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

N° attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

THÈSE
pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ AIX-MARSEILLE I
présentée et soutenue publiquement
le 29 septembre 2006
par

Irina Nesterenko

Analyse formelle et implémentation phonétique de l'intonation du parler russe spontané en vue d'une application à la synthèse vocale

Sous la direction de
Monsieur Albert Di Cristo,
Professeur à l'Université de Provence,
et
Monsieur Pavel A.Skreline,
Professeur à l'Université d'État de Saint-Pétersbourg

JURY

Pilar Prieto, INCREA Research Professor, Universitat Autònoma de Barcelona
Nina B. Volskaja, Professeur associé au Département de Phonétique, Université d'Etat de Saint-Pétersbourg
Monsieur Albert Di Cristo, Professeur à l'Université de Provence
Daniel Hirst, Directeur de recherche CNRS à l'Université de Provence
Pavel A. Skreline, Professeur à l'Université d'État de Saint-Pétersbourg
Marc Swerts, Professeur à l'Université de Tilburg

À mes parents et mes grands-parents,
à Benoît

REMERCIEMENTS

J'ai bénéficié d'une aide inestimable de nombreuses personnes tout au long de ce travail de thèse et je tiens à les remercier du fond de cœur.

Mes remerciements les plus chaleureux vont en tout premier lieu à *mes directeurs de thèse*. Je les remercie pour avoir partagé leurs connaissances lors de nos nombreuses discussions, pour leur contribution inestimable à ce travail et à mon développement en tant que linguiste.

Je remercie les chercheurs qui m'ont fait l'honneur d'accepter de faire partie de mon jury de thèse. Je remercie tout particulièrement *Marc Swerts* et *Pilar Prieto* qui ont aimablement accepté de rapporter ma thèse.

Je remercie également *mes enseignants de l'Université d'État de Saint-Pétersbourg* qui ont assuré ma formation initiale dans le domaine de la linguistique et de la phonétique. Je remercie tout particulièrement *Lya Vassilievna Bondarko,* mon premier professeur de phonétique, dont les cours ont incité en moi un intérêt qui n'a fait que grandir dans ce domaine.

Dès ma venue en l'an 2000 pour mon DEA, j'ai reçu un accueil chaleureux au sein du *Laboratoire Parole et Langage* et j'ai profité de son immense potentiel scientifique. Je remercie donc tous les membres du laboratoire qui, chacun à leur façon m'ont apporté aide et soutien. Je ne saurais jamais oublier que cette thèse est le fruit de toute une série de nos rencontres enrichissantes. Je me limite à ces remerciements un peu généraux par peur d'oublier quelqu'un. Je tiens à remercier tout spécialement

Daniel Hirst pour ses discussions rassurantes dans les moments de doutes, pour ses conseils et ses encouragements.

Martine Faraco pour sa gentillesse, son soutien, sa patience, son amitié et pour ses relectures inestimables.

Isabelle Marlien pour ses relectures inestimables et pour ses encouragements tout au long de ces années.

Un grand merci ira à *Robert Espesser* pour son aide constante avec les analyses statistiques, pour ses conseils précieux et judicieux et pour des discussions passionnantes sur le fonctionnement de la langue française.

Merci à *Stéphane Rauzy* pour son aide et ses explications simples de procédés compliqués.

Merci à *tous mes amis du labo*, et à *des collègues doctorants*, d'ici et d'ailleurs, à qui je souhaite un parcours lisse et une grande réussite dans leur futur.

Une pensée pour *les participants* à mes expériences sans qui cette thèse ne serait pas.

Merci à mes collègues du *Département des Langues Slaves* pour m'avoir permis de finir cette thèse dans des meilleures conditions.

Un immense merci à mon proche entourage et à mes amis qui ont tant compté pendant ces années-là.

Merci à *mes parents*, pour tout leur amour et tout leur soutien, pour m'avoir toujours laissée libre dans mes choix de vie. Ce n'est pas facile de vivre des kilomètres qui nous séparent et je ne saurais jamais vous remercier suffisamment, mais j'ai toujours une pensée pour vous dans mon cœur.

Merci à *ma grand-mère* pour tout l'amour qu'elle me donne depuis mon enfance.

Merci à *Geneviève et Jean-Marie Gibergues* pour leur accueil chaleureux et leurs encouragements.

Et surtout un immense merci à *Benoît* qui a toujours su trouver des mots et des gestes salutaires et qui était présent dans les moments les plus difficiles. Tu as beaucoup de qualités que j'admire et j'apprends beaucoup de toi sur cette vie. Merci pour ton amour, ta gentillesse et ta présence. Je te dois beaucoup.

Enfin, j'ai une pensée toute particulière pour *tous les êtres chers qui m'ont quittée*.

TABLE DES MATIÈRES

TABLES DES ILLUSTRATIONS 13

INTRODUCTION 21

PARTIE I. ASPECTS THÉORIQUES 31

CHAPITRE I. CONCEPTIONS PROSODIQUES, UN ÉTAT DE LA QUESTION 33

1.1. CONCEPTIONS DE LA PROSODIE 36
1.1.1. Une distinction terminologique conceptuelle importante : proéminence mélodique (accent) et accent (stress) 39
1.2. LES NIVEAUX DE REPRÉSENTATION DES PHÉNOMÈNES PROSODIQUES 42
1.3. PLAN DU CONTENU DES UNITÉS INTONATIVES : FORME, FONCTION ET SIGNIFICATION 47
1.4. PROLÉGOMÈNES À LA PHONOLOGIE INTONATIVE 52
1.4.1. Problématique de l'unité minimale de description de l'organisation tonale. 53
1.4.2. Les modèles linéaires versus superpositionnels de l'intonation 57
1.5. APPROCHE AUTOSEGMENTALE-MÉTRIQUE DE L'INTONATION (PHONOLOGIE INTONATIVE) 59
1.6. DISCUSSION 60

CHAPITRE II. SYSTÈMES D'ANNOTATION DE L'INTONATION 63

2.1. INTRODUCTION 65
2.2. APPROCHE IPO 67
2.2.1. Procédure analytique 67
2.2.2. Grammaire intonative de l'approche IPO 69
2.3. TOBI : TONES AND BREAKS INDICES 71
2.3.1. Modèle de J. Pierrehumbert (1980 et postérieurs) 72
2.3.2. Association et alignement dans le modèle de Pierrehumbert 74
2.3.3. Annotation ToBI 76
2.3.4. Annotation du degré de jonction (indices de frontières) 77
2.4. TODI : TONAL DESCRIPTION OF DUTCH INTONATION (GUSSENHOVEN, 2005) 80
2.4.1. Tons de frontières 81
2.4.2. Les accents mélodiques 81

2.5. IViE : INTONATIONAL VARIATIONS IN (BRITISH) ENGLISH ... 83
2.5.1. Procédure analytique ... 83
2.5.2. Palier rythmique ... 84
2.5.3. Analyse phonétique des proéminences mélodiques ... 85
2.5.4. Annotation phonologique de IViE ... 86
2.6. TILT (TAYLOR, 2000) ... 88
2.6.1. Unité d'analyse ... 88
2.6.2. Paramètres descriptifs ... 89
2.6.3. A-t-on besoin d'une représentation phonologique ? ... 91
2.7. INTSINT (INTERNATIONAL TRANSCRIPTION SYSTEM FOR INTONATION, HIRST ET DI CRISTO 1998) ... 92
2.8. RÉSUMÉ ET DISCUSSION ... 96

CHAPITRE III. ANALYSES DE L'ORGANISATION PROSODIQUE DU RUSSE : ÉTAT DE LA QUESTION ... 99

3.1. INTRODUCTION ... 101
3.2. LES APPROCHES FONCTIONNELLES DE L'INTONATION DU RUSSE ... 103
3.3. LES DESCRIPTIONS FORMELLES DE L'INTONATION ... 108
3.3.1. Travaux de C. Boyanus ... 109
3.3.2. Modèle de Bryzgounova (1977, et postérieurs) ... 109
3.3.3. Analyse perceptive de l'intonation du russe de Cécilia Odé (1989, 1991) ... 112
3.3.4. Système d'annotation formelle de S. Kodzasov (1996,1999) ... 116
3.4. RÉSUMÉ ET DISCUSSION ... 119

CHAPITRE IV. LE PHRASÉ PROSODIQUE ... 123

4.1. INTRODUCTION ... 125
4.2. L'INTERFACE PROSODIE-SYNTAXE DANS LA TRADITION GÉNÉRATIVE LINÉAIRE (SPE) ... 127
4.3. PHONOLOGIE PROSODIQUE : APERÇU FORMEL DE LA PROSODIE ... 129
4.4. LA RÉCURSIVITÉ ET L'HYPOTHÈSE DE L'ÉTAGEMENT STRICT DANS LES THÉORIES PROSODIQUES ... 133
4.5. LES CONSTITUANTS PROSODIQUES AU-DESSOUS DU MOT ... 135
4.5.1. Thèses de la théorie générale ... 135
4.5.2. La constituance prosodique du russe au-dessous du mot ... 138
4.6. LES CONSTITUANTS PROSODIQUES AU-DESSUS DU MOT ... 141
4.6.1. Présentation générale ... 141
4.6.2. Structuration prosodique au-delà de l'énoncé ... 147
4.6.3. Constituance prosodique du russe : les niveaux supérieurs au mot ... 148
4.7. LE MÉCANISME DE LA RESTRUCTURATION ... 150
4.8. LE PHRASÉ PROSODIQUE ET LA THÉORIE DE L'OPTIMALITÉ ... 152
4.9. LES CRITIQUES DE LA PHONOLOGIE PROSODIQUE ... 153
4.10. RÉSUMÉ ET DISCUSSION ... 156

CHAPITRE V. LA MATERIALITÉ DE LA CONSTITUANCE PROSODIQUE ... 159

5.1. LA PHONÉTIQUE DE LA CONSTITUANCE PROSODIQUE ... 161
5.1.1. Études articulatoires ... 162
5.1.2. Les études acoustiques ... 166
5.2. LA RÉALITÉ COGNITIVE DES CONSTITUANTS PROSODIQUES ... 180

5.2.1. Les études neurolinguistiques ... 181
5.2.2. Les études psycholinguistiques ... 182
5.3. RÉSUMÉ ET CONCLUSIONS ... 187

CHAPITRE VI. LES MODÈLES PROSODIQUES EN SYNTHÈSE VOCALE ... 189

6.1. PREMIER AXE TYPOLOGIQUE : MÉTHODES DE GÉNÉRATION DU SIGNAL ... 192
6.1.1. Taille de l'unité dans les systèmes concaténatifs ... 192
6.2. DEUXIÈME AXE TYPOLOGIQUE : PROPRIÉTÉS DE LA FORME D'ENTRÉE ... 194
6.2.1. Systèmes TTS : Text-to-Speech ... 196
6.2.2. Systèmes de synthèse intégrant le module de la génération du dialogue (Concept-to-Speech) ... 200
6.2.3. Le rôle des corpus dans la synthèse de la parole ... 202
6.3. MODÈLES PROSODIQUES EN SYNTHÈSE VOCALE ... 207
6.3.1. Prédiction du phrasé prosodique ... 208
6.3.2. Prédiction de la distribution des proéminences mélodiques ... 215
6.3.3. Modèles de la génération des courbes de f0 ... 216
6.3.4. Méthodologie d'une double sélection des unités (van Santen et al. 2005) ... 223
6.4. RÉSUMÉ ET DISCUSSIONS ... 225

PARTIE II. ANALYSES EMPIRIQUES ... 229

CHAPITRE VII. CHOIX DU CORPUS ET ANNOTATIONS ... 231

7.1. INTRODUCTION ... 233
7.2. CHOIX DU CORPUS ... 234
7.2.1. Travailler sur la parole naturelle ou sur la parole de laboratoire ... 234
7.2.2. Qu'entend-on par style de parole ? ... 235
7.2.3. Styles de parole et spontanéité ... 236
7.2.4. Complémentarité épistémologique des différents types de données ... 237
7.2.5. Comment recueillir de la parole spontanée ? ... 239
7.2.6. Présentation du corpus ... 240
7.3. ANNOTATION DU CORPUS ... 242
7.3.1. Remarques générales ... 242
7.3.2. Traitement préalable du corpus : aspects acoustiques et alignement ... 245
7.3.3. Autres niveaux d'annotation ... 248

CHAPITRE VIII. EXPÉRIENCE EN PERCEPTION SUR LE PHRASÉ PROSODIQUE ... 263

8.1. INTRODUCTION ... 265
8.2. POINTS MÉTHODOLOGIQUES ... 266
8.2.1. Population d'experts versus population de locuteurs naïfs : solution mixte ... 267
8.2.2. Pourquoi une étude perceptive ? (choix de la tâche expérimentale) ... 269
8.2.3. Comment formuler la tâche expérimentale ? ... 270
8.3. PRÉSENTATION DE L'EXPÉRIENCE ... 273
8.3.1. Conditions expérimentales ... 273
8.3.2. Application du procédé de désémantisation dans les travaux phonétiques ... 274

8.3.3. Déroulement du test ... *277*
8.3.4. Stimuli et sujets ... *278*
8.4. EXPLORATION DES RÉSULTATS DE L'EXPÉRIENCE EN PERCEPTION ... 278
8.4.1. Problématique de l'accord entre les auditeurs ... *278*
8.4.2. Évaluation globale : nombre de frontières ... *293*
8.4.3. Détails de l'organisation rythmique et temporelle ... *300*
8.5. CONCLUSIONS ET DISCUSSION ... 309

CHAPITRE IX. CONTRAINTES SUR LE PHRASÉ PROSODIQUE ... 311

9.1. LE CADRE EXPLICATIF DE LA THÉORIE DE L'OPTIMALITÉ ... 314
9.1.1. Études précédentes ... *315*
9.2. MÉTHODOLOGIE ... 320
9.3. INTERPRÉTATION DES RÉSULTATS ... 323
9.3.1. Contraintes d'alignement émanant de l'interface « prosodie ~ syntaxe » ... *323*
9.4.2. Contrainte discursive AlignFocus,R ... *331*
9.4.4. Étude de cas : les pro-phrases ... *334*

CHAPITRE X. MODÉLISATION ACOUSTIQUE DES FRONTIÈRES PERÇUES ... 339

10.1. INTRODUCTION ... 341
10.2. MÉTHODOLOGIE ... 343
10.2.1. Extraction d'indices acoustiques ... *344*
10.2.2. Analyse discriminante ... *350*
10.3. RÉSULTATS ... 352
10.3.1. Phase d'entraînement : corpus de test ... *352*
10.3.2. Évaluation en prédiction ... *358*
10.3.3. Totalité des données ... *365*
10.4. DISCUSSION ET CONCLUSIONS ... 368

CHAPITRE XI. ARTICULATION « FORME ~ FONCTION » : LE CAS DE LA PRIMITIVE FONCTIONNELLE « ACCENTUÉ » ... 373

11.1.1. Sur les approches probabilistes en linguistique ... *376*
11.2. MÉTHODOLOGIE ... 377
11.2.1. Unité d'analyse ... *377*
11.2.2. Outils statistiques ... *378*
11.3. RÉSULTATS : VERS UNE GRAMMAIRE PROBABILISTE DES CONTOURS INTSINT ... 379
11.4. MODÉLISATION QUANTITATIVE DES DONNÉES ... 384
11.4.1. Méthodologie ... *385*
11.4.2. Application de la procédure de normalisation ... *385*
11.4.3. Résultats ... *388*
11.5. RÉSUMÉ ET DISCUSSION ... 392

CONCLUSIONS ET PERSPECTIVES D'AVENIR ... 399

RÉFÉRENCES BIBLIOGRAPHIQUES ... 405

LES ANNEXES ... 435

TABLES DES ILLUSTRATIONS

LISTE DE FIGURES

Figure 1.1 : Courbes de f0 pour un énoncé affirmatif « Они разговаривают » (ils discutent) et une question générale « Они разговаривают ? » 36

Figure 1.2 : Structure prosodique hiérarchique selon Beckman et Edwards (1990) 40

Figure 1.3 : Niveaux de la représentation des faits prosodiques 44

Figure 1.4 : Transcription parallèle d'un énoncé anglais avec ToBI et INTSINT (empruntée à Beckman et al. 2005) 45

Figure 1.5 : Représentation schématique de différents paliers de la phonologie intonative 60

Figure 2.1 : Classification des systèmes d'annotation de l'intonation 66

Figure 2.2 : Grammaire intonative de l'anglais américain (Pierrehumbert, 1980) 74

Figure 2.3 : Grammaire intonative ToBI 77

Figure 2.4 : Grammaire intonative du hollandais selon l'annotation ToDI 83

Figure 2.5 : Paliers de l'annotation IViE 86

Figure 2.6 : Exemples des courbes mélodiques avec différentes valeurs de tilt 90

Figure 2.7 : Comparaison entre les systèmes d'annotation prosodique ToBI et Tilt (empruntée à Taylor 2000) 92

Figure 2.8 : Exemple d'une courbe modélisée par l'application de l'algorithme MOMEL 93

Figure 2.9 : Configurations tonales et leurs annotations INTSINT 94

Figure 2.10 : Représentation hiérarchique de l'organisation tonale pour l'énoncé « It's almost impossible » 95

Figure 3.1 : Cadre discursif général régissant l'implémentation des indices acoustiques selon Fougeron (1989) 108

Figure 3.2 : Quatre tons distinctifs dans l'approche formelle de Boyanus 109

Figure 3.3 : Sept constructions intonatives de Bryzgounova (1977) 111

Figure 4.1 : Correspondances possibles entre différentes constituances prosodiques proposées dans la littérature (tableau inspiré de Shattuck-Hufnagel et Turk 1996 et de Di Cristo 2004) 131

Figure 4.2 : Analyse des clitiques de Selkirk, transgressant le principe de la hiérarchie stricte 134

Figure 4.3 : Traitement des enclitiques chez Selkirk (1995) 142

Figure 4.4 : Hiérarchie prosodique selon Cahn (1999) 147

Figure 5.1 : Structures syntaxiques analysées dans l'étude de Ladd (1988) 178

Figure 5.2 : Modèle du traitement phonologique du langage parlé (LSCP) 182

Figure 6.1 : Schéma d'un système de synthèse vocale prototype 196

Figure 6.2 : Schéma d'un système CTS prototype 201

Figure 6.3 : Représentation schématique de l'approche d'une double sélection d'unités 223

Figure 7.1 : Représentation du continuum des types de données 234

Figure 7.2 : Organisation du continuum stylistique selon trois dimensions 236

Figure 7.3 : Variabilité dans les réalisations de la voyelle /a/, 247

Figure 7.4 : Omission des voyelles atones dans le mot /fpˈɪtʃətlˈenˈɪ/ 249

Figure 7.5 : Exemple d'annotation du degré de proéminence 251

Figure 7.6 : Exemple d'annotation d'une unité interpausale en termes d'unités intonatives (palier 2 : expert français ; palier 3 : experts russes) 252

Figure 7.7 : Exemples d'erreur de détection de MOMEL; panel a : erreur systématique associée avec la configuration de la courbe de f0 ; panel b : erreur due à la présence de voix craquée 255

Figure 7.8 : Histogrammes des fréquences de f0 selon la catégorie INTSINT 258

Figure 7.9 : Histogramme des valeurs de f0 de tous les points-cibles 259

Figure 7.10 : Exemple d'un TextGrid d'annotation du corpus 261

Figure 8.1 : Illustration du procédé de filtrage 275

Figure 8.2 : Illustration de la courbe mélodique sous la première condition expérimentale (version rectotonale, stimulus 23) 276

Figure 8.3 : Format de réponses 277

Figure 8.4 : Exemple de TextGrid sommaire (auditeurs russes, troisième condition expérimentale). Le panel de droite : codage des auditeurs. 282

Figure 8.5 : Prétraitement des données 284

Figure 8.6 : Pourcentage d'accord entre les paires d'auditeurs (auditeurs russes) par condition 289

Figure 8.7 : Pourcentage d'accord entre les paires d'auditeurs (auditeurs francophones) par condition 290

Figure 8.8 : Nombre total des frontières marquées par auditeur (auditeurs russes) 294

Figure 8.9 : Nombre total des frontières marquées par auditeur (auditeurs francophones) 294

Figure 8.10 : Box-plot par locuteur de la longueur des unités prosodiques perçues selon la condition expérimentale et la langue maternelle des sujets (en ordonnée – longueur du syntagme en syllabes). Panel gauche : auditeurs russes ; panel droit : auditeurs français) 302

Figure 8.11 : Répartitions des unités prosodiques perçues en fonction de leur nombre de syllabes. Auditeurs russes 303

Figure 8.12 : Répartitions des unités prosodiques perçues en fonction de leur nombre de syllabes. Auditeurs français 303

Figure 8.13 : TextGrids sommaires des annotations proposées pour le stimulus 12 sous la deuxième condition expérimentale (version désémantisée, préservant les propriétés prosodiques). Panel gauche : auditeurs russes ; panel droit : auditeurs francophones 305

Figure 8.14 : Box-plots par condition des durées des unités prosodiques perçues. Auditeurs russes. Légende : nf – unités non-terminales, f – unités terminales 306

Figure 8.15 : Box-plots par condition des durées des unités prosodiques perçues. Auditeurs francophones. Légende : nf – unités non-terminales, f – unités terminales 307

Figure 9.1 : Une représentation générale de la syntaxe X-barre (X° : projection minimale (tête lexicale), X' : projection intermédaire, ψ : complément de X, ϕ :spécifieur de X) _____ *319*

Figure 9.2 : Exemple de la représentation selon la syntaxe X-barre de la structure syntaxique de l'énoncé (7) (I : catégorie fonctionnelle 'inflexion' ; C : catégorie fonctionnelle 'complémenteur') _____ *321*

Figure 9.3 : Le rôle des contraintes ancrées syntaxe dans la prédiction du phrasé prosodique perçu selon la condition expérimentale (auditeurs francophones : panel gauche, auditeurs russes : panel droit). _ *323*

Figure 9.4 : Le rôle des contraintes ancrées syntaxe pour les frontières consensuelles selon la condition expérimentale (auditeurs francophones : panel gauche, auditeurs russes : panel droit). _____ *324*

Figure 9.5 : Courbe mélodique de l'énoncé « Я думаю, что мне стоит еще этим позаниматься » *332*

Figure 9.6 : Représentation d'une Macro-unité Intonative _____ *334*

Figure 9.7 : Courbe mélodique de « Да, и в принципе, за тебя всё решает инструктор, когда ты с ним едешь » : illustration du cas avec l'absence d'une frontière prosodique après la pro-phrase ___ *336*

Figure 9.8 : Courbe mélodique de « нет, сейчас ничего не смотрю » : illustration du cas une rupture mélodique après la pro-phrase _____ *336*

Figure 10.1 : Histogramme et courbe de densité des valeurs brutes de f0 pour la mesure de la f0 moyenne dans l'empan pré-frontalier. _____ *346*

Figure 10.2 : Histogramme et courbe de densité des valeurs de f0 après la transformation logarithmique (pour la mesure de la f0 moyenne dans l'empan pré-frontalier) _____ *347*

*Figure 10.3 : Amplitude totale*_____ *350*

Figure 10.4 : Distributions des pourcentages des prédictions correctes : _____ *354*

*Figure 10.5 : Distributions des pourcentages des prédictions correctes*_____ *355*

*Figure 10.6 : Distributions des pourcentages des prédictions correctes*_____ *355*

*Figure 10.7 : Exemple de variabilité dans l'annotation du phrasé prosodique par l'expert français (EF), des experts russes (ER), des auditeurs russes et francophones dans l'expérience de perception, et par l'application de la procédure d'annotation automatique*_____*360*

Figure 10.8 : Évaluation de la démarche prédictive _____ *362*

*Figure 10.9 : Segmentation en unités intonatives avec une frontière après le mot porteur de l'accent mélodique*_____ *363*

Figure 10.10 : Exemple d'une surprédiction des frontières prosodiques sous l'effet du rythme variable de la parole spontanée _____ *364*

*Figure 10.11 : Comparaison des annotations par deux modèles acoustiques : cas de dysfluences. (Légende : A1 : modèle équilibré ; A2 : modèle avec la totalité des positions « absence de frontières » incluse dans l'analyse)*_____ *366*

*Figure 10.12 : Comparaison des annotations par deux modèles acoustiques*_____ *367*

Figure 10.13 : Structure syntaxique de l'énoncé « кстати ты смотришь чемпионат мира по футболу » (« à propos, tu regardes le mondial de foot ») _____ *368*

Figure 10.14 : Illustration de l'absence de frontière après une pro-phrase (« Да на каникулах я была в Финляндии и в Швеции », Oui en vacances j'ai visité la Finlande et la Suède) : s'agit-il d'une erreur de prédiction ? _____ *370*

Figure 11.1 : Grammaire INTSINT pour le français (Hirst et al. 1991) _____ *376*

*Figure 11.2 : Courbe mélodique correspondant à la séquence la plus longue de symboles INTSINT*___ *379*

Figure 11.3 : Ébauche d'une grammaire probabiliste INTSINT pour la primitive fonctionnelle « accentué _____ *383*

Figure 11.4 : Localisation relative du pic mélodique exprimée en fraction de la durée absolue de l'unité (mot prosodique) _____ *387*

Figure 11.5 : Contours prototypes des agrégations obtenues 391

Figure 11.6 : Variabilité interne de chaque agrégation 396

Figure 11.7 : Illustration de la variabilité dans l'encodage INTSINT 397

TABLEAUX

Tableau III.1 : Recensement de quelques analyses fonctionnelles de l'intonation du russe 105

Tableau III.2 : Types d'accents mélodiques du russe selon Odé (1989). 113

Tableau III.3 : Accents toniques du russe selon Kodzasov 1996 118

Tableau IV.1 : Hiérarchie des constituants prosodiques 130

Tableau VII.1 : Nombre de symboles utilisés pour chaque phonème par le phonétiseur automatique et par des phonéticiens dans la tâche de la transcription perceptive (emprunté à Bondarko, 2000 :31) 244

Tableau VII.2 : Liste de sept classes de symboles vocaliques utilisées lors d'annotation du corpus au niveau phonémique 248

Tableau VII.3 : Données sur l'intervention d'annotateur lors de la modélisation MOMEL 255

Tableau VII.4 : Résultats de l'évaluation d'algorithme MOMEL-INTSINT 256

Tableau VII.5 : Informations sur les points cibles déplacés 256

Tableau VII.6 : Moyenne et Écart type des valeurs de f0 prédites 257

pour les catégories INTSINT 257

Tableau VII.7 : Dissymétrie et aplatissement selon la catégorie INTSINT 259

Tableau VII.8 : Comparaison des valeurs de points-cibles détectés et des valeurs prédites selon la catégorie INTSINT 260

Tableau VIII.1 : Fréquences moyennes par stimulus 275

Tableau VIII.2 : Caractéristiques temporelles des stimuli 278

Tableau VIII.3 : Degré d'accord et valeur de Kappa 287

Tableau VIII.4 : Pourcentage d'accord entre les paires d'auditeurs 289

Tableau VIII.5 : Coefficients de Kappa 291

Tableau VIII.6 : Effets fixes du coefficient de Kappa en fonction de la condition expérimentale (auditeurs russes) 292

Tableau VIII.7 : Effets fixes du coefficient de Kappa en fonction de la condition expérimentale (auditeurs francophones) 292

Tableau VIII.8 : Tableau de recensement des valeurs de cœfficient de Kappa communiquées dans la littérature 292

Tableau VIII.9. Pourcentage de frontières syllabiques marquées 293

Tableau VIII.10 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes ; intercept = CondR) 295

Tableau VIII.11 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes, intercept = CondD) 296

Tableau VIII.12 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes, intecept = CondR) 296

Tableau VIII.13 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes, intecept = CondD) 296

Tableau VIII.14 : Exemple d'une matrice de correspondances 297

Tableau VIII.15 : Exemple d'une matrice de correspondances 298

Tableau VIII.16 : Pourcentage et pourcentage corrigé pour l'effet de hasard des localisations exactes des frontières à travers les conditions expérimentales 298

Tableau VIII.17 : Pourcentage et pourcentage corrigé pour l'effet de hasard dans une comparaison trans-linguistique 298

Tableau VIII.18 : Différents test chi² qui estiment si les proportions de trois catégories de frontières sont les mêmes entre les conditions comparées 299

Tableau VIII.19 : Différents test chi² qui estiment si les proportions de trois catégories de frontières sont les mêmes entre les conditions comparées (perspective interlangue) 299

Tableau VIII.20 : Données sur la longueur des unités prosodiques perçues 300

Tableau VIII.21 : Pourcentages des unités de longueur spécifiée 304

Tableau VIII.22 : Données de la validation statistique des hypothèses sur la différence des durées entre les unités intonatives terminales versus non-terminales (auditeurs russes) 306

Tableau VIII.23 : Comparaison des durées des syntagmes non-terminaux (auditeurs russes) 307

Tableau VIII.24 : La validation statistique des hypothèses sur l'écart de durée entre les unités intonatives terminales (T) et non-terminales(NT) (auditeurs francophones) 308

Tableau VIII.25 : Comparaison des durées des syntagmes non-terminaux (auditeurs francophones) 308

Tableau VIII.26 : Durée moyenne et écart type par condition des syntagmes prosodiques non-terminaux 309

Tableau IX.1 : Exemple d'un tableau de contraintes. Dans cet exemple, la contrainte A domine la contrainte B 315

Tableau IX.2 : Exemple d'un tableau de contraintes : deux contraintes à prédictions contradictoires 316

Tableau IX.3 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe. Auditeurs russes 325

Tableau IX.4 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe : frontières consensuelles. Auditeurs russes 325

Tableau IX.5 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe. Auditeurs francophones 326

Tableau IX.6 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe : frontières consensuelles. Auditeurs francophones 326

Tableau IX.7 : Évaluation du rôle des contraintes ancrées syntaxe selon la condition expérimentale 328

Tableau IX.8 : Évaluation du rôle des contraintes ancrées syntaxes dans une perspective translinguistique 328

Tableau IX.9 Évaluation du rôle des contraintes ancrées syntaxe selon la condition expérimentale (frontières consensuelles) 329

Tableau IX.10 : Évaluation du rôle des contraintes ancrées syntaxe dans une perspective translinguistique (frontières consensuelles) 330

Tableau IX.11 : Évaluation du rôle des contraintes ancrées syntaxe dans la prédiction des frontières consensuelles versus optionnelles 330

Tableau IX.12 : Perception du phrasé prosodique : distribution des frontières 332

Tableau IX.13 : Influence de la distribution des accents mélodiques sur la perception du phrasé 333

Tableau IX.14 : Influence de la distribution des accents mélodiques sur la perception du phrasé (frontières consensuelles) 333

Tableau IX.15 : AlignFocus,R >> WrapXP, Align-XP,R ____ *334*

Tableau IX.16 : Distribution des forces de frontière perçue après les pro-phrases (nous indiquons entre parenthèses le nombre de cas de déplacement de la frontière d'une syllabe) ____ *335*

Tableau X.1 : Mesures dérivées de f0 ____ *348*

Tableau X.2 : Mise en correspondance des mesures de f0 dérivées avec des phénomènes mélodiques frontaliers, débattus dans la littérature ____ *349*

Tableau X.3 : Paramètres acoustiques significatifs par condition ____ *352*

Tableau X.4 : Matrice d'effectifs pour des frontières prosodiques selon que ces dernières sont attestées dans l'expérience et prédites par le modèle acoustique ____ *353*

Tableau X.5 : Évaluation de la démarche prédictive pour le modèle avec 37 indices acoustiques retenus ____ *353*

Tableau X.6 : Paramètres individuels assurant une meilleure prédiction ____ *356*

Tableau X.7 : Évaluation de la démarche prédictive pour les modèles avec trois paramètres ____ *356*

Tableau X.8 : Lien entre la variable discriminante dérivée et le facteur 'Frontière' ____ *357*

Tableau X.9 : Effectifs de deux catégories « présence » versus « absence d'une frontière prosodique » : analyse en prédiction sur la totalité du corpus ____ *361*

Tableau X.10 : Évaluation de la démarche prédictive pour les modèles avec trois paramètres : totalité des positions « absence de frontières » incluse dans l'analyse ____ *365*

Tableau X.11 : Lien entre la variable discriminante dérivée et le facteur 'Frontière' : totalité des positions « absence de frontières » incluse dans l'analyse ____ *366*

Tableau XI.1 : Fréquences des séquences de différentes tailles ____ *380*

Tableau XI.2 : Distribution des probabilités pour des étiquettes initiales dans la séquence ____ *380*

Tableau XI.3 : Entropie normalisée du système des bigrammes ____ *381*

Tableau XI.4 : Les séquences les plus informatives ____ *382*

Tableau XI.5 : Fréquences d'unités avec le nombre de syllabes désigné ____ *386*

Tableau XI.6 : Distirubtion des fréquences de différentes structures rythmiques dans le corpus ____ *386*

Tableau XI.7 : Validation statistique d'influence des facteurs rythmiques sur la durée du mot ____ *387*

Tableau XI.8 : Position du maximum de f0 en fraction de la durée absolue de l'unité ____ *388*

Tableau XI.9 : Distribution des unités de différentes structures rythmiques entre agrégations résultantes ____ *389*

Tableau XI.10 : Distribution des structures prosodiques de trois syllabes entre les agrégations des contours ascendants ____ *390*

INTRODUCTION

Le don de parole, conçu comme une maîtrise des mécanismes cognitifs, linguistiques et articulatoires complexes, est une des caractéristiques spécifiques de l'espèce humaine. On communique beaucoup dans la vie courante, sans pourtant être conscient de la machinerie interne de ce processus qui assure la transformation d'une suite d'idées et de concepts en signal acoustique intelligible. Cependant, dès les temps les plus reculés, on comprend que les mots ne se forment pas par la simple concaténation des sons, que l'on ne parle pas avec des formes de citation juxtaposées, mais que l'on obtient une séquence intelligible de syllabes grâce à la manipulation consciente et inconsciente des propriétés prosodiques. Un jour, on a décidé d'apprendre à parler aux machines (cf. Wolfgang von Kempelen 1791[1], Euphonia de Joseph Faber 1846), ce qui a demandé d'élucider ce mécanisme caché de la production de la parole ; depuis, le chemin parcouru et les progrès atteints ont été considérables, de nombreuses applications ont été conçues et commercialisées. Néanmoins, la génération de l'intonation reste toujours un maillon faible dans les systèmes de synthèse vocale, et ceci pour différentes raisons.

Premièrement, on cherche souvent à mimer les conséquences sans vraiment maîtriser les causes. La tradition montre que le champ des recherches sur la synthèse vocale a été principalement occupé par les ingénieurs en quête d'applications industrielles. Dans cette perspective, les méthodes scientifiques employées (algorithmie, méthodes stochastiques) ont permis des avancées remarquables. Mais il semble qu'il soit difficile d'aller beaucoup plus loin dans cette voie. Or, les tests d'évaluation des systèmes de synthèse montrent que l'intelligibilité et surtout le naturel de la voix de synthèse sont jugés insuffisants par les experts. Dans cette situation, il appartient maintenant aux linguistes de proposer des solutions nouvelles fondées sur une connaissance approfondie des processus cognitifs et des modalités de traitement mis en œuvre par l'humain dans l'usage de la parole. Notons cependant que décrire la prosodie d'une langue est une tâche complexe, car la prosodie a la particularité d'être à la fois la composante du langage la plus universelle et la plus spécifique à une langue donnée. Pour qu'une entreprise de description de l'organisation prosodique d'une langue

[1] Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine, 1791 (au Deutsches Museum), références empruntées à Julia Hirschberg (http://www1.cs.columbia.edu/~julia/cs4706/syllabus.html).

aboutisse à un résultat satisfaisant, nous jugeons indispensable de se doter d'un cadre d'analyse, au sein duquel les variations prosodiques sont discutées et formalisées. En raison des considérations présentées, nous nous sommes fixé un premier objectif de recherche, à savoir : d'entreprendre une analyse contrainte par un cadre théorique, en l'occurrence, celui de la phonologie intonative et de la phonologie des domaines[2], de réfléchir et de tester, dans une étude expérimentale, en quoi l'adoption de ce cadre peut contribuer à l'amélioration de notre compréhension de l'organisation prosodique du message et par conséquent, à l'amélioration de la qualité de la parole synthétisée.

En second lieu, les recherches antérieures (destinées à améliorer la qualité de la parole synthétisée) se sont fondées principalement sur de la parole lue. En tant qu'initial et initiatique, ce choix est cohérent : en lecture, les variations prosodiques sont contrôlées et restreintes, ce qui permet d'appréhender au mieux les rapports entre les dimensions et les structures prosodique et syntaxique de l'énoncé[3]. Or, même dans ce cadre simplifié, il apparaît que la prédiction de l'emplacement des frontières prosodiques et du degré de saillance des éléments structurels de l'énoncé n'est pas si aisée qu'il y paraît. Qui plus est, on aspire maintenant à des applications fondées sur le style conversationnel (à buts lucratifs comme thérapeutiques, cf. van Santen *et al.* 2003). Pourtant, le transfert direct des connaissances acquises pour le style lu se trouve inadéquat, surtout dans le contexte d'une prise de conscience sur les mécanismes régissant la production de la parole spontanée.

Dans le domaine de la synthèse vocale, un changement de paradigme majeur s'opère dans les années 1990, avec l'introduction par Sagisaka (Takeda *et al.* 1992) de la méthodologie de sélection d'unités : on s'appuie désormais sur un grand corpus annoté pour sélectionner les unités pertinentes et les concaténer en des formes plus grandes (clauses, énoncés, paragraphes etc.)[4]. Le degré de référence à la structure prosodique de l'énoncé dépend des systèmes. Plus récemment, une extension de cette méthodologie a été proposée par van Santen *et al.* (2005) sous la forme de l'approche d'une double sélection d'unités, qui dissocie la quête des unités segmentales de la recherche des propriétés prosodiques visées. Pour notre part, nous voyons dans ce paradigme un outil permettant de mettre en évidence et d'explorer le mécanisme du détail phonétique fin (*fine phonetic detail*, Local 2003) : ce terme désigne une variation subtile dans la structure acoustique de la parole, qui est cependant systématique et régie par la structure linguistique multiniveaux. De ce fait, disposer des corpus dotés d'une annotation réfléchie et argumentée devient le point crucial dans l'application des méthodes de corpus en synthèse.

Les systèmes de synthèse qui ont été proposés pour le russe (Svetozarova 2000, Krivnova 1999) se fondent sur des méthodologies autres que la sélection d'unités : il s'agit de la méthode de concaténation d'unités pré-enregistrées pour le niveau

[2] Ce choix est argumenté dans les chapitres introductifs de notre thèse.

[3] En outre, l'intérêt porté aux technologies de la parole remonte aux années 1970, le temps de la dominance en linguistique de la théorie générative, qui est caractérisée par une attention aiguë vers le niveau de l'organisation syntaxique.

[4] Notons que les caractéristiques prosodiques se retrouvent ensuite manipulées pour convenir à la structure voulue, bien que les manipulations soient moindres.

segmental[5] ; ensuite, les caractéristiques prosodiques des allophones sont déterminées par des règles, qui sont des heuristiques soigneusement construites par des linguistes. Pour notre part, nous ne connaissons aucun système proposé pour le russe utilisant la méthodologie de la sélection d'unités.

Les remarques précédentes nous amènent à la définition de l'objet de notre étude qui concerne l'élaboration du cadre théorique que nous proposons d'appliquer et la discussion des problématiques qui se rapportent à notre recherche.

Objet d'étude

L'objet de notre étude est l'intonation du parler russe spontané. La tradition des études intonatives du russe commence au début du 20e siècle avec les travaux de Vsevolodskyj-Gerngross (1922), Pechkovskij (1928), Karcevskij (1931). À la fois l'intonation syntaxique et l'intonation pragmatique (termes empruntés à Vaissière et Michaud, *à paraître*) reçoivent une attention aiguë des chercheurs. Cependant, les descriptions formelles entreprises dans l'objectif de cerner le système intonatif du russe, sont peu nombreuses (Boyanus 1936, 1955, Bryzgounova 1977, cf. également chapitre 3 de la thèse). Fréquemment, les analyses ne se limitent pas aux observations et aux encodages de la fréquence fondamentale (f0) seule, d'autres indices acoustiques suprasegmentaux entrent en jeu. Simultanément, l'articulation entre le plan du contenu et le plan de l'expression se dispense d'une étape intermédiaire, celle de la représentation phonologique, ce qui engendre une interprétation simplifiée de l'articulation entre la forme et la fonction. Or, dans la perspective de la synthèse avec la méthode de sélection d'unités, la représentation phonologique est indispensable.

Notre travail porte sur la langue russe et est réalisé à partir d'un corpus de parole spontanée (conversation familière d'une locutrice). Outre l'intérêt des applications technologiques pour le style conversationnel, notre choix de fonder ce travail de recherche sur de la parole spontanée est motivé par le constat que des modèles intonatifs proposés auparavant ont été bâtis à partir des observations sur le fonctionnement du système prosodique dans le style lu.

Parallèlement, nous nous positionnons dans la perspective analytique guidée par les données. Dans le contexte d'une confrontation (cf. Lickley *et al.* 2005) entre les méthodes de la phonologie de laboratoire (Pierrehumbert *et al.* 2001, D'Imperio, 2005) et de la linguistique de corpus, que nous, pour notre part, considérons comme complémentaires, nous jugeons souhaitable, à l'étape initiale, de fonder notre étude sur l'observation des usages que font des locuteurs des moyens mis à leur disposition par le système prosodique de la langue. De plus, la méthode de corpus permet de conjuguer les démarches analytiques descendante (*top-down)* et ascendante (*bottom-up)* : une fois les entités prosodiques définies sur la base des critères formels, la tâche est de mettre en évidence des contraintes de haut niveau (contraintes linguistiques), influant leur

[5] Le choix s'effectue dans une base d'unités (allophones, diphones) pré-enregistrées dans des contextes voulus, le matériau linguistique étant construit au préalable par des linguistes : on dispose donc d'un inventaire assez restreint.

formation et la variabilité observée. En d'autres termes, les unités prosodiques formelles doivent être mises en relation avec des unités linguistiques et discursives.

Objectif de travail

L'objectif de notre travail est de recueillir des données et des arguments qui mèneront à une description formelle de l'intonation du parler russe spontané après une observation du fonctionnement du système prosodique du russe à partir d'un corpus d'étude.

Le cadre de référence dans les études intonatives est de nos jours celui de la phonologie intonative (Ladd 1992), qui se construit à la jonction de la phonologie métrique (Liberman et Prince 1977) et de la phonologie autosegmentale (Goldsmith 1976), repensée pour être appliquée aux langues à intonation. À l'heure où plusieurs propositions de formalisation de l'intonation sont disponibles pour les langues romanes (dont le français), l'anglais, l'américain, le hollandais, il se trouve que le russe n'a pas bénéficié d'une telle attention. À notre connaissance, seule la description de Yokoyama (2003) emprunte ce cadre méthodologique : mais le travail n'est pas achevé.

Cependant, nous estimons qu'un système intégral de l'organisation prosodique ne peut être édifié qu'à partir d'une réflexion conjuguée sur l'organisation tonale et sur le phrasé prosodique. Bien qu'elle définisse des primitives plus petites de la description du niveau tonal (accents mélodiques, tons de frontière, autosegments tonals), la phonologie intonative préserve le concept de contour intonatif. Par un contour intonatif, on entend un segment du continuum mélodique aligné avec un segment de parole, et qui, de ce fait, confère à ce dernier une signification particulière. Le problème se pose de la définition du statut du segment de parole porteur de ce contour mélodique : on le nomme « синтагма » (syntagme) en russe, unité intonative, syntagme intonatif, etc. Il est regrettable que la question épineuse de l'annotation du corpus en termes de syntagmes ne soit pas toujours clarifiée dans les études expérimentales (remarquons que le phrasé prosodique au niveau de syntagmes entretient les relations complexes avec la structure sémantique, pragmatique et syntaxique du message). D'autre part, la phonologie des domaines (Nespor & Vogel 1986, Selkirk 1984, Hayes 1989) a établi toute une hiérarchie des constituants qui régissent un ensemble de processus segmentaux (sandhi) ainsi que l'implémentation de différents indices suprasegmentaux (Price et al. 1991, Campbell & Ladd 1991). Cependant, à notre connaissance, des raisonnements et des procédures analytiques de la phonologie de domaines n'ont jamais été appliqués à l'analyse du russe.

Comme nous l'avons argumenté dans le paragraphe précédent, un problème inhérent à l'analyse formelle de l'organisation tonale est celui de la segmentation du corpus en unités porteuses des contours mélodiques. La question est particulièrement épineuse car, dans une perspective de synthèse à partir d'un corpus annoté, la contrainte de la reproductibilité du schéma d'annotation s'impose (critère de validation pour toute théorie linguistique, cf. Isard et Carletta 1995). Par conséquent, le deuxième objectif de notre travail est d'établir des schémas d'emplacement des frontières prosodiques, de décrire les influences des contraintes de différents niveaux (prosodiques, syntaxiques et

discursives) sur le phrasé prosodique et, à travers l'analyse des détails de son implémentation, proposer une méthode de segmentation (semi-)automatique.

Nos analyses sont largement guidées par les données de la perception : nous nous intéressons davantage aux effets produits par le message sur l'auditeur dans des contextes non-ambigus de la communication courante. Nous adhérons complètement à une vision des phénomènes prosodiques comme signalés par des faisceaux d'indices acoustiques. Dans cette perspective, Vaissière (2005) considère que

> *« perception (rather than acoustics) has strong claims to being the best starting-point for the study of intonation, a phenomenon that involves multiple acoustic cues. »*

Nous tenons à présenter, dès à présent, les limites de notre travail. Il apparaîtra en effet que des analyses menées soulèvent un ensemble de questions adjacentes, auxquelles nous n'apporterons pas de réponse dans le cadre du présent travail, mais qui nous serviront de point de départ pour les recherches à venir : nous pensons notamment à la question soulevée des indices acoustiques des frontières des mots, celle du statut du syntagme de la tradition russe par référence aux constituants du syntagme intonatif et du syntagme phonologique (intermédiaire) de la phonologie des domaines ; ou encore, la quête d'un indice viable de la proéminence perceptive du mot en parole spontanée. Nous n'aboutirons pas non plus à un véritable système de synthèse, de même que nous n'effectuerons pas de tests perceptifs pour valider expérimentalement les améliorations que nous proposons. Comme nous l'avons déjà précisé, il n'existe pas de système de synthèse à partir du corpus pour le russe : nous ne sommes pas en mesure d'en proposer un dans le cadre du présent travail. Certes, beaucoup d'études se servent de bases de données MBROLA pour réaliser des tests de perception spécifiés, mais MBROLA n'existe pas pour le russe. Vu le nombre de questions non résolues que l'analyse expérimentale effectuée soulève, nous nous sommes contentée, dans cette étape initiale d'un plus vaste programme que représente notre recherche, de démontrer comment les informations que nous avons mises en évidence pourront contribuer à obtenir en synthèse une intonation fluide, proche de celle de la parole naturelle conversationnelle.

Plan de thèse

Le présent document est organisé en deux parties qui regroupent 11 chapitres. Les réflexions précédentes nous ont permis de définir deux axes principaux de notre travail : formalisation de l'organisation tonale et phrasé prosodique. Par conséquent, après avoir discuté des prémisses de la phonologie intonative et des descriptions développées pour la langue russe, nous discuterons des systèmes de la représentation symbolique de l'intonation proposés auparavant, avant d'aborder les raisonnements de la phonologie des domaines, qui ont guidé de nombreuses études phonétiques ancrées substance. Nos propres analyses empiriques s'articuleront également autour des grands axes définis : une étude expérimentale en perception nous permettra de mener des discussions argumentées sur la reproductibilité des schémas d'annotation du phrasé prosodique, avant de nous pencher sur des contraintes de haut niveau ainsi que sur les détails

acoustiques et phonétiques qui guident la perception du phrasé. Le modèle[6] acoustique du phrasé résultant est ensuite appliqué pour l'annotation automatique du corpus en unités intonatives et nous débatterons des améliorations à apporter. Nous proposons également une méthodologie d'étude de l'articulation entre la forme et la fonction des phénomènes intonatifs dans le cadre d'une grammaire probabiliste et d'une méthode de génération hiérarchique des agrégations des contours.

La première partie est consacrée à l'exposé des fondements théoriques, sur lesquels nous appuyons notre recherche.

Le CHAPITRE I nous permet de définir ce que nous entendons par la dimension prosodique de la parole et la place réservée au niveau tonal (l'intonation). Nous discutons également de la position prise vis-à-vis des problématiques du statut de la représentation formelle et de l'articulation « forme ~ fonction et signification » pour des phénomènes intonatifs. Enfin, nous cernons des postulats sous-jacents de la phonologie intonative.

Dans le domaine de la prosodie, nous ne disposons pas d'un outil de notation aussi performant et universellement accepté que l'Alphabet Phonétique International pour la classification des unités segmentales. Cependant, au fil des années écoulées depuis l'émergence de la phonologie intonative (dans le travail de Pierrehumbert 1980), de nombreux systèmes d'annotation ont été proposés, soit pour le niveau tonal seul, soit conjuguant l'annotation du phrasé et de l'intonation. Nous consacrons le CHAPITRE II à la discussion de certains systèmes d'annotation.

Le CHAPITRE III présente les différentes approches tournées vers la description du système intonatif du russe.

L'étude du phrasé prosodique étant un des objectifs du présent projet, nous présentons, dans le CHAPITRE IV, les hypothèses et les concepts de la phonologie des domaines, et nous exposons un ensemble de réflexions sur la hiérarchie prosodique pour le russe, en nous fondant sur les données de la littérature. Ces réflexions nous permettront d'aborder des problématiques qui pourront faire l'objet de recherches approfondies ultérieurement (comme la question du statut du constituant « pied » en russe, par exemple).

Tout au long de ce travail, nous défendons la position selon laquelle il est indispensable de trouver la raison d'être de tout constituant conçu par la phonologie des domaines dans la matérialité du signal acoustique[7]. De ce fait, nous partageons le point de vue de Shattuck-Hufnagel (2000) qui interprète la structure prosodique comme un cadre de gestion de l'implémentation acoustique du message (aux niveaux tonal et temporel). Le CHAPITRE V, est consacré à la présentation des études menées dans l'objectif de mettre au clair la phonétique du phrasé prosodique ; nous évoquerons également quelques travaux en psycholinguistique, dont le phrasé prosodique a fait l'objet.

[6] Au sens mathématique du terme.
[7] Dans un premier temps, comme le préconisent les pères fondateurs de cette approche, les constituants prosodiques servent des domaines d'applications de règles segmentales.

Le CHAPITRE VI, qui clôt la première partie, établit un inventaire des algorithmes appliqués dans les systèmes de synthèse. Nous réfléchissons aux modèles de l'implémentation de l'organisation prosodique et au rôle des corpus. Nous terminons ce chapitre avec la présentation du modèle de la double sélection d'unités dont les présupposés guideront nos recherches expérimentales.

La deuxième partie expose en cinq chapitres le cheminement des analyses empiriques que nous avons réalisées.

Le CHAPITRE VII propose une description de notre corpus d'études qui commence par une réflexion sur le concept de spontanéité avant de préciser les modalités de recueil de données et de détailler des annotations que nous avons effectuées pour en préparer des traitements postérieurs. En particulier, nous procédons à l'évaluation de la performance de l'algorithme MOMEL-INTSINT sur notre corpus du parler russe spontané.

Le CHAPITRE VIII détaille le protocole expérimental du test de perception à travers lequel nous cherchons à appréhender la complexité du phénomène du phrasé prosodique. Nous avons proposé à 14 auditeurs (7 russophones et 7 francophones) une tâche d'annotation des stimuli tirés du corpus en syntagmes intonatifs, et ceci sous trois conditions expérimentales, avec une variation de la quantité des informations linguistiques et prosodiques disponibles. Nous en proposons de nombreuses analyses et nous commençons par la discussion de la problématique de la concordance dans les réponses d'auditeurs (*alias* de la reproductibilité du schéma sous-jacent) pour ensuite décrire quelques propriétés rythmico-temporelles des unités structurelles perçues.

Nous avons déjà précisé notre position qui demande de conjuguer dans l'analyse des données les effets des approches *top-down* et *bottom-up*. Dans le CHAPITRE IX nous abordons la problématique des contraintes de haut niveau (syntaxiques et discursives) qui influencent le phrasé prosodique. Pour cette analyse, nous avons choisi le cadre explicatif de la théorie de l'optimalité (Prince et Smolensky 1993). Nous concluons par une étude de cas qui porte sur le phrasé prosodique des structures avec des pro-phrases.

Le CHAPITRE X vise à proposer une technique de la segmentation automatique du corpus en unités intonatives. Nous limitons nos analyses à ce que nous désignons comme des frontières consensuelles, c'est-à-dire des frontières annotées par plus de 4 auditeurs dans l'expérience. À partir d'un vaste ensemble d'indices acoustiques, dont nous spécifions le choix et la procédure d'extraction du corpus, nous procédons à une série de tests par analyse discriminante afin de sélectionner des indices assurant une meilleure prédiction du phrasé perçu. Le modèle acoustique le plus performant, fondé sur le corpus du test de perception, sera par la suite appliqué à la totalité des données. Comme nous estimons ne pas disposer d'une annotation du phrasé de référence, nous procédons surtout à une évaluation qualitative de la performance de ce modèle sur la totalité des données.

Le CHAPITRE XI est consacré à l'exploration de l'articulation « forme ~ fonction ». En focalisant notre attention sur la primitive fonctionnelle « accentué », nous avons d'abord cherché une grammaire probabiliste qui spécifierait l'enchaînement des symboles INTSINT dans l'encodage de la courbe de f0. Ensuite, nous avons appliqué une technique

de classement hiérarchique dans l'objectif de définir des contours prototypes, démarche qui nous permet d'observer la variabilité dans l'implémentation phonétique ainsi que dans l'encodage INTSINT des contours d'une même agrégation.

Nous espérons avoir contribué au terme de ce travail à l'avancement du développement d'une description formelle de la prosodie du russe. En abordant la question épineuse du phrasé prosodique, nous pensons que nous avons effectué un parcours théorique de conceptualisation de la phonologie des domaines dans son application à la langue russe et que, d'autre part, nous avons réalisé un ensemble d'analyses empiriques dans le but de spécifier la phonétique et la phonologie des frontières intonatives. Nous avons également ébauché un paradigme analytique objectif pour l'étude de l'articulation « forme ~ fonction » : toutes ces propositions sont destinées à servir de point de départ à des travaux à venir.

PARTIE I

ASPECTS THÉORIQUES

CHAPITRE I

CONCEPTIONS PROSODIQUES, UN ÉTAT DE LA QUESTION

Cette étude ayant pour objectif de proposer un modèle formel de la description de l'intonation du parler russe spontané, nous désirons exposer, dès le début de ce travail, les présupposés théoriques au fondement de notre réflexion. Dans ce premier chapitre, nous nous attachons à définir ce qu'est la dimension prosodique de la parole et la place qu'elle accorde aux phénomènes tonals (section 1.1). Nous nous penchons ensuite sur la lecture possible du terme « formel » : le formel s'oppose, d'une part, à ce qui est physiologique, physique et mesurable, et d'autre part, à ce qui fonctionnel. La description formelle fait référence à la dimension phonologique des faits prosodiques dont la reconnaissance suscite toujours des débats. Nous consacrons donc la section 1.2 à la discussion des raisons d'être du niveau phonologique dans les analyses prosodiques pour nous tourner ensuite vers l'herméneutique du plan du contenu et l'interprétation des significations que l'on attribue aux unités prosodiques (section 1.3). Enfin nous argumentons notre choix de procéder au traitement des données dans le cadre de la phonologie intonative. Nous accordons une attention particulière aux présupposés fondamentaux de la phonologie intonative qui émergent à la jonction des principes émanant de la phonologie autosegmentale et de la phonologie métrique (section 1.4 et 1.5).

Dans le langage parlé, la signification finale d'un énoncé et le sens qu'il véhicule chez l'interlocuteur ne dépendent pas exclusivement du choix des mots et de l'ordre de leur enchaînement dans l'énoncé. C'est l'intonation qui contribue à la finalisation du contenu du message. Par exemple, considérons le contraste classique entre une assertion et une question : en anglais, ce contraste peut être signalé par des moyens grammaticaux[8] (ordre des mots, présence d'un verbe auxiliaire). En russe l'ordre des mots étant (relativement) libre, dans le cas des questions générales (*yes-no question*) c'est l'intonation qui différencie les deux groupes d'énoncés. Comme l'illustre la figure 1.1, les deux énoncés sont accompagnés de contours mélodiques distincts ; ces différences se manifestent a) par la direction du mouvement mélodique proéminent, b)

[8] Certains moyens grammaticaux mentionnés sont cependant facultatifs en parole spontanée.

par l'organisation temporelle, c) par la synchronisation d'événements tonals et de la chaîne des segments.

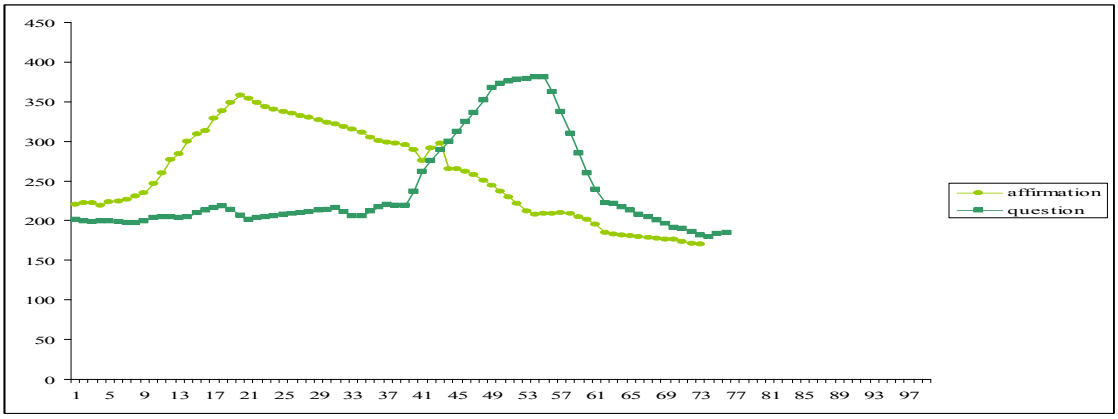


Figure 1.1 : Courbes de f0 pour un énoncé affirmatif « Они разговаривают » (ils discutent) et une question générale « Они разговаривают ? »

L'exemple analysé n'est qu'une illustration du rôle majeur de l'intonation et de la prosodie dans la communication. Dans la lignée de cette réflexion, il nous paraît cohérent au début de notre étude de nous interroger sur la place de l'intonation dans les théories linguistiques modernes. L'intonation faisant partie du domaine de la prosodie, nous nous orientons donc dès maintenant vers la discussion de la conception moderne du domaine prosodique.

1.1. CONCEPTIONS DE LA PROSODIE

Le terme « prosodie » (*prosôdia*) étant d'origine grecque, il désignait à la période de l'Antiquité les traits de la parole absents dans l'orthographe, et en particulier, l'accent mélodique lexical du grec ancien. Plus tard, le même terme est employé en métrique, désignant la structure particulière du vers. Pour notre part, nous fondons nos réflexions sur une définition technique de la prosodie en tant que domaine de recherche complexe en phonétique et phonologie, proposée par A. Di Cristo (2000) :

> *« La prosodie (ou la prosodologie) est une branche de la linguistique consacrée à la description (aspect phonétique) et à la représentation formelle (aspect phonologique) des éléments de l'expression orale tels que les accents, les tons, l'intonation, et la quantité, dont la manifestation concrète, dans la production de la parole, est associée aux variations de la fréquence fondamentale (F0), de la durée et de l'intensité (paramètres prosodiques physiques), ces variations étant perçues par l'auditeur comme des changements de hauteur (ou de mélodie), de longueur et de sonie (paramètres prosodiques subjectifs). »*

Di Cristo voit dans la prosodie un domaine complexe du langage qui doit être abordé dans le cadre de recherches phonétiques (acoustiques), phonologiques et perceptives.

Si nous nous positionnons exclusivement au niveau phonologique de l'analyse des phénomènes prosodiques, différents chercheurs reconnaissent que le terme de prosodie est un terme générique : selon Selkirk (1995a), sous ce terme on entend l'ensemble des

phénomènes en rapport avec trois aspects de l'organisation phonologique de l'énoncé, à savoir l'intonation, le rythme et le phrasé prosodique. Pour Beckman (1996), la prosodie est une structure phonologique par excellence, et l'étude de la prosodie doit comprendre la définition des éléments constituants de cette structure, la formulation des principes régissant la segmentation et l'extraction de ces éléments, ainsi que la définition des relations entre les constituants résultants. C'est la démarche que l'auteur adopte dans son travail, décrivant pour l'anglais et le japonais les catégories prosodiques qui déterminent les patrons intonatifs (en l'occurrence, pour Beckman qui s'associe à l'approche séquentielle de l'intonation, ce sont les proéminences mélodiques et les constituants de la hiérarchie prosodique) ainsi que d'éventuelles ambiguïtés dans cette analyse. Dans la même lignée, Di Cristo (2000) parle du système prosodique d'une langue comme d'un système organisateur complexe ; il y distingue trois ordres structurels, métrique, tonal et temporel, tous trois contribuant à la formation et à la perception de la hiérarchie des constituants prosodiques. Dans cette conception, l'ordre tonal fait référence aux variations de la hauteur mélodique de la voix ; l'ordre métrique décrit l'organisation rythmique d'une langue, avec les patrons propres de récurrence des unités plus saillantes (on parle généralement des syllabes accentuées) ; quant à l'ordre temporel, il rend compte de différents types de pauses et d'allongements syllabiques. Simultanément, dans la description intégrale, une analyse formelle et quantitative est recherchée pour tout niveau défini. Dans cette conception tripartite, la description d'un des aspects séparément des autres aboutira à une vision incomplète du phénomène.

Une distinction universellement admise dans les études prosodiques concerne la séparation entre l'accentuation et l'intonation, les termes (et les phénomènes) qui se réfèrent à deux niveaux de structuration de la parole. La distinction entre les deux phénomènes est faite selon leur domaine d'application respectif. Si l'accentuation est étroitement liée au niveau morpho-lexical et a pour domaine le mot ou le groupe de mots[9], l'intonation, quant à elle, est gérée par des contraintes syntactico-pragmatiques, se référant à un domaine plus large, le syntagme ou l'énoncé.

Remarquons que la dichotomie entre l'intonation et l'accentuation est orthogonale aux trois ordres structurels définis plus haut. En effet, si nous parlons de l'ordre tonal au niveau phonétique le paramètre acoustique clé qui entre en jeu est la fréquence fondamentale (f0) : les variations mélodiques sont associées aux syntagmes (domaine de l'intonation), mais ont pour domaine le mot (domaine d'accentuation) dans les langues à tons. Conjointement, au niveau phonologique, on distingue le ton et la proéminence mélodique.

Il importe de remarquer que, dans certaines études, les termes prosodie et intonation se retrouvent dans les rapports d'interchangeabilité, conception que nous réfutons. Par contre, dans de nombreuses études, c'est l'intonation qui est considérée comme étant le composant principal de la prosodie. Pour notre part, nous insisterons tout au long de cette thèse sur une conception de l'intonation en tant qu'un des niveaux structurels dans le domaine prosodique dont l'étude doit tenir compte du système intégral. Nous sommes persuadée que, dans notre entreprise d'étude formelle de l'intonation, cette dernière doit être analysée à la jonction des cadres théoriques proposés par la phonologie métrique, la

[9] Les options mentionnées étant spécifiques à la langue.

phonologie autosegmentale et la phonologie des domaines[10]. Dans la même lignée, Halliday (1967 : 29-30) parle de trois composants de l'intonation, différenciant *tone* (contour mélodique), *tonicity* (proéminence accentuelle associée) et *tonality* (division de l'énoncé en groupes tonals) :

> *« The set of phonological systems in English that are referred to collectively as "intonation" can be summed up in broad terms as follows:*
> *Tonality marks one kind of unit of language activity, and roughly where each such unit begins and ends: one tone group is as it were one move in a speech act.*
> *Tonicity marks the focal point of each such unit of activity: every move has one (major), or one major and one minor, concentration point, shown by the location of the tonic syllable, the start of the tonic.*
> *Tone marks the kind of activity involved, by a complex pattern built out of a simple opposition between certain and uncertain polarity. If polarity is certain, the pitch of the tonic falls, if uncertain, it rises. »*

Les études intonatives du russe adhèrent à cette conception composite des phénomènes mélodiques (cf. Bruzgounova 1977), or, nous considérons qu'une place moindre est accordée à la réflexion sur le phrasé prosodique.

Les ordres prosodiques structurels retrouvent une manifestation physique à l'oral moyennant un jeu de paramètres acoustiques. Il est considéré pourtant que chacun d'eux possède un indice acoustique marqueur de prédilection : pour l'ordre tonal un tel indice est la courbe de f0. Notons pourtant que la conception actuelle des faits prosodiques diverge de cette position largement répandue. Dès les années 1950, les chercheurs ont démontré l'importance du composant dynamique (l'intensité) (Fry, 1955, 1958; Bolinger, 1958; Lieberman, 1960). Les études récentes (Sluijter and van Heuven 1996; Sluijter et al., 1997; Turk and Sawusch 1996) corroborent cette position, en insistant sur la prise en compte dans les études prosodiques des paramètres d'amplitude ou de tilt spectral. Enfin, la prosodie fait appel à des paramètres qui font varier la qualité de la voix (cf. Laver 1980).

À partir de ces considérations, nous insistons tout au long de ce travail sur l'importance de définir le domaine d'actualisation pour tous les phénomènes prosodiques et en particulier pour les phénomènes intonatifs. Le concept de domaine d'actualisation renvoie au concept du phrasé prosodique et à celui de la hiérarchie des constituants en phonologie des domaines ; en outre, nous y voyons un des axes de réflexion lors de la présentation des travaux sur l'intonation du russe.

Nous avons évoqué dans le présent paragraphe à la fois le phénomène d'accentuation et l'ordre métrique de la structuration prosodique. Or, la réflexion future sur le sujet incite, selon nous, à procéder à un éclaircissement terminologique : il s'agit de la distinction faite en anglais entre deux termes (et deux phénomènes associés) « *stress* » et « *accent* ». C'est ce sujet que nous abordons dans le paragraphe suivant.

[10] Le terme alternatif, plus répandu, étant phonologie prosodique.

1.1.1. Une distinction terminologique conceptuelle importante : proéminence mélodique (accent) et accent (stress)

L'anglais disposant de deux termes, *stress* et *accent*[11], la question terminologique en français ainsi qu'en russe n'est pas aussi évidente qu'il y paraît. Sur le plan conceptuel, la présente dichotomie remonte aux travaux de Beckman (1986). Pour l'aspect terminologique, nous adoptons la distinction proposée par Di Cristo (1999) entre l'accent lexical (*stress*) et la proéminence mélodique (*accent*). Or, nous considérons également nécessaire de nous pencher sur la raison d'être de deux phénomènes, sous-jacente à la problématique terminologique.

La problématique terminologique remonte, en tout premier lieu, à celle des corrélats acoustiques de l'accent. Jensen (2004) propose une analyse historique de l'interprétation du terme *stress* : on constate une dominance de l'interprétation physiologique qui associe l'accent avec un effort délibéré de la part du locuteur en production de parole. Énumérant différentes possibilités, la littérature sur le sujet associe l'accent lexical à une augmentation systématique de la fréquence fondamentale, de la durée et de l'intensité ; on note également des restrictions observées sur la distribution des voyelles relâchées (*lax*) dans les syllabes accentuées. Remarquons également que les différentes hiérarchies des indices acoustiques ont été proposées selon leur pertinence pour le signalement de l'accent lexical et de la proéminence.

Pour Bolinger (1958), l'indice primaire de l'accent est la fréquence fondamentale (en particulier, l'excursion du mouvement de la f0), allant de pair avec la durée ; l'auteur nie toute implication de l'intensité. Ces considérations résultant des analyses acoustiques ont pour conséquence l'introduction par Bolinger du terme « *pitch accent* » (proéminence mélodique au niveau de l'énoncé). Le traitement proposé des indices acoustiques confirme le dispositif théorique de l'auteur pour qui l'accent lexical n'est qu'une entité potentielle au niveau abstrait alors que la proéminence mélodique (*pitch accent*) fait référence à l'actualisation effective de l'accent en parole.

Certains travaux ont tenté de démontrer que d'autres indices acoustiques sont plus pertinents pour caractériser les syllabes accentuées, selon la langue. Slujter et van Heuven (1993) démontrent que, pour caractériser l'accent en néerlandais, il est important de prendre en considération l'intensité. Leur étude établit que l'augmentation spectrale n'est pas régulière et que dans la gestion de l'opposition entre les syllabes accentuées et inaccentuées, la variabilité de l'énergie dans les bandes hautes du spectre (0.5-1, 1-2, 2-4 kHz) est significative.

Une étude récente de Grabe, Koshansky *et al.* (2005) analyse le rôle de différents indices acoustiques pour la prédiction de la distribution des proéminences mélodiques à partir du signal de parole. Le résultat obtenu est assez surprenant : c'est avec l'indice de l'intensité que l'on obtient le score le plus élevé des prédictions correctes. Notons pourtant que l'impact des différents indices acoustiques a été testé séparément dans la

[11] Ou plutôt *pitch accent* : comme le titre de la monographie de M. Beckman (1986) *Pitch and Non-pitch accent* l'indique, la fonction accentuelle étant une fonction universelle, les corrélats acoustiques peuvent y être différents et présupposer une stratification selon le domaine d'actualisation. Cette vision se voit reflétée dans l'opposition terminologique discutée.

présente étude : pour notre part, nous considérons qu'une approche composite (qui présuppose la prise en compte simultanée de différents paramètres) est plus en adéquation avec la nature des phénomènes prosodiques.

Dans l'approche traditionnelle de l'accentuation en russe, il est considéré que la syllabe accentuée possède une sonorité plus importante. Comme il a été démontré, la sonorité, le volume sonore, dépend de deux composants : la qualité vocalique et la durée des voyelles ; il est donc plus justifié de caractériser l'accentuation du russe dans le domaine acoustique comme à la fois quantitatif et qualitatif. Le terme de la proéminence mélodique est récent dans la description des phénomènes mélodiques en russe et est souvent utilisé par référence à une proéminence nucléaire. Cependant, Svetozarova (1973) signale les variations mélodiques associées systématiquement avec des syllabes métriquement fortes. Or, se pose la question de savoir dans quelle mesure on définit l'unité porteuse de proéminence mélodique comme étant équivalente à un mot lexical (l'approche adoptée dans l'étude de Odé 1989).

Le débat terminologique dont nous cherchons à présenter des raisons objectives remonte également à la distinction faite entre les processus de l'accentuation et de l'intonation : ainsi dans la conception de Beckman & Edwards (1990 : 155-156), ce sont les domaines de l'accent lexical et de l'accent mélodique qui diffèrent. Ce qui nous séduit dans l'analyse proposée par ces auteurs, c'est leur souci de lier les indices acoustiques susceptibles d'induire la perception d'une proéminence à la structure prosodique hiérarchique :

> « *We consider the prosodic structure of an utterance to be a hierarchical arrangement of various prominence-lending phonological properties* ».

Les auteurs adoptent pour cette structure une représentation métrique parenthétisée que nous reproduisons dans la figure 1.2.

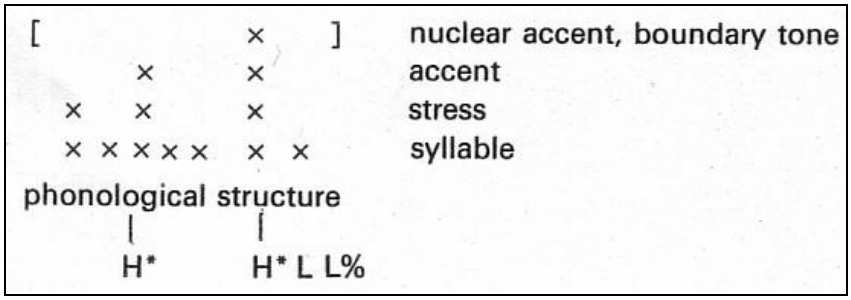


Figure 1.2 : Structure prosodique hiérarchique selon Beckman et Edwards (1990)

Les pics de proéminence à chaque niveau de représentation possèdent des marquages phonético-phonologiques bien documentés. Ainsi, au niveau le plus bas, les syllabes sont définies en relation avec des pics de sonorités ; les accents lexicaux sont associés avec des syllabes comportant des voyelles non réduites, marquées par des allongements et/ou des valeurs d'intensité élevées. Au niveau supérieur, les proéminences mélodiques sont associées avec les mouvements mélodiques induisant la perception d'une saillance particulière (*prominence-lending pitch movments*). Enfin, la juxtaposition de la proéminence nucléaire et de l'accent du syntagme promeut la dernière syllabe proéminente au niveau supérieur de la proéminence, celui de l'accent nucléaire. De

manière générale, l'hypothèse qui sous-tend l'approche de Beckman et Edwards est la suivante : la structure prosodique de l'anglais contient quatre niveaux, un fait dont on doit rendre compte dans la structure de surface ; cet objectif peut être atteint par l'intermédiaire du jeu des indices acoustiques (cf. également le travail de Astesano (2001) sur les propriétés de différents types d'accents). Remarquons que dans ce modèle il ne s'agit pas de différences quantitatives dans la manifestation de différents indices acoustiques, une syllabe porteuse de la proéminence mélodique étant caractérisée par une intensité plus importante que la syllabe porteuse de l'accent lexical. Avec Beckman et Edwards, nous nous associons donc à l'hypothèse que différents niveaux de proéminences possèdent des corrélats acoustiques différents.

La distinction entre l'accent lexical et la proéminence mélodique se manifeste également si on se place dans l'optique de la prédiction de leur position et leur distribution. Dans des langues telles que l'anglais et le russe, la prédiction de l'accent lexical est plus aisée et se fait à l'aide de règles qui prennent en compte la structure morphologique du mot (cf. le concept des morphèmes accentogènes chez Garde 1968, Kodzasov 1999). Dans le même esprit, dans l'analyse formelle de Zaliznjak (1977) le patron accentuel de chaque mot est une de ses propriétés phonologiques (cf. la position similaire de Cutler 1984). Or, une restructuration est possible dans la parole continue où les mots s'enchaînent (Kassevitch, 1999) : dans cette perspective le concept du mot phonologique et phonétique n'est pas si anodin, ce sujet faisant le thème d'une discussion approfondie dans le chapitre IV.

La prédiction de la distribution des proéminences mélodiques suscite beaucoup plus de débats. Le titre de l'article de Bolinger (1972) en est révélateur : « Accents are predictables (if you are a mind-reader)[12] ». En effet, on considère que la distribution des proéminences mélodiques est régie par des contraintes pragmatico-discursives : il n'est donc pas trivial de proposer les modélisations et les algorithmes adéquats. Nous y consacrons un paragraphe dans la discussion des algorithmes mis en place dans les systèmes de synthèse. Par ailleurs, dans le cadre de l'approche autosegmentale de l'intonation, une proéminence mélodique est nécessairement associée avec une syllabe porteuse de l'accent lexical.

Dans la mesure où nous entreprenons une étude formelle de l'intonation dans le cadre qui vient d'être cerné, nous nous sommes posé la question de la validité du niveau de la représentation formelle de l'intonation que nous traitons dans le paragraphe suivant.

[12] Les proéminences mélodiques sont prédictibles (si vous êtes un télépathe).

1.2. LES NIVEAUX DE REPRÉSENTATION DES PHÉNOMÈNES PROSODIQUES

Notre travail ayant pour objectif de proposer une description formelle de l'intonation, il nous semble important de nous pencher sur la problématique des niveaux de représentations des phénomènes prosodiques.

Les termes « REPRÉSENTATION FORMELLE » et « REPRÉSENTATION PHONOLOGIQUE » sont considérés comme des synonymes. La question est de savoir dans quelle mesure l'intonation est phonologique ? En refusant aux phénomènes intonatifs le statut phonologique, on insiste beaucoup sur une association forte entre l'intonation et les aspects extralinguistiques et paralinguistiques de l'activité langagière[13]. Subséquemment, dans la quête d'une description phonologique de l'intonation, une des démarches importantes est de construire son objet d'études, ce qui se traduit par une démarche de conceptualisation de ce qui est linguistique en intonation et ce qui ne l'est pas.

Ladd (1996 : 36), suivant Bolinger (1961), propose de concevoir cette distinction en termes d'opposition entre la structure catégorielle des significations linguistiques discrètes (« ***quantal*** *or* ***categorical*** *structure of linguistic signalling »*) et le caractère graduel du paralangage : les traits linguistiques sont catégoriels, discrets et arbitraires dans leur relation avec leurs fonctions. Dans la même lignée s'inscrit la théorie des codes biologiques proposée par Gussenhoven (2002a, cf. infra).

D'autre part, la longue période de négligence des phénomènes intonatifs dans les études phonologiques vient de l'ambiguïté du statut sémiotique des unités définies par l'intonation. Plusieurs chercheurs considèrent ces unités, les intonèmes, comme étant de vrais signes linguistiques, possédant un plan d'expression conjointement avec un plan de contenu. Dans ces approches, les intonèmes ont donc le statut de morphèmes intonatifs[14]. De ce débat théorique résulte une première classification possible des différentes approches des phénomènes intonatifs : la distinction s'opère entre les modèles phonologiques et les modèles morphologiques. Ainsi les modèles phonologiques de l'intonation (Di Cristo 2004 pour une présentation critique et argumentative) reconnaissent l'autonomie du module prosodique dans la grammaire de la langue. Les modèles morphologiques[15] conçoivent les unités élémentaires de la description comme des signes linguistiques associant une facette « signifiante » (plan de l'expression) et une facette « signifié » (plan du contenu, exprimant une fonction syntaxique, pragmatique ou phonologique), cf. Rossi (1999 : 50) :

> *« Les morphèmes prosodiques ont un contenu déterminé soit par la syntaxe des constituants (frontières intonatives), soit par le module sémantique qui inclut le contenu pragmatique étranger à la syntaxe et à la 'grammaire lexicale', soit par le lexique. »*

[13] On distingue systématiquement les aspects paralinguistiques (affects et émotions notamment) et non-linguistiques (âge du locuteur, origine régionale) dont la prosodie et l'intonation sont le moyen d'expression.
[14] Par opposition aux phonèmes, les unités dotées du plan d'expression, mais n'ayant pas de signification linguistique par elles-mêmes.
[15] L'approche morphologique, quant à elle, est représentée par les travaux de Rossi (1999), de Morel et Danon-Boileau (1998) pour le français, par les modèles fonctionnels de l'intonation du russe (cf. *infra*).

La problématique de la représentation phonologique des faits prosodiques (intonatifs) surgit donc dans le cadre d'une discussion méthodologique visant à savoir s'il est possible de relier les paramètres acoustiques directement aux valeurs sémantiques et pragmatiques, ou bien si l'on a besoin d'un niveau intermédiaire de représentation abstraite. Ladd (1992) caricature l'idée d'une association directe du niveau phonétique à la valeur pragmatique, citant l'exemple des formes de négation dans plusieurs langues : on remarque que dans différentes langues, le plan signifiant de la particule négative contient un segment nasal, or, on n'en conclut pas que la nasalité est associée avec la valeur sémantique de négation. Selon la réponse donnée à la question de la nécessité d'une représentation intermédiaire, on distingue les modèles acoustiques *versus* phonologiques de l'intonation : cette dichotomie traduit l'importance accordée dans chaque modèle à la représentation linguistique par rapport aux observables physiques[16]. Comme le remarquent Hirst *et al*. (2000) :

> *« Although the distinction between abstract linguistic representations and concrete physical representations is fairly uncontroversial, different approaches tend to differ in the relative importance which they attach to each of these two levels. »*

Pour Hirst et ses collègues, le terme de la *représentation linguistique* fait référence aux représentations cognitives que le locuteur se forge afin de s'approprier des informations prosodiques ; d'autre part, la *représentation physique* fait surtout référence aux choix méthodologiques des chercheurs qui analysent les données observables. Les deux représentations peuvent être différenciées sous les termes de représentation cognitive et de représentation analytique (*cognitive* versus *analytical representation*). Pourtant la structure peut être plus complexe que dichotomique : d'autres niveaux intermédiaires peuvent être inclus dans la description, dont le degré d'abstraction est variable. Ainsi Couper-Kuhlen (1986 : 63) discute le modèle à trois niveaux proposé par 't Hart & Collier (1975) :

> *« As 't Hart&Collier (1975) have pointed out, there are three different levels at which intonation can be analyzed, each reflecting a different degree of abstraction. At a concrete,* ***acoustic*** *level intonation can be seen as a succession of fundamental frequency curves in time. Since, however, many of these acoustic phenomena are not perceived at all by the human ear or only selectively perceived, a second,* ***phonetic*** *level must be distinguished, at which intonation can be viewed as a succession of perceivable pitch 'events'. However, not even all the pitch events which are in principle capable of being distinguished by the human ear are necessarily relevant in understanding the utterances of a given language. That is, a third, more abstract* ***phonological*** *level of intonation analysis can be identified at which potentially distinct pitch events are grouped together into 'meaningful' categories. »*

Nous retiendrons de l'analyse de 't Hart et Collier la nature de la distinction faite entre les niveaux phonétique et phonologique, en particulier par rapport au fait que le niveau phonologique contient quant à lui une séquence *discrète* des mouvements mélodiques.

[16] Cf. la différence faite dans l'ouvrage coordonné par Cutler et Ladd (1983) entre les mesureurs et les constructeurs de modèles.

La distinction telle qu'elle est faite évoque par analogie celle qui est faite entre un allophone et un phonème dans la phonologie segmentale.

Appartenant à l'école hollandaise de l'intonation caractérisée par une attention particulière aux données perceptives, 't Hart et Collier définissent les niveaux phonétique et phonologique par référence à la pertinence en perception des mouvements mélodiques. En revaanche, chez Hirst et collègues (Hirst *et al*. 2000), la distinction entre le niveau phonétique et le niveau phonologique est conçue selon le caractère discret ou continu de la représentation adoptée[17]. Dans leur modèle, quatre niveaux sont distingués, avec deux niveaux intermédiaires entre la représentation physique et la représentation cognitive (cf. figure 1.3). Ces deux niveaux intermédiaires sont le niveau phonétique et le niveau phonologique de surface. Le niveau phonétique, quant à lui, correspond à une représentation continue, reflétée par la courbe de f0 après la soustraction de la composante microprosodique qui reflète ce qui est physiologique, issu des contraintes universelles sur la production et la perception de la parole (Di Cristo 1985).

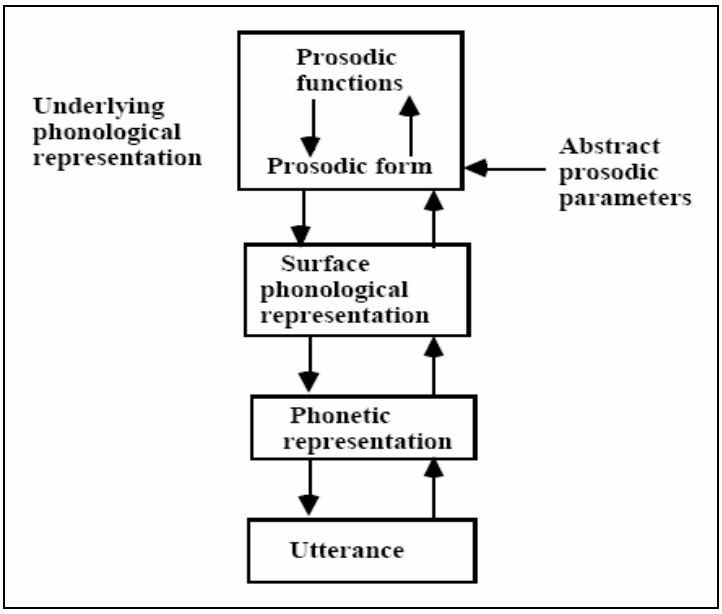


Figure 1.3 : Niveaux de la représentation des faits prosodiques selon Hirst et al. (2000 :55)

La question de reconnaissance du niveau phonologique de l'intonation fait débat à propos des différents systèmes d'annotation prosodique développés. Dans l'approche de Hirst et coll., c'est la reconnaissance du niveau phonologique de surface et sa formalisation en termes d'un alphabet d'annotation prosodique INTSINT que nous considérons comme étant de la plus grande importance. Ce niveau est phonologique car il contient une représentation de la courbe mélodique en termes de points-cibles codés. En revanche, les auteurs du système d'annotation prosodique ToBI (Beckman et Ayers 1997) refusent au système INTSINT le statut de représentation phonologique (Beckman *et al.* 2005) : la représentation fournie par l'annotation INTSINT est tenue pour un encodage symbolique de la courbe mélodique, un procédé de réduction de l'information qui ne parvient pas à atteindre le niveau phonologique, sans être vraiment une représentation phonétique non plus *(« a continuous phonetic representation can never*

[17] Critère adopté à la suite de Troubetzkoj (1939).

be replaced by even the most detailed symbolic encoding of pitch events »). Le système ToBI contient lui aussi plusieurs niveaux de représentation, l'un d'entre eux étant conçu sous la forme de la courbe de f0 brute[18]. Au niveau phonologique, nous retrouvons les étiquettes qui signalent la présence d'un événement mélodique important (proéminence mélodique ou frontière) sans encoder directement le mouvement accompagnant de la courbe de f0[19]. Procédant par une comparaison entre les systèmes d'annotation INTSINT et ToBI à partir d'une transcription parallèle du même matériau linguistique (que nous reproduisons sur la figure 1.4), Beckman et ses collègues remarquent qu'à l'encontre de ToBI, qui atteste la présence d'une séquence d'accents mélodiques descendants, INTSINT n'encode que la séquence des mouvements ascendants et descendants sans attribuer de valeur contrastive à quelque événement que ce soit et sans analyser le lien entre le contour mélodique et le texte.

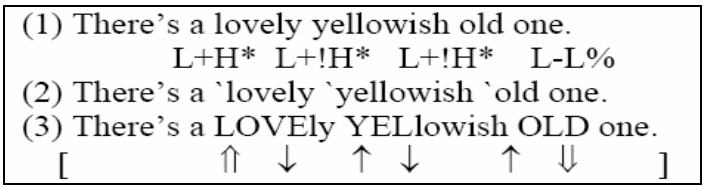


Figure 1.4 : Transcription parallèle d'un énoncé anglais avec ToBI et INTSINT (empruntée à Beckman et al. *2005)*

Pour notre part, nous voyons quelques contradictions dans les critiques de Beckman *et al.* (2005). Les auteurs d' INTSINT (Hirst et Di Cristo 1998) y voyaient un système analogue à la transcription phonétique étroite qui permettrait d'encoder les points importants *(turning points)* de la courbe mélodique. Remarquons que le même principe d'encodage de la courbe de f0 a été adopté dans le travail de Bruce (1977), un des travaux à l'origine de l'approche autosegmentale de l'intonation et donc, précurseur de ToBI. Également une démarche théorique y a été proposée, laquelle relie et rend réversible la représentation cognitive de la structure intonative et l'encodage INTSINT (1998, pour l'application en synthèse Di Cristo et Di Cristo 2001).

Beckman *et al.* (2005) entreprennent dans leur travail une analyse critique de plusieurs systèmes d'annotation prosodique et ceci en relation avec la distinction entre ce qui relève de la phonologie et ce qui relève de la phonétique. Par rapport au système IViE, qui procède par une distinction systématique des niveaux phonétique et phonologique mais qui adopte une représentation discrète[20] pour le niveau phonétique, les critiques sont similaires à celles qui ont été émises pour INTSINT. On y voit les systèmes se limitant à encoder la courbe de f0 en termes de symboles discrets, sans vraiment aboutir à une

[18] En même temps, le système d'annotation ToBI accepte l'existence du niveau phonétique dynamique, qui intervient à l'étape où l'inventaire des étiquettes pertinentes pour la langue n'a pas encore été établi. Ce niveau garde la trace d'une analyse phonologique potentielle : les étiquettes sont discrètes parce qu'elles sont associées à des catégories phonologiques potentielles. Cette approche du niveau phonétique suit la conception proposée par Beckman et Pierrehumbert qui considère la représentation phonétique comme un processus interprétatif dynamique (*dynamic interpretative process*) dont le rôle est de mettre en lien direct une représentation phonologique avec les paramètres acoustiques et physiologiques.

[19] Nous tenons à préciser qu'il s'agit ici du point de vue des auteurs du travail analysé. Pour notre part, nous considérons que les étiquettes ToBI contiennent un encodage partiel de la forme de la courbe de f0 sous la forme de tons abstraits (cf. infra).

[20] Cf. l'analyse *infra*. ('H*+L, for instance, can be realised as hM-l, as mH-l, as hH-l, as mH-l (peak lag) or lH (truncation)').

description phonologique dans la mesure où les deux systèmes d'annotation sont guidés par la courbe de f0. Remarquons que le niveau phonologique est conçu dans la tradition aixoise comme un niveau de la transcription prosodique par intermédiaire d'un ensemble restreint d'étiquettes : ce niveau ne reflète pas les choix linguistiques significatifs spécifiques à la langue en considération (rôle du niveau phonologique sous-jacent) :

> *« The level of deep phonological representation can be thought of as the level of linguistically significant choices which interact with a number of language-specific abstract prosodic parameters to generate observed intonation patterns ».*

Pour les auteurs d' INTSINT la raison d'être de ce niveau de représentation réside dans le besoin de distinguer systématiquement les aspects formel et fonctionnel dans l'analyse prosodique, ce qui n'est pas le cas dans l'annotation ToBI. En effet, dans l'étiquette H*+L ou H*L%, les symboles '*' et '%' encodent les aspects fonctionnels (marquage d'une proéminence et d'une frontière prosodique) tandis que les symboles L et H, dans leur enchaînement, décrivent la forme de la courbe mélodique correspondante[21]. INTSINT parvient à la séparation des deux par l'introduction de deux niveaux de la représentation.

De son côté, le niveau phonologique profond doit posséder d'une architecture lui permettant d'assumer sa double fonction : fournir les informations à la fois pour la prononciation de l'énoncé et pour la reconstruction de son organisation sémantique et syntaxique, cf. :

> *« […] a phonological representation must fulfil two purposes: it must provide both the information necessary for the pronunciation of an utterance and the information necessary for its syntactic and semantic interpretation. »*

Comme le démontre le schéma de la figure 1.3, il résulte de cette double fonctionnalité une organisation complexe du niveau phonologique profond : c'est à ce niveau que s'opère l'interface entre la représentation formelle la plus abstraite et la représentation fonctionnelle des significations prosodiques.

Suivant un processus *bottom-up,* nous avons proposé dans la présente section un ensemble de réflexions qui ont abouti à l'intégration dans le schéma général d'un niveau de représentation phonologique pour les faits prosodiques. Nous avons également évoqué les différentes architectures adoptées par les systèmes d'annotation prosodique. Dans le modèle proposé par Hirst et ses collègues, au niveau phonologique profond l'interface s'opère entre le niveau formel le plus abstrait et le niveau fonctionnel des significations prosodiques. Par ailleurs, la question des sens véhiculés par des formes prosodiques (intonatives) sous-tend de nombreuses descriptions de l'intonation du russe que nous discutons dans le chapitre III. La section suivant est donc consacré à la présentation de la problématique des fonctions et des significations prosodiques.

[21] De plus, l'annotation à l'aide du symbole '*' suscite de nouveau toute la problématique inhérente des concepts « association » et « alignement » dans les descriptions autosegmentales.

1.3. PLAN DU CONTENU DES UNITÉS INTONATIVES : FORME, FONCTION ET SIGNIFICATION

Dans le contexte des réflexions et des approches théoriques présentées dans les sections précédentes nous considérons que les trois aspects, à savoir le phrasé, la distribution des proéminences et le contour mélodique, font partie de la théorie phonologique de la prosodie au niveau de la représentation abstraite. Nous adhérons également au postulat selon lequel les structures prosodiques, et en particulier les structures intonatives, apportent au message des valeurs sémantico-pragmatiques spécifiques ; en d'autres termes, les formes prosodiques sont intrinsèquement associées avec un ensemble de significations par des relations non-bijectives. La question est de savoir quels sens (significations, valeurs pragmatiques) véhiculent les contours mélodiques // les structures prosodiques. Remarquons que ce domaine a reçu moins de systématisation que l'analyse de la partie formelle de l'intonation.

Dans la discussion du plan de contenu des unités intonatives, se pose la question des unités minimales associées à des significations particulières ainsi que sur la typologie de ces significations. Par référence à la question des unités intonatives dotées de sens, la rupture s'opère entre les approches holistiques et autosegmentales de l'intonation. Dans les premières, ce sont les contours mélodiques dans leur intégralité qui expriment une seule fonction que l'on définit soit par référence à l'acte de discours (on distingue alors les assertions, les demandes, les contradictions, Sag & Liberman 1974) soit par référence à l'attitude propositionnelle (conviction, ignorance, incertitude, Ward et Hirschberg 1985). Cette interprétation contraste avec la théorie compositionnelle de la signification des formes intonatives, défendue dans le cadre de l'analyse autosegmentale de l'intonation (Gussenhoven 1984, Pierrehumbert & Hirschberg 1990) : on suppose dans cette approche que la fonction générale du contour mélodique est une structure composite à partir des fonctions associées aux unités élémentaires (proéminences mélodiques, tons de frontières) :

> *« Our idea of the compositionality of tune meanings is based upon a hierarchical model of phonological domain, in which the scope of interpretation of tones is the node to which they are attached. So, the components of tune— pitch accents, phrase accents, and boundary tones— are each interpreted with respect to their distinct phonological domains. Pitch accents, phrase accents, and boundary tones each operate on a (progressively higher) domain of interpretation. Not only is each of these types of tone interpreted over a distinct domain, but each contributes a distinct type of information to the overall interpretation of a tune. »* (Pierrehumbert & Hirschberg 1990:286)

Dans cette approche, les accents mélodiques transmettent les informations sur le statut des différents référents discursifs et sur leur saillance relative ; les accents de phrase et les tons de frontière démontrent le degré de jonction entre deux syntagmes. Simultanément, l'approche compositionnelle du sens d'un contour mélodique est étroitement liée avec la conception hiérarchique de la structure prosodique. On spécifie ainsi que les tons de frontière sont associés avec les actes de parole, l'idée reformulée par Pierrehumbert & Hirschberg en des termes plus abstraits : « *Boundary tones convey information about the directionality of interpretation for the current intonational phrase*

– *whether it is 'forward-looking' or not* ». Les auteurs spécifient également que la structuration de l'énoncé en termes de syntagmes intermédiaires (exprimée dans le cadre de la théorie intonative adoptée en termes de distribution des accents de syntagme) traduit les relations entre les constituants adjacents : si le contenu propositionnel d'un syntagme doit former le tout avec le syntagme suivant (*« whether (the propositional content of) one intermediate phrase is to form part of a larger interpretative unit with another »*).

Di Cristo (2004) analyse la dimension sémiologique des unités prosodiques dans le cadre du modèle proposé par Hjelmslev, selon lequel tout niveau langagier s'organise autour d'une articulation entre le plan de l'expression et le plan de contenu. L'articulation entre les deux est rendue possible du fait que les unités prosodiques exercent plusieurs fonctions dans la langue. Parlant de la plurifonctionnalité de la prosodie, Di Cristo (2000) propose de distinguer six fonctions principales : structurale, de contextualisation, d'expression de l'affect (émotions et attitudes), énonciative, interactionnelle et identificatrice. Remarquons que les cinq dernières fonctions spécifiées sont définies et illustrées dans le cadre de la parole spontanée, la fonction structurale étant indépendante des contraintes stylistiques.

La classe des fonctions structurales est liée au rôle de la prosodie dans la mise en forme des informations et des structures linguistiques hiérarchiques. À cette problématique est étroitement liée la question des interfaces entre les différents niveaux langagiers, dont la théorie formelle a été proposée dans le cadre de la phonologie prosodique. Nous traiterons cette question plus en profondeur dans le chapitre V. Cette facette intervient à la fois en production et en perception de la parole. En production, si nous suivons le modèle modulaire proposé par Levelt (1989), la prosodie est gérée par le module de l'organisation profonde : on prépare des espaces (*slots*) dans la forme de surface, lesquelles permettront par la suite la transmission du message. Insistant sur le rôle de la prosodie dans l'acquisition du langage, Konopczynski (1999:68) la qualifie de « structure d'accueil » de tous les autres éléments de la parole. L'importance des structures prosodiques est également exploitée dans les études en psycholinguistique en relation avec les mécanismes de compréhension et du traitement cognitif du message linguistique (cf. Cutler et al. 1997, Peperkamp 2002). L'actualisation de cette fonction est soutenue par les fonctions adjointes de segmentation, de liage et de saillance.

Le rôle de la prosodie dans la gestion du dialogue a été exploité, entre autres, dans les travaux de Couper-Kuhlen et Selting (1996), Swerts (1997), Local (2003) : il s'agit de rendre compte du fonctionnement des indices prosodiques aux frontières discursives importantes (les tours de parole).

Nous évoquons également la fonction d'indentification du locuteur : les indices prosodiques peuvent former des structures et des patrons récurrents non-codifiés qui permettent l'identification du locuteur. Si le timbre est une caractéristique spécifique de chaque locuteur, les variations d'autres paramètres acoustiques peuvent communiquer les informations plus générales sur l'âge et le sexe du locuteur, sur son statut social (classe sociale et profession), sur son état affectif du moment.

L'un des contextes dans lequel on aborde la problématique des significations des structures prosodiques est la discussion de ce qui est universel et de ce qui est spécifique à la langue dans le domaine de l'intonation. La théorie des codes biologiques de Gussenhoven (2002a) s'inscrit dans cette lignée. L'auteur considère que les unités du plan de contenu discrètes associées aux phénomènes prosodiques (et intonatives en particulier) se forment lors du processus de grammaticalisation progressive des codes biologiques fondamentaux. Nous proposons une brève présentation de sa théorie des codes grammaticaux. Ces codes (de fréquence, d'effort et de production) gouvernent l'implémentation phonétique de l'intonation et sont responsables des aspects universaux que les chercheurs trouvent en relation avec des sens véhiculés par l'intonation. Les significations associées aux codes biologiques sont de nature iconique et relèvent du domaine du paralangage ; ils sont de deux types : les significations affectives transmettent les informations sur le locuteur, et les significations informatives portent sur le message lui-même. La mesure dans laquelle les codes biologiques sont exploités par une langue varie d'une langue à l'autre. À leur tour, les significations linguistiques forgées lors du processus de grammaticalisation sont arbitraires dans leur association entre la forme et la fonction quoiqu'elles suivent les lignes indiquées par les significations paralinguistiques. Le processus de grammaticalisation présuppose non seulement que les formes phonologiques sont discrètes mais que les sens véhiculés sont également systématisés. Une fois grammaticalisées, c'est-à-dire intégrées dans le système phonologique et morphologique de la langue, ces significations sont sujettes à des changements langagiers, ce qui explique les cas de divorce observés dans différentes langues entre les significations issues des codes biologiques et les significations actuelles.

Les codes biologiques proposés par Gussenhoven sont :

a) CODE DE FRÉQUENCE : la caractéristique physiologique sous-jacente à ce code est la taille des cordes vocales qui affecte la fréquence des vibrations. Dans le domaine du paralangage on l'associe à l'exploitation dans l'expression des relations du pouvoir. Dans l'interprétation affective, les valeurs de f0 élevées sont associées à la politesse et à une attitude amicale. Également, du point de vue informatif, elles sont corrélées avec le degré d'incertitude, et donc avec l'opposition « question – assertion ». La conséquence de la grammaticalisation de ce code consiste en ce que la majorité des langues du monde ont des contours montants, or leur utilisation dans les assertions est exceptionnelle[22]. Gussenhoven remarque pourtant que le code de fréquence perd plus fréquemment et plus facilement son caractère iconique en comparaison des deux autres codes.

b) CODE D'EFFORT : l'énergie apportée lors de la production du message explique une articulation plus nette et pour l'intonation des mouvements mélodiques plus clairs et distincts. Ce code se manifeste en parole à travers une variation de l'étendue tonale. L'interprétation affective associée met l'accent sur l'attitude du locuteur envers les informations communiquées ainsi que sur l'attention envers

[22] Rialland (2005) cite les exemples des langues africaines.

l'interlocuteur[23] : le locuteur cherche à transmettre les informations qu'il juge importantes (l'étendue compressée est utilisée dans les énoncés négatifs) ; ce code étant également associé avec l'expression de la surprise et de l'agitation. Une attitude pareille se traduit donc par une emphase (interprétation informative). On retrouve la grammaticalisation de ce code dans l'expression du focus (approche « accent-to-focus »).

c) CODE DE PRODUCTION : il est conçu en association avec les phases d'expiration, ce qui se manifeste par le phénomène de déclinaison. On observe donc la corrélation dans la variation de la hauteur mélodique aux frontières des segments discursifs (inteprétation informative) : les attaques basses indiquent la continuité par rapport au discours précédent, les attaques hautes évoquent l'introduction d'un nouveau topique ; les rapports inverses s'actualisent aux frontières terminales des segments.

Les codes sont biologiques dans la mesure où ils proviennent des mécanismes de la production de la parole responsables de la fréquence de vibration des cordes vocales. Or, avec le temps, le lien direct entre les codes et les conditions physiologiques associées (longueur des cordes vocales etc.) s'est rompu. Gussenhoven développe (Gussenhoven 2002, Gussenhoven et Driessen 2004) une théorie de substitution : comme les associations forme-fonction décrites sont intériorisées par les locuteurs, les formes produites ne doivent pas aller à l'encontre de ces intuitions, quels que soient les moyens utilisés. Il existe un ensemble de formes de substitution associées indirectement à des formes premières : les exemples sont le pic retardé qui induit les mêmes impressions que le pic mélodique haut, le registre haut comme le substitut de l'empan tonal large, la qualité vocalique.

Toujours dans le cadre de la réflexion sur l'universel dans le domaine prosodique, Hirst (2005) constate le caractère quasi-universel de la majorité des formes et des fonctions prosodiques si l'on considère les deux ensembles séparément ; la spécificité de la langue se manifeste quand on se penche sur les liens d'association qui se tissent entre les deux.

Un axe secondaire de réflexion sur le plan du contenu des unités prosodiques porte sur l'analyse des rapports établis entre la hiérarchie des constituants prosodiques et la structure informationnelle du message. Un exemple de la hiérarchie des unités informationnelles est proposé par Navarro (1997, cité par Simon 2004). Cette hiérarchie comporte quatre unités définies sur la base des critères hétérogènes : a) l'unité intonative définie par la présence d'une proéminence nucléaire ; b) une unité définie par la présence d'une pause (pouvant regrouper plus d'une unité intonative) ; c) un énoncé défini par la présence d'une pause et par l'expression d'un acte illocutoire ; d) le paraton conçu comme le domaine d'un seul topique discursif. Nous voyons donc que cette structure hiérarchique s'entrecroise avec la hiérarchie des constituants prosodiques avec l'unité intonative faisant partie des deux. Notons également que dans les premières

[23] Il a été démontré dans les expériences en perception (Gussenhoven et al. 1997) que le jugement sur la hauteur du pic mélodique dépend du registre global de la phrase ; orienté vers l'interlocuteur, le code d'effort porte donc sur l'excursion inférée de la f0.

définitions du concept de l'unité intonative, c'est sa dimension informationnelle qui était mise en avant, cf. Halliday (1970) :

> *« In general […] tonic prominence expresses the structure of information. [...] The choice of tonic prominence – where to put the tonic, also in fact, where to divide up into tone groups – relates to how the message is divided into units of information, where the main 'new information' lies, and how it ties up with what has been said before: anything that contributes to the structure of the discourse, in other words. »* (Halliday 1970:22)

On constate une perspective similaire dans l'introduction de la contrainte de la cohérence sémantique (*Sens Unit Condition*) dans la définition des unités intonatives dans l'approche de Selkirk (1984). De ce travail on hérite du concept de l'unité intonative comme unité de sens, ce qui présuppose dans la définition de ces unités de faire référence non seulement à des propriétés phonologiques ou syntaxiques du matériau linguistique analysé mais obligatoirement de prendre en compte la dimension sémantique :

> *« Our position, then – again following Halliday 1967a,b – is that there are no strictly syntactic conditions on intonational phrasing. Any apparently syntactic conditions on where 'breaks' in intonational phrasing may occur are, we claim, ultimately to be attributed to the requirement that the elements of an intonational phrase must make a certain kind of semantic sense. »* (Selkirk 1984:286)

Dans la réflexion de Selkirk, c'est la structure argumentative de l'énoncé qui se retrouve étroitement liée à la notion de l'unité de sens.

Dans les études portant sur l'articulation syntagmatique du message en russe, la segmentation en syntagmes est tout d'abord une segmentation sémantique. La définition de base remonte aux travaux de L.V. Scherba (1963) qui définissait le syntagme tout premièrement comme une unité de la structure sémantique de l'énoncé (structure de sens) possédant tout de même des caractéristiques formelles (« *единица речи-мысли* »[24]). La Grammaire russe (1980) définit les syntagmes comme « *интонационно нечленимые смысловые отрезки речи*[25] ». Parmi les caractéristiques formelles du syntagme, on parle de la présence d'un accent particulier, accent de syntagme sur le dernier mot lexical du syntagme ; un autre aspect formel à considérer est le fait que le contour mélodique a pour domaine d'application un syntagme. Néanmoins, l'aspect formel est parfois omis des considérations sur la structuration prosodique de l'énoncé, les aspects sémantico-syntaxiques s'appropriant le rôle central ; dans cette perspective, le terme de syntagme devient un synonyme de la phrase syntaxique (*словосочетание*)[26]. De cet état de choses émane la démarche méthodologique adoptée dans les travaux expérimentaux : on procède tout d'abord à la segmentation du texte en syntagmes à partir des relations sémantiques et syntaxiques ;

[24] *Unité de parole et de pensée* (notre traduction).
[25] *Fragments de parole dotés du sens et non-segmentables dans le domaine intonatif.*
[26] On retrouve également le concept du syntagme dans les travaux des spécialistes de la versification, ceci accorde au syntagme un rôle dans l'organisation de la structure rythmique et métrique (du vers, en particulier) : pourtant, ce concept n'est pas largement exploité dans les travaux phonétiques portant sur l'organisation rythmique de l'énoncé.

la segmentation attestée est ensuite confirmée dans les études expérimentales. Pour notre part, nous considérons que la prise en compte exclusive des rapports sémantico-syntaxiques ne permet pas de transposer le procédé aux tâches d'annotation automatique ni aux analyses de la parole spontanée où la cohérence des structures syntaxiques est fréquemment perturbée. Nous proposons donc de conjuguer les démarches ascendante (*bottom-up*) et descendante (*top-down*) dans le travail sur l'organisation prosodique du parler russe spontané et de confronter une segmentation issue des informations acoustiques et formelles avec l'articulation de l'interface « prosodie (intonation) ~ syntaxe ». En effet, si un ensemble d'indices acoustiques est systématiquement associé à la perception d'une frontière prosodique, la démarche consistant à examiner si l'on ne le retrouve pas ailleurs paraît cohérente ; en cas de réponse positive c'est le lien avec la dimension linguistique que l'on place au centre des réflexions.

Par conséquent, après avoir défini les fonctions prosodiques et avoir insisté sur le rôle important joué par la fonction structurale, nous proposons de réfléchir sur les primitives fonctionnelles qui participent à son actualisation dans la parole. Hirst (1977, 2005) propose de reconnaître les primitives telles que « accentué », « inaccentué », « emphatique », « nucléaire », « frontière », « terminal », « non-terminal ». Ces fonctions s'organisent autour des concepts d'accent et de frontière : une unité peut être accentuée ou inaccentuée, porter une proéminence nucléaire ou encore emphatique ; dans le même esprit, une frontière, si elle est présente, est caractérisée selon sa localisation au sein d'une unité plus large (frontière terminale versus non-terminale). Dans le déroulement du schéma, les primitives fonctionnelles reçoivent une spécification au niveau phonologique et ensuite une implémentation phonétique. La démarche que nous adoptons consistera donc à décrire l'organisation intonative du russe en faisant une référence systématique aux concepts d'accent (de proéminence mélodique) et de frontière.

1.4. PROLÉGOMÈNES À LA PHONOLOGIE INTONATIVE

Dans les sections précédentes, nous avons abordé certaines problématiques théoriques dont la discussion nous était indispensable pour justifier la perspective de l'analyse phonologique (formelle) de l'intonation qui est la nôtre. En particulier, nous nous sommes penchée sur les questions de l'association entre la forme et la fonction dans les analyses intonatives, mais également sur le rôle du niveau phonologique dans ces études. Il découle des discussions proposées que l'approche formelle n'est pas homogène en soi ni acceptée à l'unanimité.

La tendance actuelle est d'adopter une approche de la phonologie autosegmentale-métrique dans l'analyse de l'organisation intonative dont nous présentons les points-clefs dans la section suivante. La présente section se donne la tâche de proposer une ébauche des réflexions sur les primitives de l'analyse intonative ; dans un deuxième temps, nous présentons une dichotomie des modèles intonatifs, opposant les modèles superpositionnels aux modèles des séquences tonales (*Tonal Sequence models*). Les acquis de la phonologie intonative se voient largement exploités dans les systèmes d'annotation prosodique (le chapitre suivant sera consacré à la présentation de certains d'entre eux).

1.4.1. Problématique de l'unité minimale de description de l'organisation tonale.

Toute description formelle d'un phénomène linguistique comportera inévitablement trois facettes :

- les unités minimales, les primitives ;
- les règles de groupement de ces éléments de base en domaines / constituants / événements des niveaux supérieurs ;
- spécifications d'association et d'alignement de ces événements prosodiques avec le matériau segmental.

En conséquence, nous nous tournons maintenant vers la problématique de l'unité minimale distinctive dans les descriptions phonologiques de l'intonation. Dans la perspective d'une analyse formelle, on est amené à analyser la courbe de f0 (correspondant au niveau de la représentation physique) en termes d'unités phonologiques discrètes. Il importe donc de préciser l'inventaire de ces unités : dans la tradition des études de l'intonation, trois primitives analytiques pourraient avoir le statut de primitive dans une description phonologique de l'intonation, à savoir :

- *contour mélodique ;*
- *niveaux (tons statiques) ;*
- *mouvements mélodiques discrets (configurations, approche dynamique[27]).*

Nous discutons ci-après des différents choix de la primitive descriptive faits au fil des études intonatives.

1.4.1.1. Le contour mélodique

Le contour mélodique peut être considéré comme une primitive ou une construction selon que l'on se positionne du côté des approches holistiques ou autosegmentales de l'intonation. On définit le contour mélodique comme un segment du continuum mélodique associé avec un segment de parole ; par cette association, la liaison s'opère avec le niveau fonctionnel et le segment de parole acquiert une signification pragmatique spécifique. Dans les approches holistiques, le contour mélodique est une unité non-décomposable en constituants plus petits et ce ni au niveau formel ni au niveau fonctionnel (cf. description du français par Delattre 1966, ainsi que le modèle de Bryzgounova 1977 pour le russe). Notons que les approches holistiques précèdent historiquement les descriptions plus phonologiques de l'intonation. Les partisans de cette approche globaliste peuvent reconnaître les particularités de l'organisation interne

[27] Remarquons que les approches morphologiques de l'intonation qui utilisent le contour mélodique entier comme unité de description peuvent également être classées comme les descriptions en termes de mouvements mélodiques ; or, nous réservons cette rubrique aux approches qui distinguent les mouvements mélodiques discrets associés avec des unités plus petites que l'unité intonative (domaine d'association du contour mélodique).

du contour (la différence en niveau mélodique entre les syllabes accentuées et inaccentuées, les différences mélodiques pour les syllabes avant le premier et après le dernier accent de l'unité intonative), pourtant ces détails sont relégués au domaine de l'implémentation phonétique. Il est également reconnu qu'un patron mélodique s'accommode au matériau linguistique disponible.

Inversement, dans les approches autosegmentales, sur le plan de l'expression, le contour mélodique est conçu comme représentable en termes d'unités plus petites. En réfutant l'approche globaliste du contour mélodique, on se pose la question des unités de base au moyen desquelles la structure d'un contour intonatif est décrite. Dans cette perspective, on oppose les tons de l'approche structuraliste américaine (Pike 1945) et les configurations de la tradition britannique. Une remarque doit être faite sur les analyses de l'intonation de l'anglais dans le cadre des travaux de l'école britannique et dans la tradition structuraliste américaine : ces deux écoles d'intonologie peuvent être considérées comme les précurseurs des approches phonologiques, dans la mesure où elles introduisent la notion de la primitive dans leurs analyses et descriptions. L'intonation est décrite comme un ensemble restreint d'éléments de base ayant le statut de phonèmes mélodiques.

Les contours mélodiques acquièrent un statut particulier dans le cadre des approches superpositionnelles de l'intonation (nous développerons davantage ce sujet dans la section 1.4.2).

1.4.1.2. Les niveaux (tons statiques) de la tradition américaine

Les structuralistes américains distinguent trois types de sous-composants en association avec le domaine intonatif : *pitch* (mélodie), *stress* (accentuation) et *juncture* (jonction, structuration). En même temps, l'élément de base du contour mélodique est représenté pour eux par un point-cible, défini par sa hauteur mélodique (niveau). Ainsi Pike (1945 : 25) propose quatre tons de niveau pour décrire l'intonation de l'anglais américain (*extra high, high, mid, low*) ; ces niveaux peuvent être regroupés ensuite en des schémas récurrents, les contours mélodiques. Notons que c'est une analyse phonémique de l'intonation : les tons de niveau n'ont pas de sens en eux-mêmes, le plan de signification est associé avec le contour mélodique entier[28] :

> *« It is the intonation contour as a whole which carries the meaning while the pitch levels contribute end points, beginning points or direction-change points to the contour – and as such are basic building blocks which contribute to the contours and hence contribute to the meaning. »* (Pike, 1945:26).

Un autre élément phonémique est rajouté aux tons de niveaux de Pike dans les travaux de Hockett (1955) et Trager & Smith (1951), à savoir l'équivalent du ton de frontière, *terminal juncture* : ce sont les transitions mélodiques à la fin d'un syntagme et avant la pause :

[28] Les contours comportant le même ton de niveau (H, le plus haut) peuvent comporter des éléments de signification en commun, mais il s'agit alors d'une signification extralinguistique ('*some element of inexpectedness*').

- ton descendant bas dans le registre avant la pause
|| - montée finale dans l'intonation
| - ton du niveau (soutenu)

De son côté, Hockett (1955) élabore une description de l'intonation de l'anglais américain avec trois phonèmes mélodiques (les tons de niveaux de Pike à l'exception d'un ton extra haut) et trois tons de frontières ('*terminal contours*' dans sa terminologie) : neutre, montant et descendant.

L'analyse de l'intonation en tons de niveau permet de réaliser une description à partir des primitives ayant un statut phonémique, étant des éléments irréductibles et contrastifs. Elle a été ensuite reprise dans les travaux fondateurs de l'approche autosegmentale : Liberman (1975) décrit l'intonation de l'anglais américain à l'aide de trois niveaux (H, L, M) tandis que les accents mélodiques et autres événements tonals dans la description de Pierrehumbert (1980) sont annotés en deux niveaux seulement.

1.4.1.3. Les mouvements mélodiques

La critique majeure de l'approche des structuralistes américains revient à D. Bolinger (1951, 1972, 1989). Pour Bolinger, l'approche par niveaux intonatifs manque de pouvoir généralisateur pour capter et refléter les différences sémantiques ou les ressemblances formelles entre deux contours mélodiques, ce qui résulte en un découpage trop fin (en termes de contours mélodiques distinctifs) du continuum mélodique. Ainsi, les contours qui reçoivent l'annotation '2-3-1' et '2-4-1' chez Pike (1945) correspondent à un mouvement mélodique ascendant-descendant étroit ou large respectivement : l'école britannique insiste tout praticulièrement sur la ressemblance générale de la forme du mouvement (ascendant-descendant) ce qui permet de regrouper deux contours sous une même étiquette, un diacritique permettant de spécifier les empans tonals en jeu et signaler une différence[29]. En outre, les niveaux intonatifs de la tradition américaine sont jugés trop arbitraires, étant donné le caractère continu, et non catégoriel, des phénomènes intonatifs.

L'école britannique de l'intonation élabore son cadre descriptif dans une perspective à vocation essentiellement pédagogique. Cette tradition est caractérisée entre autres par une intégration des phénomènes appartenant aux domaines de l'intonation et de l'accentuation dans une unique description de l'organisation tonale de l'énoncé, et ceci par l'intermédiaire du concept de l'accent mélodique. On considère que ce sont les syllabes dotées d'une proéminence mélodique (*accented*) qui jouent le rôle primordial dans l'organisation du contour mélodique associé avec un énoncé ou une de ses parties. D'autre part, l'école britannique reconnaît que les détails les plus importants de l'organisation intonative sont associés avec la syllabe porteuse de l'accent principal de l'unité intonative (accent nucléaire). De cette considération résulte une structuration

[29] Dans la réflexion sur l'association du plan d'expression et du plan de contenu pour les unités intonatives, les auteurs de l'école britannique considèrent que c'est surtout la forme générale du mouvement qui a son rôle à jouer.

interne du contour mélodique (Palmer 1922, O'Connor et Arnold 1973, Halliday 1967) : dans un contour mélodique on distingue ainsi la tête (*head*), qui décrit les mouvements mélodiques dans la partie de l'unité intonative précédant l'accent nucléaire, et le noyau (*nucleus*) qui correspond à la partie du contour mélodique commencée par l'accent nucléaire. Ce sont les configurations mélodiques globales associées avec le nucleus, les tons nucléaires, qui sont les unités phonologiquement distinctives ; la partie '*tail*'[30] qui décrit le mouvement mélodique sur des syllabes inaccentuées postérieures à l'accent nucléaire, spécifie uniquement la forme phonétique de la courbe et n'a pas de statut phonologique. L'annotation de l'organisation tonale développée par l'école britannique contient les étiquettes qui décrivent la forme des mouvements mélodiques[31] ainsi que les spécifications de la hauteur initiale et de l'amplitude du mouvement mélodique[32]. Dans cette perspective, la description proposée par O'Connor et Arnold (1973) aboutit à dix combinaisons possibles des mouvements mélodiques associés aux parties structurelles du contour mélodique qui forment le noyau du système intonatif de l'anglais britannique.

Les configurations mélodiques sont également reconnues comme des unités de base par l'école hollandaise de l'analyse intonative. Les mouvements mélodiques y sont considérés comme les unités minimales de l'analyse perceptive réalisée par l'auditeur lors de la perception du message de son interlocuteur. On définit comme un mouvement mélodique les différents types de tons dynamiques montants ou descendants qui se regroupent ensuite en configurations (formes plus complexes), auxquelles on attribue un des trois rôles structurels (préfixe, racine ou suffixe). Les séquences des configurations mélodiques définies en fonction de leur rôle structurel forment les contours mélodiques. Le choix méthodologique consistant à effectuer la description de l'intonation à partir des mouvements mélodiques primitifs s'inscrit pour l'approche hollandaise dans son orientation générale perceptive : les chercheurs travaillant dans le cadre de cette approche considèrent que ce sont les mouvements mélodiques, et non pas les niveaux atteints, qui sont significatifs en perception, cf. Proposition 2 de leur modèle :

> *« At the first level of description, the smallest unit of perceptual analysis is the pitch movement. »* ('t Hart *et al*., 1990:72)[33]

1.4.1.4. Les accents mélodiques, les accents de syntagme et les tons de frontière dans le modèle de J.Pierrehumbert

La thèse de Janet Pierrehumbert (1980) visant à proposer une description phonétique et phonologique de l'intonation de l'anglais américain est un des travaux fondateurs de l'approche autosegmentale de l'intonation. En référence à la problématique de l'unité minimale de description dont nous discutons dans cette section, ce travail nous intéresse

[30] Certains chercheurs distinguent également l'avant-tête (*pre-head*) : partie structurelle du contour mélodique qui décrit le mouvement mélodique associé avec des syllabes inaccentuées précédant le premier accent de l'unité intonative.
[31] *'rise', 'fall', 'fall-rise', 'rise-fall', 'rise-fall-rise'* etc.
[32] *'high fall', 'narrow rise'*.
[33] La Proposition 3 de l'approche hollandaise contient une affirmation encore plus catégorique : « *There are no pitch levels* » (1990 : 75).

dans la mesure où il aboutit à une description qui intègre à la fois le concept de niveaux mélodiques et celui des configurations mélodiques.

