Etape 4 Detection fonctions a l interieur d un IU (FD)(detect_function=typI
U+typTU)
8t = Etape 4 Detection types de TU a l interieur d un IU connu (TD)(detect_typTU
)
8i = Etape 4 Detection types d'IU avec types TU connus (ID)(detect_typIU)
9  = Etape 5 Detection frontieres d IU (BD)(niv analyse iu et tu : detect_bounda
ries)
10 = Etape 5 Detection frontieres d IU et TTU (BD)(niv analyse TTU : detect_boun
d_IUTTU)
10t= Etape 5 Detection frontieres de TU/TTU a l interieur d un IU connu (TBD)(de
tect_bound_TU)
11 = Etape 6 Synthese, simulation avec 1 seule combi de tons et alignement (synt
h) (synthese)

12  = Preparation des donnees pour analyse detection fonctions    *
13  = Grammaire (a partir resultat du 12)                        *
14  = Calcul correlation pour tous les modeles                    *
15  = Affichage des parametres d alignement et detection            *
16  = Agregation des fichiers ...optimize...best pour stats         *

*................... creation tier syllabe pour EUROM1.............*
20  = Generation du fichier EUROM1_wordpho (pour saisie syllabes)  *
21  = Generation du fichier EUROM1_wordsyll (aprÞs saisie syllabes)*
22  = et insertion de la tier syllables dans les textgrid          *
*................... divers.........................................*
24  = Generation parametres optimize_alignement ou detect...       *
25  = Praat:Recalcul intsint en serie  + maj textgrid             *
26  = Recup des paramÞtres key, range ... du calcul de intsint     *
27  = Praat:Genere IU et TU sur MARSEC a partir de TSM (textgrids de corrige==>t
est) *
29  = Analyse annotation manuelle G04 avec annotation automatique a partir de TS
M
29b = Analyse annotation corrigee avec annotation automatique a partir de TSM
30  = Recuperation annotation manuelle textgrid de Test_annotation => NewIUTU
32  = Comparaison Intsint, Intsint1, Intsint2, Intsint3          *

Taper nO correspondant, X= Arret ==> _
```

Menu pour lancement des applications

Ce menu sert à lancer des applications Perl et certains scripts Praat qui effectuent des traitements batch. Il suffit de sélectionner l'application choisie, l'application menu demande alors la saisie des données nécessaires : corpus, sous-dossier, fichiers TextGrid à traiter, etc.

3.2.Le paramétrage

L'application est totalement paramétrée : la structure des dossiers, les paramètres d'étendue et de clé pour chacun des TextGrid, les paramètres de modélisation, etc. sont renseignés dans des fichiers au format texte (« .txt »).

La structure des dossiers

Le fichier PARAMFIC_cccc.txt (où cccc le nom du corpus, exemple PARAMFIC_MARSEC.txt) contient les informations suivantes :

- Ligne 1=dossier des programmes
- Ligne 2= dossier des TextGrid
- Ligne 3= dossier des fichiers sons
- Ligne 4= dossier des paramètres
- Ligne 5= dossiers des résultats
- Ligne 6= les fichiers sons sont-ils dans des sous-dossiers (pour EUROM1)
- Ligne 7= dossier résultat du calcul de MOMEL_INTSINT

Le contenu du fichier PARAMFIC_MARSEC.txt :

C:\THESE\Scripts
C:\THESE\Experimental\TextGrids
C:\THESE\Corpus MARSEC\Wav
C:\THESE\Experimental\Parametres
C:\THESE\Experimental\Matrices
non
C:\THESE\Experimental\MOMEL_INTSINT

Les paramètres d'étendue « range » et clé « key »

Un calcul automatique des paramètres « range » et « key » est effectué en série pour chaque fichier TextGrid (application 25 du menu qui lance INTSINT ou INTSINT2 ou INTSINT3 selon l'option choisie). Cette application, décrite plus bas, crée une rangée INTSINT ou INTSINT2 ou INTSINT3 dans chaque TextGrid et stocke aussi les résultats du calcul dans un fichier texte.

L'application 26 permet de récapituler ces résultats dans un fichier paramètres accessible par tous les scripts de l'application (pitch_settings_cccc_iiii.txt oùcccc = corpus et iiii = le type de calcul d'INTSINT utilisé).

Extrait de pitch_settings_MARSEC_INTSINT3.txt :

speaker	key	minf0	maxf0	range
G0101B	188	117	330	1.2
G0101G	155	70	326	1.2
G0102B	204	104	465	1.4
G0102G	186	92	428	1.4
G0103B	187	70	429	1.4
G0103G	164	60	272	1.1
G0104B	183	80	427	1.2
G0104G	173	90	293	1
G0105B	189	80	428	1.3
G0105G	171	80	304	1.2
G0106B	160	60	364	1.4
G0106G	176	97	313	1.2
G0107B	180	60	317	1.1

Extrait de pitch_settings_MARSEC_INTSINT3.txt.

Les paramètres de modélisation

Ces paramètres décrivent les différents modèles, les formes et alignements à utiliser à chaque étape de l'optimisation.

Certains sont renseignés manuellement (dans EXCEL ou le Bloc Note), d'autres de façon complètement automatique.

La description des modèles :

Le fichier **param_models_cccc.txt** où cccc = le corpus (exemple param_models_MARSEC.txt) est le fichier de base, il contient tous les modèles et décrit la spécificité de chacun :

Pour chaque modèle, on a une entête identifiée « »qui donne les caractéristiques générales du modèle et des lignes identifiées « position » qui donnent le type d'alignement par position et par point.

Exemple pour MARSEC :

0	1	2	3	4	5	6	7	8

<type>	11	TU	INTSINT	reglenucléus	1	8	7	4
<position>	0	1	start0					
<position>	1	2	firstsyll%	end				
<position>	M	2	firstsyll%	end				
<position>	F	3	firstsyll%	lastsyll%	endF			
<position>	1F	3	firstsyll%	lastsyll%	endF			

Fichier param_models_MARSEC.txt : description du modèle 11.

<type>	62	TU	INTSINT3	reglenucléus	1	8	7	4
<position>	0	1	start0					
<position>	1	2	firstsyll%	end				
<position>	M	2	firstsyll%	end				
<position>	F	3	firstsyll%	lastsyll%	endF			
<position>	1F	3	firstsyll%	lastsyll%	endF			

Fichier param_models_MARSEC.txt : description du modèle 62.

Description de l'entête :

- Colonne 0 = <type> ;
- Colonne 1 = le numéro du type ;
- Colonne 2 = niveau d'analyse (pause, IU, TU, TTU, TTU2) ;
- Colonne3 = INTSINT ou INTSINT2 ou INTSINT3 selon que l'on veut utiliser INTSINT ou INTSINT2 ou INTSINT3 pour calculer la valeur des tons ;
- Colonne 4 = prise en compte de la règle du nucléus ou non (reglenucleus ou 0) ;
- Colonne 5 = application de la règle de la tonalité (calcul de la tonalité en début d'IU) (0 ou 1) ;
- Colonne 6 = score minimum que la combinaison doit atteindre pour être tracée dans le fichier de sortie pour les statistiques (pour éviter que les fichiers soient trop volumineux) ;
- Colonne 7 = score minimal pour qu'une séquence de tons soit acceptable pour le choix de la meilleure séquence qui est conservée sur le TextGrid, si aucune séquence de tons ne donne un score suffisant, on est dans le cas de « bad » score (cf. description de l'optimisation des tons et alignement) ;
- Colonne8 = type de calcul pour le score : 0=corrélation seule ; 1=moyenne du carré des écarts seule ; 2= corrélation arrondie + moyenne du carré des écarts ; 4= corrélation arrondie au supérieur ou standard ou inférieur selon l'importance du carré des écarts + carré des écarts.

Description des lignes « position » :

- Colonne 0 = « <position> »,
- Colonne1 = position du TU dans l'IU : 1, M, F, 1F (une ligne par position),
- Colonne2 = Nombre de points pour modéliser les segments occupant la position considérée,
- Colonne3 = type d'alignement pour 1er point,
- Colonne4 = type d'alignement pour 2eme point,
- Colonne5 = type d'alignement pour 3eme point (s'il y a),
- Colonne6 = type d'alignement pour 4eme point (s'il y a).

La description des tons possibles lors de la phase « optimisation des tons »

Ce fichier paramètre « param_tones_by_formcoding_*cccc*.txt où *cccc* représente le corpus est renseigné manuellement sous EXCEL. Il donne par type de codage des formes (INTSINT, INTSINT2, INTSINT3), par position du TU dans l'IU (0, 1, M, F, 1F) et par type de point (1 = 1er point, F = dernier point, M = les points du milieu), tous les tons possibles :

formcoding	position	point	tonality	tones							
INTSINT	0	1	0	M	T	B	_				
INTSINT	0	1	1	_							
INTSINT	0	F		T	B	H	L	U	D	S	_
INTSINT	1	1	0	M	T	B	_				
INTSINT	1	1	1	T	B	H	L	U	D	S	_
INTSINT	1	M		T	B	H	L	U	D	S	_
INTSINT	1	F		T	B	H	L	U	D	S	
INTSINT	M	1		T	B	H	L	U	D	S	_
INTSINT	M	M		T	B	H	L	U	D	S	_
INTSINT	M	F		T	B	H	L	U	D	S	
INTSINT	F	1		T	B	H	L	U	D	S	_
INTSINT	F	M		T	B	H	L	U	D	S	_
INTSINT	F	F		T	B	H	L	U	D	S	
INTSINT	1F	1	0	T	B	H	_				
INTSINT	1F	1	1	T	B	H	L	U	D	S	_
INTSINT	1F	M		T	B	H	L	U	D	S	_
INTSINT	1F	F		T	B	H	L	U	D	S	

Fichier param_tones_by_formcoding_MARSEC.txt : les tons possibles pour INTSINT par position et type de point.

formcoding	position	point	tonality	tones									
INTSINT2	0	1	0	M	T	B	_						
INTSINT2	0	1	1	_									
INTSINT2	0	F		T	B	H	L	h	l	u	d	S	_
INTSINT2	1	1	0	M	T	B	_						
INTSINT2	1	1	1	T	B	H	L	h	l	u	d	S	_
INTSINT2	1	M		T	B	H	L	h	l	u	d	S	_
INTSINT2	1	F		T	B	H	L	h	l	u	d	S	
INTSINT2	M	1		T	B	H	L	h	l	u	d	S	_
INTSINT2	M	M		T	B	H	L	h	l	u	d	S	_
INTSINT2	M	F		T	B	H	L	h	l	u	d	S	
INTSINT2	F	1		T	B	H	L	h	l	u	d	S	_
INTSINT2	F	M		T	B	H	L	h	l	u	d	S	_
INTSINT2	F	F		T	B	H	L	h	l	u	d	S	
INTSINT2	1F	1	0	T	B	H	_						
INTSINT2	1F	1	1	T	B	H	L	h	l	u	d	S	_
INTSINT2	1F	M		T	B	H	L	h	l	u	d	S	_
INTSINT2	1F	F		T	B	H	L	h	l	u	d	S	

Fichier param_tones_by_formcoding_MARSEC.txt : les tons possibles pour INTSINT2 par position et type de point.

INTSINT3	0	1	0	M	T	B	_						
INTSINT3	0	1	1	_									
INTSINT3	0	F		T	B	H	L	U	D	S	_		
INTSINT3	1	1	0	M	T	B	_						
INTSINT3	1	1	1	T	B	H	L	U	D	S	_		
INTSINT3	1	M		T	B	H	L	U	D	S	_		
INTSINT3	1	F		T	B	H	L	U	D	S			
INTSINT3	M	1		T	B	H	L	U	D	S	_		
INTSINT3	M	M		T	B	H	L	U	D	S	_		
INTSINT3	M	F		T	B	H	L	U	D	S			
INTSINT3	F	1		T	B	H	L	U	D	S	_		
INTSINT3	F	M		T	B	H	L	U	D	S	_		
INTSINT3	F	F		T	B	H	L	U	D	S			
INTSINT3	1F	1	0	T	B	H	_						
INTSINT3	1F	1	1	T	B	H	L	U	D	S	_		
INTSINT3	1F	M		T	B	H	L	U	D	S	_		
INTSINT3	1F	F		T	B	H	L	U	D	S			

Fichier param_tones_by_formcoding_MARSEC.txt : les tons possibles pour INTSINT3 par position et type de point.

La description des types d'alignements

Le fichier paramètres param_types_alignment_cccc.txt où cccc identifie le corpus est renseigné manuellement dans EXCEL. Il permet de décrire tous les types d'alignement utilisés dans l'application, A chaque type d'alignement correspond un traitement particulier dans l'application : il n'est donc pas possible de créer un nouveau type d'alignement sans modifier le script. Les types d'alignement en pourcentage ont toujours comme préfixe « % ».

Ce fichier est composé de 4 parties identifiées par la colonne 0 (« general », « tones », « alignment », « model »).

La partie « general » correspond à la liste des types d'alignement suivie de leurs libellés utilisés dans les statistiques.

general	start0	valeur	debut_tu
general	start	valeur	debut_tu
general	startiu	valeur	debut_iu
general	prec	valeur	point_precedent
general	end	valeur	fin_tu
general	endL	valeur	fin_tu
general	endF	valeur	fin_tu
general	endiu	valeur	fin_iu
general	start%	%	duree_tu
general	med%	%	duree_tu
general	end%	%	duree_tu
general	firstsyll%	%	duree_1ere_syllabe
general	secondsyll%	%	duree_2eme_syllabe
general	lastsyll%	%	duree_derniere_syllabe
general	lastSON	nb_syllabe	debut_derniere_syllabe
general	lastSOFF	nb_syllabe	fin_derniere_syllabe
general	firstvowel	valeur	debut_1ere_voyelle
general	lastVON	valeur	debut_1ere_voyelle_derniere_syllabe

param_types_alignment_MARSEC.txt : description générale.

Les parties « tones », « alignment » et « model » décrivent comment on fera varier l'alignement des tons pour chaque type d'alignement, en fonction du niveau d'analyse (TU, TTU, TTU2). Si l'alignement est le même quel que soit le niveau d'analyse, on pourra le noter ainsi : « TU,TTU,TTU2 ».

La colonne « deltamin » indique le positionnement minimum par rapport au repère, « deltamax » indique le positionnement maximum par rapport au repère et la colonne « nbiter » indique combien on fera d'itérations. Exemples : pour le type d'alignement « start » décrit dans le tableau ci-dessous : on fera varier le positionnement du point aligné avec le type « start » à

100 et 200 ms du début du TU ; pour le type d'alignement « firstsyll% » : on fera varier le positionnement du point aligné avec le type « firstsyll%» à 0%, 50% et 100% de la 1ère syllabe du TU.

L'option « tones » est utilisée lors de la première étape de l'optimisation nommée « optimisation des tons ». À ce niveau le nombre d'itérations est restreint pour éviter un trop grand nombre de combinaisons.

Option	position	niv_analyse	Type d'alignement	deltamin	deltamax	nbiter
tones		TU,TTU,TTU2	start0	0	0	1
tones		TU,TTU,TTU2	start	0	200	3
tones		TU,TTU,TTU2	prec	100	200	2
tones		TU,TTU,TTU2	end	-60	-10	2
tones		TU,TTU,TTU2	endF	-10	-10	1
tones		TU,TTU,TTU2	start%	10	40	3
tones		TU,TTU,TTU2	med%	50	75	3
tones		TU,TTU,TTU2	end%	80	95	3
tones		TU,TTU,TTU2	firstsyll%	0	100	3
tones		TU,TTU,TTU2	lastsyll%	-25	75	3
tones		TU,TTU,TTU2	lastSON	-1	0	2
tones		TU,TTU,TTU2	lastSOFF	-1	0	2
tones		TU,TTU,TTU2	firstvowel	0	200	3
tones		TU,TTU,TTU2	lastVON	0	200	3
tones		TU,TTU,TTU2	lastpik	0	0	1
tones		IU,pause	startiu	100	300	3
tones		IU,pause	endiu	-300	-200	2

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour la 1ére phase (optimisation des tons).

L'option « alignment » est utilisée lors de la deuxième étape de l'optimisation nommée « optimisation de l'alignement ». Ici le nombre d'itérations est plus conséquent.

Option	position	niv_analyse	Type d'alignement	deltamin	deltamax	nbiter
alignment		TU,TTU,TTU2	start0	0	0	1
alignment		TU,TTU,TTU2	start	0	300	7
alignment		TU,TTU,TTU2	prec	50	300	6
alignment		TU,TTU,TTU2	end	-160	-10	4
alignment		TU,TTU,TTU2	endF	-10	-10	1
alignment		TU,TTU,TTU2	start%	10	40	7
alignment		TU,TTU,TTU2	med%	60	75	4
alignment		TU,TTU,TTU2	end%	80	95	4
alignment		TU,TTU,TTU2	firstsyll%	0	150	7
alignment		TU,TTU,TTU2	lastsyll%	-40	80	7
alignment		TU,TTU,TTU2	lastSON	-2	0	3
alignment		TU,TTU,TTU2	lastSOFF	-2	0	3
alignment		TU,TTU,TTU2	firstvowel	-100	300	9
alignment		TU,TTU,TTU2	lastVON	-100	300	9
alignment		IU,pause	startiu	100	500	9
alignment		IU,pause	endiu	-500	-100	9

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour la 2ème phase (optimisation de l'alignement).

L'option « model » est utilisée lors de la troisième étape de l'optimisation nommée « ultimate_model » et aussi pour la détection des fonctions et des frontières. Ici les paramètres d'alignement ont été « travaillés » en fonction des résultats obtenus lors des étapes précédentes. Ils peuvent varier en fonction de la position du TU dans l'IU.

Option	position	niv_analyse	Type d'alignement	deltamin	deltamax	nbiter
model	0	TU,TTU,TTU2	start0	0	0	1
model	0	TU,TTU,TTU2	end	-60	-10	2

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour le model final et en détection, pour les TU en position 0.

model	1	TU	start0	0	0	1
model	1	TU	start	0	200	5
model	1	TU	firstsyll%	0	70	3
model	1	TU	end	-60	-10	2
model	1	TU,TTU,TTU2	lastSOFF	0	0	1
model	1	TU	start%	10	40	3
model	1	TU	end%	95	95	1
model	1	TU,TTU,TTU2	firstvowel	-50	150	5
model	1	TTU,TTU2	start0	0	0	1
model	1	TTU,TTU2	start	0	100	3
model	1	TTU,TTU2	firstsyll%	0	70	3
model	1	TTU,TTU2	prec	100	150	2
model	1	TTU,TTU2	end	-60	-10	2

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour le model final et en détection, pour les TU en position 1.

Option	position	niv_analyse	Type d'alignement	deltamin	deltamax	nbiter
model	M	TU	start	0	300	4
model	M	TU	firstsyll%	0	70	3
model	M	TU	end	-60	-10	2
model	M	TU,TTU,TTU2	lastSOFF	0	0	1
model	M	TU,TTU,TTU2	lastSON	0	0	1
model	M	TU	start%	10	40	3
model	M	TU	end%	95	95	1
model	M	TTU,TTU2	start	0	100	3
model	M	TTU,TTU2	firstsyll%	0	120	4
model	M	TTU,TTU2	prec	100	300	3
model	M	TTU,TTU2	lastsyll%	0	70	3
model	M	TTU,TTU2	end	-60	-10	2
model	M	TU,TTU,TTU2	firstvowel	-50	150	5

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour le model final et en détection, pour les TU en position M.

Option	position	niv_analyse	Type d'alignement	deltamin	deltamax	nbiter
model	F	TU	start	0	150	3
model	F	TU	firstsyll%	0	70	3
model	F	TU	prec	100	300	3
model	F	TU	lastsyll%	0	70	3
model	F	TU	endF	-10	-10	1
model	F	TU,TTU,TTU2	lastSOFF	0	0	1
model	F	TU,TTU,TTU2	lastSON	0	0	1
model	F	TU	start%	10	40	3
model	F	TU	end%	75	95	3
model	F	TU,TTU,TTU2	lastVON	-50	150	5
model	F	TU,TTU,TTU2	firstvowel	-50	150	5
model	F	TTU,TTU2	start	0	150	3
model	F	TTU,TTU2	firstsyll%	0	70	3
model	F	TTU,TTU2	prec	100	300	3
model	F	TTU,TTU2	lastsyll%	0	70	3
model	F	TTU,TTU2	endF	-10	-10	1

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour le model final et en détection, pour les TU en position F.

model	1F	TU	start0	0	0	1
model	1F	TU	start	0	200	4
model	1F	TU	firstsyll%	0	70	3
model	1F	TU	prec	100	150	3
model	1F	TU	lastsyll%	0	70	3
model	1F	TU	endF	-10	-10	1
model	1F	TU,TTU,TTU2	lastSOFF	0	0	1
model	1F	TU,TTU,TTU2	lastSON	0	0	1
model	1F	TU	start%	10	40	3
model	1F	TU	end%	75	95	3
model	1F	TU,TTU,TTU2	lastVON	-50	150	5
model	1F	TU,TTU,TTU2	firstvowel	-50	150	5
model	1F	TTU,TTU2	start0	0	0	1
model	1F	TTU,TTU2	start	0	150	3
model	1F	TTU,TTU2	firstsyll%	0	70	3
model	1F	TTU,TTU2	prec	100	150	3
model	1F	TTU,TTU2	end	-200	-100	3
model	1F	TTU,TTU2	lastsyll%	0	70	3
model	1F	TTU,TTU2	endF	-10	-10	1

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour le model final et en détection, pour les TU en position 1F .

model	1F	IU,pause	startiu	100	500	9
model	1F	IU,pause	endiu	-500	-100	9

param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour le model final et en détection, pour les modèles au niveau d'analyse IU ou pause.

La description des séquences de tons retenues par modèles et par fonctions aux différentes étapes de l'optimisation

Les fichiers Param_optimize_alignment_*cccc*.txt et Param_ultimate_model_detect_*cccc*.txt (où « *cccc* » représente le corpus) sont générés automatiquement par la transaction 24 du menu, à partir des résultats de l'application R qui calcule les formes les plus courantes pour chaque fonction.