Comme nous l'avons déjà dit, dans les approches autosegmentales, le contour mélodique est une entité dérivée, compositionnelle, au moins sur le plan formel, et peut donc être décomposé en des événements (ou unités) plus petit(e)s. Pour l'anglais, ainsi que pour un ensemble de langues européennes, on distingue deux types d'événements phonologiques minimaux pertinents : les accents mélodiques et les tons de frontière, cf. Ladd (1996 : 42) où l'auteur présente un résumé des principes de base des approches autosegmentales métriques de l'intonation :

> *« In languages like English, the most important events of the tonal string are pitch accents, which are associated with the prominent syllables in the segmental string and the edge tones which are associated with the edges of prosodic domains of various sizes. »*

Par la suite, les accents mélodiques, définis comme des mouvements mélodiques associés avec des syllabes porteuses de l'accent lexical leur conférant une proéminence perceptive supplémentaire, sont analysés par Pierrehumbert en termes de tons de niveaux : chaque accent mélodique reçoit son annotation en deux tons H (*high*) et L (*low*), correspondant aux cibles mélodiques sous-jacentes. Remarquons que les tons H et L sont définis de manière relative, et non pas absolue ; par conséquent, dans l'analyse des mouvements descendants, la différence n'est pas faite entre la descente du niveau haut vers le niveau médian ou bas ni du niveau médian vers le bas : tous ces mouvements descendants reçoivent une description HL.

C'est donc que, dans le modèle de Pierrehumbert, on reconnaît à la fois la pertinence des niveaux et celle des configurations mélodiques (pour les accents mélodiques bitonaux qui possèdent une structure interne). Le modèle contient ainsi deux types de contrastes paradigmatiques : entre les tons H et L (binaire) et entre les différents accents mélodiques. En même temps, le contraste entre différentes catégories d'accents mélodiques reçoit une double interprétation : d'une part, il s'agit d'un contraste de la forme globale de l'accent mélodique (H* et H*+L) ; en même temps, la différence dans l'organisation structurelle interne de l'accent entre en jeu (opposition entre les deux accents H+L* et H*+L).

1.4.2. Les modèles linéaires versus superpositionnels de l'intonation

Du choix des primitives dans l'analyse de l'intonation émane la première typologie des systèmes intonatifs.

Empruntant la terminologie à Ladd (1983), les rapports d'opposition se créent entre les modèles intonatifs séquentiels, qui interprètent le contour mélodique comme une séquence de tons (*Tone Sequence Models*), et les modèles superpositionnels, fondés sur les interactions entre les contours mélodiques associés aux différents domaines (*Contour Interaction Models*). Cette dichotomie des approches de l'intonation rejoint la problématique des rapports entre deux mécanismes qui contribuent à l'actualisation en

parole de la fonction du groupement : il s'agit des mécanismes de l'accentuation et de l'intonation. Nous avons signalé auparavant que ce sont les domaines d'application de deux mécanismes qui diffèrent. Or, à l'oral, ces mécanismes se voient impliqués dans les rapports d'une superposition structurelle : un syntagme, étant le domaine du marquage intonatif, contient un ou plusieurs mots prosodiques, les unités résultant d'un groupement accentuel[34].

Dans les modèles superpositionnels, le contour mélodique est traité comme une fonction complexe qui peut être décomposée en plusieurs fonctions élémentaires : ainsi les perturbations mélodiques locales, associées avec des proéminences mélodiques sont-elles superposées sur une courbe globale associée avec un syntagme entier. Cette architecture est propre aux modèles de Fujisaki (1983), de Garding (1983) et de l'école de Lund, de Vaissière (1983), Mobius (1993), Grønnum (1992). Ainsi le modèle hiérarchique de Grønnum, pour l'intonation du danois, considère que la forme finale d'un contour mélodique est le résultat de l'interaction des composants dont les domaines d'association diffèrent sur l'axe temporel : ainsi un composant gère l'organisation mélodique au niveau du paragraphe (le niveau supérieur), un deuxième agit au niveau d'un énoncé, et ainsi de suite jusqu'au niveau du groupe rythmique (*stress group*).

Dans un modèle superpositionel à deux composants, on se questionne sur le rapport entre la courbe syntagmatique et le phénomène de déclinaison, un abaissement global de la courbe intonative à travers l'énoncé en entier ('t Hart & Collier 1967). Par ailleurs, même si l'on reconnaît que le composant syntagmatique reflète la déclinaison, le problème de sa modélisation quantitative persiste. Ladd (1996 : 29) remarque que le problème majeur de la modélisation quantitative de l'architecture superpositionelle réside dans la quantité importante nécessaire de la projection dans ce qui serait dit ensuite (*lookahead*) : ainsi dans la grille structurante de Garding (1983), les paramètres initiaux du composant syntagmatique dépendent de la longueur du domaine auquel ce dernier s'applique.

Dans les modèles linéaires, le contour mélodique représente une séquence linéaire d'événements mélodiques pertinents et de tons. Les événements mélodiques jugés pertinents et donc pris en compte par les modèles sont traditionnellement les accents mélodiques et les tons de frontières : cette union dans la même représentation des tons lexicaux et intonatifs constitue un des aspects novateurs des descriptions de l'intonation du suédois (Bruce 1977) et de l'anglais américain (Pierrehumbert 1980)[35]. Dans ces approches, chaque événement mélodique reçoit sa spécification tonale, et la différence entre ce qui relève du domaine de l'accentuation et ce qui est intonatif résulte de l'alignement des catégories tonales avec les syllabes proéminentes ou les frontières des constituants prosodiques.

[34] Remarquons que les deux phénomènes s'actualisent dans le discours à l'aide des mêmes indices acoustiques sur le plan de la substance.

[35] Cf. la remarque de Gussenhoven (2002b) sur l'approche IPO, qu'il juge être autosegmentale dans sa logique interne, « avant la lettre », bien que les termes « proéminence mélodique » et « ton de frontière » n'y soient pas explicitement utilisés ('t Hart et ses collègues font plutôt la distinction entre mouvements mélodiques communiquant la proéminence et mouvements mélodiques qui ne la communiquent pas *(prominence-lending & non-prominence lending pitch movements)*).

Le terme d'alignement que nous venons d'utiliser évoque la tradition autosegmentale-métrique de l'analyse de l'organisation tonale de l'énoncé, que nous résumons brièvement dans le paragraphe suivant.

1.5. APPROCHE AUTOSEGMENTALE-MÉTRIQUE DE L'INTONATION (PHONOLOGIE INTONATIVE)

La plupart des descriptions récentes de l'intonation des langues européennes se situent dans le cadre de la phonologie métrique et autosegmentale. L'approche autosegmentale dans l'analyse de l'organisation tonale s'inspire des travaux en tonologie des langues africaines. Cette approche a pour unité de description un ton doté du statut d'autosegment (une valeur phonologique en soi), et une attention particulière est portée sur l'association entre les paliers tonal et segmental (Leben 1973, Goldsmith 1976). Dans l'optique autosegmentale, les informations phonologiques sont représentées sur des paliers séparés qui sont autonomes et qui regroupent des objets comme les segments, les tons, le trait de nasalité etc. Pourtant, les informations phonologiques étant réparties entre plusieurs niveaux d'une représentation, ces derniers sont reliés entre eux à l'aide du mécanisme d'association, qui, dans les premiers travaux sur la phonologie du ton en langues africaines, relie les tons aux voyelles[36] en conformité avec les contraintes de bonne formation. Dans le transfert de cette méthodologie analytique vers les langues à intonation, on a gardé le concept d'un niveau à part pour représenter les phénomènes mélodiques ainsi que le concept d'association : par association, on entend une relation définie au niveau phonologique entre un élément du niveau structurel (une syllabe, par exemple) et un élément du niveau tonal (Pierrehumbert et Beckman, 1988:153)[37,38]. Les événements mélodiques qui composent le niveau tonal sont les accents mélodiques et les tons de frontières, chacun d'eux étant composé d'un ou plusieurs segments tonals (les tons H et B). Les approches autosegmentales de l'organisation tonale/mélodique des langues à intonation remontent aux études de Bruce (1977), Liberman (1975) et Pierrehumbert (1980). Pourtant, comme le souligne Gussenhoven (2002b) la séparation des niveaux segmental et tonal dans la mesure où les traits mélodiques n'ont pas été inclus dans la matrice de traits d'un segment, est implicite dans la tradition britannique (O'Connor & Arnold, 1973) ou dans les travaux de l'école hollandaise de l'intonation ('t Hart, Collier, & Cohen, 1990).

Simultanément, avec la gestion de l'organisation tonale du message, on considère que son contenu segmental est également organisé sous la forme d'une structure (grille)

[36] Ces dernières étant considérées comme les unités porteuses des tons, *tonal bearing units, TBU.*

[37] Il faut préciser que la phonologie générative n'a pas prêté beaucoup d'attention à la forme phonologique des phénomènes intonatifs (les problématiques de la dérivation à partir de la syntaxe de la structuration intonative et du placement de l'accent nucléaire ont suscité plus d'intérêt). Stockwell (1960, cité par Fox 2000) retranscrit le modèle de Trager et Smith (1951) en termes du formalisme proposé par Chomsky (1957) ce qui résulte en 5 règles. La première règle dissocie la structure syntaxique de l'énoncé de son patron mélodique ; ce dernier est ensuite dérivé par l'application de quatre règles qui énumèrent les différentes possibilités. Il est remarquable de noter que la syntaxe et la dérivation du patron intonatif sont dissociées à l'étape initiale de la dérivation du contour mélodique : par ce procédé, l'organisation mélodique est traitée comme autonome par rapport à tout autre aspect de l'énoncé.

[38] Les études sur le détail d'alignement entre les niveaux tonal et segmental constituent actuellement une branche importante d'études en prosodie.

métrique (Liberman et Prince 1977) et d'une structure prosodique. Par conséquent, on postule que les proéminences mélodiques ne peuvent être associées qu'avec les syllabes métriquement fortes, accentuées dans le lexique mental. Cette contrainte marque une rupture avec l'analyse tonale des langues africaines où chaque unité porteuse du ton (une voyelle) doit obligatoirement avoir une spécification tonale.

La figure 1.5 résume schématiquement le type de représentation résultante, acceptée en phonologie intonative autosegmentale et métrique.

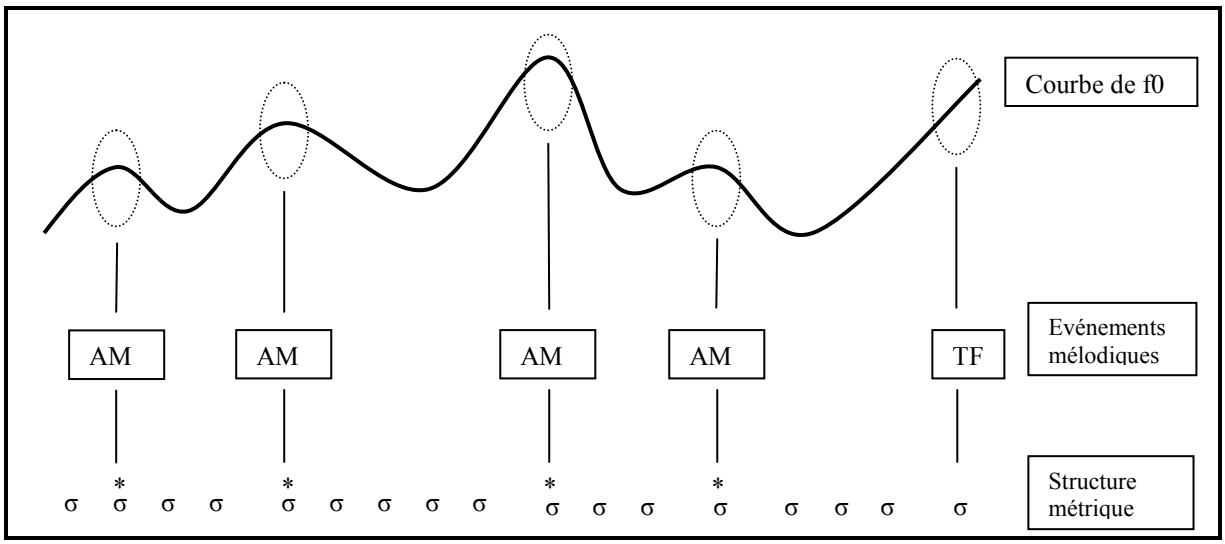


Figure 1.5 : Représentation schématique de différents paliers de la phonologie intonative

1.6. DISCUSSION

Jusqu'à présent nous avons organisé notre réflexion autour de problématiques générales liées à l'analyse prosodique. La prosodie étant une structure complexe, nous avons entrepris la présentation de ses trois ordres structurels ; nous avons également présenté deux mécanismes prosodiques structuraux, à savoir l'accentuation et l'intonation, jugeant cette dichotomie comme orthogonale à la représentation tripartite. Ces réflexions nous ont permis d'insister sur le rôle majeur joué par la dimension localisante et le phrasé prosodique (conçu comme une hiérarchie multiniveaux, cf. *infra*) : en effet, la différence entre l'accentuation et l'intonation est celle du domaine d'actualisation des phénomènes respectifs.

Nous avons également cherché à présenter des arguments en faveur du niveau formel dans les analyses et les descriptions de l'intonation ; notons pourtant que différentes structures comportant le niveau formel, phonologique, ont été proposées pour le domaine prosodique, et la lecture du terme « phonologique » fait toujours débat (cf. les critiques de Beckman et collègues émises à l'encontre des systèmes d'annotation INTSINT et IViE). Pour notre part, nous ne concevons pas le système d'analyse des faits prosodiques, et tonals en particulier, sans recourir au niveau phonologique abstrait et discret, qui assure l'articulation entre le plan de l'expression et le plan fonctionnel. Suite à l'examen de la conception proposée par Beckman et Edwards (1990), nous retiendrons comme hypothèse de travail l'idée d'une structure phonologique multiniveaux, chaque niveau possédant des corrélats acoustiques différents.

La troisième section de ce chapitre est consacrée à la discussion de la dimension fonctionnelle, du plan de contenu des unités prosodiques et intonatives. C'est le niveau phonologique dans la représentation générale qui assure la gestion de l'interface « forme ~ fonction et signification ». Nous avons argumenté, à la suite de Hirst (2005), en faveur de la définition des primitives fonctionnelles, telles les frontières et les proéminences (chacune ayant un ensemble d'attributs), qui servent de médiateurs entre la forme et le sens dans les descriptions prosodiques. D'autre part, interagissant avec le niveau phonologique de surface, ces primitives assurent le lien avec le plan de l'expression et les paramètres physiques observables.

Enfin, nous avons cerné les présupposés principaux de la phonologie intonative qui émergent à la jonction des principes émanant de la phonologie autosegmentale et de la phonologie métrique. Il s'agit donc d'un système multiniveaux. S'inscrivant dans le cadre des modèles séquentiels de l'intonation (*Tone Sequence Models*), la phonologie intonative reconnaît deux types de phénomènes intonatifs : les proéminences (accents) mélodiques et les tons de frontière qui sont à leur tour analysés en termes de tons ayant le statut d'autosegment (remarquons qu'une analyse alternative avec les mouvements mélodiques en guise d'unités minimales est tout à fait possible dans le cadre général de l'approche autosegmentale métrique).

Tout au long de ce chapitre nous avons évoqué et discuté des principaux postulats de l'approche autosegmentale-métrique qui déterminent la conception moderne de la prosodie et définissent les axes des travaux expérimentaux menés (et à mener) sur les données observables. Ce sont les principes qui sous-tendront notre travail expérimental.

Or, notre objectif étant de décrire l'intonation du point de vue formel, il nous semble nécessaire de passer en revue un ensemble de systèmes d'annotation prosodique, en prêtant attention aux modèles phonologiques sous-jacents à ces systèmes. La plupart d'entre eux adoptent l'approche autosegmentale métrique, initialement développée par Janet Pierrehumbert dans sa thèse sur l'intonation de l'anglais américain (Pierrehumbert 1980). Nous insistons donc dans la présentation de ces systèmes sur les nuances dans l'adoption des principes généraux de l'approche initiale de Pierrehumbert.

CHAPITRE II

SYSTÈMES D'ANNOTATION DE L'INTONATION

Dans le présent chapitre nous dresserons un panorama des systèmes d'annotation développés pour représenter les propriétés tonales de la structure prosodique. Le principe de base de toute annotation est d'enrichir les données brutes observables avec une description linguistique symbolique. Simultanément, on cherche à annoter ce qui relève de la volonté et de l'intention du locuteur, les choix qu'il fait à tout moment du discours, et par conséquent, ce qui constitue le fond commun des représentations cognitives que les interlocuteurs possèdent pour le fonctionnement de la prosodie en général et de celle de la composante tonale, en particulier. Comme il s'agit d'une annotation symbolique, donc discrète, tous les systèmes que nous exposons disposent d'une interprétation qui leur est propre du rapport entre les représentations physique, phonétique et phonologique dans l'analyse des faits tonals.

Nous abordons la discussion avec le système proposé à l'Institut d'Études pour la Perception de Eindhoven : l'originalité de ce système réside dans son orientation perceptive et dans son centrage sur l'auditeur.

La plupart des systèmes que nous présentons par la suite adoptent l'approche autosegmentale métrique, initialement développée par Janet Pierrehumbert dans sa thèse sur l'intonation de l'anglais américain : ce travail de thèse a donné naissance au système d'annotation prosodique ToBI. Enfin, nous argumentons notre choix de travailler avec le système INTSINT (Hirst et Di Cristo 1998).

2.1. INTRODUCTION

Dans le premier chapitre nous avons esquissé les principes (axiomes fondateurs de la phonologie intonative) et les problématiques (statut du niveau phonologique, rapport « forme ~ fonction et signification ») qui sous-tendent les études modernes en intonation.

Dans le présent chapitre nous proposons de dresser un panorama des systèmes d'annotation développés pour la strate tonale de la structure prosodique. Le principe de base de toute annotation est d'enrichir les données brutes observables avec une

description linguistique symbolique. Simultanément, on cherche à annoter ce qui relève de la volonté et de l'intention du locuteur, les choix qu'il fait à tout moment du discours, et par conséquent, ce qui constitue le fond commun des représentations cognitives que les interlocuteurs possèdent pour le fonctionnement de la prosodie en général et de la composante tonale, en particulier. Comme il s'agit d'une annotation symbolique, donc discrète, tous les systèmes que nous exposons disposent d'une interprétation qui leur est propre du lien entre les représentations physique, phonétique et phonologique dans l'analyse des faits tonals. Nous essayerons d'en souligner les points-clefs lors de la présentation que nous entreprenons.

Selon Hirst (1981) trois critères doivent être mis en évidence pour tout système d'annotation : l'alphabet des symboles employés, la séquence des symboles autorisée et les conventions assurant l'interprétation de ces symboles. Dans la limite des informations que fournissent les études discutées, nous essayerons de mieux cerner ces trois critères.

Kochansky & Shih (2005) proposent une classification partielle des approches formelles de l'intonation (la figure 2.1 reproduit le schéma de l'étude discutée) : cette classification possède deux dimensions, l'une portant sur le degré de spécification de la courbe de f0 dans la représentation symbolique par le modèle[39], l'autre précisant si une vision superpositionnelle de l'intonation est adoptée. Sans doute, cette typologie est-elle partielle. Dans la présentation que nous entreprenons, nous avons choisi de mettre l'accent sur d'autres propriétés des systèmes d'annotation. Ainsi, un axe typologique complémentaire résulte de l'orientation générale du système, si l'accent est mis lors de la modélisation sur les aspects de la production (ToBI, INTSINT) ou de la perception (IPO). Le critère important dans la réflexion que nous avons ouverte précédemment porte sur le rôle et l'importance accordée à la représentation phonologique au sein de tout modèle.

Table 1
Intonation Schools classified by the way they describe prosody

	Under-specified	→	→	→	Fully specified
Single component	INTSINT	ToBI Xu		Tilt, IPO	Olive, Machine learning
Two components	Grønnum			Fujisaki	
Multiple components					Van Santen

Figure 2.1 : Classification des systèmes d'annotation de l'intonation selon Kochansky et Shih (2005)

Nous commençons notre discussion par le système proposé à l'Institut d'Études pour la Perception : l'originalité de ce système réside dans son orientation perceptive et dans son centrage sur l'auditeur. Notons pourtant que d'autres systèmes s'exercent également à prendre en considération les aspects perceptifs, qui sont pourtant secondaires dans leurs procédures analytiques (cf. Vaissière 2005 sur l'importance de la dimension perceptive dans les études prosodiques). Nous nous tournons ensuite vers la présentation du système ToBI, ayant le statut d'un standard dans les recherches

[39] La courbe de f0 associée à un événement prosodique (accent mélodique ou frontière) peut être spécifiée de manière continue ou être sous-spécifiée (spécification comportant des blancs ou spécification sous la forme d'une étiquette abstraite discrète).

prosodiques, et des schémas d'annotation dérivés (ToDI, IViE). Nous insistons donc dans la présentation de ces systèmes sur les nuances apportés dans l'adoption des principes généraux de l'approche autosegmentale métrique introduite par l'étude de Pierrehumbert (1980).

2.2. APPROCHE IPO

Le modèle de l'intonation et le système d'annotation associé, proposés par les chercheurs travaillant à l'Institut pour la Recherche en Perception (IPO) aux Pays-Bas ('t Hart *et al.*, 1990, pour un résumé) se différencient d'autres approches par la position prise sur l'orientation générale du travail analytique mené : ce n'est pas la question des catégories intonatives phonologiquement distinctes qui est le fondement de cette analyse, mais les caractéristiques de la courbe de f0 pertinentes en perception. C'est la réflexion sur l'imperfection des méthodes disponibles dans la recherche intonative[40] qui fait naître chez 't Hart et ses collaborateurs l'idée de se positionner dans leur travail du point de vue de l'auditeur : cette position prise émane du souci de minimiser le risque d'omettre des détails caractéristiques des contours mélodiques sur le plan formel comme fonctionnel.

L'aboutissement d'un long travail collectif étant présenté dans l'ouvrage de 1990 ('t Hart *et al*., 1990), la théorie de l'intonation y prend la forme de 10 propositions : les propositions 1 à 6 portent sur l'organisation perceptive de la courbe de f0 et d'autres aspects phonétiques (il s'agit en quelque sorte de la phonétique perceptive de l'intonation). Les propositions 7 à 9 éclaircissent le statut linguistique de l'intonation, dans sa relation avec la syntaxe, l'accentuation et la signification ; enfin, la proposition 10 discute de l'intégration effectuée par le locuteur des aspects phonétiques et linguistiques de l'intonation à l'étape de planification du message et de ces caractéristiques prosodiques.

2.2.1. Procédure analytique

Il s'agit d'un modèle ascendant (*bottom-up*) de l'intonation dans la mesure où l'on part de la manifestation concrète de l'intonation sous la forme des tracés de la fréquence fondamentale et où l'on aboutit à un ensemble de catégories abstraites. La particularité de ces catégories abstraites réside dans le fait qu'elles ne sont pas imposées *a priori* mais sont ancrées dans la substance phonétique. À l'issue de l'analyse, on aboutit à une description qui rend compte de la structuration perceptive des traits intonatifs à partir d'un modèle de l'auditeur *(« the perceptual structure of intonation, captured in a model of the listener »*, *op.cit.* : 5).

Dans la démarche proposée par IPO, on part de l'étude des phénomènes acoustiques pour développer une méthode de réduction des données sur le critère de la tolérance

[40] Les auteurs procèdent par comparaison des méthodes à vocation linguistique, plus ou moins impressionnistes, dans la lignée du test des paires minimales en phonologie structuraliste segmentale, avec les méthodes instrumentales et acoustiques.

perceptive et de l'adéquation perceptive. Deux procédures analytiques sont définies : celle de l'égalité perceptive (*perceptual equality*) et celle de l'équivalence perceptive (*perceptual equivalence*). On procède par confrontation systématique de l'observation de la courbe de f0 avec les résultats d'une analyse perceptive : à l'issue de cette démarche, on procède à une stylisation (simplification) du contour mélodique adoptant la méthode d'une approximation linéaire par les segments de droite (*piecewhile linear approximation*)[41]. Une version stylisée de la courbe mélodique est donc soumise à la contrainte de contenir un nombre minimal de segments de droite et d'être perceptivement égale à celle de l'énoncé initial[42]. La procédure de stylisation est motivée par la considération que ce qui est important dans l'intonation consiste en des modifications systématiques et programmées de la courbe de f0 : à cette étape, on adopte, donc, une première dichotomie entre ce qui est pertinent en perception et ce qui relève de détails non-significatifs.

Une fois la première stylisation réalisée, on procède à une autre qui vise à remplacer la courbe stylisée par les mouvements mélodiques avec des spécifications standardisées de leurs différents paramètres (pente, amplitude etc.) : c'est à cette étape de l'analyse que l'on définit les rapports d'équivalence perceptive entre deux énoncés :

> *« If for a speech utterance two different courses of f0 are similar to such extent that one is judged as a successful imitation of the other, there is a perceptual equivalence between the two. »* (*op.cit.* : 47)

L'objectif d'une telle manipulation est d'aboutir à une analyse généralisée du contour mélodique attesté dans la parole en termes d'un nombre réduit de catégories de base. Il faut préciser que les catégories de base résultant des procédures de stylisation et de standardisation ne sont pas associées à une fonction distinctive (*op. cit.* : 51).

L'unité élémentaire de description dans le modèle IPO est donc un mouvement mélodique. Or, un mouvement mélodique peut recevoir également une spécification en termes de traits (perceptifs) : les différents mouvements montants et descendants sont décrits en termes de traits distinctifs binaires qui caractérisent leurs forme, amplitude, alignement par rapport aux frontières syllabiques, taille et vitesse de changement ([± *rise*], [± *early*], [± *late*], [± *spread*], [± *full*]).

Comme nous en avons discuté précédemment, les partisans de l'approche IPO excluent de leur inventaire d'unités phonologiques le concept de niveau mélodique. Pourtant, une modélisation du registre mélodique du locuteur s'y voit intégrée sous la forme d'une grille comportant la ligne des sommets (*topline*) et la ligne de base (*baseline*)[43,44]. Par ce

[41] Les variations de f0 pertinentes en perception des aspects mélodiques peuvent être modélisées comme les segments de droite dans l'espace log (f0) ~ temps.

[42] Dans leur démarche, les auteurs refusent une approche psychoacoustique, fondée sur le concept des seuils en perception, considérant justement que tous les éléments perceptibles de la mélodie en parole ne remplissent pas de fonction linguistique.

[43] Le modèle développé pour le hollandais se limite à deux niveaux préétablis, tandis que dans les modèles pour l'anglais et l'allemand la définition d'un niveau médian paraît nécessaire.

[44] Par rapport aux concepts de *topline* et *baseline* dans l'approche IPO, nous tenons à faire quelques remarques sur le traitement du phénomène de déclinaison dans ce modèle : il est considéré que la déclinaison étant un phénomène universel, elle a pour effet d'abaisser la f0 avec la progression de l'énoncé ; elle touche à la fois la ligne des sommets (*topline*) et la ligne de base (*baseline*).

fait le concept du niveau intonatif est reconnu dans le modèle, mais il n'a pas le statut d'une primitive descriptive.

2.2.2. Grammaire intonative de l'approche IPO

L'objectif final de l'approche IPO est de proposer une grammaire générative qui spécifiera quels mouvements mélodiques peuvent être utilisés à un moment donné et permettra donc de générer l'ensemble des contours grammaticaux dans une langue, bloquant la production des contours agrammatiques. Pour aboutir à une grammaire intonative[45], après avoir établi l'inventaire des mouvements mélodiques perceptivement pertinents, la distribution de ces derniers est étudiée dans un corpus d'études[46].

Les mouvements mélodiques élémentaires (les unités descriptives de base dans le modèle) peuvent former des unités plus larges, des configurations mélodiques : ainsi, deux mouvements mélodiques forment-ils une configuration si les possibilités combinatoires à l'intérieur de la configuration sont plus réduites qu'à ses frontières. Dans l'étape suivante de l'analyse, les configurations mélodiques s'enchaînent pour former un contour mélodique[47] selon les contraintes syntagmatiques. De cette analyse de la distribution de différentes configurations mélodiques, une observation importante est tirée : les auteurs observent que par rapport à l'unité intonative, on ne retrouve certaines configurations mélodiques qu'à ses frontières ; d'autre part, il y a des mouvements mélodiques récursifs ou au contraire optionnels. Par conséquent, on définit trois classes paradigmatiques de configurations mélodiques par rapport à leurs rôles structurels dans un contour mélodique, les *Préfixes*, les *Racines* et les *Suffixes* :

- les *racines* sont des mouvements/configurations mélodiques obligatoires, ce sont elles qui déterminent l'identité du contour mélodique. Tout contour mélodique ne contient qu'une seule configuration mélodique de cette classe, configuration qui est de ce fait non récursive ;

- les *préfixes* sont les configurations mélodiques qui précèdent la Racine ; elles sont optionnelles et peuvent être récursives ;

- enfin, les *suffixes* caractérisent l'organisation mélodique après la Racine et sont optionnels et non récursifs.

Contour ➞ (Préfixe)^{n} Racine (Suffixe)

[45] Interprétée comme un ensemble de contraintes sur la concaténation des mouvements mélodiques de base.
[46] Par la suite, la première version de la grammaire intonative est confrontée à une deuxième partie de données pour une vérification de sa couverture et de la force prédictive des règles établies.
[47] Un contour mélodique est défini comme une séquence des mouvements mélodiques élémentaires co-extensive avec une unité intonative qui correspond, dans l'approche IPO, à une *clause*. Ce choix est motivé par le fait qu'une clause en tant qu'unité linguistique y obtient une définition plus précise que d'autres unités attestées dans la littérature : *groupe de souffle*, *unité d'information*, ou *unité de sens.*

Une deuxième typologie que 't Hart et ses collègues appliquent dans leur analyse subdivise les configurations mélodiques selon leur tendance à être associée avec une syllabe métriquement forte. Deux classes en résultent : d'une part, on définit un ensemble de configurations mélodiques 'contribuant à la proéminence' d'une syllabe (*prominence lending pitch movements,* hérités de la tradition britannique, Bolinger 1958). D'autre part, la classe opposée regroupe les configurations privées du pouvoir de contribuer à une proéminence supplémentaire du mot : ce sont les mouvements marquant les frontières prosodiques[48] ou bien les mouvements qui englobent plusieurs syllabes (les transitions).

Pour chaque contour mélodique, on détermine également la famille à laquelle il appartient : une famille résulte du groupement des contours mélodiques en des catégories sous-jacentes (les patrons intonatifs). Subséquemment, un contour mélodique est dans cette approche une structure phonétique concrète qui est reliée au patron mélodique abstrait ; c'est le patron intonatif qui détermine la nature (forme) et l'ordre des configurations mélodiques dans le contour. La classification des contours en patrons mélodiques est fondée sur les propriétés de la configuration mélodique de Racine ; les différences en nombre de Préfixes ou par rapport à la présence ou absence des Préfixes ou de Suffixes ne sont pas prises en compte. De ce fait, le modèle IPO se rapproche de la tradition britannique avec la classification des contours en termes des tons nucléaires.

Pour résumer, dans l'approche IPO, une courbe de f0 se voit « remplacée » par une série de mouvements mélodiques stylisés qui forment des constructions mélodiques plus larges (les configurations mélodiques, les contours). Dans cette lignée, l'infinité des contours mélodiques est la manifestation d'un nombre limité des catégories sous-jacentes.

Nous nous arrêtons brièvement sur les présupposés linguistiques adoptés par les auteurs du modèle IPO. Le composant linguistique du modèle étudie, d'une part, le rapport entre le contour mélodique concret et la famille des contours mélodiques à laquelle il appartient ; d'autre part, du domaine de la linguistique relèvent les problématiques de l'articulation des interfaces « intonation ~ accentuation » et « intonation ~ syntaxe ». De manière générale, la problématique des rapports « intonation ~ accentuation » et « intonation ~ syntaxe » intervient à l'étape d'analyse de l'association entre le contour mélodique et le contenu segmental[49]. L'interface « syntaxe ~ prosodie » reçoit un traitement standard, avec une constatation de l'absence d'un lien biunivoque entre les structures syntaxiques et mélodiques ; or, le traitement des rapports entre l'accentuation et l'intonation nous paraît mériter une précision.

Dans l'analyse distributionnelle, on remarque, entre autres, que certains mouvements mélodiques déterminés correspondent toujours aux syllabes accentuées tandis que d'autres n'y correspondent jamais ; ce phénomène peut recevoir une explication soit en termes de co-occurrences entre les processus relevant respectivement du domaine de

[48] 't Hart et ses collègues n'associent pourtant pas certains mouvements mélodiques avec la structuration prosodique de l'énoncé en termes de domaines ; remarquons que l'unité associée à un contour mélodique dans leur approche est une clause (grammaticale).

[49] On reconnaît également l'importance d'influences pragmatiques sur l'organisation mélodique des énoncés, mais on considère que l'influence de la pragmatique est prise en compte à l'étape du choix entre différentes formes mélodiques, distribution des accents, etc.

l'intonation ou du domaine de l'accentuation, soit par intermédiaire de l'adoption de la notion de la proéminence mélodique. Dans la première interprétation, on considère qu'un contour mélodique est une séquence de mouvements mélodiques qui relèvent ou du domaine de l'accentuation ou du domaine de l'intonation mais jamais des deux à la fois. Or, dans la deuxième interprétation, intégrée dans le modèle IPO, on stipule que l'intonation préside à l'accentuation dans la détermination de la forme du contour mélodique ; dans un deuxième temps, il est admis que la distribution et l'alignement des mouvements mélodiques suivent les contraintes imposées par l'accentuation : les mouvements mélodiques seront donc associés à des syllabes accentuées des mots qui doivent recevoir une proéminence mélodique. Dans cette perspective, l'approche IPO ne peut être considérée comme complètement linéaire : on y retrouve des mécanismes propres aux modèles superpositionnels.

Le système IPO, orienté vers la perception de la parole, se propose de décrire des phénomènes perceptivement significatifs, en termes de mouvements mélodiques élémentaires. Ces derniers sont supposés caractériser le système intonatif de la langue. Pourtant, on observe que le nombre des mouvements mélodiques distingués dans la perception est très large, et que l'application de cette méthodologie dépourvue d'une analyse fonctionnelle concomitante conduit à une précision exagérée de la description (cf. les résultats obtenus dans les études expérimentales récentes par Gussenhoven et Rietveld 2000).

Les descriptions adoptant l'approche de l'école hollandaise de l'intonation ont été élaborés pour le hollandais, l'anglais (de Pijper, 1983), l'allemand et le russe (travaux de Cécilia Odé, 1989, que nous présentons plus loin).

Comme nous l'avons mentionné dans l'introduction, l'approche IPO est la seule approche des faits intonatifs orientée perception qui adopte la démarche analytique ascendante. Nous nous tournons maintenant vers la présentation du système d'annotation prosodique ToBI, fondé sur le travail de Pierrehumbert (1980), première tentative d'analyse autosegmentale-métrique des faits intonatifs.

2.3. TOBI : TONES AND BREAKS INDICES

Le système d'annotation ToBI devient un standard dans les recherches prosodiques et a été déjà appliqué à l'analyse de plusieurs langues (Jun, 2005).

Ce système d'annotation est le résultat d'un travail collectif de chercheurs américains issus de différentes branches de la recherche théorique et appliquée, dont les intérêts scientifiques ont porté sur l'analyse de l'intonation et la modélisation du phrasé prosodique. Pour l'analyse de l'organisation tonale ce système puise ses idées-clés dans le travail de thèse de Janet Pierrehumbert (1980) et les précisions et développements apportés par des travaux postérieurs (Beckman et Pierrehumbert, 1986, Pierrehumbert et Beckman, 1988).

2.3.1. Modèle de J. Pierrehumbert (1980 et postérieurs)

La description de l'intonation chez Pierrehumbert comprend trois sous-parties : une grammaire des contours mélodiques associés à un syntagme intonatif, cette grammaire ayant la forme d'un automate à état fini ; une représentation métrique du texte ; un ensemble des règles d'association assurant l'implémentation phonétique des étiquettes de la description phonologique :

> *« The phonological characterization of intonation has three components. The first is a grammar of allowable phrasal tunes. This grammar generates sequences of L and H tones (...) The second component is a metrical representation of the text. (...) Lastly, we have rules for lining up the tune with the text. The complete phonological representation for intonation is thus a metrical representation of the text with tones lined up in accordance to the rules. »* (Pierrehumbert 1980:10)

La description de l'organisation tonale de l'énoncé n'utilise que deux tons mélodiques Haut (H, *high*) et Bas (L, *low*), qui participent à la spécification des configurations de trois types d'événements mélodiques : proéminences mélodiques, accents de syntagme et tons de frontière.

Les accents mélodiques peuvent être mono- et bitonaux (composés de deux tons H et L liés par le signe '+'). Simultanément, outre la spécification en termes de tons, les accents mélodiques reçoivent une spécification pour le paramètre qui contrôle l'association entre le ton et l'unité porteuse du ton : l'astérisque marque celui des tons de l'accent bitonal qui est associé avec la syllabe métriquement forte. De cette différence d'association provient la distinction entre les configurations mélodiques reliées aux étiquettes L+H* et L*+H : dans le premier cas, c'est le ton haut qui est réalisé simultanément avec la syllabe accentuée du mot, le ton L étant réalisé à une distance fixe par rapport au ton principal[50].

Dans la perspective d'une analyse comparative de différents systèmes d'annotation prosodique qui est la nôtre, il importe de retenir que, dans l'approche de Pierrehumbert, la notion de ton se dissocie de la conception introduite par Bruce (1977), dans la mesure où cette dernière interprète toute étiquette de ton comme spécification d'une cible mélodique. Pour Bruce, tous les tons ont une implémentation phonétique sous la forme d'un point tournant dans le contour mélodique (donc, chaque ton peut être spécifié dans les coordonnées « fréquence – temps »). La position de Pierrehumbert est différente : premièrement, dans l'inventaire des accents mélodiques proposé, on retrouve l'accent H*+L pour lequel L ne reçoit pas d'implémentation phonétique mais est présent dans la spécification phonologique pour assurer l'abaissement du ton H suivant. Ensuite, les vallées et sommets visibles du contour mélodique ne reçoivent pas obligatoirement de spécification tonale (par exemple, une vallée entre deux accents H* qui est le résultat d'une interpolation non linéaire entre les deux, Pierrehumbert, 1980 :379 et 381, figures 5.17 et 5.19).

[50] Nous utilisons la notion de ton principal dans la mesure où il est considéré que les deux tons d'un accent bitonal maintiennent les mêmes rapports de force qu'on retrouve entre les syllabes accentuée et inaccentuées dans un pied. La notion de la force métrique pour les tons est analysée par Grice (1995)

La description proposée par Pierrehumbert et ses collègues n'utilise que deux tons. Simultanément, pour un fonctionnement productif de ce modèle, on y introduit le mécanisme de catathèse (les termes synonymiques étant le *downstep* ou l'abaissement tonal). On conçoit le catathèse comme compression et abaissement du registre mélodique, induit par la présence d'un ton bas dans le contexte de deux tons hauts (l'abaissement tonal est donc spécifique au contexte H L H). Dans ce contexte mélodique, le deuxième ton H est réalisé plus bas dans le registre mélodique par rapport à ce qui pourrait être attendu à ce point de l'énoncé[51,52].

L'introduction des tons de frontières dans la représentation de l'organisation mélodique est une conséquence directe du présupposé du modèle selon lequel la représentation phonologique de la constituance prosodique retrouve une manifestation à l'oral dans la forme de surface de l'énoncé. En même temps, au moins un niveau de cette constituance est défini en termes de propriétés mélodiques[53].

On reconnaît également la troisième classe d'événements mélodiques : les accents de syntagme. La fonction de l'accent de syntagme est de rendre compte des détails du contour mélodique après le dernier accent mélodique (accent nucléaire) jusqu'au ton de frontière. Le concept d'accent de syntagme est très controversé et induit beaucoup de débats théoriques. Ainsi, Taylor (2000) considère que les accents de syntagme ne sont que des entités purement phonologiques : c'est le mécanisme permettant au modèle de distinguer quatre configurations de la courbe de f0 après l'accent mélodique nucléaire (avec les configurations L- L%, L- H%, H- L%, H- H%) et très fréquemment, aucune réalisation concrète ne leur correspond dans la courbe de f0. Historiquement, l'accent du syntagme dans le modèle de Pierrehumbert (1980) fait écho à l'accent de l'énoncé (*sentence accent*) introduit par Bruce (1977) dans sa description de l'organisation tonale de l'énoncé en suédois : pourtant, comme le remarque Gussenhoven (2002b), sur le plan fonctionnel, l'accent d'énoncé de Bruce est équivalent à l'accent mélodique comme il est défini dans l'approche pierrhumbertienne. Dans l'analyse proposée par Grice *et al*. (2000), l'accent du syntagme est un ton de frontière avec une double association[54]. Cette analyse explique entre autres pourquoi l'accent de syntagme a tendance à être associé, dans certaines langues et dans certains contextes, à une syllabe accentuée à la proximité du dernier accent mélodique du syntagme ; c'est le cas du focus étroit pour l'anglais britannique, ou bien de la question à morphème interrogatif en grec par contraste avec le grec de Chypre.

[51] Ladd (1983) propose un traitement alternatif d'abaissement tonal, à l'aide d'un trait binaire [+*downstepped*].

[52] Le mécanisme d'abaissement tonal est un des points qui distingue les modèles autosegmentaux des modèles superpositionnels : dans les premiers, le *downdrift* n'est pas modélisé en termes de patron global associé avec le syntagme, mais est géré par la combinaison des mécanismes d'abaissement des accents mélodiques et par un rétrécissement du registre mélodique.

[53] Il faut préciser qu'avec l'élaboration des descriptions autosegmentales de l'intonation pour les langues autres que l'anglais, certains modèles adoptent des tons de frontières complexes, par exemple, le ton de frontière LH en coréen qui signale la frontière d'un syntagme accentuel (Jun 1993).

[54] Pierrehumbert et Beckman (1988) introduisent le concept d'une double association. Certains accents de syntagme, à part être associé avec un constituant dans l'hiérarchie prosodique, peuvent recevoir une deuxième association avec une syllabe porteuse du ton précis

À partir d'éléments que nous venons de présenter, Pierrehumbert construit une grammaire intonative permettant de générer les contours mélodiques bien formés de l'anglais américain (figure 2.2). Le contour mélodique dans le modèle de Pierrehumbert contient donc au moins un accent mélodique, un accent de syntagme et un ton de frontière terminal (et de façon optionnelle, un ton de frontière initial).

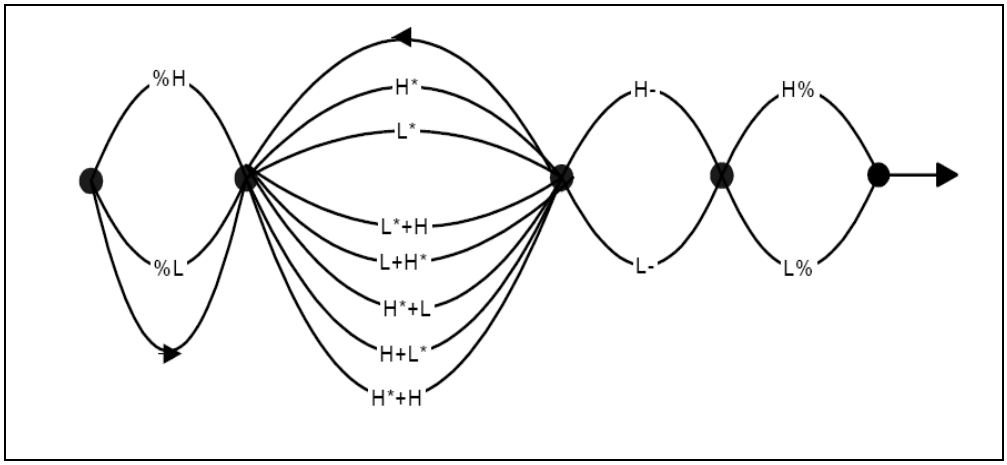


Figure 2.2 : Grammaire intonative de l'anglais américain (Pierrehumbert, 1980)

Il est important de préciser que contrairement à la tradition britannique qui procède par une distinction systématique entre les proéminences mélodiques nucléaires et pré-nucléaires en termes d'inventaires de configurations autorisées, cette opposition n'est pas maintenue dans le modèle de Pierrehumbert (cf. étude de Pierrehumbert et Steele 1989). Cependant, l'étude de Dainora (2001) remet en cause l'indépendance des choix des catégories d'accents mélodiques et de tons de frontière dans une séquence.

Dans le cadre de l'approche que nous venons d'ébaucher, si l'on envisage sa transposition dans les modèles appliqués, le concept d'association nécessite une analyse plus fine dans le cadre d'une réflexion sur l'implémentation phonétique des étiquettes phonologiques. Nous en donnons quelques détails dans la section qui suit.

2.3.2. Association et alignement dans le modèle de Pierrehumbert

Dans la section précédente nous avons présenté la grammaire intonative proposée par Pierrehumbert, et nous avons essayé d'expliciter les présupposés théoriques qui sous-tendaient l'élaboration de ce modèle. Nous avons également brièvement discuté des aspects d'organisation métrique des énoncés tels qu'on les appréhende dans la tradition métrique. Nous nous tournons maintenant vers le concept central des approches autosegmentales-métriques, troisième facette du modèle de l'intonation de Pierrehumbert (1980), à savoir la problématique d'association et d'alignement[55].

[55] La littérature sur le sujet étant très abondante, nous ne prétendons pas à l'exhaustivité de l'analyse présentée : nous nous donnons la tâche d'ébaucher les principes théoriques sous-jacents à la problématique et non pas de faire l'état de question sur les travaux portant sur différentes langues du monde.

Nous tenons à préciser la distinction faite entre les concepts d'association et d'alignement dans les recherches intonatives actuelles. Sous le terme d'association on entend un processus abstrait, conçu au niveau phonologique de la représentation, de la mise en relation du niveau tonal et du contenu segmental. De son côté, l'analyse d'alignement met en évidence les détails phonétiques de cette association : on se penche sur la corrélation entre le sommet (ou la vallée) mélodique et le début de la syllabe/du noyau syllabique/de la partie sonore de la syllabe (van Santen *et al.* 1998). La problématique d'alignement a reçu deux traitements dans les études phonologiques de l'intonation : d'une part, on analyse l'alignement comme une manifestation au niveau phonétique de l'association phonologique, exprimée à l'aide de diacritiques dans le système d'étiquetage proposé par Pierrehumbert et ses successeurs. Une autre description possible présuppose l'introduction d'un trait binaire associé au ton ou à la configuration mélodique, par exemple, Ladd (1983) propose le trait [+*delayed peak*] (on retrouve d'autres analyses sous la forme de traits chez Bruce 1977, Gussenhoven 1984, 't Hart et al. 1990).

La relation structurelle exprimée par le lien d'association entre un événement mélodique et le niveau segmental se traduit au niveau phonétique par une coïncidence des domaines temporels de deux éléments. La coordination spécifique de cette coïncidence temporelle est reléguée au domaine de l'alignement phonétique ; selon Pierrehumbert et Beckman « *the left-peripheral, central, and right peripheral alignments cannot be distinguish by association relation alone* » (Pierrehumbert et Beckman, 1988 : 153). L'alignement phonétique prend donc la forme d'un ensemble de règles de l'interprétation phonétique qui sont spécifiques à chaque langue.

La question du domaine d'alignement se pose : au niveau phonologique, une association est établie entre un événement mélodique et une syllabe métriquement forte ; cette information est-elle suffisante pour traiter la problématique de l'alignement ? Doit-on limiter le domaine d'analyse à la seule syllabe proéminente, ou bien a-t-on besoin de prendre en considération les syllabes adjacentes, ou bien est-ce le pied accentuel ou le mot prosodique qui constitue le domaine d'alignement ? Comme le démontrent Hermes *et al*. (1997), pour qu'un mouvement mélodique contribue à la proéminence perceptive, il ne doit pas être obligatoirement accompli dans les limites de la syllabe métriquement forte. Arvaniti *et al.* (2000) analysent l'alignement des tons d'un accent montant bitonal pré-nucléaire en grec et s'aperçoivent qu'aucun des deux tons n'est réalisé simultanément avec la syllabe métriquement forte (cf. également Ladd 1996 sur les accents mélodiques de l'anglais à Glasgow).

Les fonctions linguistiques de l'alignement sont à la fois universelles et spécifiques à la langue. Ainsi D'Imperio (2002) démontre que l'alignement est contrastif en italien napolitain car ce sont les différences dans l'alignement de la séquence LH qui distinguent les assertions et les interrogations dans ce dialecte. De manière générale, l'alignement phonétique définit l'appartenance d'un accent mélodique à une catégorie phonologique établie (Pierrehumbert & Steele 1989, D'Imperio & House 1997). Pourtant, cette position a éveillé de nombreuses critiques. Taylor (2000) considère, entre autres, qu'une description assidûment catégorielle, que l'on trouve dans le travail de Pierrehumbert et ses successeurs élude une question-clé de la définition des

frontières entre les catégories. Ainsi dans la prononciation soignée, tous les types d'accents mélodiques définis en théorie se différencient facilement en ce qui concerne leurs réalisations phonétiques, or, dans le discours libre les frontières entre les catégories accentuelles deviennent floues. Pour les successeurs de Pierrehumbert, pourtant, cette remarque n'est pas pertinente, étant donné leur interprétation de la notion de la cible tonale : un accent mélodique peut être annoté même dans le cas d'absence d'une excursion de f0 dans le signal.

2.3.3. Annotation ToBI

Le système d'annotation prosodique ToBI s'inspire largement de l'approche cernée de l'organisation mélodique et nous entreprenons maintenant sa présentation proprement dite. Toute transcription ToBI contient six composants, à savoir un fichier son, une courbe brute de f0, et 4 niveaux d'annotation pour les accents mélodiques, la transcription orthographique, l'annotation du degré de jonction entre les mots adjacents (*break indices*), et les commentaires éventuels.

ToBI suit l'hypothèse de l'approche autosegmentale métrique, selon laquelle les accents mélodiques marquent les syllabes métriquement fortes : on applique ce principe même dans les cas où la cible mélodique est alignée avec la syllabe inaccentuée suivante (cf. notre analyse des problèmes d'alignement dans la section précédente). Pour signaler ces discordances, ToBI dispose des diacritiques '<' et '>' qui marquent les syllabes avec lesquelles la cible tonale est effectivement alignée.

Contrairement à la grammaire intonative proposée par Pierrehumbert, le système ToBI dans la version développée pour l'anglais américain, comporte cinq accents mélodiques (2 accents monotonaux, H* et L*, et 3 accents bitonaux, L+H*, L*+H, H+!H), deux accents de syntagme (L- et H-) et trois tons de frontières (L%, H%, et un ton optionnel initial %H). Nous résumons les descriptions de chaque accent dans le tableau ci-dessous (la grammaire intonative associée avec ToBI est présentée sur la figure 2.3). Les accents mélodiques bitonaux présentés ci-dessus peuvent se manifester sous la forme abaissée suite aux effets de catathèse : ce détail d'implémentation phonétique est indiqué à l'aide de diacritique ' !'.

Accent mélodique	Description
H*	*Peak accent* – une cible mélodique apparente sur la syllabe accentuée, située dans la partie supérieure du registre mélodique du locuteur pour le syntagme en question. L'annotation peut s'appliquer aux fréquences médianes dans le registre du locuteur, mais jamais pour les basses fréquences.
L*	*Low accent* – une cible mélodique apparente sur la syllabe accentuée dans la partie inférieure du registre mélodique du locuteur

L+H*	*Rising peak accent* – une cible mélodique haute (sommet) sur la syllabe accentuée précédée immédiatement par une montée rapide d'une vallée basse dans le registre du locuteur
L*+H	*Scooped accent* – une cible mélodique basse sur la syllabe accentuée suivie immédiatement d'une montée rapide jusqu'à un pic dans la partie supérieure du registre mélodique du locuteur.
H+!H*	Un abaissement marqué sur la syllabe accentuée à partir d'un ton haut qui, lui, ne peut pas être pris en compte par un accent de phrase du syntagme précédent H ni par un accent mélodique H* dans le même syntagme.

Remarquons qu'avec la distinction faite entre les accents de syntagme et les tons de frontière, le système ToBI intègre une vision de la structuration prosodique de l'oral avec deux niveaux de constituance définis en termes mélodiques[56] : tout énoncé est structuré en termes de syntagmes intermédiaires et, au niveau supérieur, en termes de syntagmes intonatifs. Le syntagme intermédiaire contient obligatoirement une proéminence mélodique ; il peut en contenir d'autres (optionnelles) et se termine par un accent de syntagme. Un syntagme intonatif contient un ou plusieurs syntagmes intermédiaires et son marquage formel est assuré par la présence d'un des deux tons de frontière (H% ou L%) : l'hypothèse de la hiérarchie stricte a pour conséquence le fait que les frontières des syntagmes intonatifs sont signalées conjointement par l'accent de syntagme et le ton de frontière.

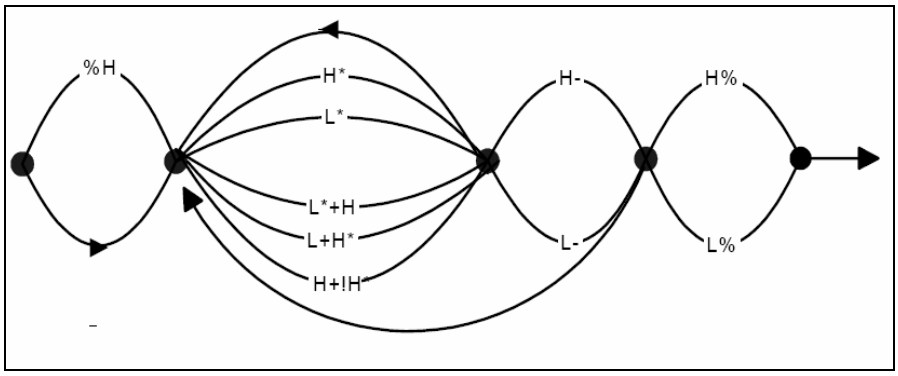


Figure 2.3 : Grammaire intonative ToBI

2.3.4. Annotation du degré de jonction (indices de frontières)

Dans le contexte de notre travail, un développement approfondi est nécessaire pour l'annotation du phrasé prosodique qui reçoit dans le modèle ToBI la forme d'une échelle de jonction entre deux mots adjacents ; on annote alors la force perçue de la frontière sur une ligne réservée dans le schéma global d'annotation. On utilise une échelle de cinq : *l'indice 4* correspond à une frontière de syntagme intonatif ; *l'indice 3*

[56] Pourtant, la hiérarchie prosodique proposée par Pierrehumbert et Beckman est, en soi, tonale : on cherche à définir les constituants en termes d'organisation tonale de l'énoncé.

à une frontière du syntagme intermédiaire ; *l'indice 1* à une frontière entre deux mots non-terminaux du syntagme ; *l'indice 0* à une frontière entre un mot clitique et son hôte. Le plus problématique, à notre avis, est *l'indice 2* : on se sert de cette étiquette dans les cas où une disjonction forte est perçue entre deux mots, correspondant à une frontière du syntagme intermédiaire/intonatif, mais laquele n'est pas accompagnée par des phénomènes tonals associés (accent du syntagme et ton de frontière)[57].

Le système d'annotation ToBI se veut cohérent et reproductible : cet objectif a servi de motivation pour réaliser un ensemble d'études sur la concordance entre les étiquettes attribuées par différents annotateurs (cf. Syrdal et MacGory 2000). Ces études suivent la même procédure expérimentale et analytique : un corpus (contenant les enregistrements de différents locuteurs avec une appartenance dialectale variée, s'exprimant en différents styles de parole) est soumis à une annotation prosodique pour le système ToBI par un groupe de transcripteurs. L'objectif de ces études d'accord inter-transcripteur est de démontrer si le système d'étiquettes et les exemples développés pour l'entraînement permettent une annotation cohérente et coordonnée d'une large base d'enregistrements. Les étiquettes attribuées à chaque mot sont ensuite comparées pour toutes les paires de transcripteurs et les pourcentages d'accord observés ainsi que les confusions fréquentes sont ensuite analysés. Notons que l'intérêt est surtout porté sur les confusions entre les différents types d'accents mélodiques ; les confusions entre différents indices de frontières reçoivent moins d'attention.

Une autre étude (Syrdal *et al.* 2001) est réalisée pour évaluer de manière expérimentale dans quelle mesure une annotation semi-automatique ToBI[58] accélère l'annotation du corpus. Le système d'annotation automatique aboutit à une annotation avec catégories par défauts : tous les accents mélodiques reçoivent l'étiquetage H* (sur les grands corpus annotés, l'étiquette H* est attribuée à 79% d'accents mélodiques, Taylor 2000) et on n'annote que les frontières des syntagmes intonatifs se servant des labels L-H% pour une frontière du syntagme intonatif interne et L-L% pour une frontière terminale. Donc, on procède par une prédiction binaire de la distribution des proéminences mélodiques et des frontières à partir du texte, les informations mélodiques n'étant pas prises en compte : on cherche les réponses aux questions

« ce mot, est-il doté d'une proéminence mélodique ? »

et

« ce mot, est-il suivi d'une frontière intonative ? ».

Dans l'étude comparative entre l'annotation du signal et la correction de l'annotation automatique, la correction des labels déjà attribués par défaut accélère le processus de transcription. Pourtant les labels par défaut peuvent créer des biais dans les décisions des annotateurs : comme les auteurs le remarquent, la précision de l'annotation par défaut varie selon les corpus et les styles de parole, ainsi que selon le type d'événement à annoter.

Collin Wightman (2002), faisant référence au contexte historique d'apparition du système ToBI dont il est un des concepteurs, souligne que ToBI a surgi au moment d'un

[57] Le système d'annotation du degré de jonction intégré dans ToBI émane des recherches menées par Patty Price et ses collègues (Price et al. 1991).
[58] Il s'agit d'une annotation automatique corrigée manuellement par la suite.

intérêt croissant pour l'utilisation de bases d'enregistrements attestés dans des buts applicatifs (dans les systèmes de synthèse ainsi que de reconnaissance vocale). Dans cette perspective, la reproductibilité de l'annotation ToBI a eu valeur de standard. Or, au fil du temps, ce système d'annotation est devenu très dépendant de la théorie intonative sous-jacente : par exemple, sous les indices de rupture on visait initialement une annotation perceptive du degré de disjonction/cohésion entre mots adjacents. Pourtant, dans sa version actuelle, les indices de frontières déterminent un ensemble de constituants prosodiques et ont perdu tout lien avec la perception. Également, le lien entre les spécifications tonales et les indices de frontière n'était pas visée initialement : savoir dans quelle mesure cet aspect affaiblit ToBI en tant que système de transcription, nécessite une étude ciblée. D'autre part, pour certaines catégories d'accent mélodiques on retrouve une consistance inter-annotateurs assez faible (au-dessous de 50%), ce qui empêche l'entraînement des algorithmes d'annotation automatique.

Dans un article-bilan du travail sur la transcription intonative, Beckman *et al.* (2005) insistent sur l'idée qu'actuellement le sigle ToBI possède deux significations, l'une faisant référence au système d'annotation de l'intonation développé pour l'anglais américain, l'autre englobant tous les systèmes d'annotation de l'intonation, développés pour un ensemble de langues qui ont adopté les principes fondateurs de ToBI et les hypothèses sous-jacentes à l'analyse autosegmentale de l'intonation :

> *« The term ToBI has come to be used in two different ways. Originally, it was the name of an annotation system, developed in the period 1991 to 1994, which was designed for use in labelling intonation and prosody in databases of spoken Mainstream American English (Beckman and Hirschberg 1994). Very quickly, however, it also came to refer to a general framework for the development of prosodic annotation systems in other varieties of English (e.g,, Mayo, Aylett and Ladd 1997 [Glasgow]) and in other languages (e.g., Grice, Reyelt, Benzmüller and Batliner 1996 [German]; Venditti 1997 [Japanese]). »*

Un autre axe du développement des principes fondateurs de ToBI a été exploité dans les études appliquées, avec la réduction de l'ensemble de catégories utilisées. Un des exemples à citer est le système *ToBI Light* (Wightman *et al.* 2000) : ce système se donne pour but de traduire sous la forme d'étiquettes les détails du fonctionnement de la perception humaine dans des situations d'interaction verbale. Il ne s'agit plus d'identifier les constituants prosodiques préétablis : ainsi la proéminence des syllabes est annotée sur une échelle de trois degrés, et on procède par jugement binaire si un mot est suivi d'une frontière prosodique. Le système a été testé dans les tâches d'annotation d'un corpus à la base du système de synthèse avec la sélection des unités : la méthode se révèle performante lors de l'évaluation perceptive de la qualité de la parole synthétisée.

Après avoir discuté de la logique interne du système ToBI qui a presque acquis le statut de standard universel dans la recherche intonative, nous procédons à une discussion succincte des systèmes d'annotation s'inspirant de ToBI qui conceptualisent différemment l'interface entre la représentation phonétique et la représentation phonologique de l'énoncé : cette démarche nous permettra ainsi de faire un bilan des critiques émises à l'égard de ToBI.

2.4. TODI : TONAL DESCRIPTION OF DUTCH INTONATION (GUSSENHOVEN, 2005)

Le système proposé pour la description de l'intonation du hollandais s'inscrit dans le paradigme autosegmental mais, à l'instar des systèmes ToBI, ne transcrit que l'organisation tonale de l'énoncé :

> *« The primary purpose of ToDI is to make a transcription system available for characterising the intonation of Dutch utterances or example sentences, including large amounts of spontaneous speech, from which subsequently generalisations about accentuation, phrasing, and tone choice could be extracted ».*

Le système ToDI a été conçu dans l'objectif de retranscrire dans le cadre autosegmental tous les contrastes mélodiques attestés dans les travaux de 't Hart *et al.* (1990), de Gussenhoven (1988) ainsi que les différences attestées entre les mouvements ascendants haut et bas (Gussenhoven et Rietveld 2000). Ce schéma d'annotation garde tout autant une ressemblance avec le système d'annotation proposée par Gussenhoven précédemment (1988, 1991).

Dans la quête d'un système d'annotation prosodique s'inscrivant dans le cadre autosegmental, la question de savoir pourquoi ne pas utiliser un cadre standardisé de ToBI est légitime. Gussenhoven présente deux objections. Premièrement, il n'est pas considéré comme pertinent de postuler l'existence de deux niveaux de constituants avec le marquage tonal[59]. Deuxièmement, l'inventaire des accents mélodiques que l'on trouve dans ToBI n'est pas suffisant pour la description de l'intonation du hollandais, un nombre important de contrastes mélodiques n'étant pas pris en compte.

Aux niveaux supérieurs de la hiérarchie prosodique, le modèle sous-jacent de ToDI postule la structuration prosodique de l'oral en hollandais en termes d'énoncés, de syntagmes intonatifs, et de syntagmes phonologiques[60]. Le constituant du syntagme intonatif correspond, dans sa définition, au groupe tonal de la tradition britannique. Ce sont donc les tons de frontière initial (%T) et final (T%) qui délimitent ce constituant. Dans l'approche de Gussenhoven, le syntagme intonatif est le seul constituant dont les frontières sont signalées dans l'organisation tonale du message par les tons de frontière ; le constituant du syntagme intermédiaire se révèle non-pertinent[61].

---

[59] La conception de la structure tonale à deux niveaux (avec la distinction entre syntagme intonatif et syntagme intermédiaire) proposée par Beckman et Pierrehumbert (1986) est présente dans le système ToBI dans l'annotation séparée de l'accent de syntagme et de ton de frontière.

[60] Inférieurs au syntagme phonologique se situent les niveaux du mot prosodique, du pied (trochaïque) et de la syllabe.

[61] Dans les études précédentes sur le hollandais (van den Berg, Gussenhoven and Rietveld 1992, Gussenhoven et Rietveld 1992), il a été considéré que l'énoncé recevait également un marquage mélodique spécifique, étant donné que c'est le seul constituant qui peut se terminer avec des tons infra-bas ou extra-haut. Dans le modèle ToDI, cette analyse est réfutée, étant donné le caractère applicatif visant la synthèse vocale des études précédentes : les observations ont été faites avec la simulation du style de parole lue et le modèle d'un locuteur idéal ; les observations à partir de la parole spontanée contredisent cette observation.

2.4.1. Tons de frontières

On reconnaît trois configurations possibles en association avec les syllabes inaccentuées initiales d'un syntagme intonatif. On distingue au niveau phonologique la prononciation avec une f0 médiane ou basse (il s'agit d'une option neutre, codée %L), la prononciation avec une f0 haute (option marquée, codée %H), et la prononciation avec un ton descendant (configuration très rare et par conséquent, très marquée, qui reçoit le codage %HL).

En outre, trois options sont envisagées (et attestées) pour la frontière finale du syntagme intonatif, qui peut se terminer avec un ton haut (H%), bas (L%)[62] ou bien ne pas être spécifiée par rapport à la configuration tonale (0%)[63]. C'est cette possibilité de ne pas être spécifiée par rapport à l'organisation tonale qui suggère l'idée du caractère optionnel du marquage mélodique des frontières terminales des syntagmes intonatifs. Cette réflexion corrobore le principe général de ToDI, adopté dans la définition des cibles mélodiques annotées en transcription :

> *« […] ToDI uses no abstract tones of the sort L-H% to mean 'mid level pitch', but has only tones for which individual phonetic targets can be identified. »*

On admet également l'existence des syntagmes mélodiques désaccentués, qui ne contiennent aucun accent mélodique et correspondent aux éléments externes à l'énoncé dans la terminologie de Astruc (2003) (il s'agit des vocatifs, des insertions parenthétiques, des appositions, de certains adverbes, des relatives non-restreintes , et du discours rapporté[64]).

2.4.2. Les accents mélodiques

Les accents mélodiques sont le deuxième type de phénomènes mélodiques qui reçoivent une annotation dans le cadre de ToDI. Les accents mono- et bitonaux sont décrits ; or, on y voit intégré le concept de l'association tonale développé dans les travaux précurseurs de Gussenhoven (1988, 1990). Ainsi, à la différence de ToBI, on se passe du concept du ton associé à gauche (*leading tone*, le H dans l'accent H+!H*) dans les accents bitonaux. Dans les systèmes s'inspirant de ToBI, le ton H de l'accent H+!H* est utilisé pour signaler, entre autres, le ton haut d'une syllabe pré-accentuée. Or,

[62] Par opposition au travail précédent de Gussenhoven (1988), le statut particulier du ton de frontière final est reconnu : dans les analyses précédentes, la courbe mélodique ne recevait que la spécification en termes d'accents mélodiques, la spécification du ton de frontière étant considéré comme automatique.

[63] Une argumentation développée en faveur de l'introduction du label 0% dans l'annotation des configurations mélodiques aux frontières est présentée chez Grabe (1998). En anglais américain, pour lequel le système ToBI a été développé, les configurations descendantes ne sont pas possibles aux frontières des syntagmes intonatifs, le ton reste au même niveau par rapport à l'accent de phrase (L%) ou bien monte (H%). Par contre, dans d'autres langues (en particulier, pour l'anglais irlandais) les trois configurations sont possibles.

[64] C'est le cas de figure que l'on retrouve dans le discours rapporté : [« … »]$_{IP}$ [elle a dit] $_{IP}$.

Gussenhoven, dans la lignée de la tradition britannique, considère que la spécification de l'accent mélodique commence par une syllabe métriquement forte[65].

Le concept des tons associés des accents bitonaux (*leading and trailing tones*) est remplacé par le principe d'association tonale (*tonal linking*). Dans le système ToDI, les deux tons d'un accent bitonal ne doivent pas être réalisés l'un près de l'autre, ils peuvent définir un mouvement mélodique descendant assez étendu[66], et la position du deuxième ton d'un accent est déterminée par la position de l'accent mélodique suivant. Dans le cas des accents prénucléaires, le mécanisme d'association tonale force donc le ton associé à droite (*trailing tone*) à être aligné avec la syllabe contiguë à la syllabe suivante dotée de la proéminence mélodique. Par conséquent, ce ton associé gère la spécification des mouvements mélodiques précédant immédiatement la proéminence mélodique. Cette conception implémente le concept du domaine d'association tonale développé par Gussenhoven dans la perspective consistant à dissocier la structure prosodique et l'organisation tonale et intonative de l'énoncé (Gussenhoven 1990, Gussenhoven et Rietveld 1992). Car l'implémentation du ton associé dépend de la distance par rapport à la proéminence mélodique suivante, on n'utilise plus le signe '+' pour joindre les deux tons de l'accent bitonal ; l'annotation L*H est donc propre au système ToDI.

Simultanément, les accents monotonaux, H* et L* décrivent les cibles mélodiques haute et basse respectivement avec une configuration plateau maintenue jusqu'à l'accent mélodique suivant. Ainsi, par opposition à ToBI, les annotations H* et L* correspondent aux configurations comportant des tons projetés (*spread)*, les configurations montantes et descendantes étant annotées L*H et H*L respectivement.

De façon générale, on reconnaît cinq types d'accent mélodique :

H*	ton haut soutenu (niveau)
L*	ton bas soutenu (niveau)
H*L	mélodie descendante
L*H	mélodie ascendante
H*!H	contour vocatif

Les formes canoniques des accents mélodiques peuvent se voir modifiées : cette modification peut refléter les effets du processus d'abaissement tonal (on obtient donc des variantes !H* et !H*L), ou bien les accents H*L et !H*L peuvent être précédés d'un L*, ce qui résulte en une configuration tritonale L*HL et L*!HL (correspondant aux *delayed falls* et *fall-rises* de Ladd (1983) et Gussenhoven (1983)). Enfin, l'annotation H*+L est réservée pour les cas où la courbe mélodique descend brusquement ; cette configuration s'oppose donc à une descente graduelle annotée H*L.

Le système ToDI est donc un système orthogonal, avec 24 contours de base (figure 2.4). Par rapport au système IPO, développé également pour le hollandais, ToDI contient

[65] Nous y voyons également l'influence du constituant de pied : en hollandais les pieds sont trochaïques.
[66] Surtout pour les accents mélodiques prénucléaires.

plus de contours. Selon Gussenhoven, cela peut être dû au fait que l'annotation IPO a été développée à partir d'un corpus restreint et n'a jamais réclamé d'exhaustivité de l'inventaire ; ensuite, IPO impose le seuil de 6% d'occurrences pour qu'un contour soit inclus dans la grammaire.

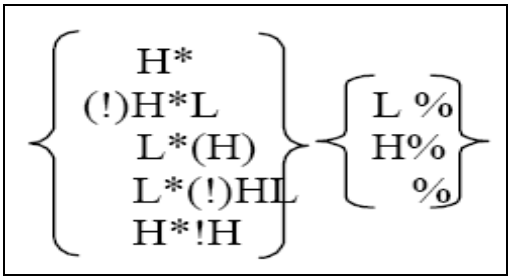


Figure 2.4 : Grammaire intonative du hollandais selon l'annotation ToDI

2.5. IVIE : INTONATIONAL VARIATIONS IN (BRITISH) ENGLISH

Le système d'annotation IViE est le résultat d'une réflexion d'un groupe de chercheurs britanniques (Cambridge) sur l'analyse et la description de la prosodie anglaise dans une perspective typologique et dialectologique. Par son contexte historique et les aspects méthodologiques, cette approche rejoint l'idée de Hirst sur la troisième génération des modèles intonatifs (Hirst 1991).

Le projet IViE porte sur l'anglais (même s'il s'agit de l'anglais britannique). La question se pose donc de savoir pourquoi ne pas utiliser le système ToBI : certes, ToBI étant développé pour l'anglais américain, les modifications éventuelles pour rendre ToBI applicable à l'anglais britannique seraient minimes. Or, les auteurs regrettent que ToBI ne permette pas de mettre en évidence toutes les différences phonétiques et phonologiques observées entre les dialectes (Nolan & Grabe 1997). Dans la même lignée, les auteurs de IViE reprochent à ToBI de mélanger dans son choix d'étiquettes les niveaux de représentation phonétique et phonologique (forme et fonction) :

> *« However, the linguistic status of ToBI has remained somewhat vague. Specifically, it is unclear whether ToBI is intended to provide phonetic or phonological transcriptions of intonation. »* (Nolan & Grabe, 1997:259)

Les auteurs refusent également le parallèle entre ToBI et IPA dans la mesure où pour transcrire l'intonation dans le cadre de ToBI, on doit prendre certaines décisions linguistiques[67]. Par conséquent, le système d'annotation IViE se positionne comme une alternative à ToBI, dans la mesure où il intègre explicitement les niveaux phonétique et phonologique d'annotation.

2.5.1. Procédure analytique

[67] L'annotation ToBI réclame de ne noter que les catégories distinctives ; donc, avant de procéder à l'étude des données, on est amené à établir la liste de ces catégories.

Le système d'annotation IViE a été développé dans le cadre d'un projet de recherche sur la dialectologie de l'anglais : le corpus sur lequel le système a été testé contient des enregistrements qui illustrent 9 dialectes urbains de l'anglais britannique contemporain. Au départ, on constate que les dialectes se différencient sur trois aspects :

a) les différents contours mélodiques sont utilisés pour exprimer la même signification/fonction ;

b) le même contour mélodique peut démontrer les différences en réalisation acoustico-phonétique d'un dialecte à l'autre ;

c) ensuite, les différences rythmiques, pertinentes en phonologie intonative à travers les concepts d'association et d'alignement contribuent à la variation interdialectale.

De ce constat, le système d'annotation prosodique doit posséder la capacité de rendre compte de toutes ces sources de variabilité. Au vu de ces considérations, le système IViE intègre le cadre autosegmental-métrique, propre à ToBI, dans l'analyse de l'intonation ; mais par rapport à ToBI, deux modifications conceptuelles importantes sont à retenir.

Tout d'abord, l'inventaire des étiquettes de IViE a été élargi pour permettre une description unifiée de plus d'une variante de l'anglais. En fait, IViE propose un ensemble ouvert d'étiquettes dans lequel on puise pour la description du corpus analysé. Les encodages pour les étiquettes utilisées s'inspirent des travaux de Gussenhoven (1988) et de Grabe (1998).

En outre, dans le système d'annotation IViE deux niveaux supplémentaires sont rajoutés : niveau rythmique et niveau des mouvements mélodiques. Cette innovation a pour conséquence de rendre plus transparent le travail de transcription.

2.5.2. Palier rythmique

On commence le travail d'annotation par identification des syllabes porteuses d'accents lexicaux. Dans la démarche d'annotation des proéminences métriques, on confronte systématiquement les données linguistiques avec les impressions perceptives. Grabe *et al.* (2001) illustrent la nécessité d'une telle démarche avec l'exemple de l'anglais de Panjabi[68] : cette variante de l'anglais est caractérisée par un renforcement significatif avant les frontières prosodiques. Subséquemment, les syllabes inaccentuées peuvent recevoir une proéminence mélodique (ainsi le mot RAIling[69] est prononcé raiLING) : les deux syllabes seront donc annotées comme proéminentes. Les syllabes 'potentiellement accentuables' sont donc marquées sur le palier rythmique avec une étiquette 'P' (pour proéminent) avec la voyelle de la syllabe accentuée. Deux autres symboles peuvent être utilisés au niveau rythmique : % pour marquer les frontières des unités rythmiques, et # pour signaler les hésitations et les interruptions.

[68] L'anglais de Panjabi (*British Panjabi English*) désigne la variété de l'anglais parlée par les locuteurs natifs de Panjabi (dans le corpus IViE, les locuteurs enregistrés sont bilingues).
[69] Les majuscules sont utilisées pour signaler la syllabe métriquement forte.

2.5.3. Analyse phonétique des proéminences mélodiques

Ensuite, il s'agit de marquer les syllabes accentuées qui ont reçu en production une proéminence mélodique supplémentaire. La particularité de la démarche IViE consiste en une distinction systématique entre la représentation phonétique et la représentation phonologique des accents mélodiques : ce choix méthodologique est pourtant la conséquence d'une observation empirique selon laquelle ce sont les détails d'implémentation des accents mélodiques qui peuvent fréquemment différencier les variantes de l'anglais britannique. On procède donc par examen de la courbe de f0 visualisée, mais également on se fie à la perception. Les mouvements mélodiques associés à la syllabe proéminente se voient annotés sur le palier des mouvements mélodiques (*pitch movement tier*), ce niveau d'annotation ayant une valeur heuristique : on précise les détails de la réalisation acoustico-phonétique d'un accent mélodique. Les étiquettes disponibles pour le niveau des mouvements mélodiques sont h (*high*, haut), l (*low*, bas) et m (*mid*, médiane), qui sont définies de manière relative. Les majuscules sont utilisées pour transcrire le niveau mélodique associé à la syllabe proéminente.

On utilise le concept du domaine d'implémentation d'un mouvement mélodique : dans ce domaine on reconnaît des parties significatives qui correspondent à une syllabe précédant immédiatement la syllabe tonique, à la syllabe tonique et à toutes les syllabes inaccentuées postérieures jusqu'à la prochaine proéminence mélodique. Par conséquent, les deux domaines d'implémentation adjacents se chevauchent d'une syllabe. Le domaine d'implémentation se différencie du domaine d'association de Gussenhoven (1990) et de l'approche de l'école britannique dans la mesure où il commence par une syllabe inaccentuée. Dans la spécification du mouvement mélodique, le niveau atteint à la frontière d'un domaine d'implémentation est précédé d'un tiret ('-'[70]) : ainsi, l'annotation 'lL-h' signifie que la syllabe proéminente et la syllabe qui la précède sont réalisées avec un ton bas, le ton haut étant atteint à la fin du domaine d'implémentation.

Dans la figure 2.5, les différents paliers du système IViE sont représentés. Ainsi cette figure nous permet-elle de voir les détails de la réalisation de l'accent mélodique H*+L dans la position initiale et finale d'un syntagme intonatif. En position finale, cet accent est réalisé avec un ton haut (H) sur la syllabe proéminente, un ton bas (l) sur les syllabes post-accentuelles et avec un ton moyen sur la syllabe inaccentuée précédente. À l'initiale, le sommet mélodique se voit retardé (*delayed*) faute du matériau segmental disponible. Par conséquent, pour le même accent on retrouve la réalisation avec un ton moyen (M) associé à la syllabe proéminente, suivi d'un sommet mélodique haut sur la syllabe immédiatement adjacente et d'un ton bas sur les syllabes inaccentuées après le sommet.

[70] De manière approximative, '-' marque une interpolation tonale ; cf. d'autres systèmes d'annotation qui sont amenés à introduire une classe des mouvements connecteurs pour rendre compte des changements de f0 entre deux accents mélodiques.

Cambridge English

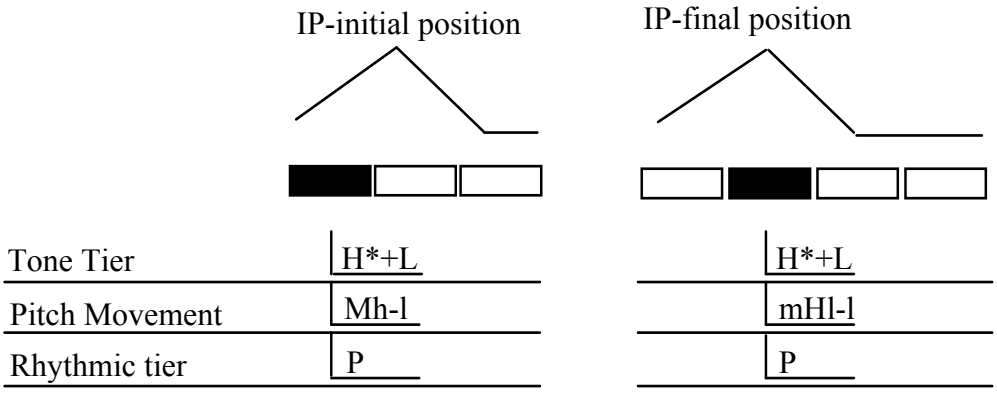


Figure 2.5 : Paliers de l'annotation IViE

L'utilité des conventions définies réside dans le fait que les informations représentées sur les paliers rythmique et mélodique rendent le système d'annotation plus transparent *(« The pitch movement tier adds transparency to the process of phonological classification »*) ce qui facilite l'utilisation des transcriptions faites par d'autres chercheurs : on fait apparaître la réflexion qui a sous-tendu le choix d'une étiquette à partir de l'inventaire phonologique. Dans cette démarche, l'implémentation phonétique des accents mélodiques affiche une variabilité remarquable mais qui est pourtant systématique : le système offre donc la possibilité d'établir une correspondance *many-to-one* entre une étiquette phonologique et l'ensemble des mouvements mélodiques associés. De plus, les étiquettes phonétiques permettent de procéder à une mise en correspondance systématique entre une annotation phonologique d'un accent mélodique et son implémentation phonétique dans les différents dialectes de l'anglais britannique et dans les différents contextes segmentaux (cf. van Santen et Hirschberg, 1994) et prosodiques.

2.5.4. Annotation phonologique de IViE

Au niveau de l'annotation phonologique, on procède à des généralisations à partir de la spécification des réalisations phonétiques sur le palier des mouvements mélodiques. L'inventaire des accents mélodiques est fondé sur les descriptions de l'intonation de l'anglais britannique standard par Gussenhoven (1984) et Grabe (1998) : on y intègre, entre autres, certaines idées de l'école britannique sur l'intonation (par exemple, on considère que tous les accents mélodiques ont la tête à gauche). Le trait caractéristique de cet inventaire réside dans le fait qu'il ne correspond à aucune variante de l'anglais britannique, mais représente plutôt une liste des configurations mélodiques potentielles : pour la description de chaque variante de l'anglais britannique, on peut en soustraire un sous-ensemble d'étiquettes, qui lui est propre.

Dans cet inventaire on retrouve :

a) 6 étiquettes pour les accents mélodiques :

H*+L	L*+H
H*	L*
L*H+L	H*L+H

b) les modificateurs[71] :

^ rehaussement d'un ton
! abaissement d'un ton
} pour indiquer la localisation du ton dans le registre du locuteur

_ déplacement à droite d'un ton associé (*trailing tone*) de l'accent mélodique bitonale (par exemple H*+_L), ce qui signale la modification de la localisation du ton dans le domaine temporel.

c) Les tons de frontières :

Frontière initiale	Frontière finale
%L	%L
%H	%H
%0	%0

L'annotation d'une frontière comme non-spécifiée pour le ton ('%0' ou '0%') signifie que le niveau mélodique de la dernière syllabe d'un syntagme intonatif est le même que le niveau mélodique des syllabes précédentes. Avec l'inclusion dans l'inventaire tonal de cette étiquette, il est possible de rendre compte d'un mouvement ascendant suivi d'un plateau dans l'anglais parlé en Irlande du Nord. Cependant, il découle de ce fait que, dans le cadre d'IViE, la structuration du flux mélodique en termes de domaines se fait à partir des informations autres que tonales : on prend en compte l'achèvement du contour mélodique, la présence d'une continuité rythmique ou d'un allongement final.

Le trait distinctif du système IViE est donc la prise en compte systématique des réalisations contextuelles (une représentation phonétique) pour les accents mélodiques : il a été démontré que certains contrastes intonatifs sont phonologiques dans un dialecte de l'anglais et phonétiques dans un autre (Grabe *et al.* 2000). Simultanément, comme nous l'avons déjà mentionné, les systèmes formels d'annotation prosodique ont pour objectif une description consensuelle des faits attestés : or, même si les transcripteurs n'arrivent pas à un consensus sur la classification phonologique d'un patron accentuel, les différences dans l'annotation phonétique sont infimes. Nous pouvons donc penser que cette unanimité permettra par la suite l'entraînement des algorithmes d'annotation automatique et de détection automatique des événements linguistiquement pertinents sur le plan mélodique.

Les systèmes ToDI et IViE, discutés dans les deux dernières sections, conjuguent les dimensions phonétique et phonologique. Le modèle que nous présentons ci-après se positionne comme le modèle phonétique de l'intonation : son auteur, Paul Taylor, discute en profondeur de ce choix analytique, proposant son interprétation du niveau phonologique dans les études intonatives.

[71] Cf. les traits binaires proposés par Gussenhoven (1984) et Ladd (1983).

2.6. TILT (TAYLOR, 2000)

Tilt est un modèle phonétique de l'intonation qui s'inscrit dans le cadre de recherches sur le traitement automatique des langues. En particulier, Taylor s'interroge sur l'absence d'un module intonatif dans les systèmes de reconnaissance automatique et considère que cela est dû à l'absence d'un modèle robuste de l'intonation[72]. Sur le plan méthodologique, on se situe dans la logique de la détection et non pas de la classification des faits prosodiques pertinents, ce qui distingue le modèle de Taylor des approches présentées auparavant. Simultanément, Taylor insiste sur le caractère phonétique de la représentation obtenue par opposition aux analyses phonologiques qui cherchent à démontrer les structures intonatives sous-jacentes. Or, selon lui, la valeur des modèles phonologiques n'est pas claire :

> *« It is only in a few practical cases that this distinction[73] actually matters much, for example with the treatment of "level accents". These are pitch accents which no observable F0 behaviour and hence should be present in a phonological transcription but not a phonetic one. »*

2.6.1. Unité d'analyse

L'unité de base dans le modèle de Taylor est un événement intonatif discret qui peut surgir au niveau de surface comme un ton de frontière ou un accent mélodique. De manière opérationnelle, on définit l'accent mélodique comme une excursion de f0 associée à une syllabe lui transmettant une proéminence discursive (on adopte l'annotation **a**). Vu cette définition, les accents mélodiques correspondant au ton de niveau ne sont pas modélisés dans le modèle Tilt, car ces accents n'ont pas de manifestation dans la courbe de f0 ; or, dans le système de la langue, ces accents peuvent avoir un statut phonologique. Les tons de frontières (annotés **b**) sont les excursions montantes de f0 localisées aux frontières des syntagmes intonatifs ; sur le plan fonctionnel, les tons de frontières, outre la délimitation du syntagme intonatif en cours, transmettent à l'auditeur l'information sur sa modalité (interrogative ou continuative). Dans le modèle Tilt seules les frontières montantes sont considérées comme des événements mélodiques, les frontières descendantes constituant une forme par défaut[74]. Un événement mélodique combiné (**ab**) a lieu quand l'accent mélodique et le ton de frontière sont rapprochés dans le temps, et par conséquent, un seul mouvement mélodique est observé. Il existe différents types d'accents mélodiques et de tons de frontière, permettant aux locuteurs de produire un ensemble de contours mélodiques à fonctions variées.

Malgré son ancrage dans la substance phonétique, le modèle Tilt s'inscrit dans le cadre autosegmental métrique : la séquence d'événements intonatifs d'un énoncé constitue le continuum intonatif (*intonational stream*) qui, une fois associé avec le continuum

[72] Cf. également les travaux d'Ostendorf (2000) dans lesquels l'auteur insiste sur l'importance de prise en compte des données prosodiques en processus de reconnaissance de parole.
[73] Entre l'approche phonétique *versus* phonologique de l'intonation.
[74] Le ton de frontière bas L% est absent dans le modèle, car l'intonation descendante après l'accent nucléaire peut être modélisée par interpolation.

segmental, résulte en une description intonative complète. Pourtant, les liens d'association dans le modèle sont bi-directionnels avec une contrainte interdisant leur croisement.

2.6.2. Paramètres descriptifs

Par opposition aux modèles phonologiques, Tilt n'a pas besoin de catégorisation précise des événements mélodiques : un ensemble de paramètres continus est utilisé pour la spécification de chaque événement mélodique. Ces paramètres sont déterminés dans un examen minutieux de la courbe de f0 associée.

On utilise trois paramètres pour la spécification de tout mouvement mélodique. Étant le successeur du modèle *Rise-Fall-Connection Model* (RFC, Taylor 1994)[75], Tilt hérite de certains principes descriptifs et analytiques. Tout contour mélodique reçoit une description en termes de mouvements montants, descendants ou de combinaison des deux. Sur le plan acoustique, chaque mouvement primitif est spécifié par sa durée et son amplitude (empan), les amplitudes des montées étant toujours positives par opposition aux amplitudes des descentes, négatives. Ainsi tout mouvement mélodique est décomposé au préalable en une partie montante et une partie descendante, pour lesquelles on récupère quatre mesures, deux pour l'empan tonal ($A_{rise,}$ $A_{fall)}$, et deux pour la durée (D_{rise} et D_{fall}). Si un événement mélodique correspond à une montée seule, la durée et l'amplitude de descente obtiennent la valeur de zéro et vise versa (cf. les formules (1) et (2)). Les fragments du contour entre deux événements mélodiques sont appelés connecteurs, et reçoivent également leur description en termes d'amplitude et de durée.

(1) $$A_{\text{événement mélodique}} = |A_{rise}| + |A_{fall}|$$

(2) $$D_{\text{événement mélodique}} = D_{rise} + D_{fall}$$

Le troisième paramètre utilisé dans le modèle de Taylor est le *tilt* lui-même qui décrit la forme de la courbe indépendamment de sa durée et de son amplitude. La valeur de tilt étant situé dans l'intervalle [-1 ; 1], son calcul comporte quelques étapes intermédiaires. On détermine, tout d'abord, l'amplitude et la durée de tilt (équations (3) et (4)) : en les calculant, on compare les tailles relatives des amplitudes et des durées pour les parties montante et descendante du contour.

(3) $$tilt_{amp} = \frac{|A_{rise}| - |A_{fall}|}{|A_{rise}| + |A_{fall}|}$$

(4) $$tilt_{dur} = \frac{D_{rise} - D_{fall}}{D_{rise} + D_{fall}}$$

[75] Avec les modifications aboutissant à l'architecture propre de Tilt, Taylor vise à proposer une description moins redondante et plus motivée du point de vue linguistique, du contour mélodique.

Les études expérimentales ont démontré que les mesures de tilt pour la durée et de tilt pour l'amplitude sont fortement corrélées, ce qui permet de n'introduire qu'une seule valeur correspondant à la moyenne des deux (équation 5).

$$\text{tilt} = \frac{D_{rise} - D_{fall}}{2\,(D_{rise} + D_{fall})} + \frac{|A_{rise}| - |A_{fall}|}{2\,(|A_{rise}| + |A_{fall}|)} \quad (5)$$

La figure 2.6 représente cinq formes possibles pour un événement mélodique et les valeurs de tilt associées.

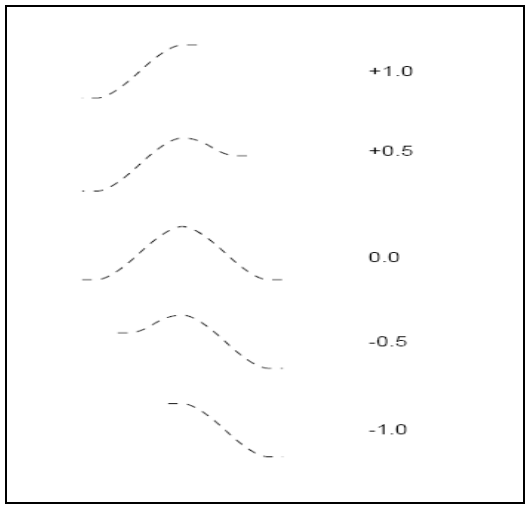


Figure 2.6 : Exemples des courbes mélodiques avec différentes valeurs de tilt (Taylor 2000)

Les trois paramètres inclus dans la spécification des événements mélodiques ont une signification phonétique : par suite l'amplitude peut être considérée comme le corrélat de la proéminence perceptive d'un événement mélodique. L'amplitude est mesurée en Hz, or, certaines corrections sont appliquées pour compenser l'effet de déclinaison[76]. Dans la même lignée, les variations en durées résultent de l'interaction entre l'organisation tonale de l'énoncé et le contenu segmental. Les paramètres de durée et d'amplitude ne peuvent pas être joints en une seule caractéristique (par exemple, *taille de mouvement mélodique*) car il en résulte une perte conséquente de l'information en synthèse. Tilt, en revanche, est la mesure de la forme mélodique de l'événement intonatif qui reflète la taille relative des composants montant et descendant (précisons encore une fois que Tilt est un modèle phonétique de l'intonation) : la valeur de +1.0 correspond à une montée, la valeur de –1.0 à une descente, et la valeur de 0 correspond à une courbe avec les composants montant et descendant de même taille (cf. figure 2.6).

La question de l'alignement entre l'événement tonal et le niveau segmental est abordé sous deux angles : premièrement, le paramètre de durée reflète déjà l'interaction entre le plan mélodique et le plan segmental ; par la suite ; on s'interroge sur le choix du point de repère pour la mesure de durée. Ainsi la durée peut-elle être mesurée de manière

[76] Le même intervalle mélodique serait perçu plus proéminent à la fin du syntagme (Liberman & Pierrehumbert 1984).

absolue, par référence au début de l'énoncé, ou de manière locale, par rapport au noyau vocalique de la syllabe proéminente. C'est la deuxième option qui est retenue et le paramètre de la position syllabique exprime la distance entre le début du noyau vocalique de la syllabe proéminente et le pic mélodique.

2.6.3. A-t-on besoin d'une représentation phonologique ?

Proposant un modèle phonétique de l'intonation, dont les spécifications des paramètres mélodiques sont continues, Taylor (*op.cit.*) se voit obligé d'aborder la problématique de la valeur linguistique du modèle qu'il propose[77]. Selon lui, avec les paramètres continus le modèle résulte en une description abstraite :

> *« we believe the success of this approach is due to the adoption of an abstract level of representation in the broad sense, rather than a necessarily phonological one in the traditional discrete sense. »*

Réfléchissant sur la possibilité de proposer un modèle phonologique (discret) de l'intonation, Taylor arrive à la conclusion qu'une représentation abstraite de l'intonation ne peut être discrète (comme l'est la représentation phonologique au niveau segmental) dans la mesure où le rapport entre la forme et la signification n'est pas le même pour les phénomènes intonatifs. Certes, les différents contours mélodiques évoquent différentes significations, or, le rapport entre la forme et le sens n'est pas tout à fait arbitraire[78] ; autrement dit, il n'existe pas d'arguments qui prouvent l'existence des frontières tranchées entre les contours mélodiques qui correspondent aux changements radicaux de la signification :

> *« Traditionally the argument about categorisation in intonation has revolved about a false dichotomy, namely that the relationship between sound and meaning, can either be as in segmental phonology, where the relationship is completely arbitrary, or as in paralanguage, where the relationship is a simple linear one. In fact this is an inappropriate application of the law of the excluded middle and one does not have to choose either of these positions: a third position is that intonation is continuous with regard to both sound and meaning, but that the relationship between the two is highly complex and non-linear. »*

Pour illustrer ces propos, Taylor procède à une analyse comparative de son modèle avec le système d'annotation ToBI. Nous reprenons la figure 2.7, issue de (Taylor, 2000) qui projette les catégories d'accents mélodiques adoptées dans ToBI dans l'espace acoustique déterminé par les paramètres de position syllabique et de tilt du modèle de Taylor.

[77] On reproche au modèle de n'être qu'un codage de la courbe de f0 par opposition aux descriptions phonologiques dans lesquelles les étiquettes correspondent aux catégories discrètes. Pourtant, comme le remarque Taylor, *« it is very difficult to justify the linguistic relevance of any existing model of intonation »*.
[78] Cf. le code biologique de Gussenhoven (2002), chapitre I.

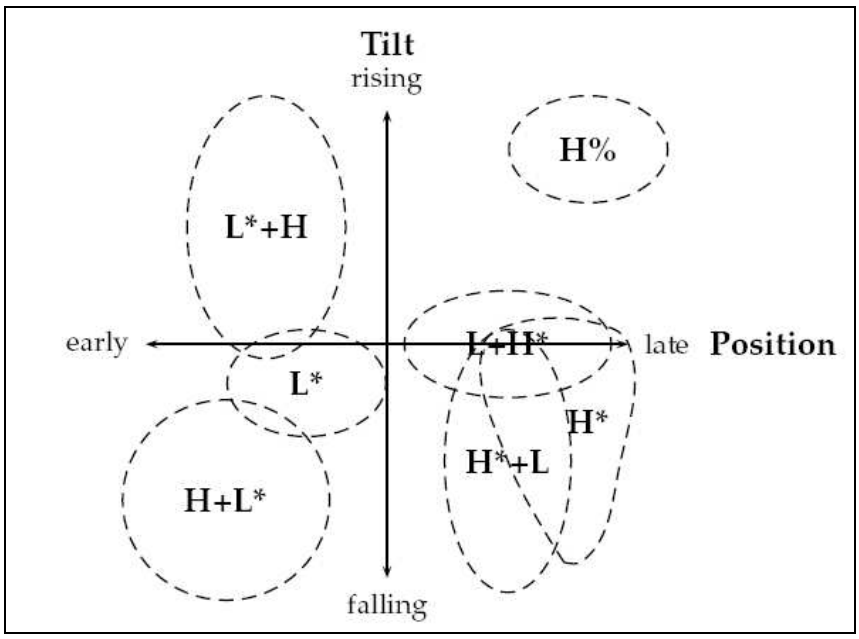


Figure 2.7 : Comparaison entre les systèmes d'annotation prosodique ToBI et Tilt (empruntée à Taylor 2000)

On y observe un chevauchement considérable entre les accents montants L+H*, H* et H*+L, ce qui démontre l'absence d'une correspondance biunivoque entre les catégories phonologiques et ses réalisations phonétiques[79].

Pourtant, Taylor considère que dans la suite de la réflexion, une représentation phonologique peut être dérivée à partir de la description phonétique élaborée : cette représentation phonologique contiendra les mêmes paramètres spécificateurs que la représentation phonétique (à l'exception de la durée), or, avec le passage au niveau plus abstrait, on procédera à une modification de l'échelle. Au niveau phonologique, l'échelle adoptée doit permettre aux événements ayant la même proéminence de recevoir une même annotation : ainsi l'amplitude phonologique reflétera la proéminence perceptive de l'événement mélodique et non pas une magnitude acoustique ; de ce fait, l'amplitude doit être normalisée par rapport au registre mélodique. La durée et le paramètre de la position syllabique reflètent l'interaction entre le niveau tonal et le contenu segmental, ceux-ci relèvent plutôt du domaine de l'alignement, et non pas de l'association. Par contre, le paramètre de tilt est déjà abstrait, sans rapport ni avec la dimension temporelle ni avec la dimension mélodique, et peut donc être adopté tel quel dans la représentation phonologique.

2.7. INTSINT (INTERNATIONAL TRANSCRIPTION SYSTEM FOR INTONATION, HIRST ET DI CRISTO 1998)

Dans la discussion sur la validité de la représentation phonologique pour des faits prosodiques et intonatifs, nous avons évoqué l'architecture proposée par Hirst et Di Cristo (1998), à savoir une distinction systématique entre les niveaux phonétique,

[79] Nous remarquons également que le cas de figure présenté peut expliquer les confusions attestées dans les études de consistance du système de transcription.

phonologique de surface et phonologique profond. Une méthodologie d'analyse et d'annotation de la strate tonale est associée à cette représentation multiniveaux sous forme d'algorithme MOMEL et du système d'annotation INTSINT que nous présentons dans cette section. Le système d'annotation INTSINT aboutit à un encodage des faits tonals au niveau phonologique de surface. De par sa position dans l'architecture générale, ce niveau interagit, d'une part, avec le niveau phonétique et, d'autre part, avec le niveau phonologique profond.

Le système INTSINT est intrinsèquement lié à la représentation proposée pour le niveau phonétique, résultant de l'algorithme MOMEL, que nous présenterons au préalable. Le présupposé premier de ce modèle consiste à considérer la courbe de la fréquence fondamentale telle qu'elle s'affiche dans un éditeur du son comme le résultat de la superposition d'une composante microprosodique et d'une composante macroprosodique (Di Cristo et Hirst 1986)[80]. Dans cette optique, la composante macroprosodique reflète les modulations volontaires imposées au signal par le locuteur dans un objectif communicatif précis[81]. L'idée sous-jacente consiste à concevoir la courbe intonative au niveau phonétique comme une courbe continue et lisse et donc modélisable à l'aide des fonctions splines quadratiques[82]. De ce choix émane la procédure de codage : une courbe peut y être représentée par un ensemble de points-cibles définissant dans l'espace « temps ~ fréquence » les changements significatifs. La mise en œuvre de ce procédé assure à INTSINT une de ses caractéristiques fondamentales, à savoir la réversibilité entre les niveaux phonétique et phonologique dans la représentation prosodique. Cette démarche s'intègre dorénavant dans l'heuristique de l'analyse par synthèse.

L'algorithme MOMEL permet d'obtenir de manière automatique la courbe stylisée modélisant au mieux la courbe de f0 donnée (Hirst et Espesser 1993, cette étude détaillant également les étapes-clefs de la modélisation mathématique sous-jacente).

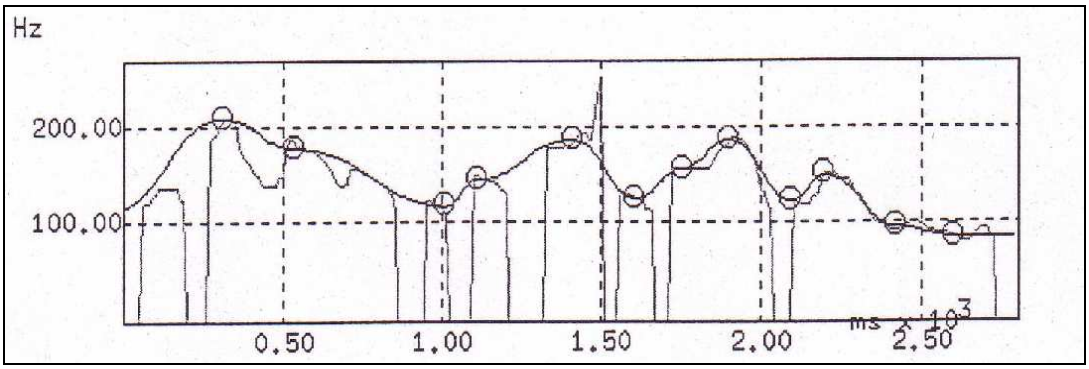


Figure 2.8 : Exemple d'une courbe modélisée par l'application de l'algorithme MOMEL

[80] Di Cristo (1978) propose également d'attribuer à la composante microprosodique une structure complexe, faisant la différence entre les effets intrinsèques (cf. également Lehiste 1970) et les effets co-intrinsèques (issus de la coarticulation, et notamment de l'influence de la consonne sur les paramètres acoustiques des voyelles adjacentes).

[81] Vaissière (2002) qualifie ce système comme étant orienté production : en effet, on considère que la courbe macroprosodique reflète l'activité de source glottique ; dans le même esprit, la raison d'être des points-cibles détectés est l'intention volontaire du locuteur.

[82] Ce choix s'explique par le fait que les fonctions splines quadratiques, c'est-à-dire les arcs de paraboles qui s'enchaînent, sont les fonctions continues les plus simples dont la dérivée première est une continue.

L'illustration présentée (figure 2.8) confirme l'interprétation de la courbe modélisée à l'aide de l'algorithme MOMEL comme étant une interpolation entre les points-cibles définis dans les coordonnées « temps ~ fréquence ». L'interprétation du concept de point-cible comme un point tournant dans la courbe mélodique y est proche de celle de Bruce (1977), et, par conséquent, opposée à celle qui se trouve intégrée dans l'approche ToBI (cf. discussion ci-dessus).

Le passage vers le niveau phonologique de surface est assuré, dans le cadre de ce modèle, à l'aide de l'alphabet INTSINT: il s'agit d'un alphabet fini de symboles tonals discrets. Le système INTSINT comprend un alphabet de huit symboles {T, M, B, H, S, L, U, D}[83,84] qui se répartissent en deux grands groupes : les tons absolus (T, M, B) réservés au codage des valeurs définissant l'étendue tonale du locuteur, au moins dans le cadre de l'unité intonative sous analyse, et les tons relatifs (H, S, L, D, U) définis en référence aux points cibles adjacents. Les symboles relatifs se divisent à leur tour en deux groupes, la distinction étant faite entre les tons non-itératifs (H, S, L) et itératifs (U, D) à la base de la considération que les tons itératifs puissent être suivis par des tons de même identité, configuration impossible et non autorisée pour des tons non-itératifs. Simultanément, les tons itératifs utilisent les intervalles moins importants pour les mouvements de montée et de descente. Les symboles non-itératifs correspondent donc aux pics ou aux vallées de la courbe de f0 stylisée tandis que les tons itératifs marquent des suspensions intermédiaires dans un mouvement global montant ou descendant (cf. la figure 2.9).

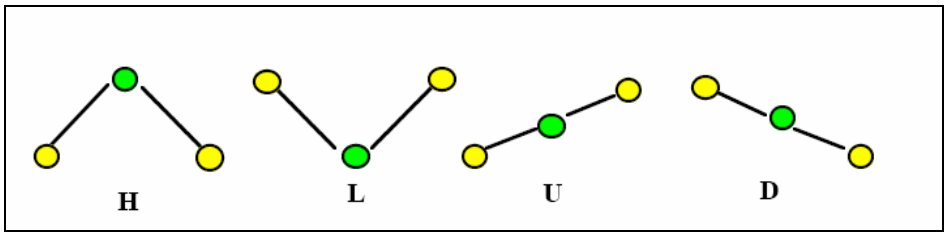


Figure 2.9 : Configurations tonales et leurs annotations INTSINT

La première version de l'algorithme INTSINT a été appliquée pour décrire l'intonation de neuf langues dans l'ouvrage de Hirst et Di Cristo (1998). Depuis, une deuxième version d'algorithme a été proposée (Hirst 2001) : l'encodage des points-cibles s'appuie sur deux paramètres, *clef* et *étendue*, dépendants du locuteur/de l'unité d'analyse. Les tons absolus T et B sont définis aux limites de l'étendue du locuteur, et sont symétriques autour de la valeur centrale (M) correspondant à la valeur de *clef*. Ensuite, les tons absolus sont définis en fonction de l'intervalle entre le point-cible précédent et les deux valeurs asymptotiques T et B selon l'équation (6)

$$(6) \qquad P_i = P_{i-1} + c*(A - P_{i-1}) ,$$

où A (asymptote) correspond à la valeur de T (pour les tons H et U) ou de B (pour les tons D et L) et *c* est une constante, égale à 0.5 pour les tons non-itératifs H et L et à 0.25 pour les tons itératifs D et U. L'application de cette deuxième procédure d'encodage

[83] Les abréviations des étiquettes *Top, Mid, Bottom, Higher, Same, Lower, Upstepped, Downstepped* respectivement.
[84] L'alphabet INTSINT existe également sous forme iconique, laquelle a été utilisée dans Hirst et Di Cristo (1998).

aboutit à une annotation en contradiction avec la version « configurationnelle » de l'algorithme présentée dans la figure 2.9.

Nous insistons tout particulièrement sur le caractère automatique de l'annotation phonologique discrète telle qu'elle résulte de l'application successive des algorithmes MOMEL et INTSINT. Notons pourtant que la représentation phonologique profonde ne peut être obtenue automatiquement et nécessite le développement d'une théorie phonologique sous-jacente spécifique pour la langue : dans cette démarche l'accent est mis sur la jonction entre l'organisation de l'espace tonal et la constituance prosodique. Dans l'application à l'anglais britannique (Hirst 1998), la structure hiérarchique est conçue autour du concept métrique du pied selon Abercrombie (1964) et du concept de l'unité tonale qui lui est parallèle dans le domaine fréquentiel. On définit également l'unité tonale du niveau supérieur qui est une unité intonative. On associe à chacune des deux unités des gabarits tonals spécifiques $_{H}[\ ]_{L}$ pour l'unité tonale et $_{L}[\ ]_{\{H;L\}}$ pour l'unité intonative. Vu la contrainte de linéarité des représentations phonologiques (émanant du caractère linéaire de la parole), nous obtiendrons la représentation présentée dans la figure 2.10, résultant de la projection sur un plan unique des schémas généraux des unités tonales définies. Le deuxième ton de l'unité tonale est dotée de la propriété d'être un segment flottant (p. 76) : ce ton n'est pas réalisé phonétiquement mais a pour fonction d'induire l'abaissement du ton suivant.

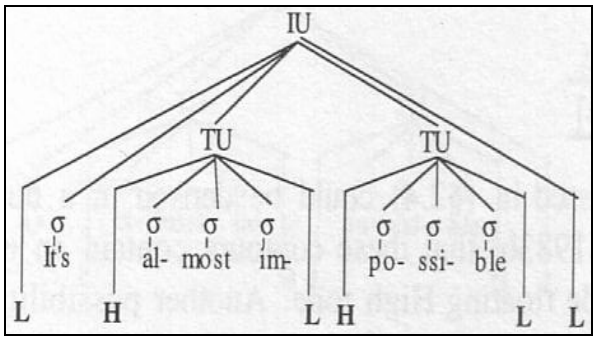


Figure 2.10 : Représentation hiérarchique de l'organisation tonale pour l'énoncé « It's almost impossible *»*

Cette conception de la hiérarchie prosodique permet d'expliquer la représentation phonologique de surface associée avec l'énoncé analysé :

It's ALmost imPOsible				
M	T		D	H

Une analyse similaire a été proposée pour le français. Cependant, nous considérons que des travaux plus approfondis sont nécessaires dans le domaine de l'alignement des points cibles d'INTSINT avec le niveau segmental : actuellement, les seuls travaux effectués dans cette perspective sont Nicolas et Hirst (1996) et Astesano (2001) pour le français.

2.8. RÉSUMÉ ET DISCUSSION

Dans le présent chapitre nous avons dressé un panorama des systèmes d'annotation développés pour représenter les propriétés tonales de la structure prosodique.

Nous avons commencé notre discussion avec la présentation de l'approche IPO, orientée perception. Cette approche est positionnée en dehors du cadre des approches autosegmentales-métriques, quoique Gussenhoven (2002b) la qualifie de phonologique avant la lettre. Cependant, il importe de spécifier que le nombre des mouvements mélodiques distingués dans la perception est très large, et que l'application de cette méthodologie sans une analyse fonctionnelle concomitante pourrait conduire à une précision exagérée de la description. Simultanément, une étude postérieure de Gussenhoven et Rietveld (2000) a mis en évidence le besoin d'élargir l'inventaire des mouvements mélodiques pertinents en hollandais.

Nous avons également discuté du système d'annotation Tilt qui reconnaît deux types d'événements mélodiques, les accents mélodiques et les tons de frontières ascendants, au niveau fonctionnel ; les frontières descendantes sont considérées comme une forme par défaut, résultant d'une interpolation. L'ancrage de ces événements dans la substance acoustique sous forme d'une modulation de f0 est indispensable dans le modèle : c'est la courbe qui reçoit une spécification en termes de paramètres clefs du modèle. Le système Tilt est jugé par son auteur (Taylor 2000) comme abstrait, sans pourtant atteindre le niveau de la description phonologique. Taylor (*op. cit.*) critique une approche assidûment catégorielle de l'école ToBI, notamment, parce que les frontières entre les catégories ne peuvent pas être spécifiées avec précision dans l'espace acoustique (cf. également van Santen et Möbius 2000).

Après avoir discuté de la démarche analytique qui sous-tend l'application de la méthodologie IPO, nous avons présenté le travail de Pierrehumbert (1980) et le système ToBI qui en a découlé. Il faut préciser qu'actuellement ToBI représente un standard de l'annotation prosodique et, initialement développé pour l'anglais américain, ce système a été transféré à de nombreuses langues (Jun 2005). Or, et les auteurs le reconnaissent eux-mêmes (Beckman *et al.* 2005), le développement d'une version d'annotation ToBI pour une nouvelle langue présume un processus de réflexion approfondie sur les contrastes phonologiques et sur les structures prosodiques propres à cette langue :

> *« A ToBI framework system devised for one language variety cannot be assumed to be applicable even to other varieties of the same language, without first establishing appropriate intonational and prosodic analyses for each variety. »*

La raison d'être de cette précaution réside dans le fait que les étiquettes ToBI n'encodent pas les mouvements mélodiques perceptibles ou des points tournants dans la courbe de f0, mais sont fondées sur de nombreux facteurs perceptifs et acoustiques et sur des considérations phonologiques. Indiquons également une dépendance qui se crée entre le niveau d'annotation tonale et celui du phrasé prosodique : le marquage d'une frontière présuppose l'annotation d'un événement tonal dans la même localisation. Les systèmes ToDI et IViE essayent de remédier aux points faibles de ToBI, en adoptant le concept du domaine d'association ou en intégrant dans l'annotation un palier rythmique ; ceci fait tout intérêt de ces deux procédés analytiques. Cependant, nous

retiendrons pour l'analyse subséquente de notre corpus le modèle INTSINT, et ceci pour les raisons suivantes :

- le système d'annotation INTSINT permet de procéder à une dissociation systématique de la forme et de la fonction prosodique (cf. Hirst 2005), les deux étant confondues dans le système ToBI ;

- la procédure d'annotation INTSINT est plus explicite ; simultanément, l'encodage est automatisé ;

- le concept-clé du système ToBI est l'accent mélodique, dont l'usage est nouveau dans les analyses de l'intonation du russe (nous développerons davantage ce point dans le chapitre suivant) ;

- nous avons mentionné une interdépendance des annotations du phrasé prosodique et de l'organisation tonale dans le cadre de ToBI. Nous défendrons tout au long de ce travail la nécessité de rendre compte du phrasé prosodique dans le cadre du projet de l'analyse formelle de l'intonation du russe. Cependant, le russe n'a pas bénéficié d'une réflexion approfondie sur la constituance prosodique telle qu'elle est proposée par la phonologie des domaines (cf. chapitre IV) ;

- nous nous attachons particulièrement à la réversibilité de l'annotation INTSINT qui sous-entend et permet, d'une part, la génération d'une représentation abstraite à partir de données physiques et acoustiques sur la f0 et, d'autre part, la démarche inverse, à savoir la génération de la courbe mélodique à partir d'une annotation formelle des faits prosodiques.

Notre premier choix méthodologique effectué, nous nous devons maintenant de proposer un résumé des études antérieures sur l'intonation du russe afin d'argumenter en faveur de notre entreprise d'analyse formelle de l'intonation du russe et en faveur de la mise en application dans cette analyse du cadre de la phonologie intonative.

CHAPITRE III

ANALYSES DE L'ORGANISATION PROSODIQUE DU RUSSE : ÉTAT DE LA QUESTION

Dans un premier chapitre, nous avons présenté les principes de la phonologie intonative qui constitue de nos jours une heuristique dominante dans la recherche sur l'intonation. Il s'agit d'une analyse formelle de l'intonation qui se réfère à des axiomes fondateurs qui sous-tendent les différents systèmes d'annotation de l'intonation que nous avons exposés dans le chapitre précédent. Dans notre recherche, nous cherchons à concevoir un schéma d'annotation de l'organisation prosodique/intonative du russe : nous nous interrogeons également sur la question de savoir comment une théorie linguistique peut être exploitée convenablement afin de contribuer à l'amélioration de la qualité de la prosodie en synthèse.

Nous avons argumenté en faveur de l'application du modèle INTSINT à nos données : ce choix est d'abord motivé par l'analyse d'autres schémas proposés et des présupposés théoriques qui leur sont attachés. Cependant, avant d'aborder la tâche de la description de nos données, nous nous devons de proposer un panorama de méthodes d'analyse et de description de l'intonation du russe qui ont été proposées auparavant : c'est ce que nous faisons dans le présent chapitre.

3.1. INTRODUCTION

Les études de l'intonation du russe ont une longue tradition qui détermine dans une grande partie les principes fondateurs des descriptions proposées pour la langue russe : historiquement, on trouve les premières observations sur les aspects syntagmatiques de l'intonation dans les travaux des syntacticiens (A.M. Peshkovskij 1928, S.O. Karcevskij 1931) ou chez les enseignants d'art déclamatoire (V.N. Vsevolodskij-Gerngoss 1922) (Svetozarova 1998 : 246). Remarquons, pourtant, que la question du système intonatif n'est abordée dans aucune de ces études.

En règle générale, on peut diviser l'ensemble des travaux descriptifs de l'intonation du russe en deux catégories majeures (cf. Svetozarova, 1982, 1998) selon que l'aspect fonctionnel ou formel est au centre des typologies proposées. Dans les approches orientées forme, un nombre limité des constructions mélodiques est proposé, construction qui ont donc le statut de phonèmes intonatifs. *A contrario*, l'objectif des

études à orientation fonctionnelle n'est pas de proposer un ensemble fini de contours mélodiques, mais de décrire les particularités de la forme mélodique associée avec différents types d'énoncés ou de constructions syntaxiques (on parle alors de l'intonation d'assertion par opposition à l'intonation d'interrogation, de l'intonation typique d'énumération ou d'appellation). Il existe également des descriptions objectives détaillées (issues des études expérimentales) de l'intonation des types communicatifs majeurs, de l'intonation des énoncés complexes (Nikolaeva, 1969), etc. Or, ces descriptions fonctionnelles ne s'accompagnent pas d'une analyse formelle des phénomènes tonals.

D'une façon générale, dans la tradition des études sur l'intonation du russe on dispose d'un ensemble de descriptions des moyens intonatifs qui signalent dans le discours les différentes catégories syntaxiques ou communicatives ; les descriptions résultantes n'ont donc pas le statut d'une classification des unités intonatives. L'adoption d'une telle perspective implique l'absence d'une liste finie des formes mélodiques pertinentes. Les intonations décrites reflètent, par ailleurs, l'hétérogénéité du domaine de l'intonation : on y retrouve les détails sur les contours mélodiques associés à un type communicatif précis, les contours mélodiques des syntagmes terminaux versus non-terminaux, les contours mélodiques accompagnant des constructions syntaxiques précises (appellation, énumération), les contours associés à l'expression des émotions (tendresse, colère). L'hétérogénéité est cependant moindre dans les descriptions formelles.

Un autre aspect, d'ordre terminologique, doit être pris en compte : quand on parle d'une description intonative, le terme « intonatif » est fréquemment synonyme de « prosodique », c'est-à-dire, qu'il s'agit de prendre en considération tout un ensemble d'indices acoustiques qui participent à l'actualisation de l'organisation prosodique du message, à savoir le plan tonal, mélodique, ainsi que l'organisation dynamique, temporelle (tempo + pauses) du message et les particularités de phonation. À titre d'exemple, dans l'évolution du modèle formel de Bryzgounova, initialement, les constructions intonatives avaient été déterminées par rapport à leur organisation mélodique, dans la zone du nucléus ; or, dans le travail de 1977, la construction intonative est définie comme

> « *тип соотношения основного тона, тембра, интенсивности, длительности, способный выразить различия по цели высказывания в предложениях с одинаковым синтаксическим строем и лексическим составом* » (Bryzgounova 1977 : 232)
>
> *le type de mise en relation de la fréquence fondamentale, du timbre, de l'intensité et de durée, permettant d'exprimer les différences associées au type communicatif dans les énoncés ayant une même structure syntaxique et le même contenu lexical.*

Cette conception complexe de l'unité tonale (le terme étant « construction intonative ») détermine l'inventaire établi d'unités dans le système de Bryzgounova (1977), appartenant au groupe des approches formelles. Pour d'autres chercheurs, même si le même contour descendant peut être utilisé à la fin des énoncés assertifs pour signaler le caractère conclusif du syntagme, ainsi que dans les syntagmes non-terminaux pour marquer un lien sémantico-discursif particulier entre les deux syntagmes adjacents,

l'équivalence entre les deux n'est pas complète, car la distinction est faite selon les détails de l'organisation temporelle des deux parties de l'énoncé, leurs profils dynamiques et le type de pause entre les syntagmes (Nikolaeva 1969, Svetozarova 1982).