<type>	11	
<function>	[!|_U0	
_		
<function>	[!|_S1	
D	T	
D	H	
L	H	
L	T	
B	T	
<function>	[!|_SM	
D	H	
S	H	
D	T	
S	U	
L	T	
<function>	[!|_STM	
_	B	
_	D	
_	L	
_	S	
B	B	
<function>	[!|_STF	
U	B	B
H	B	U
S	B	B
D	B	B
_	B	B

Extrait de Param_ultimate_model_detect_MARSEC.txt pour le modèle 11.

L'application 15 permet de lister ces paramètres sous format html

Type 11 intsint Niveau analyse=tu Back to TOP

Score mini pour stats/detect : 8 Avec regle du nucleus

Type calcul score=4 Score mini pour OK=7 Marge acceptable en 1/2 tons=4

Positionnement model

0	1	M	F	1F
start0 0 0 0x1	firstsyll* 0 120 40x4 end -6 -1 5x2	firstsyll* 0 120 40x4 end -6 -1 5x2	firstsyll* 0 70 35x3 lastsyll* 0 70 35x3 endF -1 -1 0x1	firstsyll* 0 70 35x3 lastsyll* 0 70 35x3 endF -1 -1 0x1

[+U0	[+S1	[+SM	[+SF	[+STM	[+STF	[+S1F	[+A1	[+AM	[+ATF	[+N1	[+!N1	[+NM	[+!NM	[+NF	[+!NF	[+N1F	[+!N1F
_,	_H, DH, DT, _T, _U,	_D, _L, DD, SD, DL,		_L, _D, DL, DD, _B,			TD, HD, TL, UD, HL,	UL, UD, HL, SD, _L,	_LU, _BU, DLU, DBU, DDU,	HL, UD, HD, UL, TL,	TL, TD, HB, UB, TB,	_D, _L, DL, DD, SL,	TL, UD, HL, UL, TD,	SLU, SBU, DBU, SDS, SLS,	HBU, UBU, TBH, HBH, ULU,	HLS, HLD, TLD, ULS, ULD,	TBU, ULD, TBH, HBU, HLU,

[|U0	[|S1	[|SM	[|SF	[|STM	[|STF	[|S1F	[|A1	[|AM	[|ATF	[|N1	[|!N1	[|NM	[|!NM	[|NF	[|!NF	[|N1F	[|!N1F
_,	_H, _T, _U, UH, UU,	SL, _L, _D, _U, DD,		UB, HB, TB, _B, DL,	DLS, _LS, DLD, SLS, _DS,		TD, HD, UD, TL, HL,	UL, HL, UD, TL, DL,		UL, HB, UB, UD, SL,	UL, HL, HB, TB, TL,	SL, SD, SB, DB, DL,	TL, UL, HL, UD, HB,	UBB, SLS, SLD, SDS, SBB,	ULD, ULS, UBB, HBB, TBB,	ULD, HLD, ULS, UDD, U_L,	HBB, TBB, UBB, ULB, HLB,

[!+U0	[!+S1	[!+SM	[!+SF	[!+STM	[!+STF	[!+S1F	[!+A1	[!+AM	[!+ATF	[!+N1	[!+!N1	[!+NM	[!+!NM	[!+NF	[!+!NF	[!+N1F	[!+!N1F
_,	BT, DT, HT, LT, UT,	_L, _D, DD, _B, BH,		_D, _L, DL, DD, _B,			HD, HL, TD, TL, UD,	UL, HB, TB, HL, UB,	LTT, DHT, LHT, LHU, BHT,	HL, HD, HB, UL, UD,	TL, HL, HD, UD, UL,		UL, HL, UD, HB, TB,	BUU, B_U, BUH, L_U, DUU,	HBU, UBU, TBU, TBH, H_B,		TBU, TBH, HBU, UBU, HBH,

[!|U0	[!|S1	[!|SM	[!|SF	[!|STM	[!|STF	[!|S1F	[!|A1	[!|AM	[!|ATF	[!|N1	[!|!N1	[!|NM	[!|!NM	[!|NF	[!|!NF	[!|N1F	[!|!N1F
_,	DT, DH, LH, LT, BT,	DH, SH, DT, SU, LT,		_B, _D, _L, _S, BB,	UBB, HBU, SBB, DBB, _BB,		HL, TL, UL, UD, HD,	UL, HB, UB, HD, UD,				DB, _B, DL, SB, UB,	HB, TB, UB, UL, HL,	UBB, HBB, U_B, TBB, H_B,	HBB, U_B, TBB, UBB, H_B,		TLD, TLS, STB, DTB, UTB,

[?+U0	[?+S1	[?+SM	[?+SF	[?+STM	[?+STF	[?+S1F	[?+A1	[?+AM	[?+ATF	[?+N1	[?+!N1	[?+NM	[?+!NM	[?+NF	[?+!NF	[?+N1F	[?+!N1F
_,	DH, LT, DT, BT, LH,	_D, _L, SD, DD, DL,		UH, UU, _H, HU, _T,			HL, UD, SD, UL, TD,	_L, DB, SB, _B, LB,	STT, UTT, SHS, UHT, UHU,	HT, _T, UT, _H, HU,	SU, LT, DT, BT, _U,	BH, BU, LU, DU, LH,	_L, _D, LD, UT, _T,	UDT, SDT, SLT, LUT, DUT,		DTD, LTT, DTT, STT, BTT,	DTT, LTT, BTT, DHU, _TD,

[?|U0	[?|S1	[?|SM	[?|SF	[?|STM	[?|STF	[?|S1F	[?|A1	[?|AM	[?|ATF	[?|N1	[?|!N1	[?|NM	[?|!NM	[?|NF	[?|!NF	[?|N1F	[?|!N1F
_,	_T, ST, UT, _H, SH,	UB, HB, TB, UL, DL,			__B, DLD, D_B, DLB, DDL,		TD, HD, HS, TT, US,	UT, _T, ST, HT, DT,				TL, TD, HD,	LD, LB, _B, LS, DB,	__BB, __B, B_B, BBB, LBB,	HLD, D_L, SDD, TDD, __L,	HLS, SLS, TLS, ULS, H_L,	

[U0	[S1	[SM	[SF	[STM	[STF	[S1F	[A1	[AM	[ATF	[N1	[!N1	[NM	[!NM	[NF	[!NF	[N1F	[!N1F
_,	TD, HD, UD, HL, TL,		SDS, UDS, S_D, SLS, ULU,			T_D, U_D, H_D, TTD, H_L,	DT, DH, LT, BT, LH,							L_U, DSU, _DU, D_U, SDU,		STT, DTT, UTT, LTT, U_T,	

+U0	+S1	+SM	+SF	+STM	+STF	+S1F	+A1	+AM	+ATF	+N1	+!N1	+NM	+!NM	+NF	+!NF	+N1F	+!N1F
_,	BU, DS, LU, LS, _D,	SU, _U, SH, _H, UH,	SUU, SUH, DHT, LHT, SHT,			DUU, U_T, DHT, LHT, S_T,								SUS, S_U, __U, _US, SUU,		_BU, _DU, _LU, _SU,	

|U0	|S1	|SM	|SF	|STM	|STF	|S1F	|A1	|AM	|ATF	|N1	|!N1	|NM	|!NM	|NF	|!NF	|N1F	|!N1F
_,	BU, LU, DU, SU, LH,	SL, _D, _L, SD, DL,	UBB, SLD, SLS, HBB, SBB,			BUU, DUU, LUU, SUU, B_U,											

Liste des paramètres « ultimate_model » pour le modèle 11.

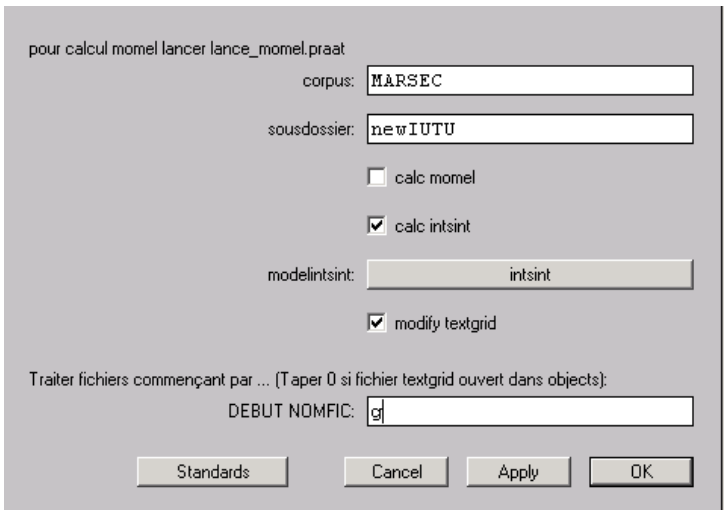
3.3. Calcul de l'étendue, de la clé et des valeurs INTSINT

Ce traitement peut être lancé à partir du menu : numéro de transaction 25, ou directement dans Praat.

C'est un script Praat, il permet de calculer en série, pour un ensemble de TextGrid les paramètres « key » et « range » et les valeurs de INTSINT en faisant appel à l'algorithme de calcul de INTSINT que l'on aura choisi au niveau du menu : INTSINT, INTSINT2 ou INTSINT3 (scripts Perl décrits ci-dessous). Les résultats du calcul sont stockés dans un fichier intermédiaire puis repris pour créer la rangée INTSINT, INTSINT2 ou INTSINT3 dans le TextGrid. Ce sont ces résultats qui serviront de base à la transaction 26 pour créer le fichier « pitch_settings_xxxx ».

Le script de lancement du calcul de INTSINT

C'est le script « Lanc_MOMEL_INTSINT.Praat », lorsqu'il est lancé directement dans Praat, la fenêtre de dialogue suivante s'affiche :



Fenêtre de dialogue du script « Lanc_MOMEL_INTSINT.Praat »

Ce script effectue les opérations suivantes :

- Si le bouton « calc MOMEL » a été activé : lancement du calcul de MOMEL pour tous les fichiers sélectionnés, puis si la touche « modify TextGrid » a été activée : création de la rangée « MOMEL » sur le TextGrid (après suppression de l'ancienne rangée si déjà présente) ;
- Si le bouton « calc INTSINT » a été activé : lancement du calcul de INTSINT pour tous les fichiers sélectionnés :
    - génération du fichier intermédiaire *ffff*.cb (où *ffff* est le nom du fichier TextGrid) à partir de la rangée « MOMEL »

- Lancement du script Perl INTSINT, INTSINT2 ou INTSINT3 selon le mode INTSINT demandé
- Si la touche « modify TextGrid » a été activée : récupération du résultat dans le fichier *ffff*.int et création de la rangée « INTSINT » ou « INTSINT2 » ou « INTSINT3 » sur le TextGrid (après suppression de l'ancienne rangée si déjà présente).

Le script de calcul de INTSINT

Ce script Perl détermine pour un fichier TextGrid les meilleurs « Key » et « range » et valeurs de INTSINT pour se rapprocher le plus de la courbe MOMEL.

C'est le script « INTSINT_v211.pl » (duplication du script « INTSINT_v210.pl » disponible sur le site du LPL auquel nous avons apporté quelques modifications mineures qui seront signalées dans la description du script).

Ce script effectue les opérations suivantes :

- Récupération des valeurs de f0 (origine MOMEL) à partir du fichier texte « *ffff*.cb » (où « *ffff* » représente le nom du TextGrid) et conversion de ces valeurs en octave (tous les calculs se font en octave) ;
- Calcul de la moyenne des valeurs de f0 en octave (mean_f0) et en hertz (linear_mean_f0) ;
- Fait varier le « range » de 0.9 à 1.1 octave (valeurs dans la version originale = 0.5 octave à 2.5 octaves) avec pas de 0.1 octave et pour chaque valeur du range :
    - fait varier le « key » entre (linear_mean_f0 -50) hertz et (linear_mean_f0 + 50) hertz avec des pas de 1 hertz et pour chaque valeur du « key » :
        - traduit le « key » en octave => « mid »
        - effectue le traitement « optimize » pour ce range et mid :
            - pour chaque valeur de f0 : cherche la valeur d'INTSINT qui donne le meilleur résultat
            - pour le 1er point ou après une pause on retient uniquement T,B,M (on considère qu'il y a une pause si la distance entre 2 points f0 est > 0.5)
            - sinon on essaie tous les tons INTSINT excepté M, soit : T, B, H, U, L, D, S ;
            - et pour chaque valeur de INTSINT, on calcule une valeur de f0 de la façon suivante :
                - M => valeur calculée= « mid »
                - T => valeur calculée = « top » (« mid »+ « range »/2)
                - B => valeur calculée = « bottom » (« mid » - « range »/2)
                - H => valeur calculée =

« last_target » + (« top »-« last_target ») *0.5

- U => valeur calculée =
  « last_target » + (« top »-« last_target ») *0.25
- L => valeur calculée =
  « last_target » - (« last_target »-« bottom ») *0.5
- D => valeur calculée =
  « last_target » - (« last_target »-« bottom ») *0.25
- S => valeur calculée = « last_target »

(« last-target » est la valeur calculée retenue pour le point précédent).

La valeur calculée retenue est celle qui s'approche le plus de la valeur de f0 (résultat MOMEL) (« error=abs(val-f0) »

Cumul du carré des erreurs (« ss_error += error*error »)

- On retient finalement le « range » et le « key » et les valeurs de « INTSINT » correspondantes pour lesquels la somme du carré des erreurs est la plus faible
- Tous ces résultats sont enregistrés sur le fichier de sortie *ffff*.int

Le script de calcul de INTSINT2

Ce script « INTSINT_v5.pl » se déroule de façon identique au script « INTSINT_v211 », nous ne reprenons ici que ce qui différencie les deux scripts :

- Le « range » varie de 0.9 à 1.4 ;
- Deux valeurs de « INTSINT » supplémentaires « u » et « d » ;
- L'étendue est divisée en 3 parties, selon la position du point précédent « last_target » dans l'étendue, les tons possibles diffèrent.
- Exemple pour un mouvement ascendant :
  - Dans tous les cas les 2 tons suivants sont possibles :
    - T => valeur calculée = « top »,
    - S => valeur calculée = « last-target »
  - Si la distance entre le dernier point connu et le « top » est >= 60% de l'étendue : 3 tons sont possibles (autres que T ou S) :
    - H => valeur calculée = « last-target » + (« top » - « last-target ») * 0.75
    - U => valeur calculée = « last-target » + (« top » - « last-target ») * 0.50
    - u => valeur calculée = « last-target » + (« top » - « last-target ») * 0.25
  - Si la distance entre le dernier point connu et le « top » est comprise entre30 et 60% de l'étendue : 2 tons sont possibles (autres que T ou S) :
    - U => valeur calculée = « last-target » + (« top » - « last-target ») * 0.50
    - u => valeur calculée = « last-target » + (« top » - « last-target ») * 0.25
  - Si la distance entre le dernier point connu et le « top » est inférieure à 30% de l'étendue : 1 seul ton est possible (autres que T ou S) :
    - u => valeur calculée = « last-target » + (« top » - « last-target ») * 0.50

- Pour un mouvement descendant :
    - Dans tous les cas les 2 tons B ou S sont possibles.
    - Si la distance entre le dernier point connu et le « bottom » est >= 60% de l'étendue : 3 tons sont possibles (autres que T ou S) : L, D,d ;
    - Si la distance entre le dernier point connu et le « bottom » est comprise entre30 et 60% de l'étendue : 2 tons sont possibles (autres que T ou S) : D et d ;
    - Si la distance entre le dernier point connu et le « bottom » est inférieure à 30% de l'étendue : 1 seul ton est possible (autres que T ou S) : d ;

Le script de calcul de INTSINT3

Ce script « INTSINT_v6.pl » se déroule de façon identique au script « INTSINT_v211 » et « INTSINT_v5 », nous ne reprenons ici que ce qui différencie les scripts :

- Le « range » varie de 0.9 à 1.4
- Les tons INTSINT possibles sont M, T, B, H, U, L, D, S
- L'étendue est divisée en 2 parties, selon la position du point précédent « last_target » dans l'étendue, les tons possibles diffèrent, exemple pour un mouvement ascendant :
    - Dans tous les cas les 2 tons suivants sont possibles :
        - T => valeur calculée = « top »,
        - S => valeur calculée = « last-target »
    - Si la distance entre le dernier point connu et le « top » est >=30% de l'étendue : 2 tons sont possibles (autres que T ou S) :
        - H => valeur calculée = « last-target » + (« top » - « last-target ») * 0.60
        - U => valeur calculée = « last-target » + (« top » - « last-target ») * 0.30
    - Si la distance entre le dernier point connu et le « top » est inférieure à 30% de l'étendue : 1 seul ton est possible (autres que T ou S) :
        - U => valeur calculée = « last-target » + (« top » - « last-target ») * 0.50
- Pour un mouvement descendant :
    - Dans tous les cas les 2 tons B ou S sont possibles.
    - Si la distance entre le dernier point connu et le « bottom » est >= 30% de l'étendue : 2 tons sont possibles (autres que T ou S) : L, D
    - Si la distance entre le dernier point connu et le « bottom » est inférieure à 30% de l'étendue : 1 seul ton est possible (autres que T ou S) : D

3.4. Le programme d'optimisation des formes et de l'alignement et de détection des types de TU et/ou IU et des frontières

Nous décrivons ici, la dernière version du script, utilisée pour le corpus MARSEC.

Principes

Les données de base sont stockées sous forme de TextGrid (format « short textfile » pour Praat).

Rangées utilisées :

- syllables
- phonemes
- word
- IU
- TU, TTU, TTU2
- MOMEL

Ce programme permet de sélectionner :

- Les TextGrid sur lesquels on veut travailler : corpus, sous-dossier et liste des fichiers commençant par…,
- L'étape du traitement,
- Le modèle d'optimisation.

Les étapes

- Optimisation des séquences de tons,
- Optimisation de l'alignement,
- « ultimate_model »,
- Détection des types de TU ou des types d'IU ou des fonctions,
- Détection des frontières/fonctions.

Les modèles d'optimisation :

Les différents modèles d'optimisation sont définis dans le fichier « param_models_*cccc*.txt » (où *cccc* représente le corpus), cf. paragraphe ci-dessus.

Ils sont décrits dans le chapitre 3.2 des annexes.

Le traitement d'optimisation est identique quel que soit le niveau choisi.

Étape 1 : l'optimisation des séquences de tons

L'optimisation se fait au niveau d'un intervalle (entre 2 pauses, à l'intérieur d'un IU ou à l'intérieur d'un TU (TU, TTU ou TTU2) : l'intervalle dépend du type d'optimisation sélectionné.

Pour expliquer le fonctionnement du script, nous allons prendre comme exemple le modèle 11, qui se situe au niveau du TU. Ci-dessous les paramètres utilisés dans la phase d'optimisation des séquences de tons pour le modèle 11 :

<type>	11	TU	INTSINT	reglenucléus	1	8	7	4
<position>	0	1	start0					
<position>	1	2	firstsyll%	end				
<position>	M	2	firstsyll%	end				
<position>	F	3	firstsyll%	lastsyll%	endF			
<position>	1F	3	firstsyll%	lastsyll%	endF			

Extrait du fichier param_models_MARSEC.txt : description du modèle 11.

Option	position	niv_analyse	Type d'alignement	deltamin	deltamax	nbiter
tones		TU,TTU,TTU2	end	-60	-10	2
tones		TU,TTU,TTU2	endF	-10	-10	1
tones		TU,TTU,TTU2	firstsyll%	0	100	3
tones		TU,TTU,TTU2	lastsyll%	-25	75	3

Extrait de param_types_alignment_MARSEC.txt : les types d'alignement utilisés pour le modèle 11, dans la phase d'optimisation des tons, pour tous lesTU, quelle que soit leur position.

formcoding	position	point	tonality	tones							
INTSINT	1	1	1	T	B	H	L	U	D	S	_
INTSINT	1	F		T	B	H	L	U	D	S	

Extrait du fichier param_tones_by_formcoding_MARSEC.txt : les tons possibles pour INTSINT pour un TU en position 1 , pour le 1er et dernier point.

Nous décrivons ci-dessous, les séquences de traitements effectuées par le script.

Détermination de la position du TU dans l'IU et du type de fonction

Pour les modèles 1 et 3 qui se situent respectivement au niveau d'un intervalle entre 2 pauses ou de l'IU, on considère qu'il y a une unité tonale dans l'IU, sa position est codifiée 1F.

Pour les autres modèles, qui se situent au niveau du TU, la position du TU dans l'IU et le type de fonction sont déterminés ainsi :

	TU Exemple	position	Type de fonction	Type du TU dans l'IU (ex. pour un IU [+)
1er TU de l'IU et TU non accentué	U	0	U0	[+_U0
1er TU accentué de l'IU	A	1	A1	[+_A1
TU situés en milieu d'IU	S	M	SM	[+_SM
Dernier TU de l'IU	N	F	NF	[+_NF
Un seul TU dans l'IU	N	1F	N1F	[+_N1F
TUs situé derrière le nucléus	S	M ou F	STM ou STF	[+_STM ou [+_STF

Détermination du type de fonction .

Essai de toutes les combinaisons d'alignement de points possibles

Tout d'abord, le script essaie toutes les combinaisons de points selon les règles données en paramétrage. Exemple, pour un TU en position 1 dans le modèle 11 :

Ce type de TU est modélisé avec 2 points, le 1er aligné en % de la durée de la 1ère syllabe et le second en valeur par rapport à la fin du TU. Nous aurons les combinaisons suivantes :

- 0%, -60
- 0%, -10
- 50%, -60
- 50%, -10
- 100%, -60
- 100%, -10

Pour chaque combinaison d'alignement, on essaie de toutes les séquences de tons possibles.

La tonalité

Dans les 1ers modèles (1 à 10), les TU en position 1 sont modélisés avec 3 points. Le 1er point peut prendre les valeurs M, T, B ou _ (_ = pas de tons). Si la durée de la pause précédant le TU est < 300 ms, le ton retenu est « _ », sinon c'est M, T, B (le ton le plus proche de la courbe MOMEL). Si le TU en position 1 est situé derrière un TU 0, le 1er ton est alors « _ ».