Les deux approches, formelle et fonctionnelle, procèdent de la description des formes intonatives par la spécification des tons nucléaires : il est reconnu que le détail de l'organisation mélodique le plus pertinent pour l'identification du contour mélodique est associé à la syllabe porteuse de l'accent principal dans l'unité intonative. Dans les descriptions entreprises, on n'utilise pas le terme d'accent mélodique nucléaire. Également, la spécification tonale de la partie pré-tonique se fait en termes de niveaux, et non pas de proéminences pré-nucléaires : on considère que dans la perspective formelle, tout syntagme est défini par la présence d'un seul centre mélodique[85]. D'autres chercheurs parlent de proéminences pré-nucléaires associées avec tout mot lexical dans la partie précédant le centre mélodique. L'étude expérimentale de Svetozarova (1973) démontre que les syllabes porteuses de l'accent lexical dans la partie pré-nucléaire ont systématiquement les valeurs de f0 les plus élevées ; par contre ces petites perturbations mélodiques sont nivelées dans les énoncés interrogatifs, ce qui témoigne d'une plus grande fusion des mots dans l'unité du niveau supérieur. Nous voyons donc que le concept de la proéminence mélodique n'a pas été adopté pour la description de l'intonation du russe.

3.2. LES APPROCHES FONCTIONNELLES DE L'INTONATION DU RUSSE

Nous résumons les différentes classifications fonctionnelles de l'intonation du russe d'après Svetozarova (1982 : 85) et les présentons dans le tableau III.1, avec quelques commentaires. Dans ce tableau, sur la même ligne, nous retrouvons les intonations analogues selon leurs propriétés formelles et fonctionnelles ; dans la partie supérieure du tableau, sont regroupées les intonations exprimant les différentes catégories communicatives et émotionnelles, tandis que dans la partie basse sont présentées les intonations participant à l'expression de différents types de liens entre les syntagmes adjacents.

Les intonations de Vsevolodsky-Gerngross (1922) ont été définies dans une perspective d'apprentissage des techniques déclamatoires et se fondent sur des catégories logico-psychologiques (l'auteur utilise des intervalles musicaux pour décrire les aspects formels de chaque catégorie). Les seize intonations se résument en sept types qui, selon Svetozarova (1982 : 86) peuvent être associés à des catégories communicatives ou syntaxiques. On retrouve la même sub-catégorisation dans le système proposé par A.N. Gvozdev (1949)[86].

La description proposée par N.V. Tcheremissina (1969) s'inscrit dans le contexte général de ses travaux sur l'intonation et la syntaxe, réalisés à partir des observations sur les œuvres littéraires prosaïques, poétiques et dramatiques : il n'est donc pas étonnant que la plupart des intonations dans sa nomenclature soient associées avec des syntagmes/des clauses d'un

[85] Volskaya (2001a) parle de syntagmes bimodaux, comportant deux centres mélodiques nucléaires.
[86] Remarquons que A.N. Gvozdev a été le premier à illustrer les différences entre les contours intonatifs du russe avec les énoncés de même contenu segmental.

énoncé complexe. Dans la classification de I.G. Torsoujeva (1974), c'est l'aspect communicatif qui est au centre de la classification élaborée, bien que les descriptions que l'on trouve dans ce travail soient plus rigoureuses pour le plan formel. Un autre détail important de l'approche réside dans la caractérisation globale des modèles intonatifs, sans association des détails du contour mélodique avec des syllabes accentuées ou inaccentuées. L'auteur cherche également une correspondance biunivoque entre les formes mélodiques distinctes et les catégories communicatives des énoncés analysés.

Nous nous arrêtons plus en détail sur la description proposée par T.M. Nikolaeva (Nikolaeva, 1977). Dans le cadre d'une étude comparative et typologique de l'intonation en langues slaves, T.M. Nikolaeva est amenée à proposer une typologie des contours intonatifs de la langue russe. La distinction entre les différentes 'intonations' se fait sur la base de critères fonctionnels, ensuite, l'analyse acoustique et l'examen minutieux des courbes de f0 permettent de mettre en évidence les différences formelles, ancrées dans la substance physique.

Nikolaeva propose une description des différentes réalisations (les variantes conditionnées par le matériau linguistique disponible ainsi que les variantes libres) pour les types d'intonation suivants :

- intonation conclusive (cadence dans sa terminologie ; l'intonation du syntagme final d'un énoncé assertif) ;
- intonation de continuation (semi-cadence) ;
- intonation d'interrogation, avec deux sub-catégories :
  - demande de confirmation
  - et question alternative
- intonation d'interrogation des questions spécialisées[87], avec une sub-catégorie
  - question-écho
- intonation interrogative avec la conjonction *'A'* (« A он? », « Et lui ? » ; cette intonation a été décrite très minutieusement par Bryzgounova (1969) qui démontre son autonomie dans l'inventaire des contours et sa spécificité) ;
- intonation d'exclamation ;
- intonation du contraste, de mise en relief (emphase) ;
- intonation de parenthèse

[87] Avec un mot interrogatif.

V.N. Vsevolodsky-Grengross (1922)	A.N. Grozdev (1949)	N.V. Cheremissina (1969)	T.M. Nikolaeva (1977)	I.G. Torsueva (1974)
1) Interrogation ;	2) Interrogation ;		3) Interrogation 4) Intonation de question spéciale 5) Intonation de question avec la particule introductive 'A' ;	2a) Intonation d'une question générale 2b) Intonation d'une question spéciale
2) Exclamation, intonation d'appel, intonation d'étonnement ;	4) Exclamation ;		6) Exclamation ;	6) Intonation d'exclamation
3) Intonation assertive intonation de persuasion, de demande, d'ordre, d'invitation, d'exhortation ;	1) Intonation assertive 3) Impérative	1) Intonation de constatation ; 8) Intonation d'impératif ;	1) Intonation conclusive ;	1a) Intonation de narration (conclusive) 4) Intonation d'incitation
4) Intonation d'explication ; 5) Intonation de confrontation, de comparaison ; 6) Intonation d'interruption du lien 7) Intonation d'énumération	3') Intonation d'explication, d'avertissement ; 4') Intonation d'opposition, du contraste ; 2') Intonation de subordination ; 1') Intonation d'énumération ; 5') Intonation d'introduction ;	2) Intonation d'explication ; 6) Intonation de précision ; 4) Intonation d'opposition ; 3) Intonation du début ; 5) Intonation d'énumération ; 7) Intonation d'introduction ;	7) Intonation de mise en relief ; 2) Intonation de continuation ; [7) Intonation de mise en relief]	1b) Intonation de narration (non-terminale, de continuation) 6) Intonation de parenthèse 5) Intonation d'implication

Tableau III.1 : Recensement de quelques analyses fonctionnelles de l'intonation du russe

On considère que les contrastes intonatifs s'opèrent dans le nucleus. Étant donné le caractère mobile de l'accentuation en russe[88], l'auteur décrit minutieusement les différences des réalisations (variations) pour les mouvements mélodiques dans la partie pré-tonique, dans le nucleus (syllabe accentuée) et dans la partie post-tonique du mot porteur de l'accent nucléaire syntagmatique.

Ainsi, à titre d'exemple, pour l'intonation conclusive, on sépare en premier lieu les cas avec et sans syllabes post-toniques. La syllabe nucléaire peut également avoir deux réalisations, avec un ton de niveau ou bien un ton descendant (l'intervalle typique équivalent à 40 Hz). Les syllabes pré-toniques peuvent être réalisées à des niveaux différents : la première syllabe pré-tonique est plus haute que la syllabe nucléaire, la deuxième syllabe prétonique peut être soit haute (plus haute que la première, cas de figure possible mais peu fréquent en russe, par rapport au polonais) soit plus basse que la première syllabe pré-tonique (on forme ainsi une figure de « zanos »). Les trois configurations (montante, descendante ou niveau) sont possibles pour les syllabes post-toniques, le choix d'un contour particulier étant facultatif. Les descriptions similaires sont élaborées pour les classes d'unités intonatives que nous avons énumérées plus haut. Dans la suite de sa réflexion, l'auteur s'interroge sur la variabilité inter-locuteurs observée dans son analyse de même qu'elle présente ses considérations sur certaines problématiques d'ordre théorique, par exemple, savoir si la distinction entre l'intonation d'interrogation et de continuité est catégorielle ou non[89]. Les intonations qui caractérisent l'enchaînement des unités sur l'axe syntagmatique sont décrites plus en détails dans l'ouvrage précédent de l'auteur (Nikolaeva, 1969) ; dans le travail analysé ne sont spécifiés que les détails de leurs réalisations acoustiques.

Le travail de Nikolaeva (1969) exploite le concept d'intonème. Traditionnellement, dans les études intonatives (appartenant au groupe des approches holistiques, cf. Rossi 1999, entre autres), ce terme désigne le contour mélodique associé avec un syntagme intonatif, conçu comme morphème, c'est-à-dire l'unité dotée d'un facette signifiante correspondant au contour mélodique, et d'une facette signifié faisant référence aux contrastes de modalité ou d'attitude. Dans la conception de Nikolaeva, les intonèmes sont des unités relatives, qui s'actualisent sur le point de jonction entre deux syntagmes

[88] Comme le signale Svetozarova (1998) cette liberté est double : d'abord par rapport à la position de la syllabe accentuée dans un mot (l'accent peut tomber sur n'importe quelle syllabe du mot) ; ensuite, dans la relation entre l'accentuation et la structure morphologique des mots (l'accent peut tomber sur n'importe quel morphème du mot – la racine, le préfixe, le suffixe ou la flexion ; on distingue pourtant le sous-ensemble des morphèmes accentogènes (Garde 1968). Une des propriétés importantes de l'accent en russe est sa mobilité, c'est-à-dire la tendance à changer de place dans le processus de dérivation et de déclinaison. Pourtant, en dépit d'une telle liberté, les chercheurs ont trouvé des régularités paradigmatiques dans le placement d'accent en russe, définies par la classe à laquelle un mot donné appartient. Svetozarova (1998) estime que la description accentuelle complète d'un mot en russe comporte des indications de son type accentuel (l'accent mobile ou bien fixe sur le thème ou la désinence), de sa structure morphologique par rapport à l'accent et de sa structure rythmique (262).

[89] Traditionnellement, on parle de deux types de contours mélodiques en russe, contours montants s'opposant aux contours descendants. Dans cette optique, on différencie l'interrogation de la continuité par la différence de l'intervalle du mouvement montant ; dans la typologie de Bryzgounova, les mêmes constructions sont associées à deux fonctions, dans la description de Svetozarova, l'intonation de continuation est caractérisée par un début plus bas, un maximum de f0 moins élevé et le niveau plus haut des syllabes post-toniques.

adjacents ; les intonèmes possèdent tout de même leur signifié propre. Dans l'étude de 1969 quatre intonèmes sont définis :

- intonème neutre : les deux syntagmes maintiennent des rapports d'égalité ; le premier présente alors un contour montant, le tempo est stable dans les deux parties, la pause est relativement brève ou complètement absente ;

- intonème avec une signification supplémentaire de comparaison ou mise en opposition : c'est une pause dont la durée est augmentée qui différencie cette configuration de la précédente ; le premier syntagme maintient une mélodie montante, ce qui assure le lien entre les deux parties ; sinon, les deux parties ont la même importance ;

- intonème avec la signification d'une précision importante : la durée d'une pause est encore augmentée, le niveau mélodique du deuxième syntagme est rabaissé, pourtant les deux parties ne se distinguent pas en tempo ;

- intonème de parenthèse (un des syntagmes a la signification d'une précision sans importance) : la pause est minimale, le syntagme introductif est prononcé avec une accélération du tempo et présente une mélodie descendante, tandis que le niveau mélodique de la parenthèse est bas.

Remarquons pourtant que c'est un lien logique et non pas un lien formel syntaxique qui constitue le point fondamental de cette analyse.

Le travail de Fougeron (1989) se positionne à part dans le cadre des approches fonctionnelles. L'attention de l'auteur porte sur la prosodie et la structure énonciative de la phrase telle qu'elle est conçue par l'école pragoise. Dans les conceptions modernes, la structure informationnelle possède au moins deux versants : le premier correspond à l'articulation « focus ~ fond », qui assure la différenciation des informations connues ou déductibles du contexte et les informations nouvelles, rapportées ; la deuxième articulation porte, quant à elle, sur l'organisation interne de l'information, à savoir la différenciation entre « le sujet logique » et « le prédicat logique » (termes de Potebnia, correspondant à la distinction entre de quoi on parle et ce qu'on dit à son propos ; d'autre termes utilisés sont thème ~ rhème). C'est ce deuxième versant qui est au centre de l'étude discutée. Procédant à l'analyse d'un corpus des phrases assertives avec le nombre et l'ordre de composants syntaxiques variables et moyennant une prise en compte systématique de la structuration de l'énoncé en thème et rhème, l'auteur aboutit à la désignation de quatre types de structures correspondant chacune à une organisation du message (p. 446, nous représentons le système résultant dans la figure 3.1). Le schéma de la figure 3.1 est considéré comme un cadre structurel gestionnaire de l'implémentation des paramètres acoustiques, et notamment de la f0. La distribution des accents nucléaires est déterminée par le lien que l'unité actuelle maintient avec le contexte général du discours.

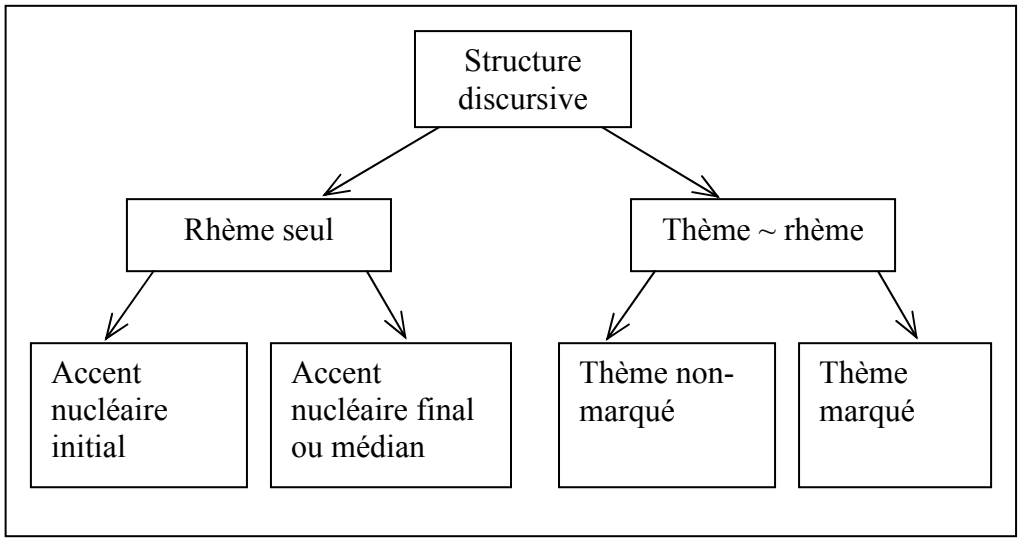


Figure 3.1 : Cadre discursif général régissant l'implémentation des indices acoustiques selon Fougeron (1989)

Spécifions que dans le travail discuté seules les phrases assertives à trois constituants maximum (sujet – verbe – objet) ont été analysées.

La nomenclature des contours mélodiques que nous venons de présenter peut être utilisée dans la perspective de l'annotation des corpus oraux. Quoique nous la considérions comme plus appropriée dans une perspective d'annotation auditive. Nous lui reprochons tout de même d'être par trop généralisatrice et d'ignorer la représentation phonologique intermédiaire (nous avons auparavant fourni des arguments en faveur de ce choix méthodologique).

Le tableau comparatif (tableau III.1) permet d'appréhender un autre défaut des approches fonctionnelles décrites, à savoir le fait de proposer des nomenclatures non-exhaustives : une classification formelle, telle que nous la concevons, doit être par essence exhaustive, prévoyant une possibilité d'ajouter de nouveaux éléments dans les cases prédéfinies (cf. le tableau d'éléments chimiques). Pourtant, dans la perspective fonctionnelle adoptée, on ne peut jamais être sûr que toutes les catégories communicatives et syntaxiques aient été prises en compte, et que toutes les variantes formelles aient été analysées et regroupées en classes globales. Les catégories utilisées dans les descriptions présentées proviennent du domaine de la syntaxe communicative et constructive, ainsi que de la psychologie : pour notre part, nous cherchons une représentation formelle phonologique de l'intonation.

3.3. LES DESCRIPTIONS FORMELLES DE L'INTONATION

Comme l'affirme Svetozarova (1982 : 89), les descriptions formelles de l'intonation, quoique assez typiques pour d'autres langues (comme l'anglais britannique, par exemple), ne sont présentes que dans deux travaux consacrés à l'intonation du russe. L'un d'entre eux est une analyse peu connue de C. Boyanus (1936, 1955), l'autre – un système largement adopté dans la description de l'intonation du russe qui a été proposé

dans les années soixante-dix par E.A. Bryzgounova. Nous entamons donc maintenant la présentation de ces deux approches, et nous conclurons la présentation des approches formelles par la présentation du travail de Odé (1989) et d'un modèle développé récemment par S.V. Kodzasov (1996, 1999).

3.3.1. Travaux de C. Boyanus[90]

Dans l'approche de Boyanus on ressent une forte influence de l'école britannique, or, les particularités de la prosodie russe y sont prises en compte : ainsi, dans sa description des principaux types intonatifs du russe, Boyanus prend en compte l'alignement du contour mélodique avec la syllabe nucléaire et les syllabes post-toniques qui la suivent. Boyanus propose un système de quatre types de contours mélodiques distinctifs pour le russe (figure 3.2) : ils sont comparables avec le ton descendant vers le registre bas (*low fall*), ton montant vers le registre haut (*high rise*), ton montant dans le registre réduit (*narrow rise*) et ton montant-descendant (*rise-fall*) de la tradition britannique.

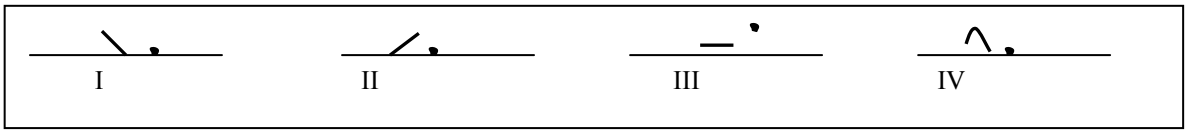


Figure 3.2 : Quatre tons distinctifs dans l'approche formelle de Boyanus

Pour Boyanus toute forme mélodique de base peut être utilisée dans des énoncés appartenant aux différentes catégories communicatives : ainsi, on retrouve le premier accent nucléaire (si l'on adopte la terminologie de l'école britannique) dans les assertions, les ordres, les questions à morphème interrogatif (*специальный вопрос*) et dans les exclamations ; le deuxième, dans les questions générales et certains groupes rythmiques ; le troisième dans les questions, les ordres, les assertions avec une nuance d'implication, d'étonnement ou de doute, et enfin, le quatrième type sert à l'expression d'une demande. En sus, les trois premiers accents nucléaires s'associent avec des syntagmes non-terminaux[91] : ainsi Boyanus efface une distinction maintenue dans d'autres classifications entre les intonations des énoncés entiers et les parties des énoncés. L'auteur analyse également les variantes de ces quatre types dans le contexte d'une emphase : on y observe alors l'augmentation de l'empan des mouvements mélodiques, accompagnée par une augmentation de l'intensité et par des variations du timbre.

3.3.2. Modèle de Bryzgounova (1977, et postérieurs)

La description proposée par E.A. Brygounova a été conçue dans des objectifs didactiques d'enseignement du russe langue étrangère et représente l'unique

[90] Nous présentons ce modèle en nous fondant largement sur la présentation de Svetozarova (1982 : 89-90).
[91] Boyanus considère que la mélodie interrogative (II) est plus propre au discours spontané (dans le cas où le locuteur argumente et présente les preuves de son point de vue, il utilise la mélodie montante (III) commençant avec un niveau relativement bas) ; les mélodies I et III sont plus spécifiques à la lecture.

classification des patrons mélodiques du russe à vocation exhaustive. Comme le remarque N.D. Svetozarova (1982 : 91),

> *« Это, по существу, единственная полная и законченная классификация интонационных типов русского языка, основанная на точно сформулированных критериях выявления интонационных единиц. »*
>
> *C'est, au fond, la seule classification complète et achevée des types intonatifs de la langue russe, fondée sur des critères précis de définition des unités intonatives.*

Pour Bryzgounova, dans toute spécification de l'intonation, les aspects suivants doivent être pris en compte :

a) la segmentation de l'énoncé en syntagmes (cette question n'est pas débattue lors de la présentation du modèle) ;

b) le type de la construction intonative ;

c) le lieu du centre mélodique.

Ce modèle adopte un parcours qui commence par spécifier le plan formel des unités intonatives et qui précise ensuite la signification linguistique associée avec chaque contour mélodique. Les unités de base dans le modèle sont des constructions intonatives (CI) ; les types de constructions intonatives sont déterminées dans le paradigme des paires minimales[92]. Dans la version initiale de la classification (Bryzgounva 1963), il n'y avait que cinq unités de base. Le nombre global a ensuite été augmenté par l'ajout des intonations émotionnelles et modales.

On adopte une structure tripartite du contour mélodique associé à un syntagme : on y distingue le centre (nucleus) et les parties pré- et post-toniques[93]. Pour caractériser toute construction intonative, il est nécessaire de préciser les points suivants :

- la direction du mouvement dans le nucleus (on distingue les mouvements descendants et ascendants, et on prend en considération la forme du mouvement, convexe, concave ou droit) ;

- la quantité et la(les) position(s) des nuclei (une construction intonative peut avoir deux centres mélodiques) ;

- les détails de l'organisation mélodique (très fréquemment, le positionnement dans le registre mélodique du locuteur) pour les parties pré- et post-toniques ;

- le rôle de l'intensité et d'autres moyens phonétiques (phonation).

[92] L'intonation véhicule les nuances du sens incompatibles dans le même contexte, les différentes significations ; cette signification se voit inhibée lors de la substitution d'une construction intonative par une autre.

[93] Du point de vue structurel, les parties pré- et post-toniques peuvent être comparées à ce que la tradition britannique nomme « queue » (*tail*) et « tête » (*head*).

Bryzgounova propose sept constructions intonatives pour la description de l'intonation de la langue russe que nous résumons dans le tableau ci-dessous.

Construction Intonative	Signification véhiculée
ИК-1 Такие у них обычаи.	Finalité
ИК-2 Какие у них обычаи?	Question à morphème interrogatif, impulsion à l'action, comparaison
ИК-3 Какие у них обычаи?	Interrogation générale
ИК-4 А у них? Какие у них обычаи?	Demande de confirmation, continuation, commandement
ИК-5 Какие у них обычаи!	Exclamation, ravissement
ИК-6 Какие у них обычаи!	Continuation, exclamation, valeur émotionnelle
ИК-7 Какие у них обычаи!	Émotions négatives

Figure 3.3 : Sept constructions intonatives de Bryzgounova (1977)

Sur la figure 3.3, la première colonne comporte le numéro de la construction qui peut servir d'identificateur ; ensuite, nous donnons l'exemple d'un énoncé en russe et la représentation de sa courbe intonative associée, puis, dans la troisième colonne, nous résumons les significations linguistiques véhiculées par chacun des contours mélodiques.

Ce modèle de l'intonation peut être utilisé lors de la transcription du corpus, par exemple, l'annotation suivante est possible :

(1) Они уезжают <u>завтра</u> <1>.

Dans l'exemple (1) on souligne le mot porteur du mouvement mélodique nucléaire (le centre mélodique du syntagme) et on indique entre crochets le numéro de la construction intonative dans la classification.

Sur le plan de la signification, les constructions de Bryzgounova sont surtout associées avec les catégories communicatives et émotionnelles. Les aspects syntaxiques y sont donc moins développés. On indique ainsi que différentes constructions intonatives (CI-I, CI-III, CI-IV) sont possibles dans les syntagmes non-terminaux : le choix y est surtout stylistique[94], et non pas conditionné par la nature du lien syntaxique entre les syntagmes, comme dans les approches fonctionnelles que nous avons présentées précédemment. Nous remarquons également la synonymie des constructions CI-3, CI-4

[94] Ainsi CI-III est plus fréquente dans le discours spontané, CI-IV en lecture ou en parole officielle.

et CI-6 toutes les trois ayant pour la signification première « intonation de continuation ».

Néanmoins, comme nous l'avons déjà souligné, le système de Bryzgounova ne peut pas être considéré comme un système d'annotation formelle de l'ordre intonatif dans le système prosodique dans la mesure où, premièrement, il fait référence aux indices acoustiques autres que la fréquence fondamentale. Cela a pour résultat une ressemblance formelle entre certaines constructions intonatives (CI-1, CI-2, CI-7). Pour les différencier dans la description, on fera référence à des plans de l'organisation prosodique autres que la mélodie : ainsi la CI-2 est caractérisée par une intensification de l'accent mélodique (dynamique), CI-7 est souvent accompagnée d'un coup de glotte. Pour notre part, nous cherchons un système d'annotation du niveau tonal qui puisse rendre compte de tous les contrastes observés mais qui puisse également proposer une analyse stratifiée en accord avec les conceptions modernes de la prosodie.

3.3.3. Analyse perceptive de l'intonation du russe de Cécilia Odé (1989, 1991)

Le travail de thèse de Cécilia Odé (Odé 1989) a été mené en collaboration étroite avec les chercheurs de l'IPO et s'inspire donc du cadre analytique que nous avons présenté dans la section 2.2. L'objectif de ce travail a été de décrire l'intonation du russe sur la base de mouvements mélodiques élémentaires qui sont nécessaires pour rendre compte des contours pertinents du point de vue de la perception. Le travail s'articule autour de deux axes. Premièrement, on procède à la définition des unités de base qui sont des mouvements mélodiques discrets et à leur description en termes de traits binaires. Ensuite, dans la perspective linguistique, on se pose la question de la combinatoire des mouvements élémentaires en contours mélodiques et celle de leur fonction linguistique.

Le travail s'effectue à partir d'un corpus comportant des enregistrements de différents styles (de parole spontanée et de lecture) avec un vaste panorama d'expériences en perception réalisées pour valider les hypothèses de recherche. C'est la perception qui guide le travail. En termes plus explicites, dans la définition des types pertinents (étape de standardisation), on procède avant tout à une tâche de classification perceptive des mouvements mélodiques. Ce n'est qu'ensuite que l'on élabore la spécification des classes établies en termes de traits, résultant entre autres de l'inspection de la courbe de f0. Comme l'accent lexical est morphologique en russe et par conséquent, mobile, on distingue dans un mouvement mélodique (accent mélodique) les parties structurelles prétonique, tonique, et post-tonique.

Le travail d'analyse d'un corpus et les expériences en perception menées auprès des ocuteurs natifs permettent d'établir treize types d'accents mélodiques, six montants et sept descendants. Les types d'accents mélodiques sont distingués à l'aide de six traits phonétiques :

- direction : permet de distinguer les accents montants *versus* descendants ;
- excursion (annotation : R vs. r ; F vs. f) : indique l'amplitude de l'intervalle mélodique ;

- *timing* (alignement, annotation '–' ou '+') : indique la position de la fin du mouvement mélodique par rapport à la syllabe proéminente, le point de référence étant fixé au début du noyau vocalique ;
- pente du mouvement mélodique : mesure la vitesse du changement des valeurs de f0. Odé considère que ce n'est pas un trait distinctif en russe, dans la mesure où ce paramètre interagit toujours avec le *timing* et la configuration de la partie post-tonique ;
- partie post-tonique : correspond aux syllabes immédiatement adjacentes à la syllabe proéminente. Certains types d'accents mélodiques sont différenciés selon le niveau atteint dans cette partie-là : on distingue le niveau haut *versus* bas *versus* moyen pour les accents montants et le niveau bas et non-bas pour les accents descendants ;
- partie prétonique : les syllabes précédant immédiatement la syllabe tonique ; la configuration mélodique associée avec cette partie du mot peut rendre le mouvement mélodique principal plus saillant.

Nous présentons les configurations mélodiques, ainsi que leurs spécifications phonétiques dans le tableau III.2.

Accents mélodiques ascendants

type	excursion	timing	posttonics	slope	register	picture
Rl-	17 ST (13-21)	89% early 11% late	low	76 ST/s (54-116)	high	
Rh-	17 ST (15-20)	95% early 5% late	high	74 ST/s (35-120)	high	
Rø-	16 ST (13-21)	84% early 16% late	ø	73 ST/s (30-86)	high	
Rm-/+	15 ST (11-17)	70% early 30% late	middle	54 ST/s (39-94)	high	
rm-/+	10 ST (8.5-12)	60% early 40% late	middle	35 ST/s (19-56)	low	
rl-/+	11 ST (9-12)	87.5% early 12.5% late	low	52 ST/s (23-95)	low	

Accents mélodiques descendants

type	excursion	slope	above low	posttonics	picture
Fl-	8 ST (6-11)	47 ST/s (39-71)	0 ST	low	
Fnl-	7 ST (3-16)	42 ST/s (15-62)	4 ST (3-5)	non-low	
Fl+	9 ST (6-13)	47 ST/s (32-60)	0 ST	low	
Fnl+	8 ST (5-11)	50 ST/s (26-83)	4 ST (2-5)	non-low	
Fh-	6 ST (5-7)	35 ST/s (10-58)	4 ST	rises 9 ST to 13 ST above low (7.5-17)	
F^n+	10 ST (8-13)	65 ST/s (55-73)		rising	
f-/+	4 ST (2-6)	28 ST/s (14-57)	6 ST (5-8)	varying	

Tableau III.2 : Types d'accents mélodiques du russe selon Odé (1989).
Légende : R,r : mouvement ascendant ; F,f : mouvement montant ; -/+ : alignement tardif/ ; m,h,l,Ø :position de la partie post-tonique ; lettres capitales :empan mélodique large.

En examinant les spécifications des types accentuels du tableau III.2, nous nous posons la question de savoir dans quelle mesure les spécifications retenues sont dépendantes du choix de l'unité d'analyse, à savoir un mot grammatical. En effet, la distinction qui est faite entre l'accent montant sans partie post-tonique et l'accent montant avec une partie post-tonique haute, est-elle phonologique ou résulte-t-elle du procédé d'alignement entre les niveaux tonal et segmental ? Dans quelle mesure la présence d'un événement mélodique postérieur (accent ou ton de frontière) influence-t-elle les réalisations observées des accents mélodiques ? Rappelons que les conclusions de cette étude sont largement fondées sur les résultats des tests de perception avec les auditeurs natifs ; or, Odé (2003) signale que les jugements d'équivalence perceptive des mouvements mélodiques sont largement influencés par les différences temporelles entre les stimuli ainsi que par les différences fonctionnelles et le nombre d'accents mélodiques dans la séquence.

Actuellement engagée dans le projet ToRI (pour Transciption of Russian Intonation, Transcription de l'intonation du russe), Odé procède à une vérification expérimentale pour savoir si les types d'accents mélodiques attestés couvrent le panorama de tous les accents mélodiques du russe. Des premières analyses (Odé 2003), l'auteur conclut que l'inventaire initial des accents mélodiques ne doit pas être réajusté, s'il s'agit des réalisations proches du prototype de la catégorie. Par contre, les réalisations frontalières induisent l'abaissement de la performance des auditeurs dans les tâches expérimentales (comparaison appariée et classification). Ces résultats suggèrent donc que dans l'espace acoustique, les frontières entre différentes catégories accentuelles s'entrecroisent. L'auteur remarque également que les stimuli utilisés provenaient d'enregistrements faits auprès de la génération des 20-25 ans, ce qui crée une différence de génération par rapport au corpus analysé dans l'étude initiale : la mesure dans laquelle cette observation doit mener à la redéfinition des catégories n'est pas établie, car il n'existe pas d'études analysant les changements intonatifs dans la période de 10 à 20 ans. Pourtant, nous considérons qu'un système d'annotation formelle doit être plus robuste par rapport aux facteurs sociolinguistiques.

3.3.3.1. ToRI (Odé, 2003)

Nous ferons quelques brèves remarques sur le projet d'annotation de l'intonation du russe ToRI.

Ce projet vise à une description détaillée de l'intonation du russe, de ses propriétés phonétiques, phonologiques et fonctionnelles ; c'est la recherche en phonétique expérimentale qui est à la base de cette description qui doit aboutir à un cours interactif de l'intonation du russe (entraînement à la prononciation et à la transcription).

Sans être adepte de l'approche autosegmentale en intonologie, Odé adopte la méthodologie développée dans les études issues de cette mouvance théorique, en s'inspirant surtout du système ToDI. Pourtant, dans la continuité des études précédentes, les symboles dans la transcription visée par Odé ont une valeur perceptive,

la catégorisation des mouvements mélodiques étant issue des expériences en perception avec des locuteurs natifs. L'auteur refuse d'introduire dans le système l'annotation du degré de jonction entre les mots adjacents, un des paliers dans l'annotation ToBI. Cependant le projet de recherche prévoit d'établir une classification perceptive des mouvements mélodiques dont la fonction est de marquer les frontières entre les séquences de mots (les constituants prosodiques, le terme n'étant pas utilisé explicitement).

Les expériences en perception au fondement du système de la transcription proposée répondent à deux problématiques :

b) quêtes des formes pertinentes dans l'intonation du russe, distinguées par les locuteurs natifs : on obtient alors les spécifications phonétiques de ces formes ;

c) codage de la fonction communicative également issu des expériences en perception.

Les étapes de ce projet ambitieux portent sur

a) une vérification expérimentale des types d'accents mélodiques établis dans l'étude précédente de l'auteur ;

b) une étude de la fonction contextuelle de différents types d'accents ;

c) une classification des accents mélodiques et de leurs combinaisons selon les fonctions contextuelles accomplies, suivie d'une vérification expérimentale ;

d) une détermination des tons de frontières pertinents.

Avec la réalisation de ce programme, l'auteur vise à aboutir à une première version du système de transcription : les symboles choisis doivent traduire de manière symbolique le rapport entre la forme phonétique et la fonction des événements intonatifs phonologiques. Dans un deuxième temps, ce système d'annotation devrait être vérifié dans deux tâches appariées d'annotation et de production, ces tâches étant menées auprès d'experts phonéticiens locuteurs natifs du russe.

À ce stade, nous ne connaissons que des résultats de la vérification préliminaire de l'inventaire des accents mélodiques (Odé 2003, 2005).

Pour résumer les réflexions sur l'intonation du russe et les mettre en perspective, Svetozarova (1982) considère le système intonatif comme un système complexe qui admet l'existence d'unités hétérogènes au service de différentes fonctions de l'intonation. Dans cette lignée, le système intégral de l'intonation comportera, selon l'auteur, trois classes d'unités :

- unités associées avec les aspects communicatifs et émotionnels[95] ;

[95] Remarquons l'inclusion dans une description linguistique de l'intonation des aspects émotionnels (cf. l'exclusion de tout ce qui est paralinguistique du domaine de la linguistique, et donc, de la phonologie).

- unités associées avec la segmentation du flux de parole, exprimant le lien entre les différentes parties de l'énoncé. Le lien entre deux syntagmes adjacents peut être fort, faible ou moyen, mais également caractérisé par référence au degré de l'importance relative des syntagmes, au type logique du lien (subordination, coordination, égalité) ou à d'autres significations générales. Une fois associées avec la structure syntaxique de l'énoncé, ces significations générales se transforment en des types concrets de liens inter-syntagmatiques, tels que énumération, comparaison, opposition, explication, avertissement etc. D'autre part, par association avec le contenu lexical ces types deviennent encore plus concrets : on aboutit donc à une description de la façon dont l'intonation permet de distinguer les relations temporelles, conditionnelles, causales, et autres entre les parties de l'énoncé composé.

- Unités permettant de rendre compte des rapports entre les différents éléments au sein d'un même syntagme.

On remarque que les classes d'unités proposées ont rarement le statut des schèmes conventionnels à cause d'un lien étroit avec le contenu sémantique du message. Dans le même esprit, nous considérons qu'une réflexion approfondie sur la définition des unités de l'espace fonctionnel serait nécessaire dans le contexte du projet ToRI, malgré une tradition abondante des analyses fonctionnelles de l'intonation du russe. Cependant, la perspective nous paraît intéressante dans le cadre d'une analyse fondée sur un corpus annoté (cf. Calhoun 2005).

3.3.4. Système d'annotation formelle de S. Kodzasov (1996,1999)

Un système d'annotation formelle a été proposé récemment par Kodzasov (1996) dans le cadre d'un projet de recherche portant sur des caractéristiques prosodiques du parler russe mené par l'Académie des Sciences (Nikolaeva *et al.* 1996, Kodzasov 1999). Dans le cadre de ce projet, on vise un système de transcription des faits prosodiques, le modèle sous-jacent étant qualifié par ses auteurs comme combinatoire. Dans sa version actuelle, ce système est le résultat d'une analyse perceptive qui doit être confirmée par la suite par les données de l'analyse acoustique. La réflexion globale porte donc sur l'inventaire des symboles à utiliser dans l'annotation : on cherche un inventaire de symboles iconiques reflétant les caractéristiques perceptives acoustiques des unités à annoter[96].

Dans le domaine tonal (mélodique), on commence par identifier une grille de référence, couvrant l'empan de deux octaves et comportant 5 niveaux : on distingue le niveau médian le plus typique en discours, le niveau bas, correspondant aux expressions parenthétiques, et le niveau haut comme étant propice à l'expression de l'étonnement. Nous voyons déjà le mélange dans le système d'annotation des traits purement formels et des présupposés relevant de l'espace fonctionnel associé aux indices prosodiques.

[96] Ce choix est en partie motivé par la taille de l'inventaire.

Le modèle dispose du concept opérationnel de la proéminence mélodique (accent mélodique), le définissant comme un mouvement ascendant-descendant associé avec la syllabe forte dans la représentation métrique du mot. Or, le domaine de l'accent mélodique dans la conception de Kodzasov n'est pas réduit à un mot mais à un constituant (le terme russe étant « *составляющая* », *composante*) dont la formation est assurée par des propriétés phonétiques : tous les membres de ce constituant sont réalisés avec une accélération, sans pauses internes, avec un accent lexical affaibli pour les mots lexicaux non associés avec la proéminence mélodique. Chaque accent mélodique possède trois caractéristiques déterminantes : le positionnement du mouvement par rapport au noyau vocalique, la vitesse du changement mélodique, et l'empan tonal déterminé par référence à la grille des niveaux mélodiques. Dans un premier temps, on différencie deux classes d'accents mélodiques : accents de niveaux et configurations accentuelles. Le système proposé comporte 6 accents principaux (nous résumons la description de ces accents dans le tableau III.3).

Lors de l'annotation des corpus oraux, on peut rencontrer des variantes des accents ascendants et descendants : une réalisation de l'accent ascendant avec un plateau haut allongé (contour continuatif), mais aussi une variante du ton descendant avec la montée étant également réalisée sur le noyau vocalique[97].

Le tableau ci-dessus résume les mouvements mélodiques de base ; des sous-classes sont définies par la suite :

a) les accents combinés (utilisés pour exprimer des significations illocutoires complexes (demande, question, affirmation)) ;

b) les doubles accents (caractérisés par un intervalle mélodique augmenté) ;

c) les accents aigus (caractérisés par une impulsion tonale brève, de faible amplitude, localisée dans le noyau vocalique) ;

d) les petits accents (avec empan réduit) ;

e) les demi-accents (le mouvement mélodique n'est pas finalisé, on ne revient pas au niveau de base).

Nous voyons donc que cette typologie n'est pas purement formelle. Elle va également à l'encontre de la vision non-phonologique du paramètre de l'amplitude du mouvement mélodique (Liberman & Pierrehumbert 1984)[98].

Les accents mélodiques dont nous avons discuté jusqu'à présent forment les primitives de la transcription intonative. À l'étape suivante, on définit les structures composées nommées compositions d'accents : il s'agit des cas de fusionnement de deux accents dans une seule figure tonale (sept figures tonales sont ainsi définies).

[97] Remarquons que dans la description proposée cette configuration ne correspond pas à une classe à part.
[98] La vision phonologique du registre est spécifique pour un domaine plus large (cf. Ladd 1992).

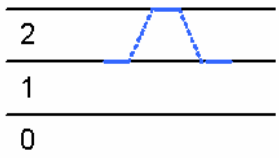
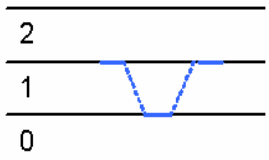
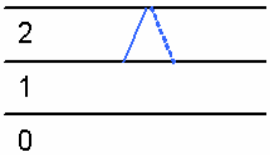
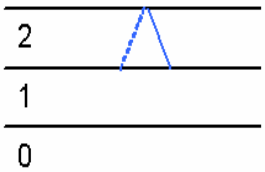
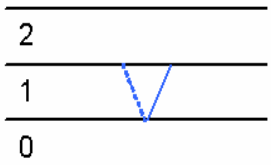
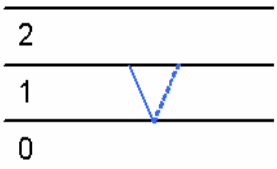
Représentation graphique	Exemple	Description (caractéristique dans le domaine perceptif)
« accent de niveau positif »		- défini comme un saut du niveau de base vers un niveau supérieur ; - typique pour des questions à morphème interrogatif ; - possède deux variantes selon l'alignement (le point culminant étant aligné avec le début ou la fin du noyau vocalique)
« accent de niveau négatif »		- est caractérisé par la perception d'un saut du niveau de base vers un niveau inférieur ; - typique pour les affirmations avec une connotation d'évidence
« accent ascendant[99] »		- est caractérisé par la perception d'un glissando ascendant sur la voyelle accentuée ; - suivi par le retour au niveau de base ; - typique pour une question générale
« accent descendant »		- est caractérisé par la perception d'un glissando descendant sur la voyelle accentuée ; - est précédé par une montée à partir du niveau de base ; - typique pour des réponses à une question générale et pour les assertions
« accent ascendant dans le registre bas »		- est précédé par un abaissement par référence au niveau de base ; - a la signification d'évidence
« accent descendant dans le registre bas »		- est suivi par le retour au niveau de base ; - a la signification d'évidence

Tableau III.3 : Accents toniques du russe selon Kodzasov 1996

[99] La distinction entre l'accent de niveau positif et l'accent ascendant réside dans la raideur des mouvements respectifs ; les différences disparaissent dans le discours peu soigné.

3.4. RÉSUMÉ ET DISCUSSION

Ce chapitre résume les principaux travaux sur l'intonation du russe. Nous avons distingué et discuté de deux groupes d'approches, les approches fonctionnelles et les approches formelles. Ces deux groupes se réfèrent à la même unité de description, le contour intonatif aligné avec un syntagme. Cependant, ils différent fondamentalement dans la gestion des rapports entre le plan de l'expression et le plan du contenu pour les unités décrites.

Les approches fonctionnelles présentent des nomenclatures non-exhaustives en puisant les étiquettes de catégories dans le domaine de la syntaxe communicative et constructive ou dans les études à orientation psychologique. Toutefois, ces descriptions n'intègrent pas un niveau phonologique de représentation. Si l'on envisage de poursuivre cette mouvance dans la démarche analytique, le danger est d'aboutir à une description très fragmentaire, dépendante du contenu sémantique du message. Il est évident que le fonctionnement des patrons intonatifs dans le discours doit être analysé dans toute sa profondeur : or, à notre avis, la valeur pragmatique des contours doit être appréhendée, ayant recours à un modèle de discours ; parallèlement, dans cette analyse, on doit intégrer des catégories phonologiques formelles pour le plan d'expression (par exemple, des primitives phonologiques de Hirst (2005) dont nous avons discuté dans la section 1.3).

Les approches formelles, quant à elles, ne sont pas nombreuses. Nous avons discuté l'approche perceptive proposée par Odé (1989) qui introduit le concept d'accent mélodique : nous considérons que la nomenclature des accents mélodiques proposée dans cette étude est fortement dépendante de l'unité minimale d'analyse, à savoir le mot. Le concept d'accent mélodique est repris dans l'étude de Kodzasov, laquelle accorde pourtant plus d'attention à l'annotation des aspects d'implémentation phonétique. De même, le système d'annotation élaboré ne se limite pas au domaine tonal ; les phénomènes temporels (allongement et distribution des pauses), comme ceux de la qualité de voix y sont également traités.

Le système des constructions intonatives de Bryzgounova (1977) se concentre sur la description du mouvement mélodique dans le nucléus. Nous y observons une certaine synonymie au niveau formel comme fonctionnel entre différentes constructions de base : cette synonymie est due au fait que les indices acoustiques autres que la fréquence fondamentale sont pris en compte dans la définition des unités. Certes, les différents indices acoustiques agissent conjointement pour véhiculer des significations différentes propres aux structures prosodiques ; cependant, nous considérons que le cadre méthodologique de la phonologie intonative, qui incorpore à la fois des concepts de la phonologie métrique et ceux de la phonologie autosegmentale, assure une gestion rigoureuse de tout détail d'implémentation acoustique, permettant en même temps de disposer d'un système d'annotation réservé au strate tonal.

Nous constatons donc qu'aucune description formelle proposée jusqu'à présent n'implique les principes sous-jacents de la phonologie intonative. De même, lorsque le

concept de la proéminence mélodique y est incorporé (citons également les travaux de Igarashi 2002), le concept du ton de frontière est un grand absent[100].

Comme le remarque Yokoyama (2003), beaucoup d'études portant sur des nuances sémantico-intonatives fines ont été réalisées sur la base du système des constructions intonatives de Bryzgounova, toutefois, les fondements théoriques de cette approche n'ont jamais fait l'objet d'une réflexion approfondie. Pour cet auteur, le paradigme autosegmental-métrique permet une introduction dans le domaine de l'intonation de l'opposition tonème ~ alloton (réalisation contextuelle de la f0), analogue à celle que l'on trouve entre phonème et allophone en phonologie segmentale.

Yokoyama (1987, 2003) propose une description formelle de l'intonation qui se rapproche le plus de la phonologie intonative. L'auteur réunit dans une même description (et annotation) les accents mélodiques et les accents de phrase (*sentential stress*, « *логическое ударение* », distribution du focus) : remarquons que les phénomènes d'emphase et de focus étaient toujours traités en dehors des descriptions du système intonatif du russe[101] (cf. Nikolaeva, 1982)[102]. Notons également que la tradition décrivait l'accent focal en termes de durée et d'intensité, sans faire référence aux phénomènes mélodiques. Dans son analyse, Yokoyama différencie le niveau phonémique et le niveau phonétique dans la description élaborée (correspondant au niveau phonologique et phonétique dans notre terminologie). L'auteur propose un schéma standard où tout syntagme non-terminal affiche un contour montant, le contour ascendant étant attribué au syntagme final d'un énoncé assertif, cf. :

$[LH]_n$ [HL]

L'auteur introduit également la notion du ton de frontière.

À partir des descriptions que nous avons abordées dans le présent chapitre, nous constatons également que le sujet du phrasé prosodique a reçu une attention moindre dans les travaux sur l'intonation du russe. Certains exemples que nous avons pu recenser dans la littérature démontrent clairement une situation floue. Ainsi nous trouvons chez Yokoyama (1987) l'analyse de l'énoncé « *Тут треножник* » (*Ici un trépied*) en deux syntagmes dans un contexte non-contrastif. Le même auteur considère que l'énoncé suivant :

[100] Sauf, peut-être, dans l'analyse de Nikolaeva (1969) ; or, cette analyse ne traite pas exclusivement de l'organisation tonale ; l'accent est également mis sur l'analyse du lien logique entre les syntagmes et sur les rapports entre les deux syntagmes adjacents (caractéristique globale du registre et de l'étendue tonale).
[101] Quoique la construction IC-2 de Brygounova puisse être interprétée comme associée à la structure informative et à la distribution du focus.
[102] Fougeron (1989) analyse la diversité de la terminologie et des interprétations proposées dans la tradition russe pour les phénomènes d'accent de phrase (dernier accent syntagmatique dans l'énoncé) et d'accent logique (contrastif, focus étroit). De son côté, procédant à une analyse acoustique, l'auteur ne maintient la distinction qu'entre accent nucléaire (dernier accent mélodique du syntagme) et accent exclusif (qui exprime l'opposition entre deux termes d'un même paradigme et possède une connotation émotionnelle).

(2) В нашу комнату || вошла пожилая женщина || в больших || мужских || сапогах.

Dans notre chambre || est entrée une femme âgée || chaussée de grandes|| bottes || masculines.

est très probablement prononcé en cinq syntagmes.

Cette situation résulte, à notre avis, de l'absence d'un cadre formel et strict pour l'analyse du phrasé prosodique, un cadre néanmoins autonome vis-à-vis de la syntaxe. Cependant, nous estimons qu'un système intégral de l'organisation prosodique ne peut résulter que d'une réflexion conjuguée sur l'organisation tonale et sur le phrasé prosodique. Or, nous sommes persuadée qu'une entreprise de formalisation du système prosodique se doit de conjuguer l'étude de la dimension tonale et celle du phrasé prosodique. Par conséquent, l'un des objectifs de notre travail sera d'établir des schémas d'emplacement des frontières prosodiques, de décrire les influences des contraintes de différents niveaux (prosodiques, syntaxiques et discursives) sur le phrasé prosodique dans le cadre de la phonologie des domaines.

CHAPITRE IV

LE PHRASÉ PROSODIQUE

Ce chapitre est consacré à la présentation et à la discussion de la phonologie des domaines dont l'objectif consiste notamment à formaliser l'interface « prosodie ~ syntaxe ». Le phrasé prosodique est conçu par la phonologie des domaines comme une structure hiérarchique de constituants qui étaye sous la forme d'une architecture multiniveaux la modélisation des interfaces entre la phonologie, d'une part, et d'autre part, la syntaxe et la pragmatique pour un degré moindre. Les constituants prosodiques ont une réalité cognitive dans la mesure où nous percevons à l'oral la segmentation du continuum sonore en blocs ('*chunks*', terme de Bolinger 1989) : cette structuration est signalée à l'oral par un ensemble d'indices acoustiques. Nous commençons par situer cette mouvance théorique par rapport à la phonologie générative classique avant de passer en revue les niveaux structurels définis : nous essayons également de déterminer la pertinence des constituants décrits dans la littérature en ce qui concerne l'analyse de la hiérarchie prosodique propre à la langue russe. Nous observons également que l'une des critiques majeures qu'il est possible de formuler à l'encontre de la phonologie prosodique porte sur les restrictions stylistiques que les différents auteurs sont amenés à faire afin de justifier leurs analyses : dans cette perspective il nous semble intéressant de tester les présupposés de la phonologie prosodique sur un corpus de parole spontanée. Nous y consacrons une partie de nos analyses empiriques.

4.1. INTRODUCTION

Quand une proposition écrite est réalisée à l'oral, elle acquiert des caractéristiques qui ne peuvent pas être directement ni exhaustivement inférées à partir du texte écrit. À l'oral, un énoncé est caractérisé par le choix du contour mélodique, par les distributions des durées segmentales et celles des pauses, et par des variations de la prononciation des mots si l'on compare la prononciation réelle d'un mot avec sa forme de citation dans le dictionnaire. Certains phénomènes phonologiques ont pour domaine d'actualisation un constituant supérieur au mot : ils peuvent relever du niveau segmental (par exemple, la liaison en français, ou le *raddoppiamento sintattico*[103] en italien) ou être liés au domaine

[103] Il s'agit d'un phénomène d'allongement que l'on observe en italien : si pour deux mots adjacents le premier se termine avec une voyelle, sa dernière syllabe porte l'accent lexical et si le second mot

tonal ou à celui de l'accentuation[104]. En même temps, le flux de parole est organisé en blocs (*chunks*)[105] et on perçoit une interruption au point de leur achèvement : la contribution au phrasé du flux de parole est une des fonctions les plus importantes de la prosodie. Les traits prosodiques, tels que la quantité, le tonal, l'accentuation, sont utilisés à l'oral pour assurer la cohérence et la mise en forme des unités lexicales et post-lexicales. En raison de notre intérêt pour le phrasé prosodique, nous sommes donc amenée à réfléchir sur l'articulation et l'actualisation à l'oral des interfaces complexes entre différents modules de la grammaire, et notamment de l'interface « prosodie – syntaxe ».

Pendant une longue période de l'histoire de la linguistique, la prosodie a été négligée par la linguistique (cf. Di Cristo, 2004) et que les modèles phonologiques des phénomènes prosodiques n'existaient pas,[106] les phénomènes post-lexicaux ont donc été décrits par des linguistes essentiellement dans les études phonético-acoustiques (Lehiste, 1973 ; Klatt, 1976). Dans le même esprit, faisant l'historique des études sur l'interface « prosodie – syntaxe », Vaissière (1997) remarque que l'orientation théorique des descriptions proposées dépend essentiellement de l'objectif final de l'étude. Ainsi, avant la venue de la théorie générative, le travail linguistique a été mené essentiellement à partir des corpus de parole spontanée, et le phénomène de la structuration prosodique a reçu une interprétation en relation avec la structure discursive ; on hérite d'ailleurs de cette époque du concept de l'unité intonative en tant qu'unité sémantiquement cohérente. Dans la tradition de la grammaire générative, le corpus de travail n'est plus de la parole spontanée mais de la lecture oralisée, ce choix méthodologique permettant une manipulation plus facile de la structure syntaxique. La recherche s'articule alors autour de la question du lien entre les composants phonologique et syntaxique de la grammaire (*Sound Patterns of English* 1968). Le changement de paradigme se produit dans les années 1970-1980 quand émergent les théories linguistiques qui caractérisent les phénomènes prosodiques et les phénomènes segmentaux post-lexicaux faisant référence à la hiérarchie des constituants prosodiques et/ou à la hiérarchie des proéminences. La phonologie prosodique émergente est donc le résultat d'une nouvelle réflexion sur l'interface « prosodie – syntaxe » : en se détachant de ses prédécesseurs, elle postule l'existence d'une hiérarchie prosodique distincte de la hiérarchie morphosyntaxique, quoique étroitement liée à cette dernière.

Les recherches phonologiques acquièrent une nouvelle orientation avec l'adoption de l'approche de la phonologie de laboratoire (Pierrehumbert *et al.* 2001 ; D'Imperio 2005) : on adopte dorénavant la pratique de la vérification expérimentale des hypothèses phonologiques, fondées par le passé sur les intuitions linguistiques des locuteurs. Nous disposons, donc, à présent, de trois types de données qui corroborent la vision hiérarchique structurelle de la prosodie : les arguments théoriques phonologiques, les données acoustiques quantitatives et les données comportementalistes, issues des

---

commence par une consonne, la consonne se voit allongée. La réalisation de ce phénomène est sensible à la présence de frontières prosodiques d'un certain niveau (syntagmes phonologiques, cf. *infra*).

[104] Sous forme de la gestion des distributions des proéminences mélodiques et/ou métriques.

[105] Pour l'instant, on laisse de côté la question de savoir si ces blocs (*chunks*) sont impérativement dotés de sens.

[106] Surtout pour le niveau tonal : la réflexion phonologique se limitait à l'établissement d'une liste des contours mélodiques distinctifs, or on y retrouvait rarement la dissociation entre forme et fonction.

expériences psycholinguistiques en perception et en compréhension de l'oral. Dans la suite du présent chapitre, nous présentons et analysons ces différents types d'arguments. Il est important de préciser que de nombreuses études en phonologie prosodique portent sur l'anglais, le japonais, les langues romanes, ainsi que les langues africaines à tons; à notre connaissance, les études sur la langue russe ont été exclues de cette mouvance. Par conséquent, nous présentons les considérations qui nous amènent, dans la présente étude[107], à accepter ou à réfuter certains axiomes de la phonologie prosodique, par rapport à leur transfert dans la langue russe.

4.2. L'INTERFACE PROSODIE-SYNTAXE DANS LA TRADITION GÉNÉRATIVE LINÉAIRE (SPE)

Avant la mise en forme de la phonologie générative linéaire, pour rendre compte des phénomènes post-lexicaux on se limitait à une simple énumération des contextes phonologiques et syntaxiques. Ce paradigme a subi une grande modification avec le développement du modèle générativiste. Une conception modulaire de la grammaire, propre à ce courant, a engendré une re-analyse et une formalisation des relations entre la phonologie et la morphosyntaxe.

Dans l'architecture de la théorie générativiste, la forme d'entrée du module phonologique correspond à la forme de sortie du module syntaxique dans lequel la structure syntaxique annotée est générée pour chaque énoncé. Ce sont les règles transformationnelles (ou de réécriture)[108] qui constituent le formalisme principal de la phonologie générative. Les informations syntaxiques sont accessibles à de telles règles sous la forme de différents types d'étiquettes de frontières qui sont incluses dans la chaîne de segments. Ainsi, dans la phonologie générative linéaire le statut des frontières est-il le même que celui des segments, et ce, même si le trait [-segment] leur est attribué. La même approche a déjà été adoptée par des linguistes de la tradition structuraliste : les frontières dans leur description avaient le même statut que les segments dans la chaîne sonore. Ce qui distingue donc la tradition générative linéaire est la définition des positions des frontières en relation directe avec la morphosyntaxe.

La règle d'insertion des frontières dans la phonologie générative a la forme suivante :

> *« The boundary # is automatically inserted at the beginning and end of every string, dominated by a major category, i.e., by one of the lexical categories "noun", "verb", "adjective", or by a category such as "sentence", "noun phrase", "verb phrase", which dominates a lexical category. »* (SPE 1968:366)

Dans cette approche, la force d'une frontière est exprimée de manière quantitative : plus il y a de symboles de frontière après un mot, plus forte est la frontière syntaxique et phonologique qui le suit. La structure résultante peut être assez complexe ; Selkirk (1972) propose une autre règle de simplification des frontières qui efface les symboles

[107] Nous considérons que des études plus approfondies sont nécessaires pour développer la phonologie prosodique de la langue russe.
[108] Les règles transformationnelles se résument à une lecture « A devient B dans le contexte X_Y ».

de frontières internes. On distingue pourtant les différents types de frontières suite à la différence faite entre la morphologie dérivationnelle et la morphologie flective[109] (Carr 1999). Ainsi on distingue les frontières clitiques, à mi-chemin entre les frontières des morphèmes et les frontières des mots lexicaux (les doubles frontières). De ce fait, on trouve une interprétation implicite du degré de jonction entre les éléments terminaux dans la structure syntaxique.

On retient donc dans l'analyse phonologique la distinction entre mots lexicaux et grammaticaux, on s'appuie sur les informations provenant de la syntaxe qui concernent la structuration syntaxique, et non pas des relations entre les différents constituants syntaxiques.

Nous pouvons dire, en résumé, que dans l'approche de la phonologie générative linéaire, le lien entre les composants phonologique et syntaxique est unilatéral : la forme de sortie du module syntaxique constitue la forme d'entrée pour le module phonologique ; en revanche, aucune information phonologique n'est accessible au module syntaxique. Par conséquent, le module phonologique n'a accès qu'à une représentation phonémique de surface avec des étiquettes de frontières insérées.

D'une telle représentation, comme le note Delais-Roussarie (2005:7), il est possible de dériver une structure intermédiaire entre la phonologie et la syntaxe :

(1) a) # # son # grand # # ami # # a # acheté # # un # livre # #

b) [(son)(grand)][(ami)][(a)(acheté)][(un)(livre)]

Ainsi, à partir d'une structure issue des règles de réajustement, ayant des étiquettes de frontières insérées, on obtient une structure à deux niveaux. Pourtant, cette vision n'est pas explicite dans SPE.

Cette structure intermédiaire entre les modules syntaxique et phonologique, plus complexe que la structure présentée ci-dessus (1), est introduite et formalisée dans les analyses de la phonologie prosodique (Nespor et Vogel 1986, Hayes 1989, Selkirk 1984) qui défend l'autonomie d'une telle représentation et sa pertinence pour l'analyse linguistique.

Dans le passage du modèle de la phonologie générative linéaire à la phonologie prosodique, la réflexion principale s'articule autour du lien entre la phonologie et la syntaxe. On s'interroge, plus particulièrement, sur le caractère médiatisé ou direct de ce lien, sa direction (s'il s'agit d'un lien unilatéral ou bilatéral : dans quelle mesure les informations syntaxiques sont accessibles en phonologie et dans quelle mesure les informations phonologiques sont accessibles pour la syntaxe), ainsi que sur la nature des informations syntaxiques pertinentes pour la formalisation des phénomènes phonologiques post-lexicaux.

[109] Prenons comme exemple de la morphologie dérivationnelle, la formation des formes du féminin pour les substantives et les adjectives en français. La morphologie flective est une morphologie régulière et explique également la formation des formes du pluriel, ainsi que les formes des temps différents pour les verbes. Dans la morphologie dérivationnelle, on procède à la formation d'un nouveau mot, comme dans la paire « volcan – volcanique ».

Dans la mouvance de la phonologie prosodique, la structure prosodique est déterminée par le composant prosodique de la grammaire ; elle est influencée par des facteurs syntaxiques, en premier lieu, or, l'influence des facteurs autres que la syntaxe, des facteurs linguistiques comme extralinguistiques, n'est pas exclue. Globalement, c'est la hiérarchie prosodique qui détermine la réalisation phonético-acoustique de l'énoncé.

On note également que bien que l'analyse de l'accentuation (= organisation des proéminences relatives) et des phénomènes segmentaux associés à des frontières morphosyntaxiques (= groupements) soit proposée dans le cadre de la phonologie générative, un intérêt porté aux phénomènes prosodiques y est assez réduit. Le changement du paradigme a donc pour conséquence l'élargissement du domaine de compétence de la phonologie.

4.3. PHONOLOGIE PROSODIQUE : APERÇU FORMEL DE LA PROSODIE

Cette section est consacrée à la présentation générale de la phonologie prosodique.

Comme nous l'avons noté précédemment, vers les années 1980 on écarte l'idée de l'isomorphisme entre l'organisation prosodique de l'énoncé et sa structuration en termes de constituants morphosyntaxiques. L'identification directe des deux structures se révèle à la fois insuffisante et surpuissante[110]. L'insuffisance du lien direct entre la prosodie et la syntaxe se manifeste, d'une part, certains concepts et relations syntaxiques importants ne sont pas signalés par la prosodie ; d'autre part, certains phénomènes prosodiques ne peuvent être prédits par la seule structure syntaxique :

> *« It has become a majority view among researchers […] that syntax does not provide domains for phonological rules in a direct fashion. Both the impoverished amount of syntactic information needed by the phonological module and the variety of mismatches between phonological rule domains and syntactic constituency argue for posting another level of representation ». (*Inkelas et Zec, 1995 : 537)

On peut considérer comme travail fondateur de la phonologie prosodique l'étude de Selkirk (1978) (les principales hypothèses ont été élaborées ensuite par Nespor et Vogel 1982) : cette étude s'inspire beaucoup de la réflexion de la phonologie métrique sur l'organisation des proéminences dans le cadre d'un énoncé (Liberman, 1975 ; Liberman et Prince, 1977). D'autres travaux présentant plus d'arguments en faveur du nouveau formalisme pour l'interface « phonologie – syntaxe » sont apparus quasi simultanément (Selkirk, 1984, 1986, 1995b ; Nespor and Vogel, 1986 ; Hayes, 1989). Leurs auteurs défendent le caractère compositionnel et complexe du composant phonologique qui, pour eux, n'est pas homogène : cette approche s'inscrit dans le cadre général de la phonologie plurilinéaire[111]. On distingue systématiquement les phénomènes lexicaux et

[110] Il est difficile d'imaginer une règle qui préconise un allongement préfrontalier dans les syntagmes adjectivaux, différent de celui que l'on observe dans des syntagmes nominaux.
[111] Le terme « plurilinéaire » signifie que les informations phonologiques sont réparties en niveaux distincts mais reliés entre eux.

post-lexicaux, le domaine du tonal et de l'accentuation, et on propose un modèle phonologique pour chacun des sous-systèmes du composant phonologique. Cette séparation des niveaux d'analyse induit le développement des différentes branches de la phonologie : phonologie lexicale, théorie de grille métrique, phonologie autosegmentale et phonologie prosodique.

Pour rendre compte de l'organisation structurelle de l'énoncé à l'oral, la phonologie prosodique propose un modèle qu'on peut qualifier de théorie des domaines phonologiques[112] : étant donné une séquence de sons de parole, ce modèle lui impose une organisation hiérarchique des constituants phonologiques qui, à leur tour, constituent le contexte d'application des règles phonologiques.

Rang	Catégorie	Symbole
1	énoncé prosodique	U/E
2	syntagme intonatif	π
3	syntagme phonologique	γ
4	groupe clitique	C
5	mot prosodique	ω
6	pied	F
7	syllabe	σ

Tableau IV.1 : Hiérarchie des constituants prosodiques

La structure hiérarchique des constituants prosodiques sert d'interface entre les composants syntaxique et phonologique de la grammaire.

Le tableau IV.1 résume les 7 catégories prosodiques proposées dans l'étude de Nespor et Vogel (1986). Les auteurs veulent que cette liste ait vocation universelle (1986 : 11) : on ne trouve pas dans les langues du monde de catégories prosodiques autres que celles qui sont énumérées ci-dessus. De la même manière, les omissions des niveaux dans la hiérarchie prosodique ne sont pas autorisées : dans la génération de chaque niveau de la hiérarchie prosodique, on met ce niveau en relation avec le niveau d'un autre composant de la grammaire ; si un niveau est absent dans la hiérarchie prosodique, cela implique la non-pertinence d'une interface entre deux modules de la grammaire pour la langue donnée, situation non autorisée dans la tradition générative. En outre, les omissions des niveaux affaiblissent le pouvoir prédictif de la théorie. Notons, pourtant, que sur ce point, les approches de Nespor et Vogel (1986) et de Selkirk (1986) divergent : pour Selkirk, s'il n'existe pas de règles phonologiques faisant référence à un niveau donné dans la hiérarchie prosodique, ce niveau de structuration n'est pas pertinent dans cette langue. Pour notre part, nous adhérons à la position de Selkirk, également défendue par Di Cristo (2004), à savoir que la pertinence de chaque niveau de la hiérarchie prosodique, sa raison d'être, doit impérativement être mise en évidence dans la recherche des propriétés formelles et des marques acoustiques qui conduisent à son identification. C'est en adoptant cette démarche que nous présenterons par la suite les différents niveaux de la

[112] Spécifions pourtant que Delais-Roussarie et Fougeron (2004) mettent en garde la communauté linguistique contre la confusion entre deux types de groupement qui sont associés respectivement avec des notions de domaine (d'application d'une règle) et de constituant (structurel).

hiérarchie prosodique et que nous avancerons des arguments pour comprendre quels niveaux de la constituance prosodique sont pertinents pour la langue russe.

Nous devons cependant faire une remarque sur les divergences dans les structures proposées au fil de la progression des recherches dans le cadre de la phonologie prosodique. La figure 4.1 résume les différentes architectures et concepts et illustre les liens établis entre eux (l'analyse s'inspire de Shattuck-Hufnagel et Turk, 1996, et de Di Cristo, 2004).

Dans une perspective comparative, évaluant l'opinion de la communauté scientifique par rapport à la pertinence des différents niveaux de la hiérarchie prosodique, les conclusions des différents auteurs divergent. Pour Shattuck-Hufnagel et Turk (1996), les niveaux inférieurs de la hiérarchie prosodique ainsi que celui du syntagme intonatif suscitent moins de débats que les niveaux médians et sont, donc, majoritairement reconnus. De son côté, Inkelas & Zec (1995:539) notent qu'une grande partie des études dans le cadre de la phonologie prosodique portait sur les niveaux du mot phonologique (prosodique) et du syntagme phonologique dans les langues du monde : ces études témoignent de plus de cohérence pour ces niveaux, surtout dans leur relation avec des catégories morphosyntaxiques. Pour ces auteurs, c'est la catégorie du syntagme intonatif qui est sujette à des désaccords et à des difficultés de formalisation dans le cadre de la phonologie prosodique, étant donné le rôle important de la sémantique et de la pragmatique dans sa formation.

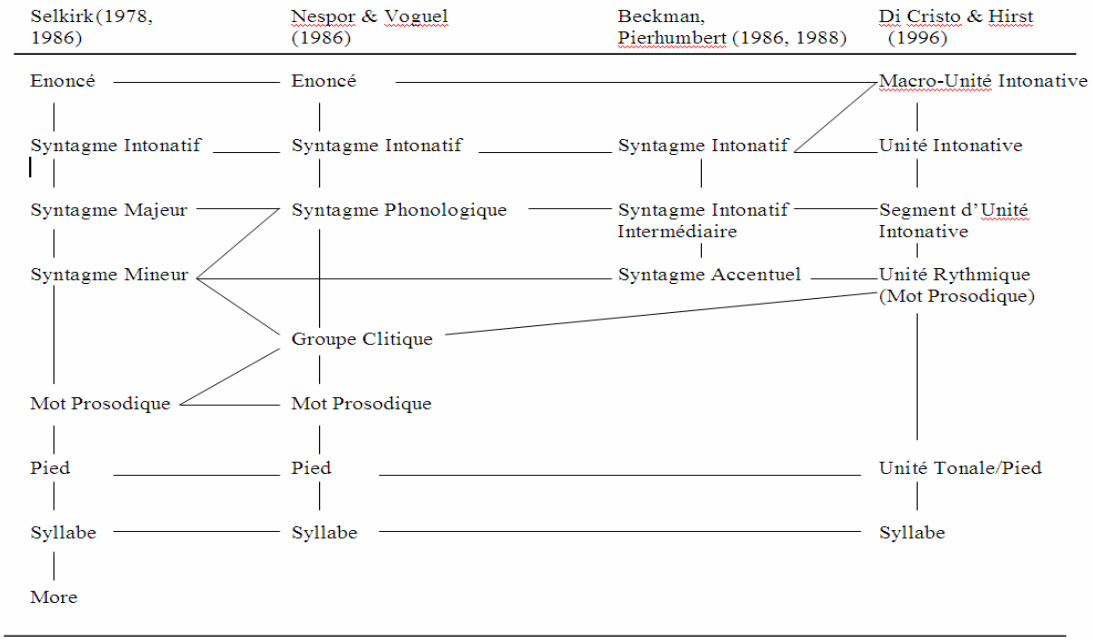


Figure 4.1 : Correspondances possibles entre différentes constituances prosodiques proposées dans la littérature (tableau inspiré de Shattuck-Hufnagel et Turk 1996 et de Di Cristo 2004)

Par rapport à la figure 4.1 qui représente les différentes architectonies de la hiérarchie prosodique, nous tenons à faire les remarques suivantes. Tout d'abord, certains chercheurs comme Inkelas et Zec (1995:549) refusent d'inclure des constituants inférieurs au mot phonologique dans la hiérarchie prosodique : leur argumentation porte

sur les différences importantes attestées dans le fonctionnement des unités métriques par rapport aux constituants des niveaux supérieurs (le rôle principal de ces derniers est de construire le domaine d'application des règles phonologiques post-lexicales). Ces auteurs proposent de distinguer deux hiérarchies, l'une métrique et l'autre touchant aux constituants prosodiques.

Les structures hiérarchiques de la figure 4.1 représentent les différentes approches vers une définition des constituants. Comme le note Gussenhoven (1988), on peut identifier deux étapes dans l'émergence de la phonologie prosodique : au début, on voit dans la hiérarchie des constituants prosodiques le mécanisme formel d'interface « syntaxe – phonologie ». Puis, dans le deuxième groupe de travaux, la problématique de la constituance prosodique est abordée à partir de l'analyse du contour mélodique et de sa structure (Selkirk 1984, Beckman et Pierrehumbert 1986, Ladd 1986)[113]. Par rapport à cette deuxième mouvance, Gussenhoven (1988, 1990) soulève la question du lien entre la hiérarchie prosodique et les constituants définis à la base des indices tonals (intonatifs). Pour lui, les constituants définis à la base des indices tonals constituent un système séparé, externe par rapport à la hiérarchie des constituants prosodiques. Dans les études expérimentales, on est amené à reconnaître l'absence de correspondance biunivoque entre les unités définies à partir des critères pausaux et celles définies à partir des configurations tonales.[114] Pour Gussenhoven, la solution consiste dans la reconnaissance comme une structure séparée de l'ensemble des domaines d'association tonale (*tonal association domaines*) : chaque unité de cette structure (mot tonal) est définie par rapport à la distribution des accents mélodiques dans un énoncé. Cette structure est mise en relation avec la hiérarchie des constituants prosodiques, de manière que les frontières des domaines d'association tonale ne correspondent pas obligatoirement aux frontières d'un niveau précis de la hiérarchie prosodique. La structure prosodique classique a pour vocation de contribuer à la structuration des événements temporels dans l'énoncé ; par son caractère arborescent, elle crée une structure hiérarchique des frontières. Du fait de la coïncidence d'une frontière de constituant prosodique avec la frontière du domaine d'association tonale, la première se retrouve renforcée, et il en résulte une modification de la hiérarchie des frontières. Dans la suite de cette présentation de la phonologie prosodique nous adoptons le point de vue d'une seule hiérarchie des constituants prosodiques avec un niveau de constituance réservé aux syntagmes (unités) intonatifs(ves).

Par conséquent, au tout premier abord, nous considérons que la constituance prosodique a une valeur explicative pour le paramètre de force de la frontière perçue dans une position précise de l'énoncé. C'est-à-dire plus haut est le rang du constituant dans la hiérarchie prosodique, plus proéminente est la frontière (et, par conséquent, plus explicite son signalement à l'aide des indices acoustiques associés).

Dans la réflexion sur la validité de la constituance prosodique, on cite souvent des arguments l'intégrant aux représentations mentales des locuteurs : on considère que

[113] C'est par la suite, quand le système de la transcription de ToBI a été proposé, qu'on a établi une relation stricte entre les caractéristiques mélodiques et les degrés de jonction entre deux mots consécutifs, qui a abouti à une mise en relation entre la syntaxe et la phonologie intonative.

[114] En particulier, c'est cette observation empirique qui a amené Beckman & Pierrehumbert (1986) à distinguer deux unités intonatives – syntagme intermédiaire et syntagme intonatif plein (*full intonation phrase*).

c'est la représentation mentale de la parole, en production comme en perception, qui est structurée en termes de composants hiérarchiquement organisés. Cette voie a été largement exploitée dans des études psycholinguistiques. Une autre branche de la recherche porte sur la quête des marques (acoustiques, segmentales et prosodiques) utilisées par les locuteurs pour signaler les différents constituants prosodiques. Nous analyserons les apports de ces études dans le chapitre V.

4.4. LA RÉCURSIVITÉ ET L'HYPOTHÈSE DE L'ÉTAGEMENT STRICT DANS LES THÉORIES PROSODIQUES[115]

Nous avons vu, dans la section précédente, la diversité des architectures des hiérarchies prosodiques, proposées par les différents auteurs. Nous en proposons dans la suite de cette section une analyse typologique.

Le critère sous-jacent de la première typologie des théories prosodiques fait référence à des structures récursives en fonction de leur admission ou de leur exclusion totale dans la hiérarchie prosodique. Le principe de non-récursivité de la hiérarchie prosodique est formalisé dans l'adoption de l'hypothèse de l'étagement strict (*Strict Layering Hypothesis*, Selkirk, 1984 ; Nespor et Vogel, 1986). Selon cette hypothèse, formalisant les rapports entre les constituants des niveaux adjacents, le constituant du niveau *i (non terminal dans la hiérarchie)* ne peut dominer que le(s) constituant(s) du niveau *i-1* :

> « **Strict Layer Hypothesis** : The categories of the Prosodic Hierarchy may be ranked in a sequence C_1, C_2,… C_n, such that
>
> a. all segmental material is directly dominated by the category Cn, and
>
> b. for all categories C_i, i ≠ n, C_i directly dominates all and only constituents of the category C_{i+1}. » (Hayes, 1989:204)[116].

Comme nous l'avons précisé plus haut, la hiérarchie prosodique se veut autonome par rapport à la hiérarchie syntaxique : l'adoption du principe d'étagement strict assure en partie cette autonomie, dans la mesure où la structure syntaxique peut être récursive. La

[115] L'hypothèse de l'étagement strict sous-tend les travaux de Beckman et Pierrehumbert (1986), Selkirk (1986), Nespor et Vogel (1986), Hayes (1989), Beckman et Edwards (1991), Wightman *et al.* (1992). Nous retrouvons les structures récursives dans les analyses de Liberman et Prince (1977), Ladd (1986, 1988, 1996). L'attrait des modèles récursifs réside dans leur caractère local ; les distinctions graduelles sont générées à partir de la distinction binaire entre le nœud fort et faible. Il en résulte une structure binaire. Or les analyses avec des structures récursives nécessitent une projection dans le futur (*lookahead*). Les structures issues de la théorie d'étagement strict sont plus simples, mais parfois le modèle théorique se révèle inadéquat par référence à des phénomènes de surface observables.

[116] Dans le travail de Nespor et Vogel (1986:7), les constituants prosodiques sont regroupés en une structure arborescente selon les quatre principes qui déterminent la géométrie d'un tel arbre :
- une unité non-terminale du niveau *i* est composée d'une ou plusieurs unités du niveau *i-1* ;
- une unité du niveau *i* est contenue (et ceci d'une manière exhaustive) dans l'unité du niveau supérieur immédiat *i+1* ;
-la structure hiérarchique est n-branchante.
-on définit le rapport de proéminence relative sur les nœuds sœurs de telle manière qu'un des nœuds porte l'étiquette « fort » et les autres nœuds – l'étiquette « faible ».
Nous considérons que les trois premiers principes sont conjugués sous la forme de l'hypothèse de l'étagement strict. À propos du principe de l'étiquetage des nœuds sœurs nous y reviendrons dans l'analyse de la hiérarchie des proéminences.

catégorie prosodique d'un niveau donné est exhaustivement segmentable en constituants du niveau inférieur adjacent, lesquels appartiennent tous à la même catégorie. Dans la structure syntaxique, un constituant est segmenté en constituants de types différents, y compris les constituants du même type (ainsi un syntagme nominal peut être composé d'un autre syntagme nominal suivi d'un syntagme prépositionnel : **NP → NP + PP**). Par conséquent, des différences émergent par rapport à la profondeur de la structure résultante : la profondeur de la hiérarchie prosodique est finie, tandis qu'elle est infinie, en théorie, pour la hiérarchie syntaxique.

Avec l'évolution de la réflexion sur la constituance prosodique, et surtout avec l'adoption de l'approche de la phonologie de laboratoire dans la recherche des indices acoustiques-marqueurs des constituants, nous sommes amenée à conclure que l'hypothèse de la hiérarchie stricte a plutôt le statut d'une forte tendance et non pas celui d'une loi universelle stricte dans son application aux langues du monde. Certains chercheurs défendent l'idée qu'un constituant d'un niveau donné peut dominer directement des constituants situés deux à trois niveaux plus bas dans la hiérarchie. Par exemple, dans l'étude de Selkirk (1995), on retrouve une analyse des éléments proclitiques qui va à l'encontre du principe de l'étagement strict. L'approche de Selkirk diverge en deux points de l'approche que nous qualifions de classique représentée par Nespor et Vogel (1986) : en premier lieu, le niveau du groupe clitique n'est pas inclus dans la hiérarchie des constituants prosodiques, puis on considère que seuls les mots lexicaux participent à la formation du niveau des mots prosodiques. De ces prémisses résulte la représentation de la figure 4.2 : un syntagme mineur domine un mot prosodique et les syllabes qui correspondent aux mots grammaticaux.

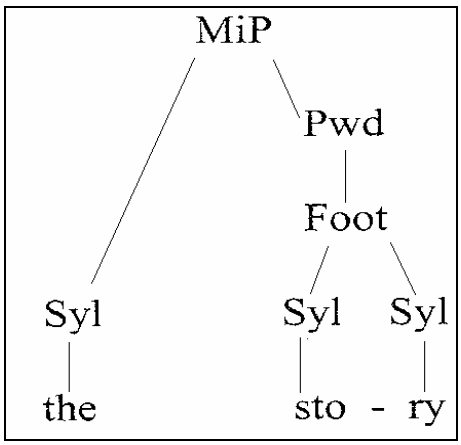


Figure 4.2 : Analyse des clitiques de Selkirk, transgressant le principe de la hiérarchie stricte

On retrouve également des structures récursives dans l'étude de Ladd (1986), structures proposées pour analyser la structuration prosodique des énoncés avec des constructions vocatives[117]. L'approche traditionnelle consiste à dire que les vocatifs forment toujours un syntagme intonatif à part ; or, les données mélodiques, à la base de la formation du constituant de syntagme intonatif, démentent souvent une telle analyse. Dans l'approche de Ladd, bien qu'écartée ensuite par ce dernier, il existe deux types d'unités : le syntagme majeur déterminé par des indices pausaux et un groupe tonal défini comme un domaine

[117] Cf. les traitements alternatifs de Gussenhoven (1990) et de Di Cristo et Hirst (1996).

de contour mélodique cohérent. Habituellement, c'est le syntagme majeur qui domine le groupe tonal dans la hiérarchie des constituants ; or il est admis qu'un groupe tonal domine un syntagme majeur quand les structurations temporelle et mélodique sont en désaccord : l'auteur analyse l'exemple des vocatifs. Ce raisonnement en adéquation avec les données observables transgresse le principe de la hiérarchie stricte.

La phonologie prosodique de Nespor et Vogel (1986) intègre le mécanisme de la restructuration des domaines : une fois les constituants prosodiques construits sur la base des principes sous-jacents, ils peuvent subir des modifications postérieures suite à l'application des contraintes phonologiques, par exemple rythmiques. On obtient donc des formes phonologiques de surface qui peuvent contredire l'hypothèse de la hiérarchie stricte. Nous reviendrons sur la présentation de la théorie de la restructuration des domaines après avoir présenté les arguments en faveur de chaque niveau de la hiérarchie prosodique proposés par différentes approches. Chaque constituant de la hiérarchie prosodique est obtenu sur la base de diverses sources d'information phonologiques comme non-phonologiques. Ces principes déterminent la mise en relation entre les constituants prosodiques et les constituants d'autres modules de la grammaire. Nous entreprenons cette démarche de présentation de chaque niveau, car, à notre connaissance, il n'existe aucune description phonologique du système prosodique de la langue russe, prenant en considération la constituance prosodique.

4.5. LES CONSTITUANTS PROSODIQUES AU-DESSOUS DU MOT

4.5.1. Thèses de la théorie générale

Comme nous l'avons déjà précisé, certaines hiérarchies prosodiques n'incluent pas les unités métriques qui se situent au-dessous du mot dans la hiérarchie des constituants prosodiques (Inkleas & Zec 1995). D'autre part, on retrouve cinq candidats potentiels au rôle des constituants prosodiques inférieurs au mot : une more, un segment, des constituants sub-syllabiques, une syllabe et un pied ; généralement, seuls deux ou trois de ces constituants sont inclus dans la hiérarchie prosodique.

LA MORE, en tant qu'unité de la hiérarchie prosodique, n'est pas universellement reconnue (elle est absente de la version de Nespor et Vogel (1986)). Fox (2000) adhère également à la position d'exclusion de la more de la hiérarchie des constituants prosodiques. Pour lui, l'inclusion de la more mène à la confusion des deux constituances prosodiques : en effet, l'auteur distingue la hiérarchie primaire qui regroupe les constituants eux-mêmes et la hiérarchie secondaire, fondée sur les propriétés des constituants. Ainsi, ce sont les caractéristiques quantitatives de la syllabe qui sont exprimées en termes du nombre de mores, c'est donc une propriété de la syllabe et non pas le constituant à part entière. Le concept de more est plus pertinent pour des langues telles que le japonais (*mora-counting languages* selon la typologie rythmique), et moins évident en anglais. Traditionnellement, on analyse en mores les phénomènes de quantité dans le système phonologique (par exemple, la distinction de trois degrés de longueur vocalique en estonien, cf. Lehiste 1960). Or, comme l'affirme Hayes (1989), pour plusieurs langues, dont l'anglais fait partie (il s'agit de

« *quantitatively sensitive languages* »), la localisation de l'accent est sensible à la structure interne de la syllabe. On retrouve d'autres arguments sur la réalité de la more dans les études expérimentales sur l'alignement des accents mélodiques. Dans l'étude de Pierrehumbert et Beckman (1988), il est démontré qu'en japonais, le nombre de mores dans le syntagme influence l'alignement de la pente de F0 dans le cadre du syntagme accentuel. Récemment, pour l'italien de Pise, la notion de more a été introduite dans l'analyse des différences d'alignement entre deux accents mélodiques (Prieto *et al.* 2005).

La more ne fait donc pas partie de toutes les hiérarchies des constituants prosodiques, pourtant ce concept se révèle pertinent pour l'analyse de différents phénomènes phonologiques comme phonétiques.

On exclut également LE SEGMENT en tant qu'élément terminal de la hiérarchie prosodique : dans cette démarche, l'argumentation est développée à partir de l'analyse de deux types de phénomènes qui corrompent les principes fondateurs de la constituance prosodique. Il s'agit des phénomènes d'ambisyllabicité et d'harmonie vocalique : dans les deux cas, c'est le principe de l'étagement strict qui se trouve transgressé.

On n'inclut pas non plus dans la hiérarchie prosodique LES CONSTITUANTS SUB-SYLLABIQUES[118], considérant que leur définition ne s'appuie pas sur les principes propres aux constituants prosodiques.

Chez la plupart des phonologues, on situe au niveau le plus bas de la hiérarchie prosodique LA SYLLABE. Après avoir été absente des analyses de la phonologie générative, la syllabe retrouve sa place dans les analyses linguistiques. Or, Shattuck-Hufnagel (1992) soulève la question de savoir si les phénomènes décrits en termes de syllabes pourraient recevoir une analyse adéquate en termes d'autres constituants prosodiques, comme le pied.

En adoptant la syllabe comme unité du niveau le plus bas dans la hiérarchie des constituants prosodiques, une réflexion est nécessaire sur les principes de syllabation universels : on retrouve une discussion générale portant sur cette problématique dans les études de Pulgram (1970), Kahn (1976), Clements et Keyser (1983), Blevins (1995), entre autres. Les syllabes résultantes ne représentent pas des groupements aléatoires de sons, mais des patrons phonémiques récurrents et réguliers. On note pourtant que dans le cadre de la phonologie prosodique, il importe de mettre en relation tous les niveaux de la structuration prosodique et le principe de la hiérarchie stricte. Dans cette perspective, on est amené à définir le domaine de la syllabation. Nespor et Vogel (1986) postulent que le domaine de la syllabation correspond toujours à un mot phonologique : la syllabation ne s'applique jamais à travers les frontières des mots, même dans le cas des mots composés, et la distinction est faite dans leur approche entre les deux

[118] En phonologie métrique, on considère de manière consensuelle que la syllabe est divisée en unités intermédiaires d'*attaque* et de *rime*, la rime étant elle-même divisée en *noyau* et *coda*. Il s'agit d'un gabarit (*template*) de syllabe dans lequel l'attaque et la coda sont facultatives.

processus de syllabation et de resyllabation[119]. Nous reviendrons sur le mécanisme de la restructuration de Nespor et Vogel dans la section 4.7.

Enfin, les arguments en faveur du PIED MÉTRIQUE comme constituant prosodique proviennent du domaine de l'accentuation. On définit le pied métrique comme le groupement d'une syllabe proéminente et de syllabes non-proéminentes sous le même nœud dans la hiérarchie prosodique. Plusieurs contraintes supplémentaires peuvent être imposées par la langue : une langue n'autorise que des pieds binaires, une autre admet des pieds illimités ; on ne retrouve dans certaines langues que des pieds ternaires (une construction très marquée qui exhibe une distribution limitée) ; le concept des pieds dégénérés (réduits à une syllabe) peut également poser un problème. Une typologie des langues a été proposée à partir de la définition de deux types majeurs des pieds : on distingue les langues à pieds trochaïques et les langues à pieds ïambiques (cf. Hirst et Di Cristo 1998, Di Cristo 2004).

Certains chercheurs insistent sur la différenciation entre un pied phonétique et un pied phonologique : la différence se traduit dans le fait qu'un pied phonétique regroupe une syllabe accentuée et toutes les syllabes inaccentuées jusqu'à la prochaine syllabe accentuée ; le nombre des syllabes inaccentuées est donc illimité. En revanche, le pied phonologique ne contient que deux à trois syllabes inaccentuées (Selkirk 1984, Hayes 1995).

Dans l'algorithme de la construction de la hiérarchie prosodique de Nespor et Vogel (1986), la restructuration est possible au niveau du pied comme au niveau de la syllabe : or, lors de la génération initiale de la structure prosodique, la formation des pieds est limitée par le domaine du mot prosodique. La tendance générale dans l'analyse des constituants au-dessous du mot est donc de préserver le mot comme unité prosodique :

> *« It is never the case that the syllables of a single foot belong to different phonological words. »* (Nespor et Vogel, 1986:109)[120]

Il est important de prendre en compte que le terme de pied a deux interprétations dans la littérature phonologique. Plus fréquemment en phonologie, on fait référence au pied interne du mot (*Within-Word foot*, ou *Rythmic stress foot*, Kiparsky 1979, Halle et Vergnaud 1987). Or, la vision la plus répandue du pied métrique remonte aux travaux d'Abercrombie (1965, 1973) : dans sa conception, le pied peut transgresser les frontières de mots et englober les fragments des mots adjacents (on parle alors du pied inter-mot, *Cross-Word-Boundary Foot*), cf. (2) :

---

[119] On note, auparavant, que Nespor et Vogel font référence au style précis de parole qui a conditionné leurs argumentations théoriques (p.64)

[120] Un modèle similaire de la structuration rythmique pour l'anglais britannique, respectant les frontières de mots, a été proposé par Jassem (1951, 1952, repris par Hirst 1998) dans l'objectif de prédire l'organisation temporelle au niveau syllabique. On reconnaît ainsi deux types d'unités qui, ensemble, forment l'unité rythmique totale de Jassem : a) l'unité rythmique étroite (*narrow rhythmic unit*) qui englobe la syllabe accentuée et toutes les syllabes inaccentuées jusqu'à la fin du mot ; b) l'anacrouse (*anacrusis*), qui comprend toutes les syllabes inaccentuées du début du mot, y compris les clitiques.
Comme il a été démontré dans une étude expérimentale à partir du corpus de l'anglais britannique (Bouzon 2004), le modèle de Jassem se révèle plus pertinent dans la prédiction de l'organisation temporelle et rythmique.

(2) |Know then thy- | -self, pre- | -sume not | God to | scan |

(exemple emprunté à Abercrombie 1973:11)

Hormis des arguments rythmiques en faveur du pied en tant que constituant prosodique, Nespor et Vogel (1986) citent des règles phonologiques qui ont le pied comme domaine d'application (par exemple, l'aspiration des consonnes plosives en anglais) ; d'autres arguments proviennent des contraintes phonotactiques et de la définition de la rime dans la poésie classique.

Dans la hiérarchie prosodique de l'école aixoise de l'intonation, une unité analogue au pied métrique est définie au niveau tonal : dans le modèle développé (Hirst et Di Cristo 1984), L'UNITÉ TONALE est une unité minimale de l'organisation tonale, assurant la synchronisation entre les gabarits tonals et la ligne segmentale ; cette unité a été également appliquée à l'analyse de l'anglais (Hirst 1983). Dans la représentation phonologique profonde, on attribue à cette unité le gabarit tonal $_{L}[\ ]_{H}$ en français et $_{H}[\ ]_{L}$ en anglais, ce qui permet d'y fonder une typologie des systèmes prosodiques. À notre avis, il est souhaitable de tester la pertinence de cette unité pour l'analyse de l'organisation mélodique du russe par une étude expérimentale, compte tenu des résultats de l'étude de Svetozarova (1973) sur l'association systématique des petites perturbations mélodiques avec l'accent lexical.

4.5.2. La constituance prosodique du russe au-dessous du mot

Dans cette section, nous passons en revue tous les constituants prosodiques trouvés dans la littérature phonologique pour déterminer leur pertinence dans l'analyse de la hiérarchie prosodique propre à la langue russe.

LA MORE : dans la section précédente, nous avons présenté les arguments en faveur de l'adoption de la more dans les analyses des langues autres que les langues à rythme moraïque : ces arguments sont liés à l'accentuation et à l'alignement des différents accents mélodiques. Comme cela a été démontré dans les études phonétiques de l'accentuation en russe, les différences entre les syllabes accentuées et les syllabes inaccentuées sont à la fois quantitatives (les syllabes inaccentuées sont plus courtes) et qualitatives (les syllabes inaccentuées contiennent une voyelle réduite). Or, ces effets quantitatifs dans la réalisation ne sont pas liés aux différences de quantité phonologique sous-jacentes : dans cette perspective, on ne peut pas considérer le russe comme étant une langue sensible à la quantité syllabique dans l'attribution de l'accent (cf. l'anglais dans l'analyse de Hayes (1989). Par ailleurs, dans le modèle combinatoire de l'intonation proposé par Kodzassov (1996), le concept de more revient pour qualifier l'alignement des tons complexes avec les noyaux vocaliques : les voyelles avec lesquelles un mouvement ascendant-descendant de f0 est associé, sont appelés bi-moraïques. Ce concept n'a pourtant pas reçu une distribution large dans la littérature phonologique sur le sujet. Remarquons que le système ToBI pour l'anglais s'en passe pour rendre compte des différences d'alignement entre différentes configurations mélodiques. Dans l'étude de Prieto *et al.* (2005), le concept de more est introduit dans l'analyse de l'italien de Pise pour rendre compte des détails d'alignement d'un accent

mélodique avec la syllabe métriquement forte. À la base de ces arguments ainsi qu'en nous fondant sur notre intuition en tant que locutrice native, nous allons exclure la more de la hiérarchie des constituants prosodiques du russe, au moins à ce stade de l'étude : nous considérons qu'une analyse empirique ainsi qu'une argumentation théorique plus développée sont nécessaires.

LA SYLLABE est une unité linguistique universelle (Fudge 1969) malgré l'existence dans la littérature d'un nombre important de théories de la syllabe[121]. Le point le plus controversé est celui de la nature de la syllabe : est-ce une unité phonétique qui caractérise les propriétés acoustiques et articulatoires particulières des sons ou bien est-ce une unité phonologique, une construction abstraite qui caractérise les combinaisons des sons admises dans une langue ? Grammont (1933) propose de distinguer la syllabe phonétique et la syllabe phonologique. Cette distinction est systématiquement maintenue dans la tradition des études prosodiques russes. La syllabe phonologique désigne alors les gabarits, les combinaisons autorisées de types de consonnes et de voyelles. Ces gabarits sont définis à partir d'une analyse combinatoire où les distributions des phonèmes consonantiques et vocaliques de la langue sont étudiées au sein d'une unité linguistique dotée du sens (morphème ou mot).

Toute syllabe peut être décrite en fonction de forme et de frontières : la problématique de la détermination des frontières d'une syllabe en russe est exploitée dans les travaux de E. Kurilowitch (1948). Cependant, Bondarko (1981:50) analyse des cas de syllabation discutables, à savoir la détermination des frontières d'une syllabe interne d'un mot dérivé[122]. Le modèle optimal de la syllabe phonologique en russe est FOSVSFO[123]. Les principes au fondement de la définition de la syllabe en tant qu'unité phonologique sont jugés ambigus.

La théorie acoustique de la syllabe, ainsi que la théorie articulatoire, font toutes les deux référence à la syllabe phonétique qui participe à la description de l'activité langagière de l'homme. La structure idéale d'une syllabe phonétique est OV (occlusive + voyelle) car cette structure représente le mouvement le plus simple de fermeture-ouverture du conduit vocal. Les études expérimentales ainsi que les réflexions théoriques sur la syllabation (Avanessov 1954, Scherba 1963) constatent une forte tendance en russe à la formation de syllabes ouvertes. Pour les théoriciens, les syllabes fermées ne peuvent être formées que dans quelques situations exceptionnelles : tout premièrement, quand la première consonne dans le groupe intervocalique est une sonnante ; ensuite, pour Scherba, les voyelles accentuées ont tendance à attirer les consonnes immédiatement adjacentes dans le groupe consonantique, tandis que Avanessov autorise la formation des syllabes fermées afin de respecter le principe de la sonorité croissante au sein d'une syllabe. Les études expérimentales démentent ces postulats théoriques, en particulier à

[121] Leur objectif premier est de fournir des précisions sur la nature de cette unité linguistique : on distingue la théorie acoustique de la syllabe, son interprétation en tant qu'unité de production ou de perception, les définitions phonologiques, cf. Meynadier 2001 pour une présentation développée de la problématique.

[122] Cf. le couple « *страх - страшный* » (*peur - effrayant*) : le premier mot est monosyllabique, le gabarit syllabique étant FOSVF. Or, comment segmenter en syllabe le second mot ? Le gabarit syllabique serait-il FOSVF-SVS, FOSV-FSVS ou encore FOSVFS-VS ?

[123] F pour 'fricative', O pour 'occlusive', S pour 'sonnante', V pour 'voyelle'.

travers l'analyse des effets de la coarticulation : à l'intérieur de la syllabe CV on observe l'interaction maximale entre les éléments constitutifs, les phénomènes de coarticulation y sont plus forts que dans la séquence VC, où les articulations de la consonne et de la voyelle sont plus indépendantes. Notons également que les opinions divergent sur le traitement des frontières morphémiques lors du processus de la syllabation. Or, les études expérimentales ont démontré que les frontières morphémiques ainsi que les frontières des mots n'empêchent pas la syllabation. Pour le mot, il est considéré que l'inclusion d'un mot au sein d'une unité du niveau supérieur entraîne la destruction de l'unité phonétique du mot.

Ainsi en russe, on reconnaît l'existence de deux processus : la syllabation, définie comme formation des syllabes par l'agencement des segments selon les structures syllabiques acceptables dans la langue, ce processus faisant référence aux mots isolés, et la resyllabation, ce terme désignant la modification de la structure syllabique lors de l'intégration du mot dans un énoncé. Les deux processus en russe privilégient la formation des syllabes ouvertes.

LE PIED MÉTRIQUE : comme son nom l'indique, le pied revient dans la description de l'organisation métrique du matériau verbal. Dans la tradition russe, le mot est considéré comme l'unité minimale de la description rythmique[124]. Faisant l'analyse de l'organisation rythmique du russe, Tcheremissina (1989) signale que pendant une longue période, pour caractériser l'organisation rythmique, on utilisait l'unité de « *такт* » [125] (cf. Peshkovsky 1930, Tomson 1910). Selon l'auteur, l'inconvénient de cette unité réside dans son rapport ambigu avec le mot : ainsi, les frontières d'un « *такт* » (mesure) peuvent soit coïncider avec les frontières des mots, soit couper un mot en deux. De même, vu la rareté des accents secondaires en russe[126], le nombre de mesures est égal au nombre de mots. En même temps, étant emprunté à la tradition de l'analyse musicale, le « *такт* » en parole ne préserve pas l'une des caractéristiques principales de cette unité en musique : avoir une même durée / longueur. Or, comme nous le signalerons dans l'analyse du concept du mot dans le discours continu[127], les frontières entre les mots s'effacent au sein des unités intonatives des niveaux supérieurs ; comme le remarque Nikolaeva (1977:261) :

> « *слова как бы растворяются во фразово-интонационных единицах, подчиняясь им.* »
>
> *les mots, se retrouvant dans la position de subordination, semblent se dissoudre dans le contexte des unités intonatives.*

En langue russe, la réduction vocalique est plus forte après l'accent lexical[128]. Or, dans le discours continu, la différenciation entre les syllabes pré-accentuées et post-

[124] Pour Tcheremissina (1989), le mot participe à l'organisation rythmique de manière indirecte, par son inclusion dans le syntagme qui est une unité principale de l'organisation rythmique et intonative.
[125] Ce terme a été emprunté à la musique : l'équivalent français est mesure.
[126] On ne le retrouve que pour quelques mots composés et certains préfixes (super).
[127] Remarquons que traditionnellement les propriétés phonologiques des mots sont analysées lors de leur prononciation isolée.
[128] Cf. la formule de Potebnia (1973) : si on attribue à la syllabe accentuée le coefficient 3, alors la structure rythmique du mot de quatre syllabes est décrite par la séquence des coefficients « 1 2 3 1 ».

accentuées est neutralisée : il semble donc qu'une unité indépendante soit nécessaire pour la description de l'organisation métrique et temporelle en russe.

Dans les études phonologiques, une partie des arguments sur la constituance prosodique proviennent des analyses de textes poétiques (cf. Nespor & Vogel, 1986, chapitre 10, Hayes 1989). Nous considérons le fait que la poésie classique en langue russe est une poésie métrique, système dans lequel une ligne poétique résulte de l'enchaînement des pieds minimaux d'une structure déterminée[129] comme argument en faveur de l'adoption de ce niveau de constituance dans la hiérarchie prosodique. Pour les analyses postérieures, il reste à déterminer si le pied en russe est de type ïambique ou trochaïque : dans les études de Gasparov (1984), il est démontré que c'est le ïambe qui est le mètre préféré de la poésie classique. Or, il nous semble qu'une étude expérimentale est nécessaire. On note que cette analyse complique son application à la langue russe : traditionnellement, on considère qu'un seul accent par mot est possible, et que les cas d'accentuation secondaire sont très rares.

4.6. LES CONSTITUANTS PROSODIQUES AU-DESSUS DU MOT

4.6.1. Présentation générale

Les constituants prosodiques au-dessus du mot sont définis par des règles de correspondance, qui font référence aux constituants d'autres composants de la grammaire.

La considération des constituants prosodiques supérieurs au mot nous ramène à la question de l'interface « prosodie – syntaxe » et à sa formalisation en phonologie prosodique : ces constituants sont construits à partir des informations morpho-syntaxiques et sémantiques. Pourtant, une fois générée la hiérarchie prosodique, le composant phonologique n'a plus accès aux informations morphosyntaxiques contenues dans la représentation de surface. Seule la hiérarchie prosodique devient accessible aux règles proprement phonologiques : cette structure sert donc de médiateur entre les modules syntaxique et phonologique.

LE MOT PHONOLOGIQUE est le constituant le plus petit dont la définition est fondée sur les notions non-phonologiques ; c'est un constituant dont le rôle est d'assurer la conjonction entre la hiérarchie prosodique et la structure morphosyntaxique : on considère qu'il n'existe pas de mots phonologiques plus grands que les éléments terminaux d'un arbre syntaxique (Nespor et Vogel 1986 :142).

Or, comme le mentionne Pepercamp (1999), le niveau du mot prosodique/phonologique est souvent cité dans les études sur la constituance prosodique pour illustrer le non-isomorphisme entre le mot morphologique et le mot phonologique. De la sorte, certaines règles phonologiques font une référence explicite à des morphèmes grammaticaux (les affixes de dérivation ou les mots composés). Un des points de discorde entre les différentes interprétations de l'interface « phonologie ~ syntaxe » concerne la

[129] On en reconnaît cinq types : trochée, ïambe, anapeste, amphibraque, dactyle.

problématique de la taille du mot prosodique : le mot prosodique peut-il être plus petit que le mot orthographique ? De manière générale, le mot prosodique est défini par rapport aux processus phonotactiques et aux règles phonologiques segmentales dont il est le domaine d'application ainsi que par rapport à son rôle de domaine d'accentuation lexicale.

Pourtant, dans l'analyse de la prosodie issue de la tradition britannique, le mot n'a pas le statut d'unité phonétique et phonologique (on distingue les unités tonales, rythmiques et intonatives) : le mot étant une unité purement grammaticale, il peut être pertinent pour la définition des unités prosodiques, mais non pas directement intégré dans la hiérarchie. Ainsi Pike (1945) remarque-t-il que les frontières des unités rythmiques correspondent aux frontières des mots).

Outre la question sur le statut de différents types de morphèmes et de mots composés, les désaccords entre les différentes approches portent sur le traitement des mots lexicaux et des mots grammaticaux monosyllabiques (clitiques). Par suite dans les études de Hayes et Nespor & Vogel, les mots grammaticaux forment des mots prosodiques à eux seuls, mais forment ensuite un seul groupe clitique avec les mots lexicaux adjacents. Ce traitement est critiqué dans les études de Selkirk (1995) : dans sa conception de la constituance prosodique, seuls les mots lexicaux forment les mots prosodiques. Les mots grammaticaux (clitiques) reçoivent un traitement différent selon leur position, avant (proclitiques) ou après (enclitiques) le mot lexical principal : les proclitiques sont attachés au mot phonologique sous le nœud d'un syntagme mineur ; on est amené à introduire la catégorie du mot prosodique supérieur pour l'intégration structurelle des enclitiques (figure 4.3). La forme de l'arbre prosodique se retrouve modifiée une fois que le mot grammatical surgit dans sa forme pleine.

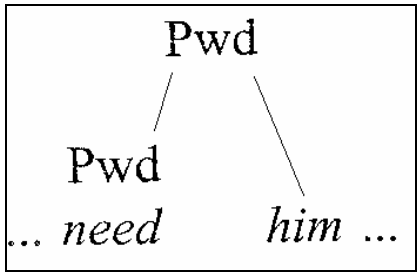


Figure 4.3 : Traitement des enclitiques chez Selkirk (1995)

Dans l'approche de Nespor et Vogel (*op.cit.*), UN GROUPE CLITIQUE contient donc un mot lexical et les mots grammaticaux monosyllabiques adjacents (clitiques). C'est le constituant dans la hiérarchie prosodique qui a pour fonction de rendre compte du sentiment intuitif d'un lien plus étroit entre le mot grammatical et le mot lexical[130], son hôte ; on y retrouve également l'indication du potentiel à la réduction, propre aux mots clitiques. Dans la typologie des constituants de Selkirk (1995) le constituant équivalent est celui du syntagme mineur, auquel les mots clitiques sont rattachés directement sans former de mot prosodique. Hayes (1989) cite deux règles phonologiques dont le domaine d'application paraît être un groupe clitique, à savoir : la palatalisation des sibilantes et la suppression du « v » avant un segment [-syllabique] en anglais.

[130] Cf. l'analyse du système de transcription ToBI (son historique) ainsi que l'analyse des études expérimentales sur la perception de la force d'une frontière.

Une des critiques que l'on peut adresser aux réflexions exposées – surtout par rapport aux règles phonologiques – consiste à dire que les arguments en faveur d'une approche ou d'une autre sont moins convaincants étant donné le caractère spéculatif des études et la présence des exemples se résumant au seul texte écrit : dans ce contexte, une étude des réalisations attestées nous paraît être plus pertinente.

L'UNITÉ RYTHMIQUE de Hirst et Di Cristo (1993) correspond à la fois aux structures analogues au mot prosodique de la hiérarchie classique et au groupe clitique. Cette unité introduite pour rendre compte de l'accentuation finale en français est de ce fait spécifique à cette langue. Ce n'est pas une unité définie dans le domaine du tonal : l'indice acoustique pertinent pour l'accentuation finale en français est l'allongement final.

LE SYNTAGME ACCENTUEL est le constituant proposé dans l'étude de Pierrehumbert et Beckman (1988) pour rendre compte de l'accentuation en japonais. Comme l'illustre le schéma de la figure 4.1, il est parfois difficile de trouver les correspondances exactes entre les unités proposées par les différents auteurs. Ainsi, pour le niveau du syntagme accentuel de Pierrehumbert et Beckman, on se demande à quel niveau dans la hiérarchie générale correspond ce constituant : correspond-il au niveau du groupe clitique (défini comme le mot porteur d'un accent mélodique et les mots inaccentués associés) ou au niveau du syntagme phonologique ? Le terme de syntagme accentuel a été repris par Jun et Fougeron (2000) pour analyser la distribution des proéminences mélodiques en français ; or, d'autres études définissent le domaine d'accentuation en français comme étant le syntagme phonologique (cf. Delais-Roussarie 2000).

Nous constatons que les deux niveaux supérieurs du phrasé prosodique sont reconnus dans la majorité des travaux ayant pour objet d'étude le phrasé prosodique. Les termes et les critères à la base desquels la définition est faite dépendent des approches. Plus précisément, la phonologie prosodique classique fait la différence entre un syntagme phonologique et un syntagme intonatif, la théorie ancrée dans l'organisation tonale (Beckman & Pierrehumbert 1986) distingue les syntagmes intermédiaires des syntagmes intonatifs.

LE SYNTAGME PHONOLOGIQUE est le niveau de constituance prosodique le plus associé à la structure syntaxique. Tous les algorithmes de la formation des syntagmes phonologiques nécessitent en tant que forme d'entrée un arbre syntaxique construit sur la base de la syntaxe X-barre (Jackendoff 1977). Par rapport aux informations syntaxiques, on fait référence à des catégories lexicales (N, V, A, Adv), à des catégories phrasales (NP, VP, PP *etc.*), et aux rapports de dominance entre ces constituants.

Par conséquent, le syntagme phonologique est le constituant dans la hiérarchie prosodique le plus contraint par la syntaxe. On note également la distinction entre un syntagme majeur et un syntagme mineur chez Selkirk (1986) : pour le syntagme phonologique majeur, les frontières sont alignées avec les frontières de la projection syntaxique maximale. L'auteur en donne un argument phonologique, à savoir, le fait qu'à la frontière du syntagme majeur, les mots grammaticaux ont la forme phonétique pleine. Les frontières des syntagmes mineurs sont alignées avec les bornes des têtes des projections

maximales. D'autres théories proposent un seul type de syntagmes phonologiques qui peuvent être les équivalents des syntagmes majeurs ou mineurs de Selkirk.

Dans leur travail sur l'interface « syntaxe – prosodie », Zec et Inkelas (1995) affirment que le niveau du syntagme phonologique est un des mieux étudiés à travers les langues du monde, avec des résultats similaires obtenus, dus aux caractéristiques morphosyntaxiques sous-jacentes à la définition de ce constituant. Or, les différences théoriques sont assez importantes dans la définition de ce niveau structurel, à savoir les différences sur la nature des informations syntaxiques nécessaires pour les règles de mise en correspondance (*mapping rules*). Trois types d'approches sont attestés :

- Mise en correspondance fondée sur des relations entre les termes (approche relationnelle) : on prend en considération les rapports « tête-complément » (cette approche est adoptée par Nespor et Vogel (1986) et Hayes (1989)). Les possibilités prédites par cette approche sont : a) la tête et le complément forment obligatoirement des syntagmes phonologiques différents ; b) la tête et le complément forment l'unique syntagme phonologique (propre au français ; prouvé à travers la règle de la liaison)[131] ; c) les compléments branchants en anglais posent un problème pour cette approche, car les deux cas de figure en attestent : si le complément est non-branchant, il forme le syntagme phonologique avec sa tête ; s'il branche, il forme le syntagme phonologique à part.

- Mise en correspondance fondée sur les frontières des constituants : cette approche est proposée dans l'objectif de réduire la sensibilité de la théorie aux informations syntaxiques nécessaires (à savoir le rang des phrases dans la structure). Les algorithmes imposent la localisation de la frontière du constituant prosodique de manière à ce qu'elle corresponde à la frontière désignée (gauche ou droite) de constituant syntaxique du rang défini. Donc, dans la présente approche, les informations syntaxiques requises dans la génération de la structure prosodique sont de nature plus généralisée, néanmoins les problèmes avec les langues à compléments branchants ne sont pas toujours résolus (cette approche est défendue dans les études d'E. Selkirk).

- Mise en correspondance prenant en compte le caractère branchant du constituant (proposé par Zec et Inkelas 1990) : dans leur algorithme, les nœuds-sœurs dans la structure syntaxique forment un seul syntagme phonologique avec les rapports d'adjacence immédiate privilégiés. Un des avantages de cette approche se manifeste dans l'analyse des rapports entre un syntagme nominal en position de sujet et un syntagme verbal (prédicat) : les sujets (SN) non branchants entrent dans le même syntagme que le verbe. Par contre, à l'inverse de l'approche relationnelle, cet algorithme n'est pas applicable aux langues où les compléments ne sont jamais inclus dans le même syntagme phonologique avec leur tête.

Par ailleurs, dès les premières études en interface « prosodie ~ syntaxe », on fait référence à la Règle Rythmique (ou la règle d'évitement de *clashes* accentuels, cf.

[131] Les possibilités mentionnées sous a) et b) sont propres aux langues avec la tête à l'initiale ainsi qu'avec la tête à la finale.

l'étude de Post 2000) dans la motivation phonologique du niveau du syntagme phonologique :

Règle rythmique (d'eurythmie)

$* \rightarrow Ø / [_{\gamma} * _ *]$

[des ENfant SAges] $_{\gamma}$

Cette règle permet, voire impose, la suppression d'un accent dans le contexte du *clash* accentuel à l'intérieur d'un syntagme phonologique. On retrouve les données empiriques, illustrant l'application de cette règle dans les études de Horne (1990) et Grabe *et al.* (1994), Shattuck-Hufnagel (2000) entre autres.

Dans la hiérarchie des constituants fondée sur des propriétés mélodiques des énoncés (Pierrehumbert et Beckman 1986), c'est LE SYNTAGME INTERMÉDIAIRE qui correspond au constituant traditionnel du syntagme phonologique (cf. les études de mise en correspondance entre les constituants de Selkirk et Tateishi (1991), Jun (1993)). Nous ne connaissons pas de travaux sur l'anglais qui aient comparé explicitement la structuration d'un texte en syntagmes phonologiques définis sur la base de critères syntaxiques avec l'annotation de l'organisation tonale en termes de syntagmes intermédiaires. Nous avons discuté de la définition tonale du syntagme intermédiaire dans le cadre du modèle de J. Pierrehumbert (cf. *supra*). Précisons tout de même qu'à l'heure actuelle les deux conceptions, la tonale et de la phonologie des domaines, se voient confondues dans la mise en pratique lors de l'annotation des corpus.

Dans la version classique de la hiérarchie des constituants, LE SYNTAGME INTONATIF est le seul constituant défini par des propriétés acoustiques (donc, physiques) : c'est le domaine d'un contour mélodique perceptivement cohérent et unifié. Le deuxième critère qui intervient dans la définition est d'ordre temporel : on associe aux frontières des syntagmes intonatifs un allongement final important et on affirme également que ses frontières représentent une localisation privilégiée des pauses silencieuses. Or, on note aussi que la formation des syntagmes intonatifs est sujette aux influences d'ordre sémantique et pragmatique, difficilement formalisables, ou d'ordre de la performance (style de parole (cf. citation), la longueur du constituant, le débit d'élocution) :

> *« In addition to the basic syntactic factors that play a role in the formation of intonational phrases, there are also semantic factors related to prominence and performance factors such as rate of speech and style that may affect the number of intonational contours contained in an utterance. »* (Nespor et Vogel 1986:186)

Comme on le voit dans la citation ci-dessus, pour Nespor et Vogel, c'est la syntaxe qui déterminera au premier abord la formation des syntagmes phonologiques : ceci dans la mesure où le principe fondamental de la hiérarchie prosodique est l'hypothèse d'une hiérarchie stricte. Une deuxième source d'influences syntaxiques sur ce niveau du

phrasé prosodique consiste dans le fait que certaines constructions syntaxiques (tels les vocatifs, les expressions disloquées)[132] forment un syntagme intonatif séparé.

Dans l'approche de Nespor et Vogel (*op. cit.*) non seulement on définit les constituants en rapport avec d'autres modules de la grammaire mais on établit également une relation de proéminence relative entre les constituants du niveau immédiatement inférieur, regroupés sous le même nœud (entre les syllabes d'un pied ainsi qu'entre les syntagmes phonologiques qui forment un syntagme intonatif). Dans cette démarche, ce sont les principes propres à la langue et à son organisation rythmique qui déterminent au niveau du pied si l'élément le plus proéminent se situe à gauche ou à droite. Pourtant, dans la définition de la proéminence relative des syntagmes phonologiques qui constituent un syntagme intonatif, la position d'élément fort ne peut être déterminée pour la catégorie entière. Comme le principe général, on admet que la position d'un élément fort (tête du syntagme intonatif) puisse varier et que ce sont les facteurs sémantico-pragmatiques qui déterminent l'élément fort (focus, distinction entre les informations nouvelles / déjà connues). Ainsi, plusieurs réalisations sont possibles pour un syntagme étant donné la variation de la position de l'élément fort : dans cette perspective, on est amené à réfléchir sur la notion de réalisation « marquée » par opposition à une réalisation « neutre ».

Dans l'approche de l'école britannique, on retrouve la décomposition d'un syntagme intonatif en termes de composants structurels de tête, avant-tête, nucleus et queue. Ces éléments ne peuvent être mis en relation ni avec la hiérarchie prosodique classique de Nespor et Vogel (*op. cit.*), ni avec la version de Beckman et Pierrehumbert (*op. cit.*), fondée sur les propriétés des contours mélodiques.

Le niveau de L'ÉNONCÉ PHONOLOGIQUE n'est pas présent dans toutes les hiérarchies de constituants prosodiques. Une définition à base tonale a été proposée : on considère qu'un énoncé se termine dans le registre infra-bas. Pour argumenter la pertinence de ce constituant, on cite, entre autres, la règle du « r battu » en anglais américain dont le domaine d'application semble être l'énoncé prosodique et le syntagme intonatif (cf. son analyse chez Nespor et Vogel (*op. cit.*) et Selkirk). C'est pour ce niveau qu'on retrouve encore une fois et avec plus d'acuité le problème de l'élaboration de la constituance prosodique à partir du seul texte écrit ; or le phrasé prosodique est sujet à de multiples influences ainsi qu'à la variabilité de la performance et on a besoin d'une annotation riche du texte pour aboutir à une prédiction adéquate.

[132]Astruc (2003) réfléchit aux différents termes que l'on retrouve dans la littérature pour désigner ces structures syntaxiques et insiste sur le lien entre le terme choisi et les présupposés théoriques sur le phrasé. Ainsi, les *tags* sont plus typiques pour l'approche autosegmentale-métrique (travaux de Ladd, Pierrehumbert etc. et nécessitent donc une explication de trois composants : distribution des proéminences, du phrasé et du patron mélodique), les termes *contour plat*, *parenthèse* dénotent l'approche orientée phonologie (on identifie le contour distinctif) ; citons ensuite des approches considérant que c'est la fonction syntaxique de ces éléments qui détermine leur prosodie (le terme utilisé étant *constituants externes*). Syntaxiquement, ce sont des éléments adjoints au niveau supérieur de la hiérarchie syntaxique (codé S dans la représentation syntaxique).

4.6.2. Structuration prosodique au-delà de l'énoncé

Cahn (1999) présente la hiérarchie des constituants phonologiques comme faisant partie d'une hiérarchie plus globale ayant la conversation comme unité maximale (cf. figure 4.4).