Cette méthode posait problème (multiplication du nombre de combinaisons de tons possibles pour le point 1). Nous avons alors pensé « externaliser » ce ton que nous avons appelé « tonalité ». C'est un ton que le script positionne au début d'un IU situé après une pause longue (> 300 ms) et qui peut prendre les valeurs M, T, B (le ton le plus proche de la courbe MOMEL).

Tous les modèles à partir du modèle 11 sont des modèles « avec tonalité », le 1er point d'un TU est modélisé avec 2 points.

Essai de toutes les combinaisons de tons possibles pour chaque combinaison d'alignement

Pour un TU modélisé avec 2 points, on essaie les combinaisons suivantes :

- TT, TB, TH, TL, TU, TD, TS (mais TH et TU sont immédiatement rejetés car interdits)
- Etc,
- _T, _B, _H, _L, _U, _D, _S

Pour chaque combinaison de tons, le script effectue les opérations suivantes :

- conversion des tons « INTSINT » en « target » (unité demi-ton, 12 demi-tons dans une octave), en utilisant l'algorithme « INTSINT » ou « INTSINT1 » ou, « INTSINT2 » ou « INTSINT3 » selon le modèle traité ;
- interpolation linéaire ➔ une valeur de « pitch » toutes les 10 ms ;
- calcul de la corrélation entre le résultat (pitch) et le résultat de l'interpolation linéaire des points MOMELs détectés sur l'intervalle (pitch_MOMEL), en appliquant la formule suivante :

$$r_p = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

  où x représente les valeurs de MOMEL et y les valeurs « pitch » de la combinaison.
- calcul de la déviation moyenne entre les 2 courbes = racine carrée du (cumul du carré des écarts / nombre de points)

  √ ((∑ (xi-yi) **2) / nb points)

- calcul du score = c,dddd : le calcul est sensiblement différent selon le modèle (paramètre« type de calcul du score) :

type de calcul du score	score
0	corrélation * 10
1	9.9999 – moyenne du carré des écarts (RMSE)
2	c=partie entière de corrélation *10, arrondie au standard dddd=(9.9999 - RMSE)/10
4	Idem 2, mais la règle d'arrondi pour « c » diffère en fonction de la valeur du carré des écarts : arrondi au supérieur si RMSE < 1, arrondi standard si RMSE < 2, sinon arrondi inférieur

Calcul du score.

Exemple, pour le modèle 11, qui utilise le mode de calcul 4, si la corrélation est égale 0,9451 et la RMSE est égale à 0,856, alors le score devient 10,9143.

- comparaison du nouveau score obtenu avec le meilleur score déjà rencontré sur l'intervalle : si le nouveau score est supérieur au meilleur score obtenu sur cet intervalle, on conserve cette valeur et la séquence de tons et d'alignement des points correspondants : il devient alors le meilleur score ;
- si le score est >= au score minimum pour les statistiques la combinaison est enregistrée sur le fichier statistiques (G_optimize_tones_11.txt pour le modèle 11), avec en parallèle le type d'IU et/ou de TU, le type de fonction [12]et le type de TU dans l'IU[13] : ces données sont reprises ensuite par un script R ;
- en final : on enregistre sur le TextGrid la meilleure séquence de tons et d'alignement (dans une rangée nommée int*nyx* (où *n* représente le numéro d'INTSINT, ici vide, *y*=l'étape et *x* le modèle d'optimisation). Pour le modèle la rangée s'appelle : inttone11 ;
- en final, on enregistre aussi la meilleure séquence de tons et alignement obtenue sur un fichier text (G_optimize_tones_11_best.txt), ce fichier est repris ensuite pour des statistiques sur l'alignement.

[12] Type de fonction (type_func) constitué de *TU +position du TU dans l'IU ex :* NF

[13] Type de TU dans l'IU (typTUIU) constitué de l'*IU+_+TU +position du TU dans l'IU ex :* [+_NF

Étape 1 suite : détermination des meilleures formes

Le fichier G_optimize_tones_11.txt qui contient toutes les combinaisons de tons et alignement dont le score est > 8 (paramètre pour le modèle11) est très volumineux. D'autre part, le fait de faire varier l'alignement peut aussi amener des séquences de tons moyennes. Aussi, nous avons décidé de ne garder, pour un TU donné, que les séquences de tons qui correspondent au meilleur alignement retenu sur le TU.

Un script Perl (prep_optimize_forR.pl), lancé automatiquement après l'optimisation des tons, permet de sélectionner ces données et de créer le fichier G_optimize_tones_11_bestalign.txt.

Un script R[14] agrège ces données, par type d'IU, par fonction et par forme, en cumulant les scores obtenus. Les formes sont alors affichées en ordre décroissant et enregistrées sur un fichier EXCEL (R8_G_optimize_tones_11_bestalign.txt) qui est présenté en annexe 2.

Puis un autre script Perl (transaction 24 au menu), permet de récupérer automatiquement les meilleures séquences de tons ainsi déterminées et de les écrire sur le fichier paramètres « Param_optimize_alignment_MARSEC.txt ». Pour cette étape, nous gardons les 30 meilleures séquences de tons par fonction, mais il est possible de changer ce nombre puisqu'il fait partie des paramètres donnés lors du lancement du script.

Étape 2 : l'optimisation de l'alignement

Cette étape fonctionne comme la précédente, mais elle s'appuie sur un nombre de formes plus restreint (surtout lorsqu'il y a 3 points possibles), ce qui permet de faire varier plus amplement l'alignement. Ci-dessous les paramètres utilisés pour faire varier l'alignement dans le modèle 11 :

Option	position	niv_analyse	Type d'alignement	deltamin	deltamax	nbiter
alignment		TU,TTU,TTU2	end	-160	-10	4
alignment		TU,TTU,TTU2	endF	-10	-10	1
alignment		TU,TTU,TTU2	firstsyll%	0	150	7
alignment		TU,TTU,TTU2	lastsyll%	-40	80	7

Extrait du fichier types d'alignements.

Les formes possibles à ce niveau ont été générées suite à l'étape 1, comme décrit ci-dessus.

[14] script R *R_optimise_tone_iu_typtu_numtu_function_newIU_newTU.txt*

Étape 2 suite : détermination des meilleures formes et meilleurs alignements

De façon identique à l'étape 1, l'application permet de récupérer automatiquement les 10 meilleurs formes par fonction. Ces données sont enregistrées sur le fichier *« Param_ultimate_model_detect_MARSEC.txt »* qui servira de base dans les étapes suivantes (ultimate_model , et détection des fonctions).

Pour connaître les meilleurs alignements, nous avons utilisé les tableaux croisés de EXCEL. Auparavant un script Perl (transaction numéro 16 au menu) permet de sélectionner plusieurs modèles afin de pouvoir les comparer, et ajoute quelques informations nécessaires à l'analyse statistique.

Les tableaux résultats ont été décrits dans cette thèse.

Étape 3 : « ultimate_model »

Cette étape fonctionne comme la précédente, mais elle s'appuie sur un nombre de formes encore plus restreint et des paramètres d'alignement plus resserrés.

Ci-dessous les paramètres utilisés pour faire varier l'alignement dans le modèle 11, dans un TU en position 1 :

model	1	TU	firstsyll%	0	70	3
model	1	TU	end	-60	-10	2

Extrait de Param_ultimate_model_detect_MARSEC.txt : les types d'alignement utilisés par le modèle11, pour les TU en position 1.

La détection des fonctions et des frontières

Cette procédure a été largement décrite dans cette thèse dans le chapitre 7.
On travaille sur un intervalle situé entre 2 pauses notées « # » dans le TextGrid ➔ segments.

Traitement général pour chaque segment

Positionnement des frontières possibles

et pour chaque combinaison de frontières :

- Décomposition en pieds ➔ foot

  Et pour chaque pied :

  - Détermination des fonctions possibles pour chaque pied ➔ function
  - Recherche de la fonction la plus proche de la courbe MOMEL
  - Conserve la séquence d'INTSINT la meilleure

- Calcul une note au niveau de chaque sous-segment ainsi défini
- et fait la moyenne pour avoir une note au niveau du segment
- si la note au niveau du segment est meilleure que la précédente on conserve toutes les meilleures séquences d'INTSINT de chaque pied qui constitue ce segment

La combinaison de frontière qui a obtenu la meilleure note a gagné, les résultats correspondants sont stockés dans le TextGrid.

En final, on ajoute dans le TextGrid les résultats du calcul de la meilleure combinaison dans les rangées BDX, intBDx, FUNBDX et CFUNBDX où x correspond au numéro de modèle, soit :

BDx ➔ intervalles délimités par les frontières optimum, contient le score obtenu au niveau du sous-segment

IntBDx ➔ les points INTSINT retenus

FUNBDx ➔ les fonctions retenues pour chaque pied

CFUNBDx ➔ le score obtenu pour chaque pied

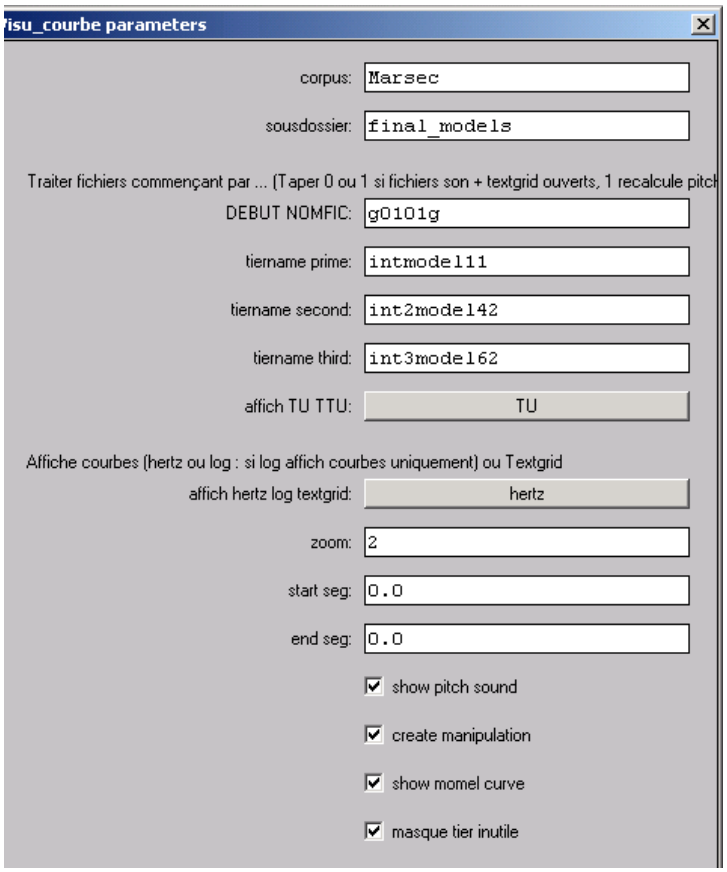
Pur le modèle 11 on crée donc BD11, intBD11, FUNBD11, CFUNBD11.

Résultats du calcul

Les résultats intermédiaires du calcul sont enregistrés dans les fichiers :

- FE_detect_boundaries_pause_x.txt
- FE_detect_boundaries_pause_x_trace.txt
- FE_detect_boundaries_pause_x_debug.txt
- (x correspond au modèle traité, et FE au locuteur sélectionné)

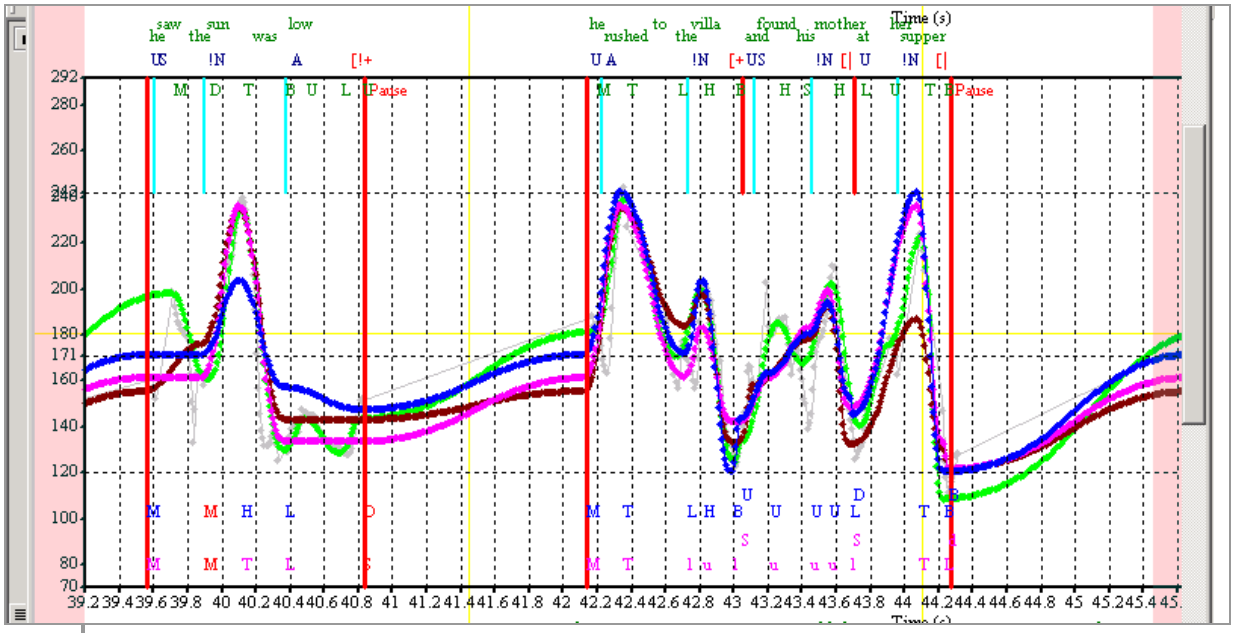
3.5. Outil pour la visualisation et l'écoute des résultats de la modélisation



Description des champs à renseigner :

- Corpus
- Sous-dossier
- Début nomfic : nom du TextGrid à visualiser ou racine : dans ce cas affichage des TextGrids correspondant en série
- Tiername prime : nom de la 1ère rangée à afficher, elle s'affichera en bleu
- Tiername second : nom de la 2ème rangée à afficher, elle s'affichera en magenta (facultatif)
- Tiername third : nom de la 3ème rangée à afficher, elle s'affichera en marron (facultatif)
- Affich TU TTU : permet de choisir le type d'unité tonale à afficher : TU, TTU1, TTU2
- Affich hertz tetgrid : permet de choisir le mode d'affichage : soit courbes uniquement sur page entière, soit courbe + TextGrid dans petite fenêtre
- Zoom : permet de choisir la précision désirée

- Start-seg : début de section de courbe à visualiser en secondes
- End-seg : fin de la section à visualiser en secondes (0 indique que on veut visualiser le fichier en entier)
- Show pitch sound : pour visualiser la courbe F0 d'origine en gris
- Show momel curve : pour afficher la courbe de momel en vert
- Create manipulation : pour générer les objets manipulation
- Masque tier inutile : pour ne conserver dans le textgrid que les rangées correspondant aux modèles demandés : ATTENTION de ne pas enregistrer alors le TextGrid!



Extrait de la fenêtre « picture » de Praat où s'affichent les courbes demandées

Axe des x : les unités de temps en secondes

Axe des y : les valeurs de F0 en hertz

Tout en haut, les mots alignés dans le temps avec les courbes

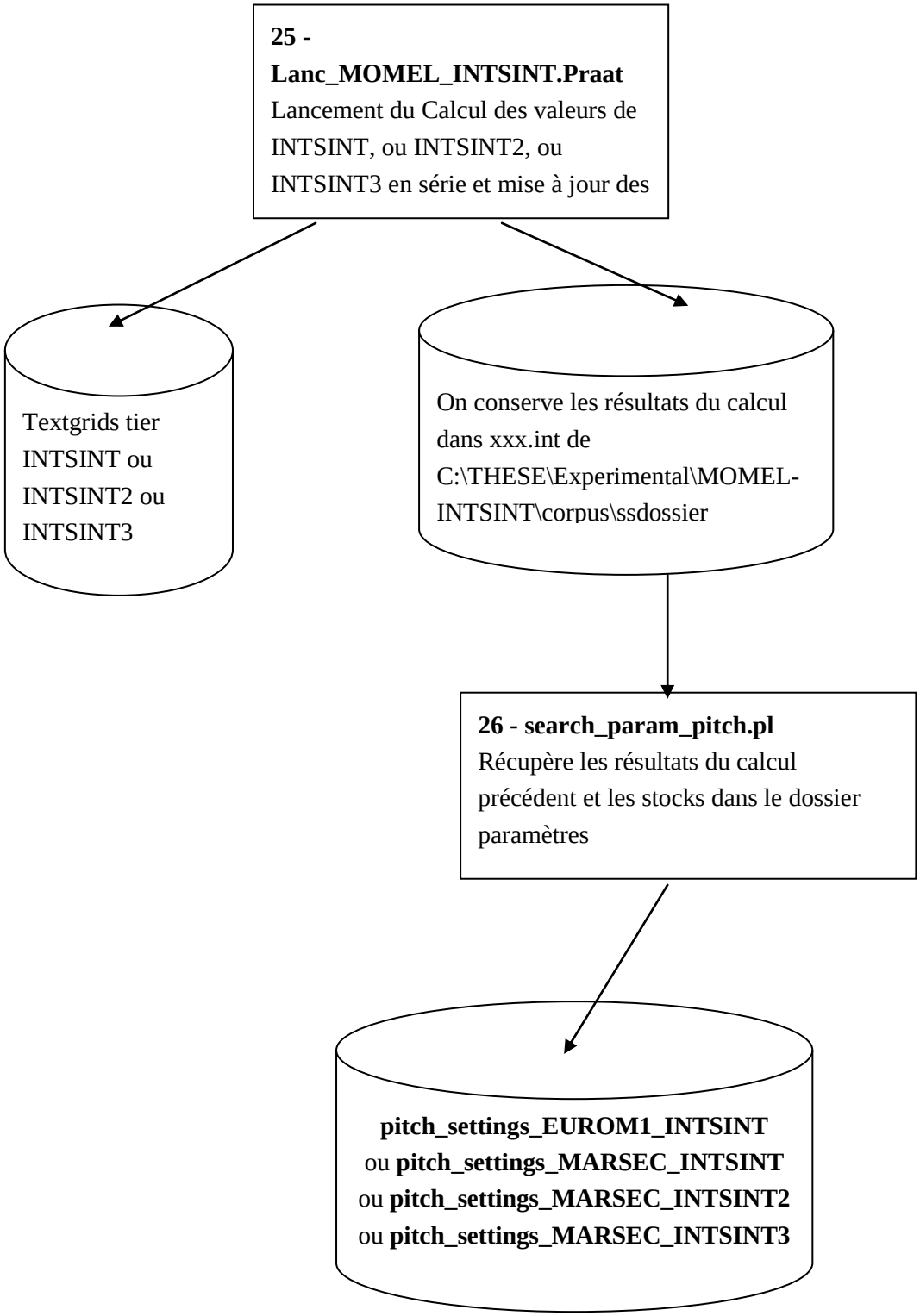
En rouge, le découpage en IU et le type d'IU annoté

En cyan, le découpage en TU et le type de TU annoté

En bas, les valeurs de INTSINT pour les courbes 1 et 2

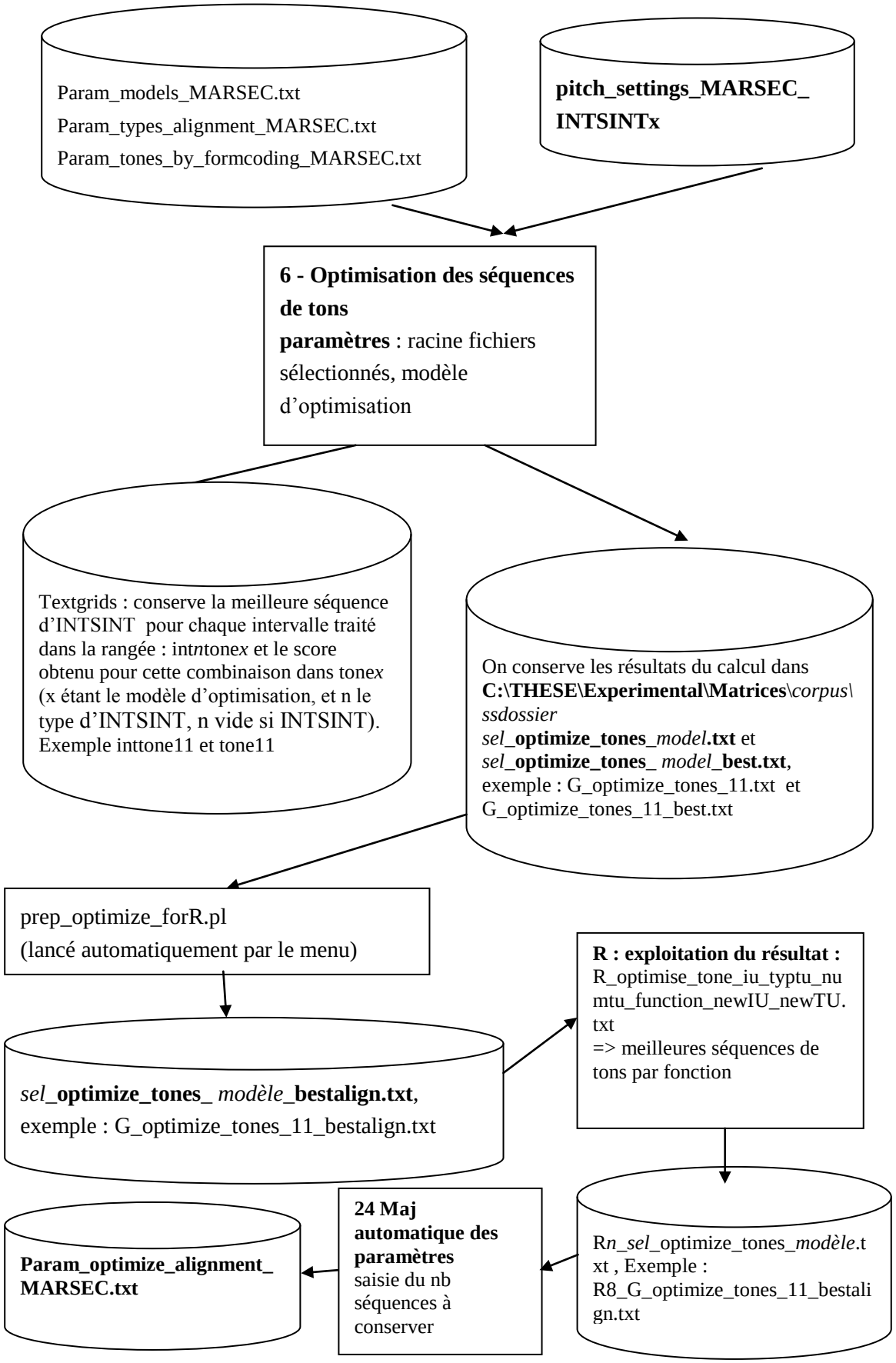
Annexe 4. ***: Schéma récapitulatif de l'applicationn***

4.1. Calcul des valeurs d'INTSINT et des paramètres « key » et « range »

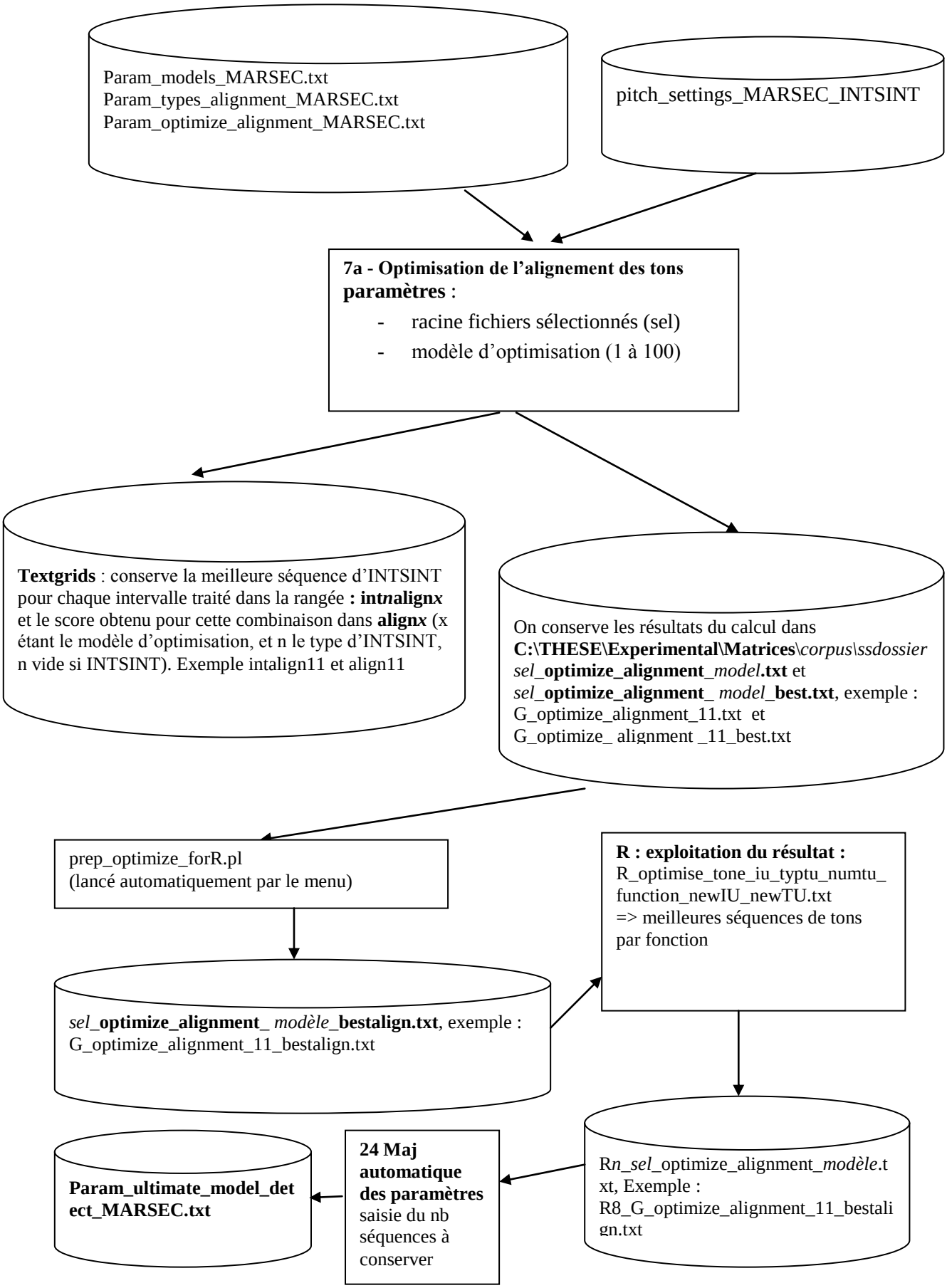


4.2. Optimisation des formes / détection des frontières et des fonctions

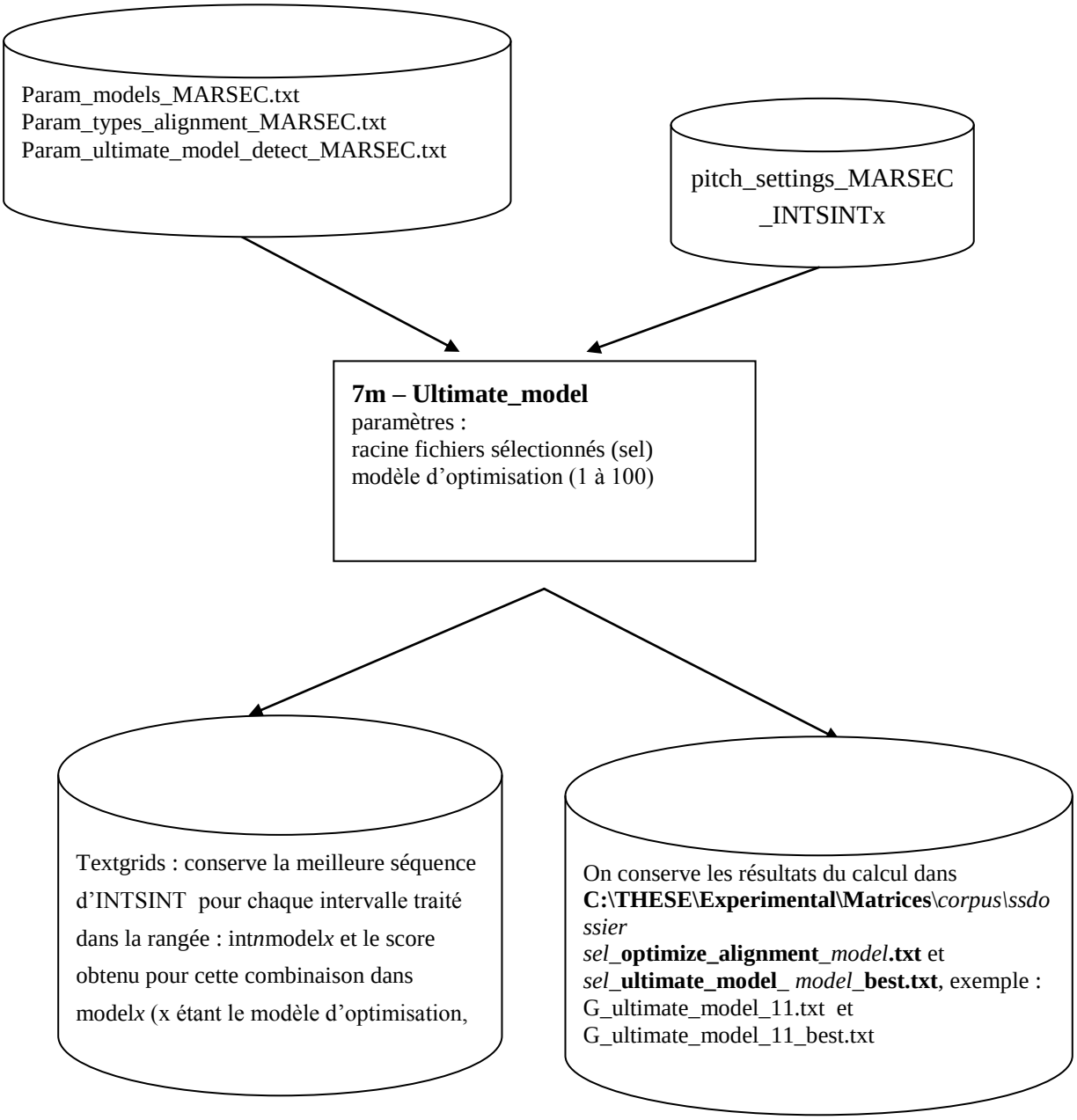
Optimisation des séquences de tons



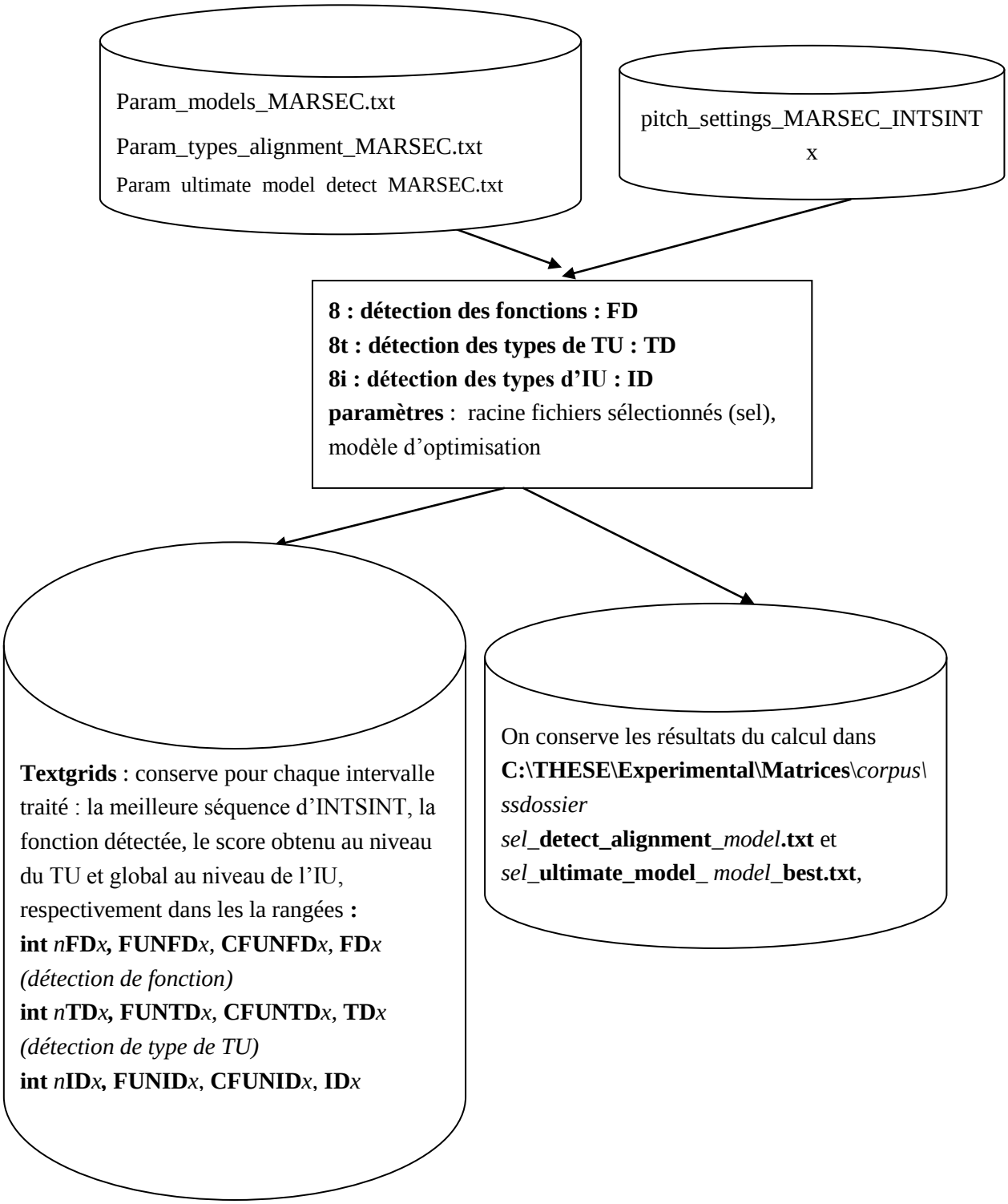
Optimisation de l'alignement des tons



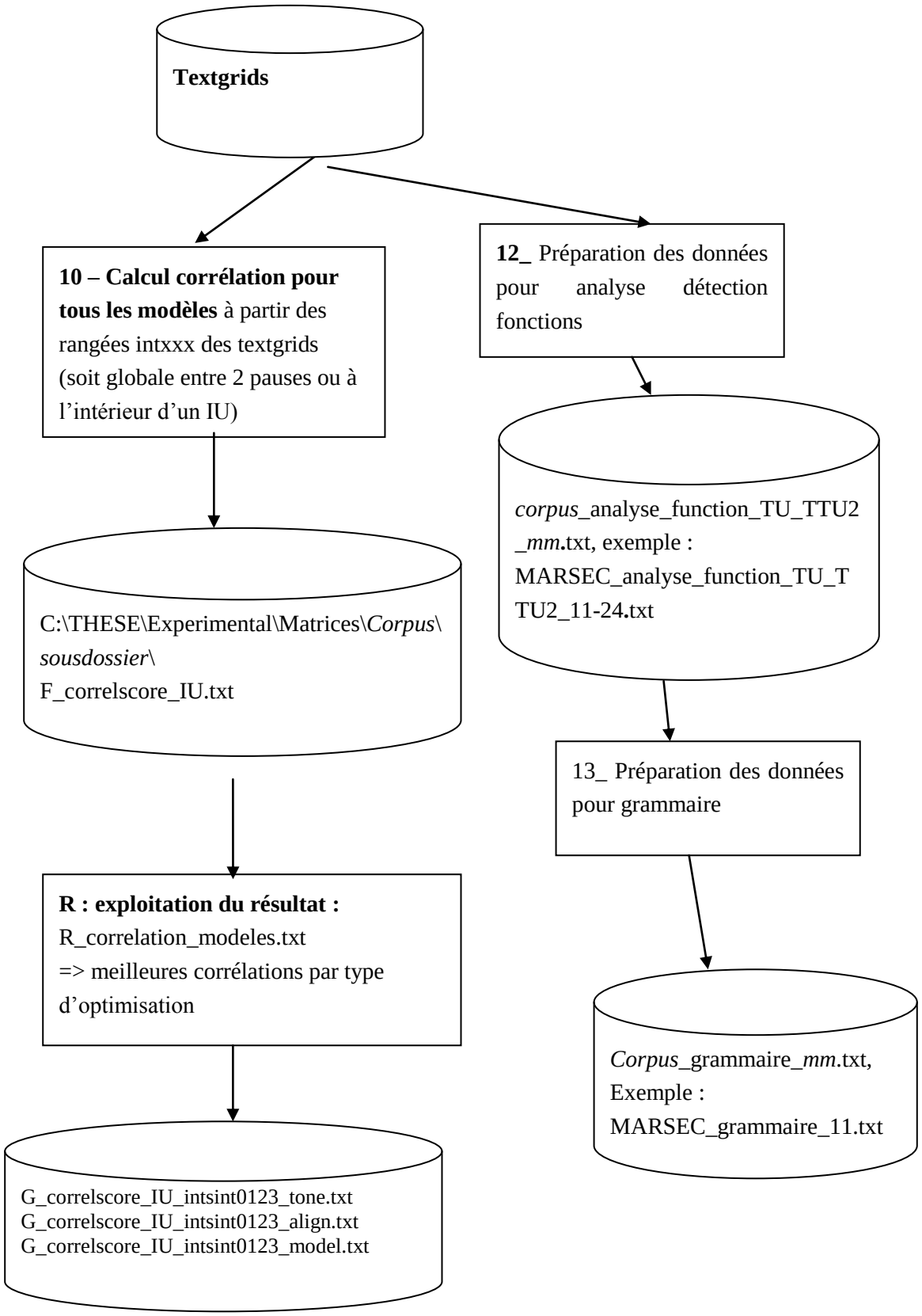
Ultimate_model



Détection des types de TU et/ou IU



4.3.Préparation des données pour les statistiques



ANNEXES CD-ROM

Annexe 5. ***: Annexes sons***

5.1. Extraits du corpus EUROM1 : locuteurs fa, fe, ff, fg

5.2. Extraits du corpus MARSEC : locuteurs G