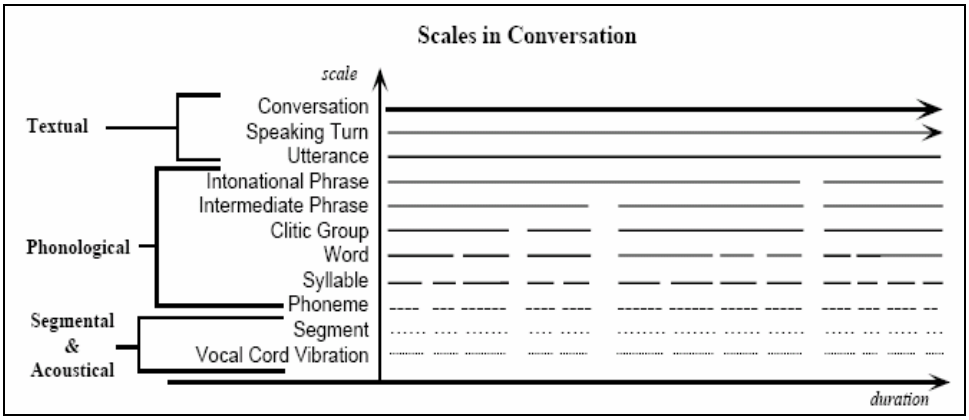


Figure 4.4 : Hiérarchie prosodique selon Cahn (1999)

Cette hiérarchie englobe les unités allant de la vibration des cordes vocales à la conversation. On note que les constituants textuels ainsi que les constituants acoustico-segmentaux ont une portée plus universelle que les constituants phonologiques et prosodiques. La conversation (discours) est l'unité maximale ; elle peut être composée d'un ou de plusieurs segments discursifs qui traduisent les différents sujets abordés. Dans le discours dialogique, chaque segment de discours comporte un certain nombre de tours de parole (en monologue, on est en présence d'un seul tour de parole bien que très développé). L'intervention de chaque locuteur peut être composée de plusieurs énoncés, énoncés ayant leur propre structuration en termes de constituants prosodiques.

Dans la perspective conversationnelle, les constituants au-dessous de l'énoncé peuvent recevoir leur propre interprétation pragmatique (en termes de structure informationnelle), car en évaluant le degré de jonction entre les mots, on détermine les mots qui vont ensemble du point de vue pragmatique et sémantique.

Au moment de sa conceptualisation, la phonologie prosodique s'est largement inspirée des études en phonologie métrique. De ce fait, dans ces premiers travaux on observe une dérivation de la structure terminale fondée sur le principe du branchement binaire et sur les relations de proéminence distinguant les nœuds/positions forts et faibles. Au niveau des unités intonatives ces rapports binaires se traduisent par des variations de registre actualisant les rapports au sein de la structuration informationnelle du message.

Il est important de préciser que les rapports entre les composants sémantique et phonologique de la grammaire ont été moins étudiés et formalisés. Dans la grammaire générative linéaire ainsi que dans la théorie du Minimalisme (Chomsky 1995), la communication n'est pas autorisée entre le calcul de la forme phonético-phonologique de l'énoncé et le calcul de sa forme logique. Or, il paraît nécessaire d'interpréter l'interaction entre deux modules interprétatifs de la grammaire, module phonologique et module sémantique. Un vaste corpus d'études sur la gestion des interactions dans une conversation démontre que certaines clés phonétiques y interviennent de manière

pertinente (Local (2003) pour une analyse dans le cadre de la théorie du détail phonétique fin) : ce qui prouve la nécessité d'intégrer des processus allant du niveau segmental au niveau discursif dans le même schéma d'annotation. D'autre part, même si on rend disponibles au module phonologique des informations sur la structure informationnelle/ logique, le fait de limiter le domaine d'analyse à un seul énoncé empêche une appréhension adéquate de certains phénomènes phonologiques segmentaux : de sorte qu'en anglais britannique l'actualisation du « r de liaison » entre deux énoncés dépend du degré du lien logique entre eux ; or cette information est absente de la structure logique de l'énoncé générée dans le composant sémantique, puisque l'empan d'analyse est limité au domaine de l'énoncé syntaxique.

4.6.3. Constituance prosodique du russe : les niveaux supérieurs au mot

LE MOT : au niveau du mot on maintient la différence faite pour les unités de la syllabe et du pied entre un mot phonologique (très fréquemment ce terme fait référence au mot grammatical, constituant de la structure morpho-syntaxique) et un mot phonétique, ce terme désignant une construction nommée groupe clitique dans la terminologie de la phonologie prosodique de Nespor et Vogel (1986) (cf. Bondarko 1998 : 217). Kassevitch (2003) définit le mot phonétique comme un segment de texte isolé du reste du discours à la base des principes phonétiques ou phonologiques. Les principes dont il est question dans la définition contribuent à l'accomplissement de deux tâches : déterminer le nombre de mots phonétiques dans l'énoncé et permettre d'établir leurs frontières.

Ainsi, en russe, le nombre des mots phonétiques dans un énoncé est en relation biunivoque avec le nombre d'accents lexicaux. Remarquons que les mots clitiques ont un accent dans le dictionnaire ; or, quand on affirme que ces clitiques n'ont pas leur propre accent lexical, on prend en compte les relations syntagmatiques.

La question des frontières des mots paraît être plus problématique. Dans l'exemple (2) la frontière entre les mots est un potentiel, dont l'actualisation par le locuteur est probable, mais non obligatoire (par exemple, par un coup de glotte pour l'attaque vocalique).

(3) Собаку видел // собак увидел /sʌˈbakuˈvʼidʼil /

voyait un chien // a vu des chiens

Pour accéder à la compétence linguistique des locuteurs, Kassevitch *et al.* (1990) ont mené une expérience en perception : des stimuli désémantisés mais avec des propriétés prosodiques authentiques ont été présentés aux auditeurs dont la tâche était de proposer un mot/ une phrase ayant une même structure prosodique que le stimulus entendu. Les résultats ont été évalués selon qu'une structure avec un ou deux accents lexicaux était proposée. Certains résultats obtenus contredisent toutes les théories phonologiques et phonétiques sur le statut du mot en tant qu'unité phonétique : comme dans le cas de la structure initiale de l'expression « *ловя птиц* » (*attrapant les oiseaux*), la réponse donnée étant « *коллектив* » (*collectif*). Un autre exemple surprenant est la paire de réponses. Pour Kassevitch (2003) ce sont les rapports de synonymie qui sont remis en

cause pour les oppositions terminologiques « accentué / inaccentuée », « fort / faible », « lourd / léger », « réduit / non-réduit ». Si l'on admet qu'une syllabe inaccentuée n'est pas toujours réduite et qu'une syllabe accentuée n'est pas toujours forte, le critère de la réduction vocalique peut également entrer en jeu pour déterminer le nombre de mots dans l'énoncé (cf. Bolinger 1981).

En tout cas, le mot phonétique cesse d'être une unité non-ambiguë : d'une part, les critères de la définition des mots phonétiques sont hétérogènes ; d'autre part, certains types des mots phonétiques ne sont que potentiellement segmentables, et leurs frontières ne sont pas déterminées du point de vue phonétique. Nous avons également discuté des cas d'intégration complète du mot dans la structure du niveau supérieur (syntagme) lors de la présentation du phénomène de la resyllabation en russe.

Dans le contexte des considérations présentées, nous tenons à insister sur deux points principaux : dans un premier temps, la distinction faite entre un mot phonologique et un mot phonétique dans la tradition russe correspond à celle qui est faite entre un mot et un groupe clitique dans la terminologie de la phonologie prosodique. Cependant, les expériences démontrent que le paramètre du nombre d'accents lexicaux ne permet pas toujours d'obtenir le nombre de mots lexicaux dans l'énoncé, la réduction vocalique étant un critère plus fiable. Nous insistons également sur le fait de la restructuration observée au niveau du mot en russe. Du point de vue prosodique, le mot phonétique peut correspondre à une structure regroupant deux mots lexicaux, donc, une structure n'ayant pas vraiment le statut de groupe clitique ni de mot phonologique : ce point nécessite pourtant d'être développé.

LE SYNTAGME : dans les études discutées au chapitre III, la seule unité à laquelle les descriptions de l'intonation russe font référence est le syntagme. Svetozarova (2000) parle pourtant d'une hiérarchie d'unités prosodiques sur l'axe syntagmatique : groupe rythmique – syntagme – phrase (analogue à l'énoncé dans l'écrit) – unité supraphrastique (paragraphe) – texte. À notre avis, cette hiérarchie démontre une association forte, voire une dépendance de l'organisation du texte écrit (quoique des analogies puissent être établies entre les paragraphes à l'écrit et les paratons (cf. Brown et Yule 1983), paragraphes à l'oral, de l'organisation informationnelle du message).

La définition fondamentale de l'unité de syntagme remonte aux travaux de Scherba (1963). Dans son approche, le syntagme est défini tout premièrement comme une unité de la structure sémantique de l'énoncé : la segmentation en syntagmes est tout premièrement une segmentation sémantique, possédant tout de même des caractéristiques formelles. La Grammaire russe (1980) définit les syntagmes comme *« интонационно нечленимые смысловые отрезки речи*[133]*»*. Parmi les caractéristiques formelles du syntagme on indique la présence d'un accent particulier, accent de syntagme (accent nucléaire), sur le dernier mot lexical du syntagme ; un autre aspect formel à prendre en compte est le fait que le contour mélodique a pour domaine d'application un syntagme. Or, l'aspect formel est parfois omis des considérations sur la structuration prosodique de l'énoncé, les aspects sémantico-syntaxiques s'appropriant le rôle central. Dans cette perspective, les travaux expérimentaux adoptent la démarche

[133] *Fragments de parole dotés de sens, non-segmentables du point de vue de l'intonation.*

suivante : on procède tout d'abord à la segmentation du texte en syntagmes en se fondant sur des relations sémantiques ; la segmentation attestée est ensuite confirmée dans les études expérimentales.

Il est considéré également qu'un texte écrit possède des marqueurs de sa structure intonative sous la forme de signes de ponctuation ; de plus, on procède fréquemment à l'établissement d'un parallèle entre les concepts de syntagme intonatif et de syntagme syntaxique (unité de la structure syntaxique). Dans cette perspective, il nous paraît nécessaire de procéder à une étude expérimentale approfondie afin de vérifier si le syntagme intonatif de la tradition russe est semblable au syntagme phonologique ou au syntagme intonatif de la phonologie prosodique classique.

Remarquons que la structure hiérarchique comportant un segment, une syllabe, un mot et un syntagme, est prise en compte dans la réflexion sur la forme qu'un allophone a dans le contexte discursif, chaque niveau hiérarchique exerçant ces propres influences. D'autre part, dans les études prosodiques le niveau du syntagme domine : la distinction est faite entre les réalisations des syntagmes terminaux et non-terminaux, mais ce contraste n'est pas interprété comme étant dû aux rapports hiérarchiques entre les constituants du syntagme et de la phrase.

4.7. LE MÉCANISME DE LA RESTRUCTURATION

En discutant de la constituance prosodique, Nespor et Vogel (1986) sont amenées à reconnaître l'existence d'un mécanisme de restructuration pour plusieurs niveaux dans leur hiérarchie : c'est un processus post-lexical, responsable de la forme de surface, que l'on récupère dans la prononciation effective d'un énoncé et qui peut être différente de celle prédite par la théorie.

La restructuration a pour fonction de formaliser les structures optionnelles qui contredisent la théorie décrite auparavant, mais que l'on observe dans les données réelles. Ainsi, dans l'approche de Nespor et Vogel, la formation d'un constituant prosodique inclut deux types de règles (1986:218) : les règles de base pour la formation du constituant (on insiste sur son caractère universel) et les règles de la restructuration (qui rendent les structures probables mais optionnelles, marginales, quoique présentes en production) ; après l'application des règles de restructuration, on obtient des structures post-lexicales dérivées.

Cette réflexion sur les formes dérivées rejoint, à notre avis, la distinction retrouvée dans la littérature phonologique entre un constituant dans la représentation phonologique et son homologue dans la représentation phonétique (cf. les distinctions systématiques entre une syllabe phonétique et une syllabe phonologique, un mot phonétique et un mot phonologique *etc.*).

Tout d'abord, on parle de la restructuration quand, dans la parole continue, les syllabes (et par conséquent, les pieds métriques) transgressent les frontières des mots. Ce phénomène se retrouve dans les études empiriques, or, du point de vue théorique, il contredit le principe de l'étagement strict. D'autre part, comme la phonologie

prosodique cherche à formaliser les interfaces entre les différents modules de la grammaire, dans la détermination des constituants au-dessous du mot, la question se pose de la relation entre la structure syntaxique et la structure syllabique de l'énoncé, comme la syllabation doit avoir lieu à l'intérieur d'un domaine syntaxique donné. Dans le souci de préserver la cohérence théorique de leur modèle, Nespor et Vogel proposent d'établir un domaine de syllabation égal au mot phonologique et de reléguer toute restructuration au domaine des processus post-lexicaux.

Pourtant, non seulement les données empiriques mais également certains phénomènes phonologiques remettent en cause le mot en tant que domaine universel de syllabation. Selkirk (1978) analyse la syllabation en anglais, laquelle possède deux types d'affixes : les affixes neutres qui constituent un domaine de la syllabation à part, et, les affixes non-neutres qui appartiennent toujours au même domaine de syllabation avec les racines. De ce fait, la règle de syllabation fait référence aux informations non-phonologiques telles que le type de suffixe.

Donc, la différence doit être faite entre la syllabation (processus limité par des frontières du mot phonologique) et la resyllabation (processus post-lexical). Ce dernier phénomène est dépendant de la langue et son domaine peut varier d'une langue à l'autre. Par exemple, on le considère comme étant équivalent à un syntagme phonologique en français tandis qu'en espagnol on parle d'un syntagme intonatif ou même d'un énoncé entier[134].

Comme nous l'avons déjà mentionné, les processus de restructuration posent un problème au principe fondateur du formalisme de la phonologie prosodique, à savoir l'hypothèse de l'étagement strict. La deuxième clause du principe de la hiérarchie stricte contraint la construction d'arbres prosodiques bien formés : on postule qu'une chaîne de sons est segmentée en domaines non-chevauchants sur tous les niveaux de la hiérarchie. Pour remédier à cette transgression, Pepercamp (1999) inclut dans la dérivation de la structure prosodique l'étape de formation des mots prosodiques post-lexicaux. L'auteur analyse l'exemple de l'italien : un syntagme « *bar aperto* » (*bar ouvert*) est syllabé comme [ba.ra.per.to]. Ainsi on distingue les deux structures prosodiques : lexicale et post-lexicale (cf. exemple 3, emprunté à Peperkamp, 1999) :

(3) *structure lexicale*
$(\text{bar})_{\text{pw}}\ (\text{aperto})_{\text{pw}}$
structure post lexicale
(a) $*(\text{ba.ra.per.to})_{\text{pw}}$
(b) $(\text{ba})_{\text{pw}}\,(\text{ra.per.to})_{\text{pw}}$

L'analyse de (3) démontre que deux structures post-lexicales sont possibles. Or, selon l'auteur, on ne retient pas la première car on ne possède pas suffisamment d'arguments pour supposer que la resyllabation induit la restructuration de deux mots prosodiques en un avec un seul accent lexical. La restructuration post-lexicale résulte en une formation

[134] Les traitements alternatifs proposés consistent à considérer les éléments à la frontière entre deux mots comme extramétriques ou ambisyllabiques.

de mots prosodiques post-lexicaux qui diffèrent au minimum[135] des mots lexicaux à partir desquels les structures post-lexicales sont obtenues. Ainsi évite-t-on la génération des structures mal-formées et on ne transgresse pas l'hypothèse de la hiérarchie stricte.

Les problèmes des processus post-lexicaux, postérieurs à la génération de la hiérarchie prosodique dont nous venons de discuter concernent les niveaux inférieurs au mot et résultent de la nécessité de déterminer leur empan en termes de constituants syntaxiques. Or, on retrouve également la restructuration aux niveaux supérieurs de la hiérarchie prosodique, à savoir, pour les syntagmes phonologiques et les syntagmes intonatifs. Cependant, au niveau des syntagmes intonatifs, la restructuration dépend de facteurs extra-linguistiques comme le débit d'élocution, le style de parole et l'emphase, mais également de contraintes rythmiques (cf. Gee et Grosjean 1983).

4.8. LE PHRASÉ PROSODIQUE ET LA THÉORIE DE L'OPTIMALITÉ

Nous avons démontré précédemment que plusieurs phénomènes observés dans les données réelles remettent en cause l'hypothèse de l'étagement strict dans sa formulation initiale, assez rigide. L'hypothèse de l'étagement strict reflète plutôt une forte tendance, et non pas une loi universelle. Une analyse alternative, plus souple, a été introduite par Selkirk (1995) dans le cadre de la théorie de l'optimalité (Prince et Smolensky, 1993). La théorie de l'optimalité propose un formalisme explicatif présenté sous la forme de contraintes universelles de bonne formation ; la propriété importante des contraintes consiste en ce qu'elles sont hiérarchisées (c'est la hiérarchie des contraintes qui reflète la spécificité de la langue) et transgressables. Ainsi un ensemble des formes de sortie est évalué contre l'ensemble des contraintes universelles et c'est le candidat qui transgresse les contraintes des niveaux inférieurs dans la hiérarchie qui est retenu. Les contraintes formulées pour expliquer le phrasé prosodique portent sur la quantité du matériau verbal qui peut être produit sans interruption et sur les relations mutuelles entre la segmentation prosodique, la structure syntaxique et la structure informationnelle.

Dans la perspective de la théorie de l'optimalité (Selkirk, 1995), l'hypothèse de la linéarité stricte est présentée sous la forme de quatre contraintes sur la domination prosodique :

> ***Layerdness*** No C_i dominates a C_j, $j > i$ (par exemple, aucun groupe clitique ne domine un syntagme phonologique).
>
> ***Headedness*** Any C_i must dominate a C_{i-1} (un syntagme intonatif domine obligatoirement un syntagme phonologique ; contrainte applicable à tout niveau de la hiérarchie prosodique, à l'exception du niveau syllabique)
>
> ***Exhaustivity*** No C_i immediately dominates a C_j, $j < i-1$ (aucun syntagme phonologique ne domine un mot prosodique)
>
> ***Nonrecursivity*** No C_i dominates a C_j, $j=i$ (aucun pied ne domine un pied)

[135] Cf. la contrainte de fidélité entre la forme d'entrée et de sortie postulée dans la théorie de l'optimalité.

Cette version décomposée rend compte, entre autres, des analyses récursives de Ladd (1986) (seule la contrainte de la non-recursivité est transgressée) ainsi que du traitement des mots clitiques dans le cadre d'un syntagme mineur de Selkirk (cette analyse transgresse la contrainte d'exhaustivité). Il n'est attesté dans la littérature phonologique aucune transgression des contraintes d'étagement et de *headedness*.

Peperkamp (1999) considère que les contraintes proposées par Selkirk (1995b) ne reflètent que la première clause du principe de la hiérarchie stricte, à savoir, la condition de la domination immédiate dans la hiérarchie prosodique. Elle propose de rajouter la contrainte de *Proper Nesting* qui permettra de formaliser la segmentation exhaustive et non chevauchante d'une séquence :

> **Proper nesting**
>
> Align(C_i, L/R; C_j, L/R)
>
> where C_i and C_j are categories of the prosodic hierarchy and C_i immediately dominates C_j.

À part l'ensemble de contraintes traduisant en termes de la théorie de l'optimalité le principe de la hiérarchie stricte, d'autres contraintes ont été proposées par différents auteurs afin de rendre compte du phrasé prosodique à partir des études spéculatives ainsi qu'expérimentales. Notre étude expérimentale comportant une analyse succincte des résultats en termes de la théorie de l'optimalité, nous réservons la présentation de ces contraintes au chapitre IX.

Nous avons présenté les prémisses théoriques de la phonologie prosodique ainsi que quelques aspects des discussions sur leurs validité et applications pour l'explication de différents phénomènes phonologiques. Nous présentons ci-dessous quelques critiques importantes de certains axiomes de la phonologie prosodique ; nous revenons ensuite aux études phonétiques et psycholinguistiques des constituants prosodiques. À la fin de ce chapitre, nous résumons nos réflexions sur le modèle de la constituance prosodique que nous appliquerons à nos données dans la suite de notre étude.

4.9. LES CRITIQUES DE LA PHONOLOGIE PROSODIQUE

Dans les théories de la phonologie prosodique que nous avons présentées dans les paragraphes précédents les positions prises se résument à trois principes :

- La hiérarchie prosodique assure le lien entre les modules syntaxique et phonologique ;

- Une démarche unidirectionnelle est adoptée, dans laquelle c'est le composant syntaxique qui élabore la forme de l'entrée du composant phonologique, les informations phonologiques n'étant pas accessibles en syntaxe :

  « As has been shown time and again, significant phonological properties of sentence follow from their syntactic properties, and not vice versa. A generative

grammar mirrors this fact by casting the set of rules comprising the phonological component as interpretative of syntax. » (Selkirk 1980);

- Les constituants prosodiques sont simultanément des primitives phonologiques.

Un premier groupe de critiques portent sur la remise en cause du lien médiatisé entre la syntaxe et la phonologie : en formalisant l'interface « prosodie – syntaxe », la phonologie prosodique propose des règles de formation des constituants prosodiques à partir des informations syntaxiques générales dans le souci d'aboutir à d'importantes généralisations. Cependant, dans l'étude de Marsi (2001) sur un corpus en hollandais, les trois algorithmes d'insertion des frontières prosodiques sont examinés et comparés avec une annotation du corpus de travail[136] par des experts phonologues. Leur évaluation révèle les classes des structures syntaxiques pour lesquelles les prédictions sont systématiquement faussées. L'analyse des erreurs de prédiction amène l'auteur à l'élaboration d'un nouvel algorithme du phrasé prosodique : par rapport aux algorithmes précédents, les erreurs ne sont plus systématiques mais individuelles. On note pourtant que l'algorithme est le résultat de la modélisation manuelle à partir du corpus de travail ; c'est un algorithme dérivé ancré syntaxe, sensible à certains types de constructions syntaxiques, mais également aux paramètres prosodiques tels que la distribution de proéminences mélodiques et la force d'une frontière. Chaque règle de cet algorithme décrit l'insertion de la frontière dans un constituant syntaxique précis[137] : cet algorithme est ainsi fortement dépendant de la syntaxe et utilise des descriptions syntaxiques beaucoup plus complexes. Donc, l'argumentation des partisans du lien direct entre la syntaxe et la prosodie se fonde sur le besoin d'avoir accès en phonologie à des informations syntaxiques plus pointues pour la construction des domaines prosodiques.

Dans la même lignée se situe la position défendue par Delais-Roussarie (2000) s'inspirant du travail de Selkirk (1984). Selon l'hypothèse défendue par l'auteur, la structure prosodique n'est plus considérée comme primaire dans l'actualisation des phénomènes suprasegmentaux, mais comme le résultat de l'interface complexe « prosodie ~ syntaxe, sémantique, discours », ce dernier s'articulant autour

- d'une structure intonative dérivée avec la prise en compte des informations syntaxiques et sémantiques telles que les relations prédicat – argument et la structure informationnelle de l'énoncé ;

- d'une grille métrique bâtie sur la base des principes phonologiques et des règles/contraintes de leur appariement.

Globalement, cette approche remet en cause la vision modulaire de la grammaire : la structure prosodique se construit dans le domaine temporel de manière dynamique, avec une intervention alternée de deux types de contraintes : morphosyntaxiques et métriques. Cette perspective a pour résultat le renversement du cadre de l'analyse de

[136] Il s'agit du corpus de domaine limité : descriptions des fleurs.
[137] Le cas d'apposition, de proposition (phrase), pour les syntagmes verbal et nominal, et pour les structures de coordination.

l'accentuation en français chez Delais-Roussarie (2000) : c'est la distribution des accents qui est première par rapport à la structure prosodique.

Pour notre démarche, à savoir, pouvoir récupérer la structure prosodique de l'énoncé automatiquement, à partir du signal acoustique, nous adhérons à la vision du phrasé prosodique comme étant le résultat final de l'application des contraintes linguistiques et discursives à plusieurs niveaux. Cela évite de passer par l'étape de la restructuration post-lexicale des constituants qui suscite, comme nous l'avons démontré précédemment, beaucoup de débats. Nous évitons, également, les restrictions stylistiques préalables à toute discussion de phrasé prosodique en phonologie.

L'une des critiques majeures que l'on puisse émettre à l'encontre de la phonologie prosodique, porte sur les restrictions stylistiques que les différents auteurs sont amenés à faire afin de justifier leurs analyses. Par exemple, dans le travail de Selkirk (2000:241), la réflexion suivante précède l'analyse du phrasé prosodique en anglais britannique :

> *« Let us consider the circumstance in which such utterances, with accents on all the content words and no VP-internal focus, would be appropriate […] The utterances lack a focus or topic structure, and all information is treated as new. These are 'neutral' pronunciations of such utterance. »*

Les restrictions stylistiques sont également fréquentes dans le travail de Nespor et Vogel, et constituent, en fait, les prémisses indispensables à leur analyse. De même que les phénomènes étudiés expérimentalement dans cette étude relèvent, eux aussi, d'un style neutre de parole (ni trop rapide, ni trop lent). Les auteurs insistent sur ce fait, car dans la parole rapide on peut observer des phénomènes prosodiques et des séquences de segments qui ne sont pas autrement autorisés par la théorie. Une autre limitation que l'on retrouve consiste à insister sur un style conversationnel neutre, avec une prononciation non emphatique et non contrastive. Nous considérons que les restrictions théoriques citées nuisent au statut de la phonologie prosodique en tant qu'outil formel puissant d'analyse et de prédiction des phénomènes prosodiques.

L'unilatéralité des rapports entre la phonologie et la syntaxe est également remise en cause par certains chercheurs. Dans l'approche configurationnelle (cf. ci-dessus, Zec & Inkelas, 1995), la syntaxe a besoin d'avoir accès aux informations phonologiques (sur la question du branchement). Également, les contraintes de l'eurythmie et de la taille minimale du syntagme phonologique imposent le choix de l'ordre des mots.

Le principe général de la phonologique prosodique fait que la pertinence d'un constituant phonologique dans une langue est assuré par l'existence des règles dont il est le domaine d'application. Néanmoins, Delais-Roussarie (1996) indique que deux règles indépendantes valident rarement la pertinence d'un même constituant dans une langue. La réflexion de la phonologie prosodique est, donc, de nature circulaire, ce qui affaiblit le pouvoir prédictif de cette approche.

Les constituants prosodiques ont une réalité cognitive dans la mesure où nous percevons à l'oral la segmentation du continuum sonore en blocs ('*chunks*', terme de Bolinger 1989). Cette structuration est signalée à l'oral par un ensemble d'indices acoustiques

(cf. le chapitre suivant). Dans la plupart des cas (70 % dans l'étude de Delais-Roussarie 1996 sur un corpus de laboratoire, 60 % dans son étude de 2000 sur un corpus « authentique » de parole lue) les frontières des constituants prosodiques récupérées à partir de l'analyse instrumentale et acoustique du signal coïncident avec les frontières des constituants morpho-syntaxiques. Pour atteindre un seuil plus élevé de prédiction correcte, on a besoin de règles de réajustement, qui, elles, sont très fréquemment spécifiques à la langue ou même à un phénomène précis. L'application des algorithmes de la prédiction des caractéristiques phonologiques et prosodiques, fondés sur la syntaxe, serait donc faussée dans 30 à 40 % des cas. Pour préciser, nous pensons que cela nuirait beaucoup à la perception dans les systèmes TTS[138], étant donné l'absence de redondance et de variabilité dans les structures générées.

Des critiques importantes ont été également présentées dans le travail de Fox (2002). Dans sa conception de la prosodie, Fox reconnaît l'existence au niveau phonologique de « traits prosodiques » (« *prosodic features* ») tels que la quantité, l'accent, le ton et l'intonation. On admet que ces traits ne sont pas complètement indépendants les uns des autres et participent conjointement à la construction d'une structure qu'ils forment et dont ils dépendent en même temps. La perception de la structure prosodique est assurée par le caractère stratifié des corrélats acoustiques des traits prosodiques.

Après avoir reconnu l'existence de la structure prosodique et ces propriétés inhérentes, l'auteur s'interroge sur les unités/constituants à intégrer dans cette structure. Il en propose la définition suivante : c'est un ensemble fini d'unités, organisées en une structure hiérarchique, qui servent des domaines d'actualisation pour les traits prosodiques (2000 :347). La hiérarchie prosodique adoptée par Fox comprend 4 niveaux, à savoir : la syllabe, le pied, l'unité intonative, l'énoncé. Par rapport à d'autres unités que l'on trouve dans les études phonologiques, certaines peuvent être reconnues avec des limites[139], d'autres sont écartées soit comme étant dépendantes de la théorie adoptée (le super-pied de Selkirk (1978) est le résultat de l'adoption du principe du branchement binaire), soit comme définies sur la base des critères hétérogènes, autres que les critères prosodiques : les niveaux entre le pied et l'unité intonative sont donc écartés car ils correspondent plutôt à des catégories grammaticales que prosodiques.

4.10. RÉSUMÉ ET DISCUSSION

Dans une perspective d'analyse formelle de l'intonation, nous insistons sur la nécessité d'une prise en compte conjuguée des problématiques de l'organisation tonale et du phrasé prosodique. Dans la conclusion du chapitre précédent, dans lequel nous avons discuté des analyses de l'intonation par l'école russe, nous avons constaté que le phrasé prosodique n'a pas reçu un traitement rigoureux dans ces études ; notamment, la question de l'indépendance des structures syntaxique et prosodique n'a pas été

[138] Comme nous l'avons discuté dans le chapitre précédent, les algorithmes de la génération de l'intonation tiennent compte du phrasé prosodique.

[139] Par exemple, la more est plutôt la propriété de la syllabe, or, cette unité peut être pertinente pour les langues à rythme moraïque. Beckman et Pierrehumbert (1988) incluent le more dans leur description de la phonologie du japonais, car en japonais la more joue le rôle d'unité de temps.

soulevée. Nous avons par conséquent choisi de mener une réflexion théorique ainsi qu'un travail empirique sur la problématique du phrasé.

Nous avons dédié le présent chapitre à la présentation des postulats et des heuristiques analytiques de l'approche de la phonologie prosodique, courant théorique, dont l'objectif premier est de réfléchir sur l'articulation et l'actualisation à l'oral des interfaces complexes entre différents modules de la grammaire, et notamment de l'interface « prosodie – syntaxe ». La réflexion menée par des chercheurs de ce courant a résulté en une hiérarchie des constituants prosodiques. L'heuristique de la phonologie prosodique n'a jamais été appliquée à l'analyse du phrasé en russe. Pour cette raison, nous avons spéculé sur la pertinence de différents niveaux du phrasé pour rendre compte de l'organisation prosodique en russe, nous fondant sur des arguments puisés dans la littérature.

Outre la discussion des différents niveaux de la hiérarchie prosodique, nous avons débattu des principes fondateurs de la phonologie prosodique, à savoir l'hypothèse de l'étagement strict et le principe de la restructuration. Des données empiriques contredisant fréquemment le principe de la hiérarchie stricte, ce dernier a été reformulé et assoupli dans le cadre de la théorie de l'optimalité ; il a acquis une forme d'une hiérarchie des contraintes universelles transgressables.

Cependant, il est regrettable que de nombreux travaux en phonologie prosodique aient un caractère spéculatif, les exemples discutés se résumant au seul texte écrit ; d'autres chercheurs insistent sur l'application des formalismes développés à une parole neutre, non emphatique et non contrastive. Ces limitations affaiblissent le pouvoir explicatif de la phonologie prosodique en tant qu'outil formel puissant d'analyse et de prédiction des phénomènes prosodiques. Or, nous considérons que la théorie approfondie du phrasé prosodique d'une part modélisera des interactions complexes entre la structure prosodique et d'autres modules de la grammaire, et, d'autre part, intégrera des effets eurythmiques, extralinguistiques et intralinguistiques, induisant une variabilité inhérente du phrasé.

Le cadre explicatif de la phonologie prosodique a été cependant exploité dans les études phonétiques dont les objectifs sont de vérifier la pertinence de différents niveaux de la hiérarchie prosodique pour différentes langues mais aussi de mettre en évidence des indices acoustiques qui contribuent à l'identification des constituants à l'oral. Nous proposons dans le chapitre suivant un panorama succinct de ces études.

CHAPITRE V

LA MATERIALITÉ DE LA CONSTITUANCE PROSODIQUE

Comme nous avons essayé de le démontrer dans le chapitre précédent, la motivation première de la hiérarchie prosodique réside dans l'élaboration d'un cadre de travail en phonologie fondée sur l'hypothèse de l'universalité de la constituance prosodique. Il existe trois types d'arguments soutenant la vision hiérarchique structurelle de la prosodie : les arguments théoriques phonologiques, les données acoustiques quantitatives et les données comportementalistes, issues des expériences psycholinguistiques en traitement perceptif de l'oral. Lors de la discussion portant sur la phonologie du phrasé prosodique, nous avons présenté les propriétés formelles à la base de l'identification et de la définition des constituants. Nous présenterons dans la suite de ce chapitre un résumé des études sur les manifestations phonétiques concrètes de la structure hiérarchique attestée. Par la suite, nous nous arrêterons sur quelques données issues des travaux psycholinguistiques.

5.1. LA PHONÉTIQUE DE LA CONSTITUANCE PROSODIQUE

Un constituant prosodique peut être abordé selon deux optiques complémentaires : d'une part, il s'agit de déterminer ses frontières, de rechercher les indices formels démarcatifs ; d'autre part, on s'intéresse à l'organisation et à la cohérence interne du constituant, aux indices qui assurent le regroupement de constituants du niveau adjacent inférieur autour d'un élément-tête. Par conséquent, il est possible de définir trois positions fortes en association avec tout constituant prosodique : sa tête et ses frontières gauche et droite.

Nous procédons, dans la suite de cette section, à une présentation des études menées dans le cadre de phonologie du laboratoire (Pierrehumbert *et al.* 2001, d'Imperio 2005) dans l'objectif de prouver empiriquement la pertinence de différents constituants prosodiques en production et en traitement perceptif de la parole. Une remarque doit pourtant être faite au préalable sur la multi-stratification des paramètres acoustiques associés aux événements prosodiques. Si l'on considère, à titre d'illustration, les événements linguistiques influençant la durée d'un segment dans la parole, on en

dénombre une multitude : structure linguistique et phonologique du message[140] (un segment est allongé à la fin du constituant, la question reste de savoir quel constituant dans la hiérarchie prosodique a contribué à l'allongement), association de la syllabe avec l'accent lexical et la proéminence mélodique (il peut être question de la syllabe à laquelle le segment appartient ou de la présence d'un accent mélodique à sa proximité[141]), les contraintes biométriques de la production de parole. Ainsi les indices acoustiques que l'on retrouve dans le signal de parole sont-ils potentiellement ambigus dans leurs relations à la structure prosodique : seule une partie de la variation des paramètres acoustiques reflète les effets de la structuration. Par conséquent, une précaution est nécessaire quant aux résultats communiqués dans les études où certains facteurs mentionnés n'ont pas été contrôlés. Dans cette perspective, un approfondissement des connaissances sur la localisation des effets induits par les différentes sources d'influence est nécessaire (cf. l'approche de Turk & White (1999) dans le cadre du modèle « domaine – locus », étude d'Astesano (2001) sur les corrélats acoustiques de l'accentuation primaire et secondaire en français). D'autre part, les données perceptives y sont également d'une importance majeure, dans la mesure où, en perception, on fait abstraction de tout ce qui n'est pas pertinent.

Nous commençons par une présentation succincte des études articulatoires traitant de la constituance prosodique pour aborder les études acoustiques dans la section 6.1.2. Notons, toutefois que cette présentation ne se veut pas exhaustive, étant donné la quantité impressionnante de travaux réalisés ayant trait à la problématique analysée.

5.1.1. Études articulatoires[142]

5.1.1.1. Constituance prosodique et renforcement articulatoire

L'ensemble des études articulatoires sur la corrélation et la covariation des articulations supraglottiques, des proéminences et des frontières prosodiques démontre que les deux événements linguistiques sont associés à une augmentation de la force articulatoire. Ainsi dans les résultats des travaux de Straka (1963), ce renforcement articulatoire se traduit par une activité musculaire (contraction des muscles) plus importante, associée à des positions syntagmatiques fortes (syllabe accentuée, début du constituant) dans la chaîne de parole. Cette vision est contemporaine à celle du travail en acoustique de Lindblom (1963) sur la réduction des voyelles en suédois.

Le travail de Lindblom (1963) aboutira au fil du temps à un modèle H&H (*Hypo&Hyper*, Lindblom 1990) de la production de parole. Les productions langagières de l'homme ont une clarté variable, cette variabilité étant de nature double ; elle peut être induite par les propriétés du système de la langue en question ; d'autre part, elle peut être externe à la langue, orientée situation de production de parole et dirigée par le souci d'être compris. Ainsi l'hypoarticuation est le résultat des stratégies d'économie

[140] Traditionnellement, on attribue à la constituance prosodique un rôle primordial dans la structuration temporelle de la parole (cf. Gussenhoven 1988, 1990 ; Gussenhoven et Rieteveld 1992).
[141] On atteste dans les études expérimentales l'allongement des durées segmentales devant les syllabes accentuées/proéminentes.
[142] Pour une analyse plus détaillée voir Meynadier (2003).

d'énergie ; d'autre part, la finalité du processus d'hyperarticulation est le renforcement des propriétés importantes, des distinctions catégorielles dans le signal de parole (dans le but de faciliter la perception et la compréhension du message). Cette stratégie est associée avec le souci de faciliter la tâche d'accès au lexique[143].

Les études articulatoires attentives aux phénomènes prosodiques, adoptant le modèle H&H, considèrent les syllabes accentuées comme étant les sites de cette hyperarticulation locale ; ce qui est reflété dans les caractéristiques spatio-temporelles des gestes des articulateurs.

5.1.1.2. Travaux de Beckman et de ses collègues (1992) : marquage articulatoire des têtes de constituants

D'autres chercheurs (notamment M. Beckman et collègues) ont nuancé cette idée générale, en se penchant sur une localisation plus précise du processus du renforcement articulatoire. Dans leur conception, la stratégie du locuteur est celle de la mise en valeur du noyau vocalique de la syllabe accentuée, associée à un événement prosodique : la syllabe accentuée est réalisée avec une ouverture maximale du conduit vocal, ce qui provoque une augmentation de l'énergie acoustique. Plusieurs études ont été réalisées dans les années 1990 sur les effets d'accentuation en association avec l'ouverture et fermeture du conduit vocal, en particulier, à travers les analyses des mouvements de la mâchoire inférieure[144]. Ces études portent sur les différents types de proéminence et analysent les réalisations des voyelles basses comme hautes. Il est démontré dans les travaux de Giot (1977) sur la voyelle /a/ en français, Rietvelt & van Lieshout (1995) sur les voyelles /i, a/ en néerlandais, Fletcher et al. (1994) sur les voyelles /a:, ʌ/ en anglais ou Harrington et al. (1995) sur la voyelle /a:/, qu'au moins pour certaines voyelles le mouvement d'ouverture de la mâchoire[145] est plus marqué, plus long et plus rapide en syllabes accentuées comparées avec les syllabes inaccentuées.

Les études présentées portent sur le marquage articulatoire des têtes de constituants prosodiques (syllabe, pied métrique, syntagme intermédiaire ou syntagme intonatif). Dans l'étude de Beckman et Edwards (1994), il est démontré que le contraste du niveau le plus bas de la hiérarchie des proéminences (entre syllabe réduite et syllabe accentuée) induit un contraste articulatoire plus stable que les niveaux plus hauts. Ce résultat fait écho aux propos défendus par Bolinger dans le travail intitulé « *Two types of vowels, two types of accent* » (1989). Dans les travaux de Beckman et collègues, les auteurs procèdent à une comparaison systématique des mouvements articulatoires dans l'espace spatio-temporel (allongements des gestes et de leurs phases, augmentation de l'énergie globale) pour des syllabes proéminentes (différents types d'accent) et pour les syllabes porteuses des tons de frontières (syllabes finales des syntagmes intonatifs). L'étude

[143] Nous présentons les travaux en psycholinguistique dans la section 6.2.

[144] Notons que l'influence de l'accent sur l'amplitude d'abaissement de la mandibule est notamment mise en évidence par Kozhevnikov et Chistovich (1965 :181) ; ces auteurs notent pour une réalisation du mot ['baba] : « *Lowering of the jaw is three to fivefold greater in accented syllables than unaccented ones* » .

[145] La référence aux mouvements mandibulaires ne représente qu'une facette des études réalisées : plus globalement, le marquage de proéminence touche le conduit supraglottique dans sa globalité, les mouvements des lèvres, de la langue et de la mandibule.

révèle que les effets et les mécanismes sont différents : ainsi l'allongement final est-il caractérisé par des mouvements ralentis, mais pas forcément par une amplitude augmentée des gestes articulatoires.

5.1.1.3. Marquage articulatoire des frontières

Nous présentons dans cette section les études en articulation qui se sont articulées autour de l'analyse des corrélats articulatoires des frontières de constituants. Dans la présentation des arguments phonologiques en faveur des constituants prosodiques, les phénomènes segmentaux et prosodiques sont essentiellement associés à la frontière finale du constituant[146]. Pourtant, à la base des données articulatoires, on inclut la position de la frontière initiale dans l'inventaire des positions syntagmatiques fortes des constituants : comme le remarque Keating (1995), cette position est un lieu privilégié du renforcement articulatoire.

Les deux positions dans la chaîne parlée, liées aux frontières des constituants, ont une fonction démarcative. Dans cette perspective, la position initiale du constituant est associée à un renforcement important des gestes articulatoires et le degré de ce renforcement est proportionnel au niveau du constituant dans la hiérarchie prosodique. En perspective, la question à résoudre est celle du caractère catégoriel ou graduel du renforcement articulatoire en relation avec la constituance prosodique. Il importe également d'éclaircir le lien entre le nombre de degrés du renforcement attesté et le nombre de niveaux dans la hiérarchie prosodique : étant donné la grande variabilité inter-locuteur, ce lien paraît être optionnel.

La position de la frontière initiale des constituants des niveaux supérieurs de la hiérarchie prosodique est caractérisée par un renforcement articulatoire relatif à la cavité laryngale (aspiration des consonnes, glottalisation des voyelles) ; l'effet adjoint est celui du temps de VOT augmenté observé à cette localisation[147].

Pour les attaques vocaliques des constituants prosodiques, Pierrehumbert et Talkin (1992) démontrent que la position à l'initiale du syntagme intonatif induit des réalisations plus fréquentes d'une glottalisation, et ceci également dans les cas où la frontière prosodique n'est pas associée à une pause.

Dans l'étude de Dilley *et al.* (1996), la tendance vers une glottalisation des segments vocaliques à l'initiale des syntagmes intonatifs majeur et intermédiaire (dans la typologie de Pierrehumbert et Beckman) est fortement présente, une fois comparée à la position médiane de syntagme ; cette recherche a été réalisée sur un corpus radiophonique d'un volume important.

D'autres études portent sur les segments nasals et les articulations linguopalatales (les études réalisées avec la méthodologie EPG) : de manière générale on constate que le

[146] À l'exception près du phénomène du *resetting* mélodique.
[147] VOT : *Voice Onset Time*. Il s'agit de la coordination différée entre le geste d'ouverture de l'occlusion linguale ou labiale et le geste de fermeture des cordes vocales.

contact linguopalatal et la durée acoustique des consonnes augmentent entre les positions interne du mot, initiale du mot et initiale d'un groupe intonatif (comme nous l'avons précisé au début de ce chapitre la relation entre les effets articulatoires et les niveaux de la hiérarchie des constituants n'est pas bi-univoque).

Outre le renforcement articulatoire, le deuxième phénomène articulatoire ayant pour rôle de signaler la frontière initiale d'un constituant est une réduction de la co-articulation à travers la frontière[148]. On constate la réduction de la coarticulation entre les éléments séparés par une frontière, et ceci en fonction de la force de la frontière. Comme une partie de ces études date des années 1960-1970, période de la dominance de la théorie générative, la réduction de la coarticulation y est analysée en fonction de la présence d'une frontière de constituant syntaxique (majeure ou mineure).

L'analyse présentée des études qui ont porté sur la question des corrélats articulatoires des frontières prosodiques (certes, très bref et incomplet[149]), nous permet de conclure qu'en plus des corrélats articulatoires de marquage de proéminence (tête d'un constituant prosodique), deux autres phénomènes articulatoires sont associés au signalement des frontières ; à savoir un renforcement articulatoire et une réduction de la coarticulation à travers les frontières. La grande variabilité inter-locuteur attestée peut suggérer le caractère facultatif du marquage articulatoire de la constituance prosodique : il reste à savoir dans quelle mesure ces variations articulatoires sont optionnelles, puisque relevant des stratégies compensatoires.

Cependant, si le constat du renforcement articulatoire spécifique des segments retrouvés dans les positions prosodiques fortes est consensuel et qu'on dispose d'une base de connaissances considérable sur le sujet, l'objectif de la recherche future est de mettre en évidence la sensibilité des phénomènes articulatoires aux influences de la hiérarchie prosodique. Il est attesté que les frontières des niveaux intermédiaires sont moins fréquemment associées à un marquage articulatoire par opposition aux contextes interne du mot et initial du syntagme intonatif. Les études d'Ellen Byrd et de ses collègues corroborent ce projet de recherche : en effet, Byrd et Saltzman (1998), dans une étude spatio-temporelle du geste d'aperture labiale, différencient deux à trois niveaux prosodiques. De même, dans une étude postérieure (Byrd, 2000), l'auteur adopte une hiérarchie prosodique plus abstraite que toutes celles dont nous avons discuté dans le chapitre précédent : cette hiérarchie comporte trois niveaux, faisant la différence entre les cas d'absence de frontière prosodique, une frontière mineure et une frontière majeure.

[148] Remarquons que les mêmes données sur la coarticulation transfrontalière servent de marqueur de la frontière terminale comme initiale d'un constituant.
[149] Nous renvoyons le lecteur intéressé au travail de Meynadier (2002)

5.1.2. Les études acoustiques

Dans les recherches appliquées, on est amené à constater que la prosodie ne peut pas être prédite à partir du seul texte : par exemple, en connaissant la position de l'accent lexical primaire pour chaque mot de l'énoncé, nous ne pouvons pas prédire la distribution des accents mélodiques (cf. Shattuck-Hufnagel 2000). En outre, la structure prosodique ne peut être obtenue directement à partir du signal, étant donné la multi-stratification des indices acoustiques associés avec des événements prosodiques. Dans l'ensemble des études sur les corrélats acoustiques de la structure prosodique, les chercheurs partent du principe que l'implémentation phonétique est guidée par une représentation phonologique que supervise la construction d'un réseau hiérarchique de constituants prosodiques (Di Cristo 2004).

Comme les indices articulatoires, les indices acoustiques contribuent à la saillance perceptive de deux types majeurs d'événements prosodiques, à savoir marquage des frontières et marquage des proéminences.

Les mesures acoustico-phonétiques ont pour objectif de mettre en évidence la nature hiérarchique de la constituance prosodique, et ceci en comparant les effets induits par deux niveaux adjacents dans la hiérarchie. Dans un deuxième temps, on cherche à démontrer expérimentalement que l'implémentation phonétique est guidée par une structure prosodique et non pas syntaxique, quoique ces deux constituances soient interliées. Ainsi, dans l'étude de Collier et 't Hart (1975), les auditeurs, ayant pour tâche de construire les énoncés dont la prosodie correspondait à celle qu'ils entendaient sous la forme d'une séquence de syllabes réitérées avec une frontière prosodique interne, ont montré une préférence pour les énoncés dans lesquels la frontière prosodique correspond à une frontière du constituant syntaxique. Les résultats de cette étude démontrent une relation forte entre les frontières prosodique et syntaxique en perception ainsi qu'en production[150]. Ce ne sont pourtant pas toutes les frontières des constituants syntaxiques qui retrouvent leur implémentation à l'aide des moyennes prosodiques. Ce constat pourrait être un argument fort en faveur de la structuration prosodique indépendante, mais la liberté laissée au locuteur en fait un phénomène non-systématique.

La seconde typologie des études repose sur le fait que les indices acoustiques contribuent à l'actualisation d'une double fonction de jonction *versus* disjonction : en effet, il est aussi important d'assurer au niveau acoustique la cohérence interne du constituant que de participer à la fonction délimitative et d'induire la perception d'une rupture.

Dans la suite de cette section, nous passons en revue un nombre d'études acoustiques en relation avec différents indices acoustiques pertinents en rapport avec la structuration du message en termes de constituants. Notons au préalable que plusieurs de ces études se fondent sur des données perceptives.

[150] Il s'agit de l'étude en production dans la deuxième partie du travail de Collier et 't Hart (1975) que nous ne présentons pas ici.

La présente section se divise désormais en 3 paragraphes qui traitent trois phénomènes :

a) pauses

b) phénomènes d'allongement

c) distribution des proéminences mélodiques.

5.1.2.1. Pauses

- *Pause, un indice préférentiel du phrasé prosodique*

Comme cela est souligné dans plusieurs études expérimentales, la pause est l'indice acoustique le plus stable dans son association avec la perception de disjonction/rupture prosodique. Ainsi dans une étude expérimentale, Strangert (1992) cherche à démontrer jusqu'à quel degré les auditeurs sont capables de déterminer à partir des clés prosodiques si la frontière qui suit est une frontière d'énoncé, de clause, de syntagme (syntaxique). Des stimuli ayant un même contenu sont utilisés avec différentes modifications des paramètres acoustiques. Sous la condition expérimentale où les informations acoustiques sont mélangées et certaines entrent en contradiction avec les autres, c'est la pause qui est l'indice le plus fort ; pourtant tous les autres indices ont également été pris en considération par les auditeurs.

L'interaction entre les différents paramètres acoustiques est démontrée dans l'étude de Duez (1985). Dans cette étude en perception les auditeurs avaient pour tâche d'indiquer les localisations des pauses ; les réponses ont été classées selon 3 types de configurations : les pauses détectées correctement, les pauses non détectées, les pauses perçues sans présence d'un silence physique dans le signal. Quant aux pauses prédites correctement, leur perception était corrélée avec la durée de la pause elle-même, mais également avec l'intensité, la durée et la valeur de F0 pour la voyelle de la syllabe précédant la pause.

Blaauw (1994), dans une étude sur la discrimination entre parole spontanée et parole lue se penche également sur la problématique de la réalisation des frontières prosodiques[151]. Les différentes combinaisons d'indices acoustiques sont analysées dans cette étude. La pause se révèle être davantage associée avec le marquage des frontières prosodiques dans la parole spontanée pour différents types de frontières (les frontières des unités mineures ou les frontières internes des unités). Notons que l'auteur utilise la notion de frontière intonative complète (*full intonational boundary*) pour désigner la combinaison des paramètres tels qu'allongement final, cadence mélodique de frontière (*boundary pitch mouvement* de l'école hollandaise) et pause.

Pourtant, nous partageons les considérations de D. Hirst (communication personnelle) suivant lesquelles lorsque est supprimée une pause dans le signal acoustique, nous

[151] Dans les travaux de Blaauw, il a été démontré que les auditeurs étaient capables de classer les extraits de quatre à six syllabes selon le style de parole (spontanée/lue) ; pourtant, bien que leur performance soit significative, elle est moins bonne que lorsque des extraits plus longs sont présentés. On peut donc tirer la conclusion que certaines informations cruciales manquent dans les extraits courts, entre autres les informations sur les réalisations des frontières prosodiques.

percevons tout de même la rupture structurelle au même endroit (il s'agit des pauses associées à la structure linguistique du message).

Une autre source d'arguments pour réduire l'importance accordée à la pause en tant que corrélat de la structuration prosodique, découle des études menées par des chercheurs suédois dans le cadre du projet GROG (Carlsson *et al.* 2002). La tâche proposée aux auditeurs dans cette étude consistait à prédire si l'extrait présenté était suivi d'une frontière prosodique mineure ou majeure alors que les auditeurs n'avaient pas accès aux indices acoustiques *a posteriori*, tels qu'une pause ou ajustement du niveau (*resetting*) mélodique. Les réponses données par des auditeurs naïfs suédois et américains démontrent une forte corrélation avec l'annotation faite par les phonéticiens pour la totalité du corpus, et ceci pour deux empan temporels : les stimuli présentés composaient deux groupes en fonction de leur durée (extrait de 2 secondes *versus* un mot préfrontalier). Un deuxième résultat majeur de cette étude réside dans la mise en évidence de la corrélation forte entre les auditeurs suédois et les auditeurs américains, suggérant ainsi un degré d'universalité dans le marquage acoustique des frontières prosodiques.

- *Pauses subjectives*

Une pause étant un phénomène acoustique complexe, elle peut être abordée du point de vue formel (on insiste alors sur la présence d'une absence physique du signal de parole) ou fonctionnelle (comme induisant une perception de rupture). La définition d'une pause paraît être problématique, étant donné la miscorrespondance observée pour ce phénomène prosodique entre les niveaux articulatoire, acoustique et perceptif. Tseplitis (1974) parle de pauses temporelles et non-temporelles, un autre terme pour les pauses non-temporelles étant les pauses subjectives, qui ont également fait l'objet d'analyses s'intéressant au phrasé prosodique[152]. Dans la même lignée, Duez (1993:23) définit la pause au niveau perceptif par rapport à la perception d'une rupture :

> *« At the perceptual level, a pause can be defined as a phonetic event which causes a listener to perceive a discourse break, usually indicating a grammatical or stylistic boundary[153]. »*

Nous considérons que toutes les frontières prosodiques satisfont la partie formelle de cette définition (perception d'une rupture). Il faut se demander maintenant dans quelle mesure la partie fonctionnelle (indication d'une frontière syntaxique ou stylistique) différencie les deux phénomènes.

Dans l'étude de Duez (1993), les auditeurs ont pour tâche d'indiquer toute pause dans un ensemble de stimuli issus d'un corpus de discours politique en français (la pause étant définie comme toute interruption dans l'énoncé). Les indices acoustiques associés avec les pauses subjectives sont analysés dans une étude comparative : la comparaison est faite entre les performances des auditeurs pour la parole normale et pour la parole à

[152] Comme le remarquent Krivnova et Tchardine (1999), dans le contexte de la recherche appliquée (synthèse vocale) le phénomène des pauses subjectives est d'importance dans la mesure où il est associé avec le concept de la frontière prosodique.
[153] Pour Goldman-Eisler (1968) une pause est une interruption du mouvement articulatoire.

spectre inversé, cette manipulation ayant pour objectif de séparer les sources d'influence potentielles : la syntaxe et la prosodie. Les résultats obtenus démontrent que, par rapport à la structure syntaxique des énoncés-stimuli, les pauses subjectives sont plus fréquentes aux frontières des syntagmes par opposition aux constituants des niveaux supérieurs (clause, énoncé). Les caractéristiques prosodiques sont analysées dans une perspective de la formation progressive du constituant : on compare les valeurs de durée, d'intensité et de f0 pour la voyelle de la syllabe adjacente à la position de la pause subjective et pour la voyelle de la syllabe précédente. Il en découle que pour leur majorité, les pauses subjectives sont associées à un mouvement mélodique, plus fréquemment montant. Simultanément, la perception d'une pause subjective est corrélée avec une grande augmentation d'intensité pour la voyelle de la syllabe adjacente à la frontière. Toutefois, les deux tendances attestées n'atteignent pas le niveau de signification statistique nécessaire. En même temps, les données temporelles sont mieux corrélées avec la performance des auditeurs dans deux expériences décrites dans l'étude : la perception d'une pause (subjective) est corrélée avec un allongement vocalique. Ce résultat fait écho à celui de Repp (1992). Dans cette étude, l'analyse de la performance des auditeurs démontre qu'ils tendent à percevoir une pause après une syllabe allongée même en l'absence d'un silence. Selon Duez (1993) la perception d'une pause subjective est associée à travers l'allongement vocalique à la présence d'une frontière syntaxique. Ce résultat va dans le sens des propos de Gussenhoven (1990) sur la hiérarchie prosodique, discutés dans le chapitre précédent.

Volskaya (2001b), dans une étude perceptive et acoustique, analyse la structuration des questions générales et partielles en russe car il s'agit, pour l'auteur, des énoncés qui, par leur nature linguistique, résistent à la structuration interne. Confrontant les données de l'expérience perceptive et de l'analyse acoustique, l'auteur constate une perception des pauses par les auditeurs là où il en existe un fort potentiel suite au contexte syntaxique, non accompagné des indices mélodiques, mais une pause n'est pas forcément réalisée par le locuteur. L'auteur voit une des explications du phénomène de la pause subjective dans la charge des contraintes cognitives, mais aussi rythmiques (longueur autorisée du syntagme).

Nous venons de voir que, bien que traditionnellement la présence d'une pause soit considérée comme l'indice le plus fort dans le marquage de la frontière prosodique, il a été démontré que la présence d'une pause est souvent associée à d'autres indices acoustiques. De cette nature acoustique composite de la rupture prosodique s'ensuit la perception des pauses subjectives. Il est donc important de déterminer la magnitude de l'influence d'autres indices acoustiques dans la structuration prosodique. Pourtant, une certaine mise en garde est nécessaire. Par exemple, dans l'étude de Bruce *et al.* (1992) sur la perception des frontières prosodiques, une attention particulière a été portée aux stratégies individuelles des auditeurs. Une observation importante, à partir de cette étude, faite par les auteurs eux-mêmes montre que les auditeurs ne sont pas tous sensibles aux mêmes indices : ils peuvent s'appuyer sur des indices temporels comme mélodiques mais le rang de chaque indice diffère selon l'auditeur. Cette considération est importante pour notre future étude en perception.

- *La distribution des pauses*

Comme nous l'avons mentionné dans l'introduction de cette section, un axe de recherche regroupe les études menées pour démontrer quelle structure, syntaxique ou prosodique, rend les meilleures prédictions sur l'implémentation phonétique du message. Dans cette lignée, et en relation avec nos interrogations sur la pertinence de la pause en tant que marqueur de la constituance prosodique, une source importante de données et de conclusions provient des travaux portant sur la distribution des pauses.

Un ensemble d'études en perception a proposé aux auditeurs la tâche de localisation des pauses dans les stimuli, et on a analysé leurs performances en relation avec la durée des pauses, leurs environnements acoustiques, les connaissances linguistiques et la structure sémantico-syntaxique des énoncés. Il est démontré, entre autres, que les pauses ne sont pas perçues de la même façon selon qu'elles apparaissent à des endroits attendus ou non ; dans ce dernier cas, elles sont moins souvent perçues (Repp 1992)[154].

Butcher (1981, cité par Duez 1993) compare la performance des sujets avec trois types de stimuli dans la tâche de localisation des pauses : les énoncés prononcés normalement, les énoncés avec une structure prosodique normale mais non-mots ; les énoncés normaux prononcés avec une fréquence fondamentale monotone. Cette étude aboutit à la conclusion que ce n'est pas la structure syntaxique qui sert de meilleur prédicteur de la distribution des pauses, mais la configuration mélodique associée.

Dans l'étude de Strangert (1990) sur le suédois, les distributions des pauses selon le type de constituant syntaxique (paragraphes, énoncés, clauses, syntagmes syntaxiques[155]) sont analysées dans une tâche de lecture. Il en résulte que seulement 53% des frontières de clauses et 3% des frontières de syntagmes sont associés à une pause.

Une variabilité interlocuteur se manifeste également dans la distribution des pauses. Comme le signalent Fant *et al.* (2003), en lecture cette variabilité est plus marquée à l'intérieur des énoncés par rapport à ses frontières. Un autre résultat obtenu dans cette étude relie la distribution des pauses à la longueur phonologique de l'énoncé : ainsi il n'y a pas de pauses internes dans les énoncés de moins de 16 syllabes ; pourtant, la présence d'une pause est obligatoire dans les énoncés de plus de 20 syllabes.

Nous avons présenté dans le paragraphe précédent les études soutenant que la présence d'une pause est un indice acoustique primordial pour la perception d'une frontière prosodique. Cependant, la situation peut être renversée comme dans l'étude de Horne *et al.* (1995). En effet, dans cette étude, c'est la force de la frontière déterminée en relation avec la structuration prosodique du message qui détermine la durée des pauses ; la pause elle-même n'est pas obligatoire pour le marquage de la frontière.

[154] Comme le démontre l'étude de Hirsh-Pasek *et al.* (1987), les enfants de 7 à 10 mois sont sensibles à la pertinence de la position d'une pause par référence à la structure prosodique de l'énoncé.
[155] Le syntagme syntaxique est défini dans cette étude comme constituant comportant une tête et tous les modificateurs qui la précèdent.

L'étude de Hansson (1998) transfère la méthodologie de Strangert (1990) sur de la parole spontanée et aboutit à la conclusion que, bien que la distribution des pauses en parole spontanée soit plus libre, elle n'est cependant pas aléatoire. Ainsi, on retrouve les pauses majeures aux frontières majeures des énoncés (25%), après les marqueurs discursifs ou les conjonctions (30%), avant les mots lexicaux accentués (18%) et dans d'autres localisations dans 27% des cas. Il est également démontré que seuls 57 % des énoncés sont précédés par une pause. Les données présentées indiquent de la sorte que les distributions des pauses sont largement influencées par la structuration informationnelle du message. Des résultats similaires sont communiqués par Blaauw (1994) sur la distribution des frontières prosodiques en relation avec les mots contextuellement importants : dans son étude, 63 % de frontières prosodiquement marquées se retrouvent avant le mot informativement saillant dans la parole spontanée et seulement 34 % en lecture.

5.1.2.2. Allongement final

Nous nous tournons maintenant vers l'analyse des phénomènes d'allongement final[156]. L'allongement final est un phénomène largement attesté et consiste en ce que les segments précédant certaines frontières linguistiques sont plus longs que les mêmes segments dans d'autres positions internes de l'énoncé. On a attesté l'allongement final non seulement devant les frontières des énoncés ou des syntagmes, mais également avant les frontières des mots, avec, toutefois, une magnitude moindre. Nous présenterons d'abord les études portant sur les constituants supérieurs de la hiérarchie prosodique. Ensuite, nous analyserons les travaux ayant pour objectif de mettre en évidence le mot en tant que constituant prosodique ou syntaxique.

- *Allongement aux niveaux supérieurs de la constituance prosodique*

Dans son approche de la structuration prosodique qu'il dissocie de la structuration du niveau tonal, Gussenhoven (1988, 1990) attribue à la première structuration la fonction de gestion de l'organisation temporelle de la production orale. Globalement, la durée du segment en production dépend du débit global d'élocution, des facteurs intrinsèques et des structures linguistiques des niveaux supérieurs.

Dans un premier temps, nous devons compter avec une longue tradition d'études des indices temporels vers le marquage des frontières (Klatt 1975, Lehiste et al. 1976). Pourtant, dans la tradition de l'époque, ces études mettaient l'allongement final en relation avec les frontières majeures de la hiérarchie syntaxique. Également, la

[156] Quand on parle d'un allongement ou d'un raccourcissement, on est face à une problématique de la durée étalon, qui, hélas, ne peut jamais être définie. Ce qui est supposé dans la terminologie en question est le fait que chaque unité segmentale a sa propre durée intrinsèque. Par rapport à la forme de surface qu'on peut observer à l'aide d'outils, comme le spectrographe ou l'ordinateur, la durée intrinsèque se retrouve modifiée sous l'influence des structures linguistiques. Turk et Shattuck-Hufnagel (2000) signalent une deuxième interprétation possible de l'ajustement temporel des segments : il se peut qu'aucune durée intrinsèque propre n'existe, que les seuls paramètres temporels disponibles soient ceux observés dans la production ; or, ils sont calculés en production et interprétés en perception avec la prise en compte de toutes les influences linguistiques potentielles.

comparaison a été faite entre la position finale d'un énoncé et la position médiane, ces analyses contrastives permettant de différencier les frontières fortes des frontières faibles.

Avec l'entrée dans la linguistique de la théorie des domaines phonologiques, la question qui suscite un fort intérêt est celle des rapports entre structures prosodique et syntaxique du message et de leurs rôles respectifs dans la gestion de l'organisation temporelle. Simultanément, on s'interroge sur l'association entre les indices temporels et les niveaux médians de la hiérarchie prosodique. Les études de Campbell et Ladd (1991), Wightman *et al*. (1992), Byrd et Saltzman (1998), entre autres, ont démontré que l'allongement final est également caractéristique pour les niveaux médians de la constituance prosodique et que le degré d'allongement est corrélé avec la force de la frontière qui suit.

L'étude de Campbell et Ladd (1991) a pour objectif premier de recueillir les preuves en faveur de la structure prosodique récursive défendue par Ladd (1986). C'est une étude prédictive utilisant l'algorithme des réseaux de neurones. Toute syllabe dans le corpus d'étude est annotée à l'aide d'étiquettes portant sur sa position à l'intérieur des constituants de différents niveaux (« une syllabe accentuée dans un pied monosyllabique », « syllabe finale d'un groupe tonal majeur » etc.) et un réseau de neurones est entraîné sur cette annotation. La première annotation est conforme aux prémisses de l'hypothèse de l'étagement strict : les résultats révèlent cinq classes de segments dont la durée est sensible à la position au sein d'un constituant ou d'un autre ; les effets ont été attestés pour des syllabes initiales comme finales des constituants. Dans un deuxième temps, le même procédé est appliqué à l'annotation avec les domaines prosodiques composés (structures récursives de Ladd), comportant quatre niveaux hiérarchiques. Les données de cette deuxième analyse confirment l'existence de trois classes de syllabes initiales. Les résultats majeurs sont pourtant recueillis pour les syllabes finales des unités déterminées : le modèle affiné (avec la structure récursive) assure une meilleure prédiction des durées syllabiques.

L'étude de Wightman *et al.* (1992) a pour objectif de valider empiriquement un modèle d'indices temporels associés au marquage de frontières. L'analyse prend en considération les présupposés issus de la théorie d'isochronie : selon l'hypothèse défendue par cette théorie en anglais (et dans d'autres langues à rythme accentuel), les syllabes accentuées ont tendance à revenir à intervalles égaux ; ou formulé différemment, les pieds rythmiques (le terme alternatif étant les intervalles inter-accentuels) ont une durée constante. En revanche, un pied contenant une frontière prosodique sera allongé (Lehiste, 1977). Dans le corpus analysé[157], la plupart des frontières prosodiques de différents niveaux se retrouvent dans des positions internes aux pieds définis à la base de critères proposés dans Lea (1980). Les mesures temporelles sont analysées dans cette étude en comparaison avec l'annotation du degré de jonction entre les mots (Price *et al.* 1991). Il résulte des analyses acoustiques que la durée du pied est corrélée avec l'indice de frontière, c'est-à-dire, plus haut est le niveau du constituant dans la hiérarchie prosodique (et plus forte est la frontière), plus important est l'allongement subi par le pied. L'analyse détaillée, portant sur la

[157] Il s'agit d'un corpus d'énoncés syntaxiquement ambigus utilisé dans l'étude de Price *et al.* (1991).

distribution de l'effet d'allongement entre les segments, révèle que c'est l'allongement des segments avant la frontière qui est davantage corrélée avec la force de frontière. Ce résultat conduit les auteurs à réfuter l'hypothèse selon laquelle le pied est une unité pertinente pour analyser le phénomène d'allongement associé avec la structuration prosodique[158].

Les résultats suggèrent également que l'allongement final n'est pas un trait catégoriel. Par contre, on en observe une certaine gradation en relation avec certains niveaux de la force de frontière perçue ; il reste donc à savoir combien de niveaux et quels niveaux de la hiérarchie prosodique ont pour corrélat l'allongement final. Il est démontré qu'il existe des différences significatives en quantité d'allongement (par rapport au noyau vocalique de la syllabe préfrontalière) en association avec les indices de frontière 1, 2, 3 et 4 dans l'annotation élaborée par Price *et al.* (1991). Toujours est-il que la quantité d'allongement ne différencie pas les frontières de la force 4, 5 et 6 même si dans une expérience en perception, ces frontières ont été différenciées par les auditeurs. Cela peut être dû au fait que l'allongement final n'est qu'un des indices acoustiques associés avec la rupture : un autre indice mentionné dans l'étude est la présence et la durée d'une pause (notons que les indices mélodiques n'ont pas été pris en compte dans cette étude). D'autre part, les indices 4, 5 et 6 séparent dans l'annotation adoptée les niveaux de la structuration discursive ; il est donc pertinent, selon les auteurs de l'étude, de penser que les indices qui gèrent la structure prosodique pure, y jouent un rôle moindre.

Dans les études que nous venons de présenter, les constituants de la hiérarchie prosodique classique (Nespor & Vogel 1986) sont mis en relation avec l'organisation temporelle du message. Comme nous l'avons signalé précédemment, une vision plus complexe des rapports entre la constituance prosodique et l'organisation tonale reste propre aux études de Carlos Gussenhoven. Ainsi l'étude de Gussenhoven et Rietveld (1992) cherche à mettre en évidence, par une étude expérimentale, les unités tonales de la tradition britannique et les constituants prosodiques de la phonologie des domaines. Le rôle de l'allongement final y est testé dans une expérience en perception : les chercheurs partent de l'hypothèse qu'un allongement final est pertinent en perception et que les auditeurs sont capables de porter les jugements (jugements de bonne formation, *well-formedness jugements*) sur les allongements disproportionnés et sur leur absence. Dans un deuxième temps, cette étude expérimentale explore deux hypothèses sur la gestion de l'organisation temporelle du message : a) l'ampleur de l'allongement dépend du niveau du constituant dans la hiérarchie : plus haut est le niveau du constituant dans la hiérarchie prosodique, plus grande est la magnitude de l'allongement observé en association avec la frontière de ce constituant ; b) en même temps, la quantité de l'allongement est plus élevée avant la frontière d'un domaine d'association. Les syllabes de même contenu segmental sont manipulées en termes de caractéristiques prosodiques de durée et de courbe mélodique associée. Les auditeurs doivent se prononcer sur des syllabes précises (syllabes préfrontalières de différents constituants) : ils doivent dire si la syllabe manipulée dans l'expérience est trop brève, trop longue ou si sa durée est satisfaisante. De manière générale, les résultats de l'étude démontrent que les frontières prosodiques des niveaux supérieurs sont accompagnées d'un allongement

[158] Il est démontré dans cette étude que l'allongement final associé avec le marquage d'une frontière prosodique est localisé dans la rime de la syllabe préfrontalière.

plus important, même si l'effet observé est moins important qu'attendu. En même temps, la coïncidence de la frontière d'un constituant prosodique et celle d'un domaine d'association (unité définie pour le niveau tonal) renforce l'effet d'allongement : en particulier, les marges d'allongement toléré se voient élargies dans l'étude menée.

- *Études sur l'allongement des mots*

Comme nous l'avons dit en présentant le mot en tant que constituant prosodique, il importe de savoir quelles sont les correspondances entre un mot constituant de la hiérarchie prosodique et un mot constituant morphosyntaxique. À part ce questionnement théorique, la problématique des indices associés avec les frontières des mots est au centre de la recherche sur le traitement perceptif de l'oral : en entendant une suite des sons de parole, on est amené à y détecter les frontières des mots à l'étape d'accès au lexique. Les études sur l'organisation temporelle des mots apportent en partie les arguments pour répondre à cette problématique[159].

Dans l'ensemble des études sur l'influence de la frontière du mot sur la durée des syllabes (Abercrombie 1965 ; Lindblom & Rapp 1973; Beckman & Edwards 1990) on compare les structures du type « poppa pose » *versus* « pop oppose » : il a été démontré que la première structure est caractérisée par une durée réduite de la syllabe « pop » et un allongement de la syllabe « a » ; la tendance est renversée pour la deuxième structure analysée.

Dans l'étude de Turk et Shattuck-Hufnagel (2000), les auteurs citent et testent expérimentalement cinq mécanismes d'ajustement de la durée des segments : deux sont liés au marquage des frontières (allongements final et initial), les trois autres étant l'allongement accentuel et les modifications des durées suite aux propriétés phonologiques des constituants de niveaux supérieurs (phénomène connu sous le terme de *polysyllabic shortening*). Remarquons que quatre sources de la variation de durée font ainsi référence à la structuration prosodique du message. Dans cet ensemble de sources de la variabilité des durées segmentales, l'allongement final paraît être le phénomène le plus étudié (Umeda 1975 ; Nakatani *et al.* 1981 ; Beckman & Edwards 1990, entre autres).

Un autre phénomène, ayant suscité un intérêt important est l'allongement accentuel (Eefting 1991 ; Sluijter et van Heuven 1995 sur le hollandais ; Turk & Sawusch 1997; Turk & White 1999 ; Cambier-Langeveld & Turk 1999 sur l'anglais et le hollandais). Toutes ces études démontrent un effet d'allongement des syllabes si le mot porte un accent mélodique. Pour les deux langues, l'anglais et le hollandais, un traitement asymétrique des syllabes inaccentuées est caractéristique : l'allongement induit par la présence d'un accent mélodique modifie plus les syllabes postérieures à la syllabe porteuse de la proéminence mélodique que pour les syllabes précédentes.

[159] Il faut également préciser que la nature prosodique ou morphosyntaxique de l'allongement accentuel n'est pas clarifiée jusqu'à présent. L'étude de Cambier-Langeveld et Turk (1999) montre que les mots clitiques monosyllabiques ne subissent pas les mêmes effets d'allongement bien qu'ils soient attachés prosodiquement au mot lexical précédent.

Le phénomène de raccourcissement polysyllabique (*polysyllabic shortening*) a été mis en évidence dans l'étude de Lehiste (1972) qui a observé que la syllabe [stɪk] était plus longue dans le mot monosyllabique « *stick* » par rapport au bisyllabique « *sticky* » ou encore au tri-syllabique « *stickiness* » (cf. également les études de Huggins, 1975, Nakatani *et al.* 1981, Port, 1981).

Parallèlement, l'étude de Turk et Shattuck-Hufnagel (2000) cherche à localiser à l'intérieur du mot et de la syllabe les effets des différents mécanismes influençant la durée du mot : il a été remarqué que l'allongement initial portait surtout sur la consonne d'attaque syllabique, tandis que l'allongement final induisait des modifications plus marquées dans la rime de syllabe. Une autre question centrale à l'étude est celle du rôle des frontières des mots grammaticaux.De manière générale, les résultats obtenus confirment les prédictions sur l'influence exercée par une frontière de mot sur l'organisation temporelle des syllabes et ceci dans des contextes prosodiques globaux différents. Ainsi la modification de la position d'une frontière induit la redistribution des durées syllabiques à gauche comme à droite de la frontière. On note également que le schéma de distribution des durées diffère au niveau du mot de celle établie pour les constituants supérieurs. Plus précisément, c'est le noyau vocalique de la syllabe préfrontalière qui subit le plus de modification et on n'observe pas d'allongement progressif. Il est aussi démontré que les locuteurs portent plus d'attention au marquage de la frontière du mot à proximité d'un accent mélodique.

Les résultats sur le rôle de la frontière d'un mot grammatical nous paraissent très importants à la lumière de notre discussion dans le chapitre précédent du groupe clitique en tant que constituant prosodique (ou du mot phonétique dans la tradition russe). En effet, les analyses effectuées démontrent que les effets d'allongement à la frontière entre un mot lexical et un mot grammatical se distinguent de ceux qui se produisent à la frontière entre deux mots lexicaux. Pourtant, on ne peut pas considérer la frontière entre un mot lexical et un mot grammatical comme inexistante. En même temps, les deux frontières du mot grammatical sont plus faibles que les frontières entre deux mots lexicaux. De ce résultat, il ressort que le mot clitique peut appartenir simultanément à deux groupes clitiques.

La présentation que nous avons entreprise dans cette section s'est articulée autour de la problématique de l'influence de la hiérarchie prosodique sur l'organisation temporelle du message linguistique. Il reste à savoir dans quelle mesure les effets d'allongement attestés peuvent servir d'indices pour la détermination des frontières des constituants ?

- *Débit d'articulation et structuration prosodique*

Une autre série d'arguments sur l'organisation temporelle du message dans sa relation avec la structure prosodique relève des études sur la variation du débit articulatoire[160]. Il est reconnu que le débit d'articulation est variable dans le discours, et ceci sous

[160] C'est une mesure de la vitesse d'élocution, toutes les pauses étant exclues (Goldman-Eisler 1968 ; Miller, Grosjean et Lomanto 1984). Le débit d'articulation est un des paramètres influençant la durée des segments.

l'influence de plusieurs facteurs : discursifs[161], syntaxiques et phonologiques[162]. Cette variabilité étant reliée avec l'organisation temporelle du message et la distribution des durées, on la retrouve au niveau du mot ainsi qu'aux niveaux supérieurs de la hiérarchie de constituants. Par rapport à l'organisation syntaxique, les résultats de l'étude d'O'Shaughnessy (1995) démontrent, entre autres, que le débit de l'articulation est réduit au début du constituant syntaxique.

Dans la perspective envisagée, Dankovicova (1997) se donne pour tâche d'élaborer un cadre théorique expliquant la variabilité du débit d'articulation. Cette étude est effectuée à partir des extraits du corpus de parole spontanée en langue tchèque. La problématique de la variabilité du débit d'articulation est abordée à travers la question du domaine qui encadre cette variabilité. Comme le note l'auteur, dans les études précédentes sur le débit d'articulation, la notion de la variation locale reste imprécise, de même que le domaine à travers lequel on étudie cette variabilité est choisi de manière arbitraire. Trois domaines se portent candidats potentiels : les unités interpausales, les clauses syntaxiques et les unités intonatives. Deux hypothèses sont alors testées : premièrement, on cherche à démontrer la stabilité du débit d'articulation à l'intérieur du domaine étudié ; dans le cas contraire, on essaie de retrouver un schéma de la variation du débit d'articulation à l'intérieur du domaine. Ainsi on observe une variation considérable au sein de l'unité intonative (ce résultat n'est pas surprenant étant donné toutes les études sur l'allongement final) ; simultanément, on constate une tendance vers un ralentissement général du débit d'élocution au sein de l'unité intonative. Notons que deux autres unités testées ne démontrent aucun patron de la variation systématique du débit d'élocution. Il semble donc, à partir de l'analyse effectuée, que la variation du débit d'articulation en tchèque soit gérée dans le cadre du syntagme intonatif.

Dans la même veine, Jun (2003) présente un projet dans lequel on étudie l'influence des quatre facteurs sur le phrasé prosodique en coréen. Pour ce faire, on analyse les rôles respectifs de la syntaxe, de la distribution du focus, de la longueur du syntagme et du débit d'élocution. On reconnaît deux unités prosodiques plus grandes qu'un mot en coréen, à savoir le syntagme intonatif, défini par la présence d'un ton de frontière et un allongement perceptible (et éventuellement, une pause postérieure) et le syntagme accentuel, caractérisé par les deux tons de frontière initiale et finale (sans aucun allongement). Au niveau du syntagme accentuel, on constate que les mots lexicaux forment les syntagmes accentuels à part ; en même temps, deux mots interliés (par exemple, dans le groupe Adjectif + Nom) forment un seul syntagme accentuel si sa longueur ne dépasse pas 5 syllabes dans la parole normale et 7 syllabes dans la parole rapide. En somme, cette étude démontre le rôle important de la longueur du syntagme dans la définition du phrasé prosodique ; pourtant, on constate un lien fort entre ce facteur et le débit d'élocution, ainsi que entre les facteurs sémantiques.

[161] Eefting (1990) travaillant sur des énoncés produits dans un contexte conversationnel et ces mêmes énoncés insérés dans une phrase porteuse, trouve que le débit ne dépend pas uniquement des propriétés phonologiques mais également de l'intention du locuteur à être compris, de son souci de collaborer avec les auditeurs. On constate entre autres, qu'un ralentissement précède une information nouvelle, importante.

[162] L'étude de Crystal et House (1990, analyse à partir des unités interpausales, portant sur la lecture en anglais,) a démontré que le débit d'articulation varie en fonction des caractéristiques phonologiques des unités analysées (complexité des syllabes, nombre d'accents mélodiques).

5.1.2.3. Corrélats mélodiques du phrasé prosodique

Nous avons présenté dans les sections précédentes des données illustrant l'influence de la structure prosodique sur l'organisation temporelle du message. Les résultats obtenus dans les études présentées ont servi d'argument pour justifier le rôle de la structure prosodique en tant que système d'interface entre la syntaxe et la phonétique. On précise que la question de la valeur prédictive des phénomènes attestés n'est pas toujours complètement exploitée. Nous nous tournons maintenant vers l'analyse des caractéristiques mélodiques sujettes aux influences de la hiérarchie prosodique.

En discutant des aspects phonologiques de la hiérarchie prosodique, nous avons présenté un ensemble de théories qui définissent les constituants en termes de courbes mélodiques et/ou accentuelles (Pierrehumbert et Beckman 1986, Jun et Fougeron 1998, Di Cristo et Hirst 1996). Par contre, les travaux présentés ci-après se focalisent sur les corrélats mélodiques du phrasé prosodique, dans sa version classique décrite par Nespor et Vogel (1986).

Le travail de de Pijper et Sanderman (1994) étudie la perception du degré de rupture entre les mots en relation avec différents corrélats acoustiques et les constituants prosodiques issus de la phonologie des domaines. Les auteurs travaillent avec la notion de force perceptive de frontière (*prosodic boundary strength*) définie comme étant la moyenne des indices de force de frontière attribués à toute frontière de mot dans une étude perceptive. En comparant la performance des auditeurs avec le phrasé prosodique défini dans le système de synthèse de parole développé à IPO à partir des analyses syntaxiques comme métriques, les auteurs constatent que les indices de force de frontières sont positivement corrélés avec le niveau de rupture dans la hiérarchie prosodique de l'énoncé (ils travaillent à partir de trois catégories, absence de frontière, frontière du syntagme phonologique et frontière du syntagme intonatif). Les corrélations entre la force perceptive de frontière et les indices acoustiques de pause, de la configuration mélodique de frontière (dans la tradition de l'approche IPO) et du *reset* de f0 sont analysées : outre la démonstration d'une grande variabilité intra- et interlocuteur, il est démontré que plus l'indice de force de frontière est élevé, plus nombreux sont les paramètres acoustiques qui y sont associés. De manière générale, c'est la présence d'une pause qui détermine le plus la performance des auditeurs, suivie par les tons de rupture. En ce qui concerne le *reset* mélodique, cet indice est corrélé avec les indices élevés de la force perceptive de frontière. Or, ce paramètre acoustique ne fonctionne jamais seul ; il est associé à la fois à la présence d'une pause et d'un ton de frontière. Travaillant avec les enregistrements de 3 auditeurs, les auteurs constatent pourtant que le *reset* mélodique est surtout propre au locuteur professionnel, les locuteurs non-professionnels ne l'utilisent jamais. La présence d'une pause dans les données de de Pijper et Sanderman (*op. cit.*) est systématiquement associée à la discontinuité mélodique (82%, 82% et 79 % de cas pour les trois locuteurs dont les productions langagières sont analysées). Il semble donc que la discontinuité mélodique soit le seul indice acoustique qu'on retrouve systématiquement en isolation.

Les effets de la hiérarchie prosodique sur l'organisation mélodique sont également analysés dans les contributions de Ladd (1988, 1990) : ce sont les phénomènes de déclinaison et d'abaissement tonal (*downstep*) qui sont au centre de l'analyse. Dans ses travaux, Ladd présente le modèle métrique de l'abaissement tonal, le modèle qu'il développe à partir des résultats de l'analyse acoustique des données expérimentales (Ladd, 1988). Cette étude est effectuée sur la base de constructions où les trois clauses grammaticales / syntagmes intonatifs sont reliées par des conjonctions *et/mais* (*and/but*) : *A and B but C* ; *A but B and C*. Les données expérimentales démontrent qu'il existe une différence significative entre les valeurs de f0 des pics mélodiques initiaux dans les trois syntagmes intonatifs selon que le syntagme est précédé par la conjonction 'et' ou 'mais'. Cette différence est interprétée comme une différence dans le registre mélodique associé avec chaque syntagme et sert d'argument en faveur d'une représentation métrique du registre (figure 5.1)

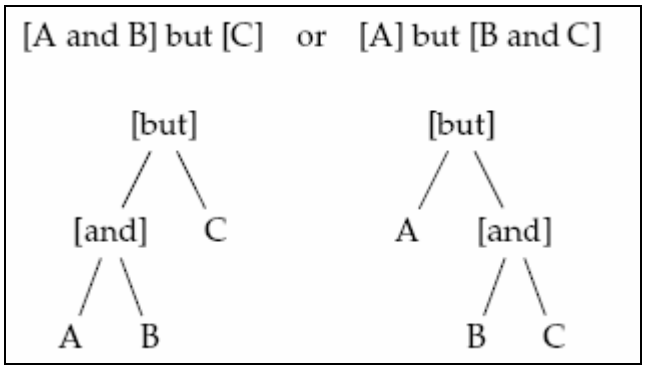


Figure 5.1 : Structures syntaxiques analysées dans l'étude de Ladd (1988)

Selon Ladd (1990) les résultats expérimentaux peuvent être interprétés comme une preuve de l'existence d'une hiérarchie prosodique sous-jacente dont une des fonctions est de contrôler les ajustements du registre mélodique aux frontières des constituants prosodiques.

Parmi les études des indices mélodiques en association avec la structuration prosodique, il est légitime, à notre avis, de classer les études sur la distribution des accents mélodiques. Un des arguments phonologiques pour justifier le constituant du syntagme phonologique provient de l'analyse du phénomène de déplacement de l'accent (*stress-shift*), attesté pour l'anglais et le français. Deux explications en ont été proposées : la première étant fondée sur la théorie de la grille rythmique (relevant de la contrainte d'eurythmie, fondée sur l'étude de Prince et Liberman, 1977) ; la deuxième relevant des contraintes sur la distribution des accents mélodiques (Bolinger, 1958[163]).

Ainsi dans la phrase « Massachusetts miracle », le mot « miracle » porte l'accent principal sur la première syllabe, ce qui crée le contexte d'un *clash* accentuel avec les syllabes accentuées des deux mots adjacents. Donc, on prédit que le mot « Massachusetts » subira un déplacement d'accent sous les contraintes rythmiques, ce qui entraînera une perception d'une proéminence plus forte sur la syllabe plus tôt dans le mot. Cette analyse est inspirée de l'approche rythmique de la résolution du *clash*

[163] Dans l'approche de Bolinger deux marqueurs intonatifs du syntagmes sont pressentis : l'un étant l'accent mélodique précoce dans le syntagme (accent thématique) et l'autre – l'accent mélodique tardif (accent rhématique).

accentuel. La deuxième approche, mélodique, est défendue dans les travaux récents de Shattuck-Hufnagel : on interprète le déplacement accentuel en termes de distribution des proéminences mélodiques et de la variabilité dont dispose le locuteur pour le placement de l'accent prénucléaire. Cette conception se fonde sur le présupposé traditionnel de la théorie autosegmentale de l'intonation, selon lequel la distribution des accents lexicaux détermine les sites potentiels pour le placement de la proéminence mélodique. Ainsi les mots avec deux syllabes fortes et plus avant la syllabe porteuse de l'accent principal du mot, peuvent être réalisés soit avec un accent mélodique précoce dans le mot, soit avec un accent mélodique sur la syllabe accentuée principale du mot ; ils peuvent également recevoir une double proéminence mélodique.

La troisième approche possible vers le phénomène du déplacement de l'accent et de résolution des *clashes* accentuels intègre les présupposés de deux approches plus récentes. Cette fois, on adopte une position rythmique et mélodique à la fois. La validité de cette approche est testée dans l'étude de Shattuck-Hufnagel *et al.* (1994) dans une étude expérimentale à partir du corpus de la parole radiophonique. Dans un premier temps, on se pose la question de savoir si la proéminence précoce dans le mot est réellement perçue par les auditeurs : dans ce but, on analyse la performance des annotateurs travaillant sur le corpus. Les résultats de cette étude démontrent que l'on observe dans le corpus un taux élevé de mots à double accentuation, même dans les contextes sans résolution du *clash* accentuel (12 parmi 14 expressions dans le corpus). Cette configuration révèle la tendance de la part des locuteurs de placer l'accent mélodique sur la première syllabe disponible ayant le potentiel de porter l'accent mélodique marquant ainsi le début du syntagme.

Globalement, les résultats obtenus dans cette étude expérimentale rejoignent les théories sur la distribution de proéminences mélodiques qui ne relient pas le déplacement accentuel avec le *clash* accentuel rythmique. Par contre, ces théories (Bolinger, Gussenhoven) intègrent des règles de placement d'accents mélodiques qui sont fondées sur la position de l'accent mélodique potentiel dans le constituant prosodique et sur la prise en compte de la rythmicité de la distribution d'accents mélodiques qui en résulte.

Les études inspirées de la problématique de la distribution des accents mélodiques fournissent donc une preuve de la pertinence de la structure prosodique hiérarchique ; avec deux autres sources d'influence, la contrainte de la régularité rythmique et les facteurs discursifs, la structuration de l'énoncé en termes de constituants prosodiques gère le placement d'accents mélodiques qui, eux, servent à démarquer des constituants en question. Nous pouvons également évoquer le travail de Post (2000) qui étudie le rôle des syntagmes phonologiques en français dans la gestion de la distribution des proéminences mélodiques.

5.2. LA RÉALITÉ COGNITIVE DES CONSTITUANTS PROSODIQUES

La question de la constituance prosodique est également exploitée en psycholinguistique et les données empiriques obtenues dans les études expérimentales soutiennent, parfois indirectement, l'hypothèse de la réalité cognitive de la hiérarchie prosodique[164].

Moyennant le jeu des proéminences, le phrasé prosodique et le choix du contour mélodique, la prosodie exerce un ensemble de fonctions dans le discours, dont nous avons discuté dans les premières pages de cette recherche. À côté des fonctions orientées vers le message communiqué et la situation discursive, on y reconnaît également les fonctions orientées vers le locuteur : la prosodie d'un énoncé transmet des informations sur le locuteur, son appartenance dialectale, son état émotionnel, son attitude envers le dit. Dans ce contexte, il semble contestable d'introduire la notion de prosodie par défaut, largement adoptée dans les systèmes appliqués, et de considérer donc la prosodie comme étant complètement indépendante du locuteur.

La personnalité de locuteur intervient également dans la mesure où ce dernier adopte ou non une attitude de coopération envers son interlocuteur. Nous caractérisons l'attitude de coopération par le souci du locuteur à transmettre son message et à être compris, et par conséquent, à structurer son discours et y mettre des marqueurs à l'attention de l'auditeur. Cette reconnaissance des stratégies individuelles des locuteurs et de la variabilité stylistique a pour conséquence que certains styles de parole ont été privilégiés dans les études prosodiques : en particulier, les études qui choisissent un style de lecture radiophonique défendent ce choix en indiquant que la structuration prosodique des énoncés y est beaucoup plus clairement marquée.

On suppose donc que la structure prosodique d'un énoncé comporte des instructions sur l'interprétation du message. En effet, pour transmettre les informations sur l'interprétation visée du message, le locuteur et l'auditeur doivent obligatoirement disposer d'un code commun, d'une convention sur le marquage des différents phénomènes. Dans le cas contraire, la coopération entre les interlocuteurs serait impossible. Les études psycholinguistiques portent donc sur l'accommodation du message par le locuteur pour faciliter ses traitement, compréhension et interprétation par l'auditeur. Les études en perception et en compréhension du message à l'oral constituent donc un premier axe d'exploration du rôle de la constituance prosodique en psycholinguistique ainsi qu'en neurolinguistique.

[164] La psycholinguistique, en tant que branche de la psychologie expérimentale, étudie les structures sonores des langues ainsi que les structures cérébrales impliquées dans le traitement de ces structures. Nous défendons donc l'idée de la nécessité de la plausibilité psychologique des modèles linguistiques intégrés dans les technologies de parole. Comme le souligne Peperkamp (2002), l'histoire commune de la phonologie et psycholinguistique a connu les périodes d'un fort attrait mutuel, par exemple dans les études portant sur l'influence de la phonologie en perception (cf. Polivanov 1931 et théorie du crible phonologique), ou dans le contexte d'une quête des données sur la réalité cognitive de l'approche de phonologie générative (Mehler 1968). La scission date des années 1970, c'est-à-dire du moment où émerge la phonologie plurilinéaire.

5.2.1. Les études neurolinguistiques

Depuis peu les études en neurolinguistique se penchent sur la question de la segmentation du signal acoustique, des indices disponibles dans le signal et associés à cette segmentation, et le rôle des informations prosodiques (frontières prosodiques) dans ce processus. Ces études utilisent la méthodologie des potentiels évoqués[165] pour étudier les réponses électrophysiologiques du cerveau aux stimuli acoustiques : ce paradigme a pour avantage d'avoir accès aux données du traitement « en ligne » du langage, la méthode EPG assurant l'accueil des informations précises sur des aspects temporels du traitement perceptif. La méthodologie de la résonance magnétique est utilisée pour compléter la représentation résultante : elle permet de localiser les processus induits dans des régions précises du cerveau.

Les stimuli dans l'étude de Steinhauer *et al.* (1999)[166] sont des énoncés syntaxiquement ambigus selon que le verbe est transitif ou non, ce qui résulte en une structure avec une frontière prosodique après le verbe transitif. On a mis en évidence dans cette étude l'existence d'une modulation positive (*Closure Positif Shift,* CPS) associée à la perception d'une frontière prosodique forte. En manipulant la structure prosodique des énoncés dans une étude postérieure, Toepel et Alter (2002) ont pu mettre en évidence que ce sont les frontières des unités intonatives et non pas les syntagmes phonologiques qui sont associées avec le CPS.

L'étude de Pannekamp *et al.* (2005) cherche à savoir si le CPS repose exclusivement sur des informations prosodiques. En manipulant le signal, ce travail analyse les interactions entre le phrasé prosodique et les informations segmentales, sémantiques et syntaxiques, manipulations résultant en quatre conditions expérimentales. L'étude précédente de Toepel et Alter (2002) explorait déjà le paradigme fondé sur des stimuli désémantisés. Dans cette étude, le CPS n'ayant pas été présent pour la condition manipulée, les auteurs concluent sur l'importance de « *sens unit condition* » (Selkirk 1984). Or, l'étude de 2005 dément ce résultat puisque dans ce travail, le CPS est attesté sous toutes les conditions expérimentales en association avec les frontières des unités intonatives. Cependant, la distribution de CPS dans les régions du cerveau n'est pas homogène et varie selon la condition. Notons que la tendance vers une localisation dans l'hémisphère droit est présente pour les stimuli composés de non-mots, mais non pour la version filtrée des stimuli.

La recherche future explorera les données neurolinguistiques en association avec d'autres niveaux de la constituance prosodique.

[165] Remarquons que la méthodologie des potentiels évoqués a tout d'abord été utilisée pour révéler les mécanismes du traitement du langage écrit (lecture silencieuse). Le langage oral possède une propriété supplémentaire : la prosodie.
[166] Cette étude est résumée dans Toepel & Alter (2002).

5.2.2. Les études psycholinguistiques

Dans le décodage de la parole, on distingue les phases suivantes : décodage, segmentation, reconnaissance, interprétation (Cutler & Clifton 1999). Ici le décodage consiste à reconnaître, dans l'environnement acoustique, ce qui est de la parole ; puis on procède par segmentation du flux de parole en mots et syntagmes intonatifs ; la troisième étape correspond à la phase d'accès au lexique. Les psycholinguistes ont modélisé les différentes étapes du traitement perceptif du signal et ont travaillé sur l'influence des propriétés phonologiques de la langue maternelle des sujets sur leur performance dans les tâches proposées. On reconnaît l'influence majeure de l'inventaire des segments de la langue, de ses propriétés phonotactiques, des structures syllabiques et métriques, et des règles phonologiques.

Le modèle du traitement du langage parlé sous-jacent aux études expérimentales du laboratoire de Sciences Cognitives et Psycholinguistique (Peperkamp 2002 ; figure 5.2) comporte le niveau de la représentation pré-lexicale, de la représentation phonologique, situé entre le signal et le lexique, ce qui constitue l'originalité de cette approche. Hormis la structuration du signal acoustique en termes des segments catégoriels (les phonèmes), la représentation prélexicale intègre une structuration du matériau en constituants prosodiques (énoncé, syntagmes intonatifs et syntagmes phonologiques). Le rôle principal de cette représentation consiste à fournir des clefs assurant la segmentation du signal en mots afin de passer au niveau suivant du traitement de l'oral : accès au lexique. Ce même modèle sert de cadre pour expliquer le processus d'acquisition du langage par le nourrisson. En effet, les premiers mots acquis seront ceux qui sont situés aux bornes des constituants prosodiques.

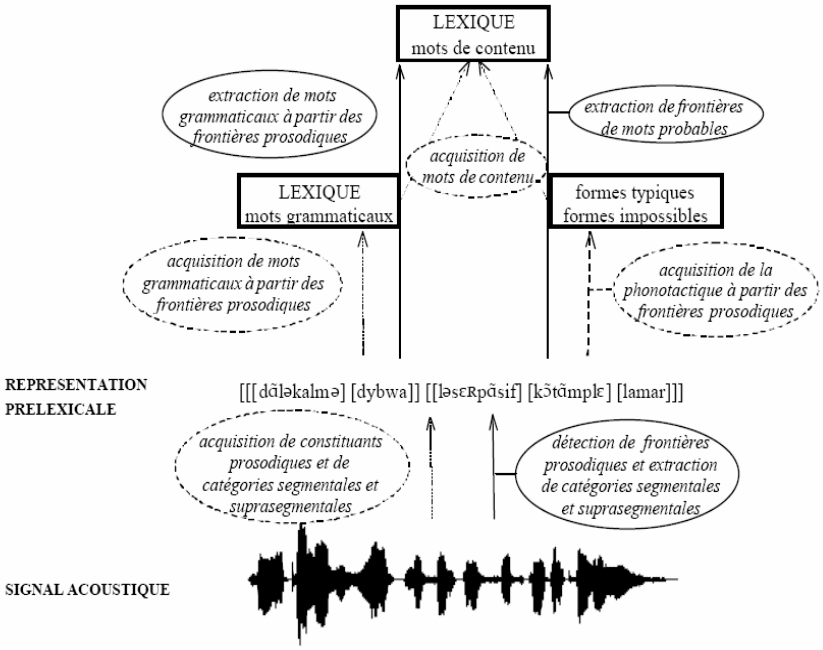


Figure 5.2 : Modèle du traitement phonologique du langage parlé (LSCP)

Dans ce groupe d'études en perception, trois grands axes de recherche portent sur le rôle de la prosodie dans l'accès au lexique, sur la désambiguïsation des structures syntaxiques, et sur la structuration du discours. Nous présentons dans la suite de cette section les principales méthodologies et les résultats obtenus pour ces trois sous-parties.

5.2.2.1. Prosodie dans l'accès au lexique

Dans le traitement auditif de la parole continue, la première question est de savoir combien de mots reconnaître ; dans cette perspective, il est donc important de comprendre comment les indices prosodiques participent à l'identification des frontières de mots. Cependant, il est important de tenir compte de deux facteurs majeurs qui proscrivent les généralisations sur le traitement perceptif. Précisons que ce sont le facteur de la langue du locuteur (certains indices sont spécifiques à la langue) et la variabilité attestée dans le discours continu (les locuteurs peuvent utiliser les indices variables pour signaler les frontières).

Les indices acoustiques ne sont pas nombreux pour signaler les frontières de mots : on y répertorie certaines clés segmentales (variations allophoniques), des contraintes phonotactiques ainsi que des traits prosodiques : le *timing* des syllabes et des segments, la distribution des proéminences. Il est important de ne pas oublier que les variations segmentales sont gouvernées par des facteurs prosodiques de haut niveau, liés à la distribution des proéminences et à la constituance prosodique.

Une série d'études de Cutler et ses collaborateurs dans les années 1990, réalisées dans le paradigme de la détection d'un phonème ou d'une syllabe, démontrent la dépendance des stratégies de segmentation lexicale par rapport aux propriétés prosodiques de la langue considérée. En particulier, ces recherches mettent en évidence que l'unité principale de la perception dépend des propriétés rythmiques de la langue ; le choix de cette unité corrobore la classification des langues par référence à leur structure rythmique, avec la distinction faite entre les langues à rythme accentuel, les langues à rythme syllabique et les langues à rythme moraïque. Ainsi la syllabe s'est-elle révélée être l'unité minimale de perception en français, mais non pas en anglais ou en japonais. Les sujets anglophones procèdent à une segmentation première à partir des informations sur la distribution des syllabes métriquement fortes. Il a été démontré dans les études postérieures que les locuteurs francophones sont capables de prêter attention à une unité aussi petite qu'un segment, mais le raisonnement est mené en termes de constituants sub-syllabiques : les attaques de syllabes constituent les points d'alignement auxquels les processus d'accès au lexique sont déclenchés. De même, les propriétés rythmiques sur lesquelles la segmentation est fondée sont propres à la langue maternelle du locuteur et non à celle du signal d'input : dans une perspective crosslinguistique, il a été prouvé que les stratégies mises en places par les Français sont les mêmes pour les stimuli en français comme en japonais.

Le modèle de Mehler, Dupoux et Ségui (1990), dérivé sur la base des résultats expérimentaux cités ci-dessus, insiste sur les différences d'encodage des mots dans le lexique mental selon la langue maternelle des locuteurs. En particulier, l'étude expérimentale de Dupoux *et al.* (1997) démontre une insensibilité des auditeurs francophones aux phénomènes accentuels, le comportement inverse étant propre aux hispanophones. L'interprétation suggérée par le modèle consiste à dire qu'en espagnol l'accent lexical est intégré dans la représentation phonologique du mot, mais non en français.

Dans les études portant sur le stade d'identification des mots, la question soulevée est de savoir si l'information prosodique participe et facilite l'activation des candidats potentiels et si, ensuite, dans la situation où plusieurs candidats obtiennent le même score, elle intervient dans le choix du « bon candidat ». Le paradigme méthodologique mis en place dans ce groupe d'études est monitoring multimodal (*cross-modal priming study,* Swinney, 1979) : les auditeurs entendent un énoncé stimulus et, lors de la reproduction auditive, une cible visuelle apparaît en un point, en association ou en contraste avec le contenu sémantique du message, et les auditeurs sont soumis à une tâche de décision lexicale (le temps de réaction étant la variable dépendante étudiée).

De manière générale, on considère que les informations prosodiques jouent un rôle lors de l'accès au lexique dans certaines langues (langues à accent libre), toutefois, leur importance reste à évaluer. Premièrement, les recherches sur le traitement perceptif des différences tonales dans les langues à tons ont démontré que les informations mélodiques sont traitées plus lentement que les informations segmentales. Les contrastes prosodiques sont omis par les auditeurs s'ils sont complètement redondants : comme le démontrent les études sur l'anglais, les locuteurs de cette langue ne se servent pas des informations sur la position de l'accent lexical (Cutler *et al.* 1986) étant donné que le contraste accentuel est soutenu par le contraste segmental entre voyelles fortes et voyelles faibles. Les chercheurs retiennent l'hypothèse selon laquelle les informations prosodiques ont un rôle à jouer lors de la phase de pré-traitement (activation des candidats) dans les langues où ces informations participent de manière significative à l'identification des mots.

5.2.2.2. Traitement perceptif et interface « prosodie – syntaxe »

Les études de ce groupe se penchent sur l'évaluation du rôle de la structure prosodique en désambiguïsation des structures syntaxiques ambiguës[167]. On y distingue les ambiguïtés locales (résolues par le contexte linguistique qui suit) et globales.

La théorie de la phonologie prosodique est une théorie des domaines en phonologie et simultanément une théorie des interactions entre différents composants de la grammaire. Si, lors de la période de dominance de la théorie générative classique, on insistait sur le fait que la structure prosodique était identique à la structure syntaxique, la conclusion était que seuls deux énoncés dont les structures syntaxiques étaient différentes, pouvaient être désambiguïsés à partir des indices prosodiques. Avec le développement de la phonologie des domaines, on adopte une approche dite « prosodique » : deux énoncés peuvent être désambiguïsés s'ils ont des structures prosodiques différentes[168,169]. Les deux hypothèses, dites « syntaxique » et « prosodique », sont

[167] L'ambiguïté en linguistique fait référence à plusieurs interprétations possibles pour une séquence de segments. L'ambiguïté peut se manifester au niveau pragmatique (on ne connaît par le référent d'un pronom) ; au niveau lexical (homonymes, plusieurs significations possibles pour un mot) ; au niveau syntaxique.

[168] Dans la lignée des considérations selon lesquelles les catégories prosodiques étant construites à la base des informations syntaxiques, il n'existe pas de correspondance bi-univoque entre les constituants prosodiques et les constituants syntaxiques.

soumises à une étude expérimentale dans le travail fondateur de la phonologie prosodique de Nespor et Vogel (1986). Les auteurs analysent la performance de sujets dans une tâche de désambiguïsation avec dix ensembles de stimuli : au niveau syntaxique, les deux énoncés ambigus peuvent être différents par rapport à la structure des constituants ainsi qu'aux étiquettes attribuées aux constituants ; au niveau prosodique les différences sont possibles au niveau des syntagmes phonologiques et à celui des syntagmes intonatifs ; il en résulte alors dix combinaisons possibles. Les résultats de cette étude confirment que ce sont les différences en structure prosodique pour les énoncés ambigus qui déterminent le degré de faisabilité de la désambiguïsation[170]. L'hypothèse prosodique permet une analyse plus différenciée des performances observées, en particulier des prédictions sur la probabilité de la désambiguïsation. On observe trois cas de figure :

a) les stimuli avec une structure prosodique différente au niveau des syntagmes intonatifs obtiennent le meilleur score de désambiguïsation ;

b) dans un deuxième temps, la désambiguïsation dépend de la structuration en termes de syntagmes phonologiques ;

c) enfin, pour les stimuli avec des structures syntaxiques différentes mais non soutenues par l'organisation prosodique, la désambiguïsation est impossible.

Dans l'étude de Pynte et Prieur (1996) sur les structures syntaxiques localement ambiguës, l'attention est portée aux frontières mineures, internes aux clauses grammaticales. Ce travail prend en considération à la fois le rôle de la structure prosodique dans la gestion de l'interface « prosodie ~ syntaxe » et la contrainte de l'unité de sens (Selkirk 1984). Une série d'expériences en perception réalisées dans le cadre de cette étude avait pour objectif de tester les hypothèses sur :

a) le rôle indépendant de la structure prosodique de l'énoncé dans le processus de compréhension du message ;

b) la gestion conjointe des contraintes syntaxiques et sémantiques.

Le paradigme expérimental de la détection du mot-cible est utilisé dans cette étude[171] : on suppose que le temps nécessaire à la détection du mot dans l'énoncé dépend de la complexité des opérations mentales nécessaires pour intégrer ce mot dans la structure de l'énoncé. Les auteurs cherchent à vérifier si la structure prosodique influence le traitement perceptif de l'attachement syntaxique dans le cas de la structure [NP + V + NP + PP]. On manipule le phrasé prosodique, selon 2 modalités, et le type d'attachement du syntagme prépositionnel, toujours selon 2 modalités ; la structure

[169] Déjà dans les réflexions premières sur les relations entre prosodie et syntaxe, on insiste sur l'absence de correspondance bi-univoque entre la structure syntaxique et la structuration prosodique, mais on attribue ce phénomène au domaine de la performance.
[170] Or, l'analyse des résultats présuppose que les deux lectures de l'énoncé ont la même probabilité, ce qui n'est pas toujours le cas. Les auteurs pensent y remédier en augmentant le nombre de stimuli.
[171] Quoi qu'il en soit, certains chercheurs considèrent ce paradigme comme peu propice pour rendre compte des stratégies du traitement en ligne.

argumentale du verbe[172] constitue également un facteur expérimental[173]. La manipulation de la structure prosodique a pour effet que, sous certaines conditions, l'isomorphisme n'est pas préservé entre le constituant prosodique et le constituant syntaxique. Dans les expériences initiales, ces difficultés du traitement sont observées chez les auditeurs dans les conditions où les structures prosodique et syntaxique se contredisent, le constituant prosodique regroupant des éléments appartenant à deux constituants syntaxiques. Or, dans la troisième expérience, les stimuli ayant la structure prosodique $(NP)_{SP}(VP+NP+PP)_{SP}$ et le syntagme prépositionnel étant rattaché au verbe, le traitement est facilité si une frontière prosodique supplémentaire apparaît devant PP. Pourtant, la présence de cette frontière n'est pas conditionnée par le souci de faire correspondre les structures prosodique et syntaxique des stimuli. On démontre également le rôle important que joue la structure de valences du verbe : dans cette perspective, l'étude met en évidence le rôle positif de la prosodie dans les cas où la structure argumentative des verbes entre en conflit avec le nombre de compléments retrouvés dans l'énoncé.

Une autre méthodologie expérimentale, que l'on retrouve dans les études de ce groupe, consiste à produire simultanément au stimulus auditif une perturbation : un clic ou dans le cas de l'écoute unilatérale changer d'oreille pour la présentation. On demande ensuite aux auditeurs de localiser cet événement interrupteur, l'hypothèse étant que les unités du traitement perceptif tendent à préserver leur unité. La tendance constatée est de localiser l'interruption aux frontières des constituants syntaxiques majeurs.

Il ressort de l'étude d'Avesani, Hirschberg et Prieto (1995) que les locuteurs d'italien, d'anglais et de l'espagnol s'appuient sur les indices tonals pour décider de l'interprétation visée des stimuli potentiellement ambigus ; pourtant, les stratégies adoptées par les sujets diffèrent en fonction de la langue de l'auditeur.

De manière générale, les études psycholinguistiques de ce groupe ne disposent pas de preuves suffisantes pour démontrer que les informations prosodiques imposent une analyse au profit de l'autre, mais leur rôle est important pour favoriser une interprétation plus appropriée. En référence avec la structure syntaxique, les informations prosodiques ont une double signification : toute frontière prosodique délimite un constituant (un constituant dans l'arbre est ainsi clôturé) ; en même temps les frontières déterminent les domaines (les éléments qui forment une unité) ce qui signifie le début du rattachement de ce constituant à l'arbre.

5.2.2.3. Dimension discursive et traitement perceptif

Ce groupe d'études exploite largement les relations entre les structures prosodiques et la structure informationnelle du message : en particulier, dans la lignée de l'approche « accentuelle » du focus (Gussenhoven 1984, entre autres), le constituant en focus est marqué par un accent focal (proéminence mélodique). Bien que la structure informationnelle du message langagier soit un phénomène universel, son marquage par

[172] Cf. « *choisir qch* » *versus* « *donner qch à qn* ».
[173] On distingue les verbes ditransitifs (« informer ») et monotransitifs (« choisir »).

la prosodie est spécifique à la langue, dans la mesure où certaines langues disposent de moyens morpho-syntaxiques (ordre des mots) pour l'actualiser.

Les tâches proposées aux auditeurs dans les études psycholinguistiques sont des tâches de détection d'un phonème ou d'une mispronunciation. Il a été démontré, en particulier pour le phénomène du focus, que les phonèmes initiaux d'un mot portant l'accent nucléaire sont traités plus rapidement (Cutler et Foss, 1977). Ce résultat pourrait être contesté, vu les observations associant une meilleure détection des phonèmes à clarté articulatoire élevée dans le contexte de la proéminence mélodique. Nous pouvons, peut-être, considérer que pour la gestion de la structure discursive, la prosodie n'y participe qu'indirectement. Par suite, dans une position prosodique forte la prononciation des phonèmes est plus claire, ce qui facilite le traitement perceptif.

L'étude de Cutler et Darwin (1981) a démontré la présence d'un même effet pour des stimuli avec une fréquence fondamentale monotone, ce qui a amené les auteurs à remettre en question les conclusions faites auparavant. Or, les études ultérieures ont mis en évidence la contrainte selon laquelle les différents indices prosodiques ne doivent pas être en contradiction pour faciliter le traitement perceptif de la structure discursive.

Une autre dimension exploitée dans les recherches psycholinguistiques est celle du lien entre la distribution des proéminences mélodiques et l'articulation « nouveau ~ connu » dans la dimension informationnelle. L'étude de Terken et Nooteboom (1987) démontre que le traitement perceptif est plus rapide si ce sont les mots nouveaux qui portent la proéminence mélodique. Pourtant ces résultats peuvent être interprétés toujours dans une perspective articulatoire, à travers la variation de l'intelligibilité entre la première désignation du référent et la répétition du mot (cf. Bard *et al.* 2000).

Citons également un ensemble de travaux sur le traitement perceptif de l'anaphore pronominale (cf. Auran 2004 pour un résumé).

5.3. RÉSUMÉ ET CONCLUSIONS

Ce chapitre nous a permis de présenter et de discuter les différents aspects de la matérialité de la constituance prosodique. Nous avons passé en revue les travaux exploitant les différents indices acoustiques qui semblent corroborer l'actualisation de la constituance prosodique à l'oral : cette analyse nous sera d'une grande utilité dans notre réflexion et dans l'étude expérimentale menée sur l'annotation automatique du signal en termes d'unités intonatives. Nous avons également présenté les études psycholinguistiques exploitant le rôle du phrasé prosodique en perception, études positionnées ainsi du point de vue de l'auditeur.

Les études discutées dans le présent chapitre concourent à démontrer que la structure prosodique représente un cadre de performance contraignant qui gère l'implémentation des paramètres acoustiques, cf. Shattuck-Hufnagel (2000) :

> *« This and other findings lend support to the hypothesis that prosodic structure provides the framework that integrates a variety of factors to determine the specific phonological and phonetic shape of a particular utterance, and that a*

traditional morphosyntactic structure of a sentence, including both syntactic and lexical specifications, are contributing factors to this prosodic structure rather than its sole and direct determiner. »

Or, notre étude portant sur l'oral spontané, la structure de surface des énoncés que l'on retrouve porte les marques du processus de planification de l'énoncé par le locuteur. En effet, la parole spontanée est caractérisée par une fluidité moindre, quand elle est comparée au style lu : de nombreuses dysfluences observées (hésitations, reprises, répétitions de mots, la liste n'étant pas exhaustive) traduisent des difficultés de planification sous-jacentes (travail de formulation, recherche de mots). Il existe un moment lors de l'intervention du locuteur où ce dernier est engagé à la fois dans la planification et la conceptualisation du discours à venir, des constituants tardifs de l'énoncé qu'il prononce déjà, il achève la construction des structures syntaxiques et prosodiques et est engagé à la prononciation des parties initiales de l'énoncé. Sous ces charges multiples, ses ressources attentionnelles se voient considérablement réduites. Au premier abord, il semble que de tels phénomènes soient entièrement sporadiques, or, une certaine régularité peut tout de même y être observée : une formalisation de ces régularités a été entreprise dans le cadre du modèle « commit & restore » (Clark & Wasow 1998). L'hypothèse clé de ce modèle consiste à stipuler que le locuteur s'impose la contrainte d'assurer une production non interrompue des constituants prosodiques ; il adopte également des stratégies de réparation conformément à cette contrainte. Il reste de savoir dans quelle mesure les dysfluences résultantes perturbent la structure prosodique et si elles possèdent des marqueurs formels propres.

Les études discutées dans le présent chapitre guideront notre travail de la modélisation acoustique du phrasé prosodique. Se pose cependant la question de savoir comment les hypothèses et les résultats des travaux fondés sur le cadre explicatif de la phonologie prosodiques ont été exploités dans des systèmes de synthèse vocale. Nous abordons cette problématique dans le chapitre suivant, après avoir discuté des principes généraux des systèmes « *text-to-speech* ».

CHAPITRE VI

LES MODÈLES PROSODIQUES EN SYNTHÈSE VOCALE

Généralement parlant, la synthèse vocale est un procédé permettant de produire du signal acoustique équivalant à la parole humaine à l'aide d'un ordinateur (il s'agit donc d'une branche de la linguistique computationnelle). Si, au début de son histoire, c'est le texte écrit qui était au centre de l'intérêt de la linguistique computationnelle (formelle, algorithmique), de nos jours, la recherche ainsi que les applications technologiques se sont réorientées vers le langage parlé. Les applications technologiques dans ce domaine sont multiples : d'une part, parmi les applications associées à la vie quotidienne nous pouvons citer premièrement les systèmes de renseignement automatique, les applications de la lecture oralisée (pour le courriel, par exemple), lecture avec des prototypes dramatiques (parole expressive), ainsi que les systèmes destinés aux personnes qui ont de la difficulté à parler. D'autre part, dans une optique plutôt en relation avec l'algorithmique interne, mentionnons les systèmes de reconnaissance vocale, et en particulier, de la parole émotionnelle, les systèmes de synthèse à partir du texte ou de la structure logique (*Concept-to-Speech Synthesis*), les systèmes dialogiques (*Spoken Dialogue Systems*), ensuite les systèmes de recherche et de traduction à commande vocale, ainsi que les applications pour l'identification du locuteur.

De ce panorama, nous voyons que le noyau de la majorité de ces applications est constitué par le couple « module de reconnaissance automatique // système de synthèse vocale ». Pourtant, malgré la longue histoire des réflexions sur l'éventualité de la parole artificielle, il reste encore des problèmes non résolus avec les méthodologies actuelles, les systèmes de reconnaissance produisant tout de même beaucoup d'erreurs et la parole synthétisée manquant du naturel. Nous sommes convaincue que c'est la prise en compte d'informations jusqu'à présent peu exploitées (un modèle plus approfondi de l'organisation prosodique, informations sur le registre mélodique, l'intensité, le débit de parole, les dysfluences…) qui permettra de remédier aux problèmes rencontrés et d'améliorer la performance des modules du traitement du langage naturel et de la synthèse vocale dans des applications technologiques.

Afin d'établir un panorama des algorithmes et des difficultés du traitement du matériau linguistique par les systèmes de synthèse vocale, nous proposons dans la suite de ce chapitre d'entreprendre une tentative de typologie des systèmes développés jusqu'à présent.

6.1. PREMIER AXE TYPOLOGIQUE : MÉTHODES DE GÉNÉRATION DU SIGNAL

À l'heure actuelle, il existe trois approches possibles de la génération du son dans les systèmes de la synthèse vocale :

- Premièrement, il est possible de se positionner du point de vue de la production de la parole et de l'articulation : ainsi les modèles de synthèse articulatoire essayent de modéliser les organes impliqués dans la production de chaque son de l'énoncé. Ces modèles se veulent reproduire la cause, et non pas l'effet (comme c'est le cas pour les modèles du groupe suivant). Remarquons pourtant que la tâche est très difficile, vu les difficultés en recueil de données sur les processus articulatoires mis en place en production de parole par les humains ainsi qu'au plan des calculs.

- Dans un deuxième temps, citons les travaux de synthèse orientés vers l'acoustique de l'onde sonore. Ces modèles se positionnent dans le cadre du modèle « source-filtre » et de la théorie acoustique de la production de parole (Fant 1960), avec la source sonore représentant l'activité des cordes vocales et une modélisation adjointe destinée à rendre compte des effets de résonance induits par le conduit vocal (modélisation des harmoniques associés avec la structure spectrale de chaque son). Les exemples de systèmes de ce groupe sont MITalk (Allen et al. 1987), YorkTalk (Coleman 1990) et PAT (Lawrence 1953)[174].

- La méthodologie actuellement la plus répandue est celle de la synthèse concaténative : dans cette approche, les briques de construction sont des fragments de parole pré-enregistrés de taille variable. Une fois joints pour former le contenu des énoncés, ces unités peuvent subir des modifications de leur organisation temporelle et tonale pour assurer des transitions plus harmonieuses entre les unités et répondre aux contraintes imposées par la structure prosodique de l'énoncé. Pourtant, toute modification provoquant une distorsion du signal initial, la qualité de la parole qui en résulte est inversement proportionnelle à la quantité de modification requise et appliquée.

Historiquement, les deux premières approches présentées ont une primauté sur les techniques de concaténation les plus répandues pourtant dans les systèmes commerciaux : cet état de choses est dû aux difficultés et à la complexité de la recherche fondamentale associée dans le cadre de la synthèse articulatoire et/ou acoustique (Klatt 1987).

6.1.1. Taille de l'unité dans les systèmes concaténatifs

Une question annexe porte sur la nature et la taille des unités utilisées en synthèse concaténative. Il n'existe pas de contrainte autre que la destination du système, sur le

[174] Les deux premiers types de systèmes sont parfois qualifiés de paramétriques, dans la mesure où les segments phonémiques y sont directement modélisés sous la forme d'états stables et de transitions.

type d'unités de base, ces unités pouvant correspondre à des énoncés entiers, phrases, mots ou encore à des unités plus petites, telles que les allophones ou les diphones.