5.3. Exemples de résultats de la resynthèse : premiers essais avec les « modèles de synthèse »

5.4. Exemples de résultats de la resynthèse : modèles EUROM1

5.5. Exemples de résultats de la resynthèse : modèles MARSEC

Annexe 6. ***: Scripts***

6.1. Scripts de calcul de l'étendue, de la clé et des valeurs d'INTSINT

6.2. Scripts divers d'optimisations des tons, de l'alignement et détection des fonctions

6.3. Scripts divers de préparation des données pour statistiques

RÉSUMÉS

Summary

One of the fundamental aims of prosodic analysis is to achieve a better understanding of what prosody contributes to meaning (prosodic function) and of how this contribution manifests itself phonetically (prosodic form).

Not only is there no consensus on what this contribution is but also there does not seem to be any unified agreement on how it should be represented. Prosodic forms can be modelled accurately by algorithms however a human input is still needed when it comes to analyzing and interpreting intonation meaning. What type of meaning is intonation responsible for? How many functional labels are needed to account for F0 patterns?

It is argued that some meaningful insights could be brought about with a clear separation of the annotation of prosodic forms and functions and a procedure of analysis by synthesis. This could enable to test and enrich different models of English intonation on the one hand and work towards an automatic version of the annotation process on the other.

This procedure is first outlined by using an extract of the EUROM1 corpus and a manual functional annotation with IF (Hirst 1977, 2005) which annotates levels of prominence (unaccented, accented, nuclear and emphatic nucleus) and boundary types (terminal *vs* non terminal boundaries). The representation of prosodic forms is optimized starting from these functional labels using the INTSINT coding system. The sequences of INTSINT tones are then converted into phonetic representations by means of the MOMEL algorithm so that the output can be compared directly to the original recordings. The quality of fit of the model is measured by linear correlation and RMSE with hand corrected modelled fundamental frequency curves, intonation unit by intonation unit.

This procedure is then further developed using 47 minutes of a large corpus of British English (the Aix-MARSEC corpus) for which the functional annotation was carried out automatically through a conversion of the TSM into IF. This allows us to enrich the IF functional annotation system and address theoretical issues such as the number of prominence levels to be taken into account, the nature of the nucleus, the best domain to analyse intonation or how to model tonal alignment.

Key words: intonation, analysis by synthesis, prosody modelling and generation, annotation systems, forms, functions, British English, MOMEL-INTSINT, IF, alignment.

Résumé

Cette thèse présente une étude expérimentale de la relation entre l'annotation des formes et fonctions prosodiques en anglais britannique. En effet l'un des objectifs fondamentaux de l'analyse et de la recherche sur la prosodie est de permettre une meilleure compréhension de la façon dont la prosodie contribue à l'interprétation des énoncés. Bien que l'on reconnaisse à l'intonation un rôle très important dans la communication orale, il n'y a pas à notre connaissance de consensus quant à la façon dont celle-ci contribue au sens ni à la façon de représenter cette contribution dans un système d'annotation.

Cette thèse défend l'idée que l'on peut proposer une meilleure analyse de l'articulation forme/fonction en utilisant un système d'annotation distinct pour chaque type d'information ainsi qu'une procédure expérimentale d'analyse par synthèse. Nous pensons qu'une telle approche pourrait permettre de tester plusieurs théories de l'intonation, d'enrichir l'annotation fonctionnelle de l'intonation et de se diriger vers une automatisation du processus d'annotation.

Cette méthode de recherche a d'abord été mise en place à partir de l'étude d'un corpus de petit calibre : un extrait du corpus EUROM1. Pour ce corpus, l'annotation fonctionnelle a été réalisée manuellement avec le système IF (« Intonation Functions », « Intonation Features », Hirst 1977, 2005). Ce système annote des niveaux de proéminence (inaccentué, accentué, nucléaire et emphatique) et deux catégories de frontières (terminale *vs* non terminale). Les représentations formelles ont été modélisées à partir de cette annotation fonctionnelle à l'aide du système de codage INTSINT et ensuite de l'algorithme MOMEL. Ce système de représentation a notamment l'avantage d'être réversible : les séquences de tons INTSINT sont converties en points cibles MOMEL que l'on peut ensuite resynthétiser afin de comparer le résultat du modèle avec les enregistrements d'origine. Cela permet de réaliser à la fois une évaluation objective et subjective du modèle proposé.

Cette base méthodologique a ensuite été utilisée pour approfondir notre travail de recherche à partir d'un grand corpus d'anglais britannique : le corpus Aix-MARSEC. L'annotation fonctionnelle a été obtenue automatiquement par une conversion des TSM (« Tonetic Stress Marks ») en IF. Cela nous a permis d'enrichir l'annotation fonctionnelle de départ et d'essayer d'apporter des réponses à plusieurs questions théoriques comme la question du nombre de proéminences nécessaires pour rendre compte des variations de F0, la question de la nature du nucléus, celle de l'unité prosodique optimale ou bien à des questions d'ordre formel comme la modélisation de l'alignement tonal.

Mots clés : intonation, articulation forme/fonction, analyse par synthèse, modélisation et génération de la prosodie, annotation prosodique, anglais britannique, MOMEL-INTSINT, IF, alignement tonal.