La taille de l'unité dépend en grande partie de l'objectif final et de l'application visée pour le système développé[175] : pour une application ciblée dans un domaine restreint l'utilisation des phrases et des mots préenregistrés n'est pas un inconvénient majeur. On retrouve cette approche dans les différentes applications commercialisées (horloge parlante, système téléphonique de la banque, réservation d'un voyage *etc.*, Aust *et al.* 1995). Cependant, pour éviter tous les problèmes avec cette méthode, un contrôle très important est requis lors la phase d'enregistrement. En effet, si les mots sont enregistrés isolément, lors de la synthèse, toutes les caractéristiques acoustiques se voient modifiées. Très fréquemment alors on insère des pauses après chaque mot pour dépasser les effets néfastes ; or, cela nuit considérablement au naturel du message qui en résulte. Ainsi, dans le système d'aide à l'apprentissage d'une langue, présenté par de Pijper (1997) les mots sont-ils enregistrés dans des phrases porteuses, c'est-à-dire, dans un contexte, étant donné l'observation suivant laquelle les mots en contexte sont plus courts que ceux qui sont prononcés isolément.

Dans un deuxième temps, la prosodie d'un mot dans un énoncé varie en fonction de sa position dans le syntagme, de la structure syntaxique et du contexte discursif. Si tous les mots sont enregistrés dans un seul exemplaire et dans un contexte neutre, les variations contextuelles ne sont pas prises en considération et des modifications plus importantes du signal sont nécessaires comme c'est le cas dans le système de de Pijper (1997) : dans ce système, une fois le mot retrouvé, les durées et les valeurs de f0 associées aux segments se voient modifiées par la technique PSOLA, procédé qui peut avoir des effets néfastes pour la qualité globale du système. Par contre, dans le système d'annuaire téléphonique de Waterworth (1983, cité par Klabbers 2000) chaque chiffre est enregistré dans trois contextes prosodiques : neutre, continuatif et terminal (cf. pour le français les analyses du phrasé prosodique dans l'annonce des numéros de téléphone de Delais-Roussarie 1995).

Les deux procédés sont combinés lors du développement d'un système de renseignements sur le trafic ferroviaire en hollandais (Klabbers, 2000)[176]. Dans ce système les mots et les phrases sont enregistrés à la fois dans le contexte d'une phrase porteuse et dans six contextes prosodiques. Ces contextes prosodiques prennent en compte les réalisations des mots/syntagmes avec ou sans proéminence mélodique ainsi que le fait de la présence d'une frontière prosodique après le mot (les options étant absence d'une frontière, présence d'une continuation majeure/mineure, et présence d'une frontière conclusive). Dans une étude d'évaluation de la performance de ce système, pour la mesure de l'erreur de reconnaissance des mots, les auteurs obtiennent un score comparable à celui de la parole naturelle. Pour d'autres paramètres utilisés pour l'évaluation (intelligibilité, facilité de parole, qualité globale et convenance pour l'application visée) ce système se présente comme étant la meilleure alternative à la parole naturelle.

[175] On distingue les systèmes à applications ciblées *versus* les systèmes à applications ouvertes.
[176] Le procédé était également exploité dans le système GoalGetter (système permettant de générer les rapports oraux sur le déroulement des matchs de foot, Klabbers *et al.* 1996).

Les unités les plus fréquemment utilisées sont les diphones (terme proposé par Dixon et Maxey 1968). Il s'agit d'une unité qui commence au milieu (au niveau de la partie stable) du son *x* et se termine au milieu du son *y* suivant ; par conséquent, les effets de coarticulation ayant lieu à la jonction de deux segments ne doivent plus être modélisés explicitement, ce qui constitue la raison d'être de l'unité de diphone. On établit la taille de la base de diphones en relation avec les règles de phonotactique propres à la langue, chaque diphone étant codé en fonction de son contexte et en fonction de ses propriétés prosodiques (temporelles et tonales) pour assurer, si nécessaire, une modification de durée et de f0. La synthèse par diphones est très souple et la base d'unités pour une langue n'a pas de taille excessivement large (environ 2000 diphones). La qualité de sortie peut pourtant être très faible, et ceci pour deux raisons :

- premièrement, les caractéristiques d'un diphone varient en fonction du contexte segmental et accentuel dans lequel il a été réalisé (fréquemment, on travaille avec des logatomes). Cela peut avoir pour effet une discontinuité spectrale perceptible qui nuit à la qualité globale de la parole synthétisée. Pour remédier à ce problème, Skreline (2001) propose une unité plus petite ;

- deuxièmement, une distorsion peut ressurgir lors des modifications en fonction de l'organisation prosodique (temporelle et mélodique) de l'énoncé. Le remède proposé induit un changement de paradigme majeur : il ne s'agit plus de la construction d'une base de données stable et finie mais c'est une méthode de sélection des unités en temps réel qui entre en jeu (cf. *infra*).

Avec la méthode de sélection des unités, une autre rupture s'opère avec la méthodologie classique de la concaténation, à savoir : on refuse l'idée d'une unité uniforme, mais on parle de l'utilisation des unités de taille variable (cf. les travaux de Sagisaka : Sagisaka 988, Takeda *et al*. 1992). De nombreux systèmes[177] sont fondés sur l'application de cette technique. L'avantage de cette approche réside dans le fait que, éventuellement, aucune modification postérieure du signal ne serait nécessaire (par exemple, au niveau d'un mot, pour le faire entrer dans le contexte prosodique requis). Pourtant, lors du traitement du texte, on doit résoudre une tâche supplémentaire, à savoir, proposer une décomposition optimale de la séquence d'entrée en unités qui seront par la suite recherchées dans la base.

6.2. DEUXIÈME AXE TYPOLOGIQUE : PROPRIÉTÉS DE LA FORME D'ENTRÉE

Le deuxième axe typologique des systèmes de synthèse vocale prend en compte les caractéristiques de la forme d'entrée, à partir de laquelle est accomplie la génération du signal de parole. La représentation à l'entrée d'un synthétiseur peut être de plusieurs types :

- une représentation acoustico-phonétique ;
- un texte ;

[177] Exemples de systèmes : Festival (Taylor, Black & Caley 1998), rVoice (Rhetorical Systems Ltd. 2000), Laureate (Page & Breen 1996), AT&T NextGen (Syrdal, Wightman, Conkie, Stylianous, Beutnagel, Schroeter, Strom, Lee & Makashay 2000) and DEMOSTHeNES (Xydas & Kouroupetroglou 2001).

- un langage d'étiquetage spécifique ;
- une représentation sémantico-conceptuelle.

Pour discuter des différences entre ces systèmes, ainsi que des avantages et des inconvénients associés, nous proposons de nous référer au cycle de la production de parole par les humains détaillé par Levelt (1989). Dans ce modèle, la production d'un énoncé commence par la conception d'une idée dans le cerveau du locuteur ; cette idée est ensuite formalisée en termes d'énoncés dans la langue de communication, énoncés qui sont ensuite prononcés par le locuteur.

De manière générale, il convient de distinguer deux types majeurs de systèmes intégrant le composant de la génération du signal de parole : les systèmes traditionnels de synthèse à partir du texte (*Text-to-Speech,* TTS) et les systèmes de synthèse intégrant la génération du message (*Concept-to-Speech*, CTS). C'est le fait de disposer du texte à l'entrée qui distingue les deux approches : dans un système de CTS le texte est généré par le système lui-même, et n'est pas disponible à l'entrée (dans sa forme finale) comme c'est le cas dans les systèmes TTS. Les systèmes CTS sont des systèmes de génération du message linguistique : le module de la génération de parole y est couplé avec le module de la génération du langage naturel ; par conséquent, les informations syntaxiques, sémantiques et pragmatiques sont disponibles directement dans le système et ne doivent pas être déduites à partir du texte. En revanche, les systèmes TTS classiques travaillent à partir du texte qui leur est imposé et procèdent tout d'abord à son analyse et à son annotation afin d'obtenir les étiquettes désignant les événements prosodiques avant de passer au stade de génération du signal acoustique proprement dite. Par référence au modèle de Levelt, les deux systèmes interviennent à des stades différents, les étapes parcourues par un système CTS étant plus en accord avec l'activité cognitive d'un être humain lors de la production de parole (Dik 1997).

De cette première différence émane celle du domaine couvert par deux types d'approches : les systèmes CTS sont orientés vers l'accomplissement d'une tâche concrète dans un domaine précisément cerné et restreint, tandis que les systèmes TTS ont pour vocation de couvrir n'importe quel texte.

Les distinctions entre les deux classes de systèmes mentionnées ci-dessus ont une portée directe sur certains choix linguistiques implémentés. Les systèmes TTS sont généralement orientés vers le style de la parole lue. Pourtant, même dans la tâche de lecture, leur performance est loin d'être similaire à celle d'un humain par manque de compréhension de la sémantique des mots et/ou du rôle discursif d'un constituant. De leur côté, les systèmes CTS ont pour modèle la parole spontanée (pourtant, contrairement aux humains, ils sont contraints à un seul domaine).

Afin de résumer les points sur lesquels l'amélioration de la génération de l'organisation prosodique est attendue dans les deux types de systèmes, nous procédons dans la suite de ce paragraphe à une ébauche succincte des deux approches.

6.2.1. Systèmes TTS : Text-to-Speech

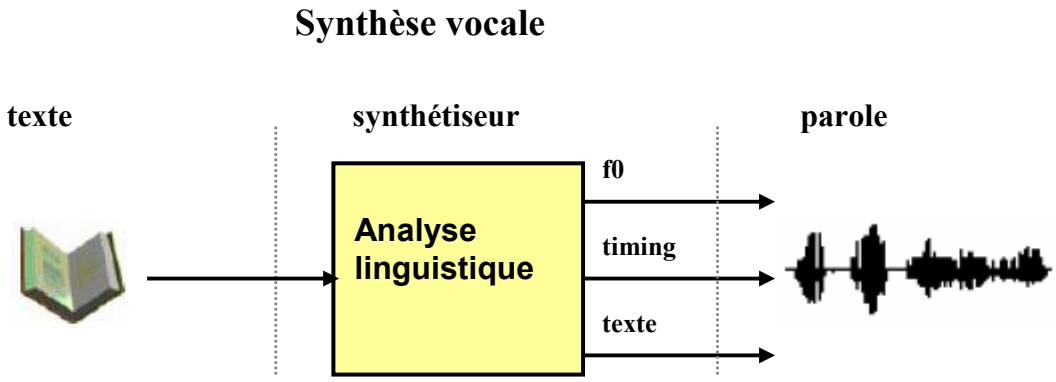


Figure 6.1 : Schéma d'un système de synthèse vocale prototype

Dans l'approche de synthèse à partir du texte (TTS classique, figure 6.1), le système dispose à l'entrée d'un texte écrit et procède au calcul de la forme prosodique à la base des informations linguistiques qui peuvent être récupérées à partir de la forme de surface. De fait, pour le phrasé prosodique on prend en compte des informations sur la ponctuation, l'organisation du matériau linguistique en paragraphes et l'annotation en termes des parties de discours. Simultanément, cette dernière annotation est prise en compte (très fréquemment de manière exclusive) dans le calcul de la distribution des proéminences mélodiques. Par conséquent, un algorithme le plus simple possible assignera la proéminence mélodique (modélisée comme un accent mélodique montant) à tous les mots lexicaux, les mots grammaticaux restant inaccentués, et le phrasé sera guidé en exclusivité par la ponctuation. Un système plus complexe fera appel au parseur syntaxique lors de l'accomplissement de deux tâches (cf. *infra*).

On cherche donc le meilleur modèle qui permettra de mettre en relation de manière efficace des informations linguistiques présentes dans la forme syntaxique de surface et l'organisation prosodique. Or, les informations à récupérer sont multiples, cf. van Santen & Sproat (1998 :1) :

> « *For a system to sound natural, the system has to have real-world knowledge ; know rules and exceptions of the language ; appropriately convey emotive states ; and accurately reproduce the acoustic correlates of intricate motor processes and turbulence phenomena involving the vocal cord, jaws, and tongue. What is known about these processes and phenomena is extremely incomplete […]* »

Outre une analyse syntaxique approfondie, il est donc indispensable de développer des algorithmes performants permettant de récupérer, à partir du texte, des informations sur la structure discursive et les valeurs pragmatiques. Or, l'appréhension de l'interface « prosodie ~ discours » n'est pas exhaustive de nos jours. Il n'est donc pas étonnant de retrouver dans les rapports de recherche sur cette problématique le concept d'organisation prosodique neutre (cf. Bachenko et Fitzpatrick 1990), indépendante du contexte discursif. Il s'ensuit donc qu'une étape préalable est nécessaire pour générer de la prosodie naturelle, à savoir, une prise en compte systématique et approfondie des associations entre la structure linguistique du texte et son organisation prosodique.

Les méthodes de corpus, remplaçant les heuristiques minutieusement ébauchées par les chercheurs, permettent d'acquérir de manière implicite des connaissances sur les multiples interfaces entre différents niveaux du langage. Les algorithmes de l'apprentissage automatique permettent au système d'acquérir des connaissances sur les fréquences (probabilités) de l'association d'une organisation prosodique particulière avec tout mot du lexique (Sproat et al. 1992, Ross & Ostendorf 1996). Une fois ces informations apprises par le système, un deuxième groupe d'algorithmes est appliqué, lesquels ont pour tâche de décider de la meilleure forme prosodique pour le texte d'entrée : on choisit donc la meilleure prononciation ou la meilleure prosodie. De manière implicite, ces méthodes permettent au système d'acquérir des connaissances sur les relations sémantiques entre les éléments ainsi que sur les contraintes discursives.

Outre le traitement robuste de la sémantique au niveau de phrase, les systèmes TTS classiques manquent d'un modèle de l'organisation discursive : le défaut majeur des algorithmes actuels réside dans leur incapacité à déduire à partir du texte les informations pragmatico-discursives, c'est-à-dire une variabilité dépendante du contexte d'élocution. Nous pensons, en particulier, au lien qui existe entre la distribution des accents mélodiques et l'organisation informationnelle du message, subdivisée en deux axes « fond – focus » et « connu – nouveau »[178]. Ainsi un concept déductible de la conversation précédente[179] ne sera pas accentué ; or, avec la progression du discours, et l'épuisement progressif des ressources attentionnelles des interlocuteurs, le même mot peut recevoir de nouveau une proéminence mélodique. Pour remédier à ce problème, Hirschberg (1990, 1993) propose la méthodologie suivante : on stocke tous les mots lexicaux dans un vecteur (ou fichier) pour procéder ensuite à une comparaison du texte d'entrée afin de déduire si un mot est connu ou nouveau dans le contexte du discours ; ce vecteur est réinitialisé à la fin de chaque paragraphe. Si un mot a été précédemment stocké dans le vecteur « discursif », on considère que ce mot, ou tout mot ayant la même racine, ne sera pas accentué au fil du paragraphe. Ce modèle a été implémenté dans le système de synthèse (Black et Taylor 1994).

Un modèle similaire a été proposé pour le système de synthèse en suédois par M. Horne et ses collègues (Horne *et al.* 1993) avec la différence que les auteurs ne se réfèrent pas au paragraphe pour modéliser les ressources attentionnelles du locuteur, mais utilisent une fenêtre dynamique de 60 mots. Par ailleurs, le système se réfère également en temps réel au thésaurus du suédois, pour assurer la désaccentuation non seulement des mots qui sont une répétition directe du mot ou de sa racine, mais également des mots désignant des concepts proches, liés par les relations de synonymie ou d'hyponymie.

Pour notre part, une maîtrise plus fine du détail phonétique (Local 2003) est nécessaire dans la modélisation de différentes interfaces.

Cahn (1999) propose de distinguer deux types d'approches de l'organisation prosodique du message : les approches de compétence, qui n'associent la prosodie du texte qu'à sa structure, et les modèles de performance, qui mettent l'accent sur l'implication du

[178] Nous nous positionnons dans la perspective de l'approche « focus-to-accent » (Gussenhoven 1983).
[179] Il peut s'agir de la répétition du mot lui-même, d'un synonyme, d'un mot appartenant au même champ lexical, d'un procédé anaphorique.

locuteur dans la production de parole. Néanmoins, tout l'effort d'amélioration de la performance des systèmes TTS s'est concentré sur l'exploitation de l'axe de compétence linguistique, c'est-à-dire, sur une modélisation du lien entre le texte et la prosodie. Par contre, l'idée de performance est étroitement associée aux les concepts de style vocal et d'idiolecte. Dans cette optique, un ensemble des travaux en synthèse vocale est consacré à la problématique du traitement de la variabilité observée dans les productions langagières chez un même locuteur ou de celle qui existe entre les différents styles de parole.

On constate entre autres que l'effort s'est concentré sur la modélisation de la parole spontanée par opposition au style lu. Les techniques mises en œuvre diffèrent tout de même selon que c'est le processus cognitif sous-jacent ou l'effet acoustique qui est modélisé. Ainsi la modélisation de Johnson (1996) est-elle fondée sur les différences en empan attentionnel entre ces deux styles de parole : on considère que la parole spontanée est caractérisée par une planification réduite, et c'est une des propriétés qui la différencie de la lecture. Cette différence en empan attentionnel affecte l'organisation prosodique du message : en lecture, le locuteur s'applique plus à la réalisation des séquences de tons abaissés (*downstepped*), les contours descendants terminaux y sont aussi plus raides. Ce sont ces effets qui se voient modélisés en synthèse.

Dans une autre étude, Abe (1996) parvient à modéliser à l'aide de méthodes statistiques les différences acoustiques (prosodiques) entre trois styles de lecture en japonais : style publicitaire, lecture d'un texte littéraire et lecture d'un texte tiré d'une encyclopédie. L'analyse du corpus effectuée au préalable met en évidence les différences dans les valeurs des formants des voyelles, dans les spectres d'énergie, ainsi que dans l'organisation temporelle et mélodique du message. Par la suite, les indices retrouvés sont implémentés dans un système TTS sous la forme de règles ; la deuxième manipulation effectuée est la conversion du style de lecture encyclopédique en deux autres styles étudiés. Les tests de perception effectués pour évaluer la performance du système mettent en évidence l'existence des distinctions systématiques au moins entre le style publicitaire et le style de lecture littéraire.

L'attention à la variabilité inter- et intra-individuelle est également présente dans les travaux sur la synthèse de la parole émotionnelle ou expressive. En effet, nous avons mentionné la position adoptée par une partie des chercheurs, consistant à générer une prosodie neutre du point de vue discursif (Bachenko et Fitzpatrick 1990) ; de la même manière la prosodie des systèmes appliqués actuels se révèle des plus neutres et « plates ». Van Santen *et al.* (2003) discutent du rôle de la parole expressive en tant qu'outil diagnostique et thérapeutique dans le cas des pathologies de parole chez l'enfant, et y voient une possibilité d'application pour la synthèse expressive. Comme l'approche thérapeutique réside dans l'utilisation d'une prosodie exagérée[180], la prosodie expressive doit être visée dans les applications destinées à ces enfants. La solution technique proposée pour la génération de la parole expressive s'appuie sur la méthode des quadruples (transplantation des schèmes mélodiques typés). Remarquons pourtant que la difficulté majeure dans la modélisation de la parole expressive est liée à

[180] Cf. le style de la parole maternelle et son rôle pour l'établissement d'un cadre des interactions sociales et conversationnelles.

une confluence dans la parole de ce qui est linguistique et de ce qui est paralinguistique/endolinguistique.

À leur tour, Shih & Kochansky (2001) démontrent les possibilités de synthétiser le chant rythmique de Dinah Shore ainsi que le discours rhétorique de Dr. Martin Luther King en appliquant leur modèle de description et de modélisation de l'intonation Stem-ML. Comme dans ce modèle les étiquettes reçoivent une définition mathématique, un algorithme leur est ensuite associé pour procéder à une transposition de l'annotation symbolique en une prosodie quantitative. Tout se joue donc à l'étape de la sélection des étiquettes à prendre en compte ; ce sont ces balises issues de l'analyse préalable du corpus qui permettent de capter la prévalence dans le discours de Dr. Martin Luther King de la composante syntagmatique globale sur la composante accentuelle (en l'occurrence, il s'agit de la balise *step_to*). Nous voyons donc que, dans l'approche proposée, un style (un idiolecte) est représenté à l'aide des traits caractéristiques, formalisés sous forme de balises et de règles permettant de localiser les traits repérés : cette approche permet de traduire l'impression d'un style particulier à travers les traits les plus saillants qui lui sont associés.

Un des objectifs du projet suédois GROG (2002) est d'acquérir des connaissances sur l'organisation prosodique de la parole conversationnelle par référence aux phénomènes de frontières et de groupements pour ensuite appliquer les procédés attestés en synthèse. Remarquons que les données ont été analysées en relation avec le modèle d'engagement du locuteur (*commit & restore model*, proposé par Clark et Wasow 1998) ; les auteurs associent donc les particularités de l'organisation prosodique du message avec les mécanismes cognitifs de la planification du message et l'objectif communicatif de toute interaction verbale. Le texte d'entrée contient l'étiquetage des frontières prosodiques fortes et faibles marquées par les experts (et confirmées dans des expériences en perception (Carlson *et al.* 2004) ainsi que l'annotation de la structure informative en termes de distribution du focus. La modélisation acoustique des frontières fortes et faibles est réalisée par le même algorithme que pour les marques de ponctuation « . » et « , » en lecture. On utilise également l'étiquetage des dysfluences : il s'agit d'une frontière faible, réalisée essentiellement avec un allongement considérable du mot entier et de sa rime. D'autres indices acoustiques[181] modélisés sont un léger abaissement tonal et un bref silence après le mot concerné. Les procédés mis en œuvre ont permis aux auteurs de mimer la production du locuteur et d'augmenter le naturel de la parole résultante.

Sagisaka *et al.* (2005) indiquent comme nouvel objectif pour les algorithmes de synthèse de la parole d'aboutir à une reproduction du style conversationnel. Il est vrai que la parole conversationnelle humaine est régie par plusieurs facteurs, dont les influences ne peuvent pas toujours être séparées ainsi qu'il n'est pas anodin de prévoir toutes les sources de la variabilité. Les auteurs proposent donc de se servir de la sémantique des mots dans la forme de sortie pour avoir une vision plus pointue de leur prosodie. En particulier, la méthode mise en œuvre dans cette étude cherche à établir

[181] Et nous considérons comme important d'attirer l'attention sur cette modélisation complexe des indices acoustiques : d'autres études ont essayé de mimer l'effet d'une pause d'hésitation seulement en en insérant une à l'intérieur du syntagme intonatif.

une correspondance entre les attributs sémantiques des mots du texte et leur prosodie. L'étude est réalisée à partir des structures comportant un adjectif et un adverbe, les adjectifs exprimant l'attitude du locuteur et les adverbes quantifiant des différents degrés de l'attitude exprimée. L'analyse acoustique du corpus permet d'établir les corrélations entre les attributs sémantiques des mots et les valeurs de f0 ; ces résultats sont ensuite confirmés dans des tests de perception. À la dernière étape, les données recueillies sont intégrées dans un synthétiseur qui applique en l'occurrence le modèle de Fujisaki pour la génération de la courbe intonative : on modifie l'amplitude de l'impulse accentuel en relation avec le degré de la caractéristique exprimée par l'adverbe.

Nous voyons donc que la majorité des études se penchant sur les phénomènes de performance et y cherchant les ressources pour l'amélioration de la qualité de la prosodie synthétisée emprunte la voie selon laquelle on essaie de reproduire l'effet observé, et non pas de modéliser le processus cognitif derrière le phénomène observé. Pourtant, nous tenons à faire le parallèle entre cet état de l'art et ce que van Santen et Sproat (1998) remarquent à propos du rôle de l'approche concaténative en synthèse : effectivement, il s'agit de la technologie la plus répandue, mais ceci pour des raisons pratiques plutôt que théoriques. Réfléchissant à l'évolution future des systèmes, van Santen et Sproat remarquent que la compréhension complète du mécanisme de la production de la parole n'est possible que dans le cadre d'un synthétiseur articulatoire :

> *« however, we feel that the complex forms of coarticulation found in human speech ultimately can only be mimicked by accurate articulatory models […]. »* (*op. cit.* : 2)

Nous pensons que, dans la même lignée, pour ce qui concerne les effets de performance, les contraintes externes imposent aux chercheurs le choix de travailler en mimant l'effet. Or, la recherche future permettra d'en dégager la cause.

6.2.2. Systèmes de synthèse intégrant le module de la génération du dialogue (Concept-to-Speech)

La figure 6.2 présente les modules usuellement inclus dans un système Concept-to-Speech. Cette approche étant très fréquemment implémentée dans les systèmes dialogiques, le module de la génération de la parole s'y voit couplé avec un module de reconnaissance vocale. À la sortie du module de reconnaissance on dispose d'un ensemble d'hypothèses sur les mots prononcés, cette variabilité étant due aux ambiguïtés linguistiques ainsi qu'aux erreurs dans l'analyse acoustico-phonétique. C'est le module du traitement automatique du langage naturel[182] qui prend la décision et choisit, à partir des informations sémantiques, syntaxiques et pragmatiques, une seule séquence de mots pour l'énoncé analysé. Lors de cette tâche, chaque composant linguistique apporte sa contribution : les informations syntaxiques permettent de

[182] Remarquons, qu'à l'heure actuelle, il existe deux approches possibles au traitement du langage naturel : le premier est fondé sur la modélisation explicite des connaissances linguistiques acquises ; le deuxième est orienté vers l'utilisation des grands corpus authentiques : à partir de ces corpus, une base de données sur les arbres (=structures) syntaxiques est obtenue ; le corpus peut couvrir le domaine restreint pour lequel le système de CTS est conçu.

déterminer si le mot en question est probable dans la position syntaxique correspondante, vu le contexte. Les informations sémantiques, quant à elles, permettent de déterminer si l'énoncé final est doté de sens ; ensuite les informations pragmatiques déterminent la plausibilité de l'énoncé analysé dans le cadre de l'échange discursif effectué. On dispose à la sortie d'une représentation abstraite de l'énoncé, qui est ensuite interprétée par le module de gestion du dialogue ; cette étape termine le cycle de reconnaissance de parole. Sans entrer dans les détails du fonctionnement de ce module, nous ferons tout de même référence aux travaux de Mary Ostendorf (2000, 2003), qui insiste à plusieurs reprises sur le rôle majeur de l'organisation prosodique dans l'amélioration de la performance des systèmes de reconnaissance vocale.

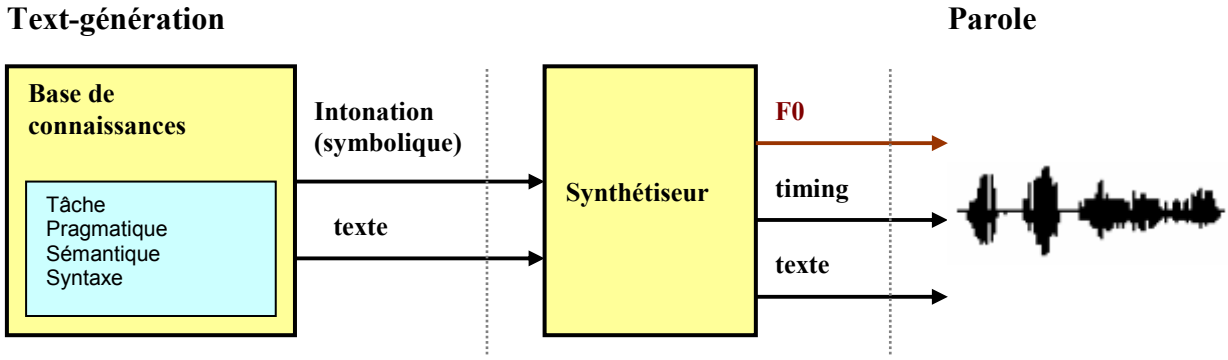


Figure 6.2 : Schéma d'un système CTS prototype

Dans l'étape suivante, le gestionnaire du dialogue détermine quelle réponse du système sera la plus adaptée à la situation, vu l'état informationnel mutuel atteint par les participants. Trois possibilités sont à prévoir : étant donné l'état informationnel, le gestionnaire du dialogue peut a) communiquer les informations demandées, que l'on retrouve dans la base de données ; b) demander des précisions, des détails supplémentaires nécessaires ; ou c) procéder à la vérification des informations reçues. La génération du message que le système produira en réponse commence donc par une représentation abstraite du contenu.

Conséquemment, le module de génération du langage naturel a pour tâche de convertir cette représentation abstraite (acte de parole) en une représentation enrichie : le texte doté d'étiquettes signalant la distribution des proéminences mélodiques et des frontières prosodiques[183]; nous rejoignons ici le protocole décrit pour les systèmes TTS classiques en ce qui concerne le traitement de la représentation symbolique par le module de génération de l'intonation. L'avantage de cette structure pour la génération de la parole réside dans le fait qu'avec les informations syntaxiques, sémantiques et pragmatiques disponibles, le calcul du phrasé prosodique et de la distribution d'accents est plus fiable, si on le compare avec le système de TTS classique. Au stade terminal, le générateur de parole convertit donc la représentation étiquetée en signal de parole.

Le système conçu par House et Youd (1992) est fondé sur une exploitation très large des informations linguistiques afin d'aboutir à une prosodie appropriée au contexte :

[183] Nous omettons les étapes intermédiaires du traitement.

dans le calcul des traits syllabiques ([focus], [emphase], [proéminence] ; ces informations étant utilisées ensuite pour accéder aux étiquettes des événements prosodiques) ce système utilise les informations sur le type d'acte de parole, le contenu sémantique et la forme de surface. Citons également le travail de Prevost et Steedman (1994) dans lequel la grammaire catégorielle est utilisée pour assurer la distinction entre le thème et le rhème, traitement qui a des conséquences sur le calcul de la distribution des proéminences mélodiques.

Ainsi les systèmes CTS ne doivent-ils pas accomplir la tâche de retrouver la structure conceptuelle et discursive du message à partir du seul texte. Nous voyons donc que les modules de génération de parole dans les systèmes CTS disposent des informations sur l'organisation du discours dont ils peuvent profiter pour le traitement de l'organisation prosodique du message. Dans ces systèmes, il est envisageable donc de transmettre une intention propositionnelle et communicative, de déterminer le rôle thématique de chaque élément/constituant, la distribution du focus, les nuances dans la sémantique des mots, les effets de contraste. La performance de cette approche dépendra par conséquent de l'adéquation des analyses et des modélisations qui relient les propriétés prosodiques (les indices acoustiques) et la structure informationnelle, propositionnelle et attentionnelle[184]. Par conséquent, les connaissances linguistiques qu'il importe d'acquérir afin d'améliorer la performance de ce type de systèmes touchent à un domaine très controversé, à savoir la mise en relation du discours et de la prosodie (cf. Auran 2004). Remarquons également que la base des connaissances spécifiques au domaine d'application est close dans ces systèmes et élaborée à l'avance.

Pour les systèmes CTS, dans la plupart des cas, il s'agit d'une application concrète destinée à des usagers multiples. Une bonne qualité de la parole synthétisée est cruciale dans les systèmes appliqués CTS-TTS dans la mesure où l'utilisateur du système y est confronté. On se doit donc de produire un signal de parole compréhensible, mais également de ne pas provoquer chez l'auditeur des sentiments de rejet.

6.2.3. Le rôle des corpus dans la synthèse de la parole

Dans les systèmes de synthèse vocale, on fait référence aux corpus oraux annotés dans deux cas de figure : premièrement, ils constituent une source d'unités primaires dans le cadre de la synthèse concaténative qui applique le paradigme de la sélection d'unités ; dans un deuxième temps, c'est à partir d'un corpus authentique que s'opère l'apprentissage automatique des associations entre les événements linguistiques de différents niveaux et le texte, étape préalable à la génération de l'intonation et de la prosodie. Nous nous procédons donc à la discussion de ces deux points dans ce qui suit.

[184] Remarquons également la tâche coûteuse en temps, laquelle consiste à rassembler les données sur le domaine spécifique pour lequel l'application est destinée.

6.2.3.1. Technique de sélection des unités

Il découle des analyses présentées précédemment, que dans le cadre d'une approche concaténative, une fois le traitement linguistique du texte achevé, deux tâches doivent être effectuées pour générer le signal de parole : d'abord, il faut choisir les unités à concaténer au niveau segmental, ensuite, y appliquer un traitement ultérieur (méthodes LPC ou PSOLA) pour aboutir à une organisation prosodique en accord avec le contenu sémantico-pragmatique du message. Les bases de diphones développées pour différentes langues contiennent environ 2000 unités codées[185]. De manière générale, les diphones ont été extraits des logatomes (non-mots) prononcés dans une phrase-porteuse. Pourtant, la question de savoir quel contexte convient le mieux pour la constitution de l'inventaire n'est pas tranchée, vu le besoin de diminuer le coût de concaténation et de modifications postérieures.

Au début des années 1990, une technique alternative a été proposée, à savoir celle de la sélection des unités (cf. Takeda *et al.* 1992). Dans cette approche, les unités sont sélectionnées de manière automatique et en temps réel à partir d'un large corpus oral annoté (segmenté et transcrit phonétiquement). La sélection est guidée par les procédés d'évaluation du coût de la concaténation : ces mesures reflètent l'adéquation du segment analysé au contexte requis, la continuité des changements spectraux à l'intérieur et entre les unités. Ainsi, vu la logique interne du paradigme de sélection d'unités, son application est guidée par trois problématiques :

- les procédures de la sélection ;
- les mesures objectives, critères de sélection ;
- la réflexion sur le type de corpus à utiliser.

La mise en œuvre de tous ces procédés permet de sélectionner le meilleur candidat pour la séquence d'entrée. L'avantage majeur de cette approche réside dans le fait qu'on dispose de plus d'un exemplaire (*token*) pour le segment en question. Par ailleurs, les unités de différentes tailles issues des contextes différents (et donc avec une variabilité inhérente) peuvent être utilisées. Remarquons pourtant que dans le cadre de l'heuristique classique de la sélection des unités, la prosodie et les durées des segments subissent tout de même des modifications *a posteriori*, selon des règles implémentées dans le système.

Une position extrême de la méthodologie est défendue par Black et Taylor (1994). En effet, les auteurs présentent le système CHATR et adoptent la stratégie de n'imposer au signal acoustique aucune modification postérieure après la concaténation des unités segmentales choisies dans la base de données. Néanmoins, les spécifications du contexte prosodique sont prises en compte à l'étape de la sélection des unités. Les auteurs considèrent que l'impact des propriétés prosodiques adéquates, dont ils refusent l'implémentation directe, sera compensé par une bonne qualité de la parole au niveau segmental.

[185] Citons le travail de Chalonova (1999), recherche menée dans l'objectif de diminuer l'inventaire des unités. Or, avec le progrès technique en ce qui concerne la vitesse de calcul et les capacités de stockage, cela n'est plus d'actualité.

L'unité de sélection de base était souvent un segment ou un diphone, on peut se demander dans quelle mesure ce choix réduit le nombre des sites de concaténation. En effet, disposant de plusieurs candidats pour une même séquence de sons, il est possible de choisir celui qui impose un coût de concaténation moindre. Ainsi dans le système proposé par Takeda *et al.* (1992), la fonction du coût de concaténation prend en compte les observations issues des analyses acoustiques et perceptives : par exemple, on considère que les consonnes fricatives sourdes représentent un meilleur contexte pour la concaténation que les consonnes fricatives voisées, ce qui est reflété dans les coûts de concaténation attribués aux unités respectives. Les algorithmes de sélection cherchent à minimiser deux valeurs de coûts : le coût de la distorsion de l'unité trouvée (*target cost*) et le coût de la continuité (*concatenation cost,* termes de Black et Hunt 1996) (Möbius 2000). Dans le premier cas, il s'agit d'une mesure de différence entre l'unité trouvée dans le corpus et l'unité requise par le texte d'entrée. Le calcul de cette mesure se fonde sur l'annotation du corpus de base et les paramètres établis lors du traitement linguistique préalable du texte d'entrée. La deuxième mesure évalue les discontinuités entre deux unités adjacentes sélectionnées au point de concaténation, et peut donc prendre en compte toutes les informations disponibles (entre autres, les mesures acoustiques). Un des axes de recherche associés porte sur le choix et la pondération des paramètres acoustiques à prendre en considération lors de l'évaluation des coûts. Or, la question de la corrélation entre les mesures de distances objectives issues des calculs que nous venons d'ébaucher et l'impact des propriétés acoustiques du signal résultant en perception, se pose.

Une méthode de prétraitement de la base des données acoustiques afin de remédier à la question délicate du choix de l'unité a été proposée par la suite : il s'agit de procéder au préalable à une formation des agrégations des contextes prosodiques et phonologiques. On établit les classes d'équivalence (agrégations) pour les réalisations d'un même segment observées dans la production du locuteur de référence. La procédure est automatique et utilise la technique des arbres de décisions dont le critère est de maximiser la ressemblance acoustique entre les unités à l'intérieur de la classe. Cette méthode est présentée et testée dans l'étude de Black et Taylor (1997). On construit un arbre de décisions en fonction des questions posées à chaque étape et les feuilles de cet arbre sont des listes de segments dans la base initiale qui correspondent le mieux aux contextes décrits par les questions. Lors de la synthèse, pour chaque segment, on cherche la liste des meilleurs candidats en se fondant sur les arbres établis. Lors de la sélection du candidat à partir de cette liste, on prend en considération le coût de la concaténation ainsi que la distance entre chaque unité de la liste et le centre de l'agrégation : ce procédé permet de se passer de la mesure du coût de distorsion de l'unité. Le calcul de différentes mesures d'évaluation des candidats est effectué d'avance, ce qui constitue un des avantages de cette méthode. Le temps de la sélection se voit par conséquent réduit, car on n'analyse qu'une seule classe d'éléments la plus appropriée aux spécifications d'entrée.

Dans la réflexion sur le rôle des représentations symboliques un autre groupe d'approches s'est forgé, approches qui prennent en considération la structure phonologique de l'énoncé. L'idée principale de ces méthodes réside dans la conviction que la majorité de la variabilité dans la parole est prédictible et que les unités

sélectionnées dans des contextes appropriés possèderont des spécifications requises par la forme d'entrée. Cette méthodologie est exploitée dans l'étude de Taylor et Black (1999) : chaque énoncé de leur corpus est annoté en un arbre phonologique comportant des niveaux entre l'énoncé et le phone ; il s'agit également d'un arbre métrique (les nœuds comportant les étiquettes « fort » et « faible »). La sélection des unités s'opère à partir des informations sur la prononciation canonique, l'accentuation et le phrasé prosodique (et non pas à partir de la transcription phonétique, ni des valeurs quantitatives de durée et de f0) ; tout nœud de l'arbre est donc annoté en attributs qui lui sont propres. À l'étape de la recherche du candidat, l'algorithme récursif (on commence par le nœud le plus haut) procède par comparaison de la structure arborescente produite pour l'énoncé à synthétiser[186] et des structures dans la base source : on détermine si la structure au-dessous du nœud est la même pour deux arbres (une correspondance stricte est demandée). Tout candidat trouvé domine un arbre de la même structure ; la sélection s'opère donc à la base des considérations sur les propriétés prosodiques (accentuation et intonation) et contextuelles (position à l'intérieur d'un syntagme, contexte phonétique). On procède ensuite à une concaténation des candidats et éventuellement à des manipulations de leurs propriétés tonales et temporelles. Möbius (2000) signale tout de même un des points faibles de ce modèle : ses auteurs considèrent les frontières des mots comme étant des frontières robustes, or, le fait de considérer une frontière de mots comme une barrière à la coarticulation reste constroversé (cf. chapitre IV).

L'étude de Klabbers *et al.* (2002) analyse plusieurs schèmes d'annotation des corpus. La réflexion des auteurs commence par la considération que le schéma d'annotation au niveau de syllabe avec des facteurs [± accentué], [± proéminent], [± final dans un syntagme] n'est pas suffisant[187], et un schème plus élaboré est nécessaire. En particulier, les auteurs insistent sur l'inclusion dans le schéma d'annotation d'un constituant similaire au pied métrique trochaïque (avec tête à gauche) défini pourtant par référence à la proéminence mélodique. La comparaison de quatre schémas de factorisation témoigne des avantages du schéma avec les pieds par référence au schéma simple.

Il ressort également de l'analyse présentée que les critères prosodiques sont explicitement utilisés[188] lors de la sélection des unités ; il est également considéré comme plus approprié et avantageux pour la performance du système d'inclure les spécifications prosodiques symboliques (en termes de systèmes d'annotation que nous avons discutés dans le chapitre II) et non pas quantitatives.

6.2.3.2. Corpora et méthodes d'apprentissage

Intéressons-nous maintenant à l'appréciation du rôle des méthodes d'apprentissage automatique des structures linguistiques à partir des corpus. L'émergence de telles

[186] Les algorithmes du traitement du texte ont plus de facilités à produire la structure phonologique que les spécifications phonétiques.
[187] Remarquons au passage que ce schéma d'annotation est très fréquemment utilisé par les méthodes d'apprentissage automatique.
[188] Sous forme de paramètres et de questions lors d'élaboration des arbres de décisions : on s'interroge, par exemple, sur le fait de savoir si le segment appartient à une syllabe accentuée/proéminente ou s'il se trouve en début du syntagme etc.

méthodes, le développement de l'appareil mathématique associé, constitue la démarche privilégiée pendant les quinze dernières années dans le domaine des technologies de parole. En effet, dans son article synthétique de l'état de l'art dans le domaine de la synthèse vocale, Klatt (1987) présente et analyse essentiellement les systèmes agissant par application des règles formulées et formalisées par les linguistes. Or, le changement majeur qui a eu lieu depuis est celui de la création et du développement des grandes bases de données orales.

Le procédé central du module de traitement du texte dans un système TTS classique consiste à produire une sortie quantitative (valeur de durée, de f0) ou qualitative (information sur la présence d'une proéminence ou d'une frontière) à partir du texte dont tout élément est doté d'un certain nombre d'attributs discrets[189]. Pour déterminer les facteurs les plus étroitement associés au phénomène en question on peut procéder soit par élaboration manuelle des heuristiques (méthode par règles) soit par application au texte annoté des méthodes d'apprentissage automatique.

En référence à l'organisation prosodique du texte, notons que conventionnellement, le phrasé prosodique et la distribution de proéminences mélodiques sont gérés séparément. Les étiquettes prosodiques sont alors associées avec des propriétés linguistiques évidentes du texte d'entrée : distance par rapport au début/fin de phrase (clause, énoncé), annotation en parties de discours, informations syntaxiques (plus ou moins précises, cf. *infra*), ponctuation. Les régularités relevées (= règles apprises) seront ensuite appliquées lors de l'annotation d'un texte vierge.

Il importe de préciser que la perception des techniques statistiques par la communauté scientifique n'est pas homogène. Fréquemment, à l'opposition « méthodes par règles versus méthodes stochastiques » on associe celle des approches d'ingénieurs *versus* celles des linguistes. Il est évident que les méthodes d'apprentissage automatique nécessitent moins d'effort humain, cf. Traber (1995:196) :

> *« The best f0 generator for the SVOX system was realized by using a recurrent neural network as f0 predictor. This method required only a limited human effort in order to obtain an f0 generator of high quality. It was much more important to know what phonological properties control the phonetic realisation of intonation that how these properties are realised quantitatively. The qualitative part of the task of building an f0 predictor was therefore based on well-known linguistic (especially phonetic) research results. The quantitative part, however, was derived automatically by optimising the parameters of a non linear model for the prediction of f0 contours, based on given statistical data. The advantage of this approach lies in the fact that a high quality of f0 generation was reached with only little research on natural f0 production. Disadvantages of this approach are that little insight is gained into the actual rules that govern the phonetic-acoustic realisation of intonation and, connected with this problem,*

[189] Cf. l'approche de « la somme des produits » (*sum-of-the-products*) proposé par van Santen : a) à chaque facteur on attribue une valeur quantitative du paramètre en question, ensuite, on applique des règles combinatoires (modèle de somme des produits).

that it is rather difficult to effect a given desired change of the behaviour of the resulting f0 generator. »

Ainsi la citation ci-dessus illustre une des positions vis-à-vis des méthodes stochastiques : l'auteur attribue l'avantage majeur des méthodes stochastiques à l'économie de l'apport humain requis. Les méthodes stochastiques sont, selon lui, plus sensibles aux propriétés/événements phonologiques qui régissent (d'après la littérature) la réalisation de l'intonation qu'aux données quantitatives. Or, un désavantage important est également mentionné : on n'acquiert aucune connaissance sur les règles précises et les processus sous-jacents ainsi que sur l'ampleur de l'effet induit par tout paramètre. Par conséquent, le système fondé sur des données stochastiques est difficile à optimiser (y incorporer une propriété nouvelle influençant la génération de f0).

De plus, dans le discours humain, il n'existe pas de relation biunivoque entre le texte et la prosodie, et plusieurs variantes peuvent être acceptables. Quoiqu'il en soit, les méthodes statistiques présentées ne permettent pas de traiter de cette variabilité.

De son côté, Sproat *et al.* (1998 :77) avancent les arguments en faveur d'une approche combinée, qui puiserait à la fois dans la réserve des connaissances linguistiques acquises et formalisées sous forme des théories prosodiques. L'adoption de cette ligne de conduite permettra de dépasser les désavantages des systèmes stochastiques cités (entre autres, la pauvreté de la performance sur des données nouvelles) :

« …We have nothing in principle against automatic or statistical methods. […] However, we do not believe such approaches can be done in absence of significant human intervention in the sense that it is, in many cases, a human who needs to decide which features are likely to be important in making decision. »

Du point de vue méthodologique, nous observons donc une opposition tranchée entre les approches statistiques *versus* symboliques dans le domaine de TTS. De ce fait, il nous paraît nécessaire, dans le cadre du présent travail, de nous pencher sur la question de la quantité et de la nature des connaissances linguistiques qui doivent être intégrées dans un système de génération de la parole.

6.3. MODÈLES PROSODIQUES EN SYNTHÈSE VOCALE

Dans la présente section, nous nous arrêterons sur quelques algorithmes développés pour la génération de l'organisation prosodique dans les systèmes de synthèse vocale. Nous avons précisé auparavant que deux étapes sont différenciées dans le processus de gestion de l'organisation prosodique du message : la génération d'une représentation symbolique (sous la forme d'une distribution des étiquettes prosodiques abstraites) suivie par le calcul de la courbe de f0 et des durées attribuées à tous les segments. Nous suivrons ce schéma dans la présentation des algorithmes ci-après. Cependant, nous prêterons plus d'attention aux modèles de calcul du phrasé prosodique, ce sujet étant au cœur de notre réflexion sur l'organisation prosodique du message en langue russe. De même, notre présentation des modèles pour la génération de la représentation

acoustique sera limitée à la problématique du calcul du contour de f0. Précisons que nous jugeons le sujet de la gestion de l'organisation temporelle de l'énoncé digne d'une analyse beaucoup plus détaillée que les limites du présent travail ne l'autorisent.

6.3.1. Prédiction du phrasé prosodique

La plupart des systèmes TTS utilisent les algorithmes assez simples pour prédire le phrasé prosodique. Les informations prises en compte sont limitées à la ponctuation, à des listes des mots-clés et à un apprivoisement plus ou moins approfondi de l'eurythmie en tant que principe régulateur de l'activité langagière (Di Cristo 1999) (les informations plus détaillées sont disponibles dans le domaine de synthèse Concept-to-Speech). Par ailleurs, les systèmes se différencient par référence aux tentatives d'intégrer les algorithmes avec une utilisation plus importante des informations syntaxiques lors de la génération du phrasé prosodique.

La plupart des systèmes sont orientés vers l'application des techniques statistiques de l'apprentissage automatique ; comme le remarquent Hirschberg et Prieto (1996:23), les heuristiques manuelles ne seront pas suffisamment performantes :

> *« Even if accurate syntactic and semantic information could be obtained automatically and in real time for text-to-speech, such handcrafted rules systems are notoriously difficult to build and to maintain*. »

Une caractéristique commune des différentes approches – et ceci est vrai pour les heuristiques manuelles comme pour l'apprentissage automatique – réside dans le procédé de travail à partir de l'analyse des productions naturelles. La variabilité stylistique des corpus analysés n'est pas importante, incluant le style lu et majoritairement, ces dernières années, celui de la présentation radiophonique[190]. Or, une approche alternative est également évoquée dans certains travaux : Hirschberg et Prieto (1996) discutent de l'approche mise en place dans les laboratoires AT&T, à savoir, le travail effectué à partir des transcriptions de textes réels annotés en termes de phrasé plausible par un expert. La raison principale de ce choix méthodologique réside dans une réduction plus que significative du temps de travail nécessaire pour avoir une quantité suffisante de matériau pour l'entraînement de l'algorithme. Les auteurs reconnaissent que la tâche d'identification des frontières prosodiques dans la production d'un locuteur est beaucoup plus difficile et nécessite plus de temps et d'implication que la prédiction des frontières plausibles à partir du texte.

Une précaution à prendre lors du développement des algorithmes est la réflexion sur la disponibilité des informations linguistiques nécessaires pour l'application de l'algorithme. Ainsi la particularité de l'algorithme proposé par Altenberg (1987) réside dans une prise en compte des informations discursives, sémantiques et syntaxiques ; pourtant, de manière générale, récupérer ces informations à partir du texte seul n'est pas

[190] L'etude de Wang et Hirschberg (1992) est réalisée à partir d'un corpus de questions spontanées, ce corpus étant obtenu dans une simulation de l'interaction avec un système informatique de commande de billets d'avion.

une tâche aisée, même si celle-ci n'est pas traitée par l'auteur. L'algorithme d'Altenberg reconnaît trois niveaux de phrasé déterminés en termes syntaxiques (on distingue les frontières des énoncés, des clauses grammaticales et des syntagmes). Or, la performance de l'algorithme est fortement dépendante du niveau (*alias* de la force) d'une frontière : il atteint le niveau de 87% pour les frontières d'énoncé (distinguant les structures avec le discours direct, les subordonnées complétives nominales introduites par *that* etc.). Or, pour les niveaux inférieurs, l'auteur conclut en soulignant l'absence de régularités d'association entre les niveaux prosodique et grammatical.

Pour notre part, nous insisterons sur une autre précaution importante, à savoir disposer d'un algorithme en adéquation avec les mécanismes cognitifs de l'humain.

6.3.1.1. Rôle des informations syntaxiques dans les systèmes TTS

Dans le chapitre II nous avons entrepris une présentation et une réflexion sur l'interprétation actuelle de la structure prosodique hiérarchique en phonologie. Nous avons particulièrement insisté sur la position prise selon laquelle la structure prosodique est considérée comme autonome par rapport à la structure syntaxique, quoique dérivée tenant compte de l'organisation syntaxique. La réflexion sur le rôle de la syntaxe en synthèse vocale est menée dans la perspective du lien supposé entre la structure syntaxique et la structure des constituants prosodiques. En effet, quelle que soit la théorie prosodique sous-jacente au modèle d'intonation implémenté dans le système TTS, une réflexion est nécessaire sur la formation des syntagmes intonatifs en tant qu'unités auxquelles sont associés les contours mélodiques (ou les tons de frontière) et les profils de durée. Nous avons évoqué précédemment les difficultés à disposer d'une représentation pragmatico-discursive pour l'énoncé à synthétiser. Pour la génération de la structure prosodique, la rupture s'opère pour la prise en compte des contraintes rythmiques. Trois classes d'approches peuvent être citées en relation avec le traitement de la syntaxe en algorithmes destinés à générer la structure prosodique :

a) prise en compte du rythme seul ;

b) postulat de la congruence entre la structure syntaxique et prosodique. Cette approche contredisant la position que nous avons défendue au préalable, nous n'entreprenons pas sa présentation détaillée ;

c) une approche mixte.

Des algorithmes ont été proposés pour construire des constituants prosodiques en s'appuyant sur des contraintes rythmiques seules : les chercheurs travaillant sur des contraintes rythmiques les associent avec les contraintes plus générales, d'ordre biologique et psychologique, sur la production rythmique (Fraisse 1974). Traditionnellement, on prend en compte la longueur des constituants et les résultats de l'analyse du corpus en mots lexicaux *versus* grammaticaux. Une telle approche a été proposée par Pasdeloup (1992) pour le français. L'auteur propose un modèle rythmique de la génération du phrasé prosodique à partir duquel on procède à l'implémentation de

la courbe de f0[191]. Les contraintes rythmiques prennent en considération la taille du groupe accentuel, le nombre de syllabes inaccentuées consécutives, la position du premier accent lexical dans l'énoncé[192], la fonction régulatrice de l'accent secondaire en français, la structure rythmique de l'énoncé entier, laquelle doit être gouvernée par un principe d'alternance. La structure prosodique adoptée dans le modèle comporte trois niveaux : les séquences rythmiques (syntagmes intonatifs majeurs, déterminés par la ponctuation), les mots rythmiques (syntagmes intonatifs mineurs) et les groupes accentuels (pieds ïambiques).

Lors de la génération de la structure prosodique pour un énoncé on applique l'approche *bottom-up* : la segmentation en mots rythmiques se fonde sur un dictionnaire limité de mots grammaticaux, les frontières étant insérées devant les mots grammaticaux. Ensuite, une série de contraintes[193] s'applique visant une structure plus harmonieuse (eurythmique) :

a) si le mot rythmique comporte moins de trois syllabes, il sera combiné avec le mot suivant, et ceci jusqu'à ce que le résultat n'excède pas sept syllabes ;

b) si le mot rythmique a plus de sept syllabes, il sera segmenté en deux mots de trois syllabes au minimum ;

c) si une séquence rythmique comporte quatre mots rythmiques et plus, elle est segmentée en deux séquences rythmiques phonotactiques.

Un avantage de cette approche réside dans le fait que l'application des contraintes rythmiques permet d'obtenir des structures différentes pour les énoncés ayant une même organisation syntaxique en fonction de la longueur des mots utilisés. Cependant, la question de la possibilité d'appliquer l'approche décrite aux langues autres que le français se pose, en particulier, aux langues à l'accentuation libre.

Dans les approches mixtes les contraintes rythmiques interviennent au même titre que d'autres informations lors de la génération de la structure prosodique ou interviennent postérieurement pour la gestion de la distribution des proéminences mélodiques. Nous évoquons cette question dans la partie consacrée à la problématique de la formation des constituants prosodiques comme l'application de la règle rythmique est limitée au domaine d'un syntagme phonologique. C'est le cas de la règle rythmique proposée par Gussenhoven (1991) : en effet, les algorithmes de prédiction de la distribution des proéminences mélodiques ont le défaut de prédire trop d'accents. Pour remédier à cela, certains accents se voient supprimés. La règle rythmique assure la suppression d'un accent si ce dernier se retrouve dans le contexte de deux autres accents et ceci à l'intérieur d'un syntagme phonologique.

[191] Cette position se situe dans la lignée de l'observation de Liberman (1975) selon laquelle la structure intonative est générée à la base de la structure rythmique.
[192] Cf. le constat de la tendance attestée pour le français (Pasdeloup 1988) et l'anglais (Shattuck-Hufnagel 1992) de placer le premier accent le plus tôt possible dans le syntagme.
[193] Remarquons que les contraintes sont sensibles à l'idéostyle du locuteur et au style de parole (conçu par référence au débit d'élocution).

Règle rythmique :

$$* \rightarrow 0 \ / \ [* \ _ \ *]_{PP}$$

Cette règle est jugée comme étant optionnelle et peut être appliquée plus d'une fois. Or, l'application récursive peut aboutir à une structure où seulement les deux accents initial et final sont présents. Le procédé décrit illustre la prise en compte des contraintes rythmiques lors de la gestion de l'organisation prosodique du message, mais non de manière exclusive.

Comme le démontre l'ensemble des travaux discutés dans le chapitre IV, les constituants prosodiques sont construits sur la base d'informations syntaxiques, discursives et rythmiques : or, la question se pose de connaître le degré de précision requise dans l'analyse syntaxique.

L'axe de réflexion récurrent dans la quête des indices linguistiques ainsi que lors de la conception des algorithmes est celui des rapports entre les structures syntaxiques et prosodiques. Comme nous l'avons déjà mentionné, les premiers algorithmes présupposent une congruence entre les deux structures : dans son étude, Silverman (1987) associe toute frontière majeure avec une marque de ponctuation ; de même, toute position entre deux mots lexicaux ou après un mot grammatical est annotée comme un site de frontière mineure.

Des constituants syntaxiques peuvent être construits à partir de l'étiquetage des mots en termes de parties du discours. Les constituants résultants sont comparables aux syntagmes syntaxiques minimaux (remarquons que pour certaines langues nous pouvons déjà obtenir des constituants prosodiques d'un certain niveau). Ainsi, pour un syntagme français « *un professeur du français canadien* », nous obtiendrons une structure parenthétisée suivante [un professeur] [du français] [canadien], les unités résultantes étant les groupes accentuels, les groupes intonatifs de Mertens (1987). La structure syntaxique peut également être obtenue par les méthodes d'apprentissage automatique.

Avant de procéder à une présentation concise de certains algorithmes proposés, nous considérons comme important d'en résumer les points communs.

Tout d'abord, il s'agit de l'uniformité de l'unité d'analyse dans tous les algorithmes : on considère tout couple de mots adjacents comme étant le site potentiel d'une frontière prosodique. Par la suite, la fenêtre centrée sur l'unité d'analyse peut être de largeur variable : par exemple, dans l'étude de Hirschberg et Prieto (1996), on analyse les informations sur les parties de discours de quatre mots, deux avant et deux après une frontière potentielle. Citons également l'étude de Tao (2004) qui exploite directement l'impact de la largeur de la fenêtre dans la performance du système ; l'auteur conclut sur le rôle plus important des informations à gauche de la frontière et propose pour les données analysées une fenêtre optimale de trois syllabes, dont une post-frontalière.

Le deuxième point commun des approches à la génération du phrasé prosodique est résumé par le procédé de produire une sortie qualitative[194] à partir d'un espace vectoriel des facteurs discrets relevant des phénomènes linguistiques variés. Ces facteurs forment 5 groupes ; nous y distinguons

a) les facteurs syntaxiques : il s'agit premièrement d'une annotation en parties de discours[195] ; ensuite, on prend en considération le signe de ponctuation interne de l'énoncé, le type de constituant syntaxique dont les frontières coïncident avec la position analysée, le constituant dominant le mot à gauche de la frontière et non pas à sa droite ;

b) les facteurs accentuels : prise en compte de la présence d'une proéminence mélodique aux alentours d'une frontière. Dans cette perspective, l'étude de Wang et Hirschberg (1992) analyse l'impact de la qualité de ces informations à la performance du système : les auteurs effectuent une comparaison entre deux algorithmes, le premier exploitant les informations précises sur la distribution des proéminences mélodiques, issues de l'annotation manuelle du corpus de test, et le deuxième se fondant sur la caractérisation quadruple du statut accentuel des mots autour de la frontière (on y inclut les catégories « accentué », « clitique », « désaccentué » et « non-accentué »). Il découle de cette comparaison que le niveau global de la performance de l'algorithme reste stable, or, la profondeur de l'arbre de classification augmente de cinq à neuf, ainsi que le paramètre « présence d'une proéminence mélodique » n'est plus pris en compte dans la classification[196].

c) les facteurs rythmiques (temporels) sont multiples. On prend en compte la longueur de l'énoncé exprimée en nombre de mots et en nombre de syllabes[197], la distance de la position analysée (exprimée en mots et en syllabes) par rapport au début et à la fin de l'énoncé, la distance par rapport à la ponctuation interne de l'énoncé, le nombre de mots dans le syntagme nominal et la distance dès son début, si le couple de mots fait partie d'un même syntagme nominal (cf. Sproat 1992). Nous avons qualifié ce groupe d'indices de rythmique, car le concept d'eurythmie leur est sous-jacent (on cherche une structure équilibrée) ; une deuxième motivation provient des contraintes sur le traitement perceptif du message délivré, et est donc fondé sur le présupposé d'une visée communicative de tout acte de parole.

[194] Avec des étiquettes possibles « absence de frontière prosodique » *versus* « présence d'une frontière prosodique du niveau *x* », le *x* variant selon les modèles.
[195] Le degré de précision de cette annotation varie entre les systèmes : on peut se limiter à la différenciation entre des mots grammaticaux *versus* lexicaux, ou bien appliquer un schéma détaillé.
[196] Les auteurs associent ce fait à la nature particulière de leur corpus (parole spontanée) : en effet, dans la production à la base de la modélisation beaucoup de mots grammaticaux sont accentués, ce qui n'est pas pris en compte lors de la prédiction de la distribution des proéminences mélodiques ; par conséquent, cela fausse la prédiction du phrasé prosodique.
[197] Selon la tradition, on mesure la longueur du constituant en syllabes. Pourtant, Baschenko et Fitzpatrick (1990) considèrent que la longueur de l'énoncé est mesurée de manière plus adéquate en mots phonologiques ou syllabes métriquement fortes.

d) les facteurs phonologiques divers : le statut métrique de la syllabe avant la frontière (forte ou faible).

e) les informations sémantico-discursives : la répétition du mot, analyse des bigrammes (fréquence de cooccurrence, *mutual information score*).

Nous avons essayé de recenser ci-dessus les différents types de facteurs linguistiques en jeu dans la prédiction du phrasé prosodique. Remarquons, pourtant, qu'il en existe un noyau dur, revenant d'une étude à l'autre. Simultanément, certains facteurs cités se voient exclus dans le modèle final parce que les tests révèlent qu'ils ne participent pas à la classification. La problématique adjacente est celle de la profondeur de l'analyse à effectuer et plus particulièrement de la prise en compte des étiquettes des frontières prosodiques prédites auparavant.

Une heuristique fondée sur la constituance syntaxique du message a été proposée par Bachenko et Fitzpatrick (1990), largement motivée par l'étude de Gee et Grosjean (1983). Leur approche présuppose l'existence d'un niveau du phrasé prosodique indépendant de la sémantique discursive, mais qui est pour autant régi par des contraintes syntaxiques et celles de la bonne formation des constituants prosodiques[198]. Cette approche intègre la vision hiérarchique de la constituance prosodique : la génération de la structure prosodique commence au niveau du mot phonologique déterminé par référence aux mots lexicaux. Ensuite, les syntagmes phonologiques sont formés : chaque syntagme phonologique comprend sa tête syntaxique et le matériau qui intervient entre la tête syntaxique précédente et la tête actuelle ; d'après l'analyse de Selkirk (1984) les auteurs considèrent que les têtes potentielles sont le nom, l'adjectif et l'adverbe si elles ne précédent pas une autre tête (le verbe a un statut particulier dans l'algorithme développé). Dans la lignée de l'hypothèse de l'étagement strict, les auteurs considèrent que toute frontière de syntagme phonologique est le site potentiel d'une frontière prosodique (*alias* d'une frontière de l'unité intonative) : ce sont des règles de saillance qui déterminent quelles frontières sont propulsées au niveau supérieur de la hiérarchie. Par conséquent, l'objectif d'application des règles de saillance est de diminuer ou d'augmenter la force d'une frontière du syntagme phonologique aboutissant ainsi à un phrasé prosodique, qui crée l'impression d'une structure équilibrée, avec des constituants qui tendent d'être d'une longueur égale. La règle de la structure verbale équilibrée (*verbal balancing rule*, héritée de l'approche par Gee et Grosjean 1983) est un exemple des règles de saillance :

> *Règle de la structure verbale équilibrée :*
> Étant donné la séquence [X V Y], C – nombre de mots phonologiques,
> si C(X)+C(V) < C(Y), alors [(XV)Y], autrement [X(VY)]

Nous avons vu que l'algorithme de Bachenko et Fitzpatrick (1990) prend en compte de manière explicite la structuration de l'énoncé en constituants prosodiques. Remarquons également que, dans cette étude, les règles sont construites par des expérimentateurs qui se fondent sur des données issues de la théorie linguistique.

[198] Il s'agit des contraintes rythmiques dans la terminologie que nous avons proposée ci-dessus.

Dans l'étude de Wang et Hirschberg (1992), les auteurs testent l'impact de différents types d'information sur la performance de la méthode statistique des arbres de classification en génération du phrasé prosodique. Ils font contraster les agrégations de paramètres récupérés de l'annotation manuelle du corpus (incluant des informations sur la position de la frontière prosodique précédente et la distribution des proéminences mélodiques) et les mêmes paramètres extraits automatiquement à partir du texte. Dans les premiers tests, l'information sur la longueur du syntagme en cours par rapport au syntagme précédent assurait déjà la classification correcte de 2403 sites parmi 2556 ; également les informations sur la structure des proéminences jouait un rôle très important. Or, si les paramètres ont été extraits automatiquement, la hiérarchie des paramètres retenus lors de la classification comprend les mesures temporelles (distances par rapport au début et à la fin de l'énoncé), étiquettes de parties du discours du mot suivant et du mot précédant la frontière, le constituant syntaxique dominant le mot avant la frontière, le débit et le type d'énoncé. Il est apparent donc que les informations sur la distribution des proéminences mélodiques n'apparaissent que bien plus tard dans la hiérarchie. Pour nous, ces résultats posent une question primordiale : dans quelle mesure les procédés mis en œuvre dans les systèmes appliqués sont-ils en cohérence avec le traitement perceptif et cognitif des informations langagières par l'humain ?

Nous tenons pour important de nous arrêter sur le travail de Ostendorf et Veilleux (1994) qui applique les méthodes de l'apprentissage automatique à une modélisation de la structure hiérarchique des constituants ; leur modèle est donc stochastique et non-déterministe. Un traitement de la structure enchâssée abstraite est d'abord proposé, sans spécification du nombre de niveaux dans la hiérarchie ; ensuite, les auteurs procèdent à une étude de cas, avec trois niveaux enchâssés, tout énoncé étant structuré en syntagmes majeurs, qui sont à leur tour composés de syntagmes mineurs[199].

Il est supposé que la relation entre les unités et leurs sub-unités (les unités du niveau immédiatement inférieur) est la même quels que soient les deux niveaux de la hiérarchie prosodique. Par conséquent, la probabilité d'une unité majeure est fonction de la probabilité de la séquence des unités mineures et de leur nombre. Dans la suite de la réflexion il est supposé que les unités du niveau inférieur suivent le principe des séquences Markoviennes, et ne dépendent que des propriétés de l'unité précédente et des propriétés de l'unité supérieure dont elles font partie. Le mécanisme statistique appliqué à l'étape d'apprentissage est celui des arbres de classification. Or, à la différence de l'approche de Wang et Hirschberg (1992), à l'issue de l'analyse on n'en récupère que la probabilité de retrouver une frontière prosodique mineure à une localisation précise. Une structure optimale est postérieurement choisie.

Un autre point d'intérêt de l'étude analysée réside dans la mise en place d'une expérimentation destinée à tester la force prédictive de différents groupes de paramètres. Trois ensembles de paramètres constituent l'entrée de l'analyseur pour la prédiction des frontières prosodiques mineures :

[199] Remarquons que la plupart des théories prosodiques excluent le niveau d'énoncé de la hiérarchie prosodique.

1) le premier set n'inclut que les informations simples sur la ponctuation, l'appartenance à la classe des mots grammaticaux ou lexicaux et de simples contraintes rythmiques ; l'arbre résultant est relativement petit, comportant 9 nœuds, et la prédiction correcte atteint le niveau de 89% pour les absences de frontière et 59% pour sa présence ;

2) à la deuxième étape on inclut les informations sur la constituance syntaxique : l'arbre résultant ne comporte que 5 nœuds, la précision de la classification et de 90% et 62% respectivement ;

3) à la troisième étape, on renforce les contraintes rythmiques en en ajoutant une au groupe 1 dans le souci d'aboutir à un équilibre entre les syntagmes mineurs adjacents : on n'observe pas de différences majeures par rapport au premier test, les prédictions correctes atteignent 87% et 66%.

Dans la phase test de l'expérimentation, Ostendorf et Veilleux trouvent que c'est le premier ensemble de paramètres qui est le plus performant ; elles en concluent donc que les informations syntaxiques ne sont pas d'une importance cruciale dans l'accomplissement de la tâche de prédiction du phrasé prosodique.

Remarquons pourtant que les auteurs ne procèdent pas à une modélisation des frontières prosodiques majeures vu que ces dernières se retrouvent dans des positions qualitativement similaires à celles des frontières mineures (en référence aux paramètres choisis).

Le résultat sur la non-pertinence des informations syntaxiques dans la prédiction de la structuration de l'énoncé en termes des syntagmes mineurs paraît assez curieux : en effet, la phonologie prosodique présuppose une congruence beaucoup plus forte entre les niveaux inférieurs des structures prosodique et syntaxique. Tout de même, nous pouvons y voir une explication dans la définition adoptée des unités prosodiques en question : en effet, l'annotation initiale du corpus se fonde sur l'approche perceptive et l'échelle de zéro à six proposée et mise en pratique par Price *et al.* (1991). Or, cette approche est moins sensible à la structuration syntaxique du message (contrairement aux pratiques appliquées postérieurement dans le cadre de ToBI, cf. Witghman 2002).

6.3.2. Prédiction de la distribution des proéminences mélodiques

Comme cela a été déjà évoqué, de manière standard la distribution des proéminences mélodiques est générée en relation avec la classification des mots en des classes plus ou moins larges, caractérisées par un degré variable d'accentogénéité. Le nombre de classes varie selon les études : un pôle est représenté par des approches comprenant une seule différenciation entre les mots lexicaux et les mots grammaticaux ; le pôle opposé prend en considération toutes les parties de discours (cette démarche est compliquée par les phénomènes d'homonymie). Les modèles plus complexes disposent de règles permettant de modifier la structure initiale en relation avec des contraintes de haut niveau (discursives et rythmiques). Tous les algorithmes ont le mot comme domaine d'analyse dans la tâche de la prédiction des distributions des proéminences mélodiques :

ce choix méthodologique complique néanmoins l'implémentation des figures avec la proéminence mélodique qui doit être placée le plus tôt possible dans le syntagme, souvent associée avec la syllabe porteuse de l'accent secondaire (cas de figure valable pour l'anglais (Shattuck-Hufnagel 1992) et pour le français (le phénomène de l'arc accentuel, Fonagy 1979)).

À titre d'exemple nous présentons l'algorithme développé par Quené et Kager (1989) pour le hollandais : la propriété importante de cette approche réside dans un traitement partiel de la constituance prosodique que l'on obtient à partir de l'analyse syntaxique de l'énoncé. Cet algorithme procède, dans un premier temps, par génération d'une version avec le nombre surévalué de proéminences mélodiques, lesquelles sont attribuées à tout mot lexical, exception faite des verbes et de certains mots vidés de sens. Cette version est ensuite ajustée avec la prise en compte de trois types de contraintes :

a) rythmiques : on désaccentue tout mot dans le contexte entre deux mots porteurs de la proéminence mélodique (cf. la règle rythmique de Gussenhoven que nous avons présentée auparavant) ;

b) discursives : on désaccentue les mots qui renvoient à des informations connues ;

c) structurelle : cette règle impose l'accentuation du verbe dans le contexte où un constituant n'étant pas l'argument de verbe intervient entre le verbe et son argument.

Certains constituants syntaxiques nécessitent un traitement particulier, ce détail étant spécifique à la langue : nous nous référons notamment au travail de Sproat (1992) qui porte sur l'analyse de la distribution des proéminences mélodiques dans des groupes nominaux complexes en anglais.

De cette brève présentation ainsi que des analyses entreprises précédemment, nous concluons que le défi majeur dans la prédiction de la distribution des proéminences mélodiques est associée avec l'absence d'un modèle discursif déductible à partir du texte seul. Comme nous l'avons déjà souligné, cette problématique paraît résolue dans les systèmes CTS. Le modèle de Quené et Kager (1989) étant un modèle par règles développé à partir d'un corpus d'énoncés isolés, le transfert de ce modèle vers des textes plus longs avec une cohérence interne paraît donc être une tâche délicate. La méthode statistique a été proposée par Hirschberg (1993).

6.3.3. Modèles de la génération des courbes de f0

Dans le paragraphe précédent nous avons brièvement passé en revue les algorithmes mis en œuvre pour prédire l'organisation prosodique de l'énoncé sous la forme d'une annotation symbolique à partir d'un texte. La deuxième étape dans le traitement de l'organisation prosodique du message en synthèse porte sur la génération de la courbe mélodique, dans le cadre de la synthèse concaténative ce procédé étant combiné avec les procédés de modification des caractéristiques temporelles et mélodiques des segments.

Pour générer l'intonation dans un système TTS, on a besoin d'un modèle formel de l'intonation et d'un algorithme permettant de convertir la représentation formelle (l'ensemble des étiquettes) en un contour mélodique (on peut également envisager certaines modifications de l'organisation temporelle). De même, la problématique au centre des réflexions sur les modèles de l'organisation temporelle ou de l'organisation tonale est l'interaction des facteurs : ainsi l'effet induit par un facteur se voit amplifié ou atténué par le deuxième facteur. Dans le domaine temporel, van Santen (1992a) a émis l'hypothèse d'une invariance directionnelle qui est dorénavant appliquée dans les systèmes de synthèse et paraît donner des prédictions correctes : selon cette hypothèse, si un facteur induit un effet quelconque, le deuxième facteur n'exerce d'influence que sur le degré de cet effet, sans renverser radicalement sa direction. Savoir dans quelle mesure ce modèle est opérationnel dans le domaine tonal reste soumis à des investigations approfondies.

Le modèle le plus simple est la reproduction de la ligne de déclinaison. Parmi les systèmes plus complexes, on différencie les modèles de la génération du contour mélodique selon leur caractère paramétrique ou non paramétrique. Le groupe de modèles paramétriques rassemble 3 types de modèles : a) les modèles adoptant la logique « source/filtre »[200], b) les modèles séquentiels, procédant par interpolation entre les points-cibles (Pierrehumbert 1981, première approche de BelLabs, Sproat *et al.* 1998), c) les modèles hybrides. Les modèles non-paramétriques, quant à eux, se fondent sur les modèles de neurones, les modèles cachés de Markov, et sur les approches par concaténation. Une vision superpositionnelle (additive) de la courbe mélodique est souvent commune à deux groupes. Pourtant, ce dernier facteur peut fournir un axe de classification supplémentaire des modèles de la génération de f0 : la question est de savoir si le registre mélodique reçoit une modélisation explicite dans le modèle[201]. Dans la plupart des modèles, on opère en deux étapes : dans un premier temps, on modélise le registre mélodique, ensuite les valeurs des points cibles associés aux proéminences mélodiques se voient insérées dans la bande de registre, définie par la ligne de base et la ligne de sommets. La modélisation du registre ne pose pas de problèmes, si les phrases sont courtes. Or, la problématique se complique lorsque les phrases deviennent plus longues avec des rapports logiques précis entre les parties : dans le domaine tonal, on associe les frontières des syntagmes intonatifs avec le phénomène de *resetting*, cependant son degré dépend largement des relations sémantico-pragmatiques entre les syntagmes (Ladd 1988).

Une modélisation du registre fondée exclusivement sur le nombre de syntagmes dans l'énoncé a été proposée par Clark (1999). Ce modèle intègre une structure prosodique hiérarchique à deux niveaux : on distingue premièrement un syntagme intonatif[202] qui comporte un ou plusieurs groupes tonals, définis comme unités structurales de base de la phonologie intonative ; les groupes tonals sont délimités du reste du discours par des configurations tonales particulières (tons du frontière). À partir de l'analyse d'un

[200] Le modèle paramétrique très connu de la production de l'intonation est celui de Fugisaki (1983)
[201] Il est à noter que la modélisation d'un composant syntagmatique, au moins en tant que la f0 moyenne dans le syntagme, est sous-jacente à tout modèle intonatif appliqué
[202] Cette unité reçoit une définition imprécise : elle est d'habitude suivie d'une frontière de force 5 et d'une pause silencieuse. Il n'est pas clair de dire si l'on peut y voir l'analogue d'un énoncé à l'oral.

corpus, l'auteur démontre que la gestion des phénomènes du registre est plus sensible à l'organisation du syntagme en termes de groupes tonals, le nombre de syntagmes intonatifs n'exerçant pas d'influence significative. Le modèle distingue trois types de groupes tonals : initiaux, médians et finaux ; les différences entre les groupes qualifiés de médians à l'intérieur d'un même syntagme intonatif ne sont pas significatives.

Une typologie supplémentaire s'opère selon le critère de la prise en compte des effets segmentaux (composante microprosodique, Di Cristo et Hirst 1986) sur la courbe mélodique : certains systèmes se donnent pour tâche d'imiter tous les détails du contour mélodique aussi minutieusement que possible ; par conséquent, ils intègrent des règles leur permettant de reproduire les perturbations segmentales de la courbe de f0 (Silverman 1987[203], van Santen *et al.* 2005). D'autres systèmes (Pierrehumbert 1981) se fixent l'objectif de développer un modèle intégral des effets prosodiques sur la courbe mélodique.

Un exemple du modèle fondé sur des règles est présenté dans le travail comparatif de Syrdal et al. (1998). Il s'agit de l'implémentation du modèle séquentiel de l'intonation : deux concepts y sont donc incorporés, celui des points cibles et celui du registre. Le registre est conçu comme une bande entre la ligne de base et la ligne des sommets. La position de chaque point cible est alors déterminée par une fraction de la distance entre les deux lignes prédéfinies. L'étendue du registre dépend de la structure discursive, une caractéristique que l'on ne peut pas déduire directement à partir du texte ; on le modélise à l'aide des arbres de régression à partir des informations sur la durée d'une pause, le nombre d'accents et la longueur de la phrase. La question de l'alignement est indirectement traitée en relation avec des informations sur la position du point cible par référence à la partie sonore de la syllabe.

Comme c'était le cas pour l'aspect symbolique de l'intonation, historiquement, les règles ont été remplacées par des méthodes stochastiques. Les techniques statistiques traditionnellement utilisées dans le domaine de TTS sont les réseaux de neurones, les arbres de classification et de régression et la régression linéaire.

Les réseaux de neurones ont été utilisés pour la synthèse par onde sonore (modèle source-filtre, Karaali, Corrigan et Gerson 1996), pour modéliser l'intonation (Sun 2001, Holm et Bailly 2002) et pour détecter les accents dans la reconnaissance automatique (Taylor 1995, Streefkerk *et al.* 1997). Cette méthode a également été appliquée à la génération de f0 (Traber, 1992) mais y reste moins populaire. Dans le modèle de Traber, l'unité d'analyse minimale est une syllabe, caractérisée par 8 points : premièrement, on accède aux demi-syllabes, qui sont ensuite segmentées en deux parties par rapport au point de l'énergie maximale de chacun. Cette analyse correspond à une stylisation du contour mélodique initial. Dans la tâche de prédiction, il s'agit de la prédiction du contour entier de f0 dans l'empan d'un énoncé (syntagme intonatif). Dans cette

[203] Le modèle de Silverman est un modèle paramétrique fondé sur des règles établies dans les expériences en perception. Les règles sont organisées en trois paliers responsables de la gestion : a) des perturbations de la courbe de f0 induites par les effets segmentaux ; b) des séquences tonales, traduisant les phénomènes de proéminences mélodiques et des tons de frontière ; c) des phénomènes de registre. L'intégration de ce troisième groupe de règles est rare : leur domaine d'application englobe plusieurs unités intonatives incorporant ainsi les effets au niveau du paragraphe.

perspective, on spécifie très minutieusement le contenu phonétique de la séquence d'entrée ainsi que la structure prosodique. L'auteur remarque pourtant que pour les données dont le patron mélodique est atypique (et qui n'était pas présent dans les données d'entraînement), la prédiction résultante peut être de très mauvaise qualité. Le travail de Traber compare deux méthodes (réseaux de neurones et méthode du lexique) du transfert d'une représentation symbolique vers les valeurs numériques de f0 appliquées dans un système de synthèse. Leur performance se révèle comparable. Or, comme le remarque l'auteur lui-même (Traber, 1992:303) « *it seems easier to find and eliminate sources of bad results in the database method than in the neural network approach* ». Comme nous l'avons déjà souligné le désavantage majeur des méthodes par apprentissage est le fait que l'on n'acquiert pas de connaissances sur des problèmes linguistiques sous-jacents auxquels on est confronté.

À son tour, dans le travail de Dusterhoff *et al.* (1999), un modèle de la génération de l'intonation par application de la méthodologie des arbres de classification et de régression est présenté. La différence entre les arbres de classification et les arbres de régressions réside dans le fait que les premiers prédisent la catégorie à laquelle appartient vraisemblablement l'événement analysé tandis que les seconds prédisent les valeurs quantitatives (en l'occurrence de f0). Le principe sous-jacent à cette technique consiste à rechercher à chaque étape une meilleure question pour la partition des données afin de minimiser l'écart type à l'intérieur de chaque classe. De cette différence émane le fait que les arbres de classification sont utilisés pour prédire la localisation des événements prosodiques, les arbres de régression – pour la prédiction des paramètres quantitatifs, des valeurs de f0.

Le présupposé général de la régression linéaire réside dans la considération que la variable dépendante peut être modélisée comme la somme des variables indépendantes pondérées ($p = w_0 + w_1f_1 + w_2f_2 + \ldots + w_nf_n$). Les facteurs dans l'équation ci-dessus sont ceux qui exercent une influence sur la variable dépendante et les poids respectifs sont entraînés dans les modèles linéaires.

Ainsi la génération du contour mélodique dans le système Festival (Black et Hunt 1996) se fonde sur trois modèles de régression linéaire, pour le début, le milieu de la voyelle accentuée et la fin de la syllabe accentuée. Les facteurs inclus dans le modèle comportent l'information sur le type d'accent présent (annotation ToBI), la position de la frontière syntagmatique, accentuation des syllabes et la position de la syllabe, et ceci dans une fenêtre de 5 syllabes centrée sur la syllabe accentuée (pour permettre au mouvement mélodique de s'achever sur la syllabe suivant la syllabe accentuée). Une fois implémentée, cette méthode permet d'augmenter la corrélation entre le contour synthétique et le contour naturel (0,62 contre 0,4 pour le système par règles) ainsi que de réduire l'erreur moyenne standard (jusqu'à 34,8 Hz). La performance du modèle est également testée pour le japonais, à partir d'un corpus utilisant les annotations J-ToBI et une amélioration par rapport à la méthodologie précédente est attestée.

Les modèles fondés sur un lexique de contours mélodiques (Traber 1992, Aubergé 1992) constituent une classe à part. Pour générer un contour mélodique en synthèse Aubergé (1992) procède à un travail préalable de construction d'un lexique fonctionnel de formes intonatives à partir d'un corpus de paires minimales. Toute paire minimale

est caractérisée par des attributs propres à chaque niveau syntaxique (au niveau de l'énoncé on distingue les assertions, les interrogations et les ordres), chaque attribut donnant lieu à des oppositions minimales, les énoncés assertifs pouvant être positifs ou négatifs. Le modèle hiérarchique mis à l'épreuve est ancré syntaxe ; on distingue donc les niveaux d'énoncé, de clause, de syntagme, de sous-syntagme (défini pour des syntagmes de plus de 4 syllabes). Le présupposé général de cette approche est le caractère additif de l'intonation : on considère que toute unité minimale définie pour chaque niveau linguistique est une unité intonative à part et possède un contour mélodique global qui lui est propre. Pour les unités non-terminales ce contour est défini par la ligne de déclinaison, les unités terminales étant définies par une séquence de valeurs (trois par syllabe). De ces présupposés théoriques sur l'organisation prosodique en parole émane la nature des unités incluses dans la base. Par exemple, on y retrouve un contour mélodique d'un syntagme nominal dans la position pré-verbale dans un énoncé assertif, comportant 3 syllabes et exerçant une fonction autre que complément. Pour chaque groupe, on calcule le contour moyen (procédure de standardisation). Le résultat d'une telle approche est un lexique des contours, organisé de manière hiérarchique en relation avec les niveaux linguistiques définis, tous les contours étant normalisés par rapport à la moyenne de la classe, et indexés avec les attributs appropriés.

À l'opposé du système proposé par Aubergé (1992), dans le travail de Malfrère *et al.* (1998) le choix du contour mélodique optimal dans une base de données s'opèresur la base du calcul du coût d'adéquation et de concaténation. Il s'agit toujours de la méthode de génération de la courbe mélodique à partir du lexique des formes minimales de base. Une autre caractéristique de ce système réside dans l'application exclusive des procédures de l'analyse automatique du corpus.

Ce modèle intègre une hiérarchie prosodique à deux niveaux (il est essentiellement fondé sur le modèle prosodique du français de Mertens 1987) : l'unité de base dans cette analyse est un groupe intonatif défini par la présence d'un et un seul accent final[204]. La formation des groupes intonatifs se fait par des règles relativement simples, nécessitant pourtant des informations sur les dépendances mutuelles et sur la structure syntaxique. Or, il est difficile d'obtenir la représentation complète nécessaire de façon automatique : par conséquent, un parseur brut n'a besoin que d'informations sur les parties du discours, la dichotomie s'opèrant entre un mot fonctionnel et un mot lexical, et des contraintes rythmiques[205].

Les méthodes d'apprentissage automatique interviennent lorsque la représentation symbolique est convertie en une représentation physique, spécifiant pour chaque phonème sa durée (on utilise la méthodologie des arbres de classification et de

[204] Dans le travail discuté, le groupe intonatif présente la structure prototypique suivante (nous notons entre crochets les éléments facultatifs) : $[NA]_nAF$ (ce modèle était développé postérieurement par avec l'inclusion dans la structure prototypique de l'accent initial ainsi que de syllabe inaccentuée en appendice).

[205] Donnons l'exemple d'une règle : si le mot lexical est suivi par un mot grammatical, la frontière du groupe intonatif passe entre les deux mots et le mot lexical reçoit un accent final. Ultérieurement, les groupes intonatifs forment les unités du niveau supérieur, l'énoncé étant une unité à part dans la hiérarchie, avec les tons de frontière particuliers (ton final bas ou haut). Les contraintes rythmiques à leur tour portent sur le nombre admissible de syllabes inaccentuées

régression. Pour la génération du contour mélodique, un dictionnaire mélodique est produit automatiquement à partir du corpus annoté. L'unité de base dans ce dictionnaire ne peut pourtant pas être considérée comme universelle : c'est le groupe intonatif plus la syllabe accentuée du groupe précédent (nous y voyons une unité dépendante de l'organisation prosodique du français, langue à accent final). Chaque forme dans le dictionnaire est codée avec une clé et pour chaque point cible sa valeur est déterminée en demi-tons par référence à la fréquence moyenne de toute la base. La clé comporte des informations sur la position de l'unité dans l'énoncé original (par référence au nombre de groupes intonatifs), le type de l'accent final précédent trouvé, la valeur moyenne (en Hz) des points cibles pour la syllabe accentuée, le nombre de syllabes inaccentuées, le type d'accent final dans l'énoncé original et la moyenne des valeurs de points cibles. Ainsi l'exemple de clé ci-dessous correspond au deuxième groupe intonatif avec trois syllabes inaccentuées et la valeur moyenne des points cibles de la syllabe accentuée de 4 demi-tons. Malfrère et Mertens (*op. cit.*) obtiennent pour leur corpus d'entraînement 1838 schèmes mélodiques avec 5 points cibles pour chaque syllabe.

2 FA1,NA1,NA2,NA3,FA2 2.5 ST 4.0 ST

Le système proposé par Kochanski et Shih (2003) cherche à combiner l'annotation du texte et la génération de l'intonation, à rendre les deux processus intrinsèquement réversibles, et par ce fait, il est proche du groupe des modèles par lexicon. L'idée clé des auteurs consiste à considérer qu'un système d'annotation développé facilitera postérieurement l'adaptation du système aux nouvelles contraintes (variations stylistiques, effets discursifs recherchés) ; on se situe dans la logique proche des systèmes CTS, quand les balises proposées sont directement associées aux formes mélodiques, cf. également l'approche de Mertens *et al.* (2001). Un ensemble d'étiquettes est défini et paramétré, lesquelles exercent une double fonction : celle d'annoter le texte et celle d'assurer la génération de la prosodie dans les systèmes appliqués ; le système est également doté d'un inventaire de correspondances entre les étiquettes et les courbes de f0. Ces auteurs considèrent que les formes idéales des courbes mélodiques accentuelles se voient modifiées dans le discours à cause de la proximité de deux événements prosodiques, et ceci dans la perspective de minimiser l'effort physiologique sous les contraintes communicatives ; la modification s'opère en relation avec la force relative associée à chaque événement prosodique. Le caractère adaptatif du système se manifeste en particulier à travers la réflexion sur la pré-planification dans le discours. Comme il a été démontré dans les recherches expérimentales, la planification accompagne le processus de la parole : elle porte sur le volume d'air inspiré ainsi que sur la hauteur mélodique initiale en fonction de la longueur de l'énoncé. On sait également que le domaine de planification se raccourcit sous la charge cognitive de la conception du message. Dans le modèle que les auteurs développent, une étiquette spécifique permet de gérer l'empan de la planification. Le modèle est défini en termes physiologiques, associant les valeurs de f0*[206] avec la tension musculaire de la source glottique.

[206] Cette annotation est adoptée pour la courbe de f0 lisse, avec des perturbations segmentales étouffées.

Les auteurs cherchent un système indépendant de la langue et neutre par rapport aux théories linguistiques sous-jacentes : ce choix méthodologique permet de se servir du modèle en tant qu'outil diagnostique et d'évaluer les avantages des différentes théories. D'autre part, la caractère neutre du modèle implique une spécification postérieure des domaines linguistiques associés aux étiquettes : ainsi l'étiquette *« stress »* peut-elle faire référence à une syllabe, un mot ou un groupe accentuel. On différencie quatre groupes d'étiquettes : a) paramètres généraux, définissant le registre mélodique du locuteur ;b) paramètres définissant la courbe de f0 ; c) paramètres associés avec les phénomènes d'accentuation ; d) paramètres associés avec les mouvements mélodiques aux frontières des constituants.

Dans le chapitre II *supra*, nous avons expliqué notre choix de travailler avec le système d'annotation INTSINT. La logique interne de ce modèle est telle qu'une fois l'étiquette connue, les valeurs de f0 des points cibles sont retrouvées très facilement[207] ; comme la fonction d'interpolation est également connue (fonction spline quadratique), les valeurs de f0 peuvent ensuite être calculées pour n'importe quel segment. La qualité dépendra donc largement du modèle de l'organisation temporelle intégrée dans le système ; autrement, l'affinement du modèle standard (note 6) est requis pour arriver à une prosodie plus naturelle, ainsi que pour l'application du modèle à une langue autre que le français. Deux travaux ont appliqué INTSINT à la modélisation de la prosodie en synthèse. Nous pensons premièrement au travail de Veronis *et al.* (1997) qui applique une méthode stochastique pour aboutir à une annotation symbolique du corpus. Il s'agit d'un modèle probabiliste de la génération de la courbe de f0 à partir d'un corpus étiqueté en catégories grammaticales. On modélise le processus de la génération de l'intonation faisant référence au modèle du canal bruité (*noise chanel*) : vu la séquence des catégories grammaticales en sortie, on se pose la question de la séquence des événements prosodiques la plus probable pour produire les classes grammaticales observées. Cependant, deux présupposés sont inhérents à cette modélisation : on présume qu'une seule étiquette (point cible) est associée au mot en tant que catégorie grammaticale ; dans un deuxième temps, on considère que l'étiquette courante ne dépend que des étiquettes voisines (précédente et suivante) et que la catégorie grammaticale dépend de l'étiquette prosodique ainsi que de la classe de l'unité précédente.

Un deuxième travail d'implémentation d'INTSINT en synthèse dans le système SYNTAIX (Di Cristo et Di Cristo 2001) apporte plutôt des modifications à l'étape de la production d'une représentation symbolique se basant largement sur l'approche autosegmentale-métrique de l'intonation.

---

[207] Di Cristo et Di Cristo (2001) citent les règles suivantes gérant l'alignement des points cibles d' INTSINT :

a) on aligne le ton M avec le début de l'attaque de la syllabe ;
b) le ton Hi est aligné avec le début de la rime dans le cas d'un accent initial de Mot Prosodique, et avec l'attaque dans le cas d'un accent initial de syntagme prosodique ;
c) le ton H de l'accent final est aligné avec le centre de la rime ;
d) les tons B et T (il s'agit des tons de frontières droites) sont alignés avec la fin de la rime ;
e) l'abaissement local est positionné au centre de la syllabe pénultième et maintenu jusqu'à son terme ;
f) l'alignement d'un ton L entre deux tons H consécutifs bénéficie d'une marge de liberté, exploitée dans l'objectif d'atteindre une certaine variabilité.

Le synthèse de BellLabs, implémentation du modèle séquentiel : la hauteur en Hz des cibles mélodiques, leur alignement précis avec le niveau segmental sont contrôlés par des règles d'implémentation phonétique qui prennent en considération la spécification tonale, les informations segmentales, les différents effets d'abaissement, et les valeurs de proéminence pour chaque accent. On procède ensuite par interpolation entre les points cibles consécutifs et par application d'une fonction de lissage. Pour implémenter les cibles tonales on établit au préalable trois lignes de référence (ligne de sommets, ligne de référence, et ligne de base) qui sont modifiées avec la progression de l'énoncé pour rendre compte des effets de la déclinaison. Les cibles H et L sont positionnées en référence à cette grille : les valeurs H se positionnent toujours au-dessus de la ligne de référence et sur la ligne des sommets si le degré de proéminence attribué à un accent égale 1 ; à leur tour, les cibles basses L sont positionnées au-dessous de la ligne de référence. La ligne de base est le minimum mélodique atteint à la fin de l'énoncé assertif.

6.3.4. Méthodologie d'une double sélection des unités (van Santen et al. 2005)

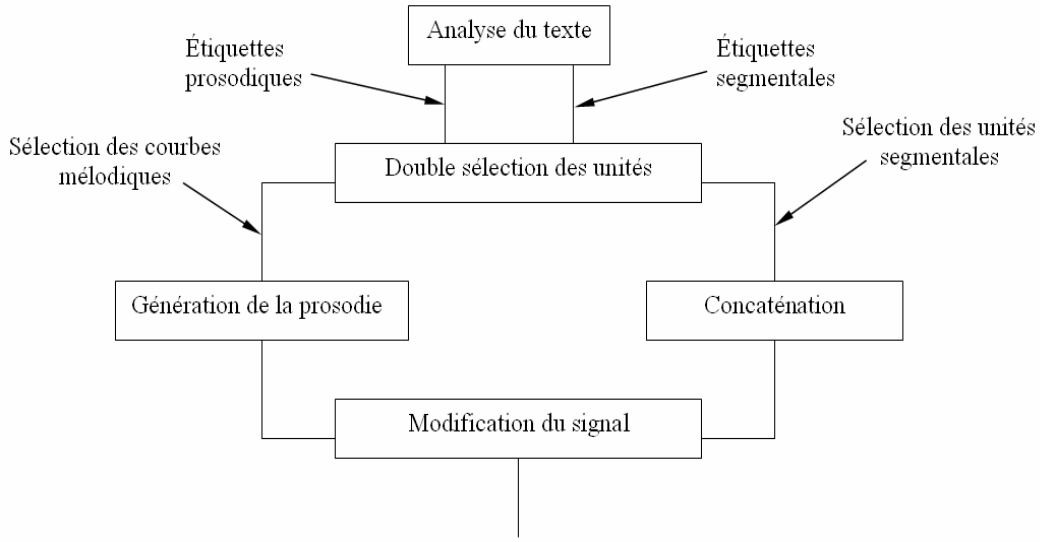


Figure 6.3 : Représentation schématique de l'approche d'une double sélection d'unités

Collaborateur permanent de Bell Laboratoires, Jan van Santen travaille depuis longtemps à l'amélioration de la qualité et du naturel de la parole synthétisée (TTS, cf. ses travaux sur la modélisation de l'organisation temporelle de l'énoncé). Le modèle que nous présentons dans ce paragraphe a été proposé récemment ; son objectif premier est de faire collaborer au sein d'un même système les avantages des deux méthodologies, celle de l'application des règles et la méthodologie classique de la sélection d'unités. Les auteurs aspirent à obtenir un contour mélodique plus naturel et à minimiser la quantité de la modification postérieure du signal afin d'aboutir à la structure prosodique requise.

Le principe sous-jacent du système proposé par van Santen réside en une combinaison des unités segmentales et des contours prosodiques naturels, puisés dans le corpus de base, en se fondant sur le modèle superpositionnel de l'intonation (van Santen et Möbius 2000). Ainsi à l'instar de la méthodologie classique de la sélection des unités,

on dissocie les deux procédures : recherche des unités segmentales convenables et satisfaction des contraintes imposées par la structure prosodique. On dispose donc des séquences multiples (*multiple unit sequences*), l'une comportant les intervalles dont les étiquettes phonémiques correspondent à la séquence d'entrée, les autres comportant les intervalles dont l'annotation prosodique correspond à celle de la forme d'entrée au niveau d'un constituant phonologique (pied ou syntagme). Le contour mélodique résultant (obtenu par l'addition des contours mélodiques retrouvés dans le corpus) est ensuite transplanté sur la séquence phonémique avec l'application de procédures standard de la modification du signal.

L'originalité de cette approche réside dans le fait que la courbe mélodique ne résulte pas d'une concaténation directe des contours mélodiques retrouvés, mais s'appuie sur le modèle intonatif sous-jacent. Le traitement des propriétés tonales y est étroitement associé à la gestion de l'organisation temporelle de l'énoncé : on insiste sur la transplantation des profils de durées. En effet, les durées segmentales ne sont pas générées à l'aide des mécanismes externes (méthodes traditionnelles, se basant sur les arbres de régression et de classification, ou modèle « somme des produits ») ; on définit le contour temporel qui est transféré sur la séquence phonémique. Il s'agit d'un moyen de s'assurer de la coordination adéquate entre l'organisation temporelle au niveau segmental et la progression dans le temps du contour mélodique.

Deux méthodes d'implémentation de cette approche sont ensuite discutées (van Santen et al. 2005) en relation avec les contraintes sur le type de corpus à utiliser et à maîtriser. Ainsi, la méthode des quadruples la plus contraignante puise dans les réflexions de van Santen sur la combinatoire entre le segmental et les contextes prosodiques dans la parole naturelle. On vise à créer un corpus comportant tous les contextes phonémico-prosodiques qui sont organisés en un sous-ensemble structuré par rapport à l'ensemble de toutes les combinaisons phonémico-prosodiques possibles. En relation avec cet objectif, les auteurs introduisent le concept des séquences phonémiques équivalentes : c'est le cas de deux séquences comportant le même nombre d'éléments, et dont les éléments respectifs appartiennent à la même classe du point de vue de la production de la parole. Il est donc nécessaire de disposer dans le corpus des couples (u_2, c_1), (u_2, c_2) et (u_1, c_1) pour obtenir (u_1, c_2) souhaité (où u_i correspond à une séquence phonique, c_i indique le contexte prosodique, et le couple (u_i, c_i) correspond à leur combinaison). On procédera alors par transposition des différences entre (u_2, c_1) et (u_2, c_2) sur (u_1, c_1).

La deuxième méthodologie n'impose pas de contraintes sur le corpus : elle présuppose la recherche des séquences multiples d'unités prosodiques, l'extraction de la courbe de f0 à partir de ces séquences, et leur recombinaison postérieure afin d'aboutir à un contour mélodique visé ; les auteurs parlent alors de la méthode superpositionnelle d'extraction des unités, en relation avec le modèle de l'organisation intonative de van Santen.

Un prétraitement du corpus est alors nécessaire, à savoir, son annotation par application d'un schéma hiérarchique comportant les constituants principaux suivants : phonème, syllabe, mot, pied, syntagme, énoncé et paragraphe (Klabbers et van Santen, 2003). En outre, chaque unité ainsi désignée, reçoit une étiquette en relation avec sa position par rapport à la distribution des proéminences ainsi que par référence à l'unité du niveau supérieur ; on précise également la taille de l'unité en constituants internes, cf.

exemple : un pied peut être caractérisé comme terminal dans un syntagme médian de l'énoncé, comportant 3 syllabes, *etc.* Pourtant, le travail cité est réalisé à partir d'un corpus lu ; si l'on vise à transférer la méthodologie proposée vers un corpus de parole spontanée, il est nécessaire de définir les unités équivalant aux énoncés et aux paragraphes dans le discours oral, cette tâche n'étant pas à négliger.

Pour une application de cette méthodologie, il est nécessaire que le module de l'analyse du texte produise une annotation multiniveaux du texte d'entrée, en plus de l'annotation des événements prosodiques traditionnels (phrasé + distribution des proéminences). À partir de cette annotation le module de synthèse cherche les séquences d'unités :

a) correspondant à la séquence-cible phonémique ;
b) correspondant à la séquence-cible en pieds : le pied contient le même nombre de phonèmes, et chaque phonème des deux séquences appartient à la même classe phonémique large (voyelle, diphtongue, consonne sonore et autre) ;
c) correspondant à la structure syntagmatique : en fonction du nombre et de la longueur (exprimée en syllabes) des pieds ;
d) correspondant à la structure de l'énoncé en termes de nombre et de la longueur des syntagmes ; *etc.*

On calcule en parallèle les différentes mesures des coûts de concaténation et de déviation par rapport à la structure interne du constituant.

Nous nous attachons particulièrement à cette approche dans la mesure où elle permet de tirer les avantages des modèles linguistiques développés pour le traitement des phénomènes prosodiques, notamment, lors de l'annotation du corpus. Le défi qui persiste concerne la modélisation des contextes discursifs dans leur variabilité (les aspects émotionnels, modalité de l'énoncé, distribution du focus).

6.4. RÉSUMÉ ET DISCUSSIONS

Dans le présent chapitre nous avons passé en revue les algorithmes mis en places dans les systèmes de synthèse vocale : ces algorithmes correspondent à différentes étapes du traitement linguistique et de la génération du signal. Dès le début, dans la réflexion sur les méthodes de la génération du signal et dans la discussion de la synthèse concaténative, nous avons constaté la nécessité de prendre en compte la structure prosodique afin de minimiser la quantité de modification ultérieure du signal (cf. l'approche vers la construction d'une base de diphones dans les travaux de de Pijper (1997) et de Klabbers (2000)).

La problématique centrale de cet état de l'art concerne la question des informations phonético-phonologiques pertinentes dans les systèmes appliqués et du lien que ces représentations phonologiques maintiennent avec d'autres niveaux linguistiques (en particulier, avec les niveaux pragmatique et syntaxique). Nous avons revu certaines méthodes dites « d'ingénieurs » qui reposent sur des techniques statistiques puissantes mais qui nous renseignent peu sur les mécanismes cognitifs sollicités par des humains en communication verbale. Pour notre part, nous adhérons à la position défendue dans des travaux de Dik (1997), selon laquelle tout algorithme formel ne doit pas aller à

l'encontre de ce que l'on sait des mécanismes cognitifs de l'homme. Par conséquent, il est nécessaire dans une recherche appliquée de se donner la tâche de respecter (sinon modéliser) ces mécanismes sous-jacentes.

Nous avons observé que la quantité des informations linguistiques dans des systèmes de synthèse varie considérablement (notamment, entre les systèmes TTS et CTS) : à l'étape du traitement et de l'analyse linguistique, un des défis réside dans le développement des méthodologies qui permettent de récupérer le plus d'informations sur l'organisation du texte/message et sur la mise en relation entre le contenu propositionnel défini dans le contexte discursif et la prosodie.

Dans un deuxième temps, nous nous sommes penchée sur la question de prédiction et de gestion de l'organisation prosodique du message en synthèse. Les chapitres IV et V ci-dessus défendent la thèse, selon laquelle la hiérarchie des constituants prosodiques est un cadre structurel qui gère l'implémentation acoustique des phénomènes prosodiques. Par référence à la synthèse vocale, il est nécessaire de préciser que rares sont les systèmes qui reconstituent une hiérarchie exhaustive des constituants prosodiques. La prédiction du phrasé est largement fondée sur les marques de ponctuation et sur les propriétés linguistiques générales des mots et des phrases (contraintes sémantico-syntaxiques et eurythmiques), ce qui permet de spécifier certains constituants de la hiérarchie prosodique, et ceci en fonction de la langue analysée. Dans cette perspective, nous défendons une approche probabiliste du phrasé (cf. Veilleux 1997), fondée sur une large étude des corpus, menée dans la perspective de confronter l'organisation syntaxique et l'annotation indépendante (perceptive) du phrasé prosodique[208]. Cependant, une réflexion approfondie est nécessaire sur la question du transfert des méthodes d'analyse syntaxique développées pour des énoncés simples (des phrases) à des corpus de la parole spontanée, beaucoup plus complexes (nous pensons notamment à des segments résultants des phénomènes de dysfluences, ou à des structures qui s'enchaînent sans avoir reçu un marquage grammatical quelconque). Dans les analyses empiriques qui suivent nous essayerons d'apporter des éléments de réponse au problème évoqué (chapitre IX).

Nous avons ensuite discuté du rôle des corpus oraux annotés pour le développement des applications technologiques (systèmes TTS et systèmes de reconnaissance vocale). Comme nous avons vu, dans les années quatre-vingt dix, un changement de paradigme s'opère dans le domaine de la synthèse concaténative. On s'appuie maintenant sur des grands corpus oraux annotés pour sélectionner les unités pertinentes et les concaténer en des formes plus grandes (clauses, énoncés, paragraphes etc.) : l'idée clé de la synthèse à partir de corpus réside dans l'utilisation de la totalité du corpus annoté comme inventaire acoustique. Les travaux de Black et Taylor (1999), Klabbers et van Santen (2003) démontrent que la prise en compte de la structure prosodique hiérarchique lors de la sélection des unités-candidates diminue les coûts de concaténation (la quantité des modifications ultérieures des propriétés acoustiques des segments). Qui plus est, van Santen *et al.* (2005) proposent le modèle d'une double sélection des unités qui repose sur le principe de la transplantation des courbes mélodiques attestées dans le corpus ; pour les auteurs, la spécification de la courbe s'inscrit dans le modèle superpositionnel de l'intonation (van Santen et Möbius 2000). Cette méthode dissocie la recherche de la

[208] La même approche nous paraît également présenter des avantages pour d'autres phénomènes prosodiques.

courbe mélodique correspondante de la recherche des unités segmentales : le lien entre les deux est assuré par une annotation prosodique du corpus.

C'est en gardant à l'esprit les présupposés de la double sélection des unités et nos intuitions sur le besoin d'une approche probabiliste du phrasé prosodique que nous avons mené des analyses expérimentales, destinées à contribuer à une description formelle de l'intonation du russe. Nous nous tournons maintenant vers la présentation des analyses empiriques qui sous-tendent notre travail.

PARTIE II

ANALYSES EMPIRIQUES

CHAPITRE VII
CHOIX DU CORPUS ET ANNOTATIONS

7.1. INTRODUCTION

Dans les chapitres qui précèdent, après avoir fait un historique des études analysant l'intonation du russe, nous avons discuté les conceptions modernes du domaine prosodique, notamment dans la perspective d'une analyse formelle, et nous avons fait un certain nombre de choix méthodologiques dans le cadre posé. Nous nous tournons dès maintenant vers les descriptions empiriques qui fondent notre travail.

Pour proposer une description formelle de l'intonation du russe et mettre en évidence les détails de la manifestation acoustique des phénomènes intonatifs nous nous situons dans le paradigme de la linguistique expérimentale : ce terme générique fait référence à la fois au travail d'analyse et d'annotation d'un corpus oral ainsi qu'aux tests de perception réalisés dans le cadre de l'étude entreprise. Il se pose donc le problème du choix des données : la réflexion sur les données appropriées est importante dans une étude expérimentale et doit prendre en compte l'orientation de l'étude en cours. Les précisions sur ce point nous paraissent d'autant plus nécessaires qu'il convient d'anticiper certaines des critiques méthodologiques qui pourront être faites aux résultats que nous présenterons dans la suite.

Lors de la présentation des données sur lesquelles une étude expérimentale est menée, deux aspects doivent être précisés : le type de données choisi et les modalités de leur recueil. Dans cette perspective, nous pouvons constater une relation orthogonale entre deux axes qui permettent de décrire les types de données, à savoir :

a) le continuum constitué entre les données construites, fréquemment nommées de parole de laboratoire et de la parole attestée ;

b) l'axe de spontanéité[209] (cf. la figure 7.1).

[209] Remarquons que Di Cristo (2000b) en attirant l'attention sur la polysémie du concept de parole spontanée, spécifie deux usages les plus fréquents de ce terme qui se situent au long des axes mentionnés : on oppose à la parole spontanée d'une part, une lecture oralisée, et d'autre part, une parole préparée : « *à cet égard, une lecture improvisée, non préparée à l'avance peut fort bien être considérée comme un exemple de parole spontanée* » (*op. cit.*: 191)

Dans la suite de ce chapitre, nous analysons les avantages et les inconvénients de différents types de données et nous concluons sur leur complémentarité. Travaillant sur de la parole spontanée, nous tenons également à préciser ce que nous entendons par la spontanéité.

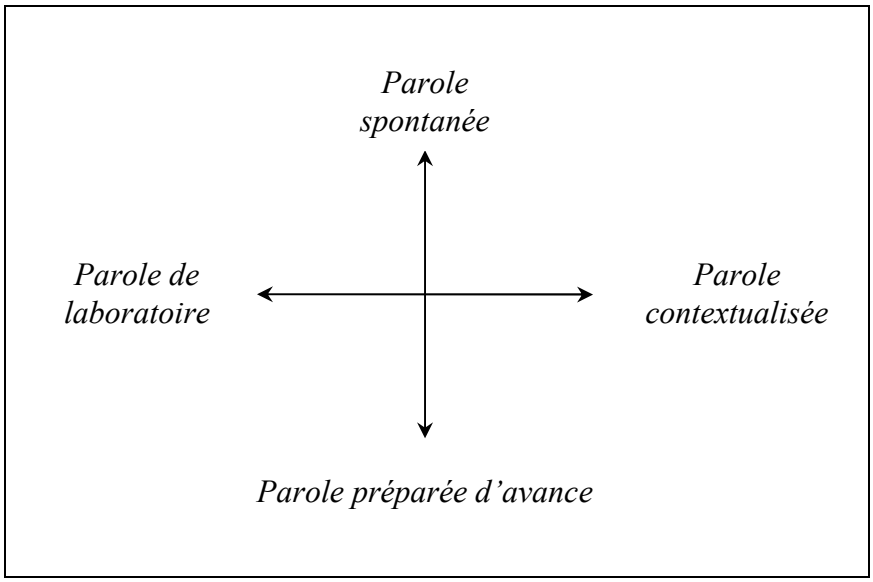


Figure 7.1 : Représentation du continuum des types de données

7.2. CHOIX DU CORPUS

7.2.1. Travailler sur la parole naturelle ou sur la parole de laboratoire

Dans les études en phonétique et en phonologie une distinction importante est faite entre les données issues d'un corpus de parole de laboratoire par opposition aux hypothèses testées dans un corpus de parole attestée/spontanée.

Par parole de laboratoire, on entend généralement les enregistrements effectués dans une chambre anéchoïque, ce qui assure une bonne qualité du signal : le matériau linguistique représente des listes de mots ou de phrases, construites au préalable par l'expérimentateur, avec le contrôle systématique de diverses sources de variabilité. En plus d'une bonne qualité de l'enregistrement, cruciale dans les études acoustiques, un atout majeur de ce type de corpus réside dans la possibilité d'élaborer des stimuli en fonction des hypothèses testées. De sorte que, travaillant à partir d'un corpus authentique, on pourrait avoir besoin de nombreuses heures d'enregistrement pour aboutir à une quantité représentative d'exemples de phénomène étudié, dans un corpus construit ce phénomène est présent de manière condensée ; conjointement, on a la possibilité de contrôler les sources d'influence potentielles afin de neutraliser leurs effets. Cette démarche est propre aux approches de la phonologie de laboratoire (Pierrehumbert *et al*. 2001 ; D'Império, 2005), cadre méthodologique des recherches empiriques qui adopte un raisonnement descendant (théorie ~ empirie) sur les conséquences des principes sous-jacents d'une théorie linguistique.

Pourtant, le désavantage des corpus construits réside dans la restriction stylistique qui leur est propre : très fréquemment, il s'agit de textes/stimuli/listes lues, et non pas de parole

produite naturellement. Précisons également que ces corpus représentent de la parole produite dans un but non communicatif.

Le corpus choisi doit donc s'accorder avec la problématique centrale de la recherche. Sans doute, quand on travaille sur la perception des voyelles médianes, on ne peut mener cette étude qu'à partir d'un corpus préfabriqué enregistré dans des conditions « idéales » (dans une chambre anéchoïque, avec un locuteur modèle). Pourtant, Eskénazi et Lacheret-Dujour (1991) analysent les variantes possibles de la prononciation d'une phrase contenant un /j/ « *que vous sachiez* » : le résultat perceptif étant le même, les quatre locuteurs utilisent différentes combinaisons de procédés articulatoires et acoustiques pour y arriver ; notons que seule une stratégie utilisée correspond à l'algorithme traditionnel implanté dans les systèmes de reconnaissance de parole. Or, nous sommes convaincue que cette variabilité ne peut être écartée mais nécessite, au contraire, une prise en compte systématique : la référence aux corpus de parole attestée permet d'avancer dans cette direction. Nous sommes persuadée que les productions d'un seul locuteur exhibent une variabilité comparable qui pourra être attestée et même expliquée (c'est-à-dire, étudiée par la mise en relation systématique avec des facteurs linguistiques et discursifs), dans la mesure où le corpus de parole naturelle est doté d'une annotation appropriée.

Le groupe opposé est représenté par les corpus de la parole attestée, authentique. Ici, la parole n'est pas préparée à l'avance par l'expérimentateur, la parole est produite dans l'intention de communiquer un message et, selon la situation, les processus de planification et de production se réalisent conjointement (et on en trouve des marques dans le signal sous la forme de différents phénomènes de dysfluences).

Cependant, historiquement la communauté scientifique linguistique interprète les différences entre les données attestées et les données construites dans la perspective des travaux en phonostylistique. Les problématiques des styles de parole et de la spontanéité interférant, nous tiendrons à fournir quelques précisions sur la notion du style de parole et sur les positions prises.

7.2.2. Qu'entend-on par style de parole ?

Réfléchissant aux tendances en recherche expérimentale sur les différents styles de parole, Eskénazi (1994) propose de faire une distinction terminologique importante entre 'type de discours' et 'style de parole'. Le concept de 'type de discours' fait référence à la tâche discursive proposée au locuteur lors d'un enregistrement : on fera alors distinction entre une lecture, une production monologique, une conversation dialogique, etc. Tandis que le 'style de parole' est défini par la situation de communication, le contexte dans lequel se trouvent les interlocuteurs ; les axes pertinents sont alors le degré de familiarité entre les interlocuteurs, le contexte social, etc.

À travers la littérature, on qualifie de variation stylistique les divers phénomènes qui se repartissent en quatre catégories : les qualités de voix, le débit de parole (qui est, peut-être, un élément supplémentaire pour donner l'impression de tel ou tel style), les dimensions de style, comme formel ou moins formel, clair, compréhensible, spontané, répété, normal,

lecture, parole de laboratoire etc. ; les styles et les activités particulières (nouvelles, sport, concert, professionnel, interview) constituent le quatrième groupe

M. Eskénazi (1994) propose une définition généralisante du concept du style, en s'appuyant sur l'analyse de la vaste littérature existant sur le sujet : elle postule que le style reflète la façon dont l'environnement agit sur l'individu et la façon dont l'individu agit sur l'environnement. Dans son choix du style le locuteur s'appuie sur la perception de différents statuts de ses interlocuteurs et du type de situation de conversation ; on y trouve un mélange d'efforts conscients et inconscients de sa part et parfois le sujet parlant ne réussit pas sa communication, son message étant interprété différemment par rapport à l'intention première.

Dans la suite de son analyse, l'auteur définit les axes permettant de mieux déceler la nature du style de parole : ces axes organisent dans un système plus ou moins cohérent le continuum stylistique présent dans notre discours. Il y en a trois :

- axe d'intelligibilité, degré de clarté[210]
- axe de familiarité dans les relations entre interlocuteurs
- strate sociale : il s'agit du degré de culture attribuée au discours par des interlocuteurs ; ici le contraste le plus évident est celui qui apparaît entre le style parlé de la classe populaire et le ton snob de la haute société.

L'auteur propose un schéma tridimensionnel où sont situés certains styles de parole qui ont été étudiés par différents chercheurs (cf. figure 7.2).

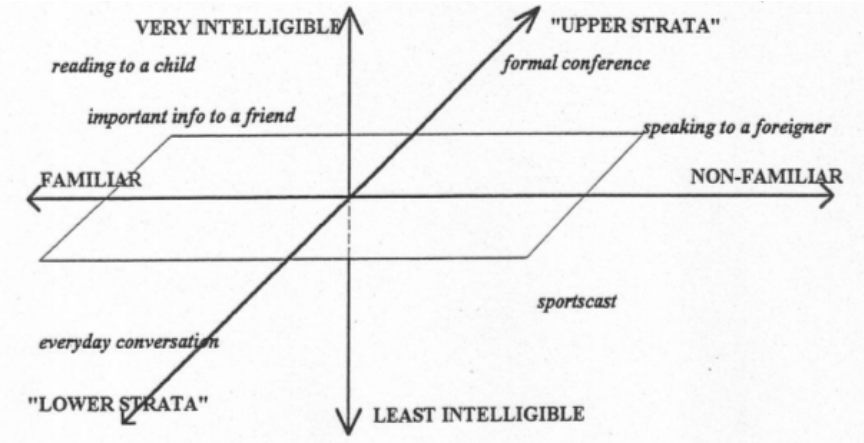


Figure 7.2 : Organisation du continuum stylistique selon trois dimensions (emprunté à Eskénazi 1994 :503)

7.2.3. Styles de parole et spontanéité

Dans le cadre interprétatif proposé par Eskénazi, nous nous associons également à la démarche consistant à caractériser la spontanéité d'une production verbale sous la forme d'une échelle continue, et non pas par référence à une catégorie bien cernée : les pôles

[210] Cf. à cet égard le modèle Hypo & Hyper de Lindblom (1990, 1995).

d'une telle échelle étant une lecture préparée *versus* une conversation familière. Cette continuité a ses impacts, dirigeant les choix méthodologiques et établissant les limites de l'étude, dont l'investigateur doit être conscient : comme le remarque Lehiste (1963:353)

> « … *the linguist who wants to use experimental methods in his investigation is forced to trade a certain amount of naturalness for rigorousness in his experimental design… The two requirements appear to occupy the opposite ends of a continuous scale of gradual mutual exclusion. At one extreme, we find complete naturalness coupled with complete lack of control; at the other end, complete control over the experimental design at the expense of linguistic relevance.* »

Analysant les différentes interprétations possibles de la notion de spontanéité, Di Cristo (2000b : 191) avance l'idée que la spontanéité est une dimension orthogonale à la notion du style de parole. Ainsi, on distinguera une lecture informative et une lecture interprétative, cette dernière mettant en œuvre la compétence du locuteur qui pourrait être qualifiée de stylistique.

Dans la perspective présentée, la notion de 'spontanéité' présuppose une approche multidimensionnelle : en plus de la modalité de production de parole (lecture *versus* expression libre), on doit prendre en compte les aspects dépendants de la situation de production et les contraintes imposées (parole privée/publique, degré de familiarité, enregistrement à l'insu du locuteur etc.), mais également les capacités socioculturelles du locuteur (compétences rhétoriques, niveau d'éducation, etc.).

7.2.4. Complémentarité épistémologique des différents types de données

Nous avons présenté dans les sections précédentes les différents approches utilisées dans des recherches expérimentales. Remarquons que certains projets de recherche sur la prononciation et les variations régionales (Phonologie du français contemporain, PFC, Durand, Laks et Lyche 2002, Intonation Variation in English, IViE, Grabe et Post 2002, entre autres) utilisent deux types de données : on combine la lecture de textes et des listes de mots avec une conversation libre. Lors de la réalisation du projet IViE, environ 36 heures d'enregistrements ont été effectuées : cette base vocale contient les enregistrements représentatifs des neufs variantes régionales de l'anglais britannique. Les différentes tâches ont été proposées aux participants à l'étape de collecte des données ; on peut les regrouper en trois types : lecture (des phrases isolées et d'un texte, Cendrillon), parole semi-spontanée (narration du contenu du texte lu) et parole spontanée (les interactions recueillies dans le paradigme de Map Task). Le même protocole expérimental a été adopté lors d'une enquête dialectologique menée en Union Soviétique dans les années 1970. Ainsi, tous ces projets visent une étude comparative des différentes variantes régionales et proposent donc un vaste panorama de données exploitables aux chercheurs dont les thèmes de recherche divergent.

Dans ce débat sur le statut et la validité des données dans les études phonétiques expérimentales, il est légitime de poser la question des objectifs auxquels se rapportent ces données. Ainsi, de manière générale, tout investigateur se penche sur les données observables soit pour construire une hypothèse de recherche, soit pour valider une

hypothèse issue de l'analyse d'autres types de données et/ou de réflexions théoriques. On distingue donc les objets d'observation et les objets de validation[211] : les deux peuvent provenir des corpus fabriqués comme de la parole attestée. Le seul contexte dans lequel la parole naturelle est préférable est le cadre d'une étude discursive et conversationnelle avec l'accent mis sur les aspects interactionnels. Ainsi Beckman (1997) conclut son plaidoyer en faveur de la parole spontanée :

> *« I have spent so much time motivating why we want to look at spontaneous speech for a better modelling of how focus domain relates to accentuation in English because this is a good example of how lab speech has served us well and of how it fails to serve our needs completely. »*

Dans la même lignée, Lickley *et al.* (2005) défendent les approches de la phonologie de laboratoire :

> *« Doubts are sometimes raised about the "ecological validity" of conclusions based on the analysis of intonation in controlled sentences read aloud under laboratory conditions, and more generally about the intonation of "decontextualised" sentences. We argue that these concerns, though by no means entirely misplaced, are not sufficiently serious to discredit laboratory phonology methods*. »

Ils remarquent, entre autres, que les différences entre la parole de laboratoire et la parole naturelle, spontanée, sont levées dès que l'on atteint les niveaux d'implémentation phonologique et phonétique de le parole (cf. le modèle de Levelt, 1989),

Pour notre part, nous considérons que l'attention aux données contextualisées, à la parole spontanée, ne remet pas en cause les principes fondateurs de l'approche de la phonologie de laboratoire. Dans l'analyse du corpus, on valide les mêmes hypothèses phonologiques (ou autres) à travers l'observation d'une masse considérable des productions langagières, on arrive même à se faire une impression sur la fréquence d'un phénomène donné, et, donc, sur sa productivité et sa valeur cognitive (c'est-à-dire, le poids des connaissances probabilistes dont disposent les locuteurs), points omis dans les travaux à partir des données construites. Le point méthodologique crucial consiste donc à disposer de corpus suffisamment grands dotés d'une annotation multi-niveaux et d'outils automatisés d'annotation et d'analyse : l'exploitation de cette piste permettra de lever la controverse entre parole de laboratoire et parole naturelle.

Nous avons abordé la problématique de la spontanéité d'une production verbale en relation avec le concept du style de parole. En réfléchissant sur la valeur épistémologique des données, nous considérons comme important de prendre en considération la dimension de la variabilité inhérente au concept de style de parole : ainsi le style est le choix parmi plusieurs possibilités fournies par la langue. Gumpertz (1971) utilise le concept de répertoire verbal qui est défini comme l'ensemble de ressources langagières à la disposition d'une communauté linguistique et partagé par tous les membres à des degrés différents. Une telle approche s'accorde avec une vision phonologique des effets stylistiques chez Troubetzkoj (1939) : dans la réflexion de Troubetzkoj, parmi les

[211] Cf. en technologies vocales, la distinction systématique faite entre le corpus d'entraînement et le corpus de test.

impressions phoniques il y en a qui, pour être exactement comprises, doivent être rapportées à des normes déterminées, établies dans la langue en question. Ces normes sont à considérer comme des valeurs linguistiques, et appartiennent, donc, à la langue, et la phonologie doit par conséquent les traiter : l'objectif de la phonologie est d'étudier les fonctions de la face phonique du langage humain, elle ne peut pas restreindre son objet d'études à une seule fonction représentative.

Dans cette perspective, l'école phonologique de Saint-Pétersbourg défend depuis des décennies le concept de la phonologie de l'activité langagière (« *фонология речевой деятельности* », cf. Bondarko 1981, Bondarko *et al.* 2000). À l'issue des études expérimentales se penchant sur les capacités physiologiques et cognitives de l'homme qui se manifestent en production et en perception de la parole, on est amené à reconnaître un lien complexe qui existe entre le système de la langue, une organisation substantielle et structurée, et une organisation beaucoup moins ordonnée des unités de parole : les cas de l'accord parfait ainsi que les divergences doivent être prises en compte dans la modélisation du système de la langue. L'analyse des propriétés réelles de la parole démontre comment les résultats de ces études entrent en conflit avec les schèmes phonologiques classiques.

Nous adhérons complètement à la position de la phonologie de l'activité langagière, et nous considérons que, dans cette perspective, les va-et-vient systématiques entre les données construites et la parole attestée sont une méthodologie à adopter.

7.2.5. Comment recueillir de la parole spontanée ?

Afin de recueillir de la parole spontanée les méthodes sont variables : on peut organiser des jeux de rôle et des débats entre les participants qui seront ensuite enregistrés.

Les enregistrements peuvent alors avoir lieu dans une chambre anéchoïque ce qui permettra d'avoir des enregistrements de bonne qualité, avec des micro-casques individuels et des magnétophones multipistes pour traiter les chevauchements éventuels (source intéressante de données, surtout dans une perspective discursive et pour une problématique de gestion de tours de parole). Toutefois, on se retrouve rapidement dans un environnement artificiel pour les locuteurs qui peut les gêner et empêcher des productions naturelles sans autorégulation.

Dans l'objectif de recueillir des données authentiques, une méthode de *Map Task* a été proposée par Brown *et al.* (1984). Deux sujets reçoivent chacun une carte géographique mais avec des divergences (présence/ absence de certains repères, changements de noms, modification des distances, *etc*.). Le trajet est tracé sur l'une de ces cartes qui, à la fin de l'interaction, doit être reproduit sur la deuxième carte. Les sujets peuvent être placés face à face ou bien être séparés par un écran ; on agit également sur le fait que les deux participants se connaissent ou non, ce qui permet à l'expérimentateur de varier le degré de familiarité dans les interactions verbales. Les sujets participant à l'expérience débattent donc du trajet possible. L'avantage de cette méthode est que, même si les participants sont

conscients d'être enregistrés, ils se retrouvent rapidement en situation de communication naturelle, leur interaction ayant un objectif.

Dans la constitution du corpus CID (projet en cours au Laboratoire Parole et Language), les enregistrements ont été menés dans la chambre sourde du laboratoire, pourtant les sujets avaient pour consigne de raconter des histoires ayant une même tonalité (des histoires insolites entre autres).

Les enregistrements radiophoniques représentent également une source potentielle de données orales authentiques, préférées dans les études prosodiques pour le caractère marqué des structures. Or, ici la variabilité stylistique dépend du type d'intervention et doit nécessairement être prise en compte : si dans les interviews on observe un style parfaitement naturel de l'interaction, d'autres types d'interventions se rapprochent plutôt de la lecture (les actualités). On peut retrouver dans les corpus radiophoniques des enregistrements situés entre ces deux extrêmes, qualifiés et perçus comme naturels.

Après avoir discuté des points méthodologiques associés avec la problématique du choix de corpus, nous présentons les données sur lesquelles nous avons fondé notre étude.

7.2.6. Présentation du corpus

L'intégralité du corpus dont une partie a été utilisée dans le cadre du présent projet de recherche a été recueillie au laboratoire de phonétique de l'université d'État de Saint-Pétersbourg lors du travail sur un projet de recherche international *« Spontaneous Speech of Typologically Inrelated Languages (Russian, Finnish and Dutch) : Comparison of phonetic Properties »* (projet subventionné par programme INTAS # 00-0915). Ce projet a été conçu dans une perspective comparative des propriétés phonétiques de la parole spontanée et de la lecture : cette orientation de projet a déterminé les modalités de recueil des données orales.

Le recueil des données comportait deux phases : d'abord, des dialogues spontanés ont été provoqués chez 10 locuteurs (5 hommes et 5 femmes) et enregistrés dans une chambre anéchoïque[212]. Les locuteurs avaient pour interlocuteur une personne connue pour induire de la parole plus familière ; un sujet neutre leur a été proposé au début (« Les vacances passées »), ensuite, ils étaient libres du choix des thèmes de conversation. Les enregistrements ont eu lieu avec un équipement professionnel ; les locuteurs ont porté un micro-casque individuel AKG HSC 200 et le mixeur Behringer 802A a été utilisé.

À la base des enregistrements de parole spontanée, les investigateurs ont sélectionné des fragments monologiques sans dysfluences (pauses d'hésitations, reprises, allongements non-structuraux) qui ont été convertis en un texte homogène. Ce texte a été proposé aux locuteurs en lecture lors de la deuxième phase d'enregistrement[213]. Les deux phases

[212] Dans le cas de figure le plus fréquent les propos d'un seul locuteur ont été inclus dans le corpus d'analyse.
[213] D'autres types de matériau enregistré incluent la lecture de mots isolés et de phrases issues d'enregistrements spontanés, ainsi que la lecture du texte représentatif de la langue russe (Stepanova 1988).

d'enregistrement ont été espacées d'une semaine. Les enregistrements ont eu lieu entre janvier et juin 2002.

La méthodologie décrite a été adoptée dans l'objectif de disposer des enregistrements de qualité permettant une analyse acoustique fine. Conjointement, le contenu segmental est le même dans les enregistrements des deux types de discours ; les caractéristiques acoustiques des différents phénomènes prosodiques et des segments peuvent donc être étudiées sans recours à une procédure de normalisation. Dans le développement du projet, seuls les passages choisis des enregistrements spontanés et leurs homologues en lecture ont été soumis à une segmentation phonémique et ont reçu une annotation et une analyse.

Pour notre part, nous avons choisi de travailler sur la totalité des propos, recueillis lors de la phase spontanée, d'une locutrice, N, âgée de 20 ans, étudiante à la faculté de philologie à l'université d'État de Saint-Pétersbourg.

La totalité du matériau spontané recueilli couvre 29,13 minutes de parole. Pourtant, nous avons restreint le matériau analysé dans le cadre de la présente étude sur la base des considérations suivantes : le corpus dont nous disposons est un corpus de parole spontanée contrainte. Certes, il s'agit d'une interaction entre locutrices et le discours est produit dans un objectif communicatif. Pourtant, les conditions d'enregistrement sont relativement artificielles, il n'est donc pas évident de parler sans à-coups[214] ou même de trouver un sujet de conversation. Nous avons, par conséquent, exclu de l'analyse des passages que nous avons jugés comme artificiels. Également, de longs passages de silences, correspondant aux interventions prolongées de l'interlocutrice n'ont pas été pris en compte. Au final, nous disposons d'un corpus de 17 minutes (temps de parole, pauses comprises) sur lequel nous avons fondé la présente étude.

Nous sommes consciente des limites de notre corpus dont la taille peut inciter à des critiques Notons pourtant que l'annotation du corpus et l'alignement au niveau phonémique ont été effectuées manuellement, ce qui représente une tâche très prenante et très longue. Simultanément, nous visons à proposer une approche généralisable sur d'autres types de données et nous concevons le travail effectué comme une étude exploratoire et évaluatrice de différents choix méthodologiques à prendre ou à laisser. Il s'agit donc tout d'abord de proposer un modèle de données. Le transfert de la méthodologie proposée sur un ensemble plus vaste des données fait partie de nos projets futurs.

Nous avons choisi de travailler sur la totalité des enregistrements en espérant aboutir à une base de données enrichie comportant des annotations multiniveaux. De manière générale, le terme d'« annotation linguistique » couvre le procédé d'une notation descriptive ou analytique des données linguistiques brutes : ainsi il s'agit d'un travail de dotation d'un objet linguistique préexistant d'une description symbolique. Dans le contexte le plus général, il peut s'agir de différents types de transcription, du marquage des parties du discours, de la structure syntaxique, annotation des cas de co-référence, etc. Dans la suite de notre travail analytique, le corpus annoté constitue la ressource de référence pour validation d'hypothèses issues d'une étude expérimentale en perception. Il nous paraît

[214] Il ne s'agit pas des phénomènes de dysfluences.

donc important de spécifier les niveaux d'annotation et les formalismes choisis pour chacun d'entre eux : cela constitue le sujet de discussion du paragraphe suivant.

7.3. ANNOTATION DU CORPUS

Nous concevons le corpus oral comme un ensemble d'enregistrements audio, numérisés, accompagnés d'annotations multiniveaux. Nous considérons que l'annotation du corpus représente une valeur ajoutée importante, toute l'originalité de la démarche résidant dans les principes qui la sous-tendent. Pourtant, l'annotation, telle que nous la concevons, doit communiquer aux données brutes une structuration particulière, dont les fondements doivent être les plus neutres possibles vis-à-vis des théories linguistiques et/ou phonologiques. Simultanément, cette annotation initiale doit mettre en œuvre un cadre dynamique, permettant l'ajout d'informations nouvelles ou bien la formalisation des théories alternatives à celles qui sous-tendent les annotations de base (nous pensons notamment à une étude comparative éventuelle entre l'annotation INTSINT et l'annotation qui s'inspire de celle de IViE-ToBI). Pour exploiter alors toute la structure des données, on a besoin de disposer d'une possibilité de formuler des requêtes ou de récupérer des données brutes observables pour des traitements statistiques postérieurs.

L'annotation multiniveaux s'avère être une ressource inestimable pour l'exploitation des rapports entre les différents niveaux linguistiques, chacun représentant une partie des informations disponibles aux locuteurs en communication. Nous proposons donc une annotation prosodique du corpus en accord avec la conception moderne de la prosodie que nous avons développée dans les chapitres I et II : nous procédons notamment à l'annotation de l'organisation tonale et du phrasé prosodique, l'alignement temporel à chaque niveau retenu de la constituance prosodique rendant exploitable la dimension de l'organisation rythmique et temporelle.

Nous présentons, dans la suite de cette section, les démarches entreprises et les choix retenus.

7.3.1. Remarques générales

Dans les études menées à partir des corpus oraux, on retrouve plusieurs types de transcription de données : le choix d'un type particulier de transcription dépend directement du type d'étude que l'on envisage. Delais-Roussarie (2003:98) propose une typologie des transcriptions possibles pour un corpus oral selon le niveau de représentation central choisi :

- transcription à partir des scripts d'enregistrement : si le corpus consiste en un matériau lu à partir des scripts ;
- transcription orthographique : encodage orthographique de la totalité du contenu du signal ;

- transcription correspondant à une représentation phonémique de citation (*citation phonemic representation*) : on retrouve alors les formes phonologiques sous-jacentes que l'on trouve dans un dictionnaire ;
- transcription phonétique large : c'est une transcription phonémique incluant les phénomènes propres à la production orale tels que la liaison, l'insertion du schwa, les phénomènes de sandhis standardisés *etc*;
- Transcription phonétique étroite : il s'agit de la transcription exacte du signal ;
- Transcription acoustique-phonétique : ce niveau repose sur une segmentation précise du signal (on sépare parfois même des phases articulatoires comme les phases de fermeture et d'explosion d'une plosive non-voisée).

Bien que le niveau de la transcription orthographique puisse paraître neutre, il fait également l'objet de choix et de conventions d'annotation, d'autant plus délicates dans le cas de l'oral spontané. Les questions qui surgissent concernent l'utilisation des signes de ponctuation, des majuscules, l'annotation des pauses remplies, des bruits extralinguistiques (claquements des dents, si observés) etc. (cf. les conventions développées pour la description de français parlé développé au sein du Groupe Aixois de Recherche en syntaxe (Blanche-Benvéniste et Jeanjean 1987). La tradition est de s'abstenir de l'utilisation des marques de ponctuation : or, si l'on envisage une étude de l'interface « prosodie – syntaxe », l'absence des marques de ponctuation ne permet pas d'utiliser des analyseurs syntaxiques automatiques, travaillant à partir des phrases. Disposer de la transcription orthographique se révèle avantageux quand il est possible de se servir par la suite des outils assurant la conversion « graphème – phonème », et éventuellement alignement automatique au niveau phonémique.

Un des problèmes d'actualité dans le domaine de la phonétique expérimentale est le conflit permanent entre le caractère continu du flux de parole et le caractère discret de ses descriptions phonétique et phonologique (cf. Zinder, 1979 : 36). Pourtant, le domaine de la transcription phonétique n'est pas non plus homogène : une distinction systématique doit être faite entre une transcription idéale, une transcription phonétique large ou étroite ; dans la tradition des études phonétiques de l'école saint-pétersbourgeoise, on parle également de transcription perceptive (cf. *infra*). L'objectif de toute transcription est de fixer les détails des réalisations de chaque unité segmentale à l'aide d'un ensemble pré-déterminé d'étiquettes. La distinction entre une transcription idéale, phonétique large ou phonétique étroite réside dans le degré de précision avec lequel on indique l'interaction entre les sons adjacents. Remarquons, que la transcription phonétique large est souvent soutenue par des données langagières, or, très fréquemment il s'agit d'une prononciation modèle d'un mot isolé dont la réalisation dans un contexte stylistique neutre ne manifeste pas de caractéristiques dialectales et/ou individuelles. Cependant, la parole en adéquation avec son contexte (et tel est le corpus de notre étude) est caractérisée par des phénomènes de réduction ; mais également, dans le flux de la parole, on observe une quantité importante de segments dont les réalisations concrètes ne sont pas conformes au modèle langagier. Dans la tradition russe, on distingue alors deux types de prononciation, prononciation complète par opposition à prononciation incomplète *(« полный и неполный типы произнесения »)* : pour la réalisation incomplète, on a besoin d'informations complémentaires (contextuelles, grammatiques, statistiques) pour prendre une décision sur

le contenu segmental du mot (cf. Bondarko, Verbitskaja, Gordina *et al.* 1974). Toutefois, une transcription phonétique, pour être opérationnelle dans le traitement automatique, doit nécessairement tenir compte des phénomènes particuliers propres à l'oral, tels que : les élisions de phonèmes, les réductions, les assimilations, les métathèses, *etc*.

Conjointement, on constate que le nombre de catégories d'unités segmentales observées dans la production langagière est plus grand que le nombre de phonèmes dans cette langue[215]. Les études expérimentales en perception (cf. Bondarko 1981, 74-96 ; 127-135) démontrent que les locuteurs de la langue russe sont capables de distinguer plus de contrastes que le nombre d'oppositions pertinents dans le système de la langue[216].

Dans un discours spontané informel, le contenu prime sur la forme, et ceci a pour conséquence le renforcement des processus de coarticulation : on constate alors des réalisations allophoniques non-conformes au modèle de la langue fondée sur la prononciation de mots isolés. Remarquons que les caractéristiques acoustiques sont modifiées non seulement pour les allophones dans des positions faibles mais également pour les positions fortes.

Une transcription perceptive résulte, quant à elle, de l'évaluation par des experts phonéticiens de la prononciation des unités de la taille d'un mot[217]. Le tableau VII.1 résume les rapports entre une transcription phonétique étroite, implémentée sous la forme d'un phonétiseur automatique, et une transcription perceptive réalisée par des experts (les données portent sur de la lecture) : on constate donc qu'une annotation automatique sur la base de règles formulées par des linguistes utilise entre 5 et 15 symboles différents pour annoter un son ; or, un phonéticien dans la tâche perceptive n'en utilise pas moins de 10, la limite supérieure étant de plus de 90.

PHONÈME	PHONÉTISEUR AUTOMATIQUE	TRANSCRIPTION PERCEPTIVE
a	15	92
o	10	76
u	7	25
e	8	27
i	5	11
1	6	15

Tableau VII.1 : Nombre de symboles utilisés pour chaque phonème par le phonétiseur automatique et par des phonéticiens dans la tâche de la transcription perceptive (emprunté à Bondarko, 2000 :31)

Outre les décisions à prendre sur le type de transcription à adopter, il reste la question de l'alphabet utilisé. Comme nous l'avons suggéré dans les discussions sur la variabilité des systèmes de transcription pour les phénomènes prosodiques, nous disposons d'un outil standardisé pour la transcription au niveau segmental, à savoir l'Alphabet Phonétique

[215] Par conséquent, la compréhension des mécanismes qui permettent identifier l'appartenance phonémique des segments ambigus est une des branches importantes de recherche.
[216] Le système phonologique de la langue russe comporte 6 phonèmes vocaliques et 36 phonèmes consonantiques auxquels correspondent 18 allophones vocaliques et 41 allophones consonantiques.
[217] Par opposition à une transcription phonétique étroite on ne se sert pas du signal acoustique, l'annotation est fondée sur la perception

International (API). Or, en vue d'un traitement automatique des données ainsi que d'une transférabilité et d'échanges de données, l'adoption de l'API pose un problème technique, étant donné l'exigence de l'installation d'une police spéciale. Dans cette perspective, l'utilisation de l'alphabet SAMPA (*Speech Assessment Methods Phonetic Alphabet,* Wells 1997) est souvent préférable : SAMPA est constitué uniquement de symboles ASCII et donc, compatible avec toutes les plate-formes sans installation préalable d'une police spécifique (cf. liste des symboles SAMPA avec leur équivalent en IPA en Annexe 1).

7.3.2. Traitement préalable du corpus : aspects acoustiques et alignement

Afin de pouvoir procéder à l'analyse de l'organisation mélodique du parler spontané, nous avons procédé à un traitement préalable des enregistrements qui comporte plusieurs étapes : ce traitement a permis de formuler un cadre général d'analyse et ensuite de pratiquer différentes mesures acoustiques. Nous présentons ci-dessous les détails des traitements effectués.

7.3.2.1. Numérisation

Un fois l'enregistrement terminé, il fallait procéder à une conversion des données analogiques en données numériques. La numérisation suit une fréquence d'échantillonnage retenue de 44100 Hz, avec un encodage mono sur 16 bits.

7.3.2.2. Segmentation phonémique

Nous avons commencé le traitement de notre corpus en accomplissant une tâche d'alignement phonémique : nous considérons que le niveau phonémique est le pivot et la base pour l'alignement d'autres niveaux de l'analyse linguistique que nous proposons pour le corpus (cf. l'interprétation du niveau de squelette en phonologie autosegmentale).

Ce passage est formalisé sous la forme d'une transcription. Effectuer une transcription correspond à un processus d'analyse du matériau sonore, lors duquel on procède à une segmentation conventionnelle du flux de parole en des segments que l'on croit être les réalisations des allophones de phonèmes propres à la langue étudiée. Or, lors de la transcription on n'a pas besoin d'indiquer précisément les frontières de ces segments, la priorité étant accordée à l'ordre linéaire des unités segmentales.

Faute de disposer d'un aligneur automatique[218], nous avons procédé à une segmentation manuelle du corpus en phonèmes bien que cette tâche prenne beaucoup de temps. Ce choix méthodologique est fondé sur les considérations que les méthodes d'alignement automatique n'ont pas été exhaustivement testées sur de la parole spontanée. Un

[218] Se referer à Malfrère *et al*. (2003) pour la discussion de deux types d'aligneurs utilisés à l'heure actuelle. Notons que pour la méthode analyse-synthèse, les traitements sont fondés sur l'utilisation d'une base de données MBROLA, pourtant il n'en existe pas pour la langue russe.

phonétiseur automatique a été développé dans le département de phonétique et il a testé et amélioré par rapport à la transcription *d'un texte lu* (Bondarko et al. 1993) : or, travaillant avec un corpus de parole spontanée, il était difficile de mesurer le niveau de performance du phonétiseur, ainsi que la quantité du travail de correction requise.

Nous avons suivi les principes de segmentation décrits par Skrelin (1999), la démarche étant fondée sur une observation systématique de spectrogramme et d'oscillogramme, accompagnée d'une vérification auditive.

Les cas les plus litigieux sont :

- les occlusives sourdes à l'initiale : nous avons appliqué une contrainte temporelle stipulant la durée de l'occlusion supérieure à 30 ms, mais dépassant jamais 100 ms (cf. les données sur la durée moyenne des phonèmes dans le corpus présenté dans l'Annexe 2) :

- séquence d'une voyelle suivie d'une consonne nasale ou liquide : nous nous sommes fondée dans un premier temps sur les impressions auditives fondées toutefois sur l'examen du spectrogramme à bande large où nous avons cherché des micro-variations dans le comportement des formants et de l'intensité ;

- les désinences atones des adjectifs : elles peuvent comporter une ou deux syllabes et subissent une réduction systématique. Très fréquemment l'identité du phonème vocalique n'est pas claire et nous avons choisi de transcrire toute la rime comme [ə] ([@] en SAMPA) ; de même, le /j/ final de la désinence du genre masculin n'était transcrit et segmenté que quand nous pouvions en observer des traces physiques sur les supports ;

- les consonnes /v/ et /l/ dans les positions intervocaliques. Nous avons pourtant appliqué le critère temporel, considérant qu'un phonème ne peut avoir une durée inférieure à 25 ms.

- Nous avons pu également constater les cas d'insertion des voyelles atones dans des groupes consonantiques, ce phénomène impliquant une resyllabation.

7.3.2.3. Choix d'étiquettes

En procédant à un alignement manuel au niveau phonémique nous aurions donc pu être très précise dans le choix d'une étiquette à partir de l'inventaire proposé par l'alphabet phonétique international. En même temps, le transcripteur automatique (Bondarko *et al.* 1993) reflète non seulement les allophones combinatoires et positionnels mentionnés dans les travaux en phonétique de la langue russe, mais également, des modifications plus fines des phonèmes, assez souvent ignorées : on précise entre autres, les influences des consonnes palatalisées, antérieures (les voyelles, elles aussi, ont alors dans ce contexte une articulation plus antérieure), labiales et nasales sur les voyelles. Pour les allophones positionnels le phénomène de réduction est au centre de la formalisation : deux degrés de réduction pour le

phonème /a/ ; la distinction entre les allophones pré- et post-accentuels des voyelles /u/, /i/, /ɪ/ ; les signes spéciaux pour les allophones du début et de la fin absolue d'un mot.

Le phonétiseur automatique incorpore les connaissances phonétiques et phonologiques acquises au niveau de la transcription phonétique étroite. Nous avons également mentionné les désaccords entre transcription phonétique étroite et transcription perceptive (acoustico-perceptive, plus exactement, comme toute décision est soutenue par une inspection des caractéristiques acoustiques du son). La figure 7.3 illustre le phénomène dans toute son ampleur. En langue russe, c'est la voyelle médiane basse /a/ qui subit les plus fortes modifications dans le flux de la parole continue : exclusivement pour ce phonème on parle de deux degrés de réductions, toutes les autres voyelles n'en démontrant qu'un seul (Avanessov 1984). On annote la réduction qualitative de ce son /a/, en adoptant des signes différents pour sa transcription dans les syllabes pré- et postaccentuées. Ainsi transcrit-on le /a/ de la syllabe pré-accentuée avec le signe [ʌ], une voyelle postérieure moyenne basse non-arrondie. Or, comme l'illustre la figure VI.2, dans une étude acoustico-perceptive, ce son est très fréquemment réalisé comme une voyelle antérieure [ɛ], parfois comme des voyelles supérieures ou comme des voyelles arrondies antérieures comme postérieures.

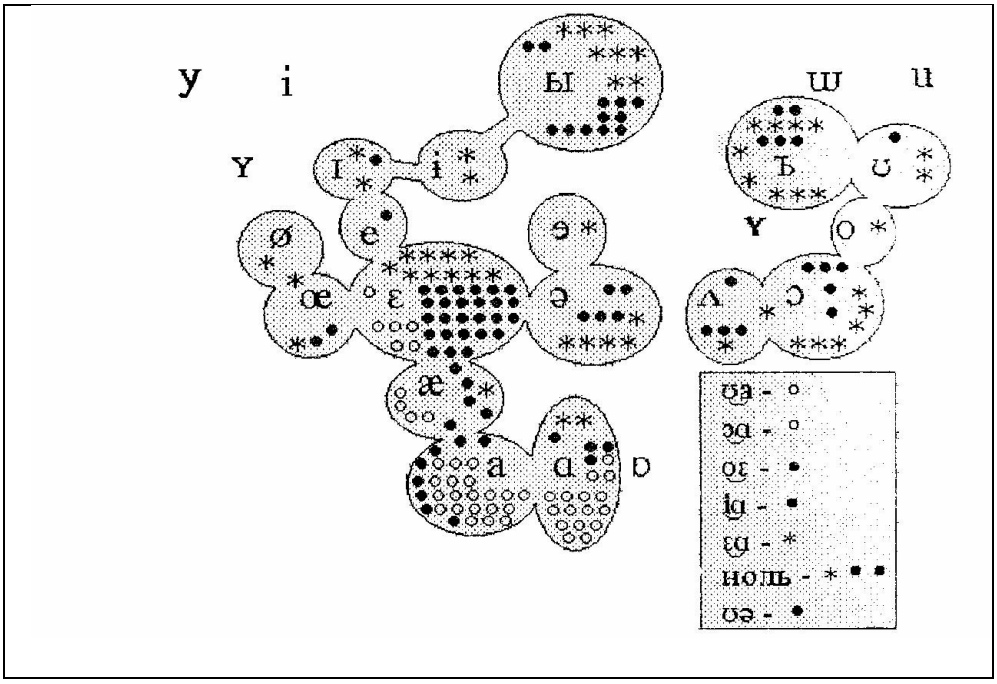


Figure 7.3 : Variabilité dans les réalisations de la voyelle /a/, données issues d'une analyse perceptive (empruntée à Kouznetsov, 2000 : 48)

La transcription phonémique étroite étant un travail intéressant en soi mais très couteux en temps, nous avons décidé de procéder plutôt à une transcription phonétique large, quasi-allophonique : nous parlons d'une transcription quasi-allophonique car nous nous sommes fiée plus à notre perception du son, qu'aux principes phonologiques généraux. Nous adoptons donc 7 classes de voyelles (les variations contextuelles des consonnes étant moins problématiques à annoter), cf. le tableau VII.2.

GROUPE	RÉALISATIONS
/a/	[ɑ, æ, a, ʌ, ɤ, ɐ], combinaisons /a/+j, /a/+V, /a/+ sonnante (en fonction de la réalisation)
/i/	[ɪ, i, j], combinaisons /i/+j, /i/+V, /i/+ sonnante (en fonction de la réalisation)
/o/	[o, ø, œ, œ, ɔ], combinaisons /o/+j, /o/+V, /o/+ sonnante (en fonction de la réalisation)
/u/	[ʊ, u, ɯ, ʏ, y], combinaisons /u/ + V
/e/	[ɛ, e], combinaisons /e/ + V
/ɨ/	[ɨ, ʉ], certaines réalisations de /a/, /u/
/ə/	[ɜ, ə, 9], certaines réalisations de /a/, /u/, /o/, /i/, /e/

Tableau VII.2 : Liste de sept classes de symboles vocaliques utilisées lors d'annotation du corpus au niveau phonémique

Nous présentons les informations sur les durées moyennes des classes de phonèmes et les écarts types dans l'Annexe 2, ces valeurs provenant de la totalité de notre corpus.

7.3.3. Autres niveaux d'annotation

Une fois l'alignement réalisé au niveau phonémique, nous avons procédé à un regroupement des phonèmes en unités supérieures, à savoir, en syllabes, en mots phonétiques, en unités intonatives. Nous avons également appliqué l'algorithme MOMEL et le codage de points cibles en sortie avec les étiquettes de l'alphabet INTSINT.

7.3.3.1. Regroupement en syllabes

Le niveau immédiatement supérieur au niveau phonémique est le niveau syllabique. Nous discutons dans la présente section, de principes qui ont guidé notre démarche dans le regroupement de phonèmes en syllabes.

En discutant de la syllabe en tant que constituant de la hiérarchie prosodique nous avons mentionné les principes de la syllabation proposés par Avanessov (1954) et Scherba (1963) : précisons que ces principes de syllabation font une référence explicite aux frontières de mots. Nous avons également mentionné une forte tendance en russe à former des syllabes ouvertes (nous pouvons y voir l'actualisation du principe de l'attaque maximale) et une tendance à la restructuration des mots phonologiques à l'oral.

Nous avons décidé de suivre au maximum le principe de la formation des syllabes ouvertes lorsque nous avons procédé au regroupement des segments en syllabes. Selon les hypothèses phonologiques, les groupes de consonnes résultants doivent obéir au principe de la sonorité croissante : ce principe n'a pas toujours été respecté dans l'annotation du corpus. Nous avons plutôt respecté le principe phonotactique selon lequel un groupe de

consonnes est autorisé au début de syllabe s'il est autorisé au début de mot. Un deuxième critère a été de nature perceptive : la syllabe résultante devait sonner juste, sans rupture interne perceptible.

Nous avons appliqué un traitement spécifique aux cas de chute des voyelles atones, notamment dans des positions post-accentuelles et dans le contexte des consonnes sibilantes (cf. la figure 7.4). Pourtant, ce processus ne provoque pas de resyllabation au niveau du mot, du moins en perception[219] : la consonne devient syllabique et les paramètres temporels (durée de la syllabe) sont conformes aux modèles standards.

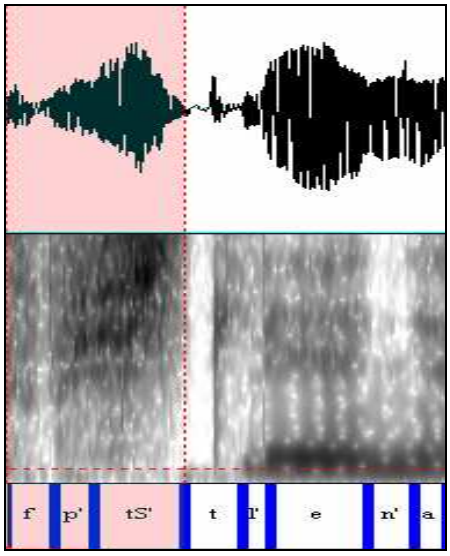


Figure 7.4 : Omission des voyelles atones dans le mot /fpˈɪtʃətlˈenˈɪ/

7.3.3.2. Les mots

Avant de procéder à l'annotation du corpus au niveau de mot nous avons été obligée de prendre des décisions sur les critères qui guideront notre démarche. Nous avons souligné dans la discussion du chapitre IV que la phonologie prosodique classique établit un parallèle entre le constituant du mot dans la hiérarchie prosodique et le mot morpho-syntaxique. À cette étape, aucune distinction n'est faite entre les mots grammaticaux, privés de l'accent lexical dans les positions syntagmatiques faibles, et les mots lexicaux : cette différence est traitée lors de la formation des groupes clitiques, niveau immédiatement supérieur de la hiérarchie. À cette structure bi-niveaux la pensée phonétique russe oppose la distinction systématique entre un mot phonologique et un mot phonétique (ce terme désignant la structure comportant un mot lexical et tous les clitiques qui lui sont associés au niveau phonétique)[220] Nous avons également souligné une forte soumission de l'organisation rythmique du mot à un contexte rythmique et intonatif plus large (révélée dans les travaux de Kassevitchь *op.cit.*). Un troisième paramètre à garder à l'esprit fait référence aux critères qui fondent la formation des syllabes, à savoir le principe

[219] Nous nous posons la question de savoir dans quelle mesure il est possible de parler du mode articulatoire chuchoté, car il est tout de même possible de distinguer deux phases dans l'articulation de la consonne.
[220] Si nous prenons l'exemple d'un énoncé conditionnel « *Я бы уже знал об этом* » (*Je le savais déjà*), dans la représentation morpho-syntaxique la particule **бы** est associée au verbe, tandis qu'elle forme le mot phonétique avec le pronom personnel.

des syllabes ouvertes : il persiste le risque que les frontières des mots ne coïncident pas avec les frontières de constituants syllabiques. Une solution possible serait l'utilisation du concept du mot phonologique restructuré, proposé par Pepperkamp (1999).

Nous avons pourtant opté pour l'annotation du corpus en termes de mots phonétiques de la tradition russe (ou des groupes clitiques) pour les raisons suivantes :

a) l'adoption du concept du mot prosodique restructuré nécessite à notre avis une analyse approfondie des données dont l'objectif sera de recueillir des preuves et de justifier l'introduction de ce concept dans l'analyse de l'organisation prosodique du russe ;

b) l'unité de mot ne rentre pas en jeu dans les analyses effectuées dans le cadre du présent travail ; une fois la décision prise, le re-alignement pourra être effectué de manière automatique.

En conséquence, au niveau du mot nous disposons d'une annotation en termes d'unités proches des groupes clitiques de la phonologie prosodique classique (Nespor & Vogel 1986, Hayes 1989) : de ce choix méthodologique émane le fait que l'on observe dans la représentation multi-niveaux les cas de la non-coïncidence des frontières des mots et de celles des syllabes.

7.3.3.3.Annotation du degré de proéminence

L'annotation du degré de proéminence nous permet de rendre compte de différents niveaux de l'organisation métrique de notre corpus. Nous avons adopté une annotation qui permet de reconstruire une grille métrique à trois niveaux (dans la lignée des propos de Beckman et Edwards 1990). Trois degrés de proéminence sont distingués qui se résument en deux oppositions : premièrement les rapports d'opposition se créent entre les syllabes inaccentuées et les syllabes porteuses de l'accent lexical ; dans un deuxième temps, parmi les syllabes métriquement fortes au niveau du mot, nous différencions celles qui sont porteuses de la proéminence mélodique. Dans cette annotation, nous défendons le principe selon lequel tout mot lexical n'est pas forcément associé avec une proéminence mélodique. Comme nous avons essayé de le démontrer dans les discussions précédentes, le concept de la proéminence mélodique, tel que le conçoit la tradition métrique-autosegmentale, n'a pas été entièrement assimilé dans les études sur l'intonation du russe. Lors de l'annotation nous nous sommes tout particulièrement fiée à notre perception. Ainsi la figure 7.5 correspond à l'annotation du degré de proéminence pour un syntagme nominatif « *очень красивый город* » ((une) *très belle ville*) : nous considérons que c'est seulement le dernier mot qui porte la proéminence mélodique ; aucune autre proéminence mélodique (accent mélodique pré-nucléaire) n'est perçue (pour la même structure Adjectif + Nom Jasinskaja (2001) annote la présence d'une proéminence mélodique prénucléaire sur l'adjectif).

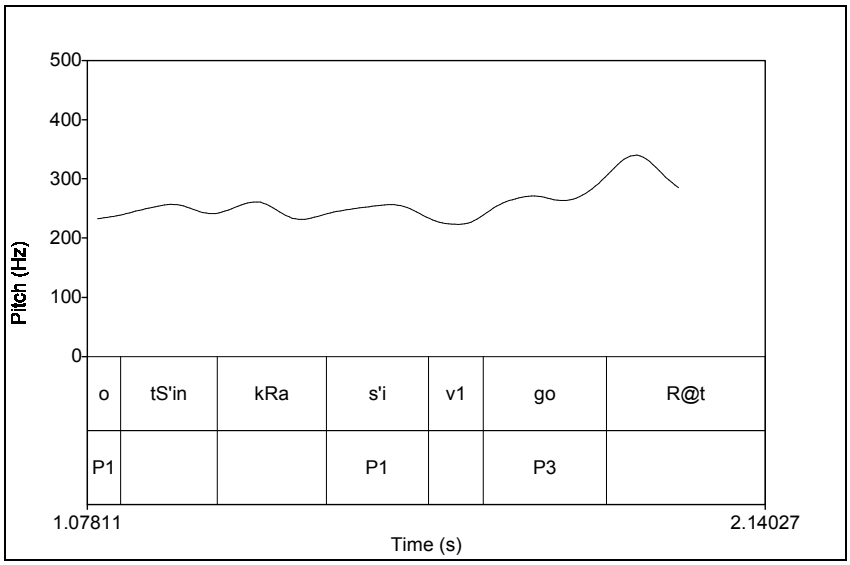


Figure 7.5 : Exemple d'annotation du degré de proéminence

Ce niveau d'annotation étant fondé sur l'analyse perceptive, nous ne pouvons pas l'exploiter à sa juste valeur lors de notre étude expérimentale : nous sommes persuadée que tout schème d'annotation doit obligatoirement être validé par une étude du degré de concordance entre les juges. Ne pouvant disposer à l'heure actuelle d'une autre annotation experte de notre corpus en termes de la distribution des proéminences mélodiques, nous espérons pouvoir exploiter cette dimension plus en profondeur dans des études postérieures.

7.3.3.4. Les unités intonatives

Comme nous l'avons discuté dans les chapitres précédents, l'unité intonative paraît être une unité universellement reconnue dans l'analyse de l'organisation sonore du langage, le concept étant étroitement lié à la fonction structurelle de la prosodie. Il a été proposé une définition formelle par référence aux tons de frontières et à la présence d'une proéminence mélodique nucléaire (centre intonatif). On insiste également sur le rôle de la structure prosodique en tant qu'interface entre d'autres modules langagiers (syntaxique, informationnel, discursif). Dans le même esprit, on constate un lien étroit entre la structuration prosodique du discours et le domaine de la phonostylistique (cf. Zlatooustova *et al*. 1986 : 79) : selon les conditions de production, le même texte peut recevoir une structuration différente. Ces considérations-là nous font adopter une attitude de réticence vis-à-vis des approches qui présupposent l'identification de la structure prosodique sur la base du texte seul.

Une partie de notre corpus a reçu l'annotation en termes de syntagmes prosodiques dans le cadre du projet INTAS mentionné ci-dessus. Nous disposons également d'une annotation du corpus par un expert francophone qui ne maîtrise par le russe : il s'agit donc d'une annotation fondée sur des propriétés formelles (mélodiques et temporelles) des unités résultantes. Nous avons pu constater des cas de divergence dans les représentations résultantes. La figure 7.6 reprend l'annotation d'une unité interpausale « *Ну у дедушки там*

своих забот хватает у него тоже там кошки всякие собака есть[221] » (« *Eh bien il ne lui manque pas de soucis à mon grand-père lui aussi il a là-bas de nombreux chats il y a un chien* »). L'expert français signale trois unités intonatives, il y en a quatre dans l'annotation des spécialistes russes. Une frontière est consensuelle, laquelle paraît séparer deux énoncés ; la dernière frontière paraît se décaler d'un mot dans les deux annotations : du point de vue de la grammaire l'adjectif « *всякие* » forme un syntagme nominal avec le nom « *кошки* » qui porte la proéminence mélodique ; du point de vue de l'organisation prosodique, aucun indice acoustique n'est associé avec la frontière syntaxique en question. L'annotation des experts russes est alors guidée par les connaissances linguistiques générales. Une annotation alternative aurait pu être proposée, établissant une distinction entre la Macro-unité Intonative, composée de l'unité intonative et le segment d'unité intonative (Di Cristo et Hirst 1996). Enfin, nous observons une seule frontière discordante. D'une part, la motivation pour l'annotation de cette frontière peut être linguistique : un syntagme prépositionnel au début de phrase induit la présence d'une frontière. Il est évident qu'une proéminence mélodique est associée avec le premier mot de cette unité de performance ; or, pour respecter la contrainte structurelle de la formation du syntagme intonatif, il est nécessaire qu'un des mots du syntagme résultant « *своих забот хватает* » soit associé avec un accent mélodique. Nous avons évoqué les controverses sur l'annotation du degré de proéminence dans le paragraphe consacré à ce sujet : à notre avis, il est donc impossible à l'étape actuelle de prendre une décision sur la présence d'une frontière prosodique dans la position désignée.

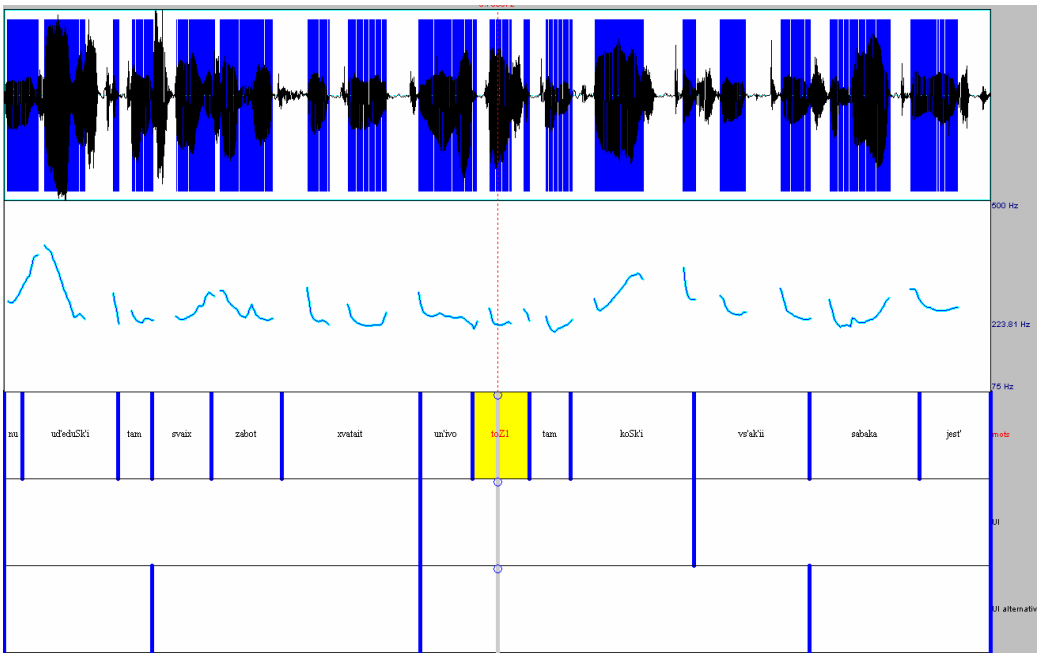


Figure 7.6 : Exemple d'annotation d'une unité interpausale en termes d'unités intonatives (palier 2 : expert français ; palier 3 : experts russes)

Nous avons donc décidé d'exploiter la question de l'annotation du corpus en termes d'unités intonatives dans une étude expérimentale en perception dont nous présenterons les détails dans le chapitre suivant ; les chapitres VIII, IX et X sont orientés sur l'exploration

[221] Nous omettons exprès les signes de ponctuation.

des résultats obtenus[222]. L'objectif de l'étude en perception que nous avons menée est donc d'éclaircir le point de la structuration de l'oral spontané en russe en termes de groupements et de frontières prosodiques.

7.3.3.5. Les unités interpausales

Au niveau le plus concret (et le plus bas dans la hiérarchie adoptée) nous disposons de la segmentation de notre corpus en unités de performance : unités interpausales (*inter pausal units*, IPU, Koiso et al. 1997). Le terme de IPU désigne un bloc de parole délimitée du reste de discours par des pauses silencieuses d'une durée supérieure à 200 ms[223]. Notons que cette unité est introduite dans une étude portant sur l'analyse de l'organisation discursive du message, et que le choix est argumenté par son caractère objectif. Ainsi, dans la transposition de la méthodologie du corpus, c'est l'unité interpausale qui peut servir de pivot du transfert des annotations pour tous les autres niveaux.

7.3.3.6. Annotation mélodique : MOMEL et INTSINT

Afin d'identifier l'organisation tonale de notre corpus, nous avons appliqué les traitements suivants : la courbe mélodique a reçu une modélisation par application de la procédure MOMEL et a été annotée par la suite à l'aide de l'alphabet INTSINT. Nous avons déjà discuté des présupposés théoriques sous-jacents de cet algorithme et nous avons argumenté notre choix pour ce système d'annotation dans le chapitre II.

Cette annotation est semi-automatique. Nous nous sommes servie d'une implémentation de l'algorithme MOMEL sous la forme d'une fonction externe en C qui est appelée par un script de l'intérieur du logiciel Praat (Boersma & Weenink 2005). Or, se pose la question de l'unité d'analyse pour le codage. Dans la version implémentée de l'algorithme, les points cibles reçoivent leur codage sur la base de deux paramètres : clé (*key*) et étendue (*range*). Ensemble, les paramètres de clé (niveau) et d'étendue définissent le registre mélodique, concept qui n'est pas anodin dans les analyses prosodiques (cf. Laver, 1994, sur cinq lectures possibles de ce terme). Le registre tonal permet d'évaluer les caractéristiques globales de la f0 d'un locuteur ou d'un contour mélodique. La dimension linguistique du phénomène du registre tonal se retrouve associée à l'organisation informationnelle du message (cf. Ladd, 1992, travaux sur l'organisation mélodique au niveau du paragraphe vocal, paraton, domaine du développement d'un topique Yule, 1980 ; Brown & Yule, 1983). Pour des raisons de facilité du traitement, notre corpus est segmenté en fichiers d'une minute environ. Or, compte tenu des remarques faites ci-dessus, pour le traitement MOMEL-INTSINT nous avons choisi de travailler avec des unités plus

[222] L'unité interpausale en question a été incluse dans l'ensemble des stimuli : sous la condition de la présentation normale des stimuli, la frontière controversée a été annotée par 3 sujets, 2 autres signalant une frontière après le mot « *дедушки* » ; 2 sujets ont proposé une annotation similaire à celle de l'expert français.
[223] Ces unités peuvent être obtenues de manière automatique pour un corpus recueilli dans des conditions d'enregistrement favorables à la base des mesures de l'énergie dans le signal.

petites correspondant aux interventions de la locutrice[224]. D'autant plus, que Hirst (2004) propose un traitement unifié pour le bloc de cinq énoncés, formant un tout du point de vue discursif dans le corpus Eurom-1.

Lors de son application, ce script prévoit une possibilité de correction manuelle des points cibles détectés. Au final, ce script permet de récupérer les informations suivantes :

- un TextGrid (grille d'annotation générale sous Praat) général comportant trois niveaux et précisant la distribution des points cibles détectés par MOMEL (version corrigée) sur l'axe temporel et leur encodage INTSINT, ainsi que les valeurs de f0 en Hz de ces points cibles tels qu'ils ont été détectés initialement, et les valeurs que les points cibles recevront étant donné leur codage INTSINT.

- un fichier PitchTier (par fichier son) permettant d'une part la représentation des points cibles dans l'espace temps/fréquence et d'autre part celle de leur interpolation à l'aide de courbes splines quadratiques, conformément à la philosophie sous-jacente de MOMEL.

La procédure de correction que nous avons appliquée se rapproche de celle de la stylisation par copie conforme intégrée dans l'approche IPO. Les procédures d'analyse par synthèse intégrées dans Praat (méthode PSOLA) permettent de reproduire le signal avec la courbe mélodique modélisée à partir des points cibles détectés. Une comparaison perceptive systématique à partir des segments de la taille d'une unité interpausale entre les courbes modélisées et les courbes originales permet de corriger d'éventuelles erreurs. Si les différences sont audibles, nous pourrons rajouter, supprimer ou déplacer les points cibles et réécouter la variante resynthétisée. Ces deux étapes ont été appliquées de manière récursive jusqu'à ce que la version resynthétisée soit perceptivement identique à la courbe initiale.

Procédant à une correction manuelle des points cibles nous avons pu constater quelques erreurs systématiques, associées avec certaines formes mélodiques (figure 7.7a) , avec les manipulations de la qualité de voix (présence de voix craquée, cf. figure 7.7b) ou avec des mouvements mélodiques finaux.

Nous avons ensuite effectué une évaluation quantitative de la détection des points cibles sur notre corpus du parler russe spontané pour laquelle nous avons suivi la procédure employée par Campione (2001). Les évaluations précédentes n'ont jamais porté sur de la langue russe. De même, en ce qui concerne la parole spontanée, l'algorithme n'a été testé que sur le corpus français.

[224] En fait, le premier traitement a été effectué à partir des fichiers d'une minute et nous nous sommes aperçue qu'aucun point cible n'avait reçu le codage B (« bottom »). Nous avons associé ce résultat avec un choix déontologiquement incorrect de l'unité d'analyse. Il se peut pourtant que notre choix actuel nécessite également un raffinement : nous pensons aux cas où, dans une intervention, le traitement d'un topique s'achève et le nouveau topique est introduit.

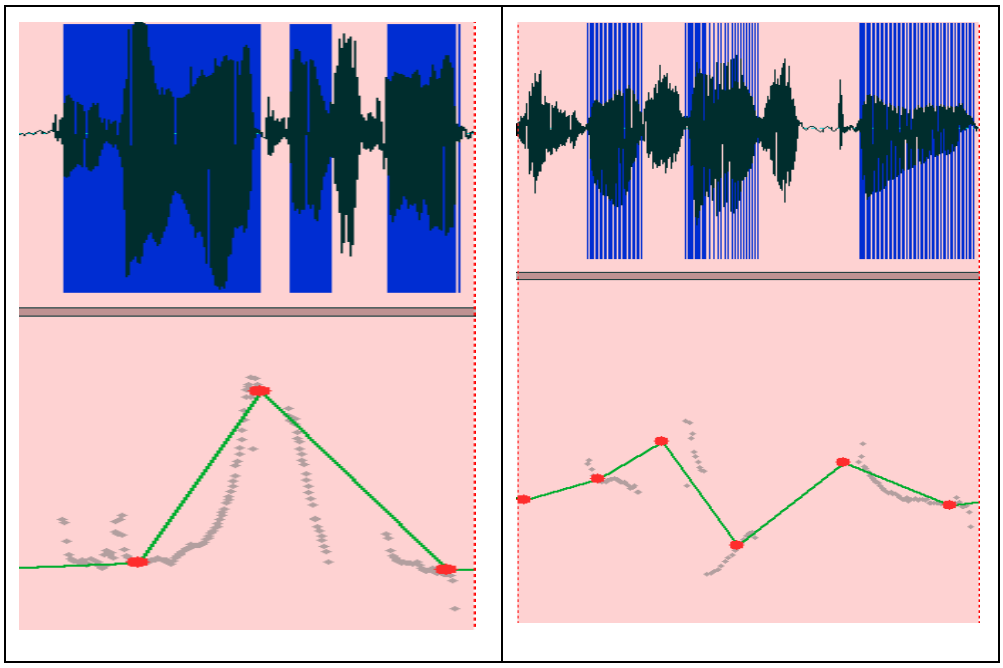


Figure 7.7 : Exemples d'erreur de détection de MOMEL; panel a : erreur systématique associée avec la configuration de la courbe de f0 ; panel b : erreur due à la présence de voix craquée

Comme nous l'avons précisé, lors de la correction manuelle, les points cibles peuvent être supprimés, ajoutés ou déplacés. En évaluation, c'est le déplacement des points cibles qui pose un problème, la démarche étant équivalente à une suppression suivie d'un ajout du point cible. Nous avons donc considéré tout point différant du point original de moins de 200 ms comme déplacé. Il se peut également que lors de la phase de correction un point ait été supprimé mais que, par la suite, nous ayons pris la décision de le garder : dans Praat, on peut revenir en arrière mais on ne peut annuler qu'une seule opération. Or, il nous est arrivé de déplacer deux points avant de décider de revenir à la version proposée par le script de détection des points cibles. Pour cette raison, nous avons décidé arbitrairement de considérer comme identiques deux points qui ne différaient que de 20 ms en temps et de 0,5 ST en fréquence. Le troisième groupe des cas représentent les points détectés correctement dans le domaine temporel, mais comportant une erreur de détection dans le domaine fréquentiel (un écart de plus de 0,5 ST). Au total, sur 1895 points détectés, nous en avons supprimé 86 et rajouté 385 (tableau VII.3).

Points détectés (déplacés inclus)	Points rajoutés	Points supprimés
1809	385	86

Tableau VII.3 : Données sur l'intervention d'annotateur lors de la modélisation MOMEL

La performance de l'algorithme est évaluée à l'aide des mesures de : rappel, précision, silence, bruit et *f-measure*, traditionnellement employées en recherche documentaire (cf. Van Rijsbergen 1979), qui caractérisent la démarche prédictive. La précision quantifie la proportion des points cibles détectés correctement sur l'ensemble de points cibles détectés par l'algorithme (sa mesure complémentaire étant le bruit). Le rappel mesure la proportion des points cibles détectés correctement sur l'ensemble de points cibles retenus dans l'annotation finale (le silence, mesure complémentaire, quantifiant la proportion des points

cibles non prédits par l'algorithme). À son tour, *f-measure*, étant une moyenne harmonique du rappel et de la précision, caractérise la performance du système dans sa globalité : cette performance est optimale lorsque les deux mesures expriment les valeurs les plus élevées.

Dans l'analyse de Campionne (2001) sur un corpus de la parole spontanée en français, les valeurs de rappel, précision et *f-measure* communiquées sont respectivement de 94,5 %, 92,4 % et 93,4%. C'est pour la valeur de rappel que l'on observe des différences importantes dans notre analyse (tableau VII.4) : plus d'un sixième (le silence étant de 17,55%) des points ont été omis dans la première version de la détection des points cibles. Ce résultat peut également être la conséquence de la méthodologie d'évaluation perceptive que nous avons appliquée : l'une des critiques que l'on adresse à la méthodologie de IPO fait le constat que le nombre des mouvements mélodiques distingués dans la perception est très large, et qu'on risque d'aboutir à une précision exagérée de la description.

MESURE	
PRÉCISION	95,46
RAPPEL	82,45
BRUIT	4,54
SILENCE	17,55
F-MEASURE	88,48

Tableau VII.4 : Résultats de l'évaluation d'algorithme MOMEL-INTSINT

Nous constatons également une proportion assez élevée du groupe de points déplacés : 38,15% de tous les points cibles détectés (22,4 % dans l'étude de Campionne, *op.cit.*). Nous présentons les détails sur les points cibles déplacés dans le tableau VII.5. Nous en distinguons 3 classes selon que le déplacement ait eu lieu dans le domaine temporel, dans le domaine fréquentiel ou dans les deux. Les données indiquent qu'un déplacement du point cible dans le domaine fréquentiel a été accompagné d'un déplacement dans le domaine temporel dans deux cas sur trois. Parallèlement, la majorité des déplacements des points cibles a affecté la dimension temporelle ; nous pouvons en proposer la lecture suivante : un point cible en tant qu'événement mélodique, volontaire de la part du locuteur a été détecté, or, son alignement avec la chaîne segmentale était erroné. Cependant, signalons également que par le fait d'intégrer une vérification perceptive et la manipulation des points-cibles, l'algorithme MOMEL se révèle sensible à la stratégie individuelle de l'annotateur/correcteur.

	NOMBRE	POURCENTAGE	ERREUR STANDARD	
			secondes	*demi-tons*
Domaine temporel	298	15,72	0,064	0,19
Domaine fréquentiel	150	7,92	0,008	2,21
Domaine temporel et fréquentiel	275	14,51	0,072	2,29

Tableau VII.5 : Informations sur les points cibles déplacés

Nous voyons également que l'erreur moyenne dans le domaine fréquentiel est de 2,25 demi-tons. Cependant, une fois la distribution de l'erreur absolue dans le domaine

fréquentiel analysée, nous pouvons y distinguer deux cas de figure : les écarts de plus de 5 demi-tons peuvent être associés au phénomène de saut d'octave (*octave jump*) auquel sont sensibles de nombreux algorithmes de la détection de f0 dans les programmes d'analyse du signal. Pour le groupe où l'erreur ne dépasse 5 demi-tons, l'erreur standard est de 1,44.

Dans la deuxième étape d'évaluation d'annotation MOMEL-INTSINT, nous nous sommes intéressée aux valeurs fréquentielles pour toute catégorie d'étiquettes. En effet, surtout pour les catégories absolues de points cibles qui caractérisent l'étendue tonale, nous nous sommes posée la question de savoir si ces catégories étaient bien délimitées. En plus, cette problématique est intrinsèquement liée au choix de l'unité d'analyse pour l'annotation INTSINT. Le tableau VII.6 résume les valeurs de moyenne et d'écart type pour 8 catégories d'étiquettes ; nous présentons les valeurs en Hz et en demi-tons.

ÉTIQUETTE	NOMBRE	MOYENNE	ÉCARTTYPE
M	180	258,7 8,01	21,57 0,117
B	174	226,6 7,816	19,62 0,181
T	151	379,1 8,545	67,93 0,249
L	282	238,7 7,895	17,98 0,11
D	393	241,1 7,906	26,85 0,144
H	154	295,4 8,197	34,36 0,162
U	395	259 8,014	17,34 0,093
S	465	244 7,921	34,732 0,169

Tableau VII.6 : Moyenne et Écart type des valeurs de f0 prédites pour les catégories INTSINT

Nous présentons également sur la figure 7.8 les histogrammes de distributions la f0 selon la catégorie d'INTSINT. L'examen de ces histogrammes, ainsi que celui des valeurs d'écart type du tableau VII.6 nous permettent de voir que la variabilité la plus élevée est propre aux étiquettes attribuées aux sommets des pics mélodiques ; les points cibles associés avec des vallées ont un comportement plus stable dans le domaine fréquentiel (cf. les observations similaires chez Patterson, 2000 ; Portes et Di Cristo, 2003). Néanmoins, même pour des catégories absolues (T, B, M) nous voyons que les bandes de fréquence qui leur sont propres se chevauchent.

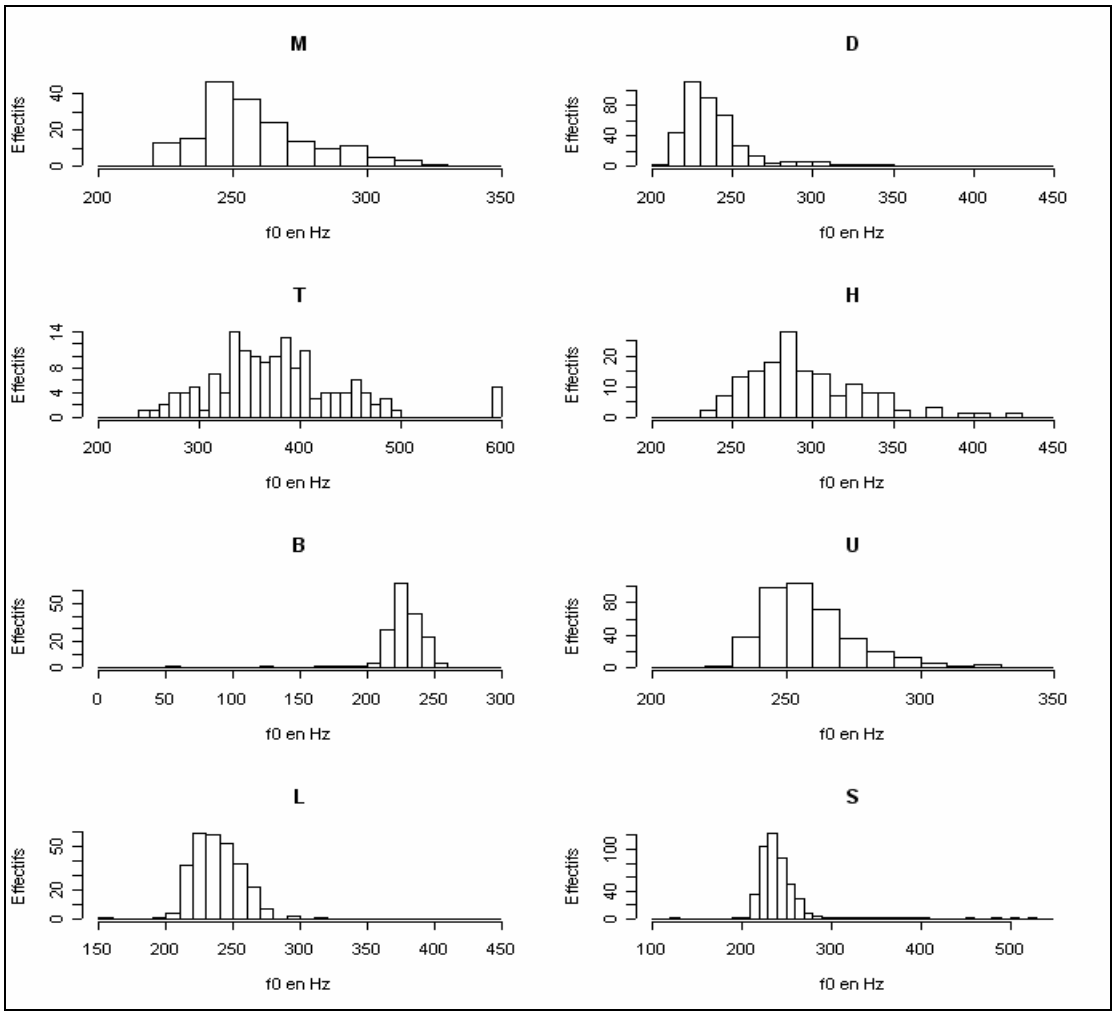


Figure 7.8 : Histogrammes des fréquences de f0 selon la catégorie INTSINT

Nous avons également recouru aux mesures de dissymétrie et d'aplatissement pour caractériser les distributions résultantes (tableau VII.7). Le coefficient de dissymétrie permet de porter un jugement sur l'égalité de répartition autour de la moyenne. Nous prêtons attention surtout aux valeurs observées pour les symboles absolus (M, T, B) : pour les symboles relatifs les valeurs éloignées de celles de la distribution normale paraissent cohérentes, vu leur caractère relatif et leur dépendance des valeurs (*settings*) des paramètres initiaux. Nous constatons tout d'abord, un coefficient de dissymétrie très élevé pour la catégorie B. En vérité, il est dû à quelques valeurs sporadiques (nombre d'observations égal à 6) dans la bande de fréquence de 100-150 Hz. Il s'agit des valeurs infra-basses pour notre locutrice, associées avec des syntagmes à valeur pragmatique conclusive. Une fois ces valeurs exclues, le coefficient de dissymétrie affiche la valeur de 0,408, le coefficient d'aplatissement étant de -0,411. Prenons en compte également les coefficients de dissymétrie positifs pour les étiquettes M et T : une dissymétrie positive est révélatrice d'une longue queue à droite dans la distribution des valeurs de f0 (figure 7.8). Ces tendances témoignent du fait que la locutrice « préfère » utiliser le registre élevé ; nous pouvons y voir l'effet des stratégies discursives adoptées (souci de préserver la parole), mais également l'effet de l'organisation discursive en relation avec le développement de l'argumentation. Cependant, ces observations n'ont que le statut d'hypothèses qui nécessitent une vérification dans une étude expérimentale spécifique.

	Dissymétrie	Aplatissement
Totalité	2,68	10,529
M	0,854	0,255
T	0,942	1,702
B	-4,558	32,766
L	0,163	2,193
D	2,594	8,832
H	0,94	1,074
U	1,233	2,25
S	4,292	25,98

Tableau VII.7 : Dissymétrie et aplatissement selon la catégorie INTSINT

Nous constatons également une dissymétrie positive importante pour la totalité des points cibles (figure 7.9) : la distribution a la forme d'une distribution de Weibul, plutôt que celle de la loi normale. Cette tendance peut être caractéristique de la locutrice de notre corpus. Or, elle fait penser également aux observations faites à propos de la dimension tonale de la parole spontanée avec la dominance des mouvements mélodiques ascendants. Il peut également s'agir d'une stratégie de la locutrice qui consiste à relever la ligne basse (ligne des vallées mélodiques) en raison de l'influence des contraintes discursives et des variations du débit de parole.

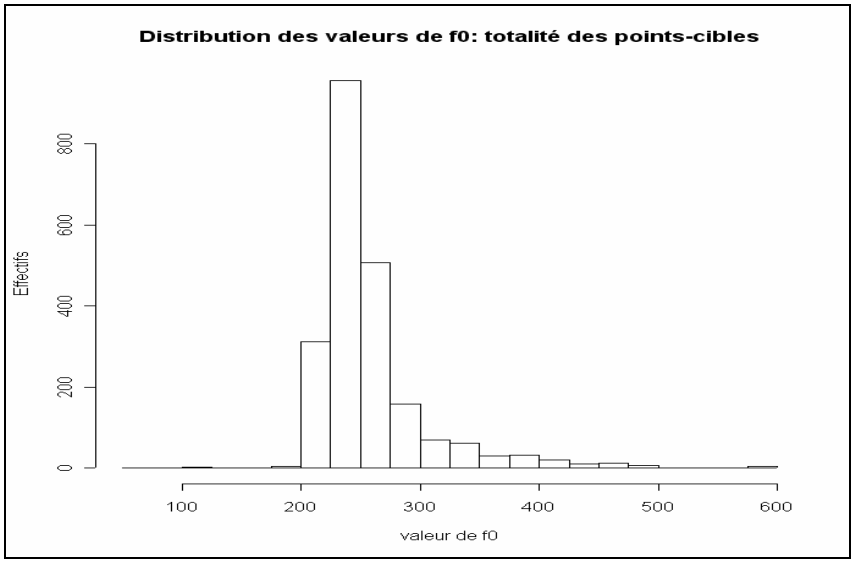


Figure 7.9 : Histogramme des valeurs de f0 de tous les points-cibles

Les auteurs d'INTSINT insistent beaucoup sur le réversibilité des représentations qu'ils adoptent pour tous les niveaux de la modélisation : il est admis que la conversion de la courbe lisse en points cibles par MOMEL et *vice versa* se passe sans aucune perte d'information ; or, la conversion d'étiquettes INTSINT en points cibles localisés dans l'espace fréquentiel implique une légère perte que nous avons cherché à évaluer à l'aide de deux mesures, largement utilisées dans l'évaluation de la performance des systèmes de synthèse : *root mean square error* (RMSE, erreur moyenne quadratique standard) et coefficient de corrélation. La RMSE caractérise la précision de la prédiction et est calculée selon la formule :

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\mathbf{freq_obs_i} - \mathbf{freq_pred_i})^2}{\mathbf{n}}}$$

Nous présentons les valeurs de RMSE en Hz et en octaves (log2 de la valeur absolue, tableau VII.8). Les valeurs de RMSE (Hz) dépassent les valeurs de seuils de perceptibilité (Rossi et Chafcouloff, 1972). Le coefficient de corrélation affiche, quant à lui, des valeurs assez élevées. Les valeurs les plus faibles sont observées pour les étiquettes M et B : en effet, la variabilité dans la distribution des valeurs étant moindre pour ces classes d'étiquettes, le choix d'une valeur standardisée entraîne davantage du décalage. D'autant plus que l'étiquette M étant associée avec tout début de l'unité intonative après une pause, le choix de la fréquence initiale est intrinsèquement lié au développement du contour intonatif dans son intégrité. Certes, la stratégie d'annotation que nous avons adoptée (traitement séparé de toute intervention de la locutrice) induit une certaine variabilité même pour les valeurs standardisées. Or, il reste à évaluer dans quelle mesure les valeurs fréquentielles de deux points adjacents sont corrélées et l'impact de cette prédétermination et des variations minimes dans les limites de RMSE en perception. Bien sûr, l'utilisation des mesures de RMSE et de corrélations ne nous renseigne pas sur l'équivalence perceptive des contours, surtout sur l'équivalence fonctionnelle et sur la valeur pragmatique véhiculée, c'est-à-dire, si les deux contours communiquent une même signification : à cette fin, une évaluation perceptive est nécessaire. Pourtant, nous jugeons que ces mesures permettent d'évaluer la composante acoustique du modèle.

	RMSE (Hz)	RMSE (octaves)	CORRÉLATION[225] (Hz)
Totalité	12,11	0,061	0,9279
M	15,93	0,28	0,6473
B	10,1	0,082	0,5698
T	27,19	0,094	0,8984
L	8,76	0,051	0,9019
D	9,05	0,051	0884
H	12,1	0,054	0,9438
U	8,55	0,047	0,8265
S	9,45	0,051	0,8345

Tableau VII.8 : Comparaison des valeurs de points-cibles détectés et des valeurs prédites selon la catégorie INTSINT

7.1. REPRÉSENTATION MULTINIVEAUX : BILAN

Le format des fichiers d'annotation que nous avons développé correspond au modèle TextGrid pour le logiciel Praat (cf. Boersma et Weenink 2000). Ce format d'annotation permet d'isoler les différentes sources d'informations sur des paliers différents (« tiers »), en se rapprochant ainsi des représentations typiques de la phonologie multilinéaire ainsi que des représentations hiérarchiques de la phonologie prosodique. Simultanément, le

[225] Sur utilisation de la corrélation des rangs de Spearman

logiciel Praat permet une formulation simple des requêtes ce qui facilite et automatise la récupération des données phonétiques (nous évoquerons ces possibilités lors de la discussion dans les chapitres suivants des analyses effectuées à partir du corpus).

Nous disposons donc d'un corpus doté d'une annotation multi-niveaux, comportant les niveaux des phonèmes (1), des syllabes (2), des mots (3), des unités intonatives (4, annotation de l'expert français pour la totalité du corpus, et 5, annotation des experts russes pour une partie du corpus), du degré de proéminence (6), des unités-interpausales (7), du codage Intsint (8) avec les valeurs en Hz récupérées initialement du corpus (9) et prédite (10). Un extrait est présenté sur la figure 7.10.

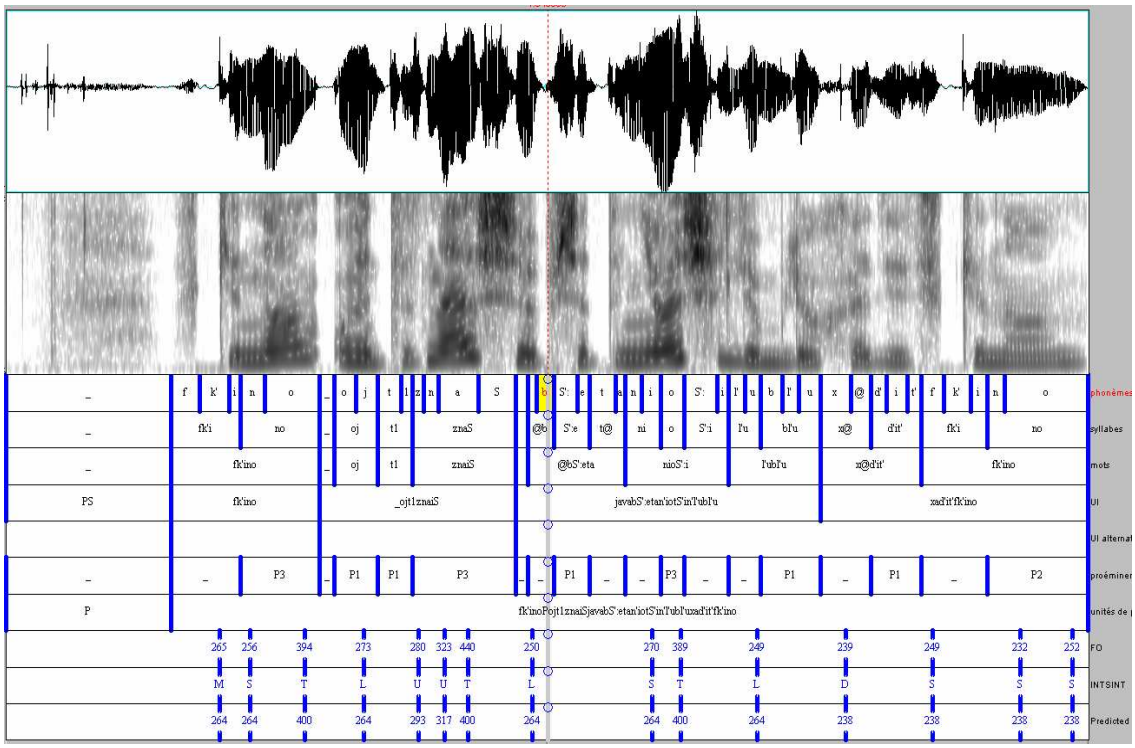


Figure 7.10 : Exemple d'un TextGrid d'annotation du corpus

7.2. RÉSUMÉ ET CONCLUSIONS

Ce chapitre détaille le corpus sur lequel porteront les analyses expérimentales de notre travail. Dans un premier temps, nous avons positionné les études à partir des corpus dans le cadre de la phonétique expérimentale, explicité la notion de la spontanéité de la parole et discuté des modalités permettant d'éliciter de la parole spontanée.

Notre corpus étant enrichi d'une annotation miltiniveaux, nous avons ensuite spécifié des principes qui ont guidé l'annotation à chaque niveau retenu. Précisons, que notre corpus est aligné temporellement pour chaque niveau d'annotation. Enfin, nous avons procédé à une évaluation quantitative de la performance de l'algorithme MOMEL-INTSINT dans l'annotation du notre corpus de parole spontanée en langue russe.

Nous avons évoqué le problème de l'annotation du corpus en termes de syntagmes intonatifs par référence aux désaccords constatées lors de la comparaison du découpage

proposé par un expert français et celui des experts russes. Dans la lignée du modèle de synthèse avec une double sélection d'unités, nous cherchons une heuristique d'annotation automatique du corpus en termes de constituants prosodiques. De ce fait, nous avons décidé d'étudier plus en profondeur des mécanismes qui participent à la perception d'une rupture prosodique. À cette fin, nous avons réalisé une étude perceptive dont nous spécifions le protocole expérimental et présentons des résultats dans les deux chapitres qui suivent.

-

CHAPITRE VIII

EXPÉRIENCE EN PERCEPTION SUR LE PHRASÉ PROSODIQUE

Après avoir analysé les différentes approches et interprétations du concept de syntagme intonatif (dans les travaux linguistiques en général, et plus particulièrement, dans le cadre de la phonologie prosodique ainsi que dans la tradition russe) nous avons trouvé enrichissant d'exploiter cette problématique dans une étude en perception. Nous avons proposé aux auditeurs (7 russes et 7 francophones) une tâche d'annotation perceptive des stimuli russes en termes d'unités prosodiques sous trois conditions expérimentales (avec la quantité manipulée des informations prosodiques disponibles aux auditeurs). Dans le présent chapitre nous explorons les résultats de l'expérience, en nous penchant sur la problématique de l'évaluation de la concordance entre les annotations, ainsi que sur l'influence des facteurs Condition et Langue maternelle sur la performance des sujets. Nous insistons en particulier sur l'importance de l'étape de validation du schéma d'annotation et de sa reproductibilité dans les études menées sur des corpus annotés. Nous analysons également quelques aspects de l'organisation rythmique et temporelle des unités prosodiques résultantes.

8.1. INTRODUCTION

Lors de la présentation des méthodes de la génération de la courbe mélodique en synthèse, nous avons particulièrement insisté sur le rôle de la structure prosodique : ayant préféré la méthode d'une double sélection d'unités, nous avons notamment insisté sur les avantages de la prise en compte de la structure prosodique. Ensuite, à l'étape d'annotation du corpus nous nous sommes retrouvée face au problème de l'appréhension du concept de l'unité intonative dans ses interactions avec les dimensions tonale, discursive et syntaxique. Dans notre démarche de la description formelle de l'intonation du parler russe spontané nous avons fait certains choix méthodologiques : de cette façon, nous avons décidé lors de cette étape préalable de nous passer du concept d'accent mélodique tel qu'il est spécifié dans l'approche de Pierrehumbert et le système ToBI. Nous avons opté pour l'utilisation de l'alphabet INTSINT, ce choix impliquant que le syntagme intonatif devient l'unité minimale recevant la spécification tonale dans notre description. Obtenir une représentation objective de notre corpus en termes d'unités intonatives est donc le point

central de notre démarche. De même, nous envisageons un algorithme permettant la reproduction de ce schème d'analyse sur d'autres données : nous cherchons une modélisation acoustique des frontières intonatives pour l'intégrer dans une procédure prédictive.

Suite aux analyses de différentes approches et interprétations du concept du syntagme intonatif – dans les travaux linguistiques en général, et plus particulièrement, dans le cadre de la phonologie prosodique ainsi que dans la tradition russe – nous avons trouvé enrichissant d'exploiter cette problématique dans une étude en perception. Nous avons donc fait du phrasé prosodique l'objet de notre étude : par phrasé prosodique nous entendons la structuration du contenu sonore du message par des moyennes prosodiques, les frontières des unités prosodiques résultantes provoquant la sensation d'une rupture plus ou moins importante. Dans la perspective conjuguant les démarches *top-down* et *bottom-up* les unités prosodiques formelles définies ci-dessus doivent nécessairement être mises en relation avec des unités linguistiques et discursives. L'étude expérimentale que nous avons menée a pour objectif final une modélisation acoustique de la segmentation de l'énoncé en unités intonatives, et ceci dans une perspective orientée perception (cf. Wightman 2002).

Nous avons opté pour une étude perceptive, proposant aux sujets une tâche métalinguistique. D'autres méthodologies auraient pu être appliquées : nous justifions nos choix dans la section consacrée aux problèmes méthodologiques.

8.2. POINTS MÉTHODOLOGIQUES

Dès le début de notre démarche, nous voudrions éviter toute circularité dans les principes fondateurs de la segmentation du flux de parole en syntagmes intonatifs ; nous pensons notamment à l'interférence forte avec la structuration syntaxique du message[226]. Lors de la présentation du système d'annotation ToBI, ayant le statut de standard dans le domaine, nous avons souligné à plusieurs reprises la circularité du schème d'annotation du phrasé prosodique et de l'organisation tonale. L'idée défendue par Gussenhoven (1990) sur la séparation du phrasé prosodique ancré syntaxe, gérant l'organisation temporelle du message, et de la structuration du domaine tonal en termes de domaines d'association nous a paru intéressant. Dans la même lignée, nous voudrions éviter le risque d'imposer aux données notre propre vision du phénomène du phrasé, pensant également à la reproductibilité du schéma d'annotation. Gardant à l'esprit toutes ces considérations, nous avons analysé les différentes techniques expérimentales utilisées dans les études du phrasé dont on retrouve la trace dans la littérature.

Il est parfois regrettable que les concepts d'unité intonative et de frontière prosodique soient devenus des lieux communs, une évidence qui ne nécessite pas une description et une formalisation dans le cadre des études prosodiques. Comme le notent de Pijper et Sanderman (1994) :

[226] Cette remarque ne contredit en aucun cas la reconnaissance de l'interface prosodie ~ syntaxe.

> *« Quite often it is left unclear on what authors base their claims about the strength of prosodic boundaries. In other cases, prosodic boundary strength is estimated through listening, but it is done by experienced phoneticians or trained listeners, who bring to the task their linguistic knowledge and ideas about prosodic organisation and make explicit use of phonetic information. »*

Pour notre part, nous sommes persuadée que tout le complexe d'informations linguistiques entre en jeu dans les jugements élicités sans que l'on appréhende leurs impacts respectifs. Or, cherchant une procédure d'annotation automatique, nous considérons qu'une transparence d'autant plus élevée est requise dans l'appréhension des interfaces « phonologie ~ implémentation phonétique » et « phonologie ~ syntaxe ».

Par référence à l'expérience en perception entreprise, les principaux points méthodologiques à préciser portent sur le choix d'auditeurs et le choix de la tâche qui leur est proposée.

8.2.1. Population d'experts versus population de locuteurs naïfs : solution mixte

Comme dans toute étude linguistique, il est possible de s'adresser à des locuteurs naïfs ou bien de prendre en compte l'opinion des spécialistes, linguistes et phonéticiens.

Dans les études linguistiques, phonétiques et phonologiques, entre autres, on fait souvent appel à la compétence linguistique des locuteurs natifs (cf. Blevins 1995) : on considère qu'ils ont des intuitions sur les faits de leur langue. Ainsi, les locuteurs natifs peuvent-ils évaluer l'acceptabilité d'une forme linguistique : l'énoncé russe « *Я ехать в Киев завтра* » (« *Moi aller à Kiev demain* ») n'est pas acceptable à cause de sa structure syntaxique. Les intuitions des locuteurs portent également sur les caractéristiques intonatives de leur langue : ainsi l'étude de Kuosmanen et de Silva (2003), traitant de la perception des intonations interrogatives en russe par des apprenants finnois, a un titre très éloquent *« Why don't Russians answer my questions ? »* : les auteurs analysent en quoi l'intonation produite par les locuteurs finnois n'est pas acceptable en tant qu'intonation d'interrogation pour leurs interlocuteurs russes. Dans l'ensemble, les théories linguistiques ont pour objectif d'expliciter ces jugements de bonne formation.

À part la structure interne appropriée, l'intonation d'un énoncé doit sonner juste dans le contexte où cet énoncé est prononcé. La distribution de proéminences mélodiques se révèle très sensible au contexte[227]. De ce fait, l'énoncé (1b) n'est pas acceptable en réponse à la question « *Кто купил книгу?* » (*Qui a acheté le livre ?*) :

(1) (a) МАША[228] купила книгу. (*C'est Marie qui a acheté le livre*)

(b) Маша купила КНИГУ. (*C'est le livre que Marie a acheté*)

[227] La raison de cet état de choses réside dans la fonction du marquage des informations importantes, propre à la proéminence mélodique.
[228] Nous désignons par des majuscules les unités porteuses de l'accent mélodique nucléaire.

Donc, les jugements de bonne formation dépendent du contexte, ce qui affecte le nombre de formes prosodiques disponibles, actives pour le contexte en conséquence.

Les réflexions ci-dessus corroborent l'idée que les locuteurs natifs peuvent émettre des jugements sur la bonne formation des expressions langagières de leur langue, et ceci également pour l'organisation prosodique de l'énoncé. Pour être acceptée, l'intonation d'un énoncé doit être bien formée par rapport à son organisation structurelle interne, et être pertinente dans le contexte de la situation de communication.

Remarquons, pourtant, que les jugements d'acceptabilité varient pour une population hétérogène d'auditeurs. Cette variabilité provient de 3 sources :

a) les jugements d'acceptabilité sont graduels et non pas catégoriels, le continuum rangeant entre « complètement inacceptable » à « parfaitement acceptable » ;

b) les formes intonatives alternatives peuvent recevoir le même degré d'acceptabilité ;

c) la présence du facteur inter-auditeur et de la problématique de variabilité a pour effet que différents locuteurs peuvent choisir différents contours mélodiques pour le même texte.

Par conséquent, dans une étude expérimentale il est indispensable lors de la conception de l'expérience de tenir compte de ces sources de variabilité en jugement d'acceptabilité ; dans une perspective prédictive la conjonction des approches *top-down* et *bottom-up* s'impose d'autant plus.

Les choix méthodologiques sur la population d'auditeurs et sur la tâche expérimentale permettent de limiter d'une manière ou d'une autre l'influence des phénomènes évoqués. Ainsi dans son étude, Marsi (2001) affirme que la modélisation entreprise doit aboutir à l'intonation la plus acceptable (*most acceptable*), ou, en d'autres termes, l'intonation préférée[229]. L'auteur considère également, que les deux formes alternatives ne présentent pas d'intérêt pour l'évaluation des deux théories : si une théorie prédit la frontière prosodique dans une position donnée, l'autre ne le fait pas, et les deux versions de l'énoncé ont un même taux d'acceptabilité, nous ne pouvons pas en tirer d'arguments pour donner notre préférence à une théorie plutôt qu'à l'autre. Nous ne partageons pas l'opinion de l'auteur et préférons raisonner en termes probabilistes (Pierrehumbert 2001) et introduire la catégorie de frontière optionnelle. Dans son étude, Marsi recourt à des évaluations faites par des experts : en le faisant, il espère surpasser la variabilité intersubjective dans les jugements d'acceptabilité pour les phénomènes intonatifs. Tout d'abord, lors de l'évaluation, les jugements de plusieurs personnes, spécialistes en intonation sont combinés. Il est considéré que les experts sont meilleurs juges de ce qui est conforme à la norme, car après avoir été exposés à une multitude des données linguistiques, ils en ont acquis une connaissance plus approfondie.

Certes, le recours à des sujets-experts facilite la tâche de chercheur, puisqu'on espère disposer ainsi d'un matériau cohérent. De même, ce choix méthodologique assure plus de flexibilité dans la formulation de la tâche expérimentale. Cependant, la question reste de savoir dans quelle mesure les résultats obtenus sont généralisables : les conclusions ne

[229] Quoique les deux termes ne soient pas toujours synonymes.

peuvent pas être faites sur le caractère général des stratégies attestées, ce qui affaiblit la portée scientifique de la recherche menée. Nous avons essayé de diminuer les biais susceptibles d'être générés par le paradigme choisi en travaillant avec deux groupes d'auditeurs et en introduisant dans le protocole trois conditions expérimentales (cf. *infra*).

Remarquons, également, que dans l'étude de Hansson (2003) les auditeurs avaient pour tâche de signaler la force d'une frontière dans la localisation précisée et ceci pour des stimuli de deux types : stimuli contenant une pause (et parfois d'une durée importante) et stimuli sans pause. L'accord entre les auditeurs est analysé et l'auteur constate que pour les stimuli sans pause, une continuité dans les réponses n'est observée que chez les sujets experts (7 paires d'annotateurs sur 10). Aucun effet significatif n'est observé dans les réponses données par les trois autres groupes ayant participé à l'expérience (des étudiants suédois, des étudiants dont la langue maternelle n'est pas le suédois, et des locuteurs naïfs). Les conclusions de Hansson portent sur la pertinence de la pause dans le marquage des frontières. Nous avons explicité notre position sur le rôle d'une pause dans le signalement des frontières prosodiques dans le paragraphe X : ces considérations ont également guidé notre choix des stimuli pour l'étude expérimentale en perception. Nous voyons donc dans les résultats communiqués par Hansson, une mise en garde vis-à-vis du choix de travailler, dans une étude perceptive, avec des auditeurs naïfs.

D'autres arguments ressurgissent si nous examinons quelles tâches ont été proposées aux auditeurs dans les études similaires à la nôtre.

8.2.2. Pourquoi une étude perceptive ? (choix de la tâche expérimentale)

Nous avons choisi de procéder à un travail en perception. Sûrement, le phrasé prosodique peut faire l'objet d'études en production ; or, dans cette perspective, les postulats théoriques initiaux jouent un rôle important. En effet, comme les indices acoustiques des frontières prosodiques sont multiples, la structure prosodique n'est plus déduite des productions observées, mais doit être déduite au préalable à partir du texte et des présupposés de l'expérimentateur. D. Hirst (*projet de recherche 145*[230]*)* propose une méthodologie objective, compatible avec l'organisation prosodique en lecture : comme en lecture, toute pause a une motivation structurelle et accompagne systématiquement une frontière prosodique[231], les positions ainsi marquées sont des sites inconditionnels des frontières prosodiques ; par conséquent, leurs réalisations acoustiques peuvent être analysées afin d'aboutir à un modèle acoustique. Cette approche est innovante dans la mesure où la méthodologie répandue consiste à confronter l'annotation du phrasé par des linguistes selon les principes théoriques, souvent à la base des transcription du texte, et les productions des locuteurs.

[230] http://teck.lpl.univ-aix.fr/projets/FMPro?-db=lpl-projets.fp5&-format=enreg_detail.htm&-lay=web&ID=145&-find

[231] Cela ne signifie pas que la présence d'une pause soit la condition nécessaire pour actualiser une frontière prosodique.

Nous nous posons la question de la structuration prosodique en termes de frontières et de groupements, de la parole spontanée. Dans les sections 7.2.2. – 7.2.3., nous avons discuté des rapports entre la parole spontanée et le concept du style de parole. Comme le conclut Barry (1995), après avoir analysé les postulats et les données de l'étude d'Eskénazi (1994), la plus grande partie des recherches dans le domaine phonostylistique s'orientent vers la production de la parole[232]. Or, pour Barry, l'approche la plus prometteuse est celle qui est orientée vers l'auditeur, c'est-à-dire, vers les effets produits sur lui par le message. Dans notre étude nous nous plaçons donc du point de vue de l'auditeur : nous voulons appréhender la dimension communicative du phrasé prosodique et analyser l'attitude coopératrice du locuteur dans les contextes non-ambigus.

La question se pose de savoir comment éliciter chez les auditeurs des jugements de bonne formation pour des structures prosodiques. Deux types de données peuvent être utilisés dans l'expérience : on peut procéder soit avec les données acoustiques soit avec une représentation symbolique[233] de l'intonation et du phrasé. Si avec des locuteurs naïfs seule la présentation acoustique est justifiée, les deux tâches peuvent être proposées aux experts. Or, étant donné que pour le russe, il n'existe pas de système standardisé de l'annotation prosodique, l'approche symbolique aurait pu induire un biais expérimental. D'autre part, avec la présentation symbolique de l'intonation, les données recueillies seront fondées sur les intuitions des experts à propos de l'acceptabilité des formes intonatives, sans vraiment nous renseigner sur des phénomènes observés dans les productions authentiques. Nous sommes donc amenée à travailler avec la présentation acoustique des stimuli.

8.2.3. Comment formuler la tâche expérimentale ?

Une fois prise la décision sur la nature de la tâche expérimentale, la question suivante se pose de savoir comment formuler la tâche expérimentale. De nombreuses études procèdent à une comparaison : des paires de stimuli sont présentées qui n'affichent que le contraste prosodique visé[234] (distribution des proéminences ou phrasé prosodique). Or, vu les limites de la mémoire, la longueur et la durée des stimuli doivent être contrôlées ; on ne peut donc pas travailler avec des passages longs (paragraphes ou énoncés complexes). D'autre part, il n'est pas toujours évident de retrouver la motivation sous-jacente à la réponse obtenue : par exemple, si l'étude porte sur la localisation des phénomènes prosodiques tels que les proéminences et les frontières, le jugement d'inacceptabilité peut être dû aux détails de l'implémentation phonétique (cf. problématique d'alignement).

---

[232] C'est pourquoi les lignes de comparaison entre la lecture et la parole spontanée, d'une part, et entre une conversation libre et négligée et une parole claire, de l'autre, ont été privilégiées par certains chercheurs.

[233] Deux tâches peuvent être proposées : soit on demande d'évaluer l'annotation prosodique déjà effectuée par un expérimentateur ou un système artificiel (analyseur syntaxique etc. ; on présente alors un texte orthographique doté d'une annotation des faits prosodiques divers (phrasé et distribution des accents, cf. Monaghan et Ladd 1990)) ; soit on demande aux sujets de prescrire les caractéristiques prosodiques au texte.

[234] On évite ainsi l'influence des facteurs comme la qualité segmentale et les caractéristiques idiolectales du locuteur (idiolecte, qualité de voix) sur les jugements d'acceptabilité.

Une autre méthodologie a été proposée dans l'étude de Fougeron et Ventsov[235] (2002). Cette méthode expérimentale est motivée, entre autres, par de nombreux travaux sur les ressources limitées de la mémoire de travail (cf. Fodor *et al.*, 1974 ; Cowan 2001) : on considère que la perception d'une frontière prosodique déclenche la re-analyse du matériau linguistique précédant, sa réinterprétation et son codage en termes de catégories plus abstraites, ce qui a pour effet une libération de l'espace dans la mémoire opérationnelle. Le travail est mené à partir d'un texte économique lu. Les auditeurs ont eu pour tâche de transcrire orthographiquement le matériau sonore ; pendant l'écoute, ils ont pu arrêter le magnétophone autant de fois qu'ils le voulaient, mais il leur était interdit de revenir en arrière. Le deuxième magnétophone fixe la performance des auditeurs. Remarquons pourtant – et les auteurs le soulignent eux-mêmes – que la tâche expérimentale ne permet pas d'évoquer les stratégies perceptives et attentionnelles, mises en place dans une situation de communication : les auditeurs ne prêtent pas attention au contenu du texte.

Compte tenu de ces controverses méthodologiques, nous avons choisi de proposer aux sujets une tâche d'annotation des stimuli.

Dans les tâches d'annotation que nous avons recensées dans la littérature, on introduit souvent le concept de la force perceptive d'une frontière (*perceptual boundary strength*). Le nombre de degrés de force à distinguer varie selon l'étude. De Pijper et Sanderman (1994) utilisent ce concept dans les expériences en perception avec des sujets naïfs qui attribuent à tout mot un indice de force chiffré. Les auteurs considèrent que les auditeurs naïfs ne fondent pas leurs jugements sur une référence explicite aux événements phonétiques ni aux structures syntaxiques ou prosodiques ; quoique ce propos nous paraisse discutable. D'autre part, cette méthodologie présuppose que les jugements soient recueillis pour toute frontière entre les mots : nous nous demandons alors dans quelle mesure les réponses ne sont pas relatives, sans appréhender l'énoncé ou fragment du discours dans sa totalité. Toujours avec des auditeurs naïfs, Hansson (2002) utilise une autre méthodologie: dans son étude, les stimuli ont été présentés en version orthographique, la frontière pour laquelle le jugement était attendu étant explicitement indiquée (toutes les frontières prosodiques analysées se retrouvaient dans des positions syntaxiquement marquées). Pour mesurer la force de frontière Hansson utilise l'échelle visuelle analogique (*visual analogical scale, VAT*, Wewers & Lowe, 1990, échelle utilisée dans la pratique clinique pour indiquer l'intensité de la douleur). L'échelle proposée aux auditeurs, a la forme d'une barre horizontale, de 100 mm, et les sujets doivent positionner le curseur pour indiquer la force de la frontière perçue. La force de la frontière est mesurée comme une distance en millimètres de l'extrémité représentant le degré le plus faible de l'échelle jusqu'à la marque indiquée par l'auditeur.

L'utilisation d'une échelle graduelle n'est tout de même pas la règle universelle. Dans l'étude de Knowles (1991) seule la localisation des frontières des groupes tonals (*tone groupes* de la tradition britannique) était déterminée par deux phonéticiens qui avaient accès aux informations phonétiques : aucune distinction entre les différentes forces de frontières n'était donc faite dans cette étude. Dans le cadre du projet GROG (Carlson *et al*.

[235] Il s'agit de la méthodologie modifiée de l'étude de Wingfield & Nolan 1980, dans laquelle les participants avaient pour tâche de répéter, après, l'avoir entendu, le segment du discours le plus long.

2005), les jugements graduels ont été recueillis auprès des auditeurs naïfs, tandis que les experts ont annoté le corpus en termes de trois catégories : « absence d'une frontière », « présence d'une frontière faible » et « présence d'une frontière forte »[236].

Terken et Collier (1992), à leur tour, s'intéressent aux traits prosodiques dans des contextes syntaxiques précis. Dans leur étude, six experts procèdent à l'annotation du matériau sélectionné, faisant la distinction entre les frontières fortes et faibles. L'analyse des scores attribués aux positions d'intérêt amène les chercheurs à prendre une décision méthodologique généralisatrice, c'est-à-dire à considérer qu'au-dessus d'un certain seuil toutes les frontières indiquées sont des frontières prosodiques (au fil de l'analyse, on omet donc toute distinction selon la force de la frontière).

Néanmoins, la distinction des différents niveaux de frontières sous l'Hypothèse de l'Étagement Strict (cf. chapitre IV) n'est possible que si les différentes forces reflètent les détails de la constituance prosodique : les différents types de frontières séparent les constituants des niveaux différents dans la structure hiérarchique. Dans l'étude de Price *et al.* (1991), on distingue 7 forces de frontières qui forment l'indice des ruptures prosodiques :

0 – est attribué à une frontière quand aucune disjonction entre deux mots n'est perçue (comme dans un groupe clitique) ;

1 – les frontières des mots prosodiques : ces frontières sont perçues comme n'étant pas marquées par des moyens acoustiques ;

2 – il s'agit d'une frontière entre deux mots au sein des constituants des niveaux supérieurs (*subordinate minor tone group* de Ladd & Cambell, 1991, ou syntagme accentuel de Pierrehumbert et Beckman, 1986) ;

3 – frontière d'un syntagme intermédiaire (ou d'un syntagme phonologique) ;

4 – frontière d'un syntagme intonatif ;

5 – cet indice sépare les groupes des syntagmes intonatifs, et est souvent associé avec une prise de souffle et une pause de durée importante (*superior major tone group* de Ladd & Cambell, 1991) ;

6 – frontière d'énoncé.

Nous voyons donc que dans l'étude de Price et ses collègues la force d'une frontière se voit explicitement associée à la constituance prosodique.

Pour notre part, nous avons jugé qu'il serait fastidieux de faire le lien entre les jugements d'auditeurs en termes de force de frontière et la structure prosodique hiérarchique : nous en avons discuté précédemment, la tradition russe reconnaît un seul constituant, le syntagme, en tant qu'unité minimale de l'organisation intonative/prosodique ; tout de même, son statut est incertain en ce qui concerne sa mise en correspondance avec le syntagme

[236] Cependant, la tradition d'analyse du phrasé prosodique en suédois (Bruce 1998, cité par Hansson 2003) veut que les frontières fortes correspondent aux frontières d'énoncés, et les frontières faibles – aux frontières des unités intonatives (le syntagme intermédiaire est exclu en tant que constituant de la hiérarchie prosodique ayant un marquage tonal à l'oral).

intonatif et le syntagme phonologique/intermédiaire, deux constituants reconnus par la phonologie des domaines. Nous avons donc demandé aux sujets-experts ayant participé à notre expérience, de marquer les frontières des unités intonatives, le terme étant utilisé explicitement dans la formulation de la tâche, sans pourtant qu'en soit proposée une définition.

Nous avons ainsi proposé aux sujets une tâche métalinguistique. Le danger inhérent de ce type de paradigme réside dans une forte dépendance de la performance des auditeurs (experts) de leur propre vision du phénomène prosodique en question : ce sont à la fois l'interprétation de la tâche expérimentale proposée et les connaissances théoriques acquises qui interviennent. Étant consciente de ce biais, nous avons réfléchi aux modalités, permettant de le dépasser et d'acquérir des données généralisables. En conséquence, les choix suivants ont été faits :

a) nous avons choisi de proposer la tâche d'annotation des stimuli aux sujets russophones et non russophones (francophones) et de comparer leurs performances ;

b) nous avons également introduit trois conditions expérimentales que nous détaillons ci-après, les performances des auditeurs étant ainsi comparées suivant les différentes conditions afin de mettre en évidence différents facteurs linguistiques influents ;

c) nous avons enfin combiné dans l'analyse de nos données les approches ascendante et descendante.

En procédant à l'analyse de nos données, nous avons toujours gardé en tête la remarque de Greenberg (2005) :

> *« With respect to speech the contrast between "pure" and "applied" research is stark indeed. Linguists and phoneticians often view spoken language through a "glass menagerie" of abstract forms, which often bear only the faintest resemblance to language spoken in the "real" world. »*

8.3. PRÉSENTATION DE L'EXPÉRIENCE

8.3.1. Conditions expérimentales

Comme nous l'avons déjà mentionné, la tâche proposée aux auditeurs dans l'étude perceptive est une tâche d'annotation des stimuli sélectionnés en unités intonatives. Les auditeurs localisent les frontières prosodiques dont le nombre n'est pas spécifié. Du point de vue formel l'unité intonative est définie par rapport à ses marques externes et à son organisation interne : c'est une unité qui comporte au moins un accent nucléaire et qui est délimitée du reste du discours par un ton de frontière Haut ou Bas. Or, comme nous l'avons démontré dans l'introduction, de nombreuses études appliquent le parallélisme entre la structure prosodique et la constituance syntaxique. Afin d'évaluer l'importance des indices prosodiques dans le marquage des frontières prosodiques, nous avons introduit dans le test des conditions

expérimentales dans lesquelles les auditeurs ne pouvaient plus s'appuyer sur des informations sémantiques et/ou syntaxiques : les stimuli y ont été désémantisés.

8.3.2. Application du procédé de désémantisation dans les travaux phonétiques

Recourir au procédé de désémantisation n'est pas nouveau dans la recherche phonétique et les méthodologies y sont multiples.

Lehiste et Wang (1977) utilisent la technique d'inversion spectrale (le signe de chaque deuxième point de l'onde numérisée est renversé).

On peut également procéder par la méthode de resynthèse qui permet de remplacer des phonèmes afin d'obtenir des non-mots. Ce type de transformation a été utilisé dans plusieurs études avec des choix de phonèmes variables : Gussenhoven & Rietveld (1988) ont procédé par réitération en « da-da », Sluijter & van Heuven (1993) ont utilisé « na-na » pour un seul mot au milieu d'une phrase ; d'autres encore choisissent « ma-ma ».

Remarquons, qu'un consensus est établi dans la littérature sur l'utilisation de la voyelle /a/ : nous voyons pourtant un inconvénient à la transposition de cette méthodologie au russe, étant donné le phénomène de la réduction vocalique qui touche surtout cette voyelle basse ouverte ; l'organisation rythmique des stimuli peut donc être perturbé.

Pour les consonnes, Ramus & Mehler (1999 : 514) recensent trois méthodes possibles (qui sont d'ailleurs toutes utilisées dans leur étude) :

- la première transformation est appelée *« saltanaj »* : toutes les voyelles sont remplacées par /a/, toutes les fricatives par /s/, les plosives par /t/, les liquides par /l/, les nasales par /n/ et les glissantes par /j/. Les phonèmes choisis sont jugés comme étant les plus universaux et représentent tous une articulation alvéolaire ou coronale ;

- le deuxième type de transformation est appelé *« sasasa »*, car toutes les voyelles sont transformées en /a/ et toutes les consonnes en /s/. La consonne /s/ est sélectionnée pour son caractère continu, qui permet à une succession de consonnes de ne sonner que comme un seul /s/ long ;

- enfin, il est possible de remplacer tous les phonèmes par /a/ : *« aaaa »*. Cette manipulation permet d'obtenir un signal, correspondant à un seul son /a/ long, variant continuellement en fonction de la hauteur mélodique et de la durée (sauf si présence éventuelle de pause).

Nous avons finalement opté pour la technique d'application du filtre passe-bas (cf. Lehiste 1979, Schaffer 1984, Frota *et al.* 2002). Cette méthode coupe les hautes fréquences dans le signal (le seuil étant spécifié par l'expérimentateur) : on aboutit ainsi à une version du stimulus dont toutes les caractéristiques prosodiques sont préservées, or, les informations sémantiques et segmentales sont anéanties ; il s'agit donc du signal comparable à ce que l'on entend à travers un mur lorsqu'une conversation est menée dans le bureau voisin. Les stimuli sont présentés dans le dossier 'Expérience en perception' sur le CD-rom. Il faut noter que cette méthode a également reçu des reproches : notamment, Ramus et Mehler

(1999) considèrent qu'il est impossible de maîtriser quels aspects du signal ont été alterés ou préservés. Une autre partie des critiques porte sur l'interprétation des résultats : est-il possible de considérer que les mécanismes mis en œuvre lors du traitement perceptif des stimuli désémantisés sont les mêmes que pour le traitement du langage (à ce propos, voir l'étude de Pannekamp *et al.* 2003).

Nous avons procédé au filtrage des stimuli à l'aide du logiciel PRAAT qui dispose d'une fonction « filtre » : cette fonction permet d'utiliser le filtre passe-bande avec une fonction de lissage aux frontières (cf. figure 8.1).

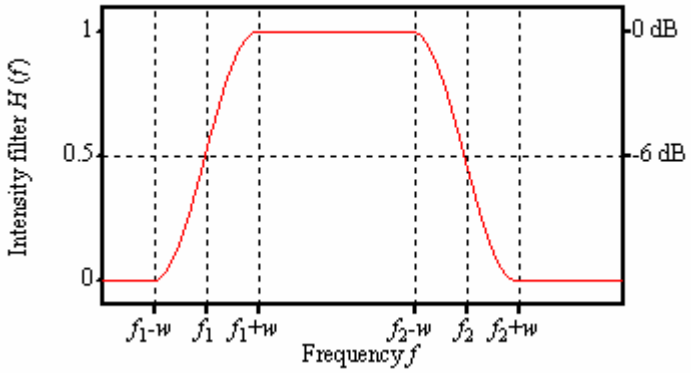


Figure 8.1 : Illustration du procédé de filtrage

Étant donné que nous travaillons avec les productions d'une seule locutrice (et donc n'étant pas confrontée aux variations de la fréquence usuelle dans le cas de locuteurs multiples), nous avons adopté le même seuil de coupure des hautes fréquences : ce seuil a été choisi après plusieurs essais allant de 300 à 600 Hz et fixé à 500 Hz (avec une bande de lissage de 100 Hz).

Dans un deuxième temps, par rapport à la composante prosodique nous avons décidé de séparer les influences rythmiques et mélodiques : pour ce faire, nous avons proposé aux auditeurs des stimuli avec une fréquence fondamentale aplatie, cette manipulation induisant uneffet de monotonie. Pour chaque stimulus nous avons imposé sa fréquence moyenne (cf. tableau VIII.1) et procédé ensuite à une resynthèse (cf. figure 8.2) et à un filtrage.

STIMULUS	F0 MOYENNE	STIMULUS	F0 MOYENNE
1	244,81	14	246,26
2	240,86	15	228,07
3	236,55	16	231,94
4	260,48	17	235,51
5	281,57	18	256,47
6	248,81	19	252,19
7	250,27	20	241,97
8	244,41	21	234,70
9	270,29	22	254,75
10	235,04	23	244,82
11	254,35	24	241,47
12	228,02	25	257,29
13	280,96		

Tableau VIII.1 : Fréquences moyennes par stimulus

(données utilisées pour la réalisation des stimuli de la condition rectotonale)

Précisons que, lors de l'expérience, l'ordre de la présentation des stimuli était inverse de celui exposé ci-dessus : c'est-à-dire que les auditeurs commençaient avec la condition la plus manipulée (et contenant ainsi le moins d'informations linguistiques et prosodiques) pour terminer le test avec les stimuli présentés sous leur forme normale. Pour résumer, les trois conditions expérimentales sont :

- Condition 1, rectotonale (CondR) : les stimuli ont une fréquence fondamentale monotone (ce ton correspondant à la valeur moyenne de la fréquence fondamentale dans le stimulus). On a appliqué par la suite un filtre passe-bas (seuil = 500 Hz), ce qui a supprimé l'information segmentale ;

- Condition 2, désémantisée (CondD) : les stimuli sont modifiés par l'application du filtre passe-bas (les énoncés ont été désémantisés, tout en préservant leur courbe mélodique initiale) ;

- Condition 3, normale (CondN) : les stimuli sont de la parole normale, sans modification.

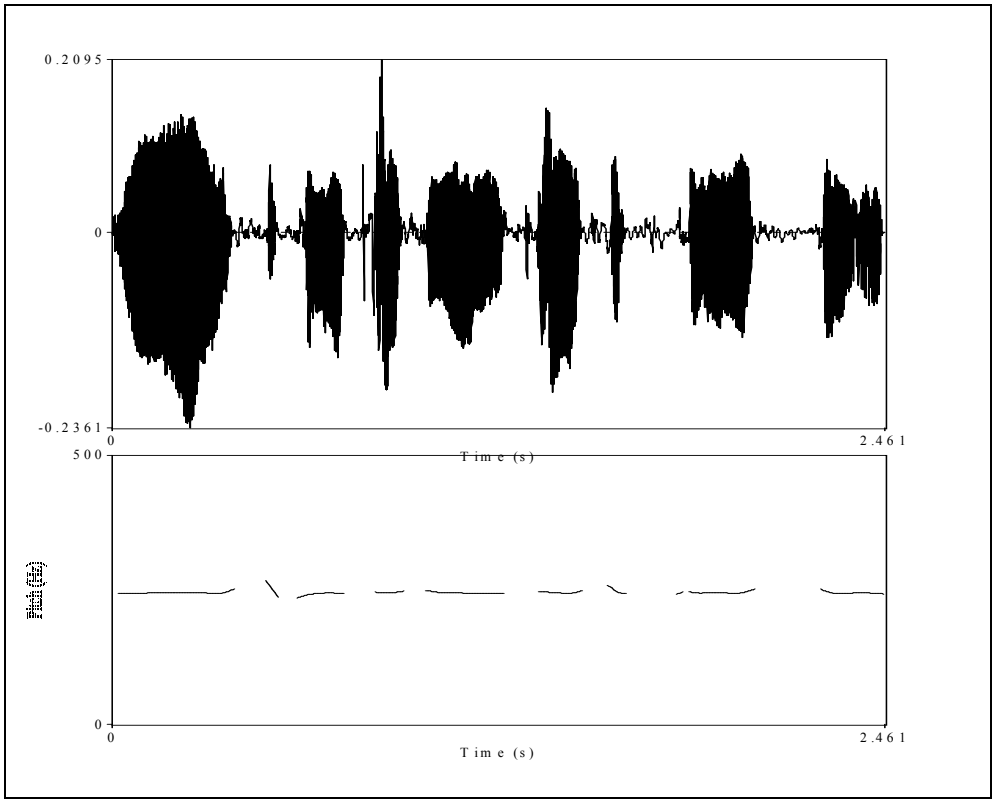


Figure 8.2 : Illustration de la courbe mélodique sous la première condition expérimentale (version rectotonale, stimulus 23)

8.3.3. Déroulement du test

Le recours aux stimuli désémantisés induit fréquemment la modification de la procédure expérimentale. Ainsi, dans l'étude de de Pijper et Sanderman (1994) la tâche proposée aux sujets était la même sous deux conditions expérimentales : exprimer la force de frontière entre deux mots sur une échelle de dix. Avec les stimuli non-modifiés, les sujets disposaient d'une transcription orthographique et notaient leurs réponses sur une feuille de papier. La procédure devait être alternée dans la condition désémentaisée car les frontières prosodiques des mots n'y ressortent pas : les frontières des mots ont été marquées sur l'écran et, afin d'évaluer sa force, les auditeurs avaient pour consigne d'écouter les deux parties, avant et après chaque frontière, ou bien d'appliquer une fenêtre centrée sur la frontière de mot en question.

De même, Fougeron (1989) modifie des consignes expérimentales quand les auditeurs se prononcent sur des stimuli en voix laryngée. Si avec la présentation normale des stimuli, les auditeurs avaient pour tâche de proposer une situation dans laquelle la phrase-stimulus pourrait être prononcée, il leur était demandé de sous-titrer les stimuli en voix laryngée, d'insérer des mots dans le squelette prosodique. Des réponses recueillies ont été ensuite analysées pour déterminer si la structure prosodique correspondait à un message complet, à une réponse, quelle était sa modalité, sa position par rapport à la fin d'un discours, s'il on percevait une connotation émotionnelle.
Pour notre part, comme il s'agit exclusivement d'une tâche d'annotation, aucune modification du protocole expérimental n'a été nécessaire.

Les sujets ont passé le test individuellement. Les stimuli étaient affichés dans le programme de traitement du signal Praat. Les sujets pouvaient écouter l'énoncé-stimulus autant de fois qu'ils en avaient besoin. Ils avaient pour tâche d'indiquer, dans les stimuli présentés, les frontières des unités intonatives qu'ils marquaient sur une ligne spécialement réservée de Praat (cf. figure 8.3).

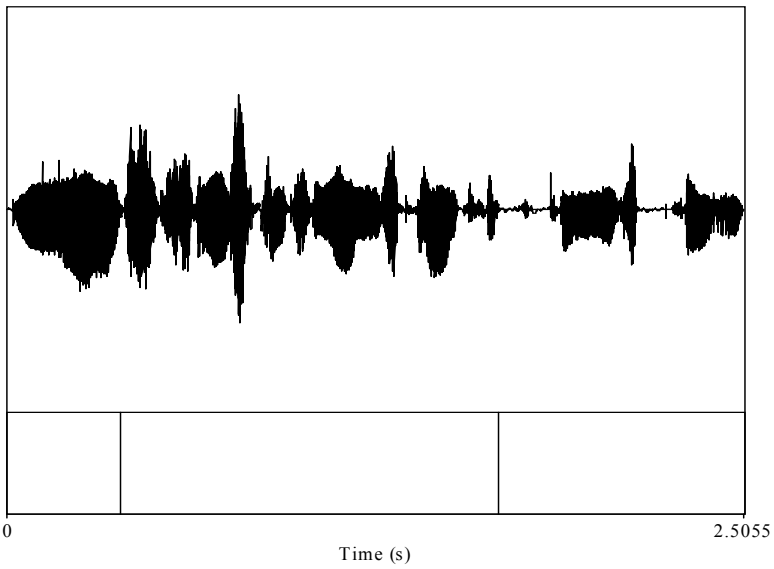


Figure 8.3 : Format de réponses

8.3.4. Stimuli et sujets

Nous avons choisi dans le corpus 25 « unités de performance », correspondant à des interventions ou parties d'une intervention de la locutrice, délimitées par des pauses et comportant au moins une unité intonative. Comme nous travaillons sur un corpus de parole spontanée, nous avons été confrontée au problème des discontinuités propres à ce style de parole (hésitations et allongements syllabiques imprédictibles, liés au processus de la construction en ligne du message). Nous avons choisi d'inclure les extraits avec ces discontinuités dans l'ensemble de nos stimuli, étant donné la perspective globale d'annotation prosodique automatique des corpus oraux.

Le tableau VIII.2 ci-après résume quelques caractéristiques temporelles des stimuli : durée absolue en secondes et longueur exprimée en nombre de mots et nombre de syllabes.

STIMULUS	DURÉE (S)	LONGUEUR (nombre de mots / nombre de syllabes)	STIMULUS	DURÉE (S)	LONGUEUR (nombre de mots / nombre de syllabes)
1	2,2	6 / 15	14	4,05	10 / 23
2	2,77	6 / 15	15	4,37	9 / 26
3	1,34	3 / 6	16	3,29	9 / 20
4	1,67	4 / 12	17	2,85	7 / 17
5	2,66	8 / 17	18	2,31	9 / 15
6	2,25	5 / 17	19	4,08	9 / 21
7	1,74	5 / 12	20	3,43	10 / 21
8	3,11	6 / 19	21	4,66	10 / 26
9	1,38	4 / 10	22	4,87	12 / 28
10	2,34	7 / 16	23	2,5	5 / 12
11	5,28	11 / 30	24	2,26	6 / 17
12	2,34	8 / 17	25	2,46	6 / 16
13	1,53	5 / 12			

Tableau VIII.2 : Caractéristiques temporelles des stimuli

14 auditeurs ayant comme langue maternelle soit le russe soit le français (6 femmes et 1 homme dans chacun des groupes, se situant dans la tranche d'âge de 25 à 50 ans) ont participé à l'expérience. Les sujets sont des phonéticiens, doctorants et enseignants-chercheurs au Département de Phonétique de l'Université d'État de Saint-Pétersbourg et au Laboratoire Parole et Langage à Aix-en-Provence.

8.4. EXPLORATION DES RÉSULTATS DE L'EXPÉRIENCE EN PERCEPTION.

8.4.1. Problématique de l'accord entre les auditeurs

Nous commençons l'exploration des données recueillies dans l'expérience de perception par l'évaluation de la concordance du jugement des auditeurs. Les analyses de l'accord entre les auditeurs sont d'autant plus importantes qu'elles nous renseignent sur la reproductibilité d'un schéma d'annotation, sur la possibilité de le généraliser et ainsi de

dépasser le caractère descriptif et impressionniste des études portant sur différents types de phénomènes linguistiques. Notamment, cette problématique est d'une grande pertinence en linguistique de corpus, dont la démarche présuppose une annotation multiniveau d'un corpus volumineux oral ou écrit. Quel que soit le niveau d'annotation, celui de l'organisation mélodique (Pitrelli *et al.*, 1994, Syrdal et McGory, 2000) ou encore de l'organisation discursive (Isard et Carletta, 1995), sur une grande échelle, il est crucial de valider l'indépendance du schéma d'annotation adopté vis-à-vis de l'annotateur[237], sa pertinence et sa reproductibilité. Donc, outre caractériser la performance des sujets lors d'une expérience concrète, l'analyse des concordances entre auditeurs permet d'évaluer de manière objective la démarche méthodologique que nous mettons en place.

Pour caractériser l'ensemble des études portant sur la concordance entre les juges, nous pouvons définir les axes typologiques suivants :

- connaissance du domaine par les participants : on distinguera les études analysant la performance des experts des études dans lesquelles des locuteurs naïfs ont été impliqués ; les deux groupes d'annotateurs peuvent d'ailleurs être comparés au sein d'une même analyse ;

- nature des corpus : traditionnellement, on oppose la parole lue (style de présentation radiophonique) et la parole spontanée ; ensuite, les différences internes peuvent être dues aux types d'interaction : MapTask, conversation téléphonique avec des objectifs fixes, discours plus formel ;

- nature de la tâche proposée aux participants : s'agit-il d'une annotation uniquement perceptive ou bien comportant aussi l'examen de la courbe de f0 ;

- évaluation de la « neutralité » de la formulation de la tâche proposée ou de l'ancrage dans une théorie linguistique précise.

Les études que nous passons en revue ci-dessous illustrent les différents choix faits dans l'espace typologique que nous venons de présenter. Ainsi la majorité des études récentes traitant de l'annotation prosodique des corpus et de la question du degré d'accord entre les annotateurs ont été effectuées dans le cadre du paradigme d'annotation ToBI (Pitrelli et al. 1994, Grice et al. 1996, Syrdal & McGory 2000) : ces études ont été menées avec des transcripteurs expérimentés et sur des échantillons de parole non-spontanée. D'autre part, l'étude de Buhmann *et al.* (2002) a été réalisée sur un corpus de parole spontanée en hollandais et évalue la concordance entre les annotations des phénomènes prosodiques faites par des non-spécialistes qui ont reçu un entraînement qualifié de minimal.

Nous avons également systématisé les informations sur les consignes données aux sujets dans les études d'annotation du phrasé prosodique :

a) on propose aux annotateurs de signaler la présence d'une frontière prosodique et d'en distinguer deux catégories. Dans l'étude de Wightman *et al.* (2000) les jugements binaires sur la présence ou l'absence d'une frontière prosodique ont été sollicités. Dans

[237] Réflexion de Pierrehumbert sur le positivisme, le caractère cumulatif des connaissances scientifiques.

l'étude de Buhmann et al. (2000), la distinction a été faite entre les frontières fortes et faibles, selon les définitions suivantes des catégories : une frontière « forte » correspond à une interruption saillante du continuum sonore, et est fréquemment accompagnée d'une pause ou même d'une prise de souffle ; en revanche, les frontières faibles sont définies comme des interruptions moins marquées mais tout de même perceptibles. Dans l'étude de Danielli et al. (2004), les annotateurs devaient distinguer les frontières terminales des non-terminales. Une frontière prosodique était définie comme une variation prosodique perceptible qui a pour fonction et effet de diviser le continuum sonore en unités prosodiques discrètes. Étant donné une séquence d'unités prosodiques, une frontière est jugée terminale si un locuteur compétent estime qu'elle peut terminer la séquence. Les auteurs considèrent que les frontières terminales servent à déterminer les frontières des énoncés en tant qu'unités de la structure informationnelle (Austin 1962). Les autres frontières sont considérées comme non-terminales[238].

b) Dans les études inspirées de ToBI, on demande aux annotateurs d'évaluer le degré de jonction entre deux mots adjacents sur l'échelle de zéro à quatre (cf. chapitre II). Ce paradigme n'est pourtant pas purement perceptif, car derrière chacun des labels on retrouve un niveau du phrasé prosodique visé (mot phonologique – syntagme intermédiaire – syntagme intonatif). Nous avons déjà mentionné un parallélisme propre à cette approche entre les annotations d'événements tonales et les annotations du degré de jonction : dans cette perspective, Yoon et al. (2004) ont refusé d'inclure dans leur système d'annotation le niveau d'indices de frontières, justifiant ce choix méthodologique par des réflexions sur le caractère redondant des indices de frontières, qui sont déductibles de l'annotation des phénomènes tonals.

De même, la réflexion de Mayo (1996)[239] nous paraît intéressante : ayant obtenu des valeurs de coefficient de kappa inférieures à celles publiées dans des travaux précédents (Pitrelli *et al.* 1994, Grice *et al.* 1996), Mayo fait une remarque très importante sur le principe fondateur du système d'annotation ToBI. Pour Mayo, les plus faibles statistiques de kappa obtenues sont dues au fait, que par ses prémisses théoriques, ToBI ne satisfait pas la condition d'indépendance des annotations : ainsi le fait qu'un mot ait été marqué avec un ton de frontière, présuppose automatiquement la présence d'un autre ton de frontière (accent de syntagme) sur le mot précédent ; conjointement, un accent mélodique doit être placé sur un des mots dans la proximité. Les mêmes dépendances sont retrouvées entre l'organisation tonale et l'annotation du degré de jonction des mots. Par conséquent, on adopte une vision compositionnelle du contour mélodique (et on n'abandonne pas définitivement ce concept) parce que tout mot dans un énoncé (dans un syntagme intonatif) est dépendant prosodiquement des mots qui l'entourent. Pour notre part, nous considérons que dans la perspective qui nous intéresse, ce ne sont pas les statistiques de kappa qui doivent être remises en cause mais l'articulation « forme-fonction » dans des systèmes inspirés de ToBI.

[238] Précisons, qu'il n'existe par de parallélisme entre les paires de catégories « faible – forte » et « terminale – non terminale », chacune relevant des objectifs généraux que les auteurs se sont fixés ainsi que de la vision de la structuration prosodique du discours qu'ils adoptent.

[239] Quoique nous considérions que les différences en valeurs de kappa dans l'étude de Mayo et les études précédentes sont dues aux différences des méthodes de calculs utilisées, cf. *infra*.

La tâche que nous avons proposée aux auditeurs se distingue du paradigme ToBI dans la mesure où nous n'avons pas demandé à nos sujets de signaler la frontière après chaque mot. Nous considérons les caractéristiques prosodiques globales des unités, et non seulement locales, comme étant pertinentes pour définir la structuration des énoncés en unités intonatives. Nous ne voulions pas non plus adopter une échelle ordinale dans l'évaluation de la force de frontière : nos stimuli provenant d'un corpus de parole spontanée, les phénomènes de dysfluences (hésitations et reprises) peuvent provoquer des jugements établissant une disjonction importante qui ne sera pourtant pas interprétée comme une rupture prosodique structurelle.

8.4.1.1. L'objectif du travail d'évaluation

L'hypothèse que nous testons dans l'évaluation du degré de concordance dans les réponses est pourtant de nature linguistique : dans un premier temps, nous voulons nous assurer que les frontières prosodiques ont une saillance perceptive importante ; ensuite, nous nous interrogeons pour savoir si les auditeurs disposent d'une image intégrale des indices acoustiques qui évoquent la perception d'une frontière prosodique. Dans l'affirmative, il serait possible de soutenir que le concept de frontière prosodique perceptive et les résultats de sa modélisation acoustique peuvent être utilisés pour l'annotation des corpus oraux. Dans une perspective trans-linguistique, nous pouvons comparer la performance de deux groupes d'auditeurs en fonction de leur langue maternelle pour mesurer le rôle des informations linguistiques des niveaux supérieurs (sémantique et syntaxe) ainsi que pour examiner la hiérarchie des indices acoustiques qui entrent en jeu.

8.4.1.2. Procédure d'évaluation

Avant d'entreprendre des analyses spécifiques pour quantifier et caractériser le degré d'accord entre les auditeurs dans notre expérience, il nous a fallu au préalable mettre en forme les données recueillies.

À l'issue de l'expérience en perception, nous avons obtenu pour chaque auditeur et sous chaque condition expérimentale, un fichier de format TextGrid avec les données sur la segmentation du stimulus en unités intonatives. Dans la perspective d'une analyse comparative, nous avons premièrement créé des TextGrids sommaires regroupant les données de sept auditeurs (nous avons adopté un traitement séparé pour chaque condition et deux groupes d'auditeurs en fonction de leur langue maternelle). Les TextGrids bruts sont présentés par condition dans le dossier 'Expérience en perception' sur le CD-rom ; nous en présentons un exemple dans la figure 8.4.

Nous avons signalé dans la partie introductive du présent chapitre que les études portant sur la concordance entre les annotations sont systématiques dans l'évaluation de systèmes de transcriptions inspirées du ToBI. La procédure d'évaluation a également été standardisée et suit toujours le même protocole : on compare l'étiquette que les deux annotateurs ont attribué à un mot (pour la présence ou l'absence d'une proéminence mélodique) ou à une frontière entre deux mots (degré de jonction).

Dans notre expérience, les auditeurs n'avaient pas toujours accès aux informations sur les frontières des mots :

- les auditeurs russophones ont évidemment disposé de cette information, effectuant la tâche de segmentation sous condition expérimentale « non-manipulée » (**CondN**) ; dans les autres conditions, le filtrage masque les frontières des mots ;

- les auditeurs francophones n'ont eu accès à cette information dans aucune condition expérimentale.

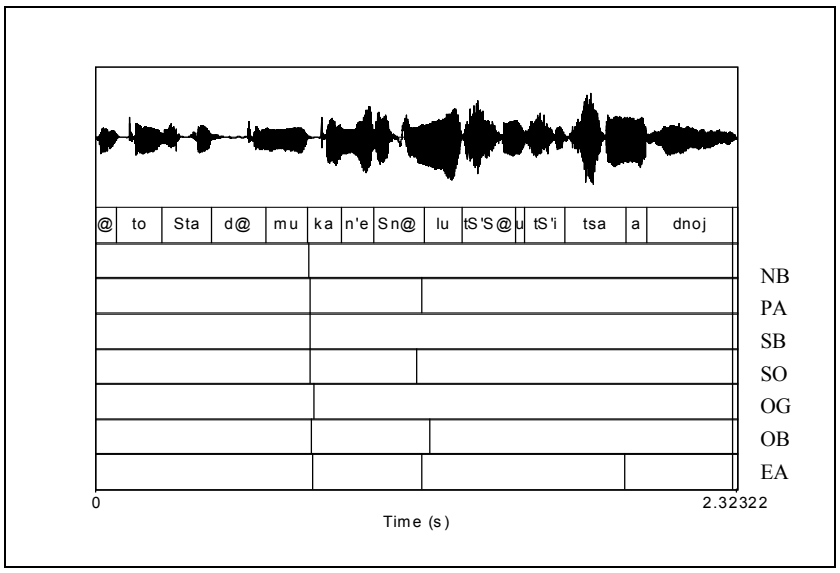


Figure 8.4 : Exemple de TextGrid sommaire (auditeurs russes, troisième condition expérimentale). Le panel de droite : codage des auditeurs.

En revanche, les propriétés rythmiques et temporelles ont été préservées : d'une part, le repérage de syllabes proéminentes (perceptibles même dans les versions filtrées) permet d'estimer le nombre de mots ; conjointement, comme nous avons pu le constater dans des travaux précédents (Nesterenko, 2001, 2002), les auditeurs francophones sont capables d'évaluer le nombre de syllabes entre deux proéminences perçues à l'aide d'un modèle temporel de la syllabe. Donc, dans notre analyse, au lieu de nous référer aux frontières des mots nous avons considéré toute frontière syllabique comme un site potentiel pour le marquage d'une frontière prosodique. En outre, ce choix méthodologique ne contredit pas l'hypothèse de l'étagement strict. Précisons que dans la tâche proposée aux auditeurs nous ne les avons pas incités à procéder à une comparaison systématique entre deux frontières syllabiques adjacentes.

Pourtant, dans les réponses recueillies auprès de nos sujets, les frontières indiquées n'ont pas toujours été alignées avec les frontières syllabiques ; par conséquent, nous avons appliqué une procédure de normalisation en deux étapes :

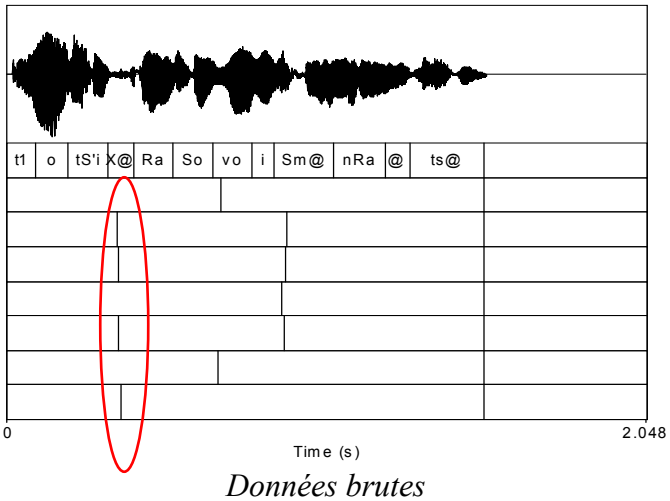
- dans un premier temps, nous avons procédé par standardisation semi-automatique des réponses recueillies : à l'aide du script *normalize_boundaries.praat* que nous avons rédigé, chaque frontière a été alignée de manière automatique avec une frontière syllabique. Les frontières des unités intonatives ont été alignées avec une

frontière initiale ou finale de syllabe selon leurs positions par rapport au milieu de la syllabe ;

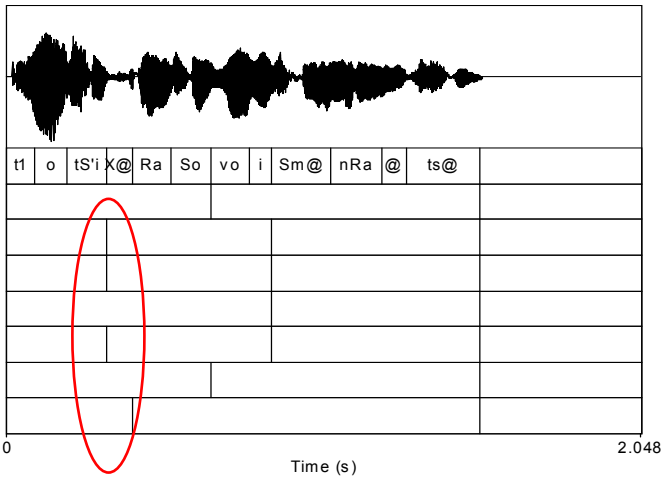
- ensuite, nous avons procédé à une correction manuelle des fichiers de segmentation ainsi obtenus. Dans certains cas, deux frontières proches dans les réponses originelles peuvent se voir alignées avec deux frontières syllabiques distinctes (début et fin de syllabe) : nous avons donc corrigé ces cas de désalignement (cf. illustration de la figure 8.5). Nous avons également pris en compte la durée relative du groupe consonantique au début de la syllabe, de son statut accentué ou inaccentué : nous considérons que si deux auditeurs signalent la présence d'une frontière, et que les positions qu'ils indiquent diffèrent d'une syllabe inaccentuée (donc, réduite), il s'agit de la perception d'une même rupture dans le continuum mélodique[240].

Une fois les données expérimentales normalisées, à l'aide du script *agreement.praat* nous avons obtenu une matrice à deux dimensions FRONTIÈRE SYLLABIQUE * MÉTHODE DE CLASSIFICATION (SUJET) (dossier 'Données' sur le CD-rom) : pour chaque frontière syllabique, cette matrice précise si un ou plusieurs auditeurs sur sept ont signalé la présence d'une frontière de l'unité intonative dans la position analysée (la présence d'une frontière est codée par « 1 », l'absence par « 0 »). Au total, nous disposons donc de six matrices de classification (2 langues maternelles * 3 conditions expérimentales) ; chaque matrice contient les informations sur 419 frontières syllabiques.

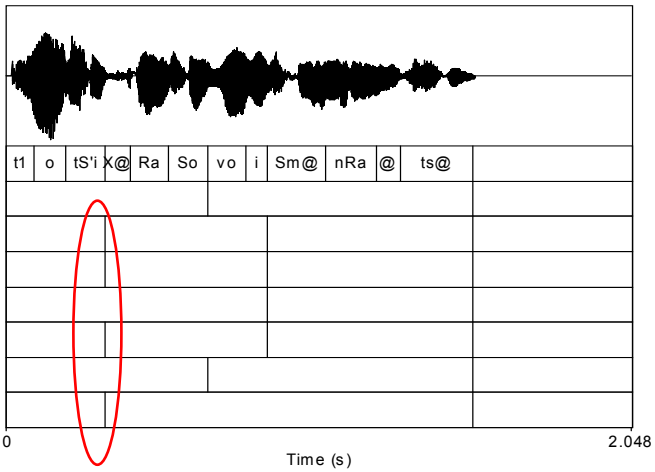
[240] On peut nous reprocher qu'à cette étape de l'analyse nous effectuions une interprétation des réponses des auditeurs et donc, les analyses que nous présentons, perdent en objectivité. Nous considérons pourtant que pour le type des manipulations expérimentales que nous avons entrepris, il n'existe pas d'environnement adapté, qui assurera une objectivité parfaite. D'autre part, les auditeurs lors de l'expérience n'ont pas reçu de consignes pour signaler une frontière prosodique le plus précisément possible. Pour notre part, lors de l'étape de la vérification manuelle de l'alignement des frontières, nous nous sommes fixée des critères strictes, motivés, que nous avons appliqués systématiquement.



Données brutes



À l'issue de l'alignement automatique



Après la correction manuelle

Figure 8.5 : Prétraitement des données

8.4.1.3. Méthodes d'évaluation de la concordance dans la performance des annotateurs

Dans la littérature, deux mesures[241] sont citées pour mesurer la pertinence d'une annotation ou la concordance entre différents annotateurs, le POURCENTAGE D'ACCORD entre les paires d'annotateurs et le coefficient de KAPPA DE COHEN.

Pour la première mesure, il s'agit du ratio entre le nombre de cas pour lesquels les annotateurs ont attribué la même étiquette et le nombre total d'observations. Pour disposer d'une mesure plus stricte, les études ToBI procèdent à une comparaison entre les paires d'annotateurs : on compare les étiquettes d'un annotateur aux étiquettes de tout autre annotateur ayant participé au test d'évaluation. Par conséquent, si dans le cas de 7 annotateurs il y en a 5 qui ont attribué la même étiquette, 71.43% de juges démontrent un accord sur le phénomène annoté ; or, 7 juges forment 21 paires de juges, et si 5 juges émettent le même jugement, on n'observe l'accord que parmi 10 paires de juges, c'est-à-dire dans 47.62% des cas. Dans cette perspective, Pitrelli et al. (1994) considèrent que, dans les travaux sur la concordance, l'unité d'analyse fondamentale est «*transcriber-pair-word* », c'est-à-dire, une paire d'étiquettes que les deux annotateurs ont attribuée au même mot ; par conséquent, la mesure de l'accord entre les annotateurs est :

> « *percentage of transcriber-pair-words exhibiting agreement on a particular element in the trancription.* »

La deuxième mesure, plus pertinente à notre avis, est le coefficient de KAPPA, autrement appelé le KAPPA DE COHEN : comme pour la mesure précédente, l'unité d'analyse est une paire d'étiquettes que deux annotateurs ont attribuée au phénomène linguistique en question. Néanmoins, cette mesure prend en considération que lors d'une mesure d'accord entre juges, on court le risque que l'accord observé soit dû au hasard. Le coefficient de KAPPA permet de mesurer l'accord entre des paires d'annotateurs qui attribuent à chaque observation une étiquette parmi les catégories proposées et corrige l'effet du hasard. Quand l'accord observé n'est dû qu'au hasard, le coefficient de kappa vaut zéro ; quand l'accord est parfait, la valeur de KAPPA vaut un. Ce propos se résume en une formule :

$$(1) \qquad K = \frac{p(A) - p(E)}{1 - p(E)} \quad ,$$

où K est le coefficient de Kappa, p(A) correspond à la proportion de l'accord observé, p(E) étant la proportion de l'accord dû au hasard.

Nous proposons ci-dessous une petite explication destinée à montrer pourquoi il est souhaitable de prendre en considération le degré d'accord que les annotateurs peuvent atteindre par hasard (emprunté à Isard & Carletta, 1995) :

> *Le degré d'accord atteint par hasard dépend du nombre et de la proportion relative des catégories que les annotateurs utilisent. Si l'on suppose que deux annotateurs*

[241] Isard et Carletta (1995) mentionnent l'étude de Passonneau et Litman (1993) où les auteurs utilisent la mesure du ratio entre l'accord observé et l'accord maximal attendu pour évaluer leur système d'annotation. Dans l'étude de Grover *et al.* (1995) pour évaluer la pertinence inter-annotateurs pour la catégorie du degré de proéminence (échelle ordinaire de 8), les auteurs ont utilisé les coefficients de corrélation de Pearson.

utilisent deux catégories dans des proportions égales pour classer un objet (une observation), mais les catégories elles-mêmes sont choisies de manière aléatoire, on attend que dans le cas de hasard l'accord sera observé dans 50 % des cas ; si les deux annotateurs utilisent quatre catégories en des proportions égales, nous espérons un accord dans 25 % de cas. Pourtant, si deux annotateurs utilisent deux catégories de manière aléatoire, mais en utilisent une 95% du temps, l'accord dû au hasard égale 90.5% ($0.95^2 + 0.05^2$).

En général, deux mesures de KAPPA sont utilisées :

- celle qui est calculée selon les principes établis par Cohen (1960), qui ne présuppose pas les proportions de classifications égales pour différents annotateurs (cette mesure est utilisée dans l'ensemble des études sur les systèmes d'annotation ToBI dans lesquelles on ne cite que l'ouvrage de Cohen (1960) ; on applique donc la formule (1) telle quelle). Remarquons d'ailleurs que les statistiques de kappa ont un rôle marginal dans ces études, les pourcentages d'accord entre les auditeurs recevant plus d'interprétation et d'attention.

- celle qui est proposée par Siegel & Castellan : dans leur réflexion méthodologique sur l'évaluation des schèmes d'annotation, Isard & Carletta (*op.cit.*) remarquent que la problématique de réplicativité pour les données nominatives rejoint celle de l'analyse du contenu (*content analysis*, Krippendorff, 1980), où on est arrivé à la conclusion que la réplicativité est évaluée par l'application de la méthode de Siegel & Castellan (1988). La méthode de Siegel & Castellan utilise les proportions de classifications jointes et assure ainsi un ajustement dans les cas où les différents annotateurs diffèrent systématiquement dans leur catégorisation. L'aspect important des statistiques utilisées réside dans l'intégration du principe d'indépendance : les objets dont plusieurs annotateurs effectuent la classification, ne sont ni rangés ni ordonnés ; la tâche se résume donc à l'attribution d'une étiquette appropriée. Utilisant la méthode de Siegel & Castellan (1988), Mayo (1996) obtient des valeurs de KAPPA moins élevées dans la tâche de l'annotation que celles obtenues dans les études du paradigme ToBI.

Pour notre part, nous remarquons que, selon la formule classique de Cohen, la probabilité d'accord par chance diminue avec l'augmentation du nombre d'auditeurs qui participent à l'expérience, surtout si les auditeurs utilisent les différentes étiquettes de catégories dans les mêmes proportions, et que les probabilités des catégories utilisées diffèrent considérablement[242] ; la méthode de Siegel et Castellan intègre une correction pour ce biais.

Notons pourtant, que dans les expériences où une des annotations peut être considérée comme la version de référence (celle effectuée par un expert, par exemple), il est possible, lors de l'évaluation, de recourir aux mesures de rappel et de précision, classiques pour l'évaluation de la qualité prédictive d'un modèle.

[242] Une catégorie est utilisée plus fréquemment que l'autre, ce qui est le cas dans la plupart des tâches d'annotation, cf. par exemple, les graphiques d'incidence des différentes catégories accentuelles comme indices de frontières dans Syrdal et McGory (2000).

La question se pose de savoir, le coefficient de kappa affichant une valeur entre 0 et 1, quelles interprétations ultérieures nous pouvons en faire ?

Dans la présentation de cette mesure du degré de concordance dans les réponses des annotateurs, Krippendorff (1980) affirme qu'il est impossible de tirer des conclusions pour deux variables dont chacune repose sur un schème d'annotation avec K inférieur à 0,7. Autrement, dans le domaine de l'analyse du contenu, le coefficient de kappa supérieur à 0.8 (K>0.8) traduit une bonne réplicativité du schéma d'annotation, une valeur de kappa entre 0.67 et 0.8 permet de tirer des conclusions provisoires et de parler des tendances.

Dans les études portant sur l'annotation ToBI, on applique une métrique qualitative proposée par Landis et Koch (1977, tableau VIII.3).

ACCORD	KAPPA
Excellent	≥ 0,81
Bon	0,80 - 0,61
Modéré	0,60 - 0,41
Médiocre	0,40 - 0,21
Mauvais	0,20 - 0,0
Très mauvais	< 0,0

Tableau VIII.3 : Degré d'accord et valeur de Kappa

Pourtant, comme le commentent Isard & Carletta (1995), les limites établies doivent être appliquées avec précaution, en prenant en compte la nature de la tâche proposée aux auditeurs : l'annotation de la structure informationnelle d'un dialogue et l'annotation des occurrences d'un mot dans un document ne sont pas du même niveau de difficulté. Dans une perspective plus générale, les auteurs soulignent le fait que, si la communauté linguistique adopte des statistiques similaires dans l'évaluation de la reproductibilité de tout schéma d'annotation, cela permettra une interprétation plus adéquate du coefficient de KAPPA.

On peut également procéder à une évaluation statistique des valeurs obtenues : le coefficient de KAPPA traduisant le degré d'accord indépendant de hasard, on peut le comparer au degré d'accord dû au hasard pour évaluer si la différence est statistiquement significative.

Pour les calculs des coefficients de kappa nous avons utilisé la fonction *cohen.kappa* du package CONCORD, destiné à la manipulation des données catégorielles (logiciel R[243]).

[243] Téléchargeable gratuitement à l'adresse suivante : http://www.r-project.org/ .

8.4.1.4. Le modèle linéaire mixte

Dans cette section, nous tenons à présenter la technique statistique de modèles linéaires mixtes qui sera appliquée à plusieurs reprises dans les évaluations de la validité statistique des résultats et des tendances que nous observerons lors du traitement des données expérimentales.

De nombreux modèles statistiques résumant les données empiriques appartiennent à la classe des modèles linéaires mais incorporent simultanément deux types d'effets : les effets fixes, qui sont des paramètres caractérisant la population entière ou certains niveaux de facteurs expérimentaux, et les facteurs aléatoires qui caractérisent l'unité expérimentale concrète. On appelle donc modèle linéaire à effets mixtes un modèle linéaire qui regroupe les deux types de facteurs spécifiés ci-dessus.

De la définition présentée il s'ensuit que les modèles mixtes s'appliquent à des données structurées en blocs : ce type de données est propre, entre autres, aux études longitudinales, aux études qui incorporent les observations sur le même individu sous différentes conditions expérimentales, aux études où des données présentent une structure hiérarchique (emboîtée).

La structure de notre expérience est telle que les données dont nous disposons manifestent une structure par bloc avec des mesures répliquées :

a) le même sujet a donné plusieurs réponses : une partie de la variabilité est donc due à ce *facteur de répétition par sujet* ;

b) chaque stimulus a été présenté trois fois (une fois par condition expérimentale) : on a donc un second facteur de répétition (par stimulus) emboîté dans le facteur sujet ;

c) enfin, le facteur fixe est Condition.

Le modèle linéaire a alors la forme de (1) :

$$(1) \qquad y(ijk) = \beta_i + b_j + b_{jk} + \varepsilon_{ijk}$$

où β_i est la moyenne estimée de la variable sous analyse dans la condition **i** (**i** € [1,3]) , b_j correspond à la variance induite par le facteur sujet (**j** € [1,7]), b_{jk} caractérise l'interaction entre les variables de bloc, caractérisant la réponse du sujet **j** pour le stimulus **k** (k € [1,25]) et ε_{ijk} est un terme d'erreur. Ainsi si nous testons l'effet du facteur 'Condition' sur la variable dépendante 'nombre de frontières', β_i nous renseigne sur la magnitude de l'effet induit par le niveau analysé du facteur fixe, b_j et b_{jk} corrigeront l'effet de la stratégie individuelle du locuteur et du stimulus.

Qui plus est, certaines de nos analyses portent sur des données déséquilibrées (par exemple, dans l'analyse des durées des syntagmes terminaux *versus* non-terminaux, le nombre de syntagmes que chaque auditeur a marqué, n'est pas constant). La fonction *lme* sous R (bibliothèque *nlme*) permet de traiter ce type de données. Pour nos analyses nous avons retenu le seuil de significativité de p=0.05 pour l'évaluation statistique de nos hypothèses.

8.4.1.5. Résultats : Pourcentage d'accord entre les paires d'auditeurs

Le tableau VIII.4 et les figures 8.6 et 8.7 présentent les données sur les pourcentages moyens d'accord entre les paires d'auditeurs sur le phénomène de présence *versus* absence d'une frontière d'unité intonative après la syllabe pour deux groupes de sujets, russes et francophones (nous présentons les tableaux plus détaillés dans l'annexe 3.1).

	AUDITEURS RUSSES	AUDITEURS FRANCOPHONES
CondR	87,49	89,84
CondD	90,43	89,7
CondN	94,98	88,77

Tableau VIII.4 : Pourcentage d'accord entre les paires d'auditeurs

Les figures 8.6 et 8.7 adoptent une représentation sous la forme de *box-plot*[244].

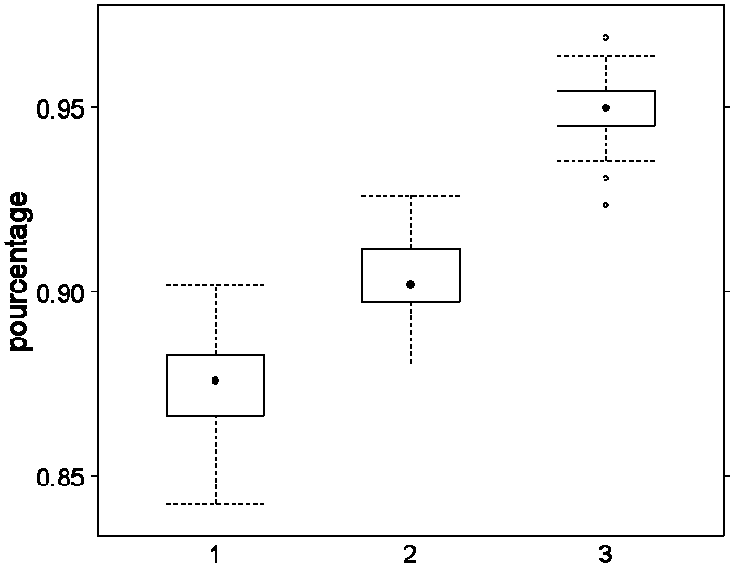


Figure 8.6 : Pourcentage d'accord entre les paires d'auditeurs (auditeurs russes) par condition

Si nous comparons les résultats obtenus avec ceux que présentent McGory et Syrdal (2000) pour le système ToBI, il semble que le degré de l'accord observé soit plus élevé dans notre étude. Nous avons proposé une tâche d'annotation des frontières d'unités intonatives, sans pourtant faire de distinction entre les syntagmes intermédiaires et les syntagmes intonatifs (nous considérons que deux catégories de constituants ont été présentes dans les stimuli). Dans l'étude de McGory et Syrdal, comme le suggèrent les graphiques présentés, l'accord entre les paires d'auditeurs n'était pas très élevé pour les syntagmes intermédiaires (à hauteur de 35%), tandis que l'accord pour l'indice de frontière 4 (syntagme intonatif) affiche des valeurs similaires (à hauteur de 88%). Il se peut pourtant que les différences

[244] Les graphiques *box-plots* permettent de présenter les caractéristiques statistiques générales des distributions sous-jacentes : le point central représente la valeur médiane, la boîte rectangulaire traduit l'écart interquartile, les barres pointillées présentent l'étendue de la distribution, les petits cercles correspondant aux valeurs atypiques.

dans l'unité d'analyse (*transcriber-pair-syllable vs. transcriber-pair-word*, accord sur l'étiquette d'une syllabe *versus* d'un mot) surestiment quelque peu nos mesures.

La comparaison entre les figures 8.6 et 8.7 laisse supposer que le facteur Condition exerce plus d'influence sur la performance des auditeurs russes et joue un rôle moindre dans les résultats obtenus auprès des auditeurs francophones. Ce résultat est tout à fait prévisible, en particulier, pour la troisième condition expérimentale. L'instabilité des réponses pour la première condition (écart type plus élevé), en comparaison avec les deux autres, va dans le même sens que les résultats présentés pour la mesure du nombre de frontières.

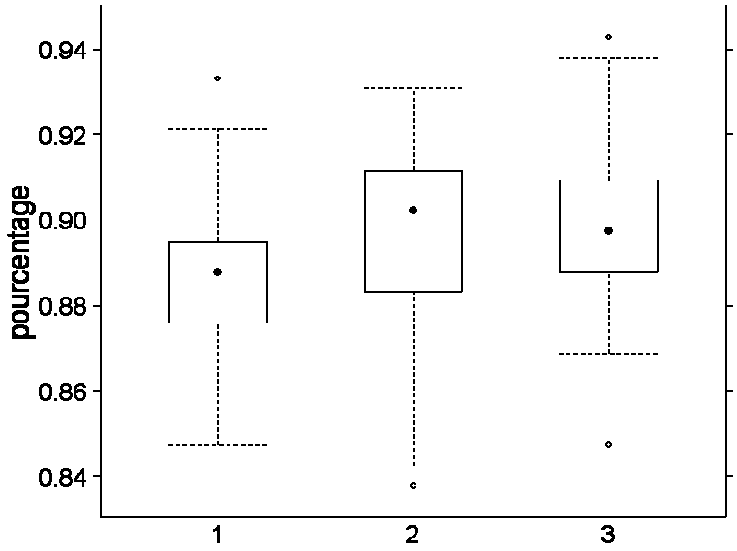


Figure 8.7 : Pourcentage d'accord entre les paires d'auditeurs (auditeurs francophones) par condition

Nous avons ensuite procédé à des tests statistiques avec les modèles linéaires mixtes de l'impact du facteur Condition sur la performance des auditeurs (pourcentage ~ Condition). Remarquons, que notre variable dépendante étant une proportion et donc contrainte à l'intervalle [0, 1], une transformation de variable est nécessaire[245]. L'anova du modèle confirme qu'il existe un effet très significatif du facteur Condition pour les valeurs d'accord des auditeurs russes ($F(2, 40) = 194.63$, $p < 0.0001$), cet effet étant absent pour les auditeurs francophones ($F(2,40) = 2.54$, $p = 0.0911$).

Dans son ensemble, pour la mesure du pourcentage d'accord entre les annotateurs, les valeurs observées sont très élevées, ce résultat pouvant être interprété comme présence d'un accord important. Nous considérons pourtant ces valeurs comme discutables, surtout, dans la mesure où cette valeur élevée est en grande partie due à une proportion importante de cas d'absence d'une frontière prosodique.

Dans un deuxième temps, pour nous assurer du résultat obtenu, nous avons appliqué aux données les statistiques de Kappa de Cohen.

[245] Nous avons appliqué une transformation classique recommandée dans les différents travaux : arcsin(√ X).

8.4.1.6.Coefficient de kappa

Comme nous l'avons mentionné ci-dessus, il existe deux méthodes de calcul de Kappa : la méthode originale de Cohen et celle proposée par Siegel et Castellan. Pour les études avec plus de deux annotateurs, les valeurs de Kappa résultant de deux méthodes peuvent différer considérablement. Le tableau VIII.5 permet d'illustrer ce cas de figure que nous avons pu observer pour nos données (la totalité des données est présentée dans l'annexe 3.2).

LANGUE	CONDITION	KAPPA COHEN	KAPPA SIEGEL
Auditeurs russes	CondR	0.802599	0.459729
	CondD	0.837035	0.542456
	CondN	0.915723	0.764485
Auditeurs francophones	CondR	0.813657	0.478595
	CondD	0.825179	0.50645
	CondN	0.83798	0.551548

Tableau VIII.5 : Coefficients de Kappa

Pour notre part, nous tenons les coefficients de kappa de Siegel et Castellan pour des statistiques plus en accord avec la réalité : les différences principales proviennent du fait que la méthode de Cohen présuppose une indépendance complète des performances des différents auditeurs dans le calcul de la probabilité de l'accord par hasard. Or, des dépendances existent entre les observations pour des paires d'auditeurs : il y a ceux qui manifestent un désaccord systématique ou un degré d'accord plus ou moins important. La méthode de Siegel et Castellan tient compte de ces liens entre juges pour la correction de la probabilité de l'accord dû au hasard.

Si nous appliquons le schème de la classification du degré d'accord exprimé en termes de statistiques Kappa selon Landis et Kokh, seule la performance des auditeurs russes sous la présentation normale des stimuli démontre un « bon » accord entre les juges ; les degrés d'accord observés sous des conditions expérimentales manipulées ainsi que ceux des auditeurs francophones seront caractérisés comme « modérés ».

Nous avons également procédé à une analyse statistique de l'influence du facteur Condition sur les valeurs de Kappa que chaque paire d'auditeurs affiche. Nous avons toujours travaillé en utilisant des modèles linéaires mixtes et avec des variables transformées (le modèle linéaire sous analyse : Kappa ~ condition).

Pour les auditeurs russes, l'anova du modèle témoigne que les valeurs du coefficient de Kappa dépendent de la condition expérimentale ($F(2, 40) = 159{,}167$, $p < 0.001$). Le tableau VIII.6 donne les coefficients estimés du modèle : ces résultats démontrent que les coefficients estimés diffèrent significativement de zéro, la première condition expérimentale étant prise pour référence. En revenant à l'échelle de pourcentage, nous obtenons respectivement des valeurs de Kappa de 0.455, 0.541 et de 0.767 : ainsi la différence entre les conditions CondR et CondN atteint 68.5 %.

	COEFFICIENTS ESTIMÉS	ERREUR STANDARD	DF	VALEUR t	VALEUR p
CondR	0.74	0.0134	40	55.119	0.0000
CondD	0.086	0.019	40	4.51	0.0001
CondN	0.327	0.019	40	17.205	0.0000

Tableau VIII.6 : Effets fixes du coefficient de Kappa en fonction de la condition expérimentale (auditeurs russes)

Pour les auditeurs francophones, l'anova du modèle montre une tendance marginale ($F(2,40) = 3.332$, $p = 0.0458$) à distinguer les différents degrés d'accord entre les paires d'auditeurs selon la condition. Si nous examinons les coefficients estimés pour le modèle (Tableau VIII.7), ils confirment qu'on ne peut parler que d'une tendance, et non pas d'un effet marqué, qui est essentiellement dû à la différence dans les valeurs observées entre la **CondR** (condition rectotonale) et la **CondN** (présentation normale des stimuli) : sur l'échelle de pourcentage, le pourcentage Kappa augmente de 0.473 à 0.54

	COEFFICIENTS ESTIMÉS	ERREUR STANDARD	DF	VALEUR t	VALEUR p
CondR	0.758	0.0223	40	33.323	0.0000
CondD	0.0212	0.27	40	0.8	0.429
CondN	0.0671	0.27	40	2.526	0.0156

Tableau VIII.7 : Effets fixes du coefficient de Kappa en fonction de la condition expérimentale (auditeurs francophones)

Nous avons aussi recensé les statistiques Kappa dans la littérature pour le degré d'accord entre les juges dans les différentes tâches d'annotation de la force / du caractère d'une frontière : ces données sont présentées dans le tableau VIII.8.

ÉTUDE	STATISTIQUES KAPPA COMMUNIQUÉES
Mayo (1996)	0.5
Grice *et al.* (1996)	NA
Syrdal & McGory (2000)	0.65 (voix féminine)[246] 0.62 (voix masculine)
Buhman et al. (2002)	~ 0.75 (frontière forte / faible)[247]
Danielli *et al.* (2004)	~ 0.6 (frontière terminale / non-terminale)

Tableau VIII.8 : Tableau de recensement des valeurs de cœfficient de Kappa communiquées dans la littérature

[246] Remarquons que nous ne disposons pas de détails sur le degré d'accord en termes des statistiques kappa pour chacun des indices de frontières : les frontières que les auditeurs ont indiqué dans notre étude correspondent aux indices de frontières 3 et 4.

[247] Dans le travail cité les valeurs de kappa pour les paires d'annotateurs sont présentées.

Nos valeurs sont moins élevées. Plus précisément, il résulte de notre analyse que dans les conditions dans lesquelles les auditeurs n'avaient pas accès aux informations linguistiques de haut niveau, le degré d'accord entre les annotateurs est plus faible. Remarquons, que nous avons essayé d'argumenter notre choix de la tâche et de la procédure expérimentales : nous n'avons proposé aux sujets ayant participé à notre expérience ni entraînement avec retour, ni définition faisant référence à des critères linguistiques/acoustiques quelconques (une pause), ni une théorie sous-jacente que nous aurions pu leur détailler. Nous avons cherché à tester les représentations que les experts se forment du concept d'unité intonative, concept comparable à celui du syntagme dans la tradition russe, mais tout de même novateur si on le conçoit dans le paradigme de la phonologie prosodique.

8.4.2. Évaluation globale : nombre de frontières

Nous résumons dans le tableau VIII.9 les données sur les pourcentages de frontières syllabiques pour lesquelles au moins un annotateur a signalé la présence d'une frontière prosodique : ces informations sont présentées en fonction de la langue maternelle des auditeurs et de la condition expérimentale.

Nous remarquons, par rapport à ces données, que si pour les sujets francophones les valeurs restent stables d'une condition expérimentale à l'autre (valeur moyenne 27.21% ; écart type 1.04%), une variabilité plus importante est observée dans la performance des auditeurs russes (valeur moyenne 25.3%, écart type 5.98%). En effet, les sujets russes signalent de plus en plus de frontières à mesure que les informations sémantico-syntaxiques et mélodiques disparaissent. Nous pouvons interpréter ces résultats en supposant que sous cette condition expérimentale, ce sont les informations sur la structuration de parole en unités rythmiques (les unités appartenant aux niveaux inférieurs de la hiérarchie prosodique) qui ressortent. Ce sont les moyens linguistiques et prosodiques autres que le rythme qui assurent le groupement des unités rythmiques en structures supérieures, telles que les syntagmes intonatifs/prosodiques.

	AUDITEURS RUSSES	AUDITEURS FRANCOPHONES
CondR	31,50	28,40
CondD	24,82	26,73
CondN	19,57	26,49
Moyenne	25,30	27,21
Écart type	5,98	1,04

Tableau VIII.9. Pourcentage de frontières syllabiques marquées comme des sites de frontières prosodiques (frontières d unité intonative)

Il est également intéressant de remarquer que le nombre de frontières indiquées (et donc, le nombre d'unités intonatives « perçues ») dépend fortement des auditeurs (cf. les figures 8.8 et 8.9 ci-après).

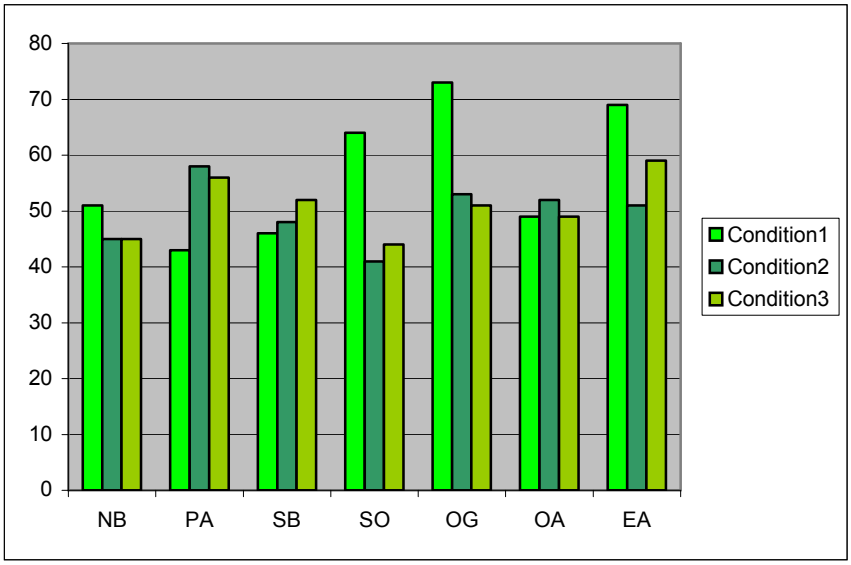


Figure 8.8 : Nombre total des frontières marquées par auditeur (auditeurs russes)

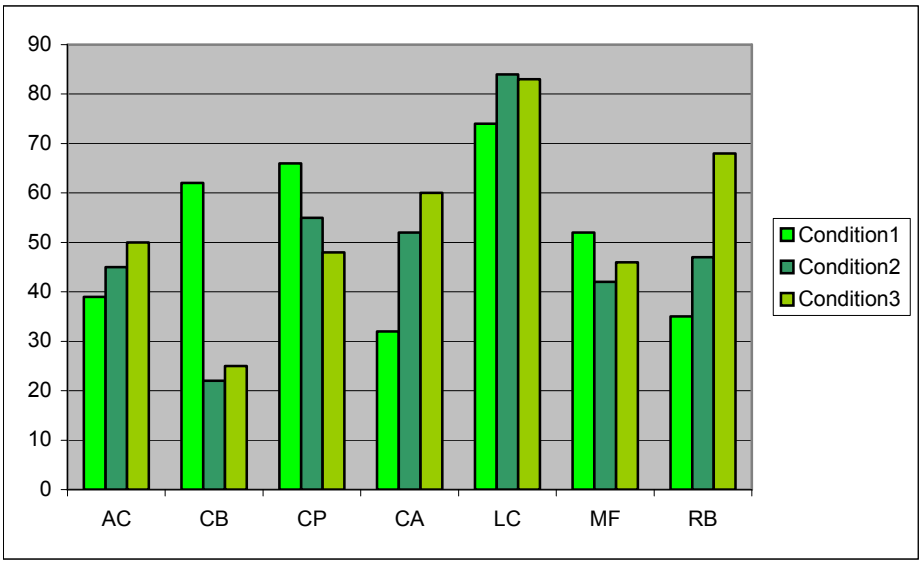


Figure 8.9 : Nombre total des frontières marquées par auditeur (auditeurs francophones)

Nous observons donc ici les stratégies individuelles différentes : le nombre de frontières indiquées diminue entre les première et troisième conditions expérimentales dans la performance des auditeurs russes NB, SO, OG et EA et des auditeurs francophones CB, CP, et MF. Pour d'autres la tendance est renversée : PA et SB chez les auditeurs russes et AC, CA, LC et RB chez les auditeurs français indiquent moins de frontières dans la première condition expérimentale. Seul l'auditeur OB marque un nombre stable de frontières entre les conditions. On observe les mêmes tendances chez les mêmes auditeurs, en comparant leurs performances entre la première et la deuxième condition expérimentale. Si nous comparons ensuite les deuxième et troisième conditions, la répartition des sujets selon les trois stratégies possibles change : pour les auditeurs PA, OG, OB (russes) et CP (francophone) le nombre de frontières diminue, par contre la majorité des auditeurs francophones (5 sur 7, AC, CB, CA, MF et RB) et les auditeurs russes SB, SO et EA indiquent moins de frontières sous présentation normale de stimuli qu'avec la version désémantisée, mais préservant les caractéristiques mélodiques ; pour une minorité des

auditeurs, NB et LC (avec une petite fluctuation tout de même), le nombre de frontières reste stable.

Des tests statistiques ont été réalisés pour valider ces analyses descriptives. Tout d'abord, pour les valeurs du nombre moyen de frontière par stimulus, nous avons appliqué le test statistique de *Kruskal-Wallis*. Ce test permet un aperçu global des résultats : il évalue si les moyennes sont les mêmes pour les trois sous-ensembles de données correspondant à trois conditions. Les valeurs p étant élevées (chi² de Kruskal-Wallis = 4.522, df = 2, p-value = 0.1042 pour les auditeurs russes, chi² de Kruskal-Wallis = 1.6766, df = 2, p-value = 0.432), l'influence du facteur Condition sur le nombre moyen de frontières par stimulus n'est pas significative. Cette analyse nivelle pourtant toute influence des stratégies individuelles, ainsi que la structure interne des données. Nous avons donc cherché à mieux expliciter la structure de nos données ; pour ce faire, nous avons utilisé le modèle linéaire mixte dont nous spécifions le fonctionnement ci-après.

Ces tests ont été accomplis à partir d'un tableau de données qui regroupe des informations sur le nombre de frontières que chacun des auditeurs a marqué pour chaque stimulus ; il contient donc quatre champs : condition expérimentale, identificateur d'auditeur, numéro du stimulus et nombre de frontières marquées.

8.4.2.1.Comparaisons des nombres de frontières indiquées

Nous commençons par une comparaison des nombres de frontières prosodiques repérées selon la condition expérimentale (nombre_de_frontières ~ condition). Nous avons analysé les réponses des auditeurs russes et francophones séparément.

Pour les auditeurs russes, l'anova du modèle linéaire mixte montre un effet significatif du facteur Condition (F(2, 348) = 5,836, p = 0.0032).

	COEFFICIENTS ESTIMÉS	ERREUR STANDARD	DF	VALEUR t	VALEUR p
CondR	2.2743	0.159	348	14.3037	0.0000
CondD	-0.23429	0.0802	348	-2.92249	0.0037
CondN	-0.240	0.0802	348	-2.99378	0.0030

Tableau VIII.10 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes ; intercept = CondR)

Le tableau VIII.10 présente les valeurs des coefficients estimés : chaque coefficient représente l'écart entre la valeur de référence (ici le nombre de frontières signalées en condition CondR) et la valeur concernée (nombre de frontières sous CondD et CondN). Nous voyons donc que les nombres de frontières signalées sous la deuxième et la troisième condition expérimentale sont significativement différents du nombre de frontières que les auditeurs russes ont indiqué sous la première condition expérimentale. Ce premier contraste ne nous permet pas de comparer les performance des auditeurs entre les deuxième et troisième conditions expérimentales, c'est-à-dire, les conditions où les informations prosodiques ont été accessibles aux auditeurs dans leur intégralité. Nous

avons donc changé de point de repère dans l'estimation des coefficients du modèle linéaire : avec ce deuxième contraste, la condition CondD (désémantisée) est prise comme référence (tableau VIII.11).

	COEFFICIENTS ESTIMÉS	ERREUR STANDARD	DF	VALEUR t	VALEUR p
CondD	2.0400	0.159	348	12.8302	0.0000
CondR	0.23429	0.0802	348	2.92249	0.0037
CondN	-0.0057	0.0802	348	-0.07128	0.9432

Tableau VIII.11 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes, intercept = CondD)

Les données présentées dans le tableau VIII.14 nous permettent de conclure, que pour les auditeurs russes, le facteur de Condition n'exerce pas d'influence significative sur la variable du nombre de frontières pour les modalités CondD et CondN.

Les mêmes analyses ont été faites pour les réponses d'auditeurs francophones. L'anova du modèle établit l'absence d'effet significatif du facteur Condition ($F(2, 348) = 1.089$, $p = 0.3377$). L'examen des coefficients et de leur (tableaux VIII.12 et VIII.13, avec les conditions CondR et CondD comme références) confirment l'absence d'effet du facteur Condition sur le nombre de frontières chez les auditeurs francophones.

	COEFFICIENTS ESTIMÉS	ERREUR STANDARD	DF	VALEUR t	VALEUR p
CondR	2.06857	0.1425	348	14.5167	0.0000
CondD	-0.06857	0.0929	348	-0.7379	0.4611
CondN	0.06857	0.0929	348	0.7379	0.4611

Tableau VIII.12 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes, intecept = CondR)

	COEFFICIENTS ESTIMÉS	ERREUR STANDARD	DF	VALEUR t	VALEUR p
CondD	2.0000	0.1425	348	14.0355	0.0000
CondR	0.06857	0.0929	348	0.7379	0.4611
CondN	0.13714	0.0929	348	1.4758	0.1409

Tableau VIII.13 : Effets fixes du nombre de frontières en fonction de la condition expérimentale (auditeurs russes, intecept = CondD)

Nous observons donc un effet significatif du facteur Condition exclusivement pour les sujets russes, la tendance étant de percevoir plus de frontières pour des stimuli filtrés privés des caractéristiques mélodiques. Ce résultat a pour nous une valeur dans le contexte des réflexions théoriques sur le concept de l'unité intonative : un premier groupement s'opère sur la base rythmique ; les indices tonals assurent la formation de structures plus globales et d'unités de(des) niveau(x) supérieur(s). Cependant, ce modèle n'est pas retrouvé chez les auditeurs francophones : ce fait a une explication ancrée dans la tradition des analyses prosodiques de deux écoles (rappelons que nous avons opté pour une tâche métalinguistique proposée aux experts). Le concept de l'unité intonative regroupant plusieurs unités tonales et

rythmiques est depuis longtemps présent dans la tradition française (cf. hiérarchie proposée par Di Cristo et Hirst, 1996). Or, nous avons discuté du statut incertain de l'unité de syntagme de l'école russe : d'après les exemples de la littérature, cette unité pourrait correspondre à un syntagme phonologique/intermédiaire ainsi qu'à un syntagme intonatif. Une structure plus complexe, avec deux niveaux de constituance définies dans le domaine tonal est également possible. Néanmoins, la définition formelle du syntagme présuppose la présence d'un seul centre mélodique (le concept d'accent pré-nucélaire est récent, Igarashi, 2004) : dans la version rectotonale, la syllabe forte du constituant rythmique pourrait s'approprier le statut du centre mélodique, l'unité rythmique étant promue au niveau de l'unité intonative.

8.4.2.2. Position des frontières

Nous avons cherché à voir si les localisations des frontières prosodiques ont été stables d'une condition à l'autre. En effet, le paramètre du nombre de frontières nous fournit une part d'information seulement : la dimension de position est exclue. Nous avons alors comparé, syllabe par syllabe, le score pour toute frontière syllabique (exprimé en nombre d'auditeurs ayant signalé une frontière prosodique) pour toutes les paires des conditions expérimentales. Ainsi dans un premier temps, nous avons obtenu une matrice de contingence, dont le tableau VIII.14 est un exemple (la totalité des matrices est présentée dans l'annexe 4). Cette matrice résume les informations sur les correspondances observées dans la performance des auditeurs russes sous la première (CondR) et la deuxième (CondD) condition expérimentale : si la cellule [0 : 0] contient le nombre 247, cela signifie que pour 247 frontières syllabiques l'absence d'une frontière prosodique à cet endroit était signalée sous les deux conditions expérimentales. Le nombre 4 pour la cellule [7:0] indique que quatre positions à l'unanimité marquées pour la présence d'une frontière prosodique sous la première condition expérimentale, ont été annoté pour l'absence de frontière sous la deuxième condition.

Condition 1 \ Condition2	0	1	2	3	4	5	6	7
0	247	11	12	3	3	4	4	3
1	28	4	5	0	2	0	1	0
2	12	2	4	2	0	1	2	1
3	9	2	1	4	4	0	1	0
4	12	0	5	0	0	1	0	2
5	2	2	1	1	0	0	1	0
6	1	0	1	0	3	2	2	2
7	4	0	1	0	0	0	1	3

Tableau VIII.14 : Exemple d'une matrice de correspondances

Cependant, nous avons jugé les informations du tableau VIII.14 trop précises et difficilement généralisables. Nous avons donc procédé à une catégorisation des frontières, en introduisant trois catégories : « absence de frontière » (si aucun auditeur ne l'a signalée), « frontière optionnelle », signalée par un à trois auditeurs, et « frontière stable, concensuelle », signalée par quatre auditeurs et plus. Ce procédé nous a permis d'obtenir d'autres matrices de contingence (tableau VIII.15), plus concises. Les données de ce tableau nous permettent de voir que parmi 47 positions associées avec une frontière

consensuelle dans la condition avec f0 monotone, 17 maintiennent leur statut lorsque l'information tonale est rajoutée, tandis que 19 positions perdent ce statut. Ce résultat évoque les observations que nous avons déjà formulées précédemment : elles concernent l'émergence des unités rythmiques dans la version rectotonale des stimuli, qui sont ensuite regroupées en des unités supérieures moyennant la distribution des proéminences mélodiques et des contours mélodiques associés.

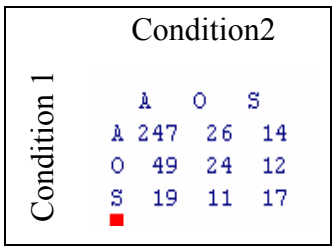


Tableau VIII.15 : Exemple d'une matrice de correspondances avec trois catégories des frontières

C'est le pourcentage des cas sur la diagonale qui nous indique si, selon la catégorie, la localisation de la frontière est la même pour les deux conditions. Le tableau VIII.16 résume cette information. Tout d'abord, nous voyons que la conservation de la frontière est observée de 60 à 72% de cas selon les conditions. Nous constatons un accord moindre entre les conditions extrêmes ; en revanche, l'accord maximal est observé pour deux conditions, dans lesquelles les indices intonatifs sont présents.

	AUDITEURS RUSSES		AUDITEURS FRANÇAIS	
	Pourcentage de diagonale	Pourcentage Kappa	Pourcentage de diagonale	Pourcentage Kappa
CondR *versus* CondD	68,74	29,58	63,48	16,18
CondR *versus* CondN	63,48	13,01	59,67	7,24
CondD *versus* CondN	71,84	24,25	71,84	33,75

Tableau VIII.16 : Pourcentage et pourcentage corrigé pour l'effet de hasard des localisations exactes des frontières à travers les conditions expérimentales

Le tableau VIII.17 ci-après reprend les analyses qui confrontent les réponses d'auditeurs russes et francophones pour des conditions expérimentales respectives. Nous voyons que pour les deuxième et troisième conditions expérimentales le même degré de concordance est observé dans la perspective interlangue que lorsque nous comparons la performance des sujets monolingues entre deux conditions.

	COMPARAISON TRANS-LINGUISTIQUE	
	Pourcentage de diagonale	Pourcentage Kappa
Condition1	60,38	14,19
Condition2	71,36	30,96
Condition3	72,79	29,04

Tableau VIII.17 : Pourcentage et pourcentage corrigé pour l'effet de hasard dans une comparaison trans-linguistique

Les pourcentages de Kappa peuvent pourtant paraître relativement faibles. Ce résultat peut en effet être biaisé par le choix d'un critère très strict, sur lequel nous avons fondé la comparaison : ainsi, nous avons mis en relation toute position syllabique sans prendre en considération les phénomènes accentuels ni les frontières de mots. Cependant, nous ne voyons pas quelle autre unité aurait pu être choisie. D'une part, comme nous l'avons souligné, les frontières des mots n'ont pas été disponibles à tous les auditeurs sous toutes les conditions expérimentales. Ensuite, pour l'unité de pied, quel est type de pied, trochaïque ou ïambique, qu'il faut choisir ? Le nombre de syllabes inaccentuées étant variable pour les pieds phonétiques, dans le cas d'une différence de deux syllabes inaccentuées, s'agit-il d'une même frontière prosodique perçue ? Puisqu'une analyse plus approfondie induisait une intervention considérable de l'expérimentateur dans les données, nous avons choisi de ne pas poursuivre dans cette voie.

Cependant, les données recueillies à ce stade, nous permettent de procéder à une autre évaluation des performances des auditeurs, c'est-à-dire nous interroger pour savoir si les proportions de trois catégories (« absence de frontière », « présence d'une frontière optionnelle », « présence d'une frontière stable ») sont les mêmes selon la condition et la langue maternelle des auditeurs. Il serait bien entendu intéressant de tester cela statistiquement. Le tableau VIII.18 résume les résultats d'un test χ^2, les réponses d'auditeurs russes et français étant séparées ; le tableau VIII.19 reprend les mêmes analyses dans une perspective interlangue. Les fréquences relatives des trois catégories d'étiquettes diffèrent significativement seulement si la comparaison est faite avec la performance des sujets russes sous présentation normale des stimuli (CondN) : la proportion des frontières optionnelles est alors très faible, les sujets montrant un meilleur accord sur le phrasé prosodique. Nous avons déjà discuté de cet accord par rapport aux influences des connaissances théoriques acquises et de l'école russe d'études intonatives.

LANGUE	HYPOTHÈSE TESTÉE	χ^2	VALEUR p
Russe	CondR *versus* CondD	5,4253	0,066
	CondR *versus* CondN	25,8741	2,407e-06
	CondD *versus.* CondN	8,6907	0,01297
Français	CondR *versus* CondD	0,4282	0,8073
	CondR *versus* CondN	1,0852	0,5812
	CondD *versus.* CondN	0,2215	0,8952

Tableau VIII.18 : Différents test chi² qui estiment si les proportions de trois catégories de frontières sont les mêmes entre les conditions comparées

CONDITION	HYPOTHÈSE TESTÉE	χ^2	VALEUR p
I	russe *vs.* français	1,0924	0,5791
II	russe *vs.* français	0,4296	0,8067
III	russe *vs.* français	9,974	0,006826

Tableau VIII.19 : Différents test chi² qui estiment si les proportions de trois catégories de frontières sont les mêmes entre les conditions comparées (perspective interlangue)

8.4.3. Détails de l'organisation rythmique et temporelle

Dans cette dernière section nous cherchons à mettre en évidence quelques caractéristiques temporelles et rythmiques des unités prosodiques qui résultent de la segmentation perceptive des stimuli en unités intonatives.

8.4.3.1.Longueur[248] des syntagmes

Lors de la discussion des méthodes de génération de la structure prosodique en synthèse, nous avons évoqué des méthodes dites rythmiques. Nous avons également cité les contraintes rythmiques sur le phrasé prosodique qui ont pour conséquence la prise en compte systématique de la longueur du syntagme résultant une fois que la décision est prise sur l'insertion d'une frontière prosodique. Évoquons encore le travail de Gee et Grosjean (1983), dont les conclusions portent sur le rôle accru des contraintes rythmiques. Compte tenu de ces arguments, nous avons analysé l'organisation rythmico-temporelle des unités prosodiques résultantes. Plus particulièrement, nous nous sommes posée la question de savoir dans quelle mesure les auditeurs avaient utilisé leurs connaissances et leur intuition sur la durée et la longueur potentielle de l'unité intonative.

Dans un premier temps, nous nous sommes intéressée à la taille des unités prosodiques. Le nombre des unités prosodiques résultant du découpage par les sujets a été collecté. Le tableau VIII.20 résume les valeurs de moyenne, d'écart type et de coefficient de variation (ratio entre l'écart type et la moyenne) pour chaque condition expérimentale et les deux groupes de sujets. La longueur moyenne de l'unité prosodique résultante était de 5,73 syllabes chez les auditeurs russes (écart type 2,998 syllabes) et de 5,77 syllabes chez les auditeurs français (écart type 2,98 syllabes). Nous constatons également une stabilité des valeurs d'une condition à l'autre.

LANGUE	CONDITION	MOYENNE	ÉCART TYPE	COEFFICIENT DE VARIATION
Russe	I	5,45	2,88	0,528
	II	5,931	2,913	0,491
	III	5,843	3,184	0,545
Français	I	5,799	2,94	0,507
	II	5,931	3,053	0,515
	III	5,59	2,908	0,520

Tableau VIII.20 : Données sur la longueur des unités prosodiques perçues (longueur exprimée en nombre de syllabes)

Par référence aux données du tableau VIII.23, remarquons que le coefficient de variation est élevé : plus la valeur du coefficient de variation est élevée plus la dispersion autour de la moyenne est grande. Un coefficient de variation d'environ 0,5 témoigne d'une dispersion

[248] Une précision terminologique nous paraît importante : nous faisons une distinction systématique entre la durée, exprimée en ms, et la longueur, mesurée en nombre d'unités linguistiques (en l'occurrence, en nombre de syllabes).

considérable des valeurs autour de la moyenne. Nous nous sommes alors intéressée aux stratégies individuelles des auditeurs : la figure 8.10[249] détaille l'analyse par locuteur et par condition. D'après ces représentations graphiques, pour les auditeurs russes, les fluctuations de la médiane sont importantes sous la première (CondR) et la troisième (CondN) condition expérimentale. De même, c'est pour la deuxième condition expérimentale que l'écart interquartile est le plus stable, la variabilité élevée étant propre à la troisième condition. Ce résultat n'est pas surprenant et s'inscrit dans la tendance dont nous avons déjà discuté : en effet, nous avons souligné la dépendance de l'analyse prosodique de la structure syntaxique pour l'école russe ; or, les constituants syntaxiques de différents niveaux démontrent une variabilité plus importante de leur longueur que ne le font les constituants prosodiques (cf. la tendance vers l'eurythmie débattue ci-dessus). Chez les sujets francophones nous tenons à attirer l'attention sur deux stratégies individuelles controversées. Tout d'abord, il s'agit du sujet CB qui exhibe une tendance vers le marquage des unités prosodiques (unités intonatives) plus longues. À contretemps, la performance du sujet LC est caractérisée par une stratégie consistant à démarquer des unités les plus petites.

[249] La représentation y a la forme d'un *box-plot* : le point noir central représente la médiane (les tests montrent que pour les données discutées les valeurs de médiane sont proches des moyennes respectives) ; la boîte rectangulaire correspond à l'écart interquartile, les lignes pointillées désignant l'étendue de la distribution et les points non-remplis – les valeurs atypiques.

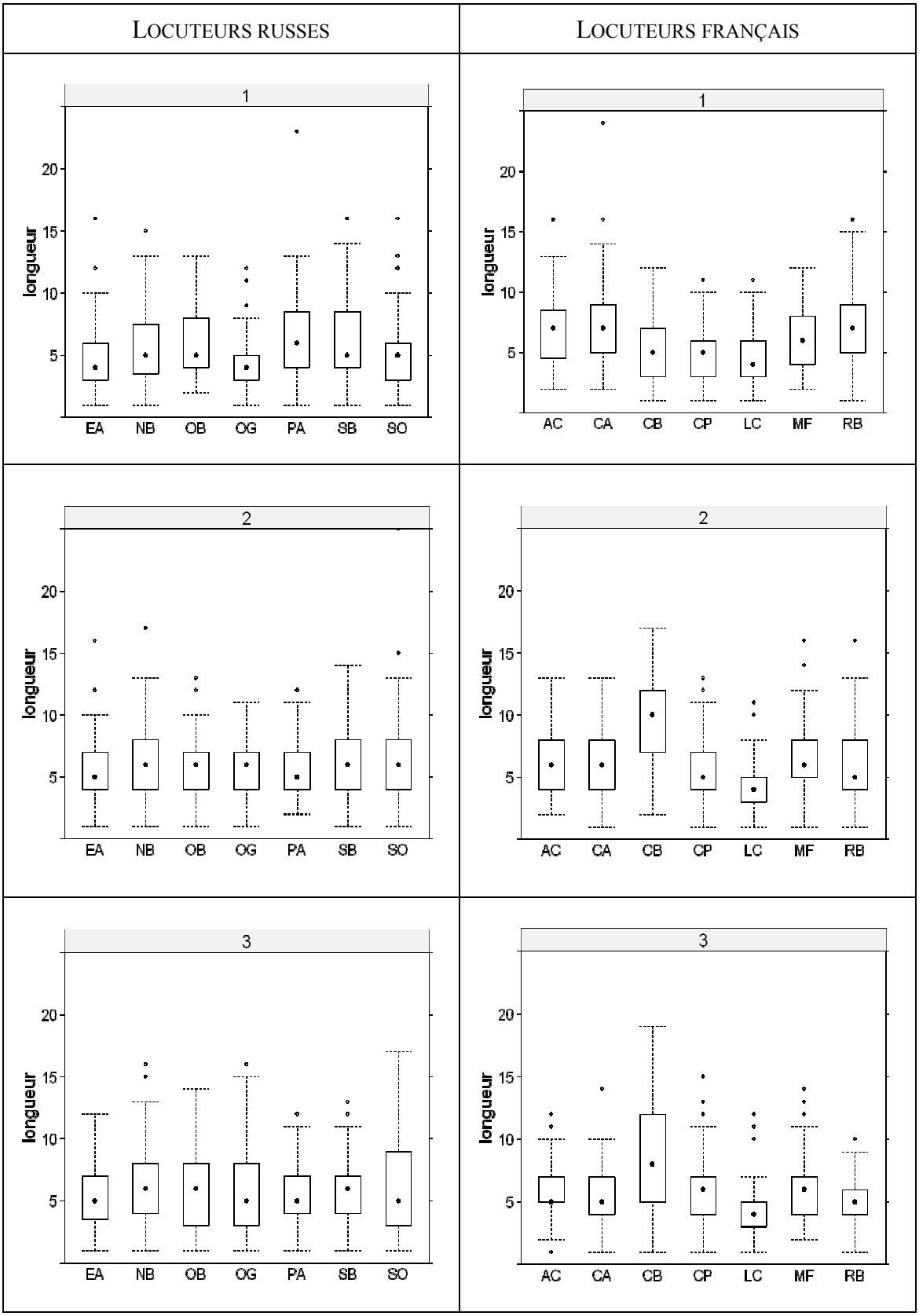


Figure 8.10 : Box-plot par locuteur de la longueur des unités prosodiques perçues selon la condition expérimentale et la langue maternelle des sujets (en ordonnée – longueur du syntagme en syllabes). Panel gauche : auditeurs russes ; panel droit : auditeurs français)

En outre, nous avons calculé la répartition des unités prosodiques résultantes en fonction du nombre de syllabes qu'elles contiennent. Les résultats sont présentés dans les figures 8.11 et 8.12.

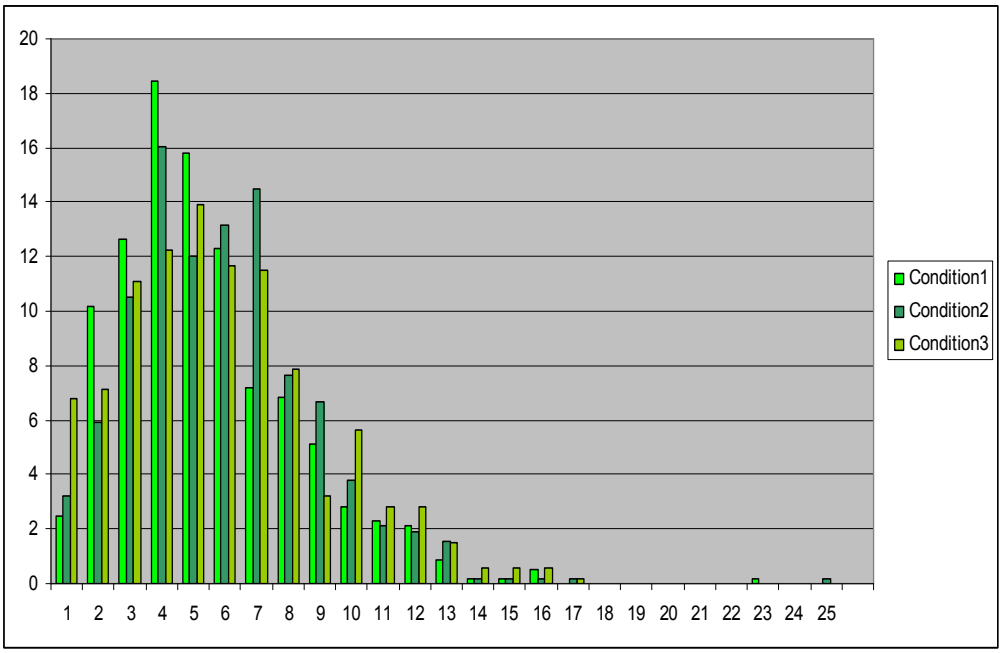


Figure 8.11 : Répartitions des unités prosodiques perçues en fonction de leur nombre de syllabes. Auditeurs russes

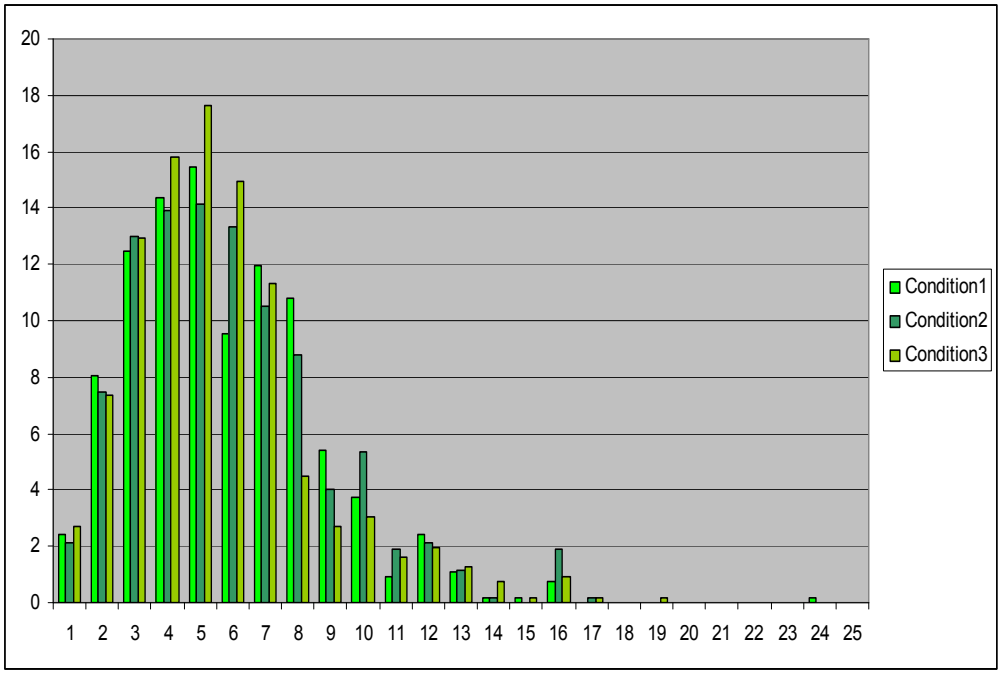


Figure 8.12 : Répartitions des unités prosodiques perçues en fonction de leur nombre de syllabes. Auditeurs français

Nous avons ensuite calculé les proportions des cas où la longueur de l'unité est inférieure à 3 syllabes ou contenue dans l'intervalle entre 3 et 6 syllabes. Ces données sont résumées dans le tableau VIII.21.

LANGUE	CONDITION	MOINS DE 3 SYLLABES	ENTRE 3 ET 6 SYLLABES	TOTAL
Russe	I	12,632	59,123	71,754
	II	9,160	51,718	60,878
	III	13,910	48,872	62,782
Français	I	10,448	51,866	62,313
	II	9,542	54,389	63,931
	III	10,448	53,777	64,225

Tableau VIII.21 : Pourcentages des unités de longueur spécifiée

En plus des remarques qui ont été faites précédemment sur la variation du nombre de frontières selon la condition, nous tenons à attirer l'attention sur un pourcentage élevé d'unités de moins de 3 syllabes sous la troisième condition expérimentale chez les auditeurs russophones. En effet, lors de l'accomplissement de la tâche, les informations prosodiques, sémantiques et syntaxiques ont été accessibles à ces auditeurs. Nous considérons donc que le nombre élevé des cas où un seul mot constitue le syntagme prosodique correspond aux cas des structures syntaxiques particulières (prophrases ou adverbes). Nous avons évoqué dans le chapitre dédié à la discussion du phrasé prosodique les considérations théoriques qui présument pour certains éléments syntaxiques le potentiel de former l'unité intonative à eux seuls (cf. Astruc 2003). Or, la question est de savoir si ces syntagmes satisfont la contrainte formelle, à savoir, un ton de frontière et un accent mélodique doivent être présents (cf. approche de Gussenhoven 1990).

Dans un deuxième temps, les données du tableau VIII.24 témoignent que plus de 60% des unités contiennent moins de 6 syllabes. Nous avons choisi ce seuil de 6 syllabes à la suite des travaux de Delais-Roussarie (1996) sur la taille minimale et maximale du syntagme accentuel (syntagme phonologique) en français : par syntagme accentuel, l'auteur entend une unité de l'organisation prosodique, et plus particulièrement de la structure rythmique, inférieure à l'unité intonative. Ainsi plus de 60% des unités issues de l'expérience en perception, correspondent aux syntagmes accentuels/phonologiques. Dans le cadre de la phonologie des domaines, à laquelle nous faisons référence tout au long de cette étude, aucun principe n'interdit qu'un syntagme phonologique forme un syntagme intonatif à lui seul. Or, la mise en pratique majoritaire de cette tendance aura un effet stylistique, à savoir, l'impression d'un style rhétorique et soutenu, éloigné de celui de notre corpus.

À partir de ces considérations, nous pouvons émettre une hypothèse selon laquelle deux unités de l'organisation prosodique reçoivent un marquage en termes d'indices acoustiques en russe au niveau supérieur à la syllabe. Cependant, il serait nécessaire de vérifier dans une étude expérimentale, dans quelle mesure ces unités peuvent être mises en relation avec les unités de mot et de syntagme selon l'école russe ; mais également quels indices acoustiques participent au démarquage de chacune de ces unités. Simultanément, nos résultats indiquent une certaine ambiguïté du statut du syntagme en tant qu'unité standard de l'analyse intonative du russe : cette unité, correspond-elle au syntagme phonologique ou au syntagme intonatif dans la hiérarchie prosodique définie en phonologie des domaines (Nespor & Vogel, 1986) ?

8.4.3.2. Durées absolues des syntagmes

Lors de différentes étapes de traitement et de pré-traitement des données, nous avons constaté, au moins chez certains auditeurs, la tendance à proposer une segmentation plus fragmentaire au début du stimulus et au contraire des unités prosodiques terminales plus longues. La figure 8.13 illustre ce phénomène : nous avons repris les segmentations que les auditeurs russes et français ont proposées pour le stimulus 12 sous la deuxième condition expérimentale.

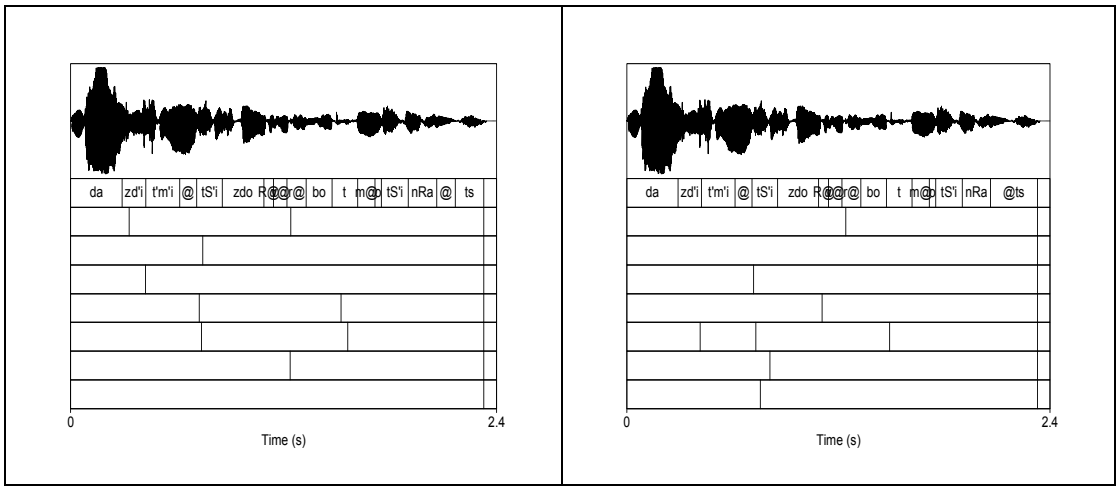


Figure 8.13 : TextGrids sommaires des annotations proposées pour le stimulus 12 sous la deuxième condition expérimentale (version désémantisée, préservant les propriétés prosodiques). Panel gauche : auditeurs russes ; panel droit : auditeurs francophones

La figure 8.14 montre la distribution de la durée de syntagmes en fonction du type d'unité et de la condition expérimentale dans la performance des auditeurs russes. La représentation graphique laisse supposer que la durée des syntagmes terminaux est légèrement plus grande.

Nous avons donc voulu vérifier statistiquement si les durées des unités intonatives terminales diffèrent significativement des durées des unités non-terminales. Pour ce faire, nous avons utilisé un modèle linéaire mixte : nous nous sommes intéressée aux effets des facteurs expérimentaux « condition » et « type d'unité intonative » (les deux niveaux de ce facteurs étant « unité terminale » et « unité non-terminale »), avec les facteurs de bloc « stimulus » et « sujet ». Nous testons au total six hypothèses : premièrement, nous comparons par condition les durées des unités terminales et non-terminales ; dans un deuxième temps, nous comparons entre elles les durées des syntagmes non-terminales en fonction de la condition expérimentale. La validation statistique de nos hypothèses nécessite le test simultané de 3 jeux de contrastes distincts (voir détails en annexe 5) ; de ce fait, après correction de Bonferroni, le seuil de significativité retenu est alors égal à 0.017 (0.05/3).

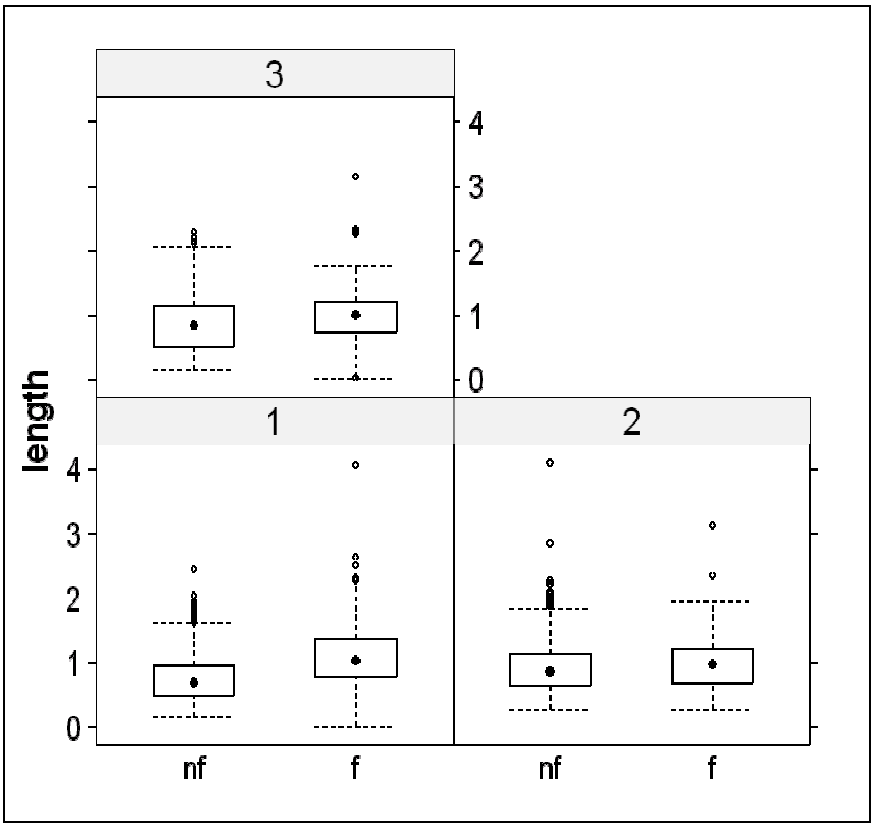


Figure 8.14 : Box-plots par condition des durées des unités prosodiques perçues. Auditeurs russes. Légende : nf – unités non-terminales, f – unités terminales

Nous résumons dans le tableau VIII.22 les résultats des tests, comparant les durées des unités intonatives terminales *versus* non-terminales.

CONDITION	Δ DE DURÉE ESTIMÉ (ms)	VALEUR t	VALEUR p
CondR	307,7	8,028	0,0000
CondD	35,33	0,904	0,366
CondN	143,4	3,688	0,0002

Tableau VIII.22 : Données de la validation statistique des hypothèses sur la différence des durées entre les unités intonatives terminales versus non-terminales (auditeurs russes)

Nous observons donc, que la différence en durée entre les syntagmes terminaux et non-terminaux est significative sous les première (CondR) et troisième (CondN) conditions expérimentales : en moyenne un syntagme terminal étant de 308 ms plus long en CondR et de 143 ms plus long en CondN. Cependant, les deux catégories de syntagmes n'ont pas des durées significativement différentes sous la deuxième condition (CondD). Nous pouvons interpréter le rapprochement entre les première et troisième conditions, en évoquant un certain parallélisme entre les structures prosodique et syntaxique que nous avons constaté chez les auditeurs russes pour la condition CondN. En effet, les mots introductifs, les prophrases, les syntagmes prépositionnels dans la fonction du complémenteur général de l'énoncé, correspondent souvent à un mot, positionné au début de l'énoncé ; de ce fait, ils induisent une variabilité importante intra-catégorie.

Nous avons ensuite comparé les durées des unités intonatives non-terminales entre les conditions (tableau VIII.23) : nous constatons une tendance à marquer des unités plus courtes dans la condition rectotonale, les unités non-terminales étant d'une durée similaire dans deux conditions où les informations mélodiques ont été accessibles aux auditeurs. Ces résultats confirment les tendances déjà remarquées pour le paramètre du nombre de frontières.

HYPOTHÈSE	Δ DE DURÉE ESTIMÉ (MS)	VALEUR t	VALEUR p
CondR *versus* CondD	-156	5,015	0,0000
CondD *versus* CondN	55	1,74	0,0828
CondR *versus* CondN	-101	3,274	0,001

Tableau VIII.23 : Comparaison des durées des syntagmes non-terminaux (auditeurs russes)

Nous avons ensuite mené les mêmes analyses sur les données des sujets francophones. Comme le montre la figure 8.15, les différences entre deux types d'unités paraissent encore plus remarquables chez les auditeurs francophones.

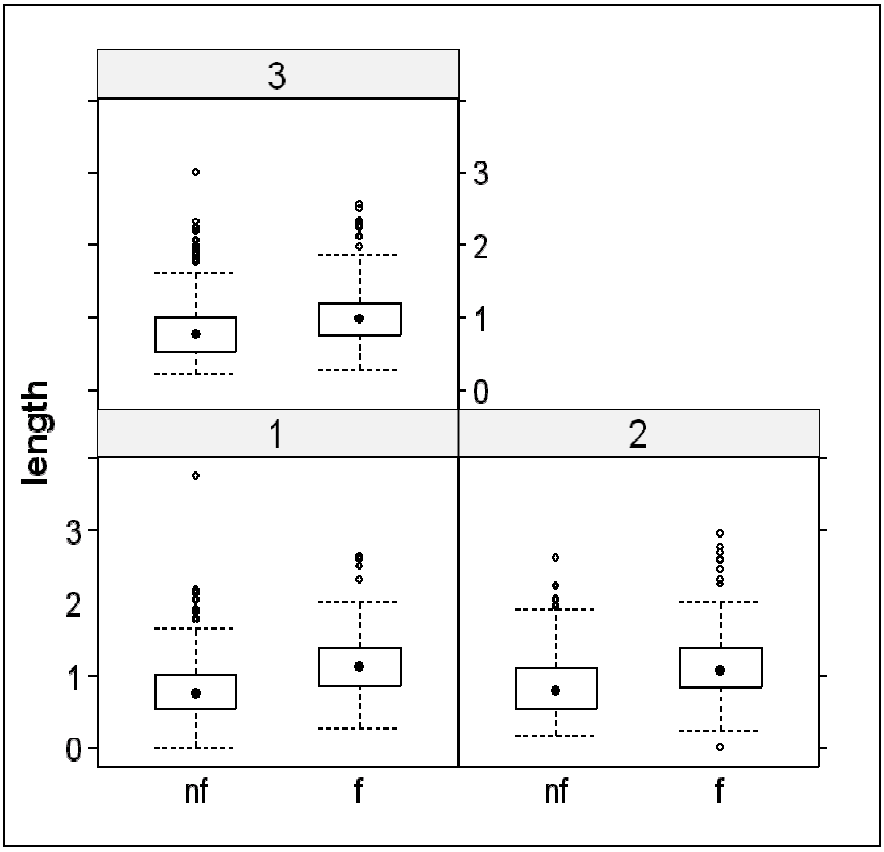


Figure 8.15 : Box-plots par condition des durées des unités prosodiques perçues. Auditeurs francophones. Légende : nf – unités non-terminales, f – unités terminales

Nous comparons d'abord les durées des unités intonatives non-terminales et terminales (tableau VIII.24) : sous toutes les conditions expérimentales, les différences de durées sont significatives, en moyenne, les unités non-terminales étant respectivement de 329, 221 et 192 ms plus courtes.

CONDITION	$\Delta_{T\text{-}NT}$ DE DURÉE ESTIMÉ (ms)	VALEUR t	VALEUR p
CondR	329,35	9,445	0,0000
CondD	221,65	6,298	0,0000
CondN	191,95	5,532	0,0000

Tableau VIII.24 : La validation statistique des hypothèses sur l'écart de durée entre les unités intonatives terminales (T) et non-terminales(NT) (auditeurs francophones)

Nous pouvons nous interroger sur cette tendance à constituer des unités intonatives plus longues à la fin d'un groupe interpausal, observée chez les deux groupes d'auditeurs. S'agit-il d'une tendance universelle, reflétant les rapports entre les structures prosodique et syntaxique que nous avons évoquées dans la discussion de la performance des auditeurs russe ? Dans le même esprit, l'énoncé français « *Oui, mon ami dont tu parles, il est parti récemment pour l'Argentine* » peut présupposer une annotation plus fragmentaire à son début (tout dépendra, bien sûr, de la prononciation réelle)[250]. Il peut s'agir aussi de l'effet induit par l'atténuation de tous les indices acoustiques vers la fin de l'énoncé/unité interpausale, ce qui rend la segmentation plus difficile et dépendante des informations autres que prosodiques : nous pensons notamment aux phénomènes de déclinaison et d'affaiblissement général de l'intensité.

Si nous comparons les unités non-terminales résultant de l'annotation des auditeurs francophones (tableau VIII.25), nous constatons d'abord une magnitude de différence moins importante que celle observée chez des auditeurs russes. Ce sont des unités non-terminales annotées sous la deuxième condition expérimentale (CondD) qui ont la durée la plus longue et significativement différente de celles de deux autres conditions.

HYPOTHÈSE	$\Delta_{T\text{-}NT}$ DE DURÉE ESTIMÉE (ms)	VALEUR t	VALEUR p
CondR *versus* CondD	- 76	2,656	0,008
CondD *versus* CondN	52	1,843	0,0655
CondR *versus* CondN	- 24	0,848	0,396

Tableau VIII.25 : Comparaison des durées des syntagmes non-terminaux (auditeurs francophones)

Dans la perspective d'une modélisation acoustique du phrasé prosodique dont nous discutons en détail dans le chapitre X, nous avons besoin d'un modèle temporel de l'unité intonative. À partir des analyses que nous avons présentées dans cette section, nous avons choisi de fonder nos calculs sur la durée moyenne des syntagmes non terminaux : ce choix est étayé par la constatation d'une forte tendance chez les auditeurs à une segmentation plus fragmentaire à l'initiale des stimuli[251]. Le tableau VIII.26 ci-après résume les informations sur la durée moyenne et l'écart type observés par condition pour les syntagmes non-terminaux. Remarquons que ces valeurs sont d'un tiers moins élevées que les valeurs communiquées traditionnellement pour la durée des syntagmes intonatifs en

[250] Ainsi que l'énoncé « *Dans un an, tu seras probablement ailleurs* », pour lequel la virgule correspond au site potentiel d'une frontière intonative.

[251] Nous avons attesté quelques différences significatives entre les durées des syntagmes non-terminaux : cependant, nous jugeons la magnitude de l'effet comparable, voire plus faible, que les valeurs de l'écart type, caractérisant la variabilité au sein de la catégorie.

langue russe (1,2 s.). Deux interprétations peuvent être proposées : d'une part, il s'agit probablement d'une unité de perception ; d'autre part, ces données vont dans le sens de l'existence en russe de deux niveaux de la structuration prosodique, similaires aux syntagmes phonologiques et syntagmes intonatifs. Cette deuxième interprétation-hypothèse nécessitera pourtant une vérification expérimentale approfondie, qui dépasse le cadre de la présente étude. Si cette hypothèse est confirmée, il serait également intéressant de s'interroger sur des indices acoustiques associés à l'actualisation de chaque niveau du phrasé à l'oral.

LANGUE	CONDITION	MOYENNE (S)	ÉCART TYPE (S)
Russe	CondR	0,785	0,395
	CondD	0,95	0,489
	CondN	0,898	0,457
Français	CondR	0,83	0,425
	CondD	0,874	0,432
	CondN	0,833	0,427

Tableau VIII.26 : Durée moyenne et écart type par condition des syntagmes prosodiques non-terminaux

8.5. CONCLUSIONS ET DISCUSSION

Dans le présent chapitre, nous avons discuté sous divers angles des résultats obtenus dans l'expérience en perception, menée dans l'objectif de disposer de données sur la perception du phrasé prosodique du russe. Nous avons proposé aux auditeurs russes et français (non-russes) une tâche d'annotation des frontières prosodiques, avec deux niveaux du facteur « absence *versus* présence d'une frontière ». Nous avons cherché à remettre en cause le parallélisme propre à la tradition russe entre les structures prosodique et syntaxique. En effet, dans la discussion présentée, nous avons souligné à plusieurs reprises le statut particulier pour des sujets russes de la condition expérimentale avec la présentation des stimuli non-manipulés. Or, les similitudes dans la performance des auditeurs français et des auditeurs russes sous d'autres conditions expérimentales sont pour nous aussi importantes : ces résultats suggèrent entre autres une possibilité d'aboutir à un modèle acoustique et à un modèle formel du syntagme en tant qu'unité de base de l'analyse intonative.

Une autre tendance émerge dans l'opposition de la performance d'auteurs russes dans les réponses obtenues pour la condition rectotonale et la condition désémantisée mais préservant des caractéristiques tonales. Nous avons constaté que les unités délimitées dans le premier cas étaient plus petites (et le nombre de frontières plus élevé) : nous y voyons les preuves d'une construction progressive de la structure prosodique. Effectivement, le processus de la formation des syntagmes intonatifs se résume à un groupement, d'unités plus petites, appartenant au domaine métrique (syntagme accentuel de Delais-Roussarie, 1996, cf. également la hiérarchie prosodique de Di Cristo et Hirst, 1996) ; ce groupement est assuré moyennant des indices acoustiques tonals.

Nous voulons tout particulièrement insister sur des résultats obtenus lors de l'analyse de certains détails de l'organisation temporelle des unités perçues. Ces réflexions sont

motivées par la discussion des contraintes rythmiques sur le phrasé prosodique dans la littérature. Notre attention a été prêtée au paramètre de longueur d'unités, exprimée en nombre de syllabes. Le bilan de cette analyse suggère que plus de 60% des unités perçues ont une longueur inférieure à six syllabes. Ce seuil a été choisi, compte tenu des réflexions sur la longueur de l'unité du groupe rythmique (Padesloup 1992), ou du syntagme accentuel (Delais-Roussarie, 2000), pour lesquelles les processus de restructuration commencent à partir de ce seuil. Comme nous l'avons déjà souligné, aucune règle n'interdit une correspondance biunivoque entre le syntagme intonatif et le syntagme accentuel ; or, ce type de phrasé est stylistiquement marqué (style rhétorique, soutenu), le style de notre corpus s'en écartant considérablement. De ce fait, nous pouvons émettre une hypothèse alternative vis-à-vis des résultats discutés : nous pouvons supposer l'existence d'un niveau intermédiaire du phrasé, recevant à l'oral un marquage moyennant un recours à des indices acoustiques. Ce sera l'objectif d'études futures de recueillir des preuves, confirmant ou contredisant cette hypothèse. Si notre hypothèse est confirmée, c'est la quête d'indices acoustiques assurant la perception de deux unités, syntagme accentuel (le terme étant provisoire) et syntagme intonatif, qui sera au centre des recherches, surtout en relation avec le modèle de Beckman et Edwards (cf. *supra*, page42).

Nous voudrions aussi attirer l'attention sur la démarche de validation des résultats d'une annotation dans une analyse de concordance entre les auditeurs. L'annotation du corpus, étant par sa logique un processus d'enrichissement des données brutes avec des représentations symboliques, peut être effectuée aux différents niveaux linguistiques. Les schémas d'annotation dépendront alors des théories linguistiques de référence. L'analyse de concordance entre les annotateurs, permettant d'évaluer la reproductibilité de l'annotation et son indépendance de l'annotateur, est alors pour nous un moyen d'expertiser la portée épistémologique de la théorie. L'adoption de cette démarche est d'autant plus importante dans une perspective du corpus multiniveaux, conjuguant plusieurs annotations : vu la nécessité de disposer de grilles d'évaluation caractérisant l'efficacité et la cohérence de tout schéma d'annotation, et vu l'objectif des tests systématiques d'hypothèses sur les rapports entre les phénomènes appartenant aux différents niveaux.

CHAPITRE IX
CONTRAINTES SUR LE PHRASÉ PROSODIQUE

En cherchant une heuristique pour une segmentation automatique de notre corpus en unités intonatives, nous avons réalisé une expérience en perception dont nous avons présenté les détails dans le chapitre précédent. Une fois les réponses des auditeurs recueillies, se pose la question de l'identification des mécanismes linguistiques qui ont pu sous-tendre la performance de chacun d'eux. Dans les chapitres introductifs nous avons insisté sur le besoin de conjuguer dans l'analyse des phénomènes prosodiques qui se veut formelle, des traitements bidirectionnels des données : la démarche ascendante (*bottom-up)* et descendante (*top-down)*. En appliquant ce principe à la lettre, nous avons décidé de confronter le phrasé prosodique qui ressort de notre étude en perception (la distribution des frontières indiquées par les auditeurs) avec les différentes prédictions théoriques.

Remarquons, que nous avons demandé à nos informateurs de déterminer les frontières des unités intonatives : il est convenu de manière générale, que, par opposition aux niveaux inférieurs de la hiérarchie prosodique où les principes de la formation de constituants sont très stricts, une variabilité est permise au locuteur quant à la structuration de l'énoncé en termes de syntagmes intonatifs, compte tenu des contraintes discursives. Cependant, dans notre étude nous avons fait le choix méthodologique de nous placer dans la perspective de l'auditeur et de la perception : les choix ont déjà été faits par la locutrice à l'étape de la planification du message, ceux-ci ont reçu une implémentation en production et il est dorénavant nécessaire de les reconnaître et d'évaluer leur efficacité dans la transmission du message.

Nous avons écarté l'idée de procéder à une présentation au cas par cas des structures prosodiques résultantes dans leurs interactions avec les niveaux syntaxique et discursif. Cherchant un moyen de résumer et de systématiser les données, nous avons opté pour une analyse dans le cadre de la Théorie de l'optimalité. Ce choix est motivé par les considérations suivantes : nous avons attesté une variabilité considérable dans la performance des auditeurs dans le cadre de l'analyse de la concordance. L'adoption d'une méthodologie par règles ne permettra pas une généralisation suffisante des données empiriques, tandis que le cadre de la théorie de l'optimalité, plus souple, permettra une interprétation plus subtile des données. Des considérations similaires sont proposées par Post (2000) dans son analyse du phrasé prosodique en français.

Dans le présent chapitre, nous présenterons un certain nombre d'hypothèses sur les contraintes postulées dans le cadre de la théorie de l'optimalité : en particulier, nous procédons à leur validation en les confrontant aux données de l'expérience perceptive cherchant ainsi à rendre compte du phrasé prosodique en parole russe spontanée. Nous entreprenons donc une mise en relation entre les structures perçues et les structures linguistiques (syntaxiques et discursives) : par la suite, ce cadre explicatif permettra d'approfondir les modèles prosodiques et, lors de la génération de parole conversationnelle de choisir la forme adéquate.

Nous commençons par cerner brièvement le cadre explicatif de la théorie de l'optimalité. Nous passerons ensuite en revue les études qui ont déjà essayé d'appliquer ce formalisme lors du traitement de la distribution des frontières prosodiques. Nous confronterons ensuite les contraintes retenues aux données observables.

9.1. LE CADRE EXPLICATIF DE LA THÉORIE DE L'OPTIMALITÉ

Dans le cadre de la théorie de l'optimalité, le phrasé prosodique, comme tout autre phénomène langagier, reçoit un traitement en termes d'interactions entre différentes contraintes. Les contraintes proposées par la théorie de l'optimalité sont considérées comme universelles, et c'est leur ordre respectif qui détermine les particularités de l'organisation prosodique d'une langue.

Le modèle procède ainsi : pour toute forme sous-jacente, le module de Génération produit un nombre de formes sous-jacentes. On cherche par la suite à ne garder que la (les) forme(s) qui satisfait (satisfont) les contraintes postulées pour la langue en question. Le processus d'évaluation s'effectue à l'aide d'un autre module EVAL (Évaluateur) : la forme qui satisfait le plus de contraintes de haut niveau (dans la hiérarchie des contraintes) est considérée comme optimale. Il est admis également qu'on n'établit pas de rapports de dominance pour deux contraintes dans la hiérarchie[252] : on aboutit alors à une structure avec un classement partiel (*partial ranking*) des contraintes (Antilla 1997, cités par Post, 2000). Un système d'annotation est adopté (cf. Tableau IX.1) : la procédure d'évaluation est présentée sous la forme d'un tableau, avec l'ordre des contraintes de gauche à droite correspondant à la hiérarchie des contraintes (ordre décroissant) ; dans le cas d'une structure partiellement non-classée, les contraintes concernées sont séparées par une ligne en pointillés. Dans le processus d'évaluation, « * » désigne la violation d'une contrainte, et « * ! » une violation cruciale après laquelle le candidat est éliminé ; le symbole «☞ » pointe le candidat optimal. Les contraintes de la théorie de l'optimalité ont la propriété importante d'être transgressables : la violation d'une contrainte est autorisée si seulement elle permet de satisfaire une contrainte supérieure dans la hiérarchie.

On postule le caractère universel des contraintes de la théorie de l'optimalité : il nous a donc paru intéressant de tester dans quelle mesure les contraintes postulées dans les études portant sur d'autres langues, sont applicables au russe, et en particulier, aux données sur la perception du phrasé prosodique dans les stimuli issus du corpus spontané. Notre réflexion

[252] Si le classement (*ranking*) de toutes les contraintes ne désigne aucun candidat optimal.

n'aboutira pas à une hiérarchie finale des contraintes sur le phrasé prosodique en russe, mais nous en fournirons quelques éléments.

Candidats	*Contrainte A*	*Contrainte B*
☞ $Cand_1$		*
$Cand_2$	* !	

Tableau IX.1 : Exemple d'un tableau de contraintes. Dans cet exemple, la contrainte A domine la contrainte B

9.1.1. Études précédentes

Nous avons évoqué dans le chapitre consacré à la formalisation du phrasé prosodique dans le cadre de la phonologie prosodique que l'hypothèse d'étagement strict (le pivot fondamental de cette théorie) a été reformulée en termes de contraintes universelles transgressables (Selkirk 1995b, cf. chapitre V).

Les contraintes à différents niveaux du phrasé prosodique ont été précédemment discutées dans les travaux de Delais-Roussarie (1996) et Post (2000) pour le français, par Selkirk (2000) pour l'anglais, Prieto (sous presse) pour l'espagnol et Truckenbrodt (2005) pour l'allemand. De manière générale, nous pouvons distinguer deux types de contraintes mises en évidence dans les études précédentes : les contraintes d'alignement généralisées, qui procèdent à une mise en correspondance systématique de la structure syntaxique ou informationnelle avec la structure prosodique, et les contraintes métriques[253], qui agissent sur la distribution des accents mélodiques ou qui déterminent la longueur du constituant en question.

L'analyse présentée par Selkirk (2000), quoique spéculative[254], aborde un large ensemble de contraintes sur le phrasé prosodique au niveau du syntagme intonatif pour l'anglais. Le premier groupe des contraintes concerne essentiellement les interfaces entre la phonologie et d'autres modules de la grammaire. Ainsi deux contraintes régissent-elles l'interface « prosodie – syntaxe » : c'est, d'une part, la contrainte ALIGN-XP, qui est dérivée à partir de la contrainte universelle d'alignement de McCarthy et Prince (1993) :

(1) Contrainte généralisée d'alignement :

ALIGN (αCat, E ; βCat, E) etc.

Cette contrainte est motivée par l'approche de mise en relation des bornes des constituants de différents modules de la grammaire développée dans les études de Selkirk. Dans une perspective modulaire, la structure syntaxique qui sert d'entrée pour le module phonologique est traditionnellement conçue par référence à la théorie de la syntaxe X-barre (Jackendoff 1977). Ainsi la contrainte d'alignement ALIGN-XP,R impose-t-elle l'insertion

[253] Les autres termes utilisés étant « contraintes rythmiques » ou « contraintes phonologiques ».
[254] Non-confirmée par l'analyse des productions réelles.

d'une frontière prosodique après toute projection XP dans la structure syntaxique (après toute tête lexicale). Or, dans l'étude de Truckenbrodt (1999) une deuxième contrainte, WRAP-XP est proposée s'inspirant des approches qui intègrent les relations « tête ~ complément » dans leurs algorithmes de gestion des interfaces « prosodie – syntaxe ». C'est une contrainte d'unité du constituant syntaxique qui demande que les éléments d'un constituant morphosyntaxique du type XP (projection maximale) dans la forme d'entrée soient obligatoirement contenus dans un seul constituant prosodique (en l'occurrence, Syntagme Majeur de la hiérarchie de Selkirk) dans la représentation de sortie.

(2) WRAP-XP

Chaque XP (projection maximale) est contenue dans un seul syntagme phonologique/intonatif[255].

Comme l'illustre l'exemple résumé dans le tableau IX.2, les prédictions faites par ces deux contraintes sont parfois contradictoires, ALIGN-XP favorisant la formation des syntagmes phonologiques et WRAP-XP la restreignant au maximum. Le phrasé actuel des énoncés, avec des syntagmes plus ou moins étendus, résulte de la hiérarchisation relative des contraintes propres à la langue ou au style de parole particulier[256].

[[verbe] [nom]$_{NP}$ [nom]$_{NP}$]$_{VP}$	***Align-XP,R***	***Wrap-XP***
(verbe nom)$_{SMa}$ (nom)$_{SMa}$		*
(verbe nom nom)$_{SMa}$	*	

Tableau IX.2 : Exemple d'un tableau de contraintes : deux contraintes à prédictions contradictoires

Watson et Gibson (2004) remarquent que les contraintes de la théorie de l'optimalité, quoique conçues pour évaluer la grammaticalité des formes prosodiques, peuvent également être interprétées comme des contraintes sur la production. Pour les deux contraintes que nous venons de présenter, les auteurs constatent que leurs prédictions sont similaires à celles qui résultent de l'application de l'algorithme proposé par Cooper et Paccia-Cooper (1980) pour la prédiction de la distribution des pauses.

La contrainte de la Catégorie Lexicale est introduite (Truckenbrodt 1999) pour refléter le postulat selon lequel seules les catégories lexicales et leurs projections (et non pas les catégories fonctionnelles) importent dans la gestion de l'interface entre la prosodie et la syntaxe.

[255] Comme nous l'avons précisé dans la présentation de différents constituants de la hiérarchie prosodique, il est difficile d'établir des équivalences entre les structures défendues par différents auteurs. Cependant, les syntagmes majeurs de Selkirk correspondent aux syntagmes intermédiaires de Pierrehumbert et Beckman (1986) et aux syntagmes phonologiques / intonatifs de Nespor et Vogel (1986).

[256] Se rapporter à l'étude de Durand & Lyche (2001) dans laquelle on analyse la liaison en français et sa fréquence selon la situation de conversation/style de parole dans le cadre de la théorie de l'optimalité.

Étant donné un syntagme verbal de la structure [V NP PP]$_{VP}$ la contrainte Align-XP en propose la structuration suivante

(3) V NP) PP) ;

or, une deuxième structuration est possible, avec un syntagme phonologique supplémentaire, non requise par la contrainte d'alignement[257] :

(4) V) NP) PP).

Pour proscrire le phrasé (4), on introduit la contrainte *P-syntagme qui a pour effet d'interdire toute formation de syntagmes phonologiques. En même temps, la contrainte de *P-syntagme permet de désactiver la contrainte d'alignement en plaçant cette contrainte plus bas dans la hiérarchie générale.

Nous qualifions les contraintes que nous venons de présenter comme dépendantes de la théorie (*theory dependent*), car à notre avis, elles représentent la reformulation des axiomes déjà présents en phonologie prosodique dans les termes de la théorie de l'optimalité et ne sont testées que sur les données purement phonologiques issues de la littérature et de l'introspection. Nous présentons ci-après quelques analyses issues des données observables et des expériences en production.

D'autres contraintes que l'on retrouve dans les études phonologiques récentes ont pour mission la gestion des influences pragmatiques et métriques sur le phrasé prosodique. Ainsi pour expliquer le phrasé prosodique de l'anglais, à partir des réflexions sur les réalisations probables des énoncés de son corpus, Selkirk (2000 :254) retient la hiérarchie des contraintes qui suit :

(5) MiPAccent, DEP(Accent) >> Align$_R$Focus >> Wrap-XP, Align$_R$-XP >> Bin (MaP)

Tout d'abord, dans cette hiérarchie, les contraintes d'alignement (Wrap-XP, Align$_R$-XP) ne sont pas hiérarchisées l'une par rapport à l'autre. Ensuite, on y voit apparaître la contrainte d'alignement de la frontière après le constituant en focus, Align$_R$Focus, à condition que le matériau phonétique qui suit, contienne des accents mélodiques. Deux autres contraintes (MiPAccent et DEP(Accent), contraintes supérieures dans la hiérarchie) reflètent également les interactions possibles entre le phrasé et la distribution des accents mélodiques. Ainsi, la contrainte MiPAccent impose la présence d'au moins un accent mélodique dans le syntagme mineur (syntagme accentuel) ; son application induit la proscription d'une frontière de syntagme majeur après le mot en focus, si dans la suite de l'énoncé, il n'y a aucun accent mélodique. La contrainte DEP appartient au groupe des contraintes de fidélité qui exigent que la représentation de surface soit au maximum conforme à la représentation sous-jacente. De ce fait, la contrainte DEP(Accent) impose cette équivalence pour les distributions des proéminences mélodiques dans les deux représentations. Ces deux contraintes ne sont pas non plus hiérarchisées l'une par rapport à l'autre. Ensuite la contrainte Bin(MaP) appartient au groupe des contraintes de la taille qui déterminent combien de constituants du niveau immédiatement inférieur doivent

[257] Remarquons, qu'une prononciation dotée du phrasé B est probable dans un style soutenu, dans le cadre rhétorique, par exemple.

obligatoirement être regroupés dans un constituant du niveau supérieur, conformément aux conditions de bonne formation (deux, en l'occurrence : un syntagme majeur consiste en deux syntagmes mineurs/syntagmes accentuels).

On retrouve les contraintes de taille maximale et minimale dans l'étude de Delais-Roussarie (1996) sur le phrasé prosodique du français : dans son étude en production à partir d'un corpus construit lu, l'auteur conclut que la seule contrainte d'alignement n'est pas suffisante pour rendre compte des occurrences observées, et que les contraintes phonologiques sont nécessaires ($Min_{AP,}$ $Max_{AP)}$ qui déterminent qu'un syntagme phonologique comporte de 3 (au minimum) à 6 (au maximum) syllabes. Comme en français, le syntagme phonologique est le domaine de distribution des accents mélodiques, ces contraintes sont qualifiées de métriques et s'inscrivent dans une approche générale qui considère que la structure prosodique initiale fondée sur l'analyse syntaxique est corrigée par des considérations rythmiques, plus universelles (cf. Di Cristo 1999).

Dans son étude Post (2000) cherche à établir la hiérarchie des contraintes qui gouverne la distribution de proéminences prosodiques en français, vu que c'est le syntagme phonologique qui est le domaine de l'accentuation dans cette langue. Les contraintes sont élaborées pour rendre compte des résultats de l'analyse perceptive du corpus de recherche (lecture de phrases préfabriquées). La hiérarchie finale comporte dix contraintes, rendant compte à la fois de l'accentuation finale et initiale des mots polysyllabiques[258].

Frazer *et al.* (2004) se penchent sur la question des contraintes qui déterminent le choix du phrasé prosodique dans une étude psycholinguistique : les contraintes syntaxiques y sont contrastées avec des contraintes sémantiques (*Sense Unit Condition* de Selkirk), dans les stimuli sélectionnés les prédictions des deux contraintes résultant en une structure prosodique divergente. Remarquons pourtant que les prédictions des contraintes sont testées avec des énoncés-stimuli construits, d'une structure syntaxique fixe (4 structures syntaxiques sont prises en compte dans l'étude analysée : subordonnées relatives extraposées, particules détachées, syntagme nominal (sujet) développé, subordonnées complémentaires). Nous voyons un autre biais méthodologique dans la procédure de la production des stimuli : c'est un phonologue expert qui enregistre les stimuli avec deux versions du phrasé, en respectant les prédictions de deux contraintes, bien que cela puisse donner lieu à des productions non-naturelles. Le test de jugements d'acceptabilité est proposé aux auditeurs sous deux formes : a) évaluation sur l'échelle de 5 et b) jugements comparatifs des deux versions. Les auditeurs ont pour instruction de ne prendre en compte que les propriétés mélodiques (tonales) et rythmiques des stimuli et reçoivent la consigne de ne pas prêter attention au contenu sémantique des phrases[259]. Les résultats démontrent que les stimuli qui respectent la condition de la cohérence sémantique reçoivent des scores de préférence plus élevés, et ceci dans les deux tâches de l'étude d'acceptabilité. Or, dans le contexte des jugements relatifs, les auditeurs ayant signalé leur préférence, indiquaient que la deuxième version du stimulus était tout de même acceptable (les pourcentages d'acceptabilité de la version « rejetée » allant de 47 à 78 %).

[258] Précisons que cette étude constate le désaccord entre les prédictions théoriques sur la formation des syntagmes phonologiques et leur restructuration et les données observables.
[259] Pour reprendre la citation de l'article « *the subjects […] were told not to rate how much sense the sentence makes* ».

Avant de procéder à la présentation des résultats obtenus, nous devons préciser que nous n'avons pas pu tester la totalité des contraintes recensées lors de l'analyse des différentes recherches. En particulier, nous ne pouvons pas tester la contrainte de la cohérence sémantique et ceci pour les raisons qui vont être développées dans ce qui suit.

La contrainte de la cohérence sémantique défend la possibilité d'insérer des frontières prosodiques au sein des constituants syntaxiques à condition que les parties résultantes aient un sens, désignent un concept uni. Si nous prenons un exemple abstrait de trois constituants syntaxiques X, Y et Z, parmi lesquels X et Y forment un tout sémantique, mais non pas Y et Z, les options suivantes du phrasé sont possibles : a) tout constituant syntaxique forme un syntagme prosodique à part ; b) X et Y forment un constituant prosodique et Z – un deuxième. Par contre, cette approche sémantique proscrit de former un constituant prosodique regroupant Y et Z sans X. Selkirk (1984) en donne la définition suivante :

The Sense Unit Condition on Intonational Phrasing :

> The immediate constituents of an intonational phrase must together form a sense unit.
>
> Two constituents, C_i et C_j, form a sense unit if (a) or (b) is true of the semantic interpretation of the sentence:
>
> a) C_i modifies C_j (a head)
>
> b) C_i is an argument of C_j (a head).

Nous considérons que cette contrainte n'est pas très explicite et que son application aux données induira une grande partie d'interprétations plus ou moins arbitraires de notre part. Nous aurions pu y appliquer la grammaire catégorielle combinatoire de Steedman (2000) afin d'aboutir à une formalisation des rapports sémantiques entre les constituants syntaxiques. Nous avons pourtant choisi de ne pas avancer dans cette direction[260] et de nous limiter à l'étude des contraintes d'alignement ALIGN-XP,R et WRAP-XP. Ce choix nous a paru d'autant plus légitime que la théorie syntaxique X-barre[261] prend en considération certains rapports sémantiques. Ainsi la théorie X-barre attribue-t-elle une structure commune aux syntagmes syntaxiques qui sont organisés autour d'une tête lexicale, qui est précédée par un spécifieur (obligatoire ou facultatif selon la catégorie) et optionnellement suivie d'un complément (cf. figure 9.1).

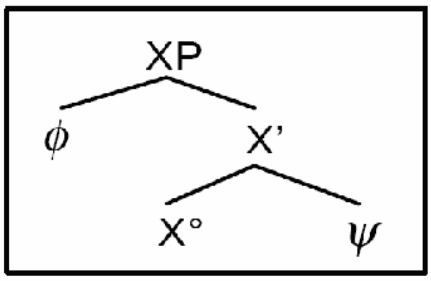


Figure 9.1 : Une représentation générale de la syntaxe X-barre (X° : projection minimale (tête lexicale), X' : projection intermédaire, ψ : complément de X, ϕ :spécifieur de X)

[260] Nous sommes persuadée que toute annotation nécessite une expertise de plusieurs spécialistes.
[261] Quoique certains syntacticiens lui refusent ce statut et la relèguent au domaine de la sémantique.

Étant donné que la mise en correspondance des structures syntaxiques et prosodiques est, d'une part, le fondement des réflexions de la phonologie prosodique et, d'autre part, largement exploitée dans les applications technologiques, nous avons dans un premier temps analysé la performance des auditeurs par rapport à la structure syntaxique : nous avons testé dans quelle mesure les contraintes « ancrées syntaxe » prédisent la performance des sujets.

9.2. MÉTHODOLOGIE

Nous avons analysé les prédictions des contraintes ALIGN-XP,R et WRAP-XP. La question peut se poser de savoir pourquoi, ayant proposé aux auditeurs la tâche d'annotation des frontières des unités intonatives, nous lui confrontons des prédictions de contraintes proposées pour le niveau des syntagmes majeurs de Selkirk, correspondant au niveau des syntagmes phonologiques de Nespor et Vogel (1986). Premièrement, dans la hiérarchie des constituants de Pierrehumbert et Beckman (1986) fondée sur l'analyse de la structure de la courbe intonative c'est le syntagme intermédiaire qui paraît être un homologue du syntagme majeur de Selkirk. Dans un deuxième temps, les frontières que les auditeurs repéraient dans l'expérience, correspondent aux niveaux 3 et 4 dans le schème d'indices de frontières du système d'annotation ToBI. Comme nous l'avons signalé aucun système de constituants prosodiques n'est développé pour la langue russe : la notion du syntagme est opérationnelle dans les travaux sur l'intonation, mais s'agit-il d'un constituant similaire à un syntagme phonologique ou à un syntagme intonatif ou une unité mixte ? De même, dans le modèle de Levelt (1989) pour la production de la parole spontanée, chaque frontière du syntagme phonologique représente une option pour l'insertion d'une frontière prosodique. Le choix de retenir cette option dépend de la disponibilité des représentations de surface du matériau verbal qui suit[262]. Étant donné les particularités du processus de planification du message en style spontané, les frontières prosodiques sont plus fréquentes dans ce style de parole (frontières avec marqueurs prosodiques). La même réflexion émane des contraintes imposées par l'hypothèse de l'étagement strict.

Afin de procéder aux analyses des relations entre la structure prosodique perçue et la structure syntaxique, il est nécessaire de disposer d'une représentation syntaxique de l'énoncé selon la syntaxe X-barre (Jackendoff, 1977). De manière générale, on suppose donc que les énoncés de la langue sont structurés en termes de constituants qui correspondent à des catégories syntaxiques, la structure résultante étant de nature hiérarchique et représentable sous la forme d'un arbre (appelé *phrase-maker*). Les nœuds de l'arbre syntaxique correspondent aux catégories syntaxiques (nœuds non-terminaux) et aux lexèmes (nœuds terminaux, qui peuvent être vides). Certes, les formalismes syntaxiques modernes (la grammaire SPSG, par exemple) sont plus puissants, or, notre

[262] Dans le modèle de Levelt (1989), la procédure décrite rend compte des pauses générées par le Compilateur Prosodique. Mais ce n'est pas la seule source des pauses réalisées ; d'autres procédures peuvent également en être la cause : telle la recherche d'un mot de basse fréquence ou d'un groupe de mots avec de faibles probabilités d'enchaînement ; à l'étape d'implémentation phonétique, les pauses résultent d'un délai dans l'accès à la forme du mot, ou lors de l'articulation (longue phase préparatoire) ; ensuite, les pauses peuvent avoir des raisons stylistiques (pauses rhétoriques). Ces pauses-là n'ont aucun rapport avec la structure prosodique et linguistique du message.

étude étant de type exploratoire, nous avons décidé de ne pas dévier de l'approche traditionnelle en phonologie prosodique.

Comme nous l'avons précisé ci-dessus, la théorie X-barre prescrit à tout syntagme une structure ternaire faisant une distinction entre le spécifieur (modificateur) et le complément. De nombreuses théories linguistiques évitant les structures ternaires, dans la théorie X-barre on introduit un niveau intermédiaire X' (appelé la projection intermédiaire) entre la tête lexicale et la projection maximale, ceci permettant de ne pas mettre au même niveau les spécificateurs et les compléments (cf. figure 9.1). La figure 9.2 présente une analyse en termes de syntaxe X-barre d'une phrase-stimulus

(6) « А раньше ты на метро ездила? »

et avant tu avec le métro venais (Et avant tu prenais le métro ?)

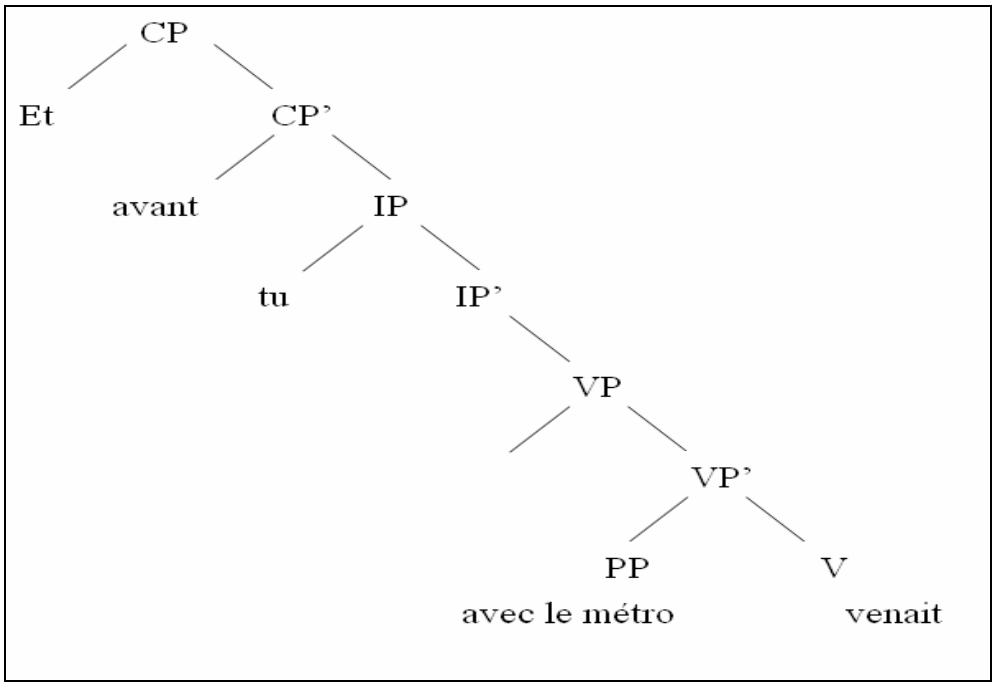


Figure 9.2 : Exemple de la représentation selon la syntaxe X-barre de la structure syntaxique de l'énoncé (7) (I : catégorie fonctionnelle 'inflexion' ; C : catégorie fonctionnelle 'complémenteur')

La difficulté majeure qu'il nous a fallu surmonter réside dans l'absence d'un cadre général et rigoureux pour l'analyse syntaxique des séquences de mots au-delà du niveau d'une phrase grammaticale, unité de base en syntaxe (définie par la présence du signe majeur de ponctuation). Or, nous travaillons sur de l'oral, et la ponctuation, si elle est présente en transcription, correspond à une interprétation imposée par le linguiste[263]. Sous cette limitation épistémologique, nous avons proposé des représentations arborescentes pour chaque clause, pour ensuite les joindre dans un seul arbre syntaxique, et ceci pour des clauses sans marquage syntaxique des relations de subordination ou de coordination, constructions

[263] Bien sûr, il aurait pu être possible de demander à un autre groupe d'auditeurs de ponctuer le texte des stimuli. Cela aurait pu constituer une autre expérience fort intéressante, avec l'attention portée aux cas de l'accord parfait et de désaccord. Nous avons, pourtant, jugé cette entreprise comme étant hors du contexte de la présente recherche.

fréquentes en russe. Nous avons considéré alors toute frontière de clause grammaticale comme un site potentiel de frontière (cf. cependant Truckenbrodt (2005) pour une analyse de la structure clausale en allemand).

A l'étape suivante, à partir de la structure syntaxique récupérée, il a fallu déterminer les positions pour lesquelles la présence des frontières prosodiques était prédite par des contraintes ALIGN-XP,R et WRAP-XP. De manière consensuelle, il est admis que les contraintes sur l'alignement entre les constituants prosodiques et syntaxiques ne s'appliquent qu'aux projections des items lexicaux (NP, VP, AP, PP, AdvP) et non pas aux éléments fonctionnels (CP, IP, DP) ou aux éléments vides et à leurs projections. Les frontières terminales des stimuli ont été exclues de l'analyse.

Ensuite, chaque frontière signalée par les auditeurs a été attribuée à une des catégories « prédite par la contrainte Wrap-XP », « prédite par la contrainte Align-XP » (ces deux catégories ne s'excluant pas mutuellement), « frontière non-prédite » et « surprédiction »[264]. Conjointement, nous avons récupéré la force de toute frontière, exprimée en fonction du nombre de juges qui l'ont indiquée. Nous avons pourtant fait une présupposition : comme nous l'avons déjà mentionné l'accent en russe est mobile et peut tomber sur n'importe quelle syllabe du mot (avec la préférence pour une accentuation centrale). Il faut ajouter que la segmentation en mots des énoncés-stimuli n'était pas disponible aux auditeurs dans toutes les conditions expérimentales. Par conséquent, nous avons fait se fusionner les frontières qui différaient de la frontière réelle du mot de plus ou moins d'une syllabe *inaccentuée*. Donc, pour la catégorie des frontières prédites, nous avons également récupéré les informations sur le décalage éventuel de la frontière signalée par rapport à la frontière réelle du mot. Ce décalage pouvait être positif comme négatif : un décalage négatif est observé dans les cas où la frontière est indiquée une syllabe avant la frontière potentielle ; un décalage positif, à l'inverse, correspond aux cas où une syllabe inaccentuée est rajoutée.

Au final, nous avons obtenu six tableaux, séparément pour chaque condition et pour la langue maternelle des auditeurs. Chaque tableau contient les informations suivantes : la première colonne correspond à la force de la frontière en question, dans les deuxième et troisième colonnes nous indiquons pour combien d'auditeurs nous avons observé un décalage positif ou négatif, ensuite, les colonnes de quatre à six, spécifient si une frontière est prédite par les contraintes WRAP-XP, ALIGN-XP,R, ou s'il s'agit d'une frontière non-prédite en théorie mais indiquée par les auditeurs. Au total, les auditeurs russophones ont marqué 117 frontières différentes dans la première condition expérimentale, 107 – dans la deuxième et 81 dans la troisième ; pour les auditeurs francophones nous avons obtenu respectivement 111, 109 et de nouveau 111 frontières.

[264] Il s'agit des frontières prédites en théorie, qui ne sont tout de même pas annotées par les auditeurs dans l'expérience.

9.3. INTERPRÉTATION DES RÉSULTATS

9.3.1. Contraintes d'alignement émanant de l'interface « prosodie ~ syntaxe »

Une fois toutes les frontières associées avec une catégorie, nous pouvons observer quelles tendances ressortent des performances des 14 auditeurs dans notre expérience.

Sur la figure 9.3 nous présentons les résultats illustrant le rôle des deux contraintes ANCRÉES syntaxe dans la prédiction du phrasé prosodique perçu.

Tout d'abord, nous voyons que le rôle des contraintes prosodiques diffère en fonction de la langue maternelle des auditeurs et des conditions expérimentales : les résultats des auditeurs francophones sont caractérisés par une homogénéité entre les conditions expérimentales, tandis que la structure syntaxique joue un rôle primordial dans les réponses des auditeurs russes sous présentation normale des stimuli, leur performance étant significativement différente des schèmes de réponses observés dans les deux autres conditions (ces observations sont confirmées par des tests chi²).

Dans un deuxième temps, si nous nous concentrons sur le rôle de la prosodie dans le signalement de la structure syntaxique, les contraintes WRAP-XP et ALIGN-XP,R ne prédisent, respectivement, que 40% et 50% de frontières perçues par les auditeurs : la conclusion portant sur le rôle moindre de la structure syntaxique dans la prédiction de la perception des ruptures prosodiques, nous paraît justifiée.

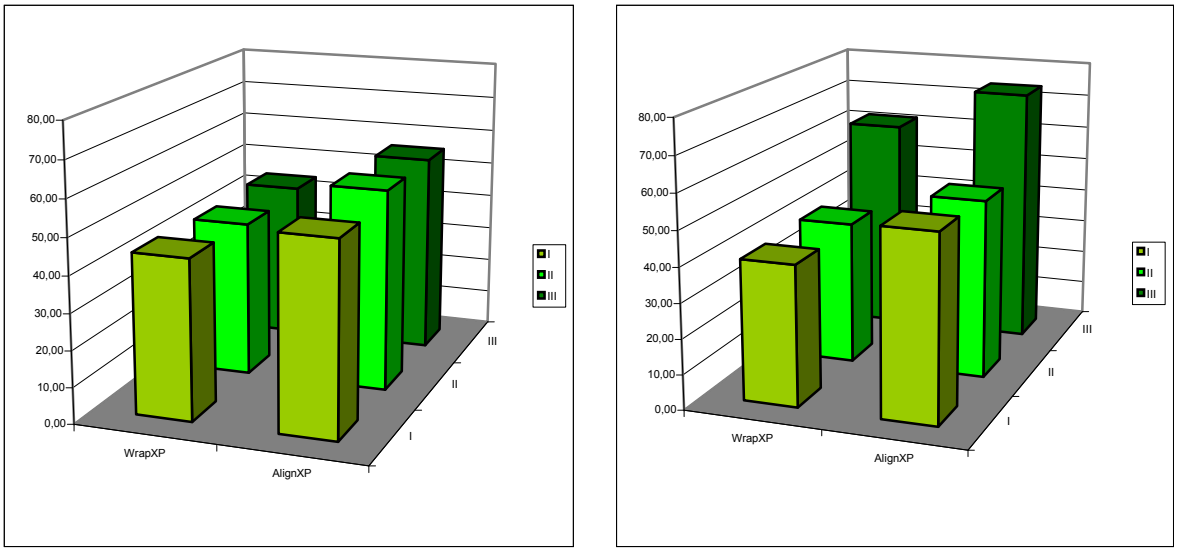


Figure 9.3 : Le rôle des contraintes ancrées syntaxe dans la prédiction du phrasé prosodique perçu selon la condition expérimentale (auditeurs francophones : panel gauche, auditeurs russes : panel droit).

Nous avons repris ensuite la distinction faite dans l'analyse de la localisation des frontières prosodiques entre les frontières consensuelles, annotées par plus de 4 auditeurs, et les frontières optionnelles. La figure 9.4 résume les observations pour la catégorie des frontières consensuelles. Nous constatons une augmentation de la force prédictive des contraintes ancrées syntaxe : le pourcentage de prédictions correctes se situe entre 50 et 65 %.

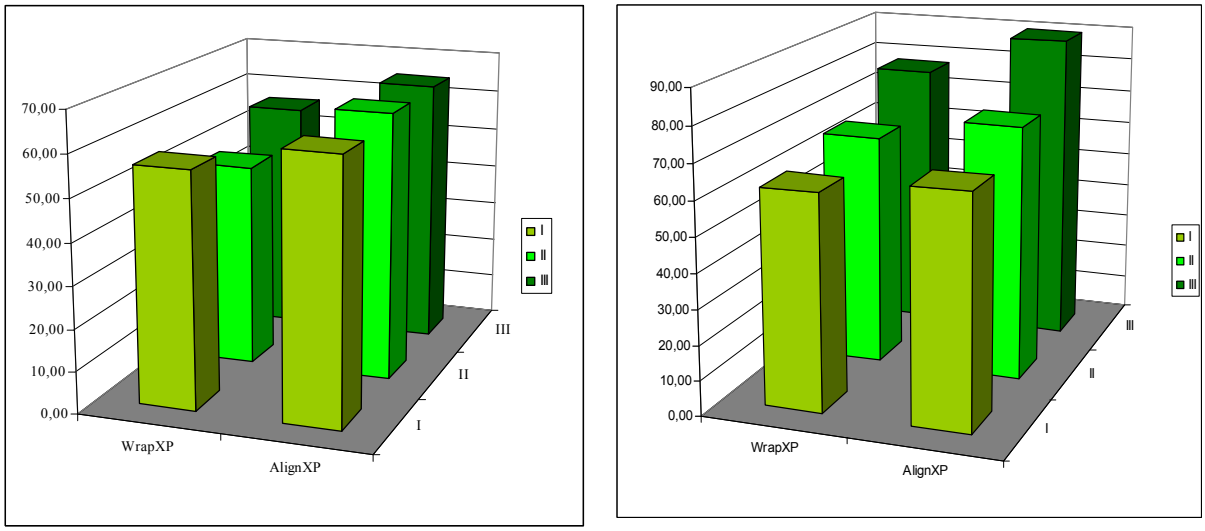


Figure 9.4 : Le rôle des contraintes ancrées syntaxe pour les frontières consensuelles selon la condition expérimentale (auditeurs francophones : panel gauche, auditeurs russes : panel droit).

Dans les analyses qui suivent, nous présentons d'abord séparément les résultats des auditeurs russes et francophones, et nous procédons ensuite à une comparaison des performances des deux groupes.

9.1.1.1. Auditeurs russophones

L'évaluation de la qualité de la démarche prédictive fondée sur l'interface « prosodie ~ syntaxe » est quantifiée à l'aide des mesures de « rappel », « précision », « silence », « bruit » et « F-mesure », traditionnellement employées en recherche documentaire (cf. Van Rijsbergen 1979) : il s'agit donc d'une analyse quantitative des erreurs de prédiction (tableaux IX.3 et IX.4). Le rappel quantifie le rapport des frontières correctement prédites par le modèle « syntaxique » sur le nombre total des frontières signalées par les auditeurs : un taux de rappel de 100 % signifierait que toutes les frontières, marquées par les auditeurs lors du test, ont été prédites par cette contrainte. La mesure complémentaire du rappel est le silence : c'est-à-dire, le silence mesure la proportion des frontières signalées par les auditeurs, non prédites par des contraintes d'origine syntaxique. D'autre part, la précision mesure le rapport des frontières correctement prédites sur le total des frontières prédites : un taux de précision de 0 % correspondrait à une prédiction complètement erronée du phrasé perçu. La mesure complémentaire, le bruit, est le rapport du nombre de frontières prédites à tort sur l'ensemble des frontières prédites.

Dans notre cas, le taux de rappel varie entre 40,17 et 61,73 % pour la contrainte WRAPXP et de 52,99 et 74,07 % pour la contrainte d'ALIGNXP,R. Pour la condition CondD, dans laquelle les informations prosodiques ressortent au premier plan, le rappel est de 41,12 et 51,4 % : cela signifie que les contraintes ancrées syntaxe prédisent correctement environ la moitié des frontières prosodiques signalées par les auditeurs.

CONDITION	CONTRAINTE	FRONTIÈRES			ÉVALUATION				
		prédites	*non-prédites*	*surprédiction*	*silence*	*bruit*	*rappel*	*précision*	*F-measure*
CondR	WrapXP	47	70	6	59,83	11,32	40,17	88,68	55,29
	AlignXP	62	55	6	47,01	8,82	52,99	91,18	67,03
CondD	WrapXP	44	63	8	58,88	15,38	41,12	84,62	55,35
	AlignXP	55	52	12	48,60	17,91	51,40	82,09	63,22
CondN	WrapXP	50	31	1	38,27	1,96	61,73	98,04	75,76
	AlignXP	60	21	5	25,93	7,69	74,07	92,31	82,19

Tableau IX.3 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe. Auditeurs russes

D'autre part, le système fondé sur des contraintes ancrées syntaxe bénéficie d'une précision très élevée, de l'ordre de 90 % (82,09 – 98,04% selon la contrainte et la condition expérimentale) : ce résultat indique qu'une majorité absolue des frontières prédites sont en effet attestées par les auditeurs.

La performance d'un système est optimale lorsque les mesures de rappel et de précision manifestent les valeurs les plus élevées ; cette prise en compte simultanée est reflétée par la F-measure qui correspond à la moyenne harmonique des deux. Un modèle du phrasé prosodique fondé sur des contraintes émanant de l'interface « prosodie ~ syntaxe » bénéficie d'une F-measure variant de 55 à 75% pour la contrainte de WRAPXP et de 63 à 82% pour la contrainte ALIGNXP,R (remarquons également que par opposition au WRAPXP, la valeur minimale de F-measure est observée dans la deuxième condition expérimentale). Il ne s'agit donc pas d'un système optimal, cette mesure démontre la qualité de la démarche prédictive : remarquons pourtant que cette valeur élevée est due à la mesure de précision. Il n'est pas étonnant que la valeur maximale de F-measure soit atteinte dans la troisième condition où les auditeurs ont eu accès à la totalité des informations linguistiques : nous considérons qu'étant donné la tradition des recherches sur l'intonation du russe, les informations syntaxiques jouaient un rôle primordial dans les réponses des auditeurs-experts.

Qui plus est, pour les frontières consensuelles nous constatons l'augmentation du rappel, et par conséquent, de F-measure du système (cf. Tableau VIII.5).

CONDITION	CONTRAINTE	FRONTIÈRES			ÉVALUATION				
		prédites	*non-prédites*	*surprédiction*	*silence*	*bruit*	*rappel*	*précision*	*F-measure*
CondR	WrapXP	29	18	6	38,30	17,14	61,70	82,86	70,73
	AlignXP	31	16	6	34,04	16,22	65,96	83,78	73,81
CondD	WrapXP	30	15	8	33,33	21,05	66,67	78,95	72,29
	AlignXP	33	12	12	26,67	26,67	73,33	73,33	73,33
CondN	WrapXP	38	11	1	22,45	2,56	77,55	97,44	86,36
	AlignXP	44	5	5	10,20	10,20	89,80	89,80	89,80

Tableau IX.4 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe : frontières consensuelles. Auditeurs russes

9.1.1.2. Auditeurs francophones

En appliquant les mêmes démarches analytiques aux données recueillies auprès des auditeurs francophones, nous pouvons faire les remarques suivantes (cf. tableaux IX.5 et IX.6). Nous voyons que les valeurs du rappel, de la précision et de F-measure sont similaires à celles que nous avons observées pour les auditeurs russes, une différence drastique n'apparaissant que pour la troisième condition expérimentale. Si pour les sujets russes cette condition avait un statut particulier, cela n'est pas le cas de la performance des auditeurs francophones.

CONDITION	CONTRAINTE	FRONTIÈRES			ÉVALUATION				
		prédites	*non-prédites*	*surprédiction*	*silence*	*bruit*	*rappel*	*précision*	*F-measure*
CondR	WrapXP	49	62	5	55,86	9,26	44,14	90,74	59,39
	AlignXP	59	52	10	46,85	14,49	53,15	85,51	65,56
CondD	WrapXP	47	62	4	56,88	7,84	43,12	92,16	58,75
	AlignXP	61	48	5	44,04	7,58	55,96	92,42	69,71
CondN	WrapXP	49	62	4	55,86	7,55	44,14	92,45	59,76
	AlignXP	64	47	4	42,34	5,88	57,66	94,12	71,51

Tableau IX.5 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe. Auditeurs francophones

CONDITION	CONTRAINTE	FRONTIÈRES			ÉVALUATION				
		prédites	*non-prédites*	*surprédiction*	*silence*	*bruit*	*rappel*	*précision*	*F-measure*
I	WrapXP	27	21	5	43,75	15,63	56,25	84,38	67,50
	AlignXP	30	18	10	37,50	25,00	62,50	75,00	68,18
II	WrapXP	22	23	4	51,11	15,38	48,89	84,62	61,97
	AlignXP	29	16	5	35,56	14,71	64,44	85,29	73,42
III	WrapXP	27	21	4	43,75	12,90	56,25	87,10	68,35
	AlignXP	31	17	4	35,42	11,43	64,58	88,57	74,70

Tableau IX.6 : Évaluation quantitative de la dimension prédictive associée avec des contraintes ancrées syntaxe : frontières consensuelles. Auditeurs francophones

Le tout premier aperçu des données nous permet de conclure que le rôle des contraintes syntaxiques diffère selon la condition expérimentale et selon la langue maternelle des auditeurs. Cependant, nous constatons un taux d'environ 60% de prédictions issues de l'application des contraintes syntaxiques dans des conditions désémantisées ou dans la performance d'auditeurs francophones : cela signifie que ces structures syntaxiques reçoivent un marquage à l'aide d'indices acoustiques. Ces résultats nous poussent à nous interroger sur le principe de substitution (de parallélisme) de Pechkovskij (1928, 1959 :181) :

> *« Чем яснее выражено какое-либо значение чисто грамматическими средствами, тем слабее может быть его интонационное выражение (вплоть до полного исчезновения), и наоборот, чем сильнее интонационное выражение, тем слабее может быть грамматическое. »*

Plus le marquage grammatical d'une catégorie est explicite, plus son marquage intonatif est faible (jusqu'à l'absence complète du marquage) ; inversement, plus le marquage intonatif est fort, plus le marquage grammatical peut être faible.

Cette problématique sera au cœur de notre discussion du bilan et des résultats de l'expérience (section 9.4).

9.1.1.3. Analyse comparative

Après cette vue d'ensemble de résultats, nous allons observer plus en détail les influences des facteurs Contrainte, Condition et Langue sur la performance des sujets. Pour tester le rôle de chaque facteur, nous calculons son coefficient chi² sur les effectifs. Ce coefficient permet de déterminer si l'effet d'un facteur donné est dû au hasard ou s'il est réellement pertinent. Nous opposons donc les conditions expérimentales deux à deux ainsi que les réponses des sujets russes et francophones pour chaque question. Nous avons ensuite créé des matrices d'effectifs pour chaque facteur avec des valeurs pour deux catégories « frontières prédites » versus « frontières non-prédites » (cf. ci-dessous). Les matrices d'effectifs permettent le calcul du coefficient chi² pour déterminer si l'effet du facteur est significatif ou non. La totalité des matrices et les calculs du coéfficient chi² sont présentés dans l'annexe 6.

	ALIGNXP,R	WRAPXP
frontières prédites	59	49
frontières non-prédites	52	62
TOTAL	111	111

Nous présentons les résultats de cette évaluation dans les tableaux IX.7-IX.10.

Dans un premier temps, nous cherchons à démontrer si le rôle des deux contraintes est similaire selon la condition expérimentale. Pour les auditeurs russes, ce test statistique permet de mettre en évidence le statut particulier de la troisième condition expérimentale, fait que nous avons déjà évoqué auparavant. Nous constatons également l'absence de différence significative dans les prédictions des contraintes ancrées syntaxe de la performance des auditeurs sous les première et deuxième conditions expérimentales. Pour les sujets francophones, les prédictions des contraintes ALIGNXP,R et WRAPXP sont similaires sous les trois conditions.

CONTRAINTE	OPPOSITION ANALYSÉE	AUDITEURS RUSSES			AUDITEURS FRANCOPHONES		
		Nombre (pourcentage)	*Chi²*	*p*	*Nombre (pourcentage)*	*Chi²*	*p*
WRAPXP	CondR *vs.* CondD	47 (40,17%) 44 (41,12%)	1e-04	0.993	49 (44,14%) 47 (43,12%)	3e-04	0.986
	CondD *vs.* CondN	44 (41,12%) 50 (61,73%)	7.028	0.008	47 (43,12%) 49 (44,14%)	3e-04	0.986
	CondR *vs.* CondN	47 (40,17%) 50 (61,73%)	8.0592	0.005	49 (44,14%) 49 (44,14%)	-	-
ALIGNXP,R	CondR *vs.* CondD	62 (52,99%) 55 (51,4%)	0.0108	0.917	59 (53,15%) 61 (55,96%)	0.0802	0.777
	CondD *vs.* CondN	55 (51,4%) 60 (74,07%)	9.0452	0.003	61 (55,96%) 64 (57,66%)	0.0138	0.906
	CondR *vs.* CondN	62 (52,99%) 60 (74,07%)	8.1259	0.004	59 (53,15%) 64 (57,15%)	0.2917	0.589

Tableau IX.7 : Évaluation du rôle des contraintes ancrées syntaxe selon la condition expérimentale

Dans un deuxième temps, nous avons voulu évaluer les interactions entre la langue maternelle des auditeurs et les contraintes ancrées syntaxe. Les données du tableau VIII.9 permettent de conclure que le facteur de la langue maternelle n'a pas joué de rôle important, sauf pour la troisième condition expérimentale.

CONDITION	OPPOSITION ANALYSÉE	CONTRAINTE WRAPXP			CONTRAINTE ALIGNXP,R		
		Nombre (pourcentage)	*Chi²*	*P*	*Nombre (pourcentage)*	*Chi²*	*P*
CondR	Français *vs.* Russe	49 (44,14%) 47 (40,17%)	0.2239	0.636	59 (53,15%) 62 (52,99%)	0.012	0.914
CondD	Français *vs.* Russe	47 (43,12%) 44 (41,12%)	0.0254	0.873	61 (55,96%) 55 (51,4%)	0.287	0.592
CondN	Français *vs.* Russe	49 (44,14%) 50 (61,73%)	5.1148	0.024	64 (57,15%) 60 (74,07%)	4.823	0.028

Tableau IX.8 : Évaluation du rôle des contraintes ancrées syntaxes dans une perspective translinguistique

Nous avons répliqué les mêmes analyses pour des frontières consensuelles (tableaux IX.9 et IX.10). Pour les auditeurs russes, nous constatons alors l'absence de l'effet du facteur

Condition pour la contrainte Wrap-XP (valeurs de p élevées), pourtant l'effet persiste pour la contrainte AlignXP,R si nous comparons les première et troisième conditions expérimentales. Nous observons par contre des valeurs de p très élevées pour les données des auditeurs français : dans la performance de ces sujets dans la tâche d'annotation du phrasé prosodique, le facteur Condition n'exerce aucune influence sur le potentiel prédictif des contraintes syntaxiques.

CONTRAINTE	OPPOSITION ANALYSÉE	AUDITEURS RUSSES			AUDITEURS FRANCOPHONES		
		Nombre (pourcentage)	*Chi²*	*p*	*Nombre (pourcentage)*	*Chi²*	*p*
WRAPXP	CondR *vs.* CondD	29 (61,7%) 30 (66,67%)	0,078	0,78	27 (56,25%) 22 (48,89%)	0,253	0,615
	CondD *vs.* CondN	30 (66,67%) 38 (77,55%)	0,898	0,343	22 (48,89%) 27 (56,25%)	0,253	0,615
	CondR *vs.* CondN	29 (61,7%) 38 (77,55%)	2,16	0,142	27 (56,25%) 27 (56,25%)	-	-
ALIGNXP,R	CondR *vs.* CondD	31 (65,96%) 33 (73,33%)	0,294	0,588	30 (62,5%) 29 (64,44%)	0,0001	0,98
	CondD *vs.* CondN	33 (65,96%) 44 (89,8%)	3,292	0,071	29 (64,44%) 31 (64,58%)	0,04	0,839
	CondR *vs.* CondN	31 (65,96%) 44 (89,8%)	6,643	0,01	30 (62,5%) 31 (64,58%)	0	1

Tableau IX.9 Évaluation du rôle des contraintes ancrées syntaxe selon la condition expérimentale (frontières consensuelles)

Dans une perspective translinguistique, les résultats issus des analyses effectuées pour des frontières consensuelles sont comparables à ceux que nous avons constatés pour la totalité des données, à savoir : l'identité des tendances pour des conditions désémantisées et le statut particulier de la condition avec la présentation non-manipulée des stimuli pour les auditeurs russes (mais uniquement pour la contrainte AlignXP,R, pour la contrainte WrapXP il ne s'agit que d'une tendance marginale).

CONDITION	OPPOSITION ANALYSÉE	CONTRAINTE WRAPXP			CONTRAINTE ALIGNXP,R		
		Nombre (pourcentage)	*Chi²*	*P*	*Nombre (pourcentage)*	*Chi²*	*P*
I	Français *vs.* Russe	27 (56,25%) 29 (61,7%)	0,11	0,74	30 (62,5%) 31 (65,96%)	0,019	0,89
II	Français *vs.* Russe	22 (48,89%) 30 (66,67%)	2,232	0,135	29 (64,44%) 33 (73,33%)	0,467	0,495
III	Français *vs.* Russe	27 (56,25%) 38 (77,55%)	4,06	0,043	31 (64,58%) 44 (89,5%)	7,41	0,006

Tableau IX.10 : Évaluation du rôle des contraintes ancrées syntaxe dans une perspective translinguistique (frontières consensuelles)

Enfin, nous voudrions savoir si la force prédictive des contraintes WrapXP et AlignXP,R est la même pour des frontières optionnelles *versus* consensuelles. Nous avons toujours appliqué le test chi². Les résultats sont présentés dans le tableau IX.11.

LANGUE	CONTRAINTE	CONDITION	PROPORTION « *frontières optionnelles prédites* »	PROPORTION « *frontières consensuelles prédites* »	χ^2	VALEUR p
Russe	WRAPXP	I	0,257	0,617	13,69	0,0002
		II	0,226	0,667	19,15	1,2e-05
		III	0,375	0,776	11,50	0,0007
	ALIGNXP,R	I	0,44	0,66	4,47	0,0346
		II	0,355	0,733	13,48	0,0002
		III	0,5	0,898	13,96	0,0002
Français	WRAPXP	I	0,349	0,563	4,20	0,0405
		II	0,39	0,489	0,68	0,4102
		III	0,349	0,563	4,20	0,0405
	ALIGNXP,R	I	0,397	0,625	5,13	0,0235
		II	0,5	0,644	1,69	0,1937
		III	0,524	0,646	1,20	0,2735

Tableau IX.11 : Évaluation du rôle des contraintes ancrées syntaxe dans la prédiction des frontières consensuelles versus optionnelles

De manière générale, les contraintes ancrées syntaxe interviennent plus dans la prédiction des frontières consensuelles : les proportions des frontières consensuelles prédites sont toujours plus élevées. Les différences sont toujours statistiquement significatives pour des auditeurs russes, mais non pas pour les auditeurs français (cf. tableau IX.11). D'une part, ce résultat a des implications vis-à-vis du principe de substitution de Pechkovskij (1928) que nous avons déjà évoqué et sur lequel nous reviendrons dans la section consacrée à la discussion et conclusions. D'autre part, la majorité de frontières optionnelles sont motivées par des considérations autres que syntaxiques, telles que les contraintes de la théorie de

l'optimalité les formalisent. Ce résultat nous permet de mesurer la difficulté de la tâche de modélisation acoustique des frontières prosodiques : il existe dans la chaîne parlée des agrégations d'indices acoustiques qui induisent la perception d'une rupture prosodique, mais ce processus n'est pas de nature déterministe, mais plutôt probabiliste.

Après avoir analysé la force prédictive des contraintes syntaxiques sur le phrasé prosodique perçu, nous avons constaté une marge permettant d'améliorer la performance du modèle ancré syntaxe. En particulier, nous avons voulu analyser l'impact de la contrainte discursive, reliant la structure prosodique et l'organisation informationnelle du message.

9.4.2. Contrainte discursive AlignFocus,R

Les analyses présentées dans la section précédente ont démontré qu'il reste une place pour l'amélioration de la démarche prédictive par rapport au modèle « syntaxique » (cf. les valeurs de rappel et de F-measure).

Remarquons aussi, que les unités perçues ont une durée moindre que celle communiquée pour des syntagmes de la tradition russe. Il se peut que les unités plus petites soient signalées sous l'influence d'une autre contrainte, appartenant au niveau discursif[265] : en effet, si nous prenons l'exemple d'un syntagme verbal complexe, un de ces composants peut être mis en focus, induisant ainsi la perception d'une frontière prosodique. Ce cas de figure est géré dans le travail de Selkirk (2000) par la contrainte d'alignement du focus (AlignFocus), qui préconise une co-occurrence de la frontière prosodique avec la frontière droite du constituant en focus dans la structure informationnelle. On définit le focus comme une partie de l'énoncé (du syntagme intonatif) qui introduit dans le discours l'information nouvelle. Un constituant plus ou moins long peut être mis en focus ; on distingue également les structures avec la focalisation large *versus* étroite. Dans la lignée de l'approche « accentuelle » du focus (Gussenhoven 1984, entre autres), le constituant en focus est marqué par un accent focal (accent mélodique).

Remarquons que le phénomène de focus (connu sous le nom d'accent logique, « *логическое ударение* ») a rarement été traité dans le contexte de l'analyse formelle du système intonatif (cf. l'analyse de Nikolaeva (1982), ainsi que les critiques émises par Yokoyama 1987). Avec la contrainte AlignFocus,R ce pas est franchi, quoique pour appréhender le phénomène de focus dans toute sa complexité, des analyses plus rigoureuses et plus approfondies soient nécessaires, analyses qui dépassent le cadre de la présente recherche.

En effet, l'analyse des frontières fortes non prédites met en évidence le rôle joué par la contrainte discursive sous discussion. L'exemple présenté ci-dessous illustre ce cas de figure.

Si on examine tout d'abord l'arbre syntaxique correspondant à l'énoncé-stimulus « *Я думаю, что мне стоит еще этим позаниматься* » (traduction littérale : *Je pense que*

[265] Remarquons, entre parenthèses, que peu nombreuses sont des contraintes discursives appliquées dans la modélisation du phrasé prosodique

(ça) me vaut (la peine) encore en faire ; Je pense que ça vaut la peine que je continue les cours), on obtient le phrasé prosodique en accord avec les contraintes syntaxiques :

(7) [Je pense]$_{MaP}$ [que *(cela)* me vaut *(la peine)* encore en faire]$_{MaP}$.

Cependant, le tableau IX.12 résume les distributions de frontières et de leurs forces (en termes du nombre d'auditeurs) pour les deuxième et troisième conditions expérimentales. Remarquons, que la frontière prédite par le modèle « syntaxique » n'est signalée dans l'expérience que par une minorité d'auditeurs. Il s'agit d'une frontière entre les clauses grammaticales : nous considérons qu'une analyse plus profonde du phrasé prosodique associée avec la structure clausale est nécessaire (cf., par exemple, le travail de Truckenbrodt 2005).

	Je pense	]	que (cela) me	]	vaut (la peine)$_{FOC}$	]	encore	]	en	]	faire
Ru3		2				1		3			
Ru2				1		5		1		1	
Fr3		1				6		4		1	
Fr2		1				6		4			

Tableau IX.12 : Perception du phrasé prosodique : distribution des frontières

La figure 8.5 présente le contour mélodique associé avec l'énoncé analysé : on observe une configuration avec un seul mouvement saillant réalisé sur la syllabe proéminente du mot « стоит » (vaut), le mot porteur du focus. Conjointement, les auditeurs attestent la présence d'une frontière « forte » (tableau IX.13) après le mot porteur de l'accent focal dans les conditions où les informations sémantico-syntaxiques ne leur sont pas disponibles.

Dans le contexte de l'approche accentuelle de la focalisation (*accent-to-focus*) nous avons analysé dans quelle mesure la distribution de proéminences mélodiques influençait la perception d'une frontière. Pour ce faire, nous avons analysé le taux de la perception d'une frontière après un mot porteur de la proéminence mélodique, avec toujours une distinction faite entre les frontières consensuelles *versus* optionnelles.

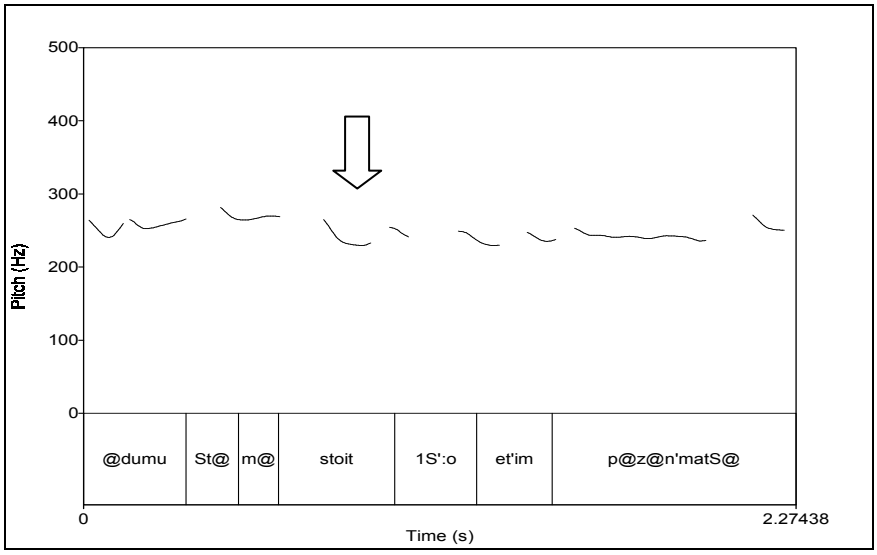


Figure 9.5 : Courbe mélodique de l'énoncé « Я думаю, что мне стоит еще этим позаниматься »

Dans cette analyse, la variable dépendante est donc binaire : nous avons voulu observer si les mots porteurs de la proéminence mélodique étaient plus fréquemment associés avec une frontière prosodique que les mots qui ne sont dotés que d'un accent lexical. Nous avons donc recouru au test de proportions (tableau IX.13). Si la comparaison est menée pour la totalité des frontières observées, le test de proportions démontre que les différences ne sont pas statistiquement significatives dans la performance des auditeurs russes. Or, pour les auditeurs français nous constatons une tendance marginale sous la troisième condition expérimentale et un effet assez marqué sous la deuxième condition (p = 0,004).

LANGUE	CONDITION	PROPORTION 1	PROPORTION 2	χ^2	VALEUR p
Russe	CondN	0.539	0.603	0.3555	0,551
	CondD	0.685	0.724	0.1006	0.751
Français	CondN	0.637	0.825	5.0848	0.024
	CondD	0.6264	0.8596	8.2708	0.004

Tableau IX.13 : Influence de la distribution des accents mélodiques sur la perception du phrasé

Si nous limitons nos analyses à des frontières consensuelles (tableau IX.14), les tendances sont différentes : aucun effet n'est constaté pour la deuxième condition chez les auditeurs russes ni pour la troisième condition chez les auditeurs français. D'autre part, l'effet est maintenu, voire augmenté, pour la troisième condition chez les auditeurs français. Or, en même temps, l'effet est dorénavant significatif pour la performance des auditeurs russes sous la troisième condition expérimentale.

LANGUE	CONDITION	PROPORTION 1	PROPORTION 2	χ^2	VALEUR p
Russe	CondN	0.236	0,466	7.4036	0.007
	CondD	0.258	0.327	0.519	0.471
Français	CondN	0.3956	0.4035	0	1
	CondD	0.1868	0.5087	15.4901	8.294e-05

Tableau IX.14 : Influence de la distribution des accents mélodiques sur la perception du phrasé (frontières consensuelles)

Quelle interprétation pourrait être proposée à ces faits ? D'une part, on distingue la focalisation de la focalisation étroite (contrastive) : de ce fait, la significativité du facteur « niveau de proéminence » pour des auditeurs russes dans le cas des frontières consensuelles peut être expliquée par la présence des cas de la focalisation large, qui présuppose l'association du dernier mot du syntagme avec la proéminence mélodique nucléaire (focale). L'absence d'effet dans certains cas nous fait penser à la théorie syntaxique de focus, développée par Selkirk (1995b). Il se peut alors que dans des conditions où les informations segmentales sont présentes, le groupement est assuré aussi par d'autres indices. Nous insistons tout de même sur le caractère exploratoire des analyses présentées.

Cependant, vu le rôle explicatif de la contrainte AlignFocus,R pour les frontières consensuelles non prédites par les contraintes syntaxiques, nous supposons que cette contrainte se situe plus haut dans la hiérarchie générale des contraintes régissant le phrasé prosodique (cf. tableau IX.15).

[je pense] [que *(cela)* me [vaut *(la peine)*]$_{FOC}$ encore en faire]	**AlignFocus,R**	**Wrap-XP**
☞(je pense)$_{IP}$ (que *(cela)* me vaut *(la peine)*) $_{IP}$ (encore en faire) $_{IP}$		*
(je pense)$_{IP}$ (que *(cela)* me vaut *(la peine)* encore en faire) $_{IP}$	!*	
(je pense que *(cela)* me vaut *(la peine)* encore en faire) $_{IP}$	!*	*

Tableau IX.15 : AlignFocus,R >> WrapXP, Align-XP,R

Ce résultat a également une implication théorique sur les propriétés de la hiérarchie des constituants prosodiques, plus précisément, sur la contrainte de non-récursivité. L'hypothèse de l'étagement strict (*Strict Layer Hypothesis*) proscrit les structures avec des constituants du type X dont les constituants immédiats sont également de type X. Cette contrainte n'est pas universellement acceptée par tous les chercheurs dans le domaine (cf. Ladd 1996). L'effet du focus sur le phrasé prosodique en perception confirme l'hypothèse que la proéminence mélodique marque la fin d'un domaine prosodique, le matériau qui suit forme un segment d'unité intonative avec un rôle discursif propre (que nous pouvons caractériser globalement comme un commentaire), le tout formant une macro-unité intonative (Di Cristo et Hirst 1996, cf. figure 9.6).

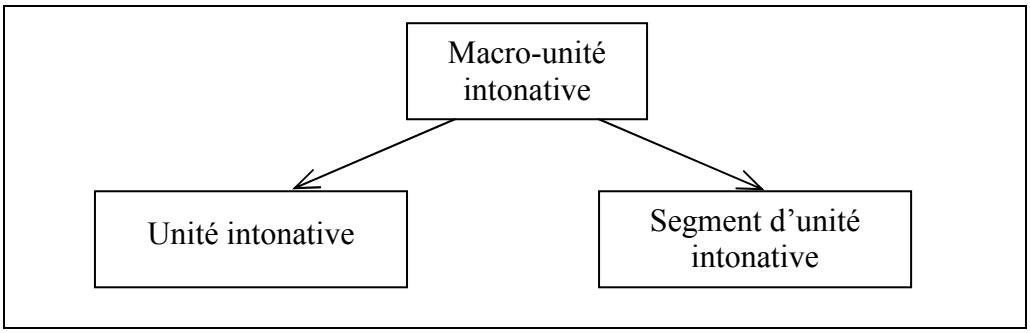


Figure 9.6 : Représentation d'une Macro-unité Intonative

9.4.4. Étude de cas : les pro-phrases

Dans la continuité des analyses entreprises dans la perspective « *top-down* », dans lesquelles notre point de départ est une hypothèse linguistique, nous nous sommes intéressée aux structures prosodiques associées avec des pro-phrases. C'est le terme générique utilisé dans l'analyse syntaxique pour désigner les mots d'interaction oui/non, qui permet de rendre tangible le fait que ces mots véhiculent un contenu propositionnel (Kerbrat-Orecchioni, 2001). Notre intérêt vers ces constituants a une double motivation. Dans un premier temps, il est considéré que certains types de constructions syntaxiques, « éléments externes à l'énoncé » dans la terminologie de Astruc (2003), forment des domaines intonatifs séparés : il s'agit des vocatifs, des insertions parenthétiques, d'appositions, de certains adverbes, des relatives non-restreintes. Dans la tradition de

l'analyse prosodique de la langue russe, c'est le cas également des pro-phrases 'oui' et 'non', ce statut externe étant soutenu en plus par les règles de ponctuation (une virgule est obligatoire à l'écrit). Dans un deuxième temps, dans l'étude de référence sur la prosodie de la parole spontanée en langue russe (Svetozarova *et al*. 1988), les auteurs ont constaté des différences entre la parole spontanée et la lecture dans la distribution du nombre de syntagmes en fonction du nombre de mots qu'ils contenaient. En particulier, on constate une prépondérance des syntagmes comportant un seul mot dans la parole spontanée. Ce fait est interprété en relation avec l'observation que le discours spontané, dialogique, comporte un nombre important de pro-phrases (oui, non), des conjonctions, des mots de liaison (les équivalents français étant *enfin*, *ben*, *quoi*) qui forment un syntagme à part. Remarquons que les deux contraintes syntaxiques prédisent une structuration à part pour les pro-phrases.

Dans notre corpus expérimental, presque un tiers des stimuli contiennent la construction désignée. Par conséquent, nous avons décidé d'analyser plus en détails la perception de leur structuration en termes d'unités intonatives.

Le tableau IX.16 résume les données sur la force de la frontière perçue après une pro-phrase. Un ensemble d'observations peuvent être faites. Tout d'abord, nous remarquons le statut particulier de la condition 3 (présentation normale des stimuli) dans la performance des sujets russes : presque tous les auditeurs marquent une frontière après les pro-phrases. Cette tendance contraste avec la performance des mêmes auditeurs dans d'autres conditions expérimentales, ainsi qu'avec la performance des sujets francophones (les nombres entre parenthèses indiquent la présence éventuelle d'une frontière décalée d'une syllabe atone).

STIMULUS	RU1	RU2	RU3	FR1	FR2	FR3
7	2	1	6	3	1	2
10	0	1	7	0	0	0
12	0 (5)	2 (1)	7	0(3)	0(1)	3
16	0	0	6	0	0	1
17	0	4	7	3	1	4
20	0	0	6	0	0	2
23	6	6	7	6	7	5
25	0(3)	0(1)	7	0(1)	3	0(1)

Tableau IX.16 : Distribution des forces de frontière perçue après les pro-phrases (nous indiquons entre parenthèses le nombre de cas de déplacement de la frontière d'une syllabe)

Il s'avère donc qu'une frontière structurelle après les pro-phrases n'est pas soutenue de manière universelle par des moyens prosodiques (cf. figure 9.7) : il est donc légitime de poser la question sur le statut des pro-phrases en tant que syntagmes intonatifs autonomes. Cette configuration peut en effet recevoir une explication du point de vue de la structure discursive : envisageant une argumentation de son point de vue, le locuteur ne veut pas être interrompu, il enchaîne donc ses propos, et l'absence d'une frontière prosodique marquée accomplit la fonction d'avertissement de l'auditeur sur l'intention de son interlocuteur.

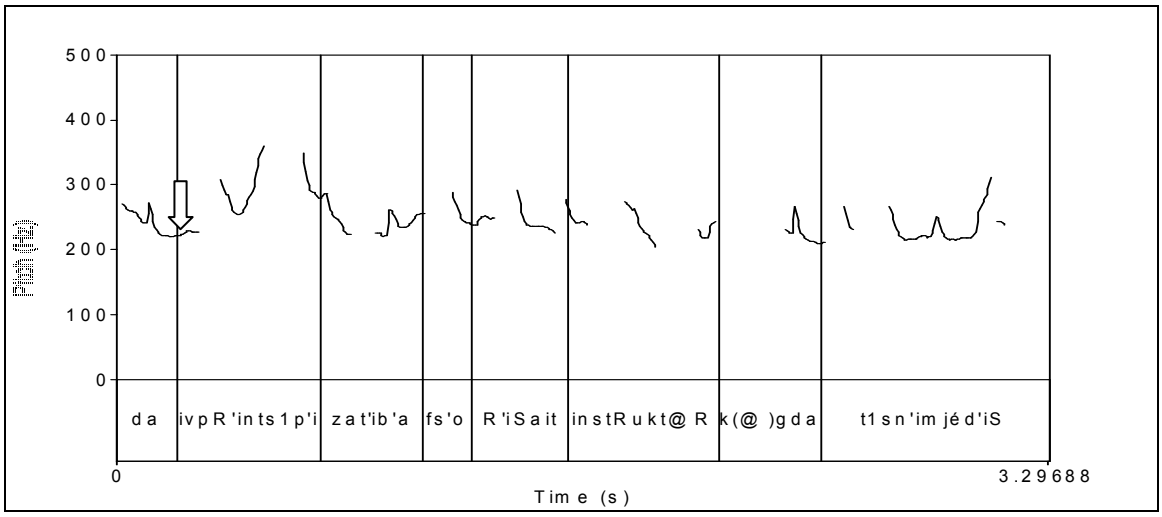


Figure 9.7 : Courbe mélodique de « Да, и в принципе, за тебя всё решает инструктор, когда ты с ним едешь[266] » : illustration du cas avec l'absence d'une frontière prosodique après la pro-phrase

Cependant, nous observons un cas divergeant de ce schème commun dans la distribution de la force des frontières perçues pour le stimulus 23 : la frontière après la pro-phrase y persiste dans toutes les conditions. Si nous analysons le tracé de f0 pour ce stimulus (figure 9.8) en le comparant avec ceux présentant une image opposée (cf. figure 9.7), nous observons que la rupture mélodique y est plus marquée.

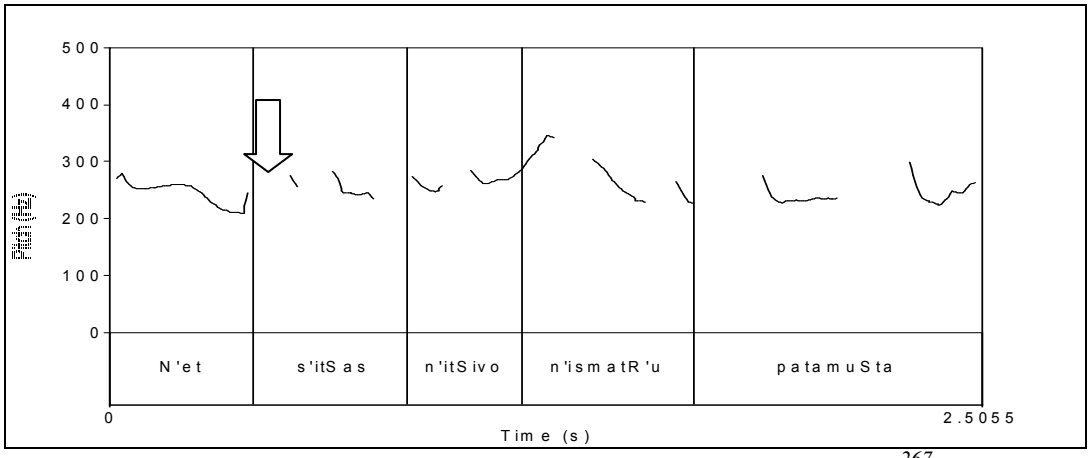


Figure 9.8 : Courbe mélodique de « нет, сейчас ничего не смотрю[267] » : illustration du cas une rupture mélodique après la pro-phrase

On observe donc une variabilité non-négligeable dans le marquage prosodique de pro-phrases, que l'on peut considérer comme important dans le traitement perceptif du message ; conjointement, nous sommes persuadée que cette variabilité nécessite d'être prise en compte dans le cadre de la synthèse de la parole conversationnelle.

[266] *« Oui, et en principe, l'est moniteur de conduite qui décide quand tu es avec lui »*.
[267] « *Non, en ce moment je ne regarde rien (à la télé)* »

9.5. DISCUSSION ET CONCLUSIONS

Dans les analyses du présent chapitre, nous sommes partie du principe selon lequel, dans le discours, les locuteurs segmentent prosodiquement un énoncé de manière à faire ressortir sa structure informationnelle. Les auditeurs, à leur tour, se voient obligés de restituer cette structure, dans l'objectif de réussir la communication. Après avoir proposé aux auditeurs la tâche de segmentation des énoncés sous différentes conditions expérimentales, nous avons analysé les données obtenues pour évaluer le rôle des différents principes explicateurs proposés auparavant.

Dans la présente contribution, nous avons fait la tentative de modéliser la perception du phrasé prosodique dans le cadre de la théorie de l'optimalité, et ceci en prenant en compte les interfaces « prosodie – syntaxe » et « prosodie – discours ». Nous avons donc vu dans quelle mesure les contraintes universelles pesant sur le phrasé prosodique, établies de manière intuitive et testées dans quelques études en production, s'appliquent aux données perceptives. Trois contraintes ont été prises en compte, toutes appartenant au groupe des contraintes d'alignement (nous réservons l'analyse du rôle des contraintes rythmiques pour des études postérieures quoique quelques indications aient été données dans le chapitre précédent).

Notre premier objectif a été d'évaluer la force prédictive du modèle ancré syntaxe. Nous avons pu constater que ce modèle permet jusqu'à 65% de prédictions correctes des frontières prosodiques perçues, et ceci pour des conditions où les informations sémantico-syntaxiques ne sont pas accessibles aux auditeurs. D'une part ce résultat corrobore l'hypothèse de la non-correspondance entre les structures prosodique et syntaxique, le pivot de la phonologie des domaines. Or, simultanément, ce résultat remet en cause, ou plutôt quantifie, le principe de superposition de Pechkovskij (1928) qui porte également sur les rapports entre syntaxe et prosodie : selon ce principe, si une relation est soutenue par des moyens grammaticaux, elle ne le serait pas par des moyens prosodiques, et inversement. Une autre interprétation consiste à dire qu'en présence d'indices grammaticaux, le marquage prosodique de la structure est optionnel. Des réflexions similaires sont menées dans le cadre de l'approche de compétition de la génération du sens avec les informations apportées par différentes composantes linguistiques (Blache et Di Cristo 2002), où les auteurs se penchent sur la question de savoir pourquoi la prosodie présente plus de variabilité dans certains contextes par rapport à d'autres. On rejoint également la problématique de la désambiguïsation des structures morpho-syntaxiques par des moyens prosodiques. Or, supposer que dans le corpus de test, 65% des structures présentent des ambiguïtés nous paraît peu probable. Nous constatons donc un certain parallélisme entre les structures prosodique et syntaxique, atteint moyennant le jeu d'indices acoustiques. Certes, le matériau sur lequel nous avons mené nos analyses n'est pas suffisamment large, mais cette hypothèse nécessitera une exploration approfondie dans le futur.

Pour les données examinées, nous considérons que la contrainte discursive de AlignFocus, R domine les contraintes émanant de l'interface « prosodie ~ syntaxe », Align-XP,R et Wrap-XP. Cette modélisation a abouti à une ébauche de la hiérarchie des contraintes, régissant le phrasé prosodique, que nous jugeons non-exhaustive pour deux raisons

majeures : premièrement, nous n'avons pas examiné le rôle des contraintes dites « phonologiques », ou « rythmiques » ; deuxièmement, la prise en compte des contraintes discursives doit recevoir, selon nous, un développement conséquent.

Nous avons également procédé à une étude de cas portant sur la problématique de l'organisation prosodique des énoncés avec des pro-phrases. Cette analyse nous a permis de mettre en évidence le rôle majeur des contraintes discursives pour la prédiction de la structure prosodique de l'énoncé. C'est ainsi que la probabilité qu'une pro-phrase forme une unité prosodique à part dépend fortement de l'intention du locuteur de continuer l'exposition de ces propos. Des analyses similaires sont selon nous nécessaires pour d'autres structures et en particulier, portant sur des structures avec des mots d'introduction (« *вводное слово* ») mais également sur la structure clausale (cf. Truckenbrodt 2005).

Nous soutiendrons donc le propos que dans la perspective d'une synthèse de la parole conversationnelle, on est amené à y intégrer – sous réserve des contraintes d'implémentation – un ensemble de contraintes, issu de l'interface « prosodie ~ discours ». Nous considérons que cette démarche permettra d'y apporter une variabilité qui est propre au discours de l'être humain (par l'intermédiaire de l'adoption d'une approche phonologique de type probabiliste) ce qui aura pour effet de dépasser la monotonie de la prosodie générée par le caractère récursif des règles phonologiques. Cependant, d'autres recherches sont nécessaires afin de nuancer le rôle et le fonctionnement de cette interface.

CHAPITRE X
MODÉLISATION ACOUSTIQUE DES FRONTIÈRES PERÇUES

10.1. INTRODUCTION

Du point de vue du signal acoustique, il semble que le phrasé prosodique constitue une structure qui permet d'élaborer une explication des plus cohérentes de la variation des indices acoustiques observés dans la parole (Ostendorf 2000:263). Dans le présent chapitre nous nous intéressons à un modèle acoustique du phrasé prosodique perçu qui pourrait aboutir à une heuristique prédictive, applicable à l'annotation automatique du corpus en termes d'unités intonatives. Plus précisément, nous cherchons les paramètres acoustiques qui semblent les plus pertinents en vue d'une prédiction fondée sur les résultats de l'expérience en perception. Notre démarche commence donc avec une mise en correspondance des annotations subjectives des frontières prosodiques et des paramètres acoustiques extraits du corpus (corpus restreint de l'étude en perception).

Dans le chapitre V, nous avons passé en revue différents travaux s'intéressant aux indices acoustiques et articulatoires du phrasé prosodique. Dans le domaine acoustique, l'indice le plus fréquemment cité étant la pause[268], nous avons précisé notre position à ce sujet : nous considérons qu'une pause acoustique n'est ni un indice obligatoire ni un indice nécessaire pour la perception d'une frontière prosodique (cf. Cho et Hirst 2006). Ce principe a été appliqué lors du choix des stimuli : aucun de nos stimuli ne comporte une pause silencieuse structurelle[269]. Cette remarque étant faite, nous avons donc essentiellement pris en considération les indices temporels et mélodiques. Le présent chapitre décrit les analyses et les réflexions, la prise de positions et le protocole de la prise de mesures sur lesquels nous nous sommes fondée pour atteindre l'objectif fixé.

Nous avons présenté dans le chapitre VI des analyses qui se penchent sur la prédiction du phrasé prosodique à partir du texte, une étape indispensable dans un système TTS. Simultanément, nous avons argumenté en faveur de la méthodologie de sélection des unités, assurant une meilleure flexibilité du module de la génération du signal. Pour son application, il est nécessaire de disposer d'un corpus oral annoté et nous cherchons donc

[268] Cf. également les Cooper et Paccia-Cooper (1980), Gee & Grosjean (1983) : la tâche que les chercheurs se donnent est celle de la prédiction de la distribution des pauses silencieuses en fonction de leur durée ; or, en réalité il s'agit de la prédiction du phrasé prosodique.
[269] Nous avons discuté de la typologie des pauses dans le chapitre V.

un procédé permettant une segmentation automatique du corpus oral. Nous nous intéressons à l'annotation prosodique du corpus en termes de syntagmes intonatifs définis d'un point de vue perceptif (à cette étape nous faisons abstraction de la problématique de la valeur linguistique de ces syntagmes, ce sujet étant abordé dans le chapitre précédent). Pratiquement, notre objectif est donc de proposer un système de classement qui permet, pour toute position analysée, de déterminer s'il s'agit d'un emplacement de la frontière prosodique. Du point de vue méthodologique nous rejoignons donc la problématique des méthodes statistiques évoquées dans la discussion de la génération de l'intonation dans les systèmes de synthèse. Nous avons présenté des critiques émises par van Santen et Sproat (1998) vis-à-vis des méthodes statistiques, critiques que nous partageons. Or, nous considérons que, dans la perspective de l'annotation des données orales, la réflexion linguistique intervient à l'étape de la construction du vecteur d'indices acoustiques. D'autre part, comme le remarque Ostendorf (2000:266),

> *« since so much is still not understood about prosodic structure and associated acoustic correlates, automatic training methods are essential for handling our lack of knowledge about specific cues and how they interact, i.e. variability that is not explained by current linguistic theories. »*

Les questions de la détection des événements prosodiques dans le signal de parole sont abordées dans le cadre de la recherche en reconnaissance vocale. Un éventail de méthodes y sont développées par rapport au phrasé prosodique, on y distingue les méthodologies modélisant le contour entier *versus* celles modélisant la séquence d'événements tonals ; une autre typologie est concevable selon que l'on travaille avec une fenêtre de taille fixe définie en unités linguistiques (en nombre de syllabes, par exemple) ou en durée absolue (cf. Braunschweiler, 2003). En plus des informations phonologiques, le système peut intégrer et profiter des hypothèses sur la séquence des mots. Ostendorf (2000) signale que les techniques proposées auparavant ont été développées et testées avec de la parole lue ; leur extension doit permettre de travailler sur de la parole spontanée. Comme l'auteur le signale,

> *« at a minimum, the « knowledge » that is incorporated via the choice of features provided to the classifier must be revised to incorporate our growing understanding of the differences betweens read and spontaneous speech ».*

Les travaux qui traitent de la problématique de l'annotation du corpus oral en termes de frontières prosodiques et qui sont menés dans une logique de prédiction, sont assez peu nombreux, la question de l'annotation des proéminences ayant bénéficié d'une plus grande attention (Streefkerk et al. 1997, Kochansky et al. 2005). Toutefois, nous pouvons prendre ces études en considération lors de l'élaboration du cahier de charges que nous mettrons en place ultérieurement.

L'originalité de notre démarche réside entre autres dans la prise en considération des données issues de l'expérience en perception. Même quand il s'agit d'une modélisation acoustique, certaines études se fondent sur l'annotation des phénomènes prosodiques à partir du texte orthographié : or, cette annotation, bien qu'elle soit effectuée par des

linguistes expérimentés, ne garantit pas que tous les phénomènes annotés ont été réalisés comme tels par le locuteur.

10.2. MÉTHODOLOGIE

Des analyses préalables ont été menées sur la totalité des données dont nous disposons (toutes les frontières annotées ont été prises en compte). Les contrastes étudiés se rapportaient à la présence *versus* l'absence d'une frontière (pour la catégorie de présence quelle que soit le nombre d'auditeurs qui l'ont marquée) et à une distinction ternaire entre absence d'une frontière, présence d'une frontière faible et présence d'une frontière forte. Les résultats obtenus n'étant pas convaincants (des tendances faibles se dégageaient sans apporter plus d'information), nous avons décidé de trier nos données. Nous avons opté pour une démarche restrictive, et nous nous sommes intéressée, dans un premier temps, aux frontières consensuelles, c'est-à-dire des frontières marquées par quatre auditeurs et plus sur sept dans l'étude perceptive.

Cependant, le reproche que l'on fait fréquemment aux études en prédiction porte sur un fort déséquilibre observé entre les effectifs selon les catégories analysées : en effet, quelle que soit l'unité d'analyse, la catégorie « absence de frontière prosodique » sera plus fréquente. Par conséquent, le modèle qui en résultera risque d'être biaisé, et risque d'afficher des valeurs élevées de bruit et de silence[270] dans le test d'évaluation, notamment pour la catégorie « présence de la frontière ». Si nous prenons l'exemple de données présentées par Heldner et Megyesi (2003), nous voyons que la performance du modèle est bonne pour la catégorie « absence d'une frontière » (93,6 % de prédictions correctes), or, pour les cas de « présence d'une frontière prosodique forte ou faible », on note un abaissement de la performance avec 42,5 et 48,8 % de prédictions correctes.

Pour remédier à ce problème nous avons cherché à équilibrer nos données. Tout d'abord pour la catégorie « absence d'une frontière prosodique » nous n'avons pris en compte que des positions incontestables : il s'agit des frontières syllabiques pour lesquelles il n'a pas été signalé de présence d'une frontière prosodique pour les deux syllabes immédiatement adjacentes (à gauche et à droite) à la syllabe analysée. Dans un deuxième temps, de cet ensemble des positions pour la catégorie « absence » nous avons choisi au hasard le même nombre d'observations dont nous disposions pour la catégorie de la présence d'une frontière prosodique.

Afin d'aboutir à une modélisation acoustique des frontières prosodiques, nous avons examiné les différentes mesures de f0, de durée et d'intensité, extraites automatiquement du signal. Ces mesures acoustiques quantifient les procédés phonétiques et phonologiques que l'on associe au phrasé : ton de frontières, réajustement du niveau (*resetting)*, allongement final, mais également proéminence relative.

[270] Voire pour définition p.324.

10.2.1. Extraction d'indices acoustiques

Nous avons choisi d'extraire toutes les données quantitatives de manière automatique après avoir réalisé une l'annotation multiniveaux de notre corpus. L'extraction a été faite à l'aide de scripts Praat (dossier 'Scripts' sur le CD-rom) que nous avons rédigés.

10.2.1.1. Domaines d'analyse

La première décision à prendre est celle du domaine dans lequel les indices acoustiques seront extraits, analysés et comparés.

Dans l'étude de Heldner *et al.* (2004) menée dans le cadre du projet GROG l'attention des chercheurs se concentre sur le dernier mot avant la frontière prosodique analysée. Tous les indices tonals sont extraits pour trois domaines dont l'empan est inférieur au mot : l'unité structurelle du mot y est tout de même présente et pertinente. Les trois domaines sont : l'attaque plus noyau de la première syllabe du mot, la rime de la syllabe accentuée, la rime finale du mot (la rime de la dernière syllabe du mot).

Dans le chapitre IV, nous avons débattu du statut incertain du mot grammatical en tant qu'unité prosodique. Lors de l'étape d'annotation de notre corpus en phonèmes et en syllabes, nous avons pu constater que les frontières de mots sont des positions faibles du point de vue acoustique (à ce propos nous avons évoqué les travaux de Kassevitch 2003, *p. X*). Nous avons donc décidé de ne pas tenir compte des frontières de mots et de nous placer dans la perspective d'une analyse non-lexicale des faits prosodiques (*Non-lexical Prosodic Analysis*, Aylett 2005)[271], c'est-à-dire, faire abstraction du contenu sémantique du message. Pour cette raison, nous avons choisi de travailler à partir de la segmentation des stimuli en syllabes, étant donné qu'il existe un ensemble d'algorithmes permettant une segmentation automatique du corpus en (pseudo-)syllabes (cf. Howitt 2000 pour le résumé récent des travaux dans le domaine).

Remarquons que les méthodes computationnelles à part entière choisiront plutôt de travailler avec des valeurs détectées toutes les 5-10 ms, sans faire référence à un concept linguistique quelconque (Braunschweiler 2003).

L'étude de Tao (2004) porte de manière explicite sur la taille d'une fenêtre d'analyse pour la prédiction de la distribution des frontières prosodiques en chinois. L'auteur détermine que, pour les paramètres acoustiques, la fenêtre optimale est donc celle de trois syllabes, avec deux syllabes précédant la frontière et une syllabe post-frontalière. Pour notre part nous n'avons pas mené une telle étude exploratoire pour déterminer la taille de la fenêtre la plus convenable. De manière arbitraire, nous avons choisi de retenir les domaines pré- et post-frontalier de deux syllabes. La comparaison des mesures acoustiques de f0 entre ces deux empans temporels est directement associée avec la notion du réajustement (*resetting*) du registre mélodique aux frontières prosodiques.

[271] Par ce choix nous ne pouvons pas non plus prendre en considération les hypothèses sur la séquence des mots qui constituent la phrase analysée, le procédé intégré dans certains systèmes de reconnaissance vocal.

Nous avons tenté de formaliser également la notion de ton de frontière. À cet effet, nous avons analysé l'organisation temporelle des unités intonatives résultant de l'expérience de perception. Nous avons pu constater que pour les unités non-terminales, les grandes moyennes par condition et par langue maternelle des auditeurs ne montrent pas beaucoup de variabilité. Nous avons donc défini une unité intonative potentielle se terminant à la frontière syllabique donnée ; cette unité est définie en se référant à la durée moyenne des syntagmes non-terminaux. Par la suite, les différentes mesures de f0 dans la zone préfrontalière sont comparées avec celles de l'unité intonative potentielle afin de rendre compte du concept du ton de frontière.

Une série d'indices acoustiques en association avec l'organisation tonale et temporelle du message a été obtenue à partir du corpus annoté. De ce fait, les analyses ont porté sur la performance des auditeurs sous les conditions expérimentales CondD et CondN, la condition rectotonale étant exclue.

10.2.1.2. Les mesures de la fréquence fondamentale

La plupart des mesures dont nous avons analysé l'impact proviennent du domaine tonal. Nous avons choisi de travailler à partir de la courbe modélisée de la f0 et non à partir de la courbe brute détectée par le logiciel PRAAT. Nous considérons les variations microprosodiques comme n'étant pas pertinentes dans l'analyse que nous entreprenons (cf. également le propos de 't Hart (1984:195), notamment en relation avec l'approche perceptive :

> *« [we should] try to draw a graph on top of the F0 curve which should represent the course of pitch as a function of time as it is perceived by the listener. »)*

Pour les valeurs de f0, nous avons travaillé à partir de la courbe mélodique modélisée par l'application de l'algorithme MOMEL. Outre l'effacement des effets microprosodiques, cette manipulation permet également d'effectuer une interpolation sur les segments non-voisés.

Pour chacun des empans temporels définis (empans pré- et post-frontaliers, unité intonative potentielle), les scripts de PRAAT nous permettent de récupérer les paramètres des minima et des maxima de f0 et leurs alignements temporels respectifs, la valeur moyenne de f0 et l'écart type : nous disposons ainsi de 12 paramètres initiaux. À côté de ces paramètres bruts, nous avons procédé au calcul des paramètres dérivés caractérisant les ratios (ou les écarts, si nous nous utilisons une échelle logarithmique) entre les différents empans. Nous présenterons ces mesures une fois la procédure de normalisation des données fréquentielles discutées.

a) <u>Normalisation</u>

Dans le cadre de la présente étude nous travaillons avec les productions d'une seule locutrice : nous évitons ainsi d'être confrontée au problème de la variabilité du registre mélodique du locuteur, auquel cas il est nécessaire de procéder à une normalisation des

valeurs absolues. Cependant, nous avons insisté à plusieurs reprises sur la nécessité de développer une méthodologie qui puisse être dupliquée sur d'autres corpus : cette quête de reproductibilité nous oblige donc à appliquer une normalisation des valeurs brutes détectées.

Nous avons procédé à une double normalisation des données fréquentielles en appliquant la même méthode que celle qui est décrite dans Auran (2004).

La figure 10.1 présente un histogramme des valeurs brutes de la f0 moyenne pour l'empan préfrontalier. Nous observons :

a) que la distribution est limitée à gauche et que les valeurs négatives, voire inférieures à un seuil, sont absentes ;

b) qu'une grande partie de données est concentrée « à gauche » de la distribution et que les données sont ainsi accumulées en début de distribution.

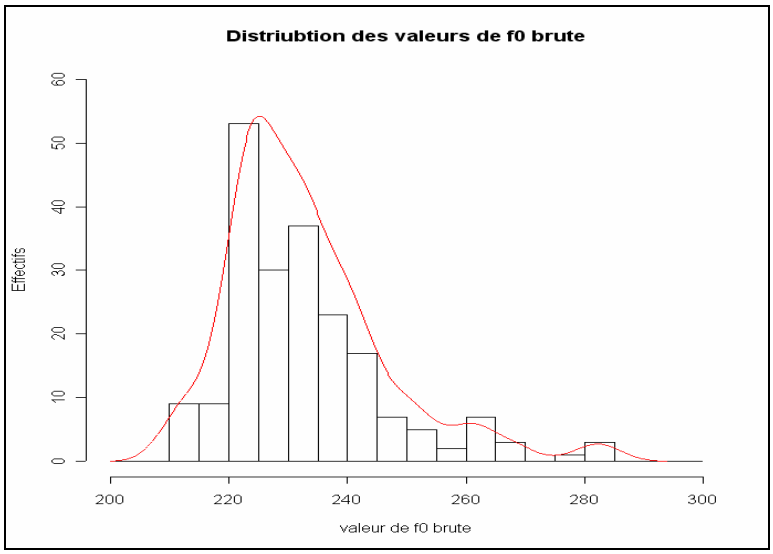


Figure 10.1 : Histogramme et courbe de densité des valeurs brutes de f0 pour la mesure de la f0 moyenne dans l'empan pré-frontalier.

Bien qu'elle soit souhaitable, la normalité de la distribution n'est pas requise par toutes les méthodes statistiques. Pour remédier aux artéfacts mentionnés ci-dessus, nous avons appliqué dans un premier temps une transformation logarithmique, qui aboutit à une distribution plus proche de la distribution normale (figure 10.2). Nous considérons que cette normalisation est légitime étant donné l'orientation de notre étude vers la perception. Cette transformation permet d'effectuer un rapprochement avec l'échelle logarithmique de demi-tons (cf. 't Hart, Collier, & Cohen, 1990).

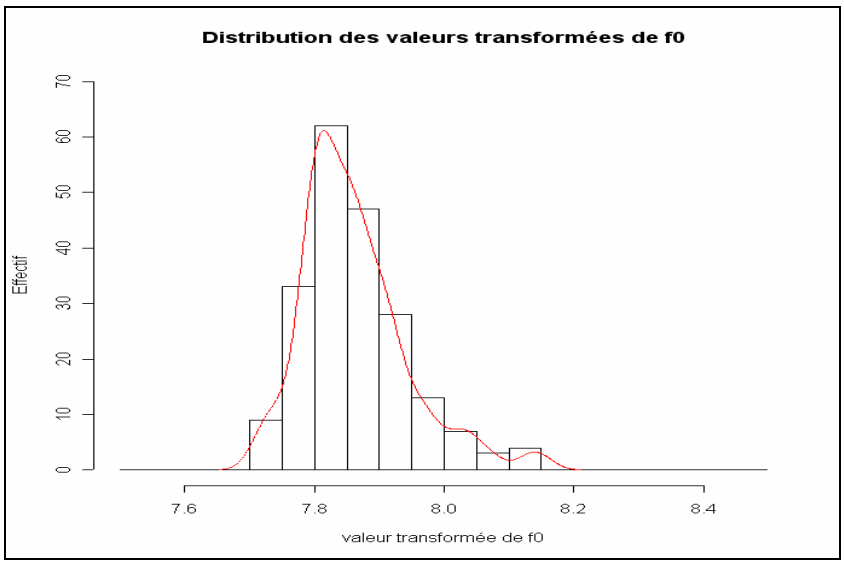


Figure 10.2 : Histogramme et courbe de densité des valeurs de f0 après la transformation logarithmique (pour la mesure de la f0 moyenne dans l'empan pré-frontalier)

Nous avons appliqué par la suite une normalisation en « transformée z » des valeurs logarithmiques. Nous exprimons ainsi une valeur donnée en termes de différence par rapport à la moyenne divisée par l'écart type. Ce procédé permet d'obtenir un ensemble de valeurs indépendantes du registre et de l'étendue tonale du locuteur, ce qui permet ainsi une vérification et une application du classificateur à des corpus différents.

b) *Mesures dérivées*

À côté des mesures calculées à partir du signal acoustique nous avons également inclus des mesures dérivées, calculées par combinaison de mesures initiales. Nous présentons le sommaire des mesures dérivées dans le tableau X.1 ci-dessous. Ce tableau peut présenter des redondances. Cependant, il nous parait nécessaire dans la mesure où rien ne justifie de privilégier les mesures de la f0 moyenne aux mesures des maxima ou des minima.

Le tableau X.2 ci-après relie les indices mélodiques aux phénomènes acoustiques à la portée de la fonction prosodique du groupement. Nous retenons trois phénomènes évoqués dans la littérature et ayant un lien avec le phrasé prosodique : tons de frontière, réajustement du niveau *(resetting)* et phénomènes du registre.

1	*slope*	Vélocité de changement de f0 sur l'empan de l'unité intonative potentielle ; le calcul est guidé par la formule : (*maxf0 – minf0*)/[abs(*temps_maxf0 – temp_minf0*)]
2	*aslope*	Vélocité de changement de f0 sur l'empan pré-frontalier
3	*bslope*	Vélocité de changement de f0 sur l'empan post-frontalier
4	*amaxmean*	Différence entre la valeur de f0 maximale sur l'empan pré-frontalier et la moyenne de f0 pour l'unité intonative potentielle[272]
5	*amaxmax*	Différence entre les maxima de f0 sur l'empan pré-frontalier et sur l'unité intonative potentielle
6	*aminmean*	Différence entre la valeur de f0 minimale sur l'empan pré-frontalier et la moyenne de f0 pour l'unité intonative potentielle
7	*aminmin*	Différence entre les minima de f0 sur l'empan pré-frontalier et sur l'unité intonative potentielle
8	*ameanmean*	Différence entre les valeurs moyennes de f0 sur l'empan pré-frontalier et sur l'unité intonative potentielle
9	*asdsd*	Différence entre les écarts type de f0 sur l'empan pré-frontalier et sur l'unité intonative potentielle
10	*abmaxmean*	Différence entre le maximum de f0 sur l'empan pré-frontalier et la moyenne de f0 sur l'empan post-frontalier
11	*abmeanmax*	Différence entre la valeur moyenne de f0 sur l'empan pré-frontalier et le maximum de f0 sur l'empan post-frontalier
12	*abmaxmax*	Différence entre les maxima de f0 sur l'empan pré-frontalier et le maximum de f0 sur l'empan post-frontalier
13	*abminmean*	Différence entre le minimum de f0 sur l'empan pré-frontalier et la moyenne de f0 sur l'empan post-frontalier
14	*abmeanmin*	Différence entre la valeur moyenne de f0 sur l'empan pré-frontalier et le minimum de f0 sur l'empan post-frontalier
15	*abminmin*	Différence entre les minima de f0 sur les empans pré- et l'empan post-frontaliers
16	*abmeanmean*	Différence entre les valeurs moyennes de f0 sur les empans pré- et l'empan post-frontaliers
17	*absdsd*	Différence entre les écarts type de f0 sur les empans pré- et l'empan post-frontaliers

Tableau X.1 : Mesures dérivées de f0

[272] Nous avons calculé la différence (résultant de la soustraction des valeurs respectives) comme nous travaillons sur l'échelle logarithmique. Pour échelle standard en Hz il s'agit du ratio.

Tons de frontières	valeurs de la pente (*slope, aslope, bslope*) ; les différences dans les maxima et minima entre l'empan pré-frontalier et l'unité intonative potentielle (*amaxmax, aminmin*) ; quantité de la variabilité dans les valeurs de f0 (*asdsd*)
Phénomènes de registre	différences entre les valeurs caractérisant l'empan préfrontalier par rapport à la moyenne de l'unité intonative potentielle (*amaxmean, aminmean, ameanmean*)
Resetting	mesures, reliant les valeurs dans l'empan pré- et post-frontalier

Tableau X.2 : Mise en correspondance des mesures de f0 dérivées avec des phénomènes mélodiques frontaliers, débattus dans la littérature

10.2.1.3. Mesures de durée

Il est bien connu (cf. Di Cristo 1985, Klatt 1987) que l'identité du phonème influence fortement sa durée, chaque phonème possédant une durée intrinsèque. C'est pour cette raison que nous travaillerons sur des durées normalisées. Nous avons appliqué la procédure de normalisation en « transformée z », nous inspirant du travail de Campbell (1992). Au niveau syllabique la durée de la syllabe est normalisée par référence à la durée moyenne des phonèmes qui le composent. On obtient la durée normalisée selon la formule :

$$Z = \frac{\mathbf{Durée_syllabe} - \sum_{i=1}^{Nb_phonèmes} (\mathbf{m_{phoni}})}{\sum_{i=1}^{Nb_phonèmes} (\mathbf{sd_{phoni}})}$$

La moyenne et l'écart type de chaque phonème sont obtenus sur l'ensemble de notre corpus ; l'application de cette procédure est d'autant plus légitime que nous travaillons sur les productions d'une seule locutrice. Notons que pour pouvoir transférer le protocole d'analyse sur d'autres corpus, une normalisation des mesures s'impose. De même, une fois le choix de normalisation fait, nous pouvons donc mener nos analyses sans tenir compte du nombre de phonèmes par syllabe ni de l'identité des phonèmes.

Le paramètre de z-score n'ayant pas de sens pour l'unité intonative potentielle, nous avons au total cinq paramètres qui caractérisent l'organisation temporelle au niveau syllabique pour toute position de frontière : z-score de la syllabe en question, les z-scores des syllabes qui lui sont immédiatement adjacentes ainsi que les ratios de z-scores entre les deux syllabes adjacentes.

Le tableau de données est organisé autour de l'unité d'analyse qui est une syllabe, ou plutôt une frontière syllabique pour laquelle nous cherchons à prédire si elle est alignée avec une frontière prosodique du niveau supérieur (celui de l'unité intonative). Outre l'annotation en termes de présence *versus* absence d'une frontière prosodique, toute entrée de ce tableau comporte des informations sur les indices acoustiques retenus.

10.2.1.4. Amplitude totale

Dans les chapitres précédents, en discutant du phrasé prosodique et des approches autosegmentales-métriques de l'intonation nous avons insisté sur un lien étroit entre les phénomènes du phrasé et le jeu de proéminences. Nous pensons notamment à l'approche défendue par Gussnehoven (1990), testée expérimentalement dans l'étude de Gussenhoven et Rietveld (1992).

Beckman (1986) propose de caractériser la proéminence perceptive d'une unité par l'amplitude totale, qui intègre l'intensité instantanée sur le temps et correspond à l'énergie dépensée pour produire l'unité en question (cf. figure 10.3). Cette mesure est introduite par l'auteur pour des langues à intonation (dans la terminologie de l'auteur *stress accent langages*) : dans ce groupe de langues plusieurs indices acoustiques concourent à produire l'impression d'une proéminence, contrairement aux langues à accent mélodique comme le japonais (*pitch accent langages*) où seule la variation de la courbe mélodique entre en jeu.

Nous avons calculé l'amplitude totale pour les empans pré- et post-frontaliers. Nous avons adopté un calcul approximatif, à savoir : nous avons multiplié la durée observée de l'empan analysé par l'intensité moyenne sur cet empan. Ensuite, nous avons également inclus dans notre ensemble de mesures le ratio entre les amplitudes totales observées avant et après la frontière présumée.

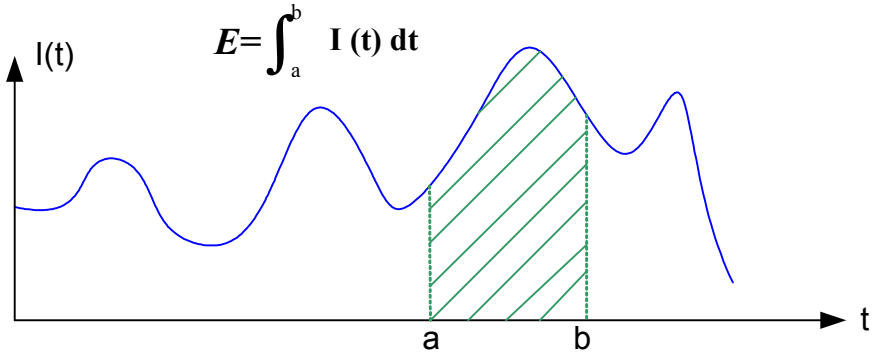


Figure 10.3 : Amplitude totale

10.2.2. Analyse discriminante

Nous utilisons un ensemble de tests basés sur la technique d'analyse discriminante. Les analyses discriminantes effectuées ont pour objectif de mettre en évidence dans quelle mesure les catégories de « présence d'une frontière prosodique » et « absence de frontière » peuvent être séparées sur la base des indices acoustiques que nous avons choisis.

Cette technique a été introduite par Fisher (1936). Pour appliquer cette technique, il est nécessaire de disposer d'un ensemble de *x* classes et de connaître pour toute entrée son appartenance à la classe prédéfinie. Cette information sur l'appartenance à une classe

est ensuite utilisée pour révéler la structure interne des données. On cherche une combinaison linéaire des variables résultant en une variable dérivée, qu'on nomme parfois discriminant linéaire, qui assure la meilleure discrimination entre les classes. Selon le théorème de Pythagore, la variance observée pour une variable peut être décomposée en variance interclasse et la variance intraclasse : si v_k est la variance de la variable x_k, b_k est la variance interclasse de la variable x_k, et w_k est la variance intraclasse de x_k (la moyenne des variances dans chaque classe pondérées par les poids des classes), alors

(a) $$v_k = b_k + w_k$$

La variable x_k a une valeur discriminante qui se mesure par l'équation (b) :

(b)
$$\frac{b_k}{v_k} = \frac{b_k}{b_k + w_k} = \frac{1}{1 + \frac{w_k}{b_k}} = \frac{v_k - w_k}{v_k} = 1 - \frac{w_k}{v_k}$$

L'équation (b) permet de voir que la valeur discriminante d'une variable est maximale si un des deux cas suivants est vrai :

a) on observe une valeur élevée de ratio entre la variance interclasse ou la variance totale (et, par conséquent, une petite valeur de ratio entre les variance inter- et intraclasse) ;

b) on observe une valeur élevée de ratio entre les variances inter- et intraclasse (et une petite valeur de ratio entre la variance interclasse ou la variance totale).

Nous avons appliqué la procédure qui maximise la variation interclasse.

Nous avons utilisé la fonction *lda* de R qui donne pour toute variable en entrée le coefficient linéaire avec lequel elle participe à la variable discriminante. La variable discriminante est retrouvée elle aussi : nous pouvons la soumettre par la suite à un test de régression linéaire, pour évaluer le pourcentage et la significativité de la variance expliquée.

Il est également possible de se situer dans la perspective d'une discrimination prédictive et de se concentrer sur la question de l'affectation d'une observation à une classe. La fonction discriminante obtenue permet de procéder à une classification pour tout *token* observé. Ce procédé permet d'adopter une autre stratégie d'évaluation du modèle. C'est ainsi qu'une matrice de classification est générée, laquelle met en rapport pour tout événement analysé sa catégorie dans la matrice initiale et la catégorie prédite. Nous avons effectué les deux types d'évaluation.

L'analyse discriminante a été appliquée dans les études de Turk et Shattuck-Hufnagel (2000), Heldner et Megyesi (2003), Aylett (2005) en relation avec les différents niveaux de la structure prosodique et dans l'étude de Hirst (2004) pour la prédiction de la langue en relation avec les paramètres mélodiques observables. L'étude de Turk et Shattuck-Hufnagel (2000) cherche à prédire trois modèles de la structuration en mots en fonction des durées observées des syllabes adjacentes à frontière. Le modèle donne une

prédiction correcte de 44,04%. Heldner et Meguesi (2003) cherchent à modéliser la force de frontière prosodique perçue (variable ayant trois modalités « absence de frontière », « présence d'une frontière faible », « présence d'une frontière forte ») en fonction des variables acoustiques (temporelles dans cette étude) et linguistiques. Leur modèle donne une prédiction correcte de 85,3 %, avec une très bonne performance pour la catégorie « absence de frontière » (92,5%) et une performance moyenne pour les deux autres catégories (42,3% et 49,3% respectivement). À son tour, Aylett (2005) cherche à distinguer les frontières intonatives structurelles des frontières associées avec des phénomènes de dysfluence en parole spontanée. Le système de classement se fonde sur des indices dérivés de la f0 et l'indice d'amplitude totale (remarquons que l'indice de présence d'une pause n'est pas pris en compte) : le modèle donne 58% de classifications correctes, ce résultat révélant une forte variabilité inter-locuteur.

10.3. RÉSULTATS

10.3.1. Phase d'entraînement : corpus de test

Tout d'abord, nous avons soumis les 37 paramètres retenus à une analyse par régression linéaire afin d'évaluer si les différences étaient significatives entre les niveaux du facteur « absence *versus* présence d'une frontière prosodique après la syllabe ». Le tableau X.3 résume les paramètres significatifs au niveau de $p < 0.01$ par condition expérimentale, l'ordre des variables correspondant à une significativité décroissante.

LANGUE	CONDITION	PARAMÈTRE	F
Russe	CondN	*abamplitot*	19,8
		aamplitot	18,38
		zscore	11,138
		azscore	6,858
		amaxmax	6,363
	CondD	*abminmin*	16,848
		abminmean	10,623
		abmeanmin	6,954
Française	CondN	*aamplitot*	25,369
		zscore	16,612
		abamplitot	15,024
	CondD[273]	*aamplitot*	19,743
		asdsd	10,240
		amaxmax	9,831
		zscore	9,907
		abamplitot	7,158

Tableau X.3 : Paramètres acoustiques significatifs par condition (degré de significativité décroissant)

[273] Pour rappel, sous la condition CondD les stimuli ont été filtrés, tout en gardant les caractéristiques prosodiques ; la condition CondN correspond à la présentation des stimuli non-manipulés.

Ensuite, nous avons soumis l'ensemble des indices retenus à une analyse discriminante. D'abord, les tests ont été effectués avec toutes les variables retenues incluses. À la fin de chaque analyse, nous obtenons une matrice de classification qui résume les effectifs pour les deux niveaux du facteur dans les données du test et à l'issue de la classification, appliquant le modèle de l'analyse discriminante. Un exemple d'une telle matrice est présenté dans le tableau X.4 pour la deuxième condition expérimentale avec des sujets francophones. Les informations résumées dans ce tableau reçoivent une interprétation suivante : sous la deuxième condition expérimentale on observe 45 frontières intonatives consensuelles chez les auditeurs francophones ; nous incluons donc dans l'extrait de test 45 positions caractérisées par l'absence de frontière. À l'issue du traitement de prédiction avec l'analyse discriminante, 40 cas d'absence de frontière ont été classés correctement, et 5 positions ont été classées de façon erronée dans la catégorie de la présence d'une frontière ; en revanche, pour la catégorie de présence d'une frontière prosodique la classification inexacte a été observée pour 9 positions, avec 36 classées correctement.

	Catégorie prédite	
	Absence de frontière	Présence de frontière
Absence de frontière	40	5
Présence de frontière	9	36

Tableau X.4 : Matrice d'effectifs pour des frontières prosodiques selon que ces dernières sont attestées dans l'expérience et prédites par le modèle acoustique

À partir de ces matrices nous avons calculé les mesures évaluatrices suivantes :

a) le pourcentage des prédictions correctes (égal au ratio des valeurs diagonales sur la totalité des cas analysés) ;

b) coefficient de Kappa pour la matrice de classification ; il s'agit d'une statistique Kappa similaire à celle que nous avons présentée dans le chapitre VIII. Nous voulons tout de même attirer l'attention sur l'interprétation du degré d'accord attendu par hasard : dans le contexte de l'étude prédictive du phrasé prosodique à partir d'un ensemble d'indices acoustiques, le p(e) est la probabilité de la classification correcte en l'absence de toute information acoustique.

c) nous présentons également les valeurs de F-measure pour les deux niveaux du facteur « frontière » (présence *versus* absence d'une frontière prosodique).

LANGUE	CONDITION	POURCENTAGE de prédictions correctes	KAPPA	F-MEASURE « absence de frontière »	F-MEASURE « présence d'une frontière prosodique »
Russe	III	80,85	61,7	0,8085	0,8085
	II	76,74	53,49	0,767	0,767
Français	III	84,37	68,75	0,8387	0,8485
	II	84,44	68,89	0,851	0,8372

Tableau X.5 : Évaluation de la démarche prédictive pour le modèle avec 37 indices acoustiques retenus

Les résultats présentés dans le tableau X.5 résument et évaluent la performance du modèle avec les 37 indices présents. Nous constatons une performance du système de classification de l'ordre de 77-84%, des valeurs comparables avec celles communiquées pour la prédiction du phrasé à partir du texte (Ostendorf 2000). Nous observons des valeurs de Kappa suffisamment élevées, mais également des *f-measures* comparables pour les deux catégories (présence *versus* absence de frontière).

Cependant, une modélisation avec 37 indices relève plutôt d'une approche statistique, de « l'approche d'ingénieur », fondée sur un calcul parfois trop puissant. Nous nous sommes posée la question de l'apport de chacun des paramètres que nous avons inclus dans le modèle. Pour ce faire, nous avons mené une série d'analyses discriminantes avec toutes les combinaisons de un, deux et trois paramètres de notre ensemble de données initial. Ces analyses ont été effectuées à l'aide des scripts Praat[274].

Sur les figures 10.4-10.6 sont présentées les distributions des pourcentages des prédictions correctes pour un, deux et trois paramètres selon les quatre conditions expérimentales analysées. Ces histogrammes permettent de constater que parmi les paramètres que nous avons inclus dans l'analyse, certains ne jouent aucun rôle dans la prédiction du phrasé. Autrement dit, la variabilité observée dans ces paramètres semble ne pas dépendre de la structuration prosodique du message, telle qu'elle se construit en perception. Nous allons donc limiter nos analyses ultérieures à des combinaisons optimales des paramètres, assurant une meilleure prédiction de la distribution des frontières.

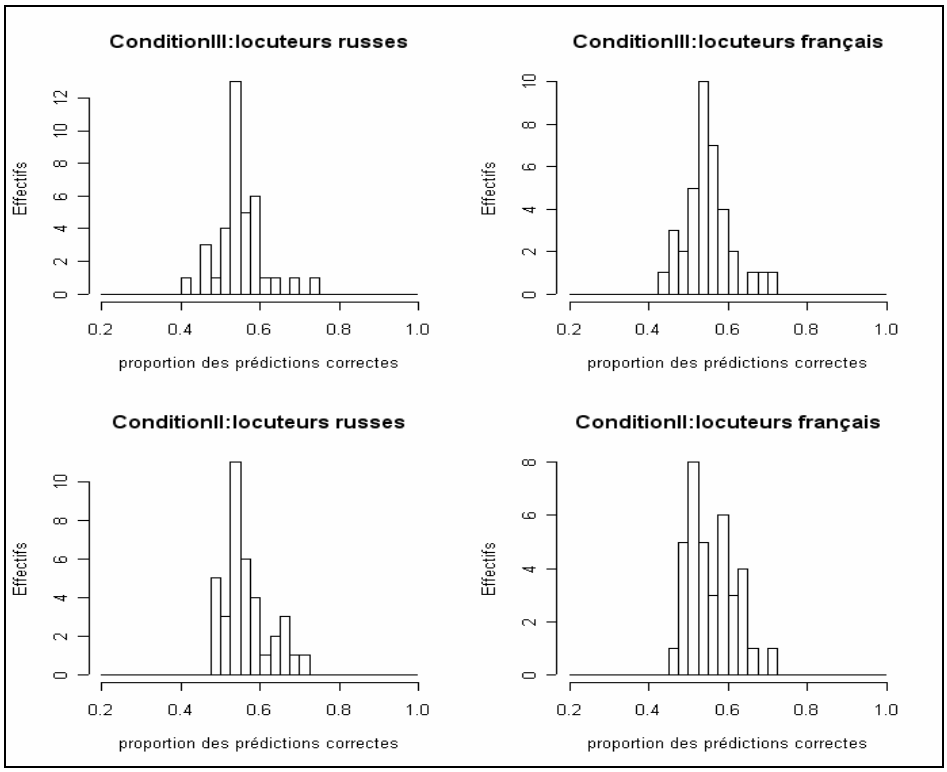


Figure 10.4 : Distributions des pourcentages des prédictions correctes : modèles à un paramètre

[274] Nous nous sommes inspirée des scripts de D. Hirst présentés dans Fernandez-Cruz (2000), avec les modifications requises, vu le caractère colinéaire de certaines de nos variables.

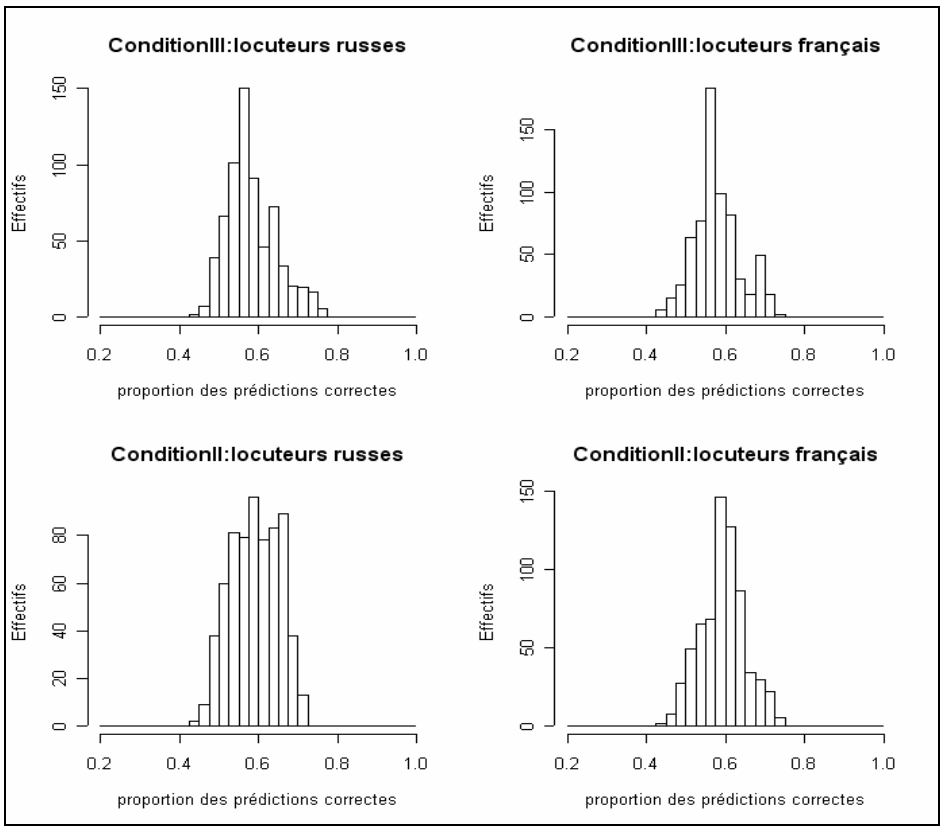


Figure 10.5 : Distributions des pourcentages des prédictions correctes modèles à deux paramètres

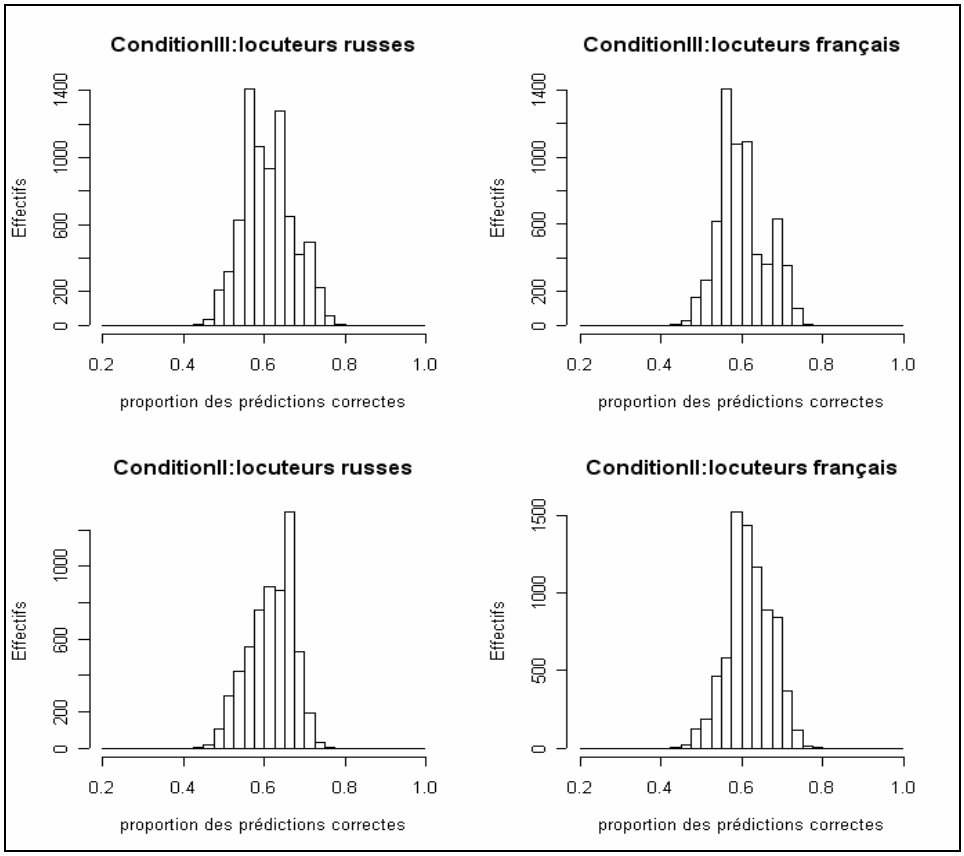


Figure 10.6 : Distributions des pourcentages des prédictions correctes modèles à trois paramètres

Nous retenons 3 paramètres, lesquels, pris isolement, assurent plus de 70% des discriminations correctes selon la condition expérimentale (Tableau X.6). Par rapport à ces résultats, nous soulignons le rôle important du paramètre de l'amplitude totale et de l'indice d'allongement.

LANGUE	CONDITION	PARAMÈTRE	POURCENTAGE CORRECT
Russe	CondN	*aamplitot*	73,4
	CondD	*aminmin*	72,09
Français	CondN	*zscore*	70,83
	CondD	*aamplitot*	71,11

Tableau X.6 : Paramètres individuels assurant une meilleure prédiction

Nous présentons dans le tableau X.7 ci-après les pourcentages de classification correcte en fonction de la condition expérimentale et du nombre de paramètres retenus. Les valeurs se situent au-dessus de 70%, avec un maximum de 77,78.

Langue	Condition	Nombre de paramètres	Pourcentage de classification correcte	Statistiques kappa pour le pourcentage de classification correcte	*F-measure* pour la catégorie "absence de frontière"	*F-measure* la catégorie "présence de frontière"
Russe	III	1	73,40	46,81	0,747	0,718
		2	75,53	51,06	0,758	0,753
		3	77,66	55,32	0,784	0,769
	II	1	72.09	44.19	0,714	0,736
		2	72.09	44.19	0,714	0,736
		3	76.74	53.48	0,767	0,767
Français	III	1	70.83	41.67	0,750	0,650
		2	71.88	43.75	0,727	0,710
		3	76.04	52.08	0,768	0,753
	II	1	71.11	42.22	0,724	0,698
		2	73.33	46.67	0,739	0,727
		3	77.78	55.56	0,787	0,767

Tableau X.7 : Évaluation de la démarche prédictive pour les modèles avec trois paramètres

Afin d'évaluer la performance du modèle, dans la lignée classique, nous avons procédé au calcul des mesures de rappel, précision et *f-measure*. Ces valeurs ont été calculées par catégorie (« absence » versus « présence »), et les statistiques de f-measure sont présentées dans le tableau X.7. Nous insistons tout particulièrement sur des valeurs comparables observées pour deux catégories : comme nous l'avons déjà mentionné, très fréquemment, l'algorithme prédictif privilégie une des catégories (absence de frontière), ce que nous avons cherché à éviter.

Dans un deuxième temps, nous nous sommes référée aux statistiques de Kappa. L'effet de ce procédé dans l'évaluation de la performance du classificateur est similaire et nous corrigeons le pourcentage de prédictions correctes par rapport à la probabilié d'une

prédiction sans qu'un indice acoustique soit pris en compte. Nous connaissons un seul travail qui prend en compte cette statistique dans l'évaluation de la performance d'un classificateur (démarche prédictive), à savoir Kochansky *et al.* (2005). Les remarques suivantes peuvent être faites et nous observons :

a) des valeurs de Kappa comparables à celles que nous avons trouvées lors de l'analyse de la concordance dans la performance des auditeurs quand les informations sémantiques et syntaxiques leur ont été inaccessibles ;

b) une amélioration de la performance du système de 18 à 30 % quand le nombre d'indices acoustiques est augmenté de un à trois.

Comme nous l'avons dit auparavant, le principe sous-jacent à la technique de l'analyse discriminante est la recherche d'une combinaison linéaire optimale de paramètres qui assure une meilleure séparation entre les groupes. Il est donc possible de calculer ce paramètre dérivé et d'évaluer quelle part de sa variance est « expliquée » par le facteur du phrasé. On peut obtenir une indication sur la qualité prédictive de ce paramètre avec un modèle linéaire, où ce paramètre dérivé est la variable dépendante et le facteur « phrasé » – la variable indépendante. Le coefficient R^2 du modèle donne la fraction de variance du paramètre dérivé, expliquée par le facteur « phrasé » (il s'agit d'un facteur à deux niveaux, présence versus absence d'une frontière). Nous avons appliqué ce procédé aux meilleures combinaisons des trois paramètres dans l'analyse discriminante. Les résultats sont résumés dans le tableau X.8.

LANGUE	CONDITION	COMBINAISON DE PARAMÈTRES	R^2	p
Russe	CondN	*bmax+bmean+aampltot*	0,2008	3.54e-06
		bzscore+abmeanmean+abamplitot	0,1813	1.12e-05
		abmeanmax+abmeanmean+aamplitot	0,2031	3.07e-06
	CondD	*bzscore+aminmin+abminmin*	0,198	1.05e-05
Français	CondN	*aslope+absdsd+aamplitot*	0,2682	3.91e-08
	CondD	*min+ameanmean+aamplitot*	0,2318	9.29e-07
		bmin+absdsd+aamplitot	0,2067	4.02e-06

Tableau X.8 : Lien entre la variable discriminante dérivée et le facteur 'Frontière'

Pour l'interprétation des résultats résumés dans le tableau X.8, nous pouvons faire les remarques suivantes. Au total, 13 paramètres acoustiques parmi 37 reviennent dans la démarche prédictive. Nous constatons également que toutes les analyses sont très significatives et, en particulier, que le phrasé prosodique exerce une influence sur des variables acoustiques dérivées. Le facteur du phrasé explique jusqu'à 27% de la variabilité de la variable dépendante. Simultanément, des valeurs légèrement plus élevées caractérisent les modèles fondés sur la performance d'auditeurs français. Les différences entre les conditions expérimentales et entre deux groupes d'auditeurs s'expliquent en partie par le fait que la localisation des frontières varie d'une condition à l'autre. Quant aux cas où plusieurs modèles atteignent un même pourcetage de prédictions correctes, nous pouvons interpréter ce résultat comme un indicateur du potentiel vers l'amélioration de la performance du modèle.

Remarquons également que les modèles acoustiques (les combinaisons d'indices retenues par l'analyse discriminante) nous paraissent pertinents d'un point de vue linguistique. Prenons comme exemple la combinaison d'indices modélisant la performance d'auditeurs francophones sous la troisième condition expérimentale (CondN). Les indices retenus sont la vitesse du changement mélodique dans l'empan pré-frontalier, la différence en variabilité dans les zones pré- et post-frontalières et l'amplitude totale avant la frontière. En effet, la vitesse du changement mélodique associée au phénomène du ton de frontière, couplée avec l'amplitude totale, nous suggère le marquage d'une frontière après une syllabe porteuse de la proéminence mélodique (accent nucléaire et/ou focus), une structure en accord avec les analyses théoriques de Di Cristo et Hirst (1996) et Gusenhoven (1990). Par ailleurs, il est établi que l'on observe une variation moindre de f0 du matériau linguistique juste après la frontière prosodique. Dans le même esprit, la combinaison des paramètres *bzscore+abmeanmean+abamplitot,* qui rend une meilleure prédiction des réponses d'auditeurs russes sous la condition CondN, prend en compte le phénomène d'allongement final (*bzscore*), celui de *resetting* (via indice *abmeanmean* qui permet la comparaison du niveau tonal avant et après la frontière) ; enfin le paramètre *abamplitot* rend compte de la saillance perceptive du matériau avant et après la frontière. Nous considérons que ces résultats fournissent une preuve de ce que la variable dérivée construite par le traitement interne de l'analyse discriminante a une signification linguistique et est reliée au facteur du phrasé.

Les dernières analyses discutées nous rassurent sur la portée linguistique des modèles acoustiques résultants. Elles nous autorisent à poursuivre les évaluations dans une perspective de prédiction du phrasé, et notamment pour la totalité de notre corpus.

10.3.2. Évaluation en prédiction

Dans un deuxième temps, nous avons choisi de tester en prédiction sur la totalité des données les modèles les plus performants de l'analyse discriminante. En effet, la fonction discriminante est entraînée sur les données issues de l'étude expérimentale en perception. Il est utile de tester les capacités prédictives de ce modèle sur la totalité de notre corpus et de comparer les résultats avec les deux annotations du phrasé prosodique dont nous disposons. Ceci en plus de nous renseigner sur les avantages de la démarche, nous permet également de mettre en évidence les sources potentielles de la variabilité qui n'ont pas été prises en compte lors de la sélection d'indices acoustiques afin de pouvoir les inclure dans une analyse ultérieure.

Pour mener à terme ce projet, nous avons fait agir la fonction *predict* dans le sens de la discrimination prédictive sur la totalité de notre corpus[275]. Nous récupérons à la sortie un fichier avec des étiquettes de catégories « présence » versus « absence » d'une frontière prosodique après la syllabe. Un script supplémentaire de PRAAT nous permet ensuite de transformer le tableau de catégories en une ligne d'annotation dans le TextGrid.

[275] Notre unité d'analyse est toujours l'unité interpausale : pour toute unité interpausale et toute syllabe, nous avons extrait à l'aide de scripts PRAAT les mêmes indices acoustiques bruts que ceux que nous avons présentés dans la section 10.2. Ces données sont résumées dans le tableau « *données_prédiction.txt* ».

Nous avons alors entrepris une évaluation qualitative de la performance du modèle sans pour autant procéder à une évaluation quantitative stricte pour les raisons suivantes. Une démarche évaluative envisageable consiste à comparer l'annotation résultant de l'application du modèle acoustique à celle proposée par des experts humains. Or, comme nous l'avons souligné, nous ne disposons pas d'une annotation de référence. D'une part, nous avons constaté des désaccords dans les annotations proposées par l'expert français et des experts russes. D'autre part, l'annotation des experts russes s'inscrivant dans la tradition de l'école russe, dans laquelle la structure syntaxique domine l'organisation prosodique, nous ne pouvons pas non plus y voir une annotation de référence par rapport aux présupposés de la phonologie prosodique. N'oublions pas non plus une éventualité de désaccord entre les annotateurs vis-à-vis de phrasé perçu. En guise d'exemple, nous avons choisi une unité interpausale

« *в кино ой ты знаешь я вообще-то не очень люблю ходить в кино* »

(aller) au cinéma ? ah tu sais moi en général j'aime pas beaucoup aller au cinéma

Nous proposons de comparer le phrasé prosodique tel qu'il a été annoté par les experts français et russes, mais également son annotation dans l'expérience en perception (cf. figure 10.7).

Nous constatons que la première frontière est consensuelle dans les annotations des experts, elle est prédite par notre classificateur acoustique ; dans l'expérience en perception cette frontière est annotée à l'unanimité par des auditeurs russes et français, à une exception près. La deuxième frontière est présente dans toutes les annotations et notre algorithme la prédit également. Or, les avis sont partagés en ce qui concerne la troisième frontière : elle est absente de l'annotation d'experts russes, tandis que l'expert français et les participants francophones de l'expérience en perception la signalent. Nous supposons que la perception d'une rupture est induite par la présence d'une proéminence mélodique focale sur le mot « *не очень* », ce qui constitue une analyse légitime du point de vue théorique. Qui plus est, la même frontière prosodique est signalée par un des auditeurs russes. Or, remarquons que sa localisation diffère dans les annotations : la frontière suit le mot-porteur de la proéminence mélodique « *не очень* » dans l'annotation issue de l'expérience en perception, mais elle est marquée après le verbe « *люблю* » chez l'expert français[276]. La création d'une annotation de référence apparaît ainsi comme une tâche d'une ampleur certaine et dépasse le cadre de notre travail. Nous avons en conséquence préféré nous limiter à l'analyse des erreurs de détections les plus marquantes, avec l'objectif d'affiner le modèle par la suite.

[276] Notons qu'une deuxième frontière apparaît chez deux auditeurs français.

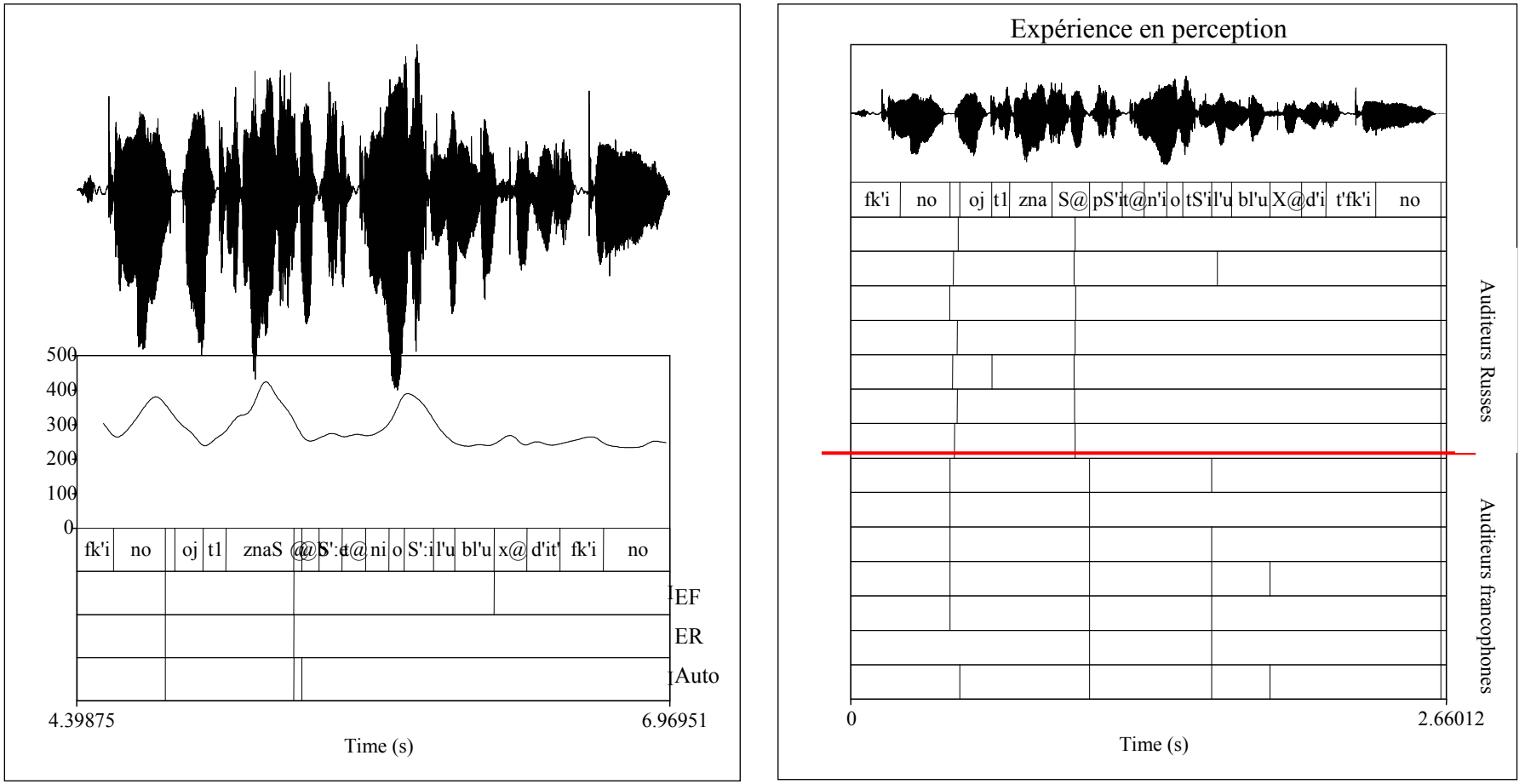


Figure 10.7 : Exemple de variabilité dans l'annotation du phrasé prosodique par l'expert français (EF), des experts russes (ER), des auditeurs russes et francophones dans l'expérience de perception, et par l'application de la procédure d'annotation automatique

Pour cette évaluation approximative de la performance de la classification, nous avons recensé pour chaque modèle les effectifs de deux classes (présence *versus* absence) et leur ratio (tableau X.9).

LANGUE	CONDITION	MODÈLE	EFFECTIFS « présence »	EFFECTIFS « absence »	RATIO
Russe	CondN	*bmax+bmean+aampltot*	726	1446	0,502
		bzscore+abmeanmean+abamplitot	859	1313	0,631
		abmeanmax+abmeanmean+aamplitot	748	1424	0,525
	CondD	*bzscore+aminmin+abminmin*	841	1331	0,632
Français	CondN	*aslope+absdsd+aamplitot*	581	1591	0,365
	CondD	*min+ameanmean+aamplitot*	763	1409	0,542
		bmin+absdsd+aamplitot	785	1387	0,566

Tableau X.9 : Effectifs de deux catégories « présence » versus *« absence d'une frontière prosodique » : analyse en prédiction sur la totalité du corpus*

Ces données nous permettent déjà de présager d'une tendance à la surprédiction de frontières. Certes, nous l'avions déjà constaté chez les auditeurs dans le test de perception, et la modélisation acoustique se fonde sur ces données. Or, cette interprétation se révèle parfois simpliste. Le modèle fondé sur la performance d'auditeurs français avec la présentation normale des stimuli (Condition III) obtient le meilleur ratio entre les catégories présence *versus* absence de frontière ; par conséquent, nous avons choisi d'analyser plus en détails l'annotation résultant de l'application de ce modèle (les fichiers d'annotation qui en résultent sont présentés dans le dossier 'Prédiction' du CD-room).

Les paramètres pris en compte sont la vitesse de changement de f0 dans la zone préfrontalière, la différence entre les écarts types des zones pré- et post-frontalières et l'amplitude totale de la syllabe précédant immédiatement la frontière. Nous présentons sur la figure 10.8 l'exemple d'une unité interpausale avec trois lignes correspondant respectivement au phrasé prosodique prédit pat expert français (ligne 1), au phrasé annoté par des experts russes (ligne 2) et au phrasé prédit par notre modèle à base d'indices acoustique (ligne 3) ; l'analyse porte sur l'énoncé « Да я больше люблю дома » (« *Oui, j'aime plus à la maison* »). Cet exemple nous permet de retenir les points suivants :

1) Nous avons discuté de la perception du phrasé prosodique pour des structures comportant des pro-phrases. Nous constatons en particulier un désaccord dans l'annotation entre les experts français et russes (certes, la segmentation en mots n'étant pas accessible à l'expert français). La frontière est prédite après le pronom personnel, ce dernier ayant une forme atone et réduite. Nous pouvons supposer que le réajustement du niveau ne s'opère qu'avec la syllabe portant un accent lexical : de ce fait, la localisation de la frontière prosodique change. Le modèle acoustique prédit une frontière après le pronom personnel en accord avec l'annotation de l'expert français.

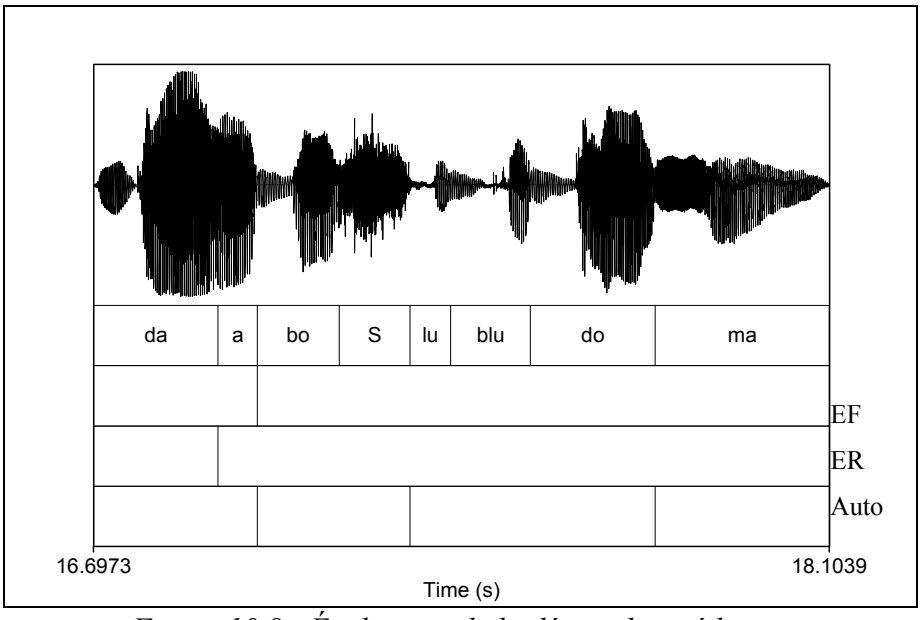


Figure 10.8 : Évaluation de la démarche prédictive

2) Deux autres frontières prédites sont absentes des annotations des experts. La frontière prosodique après la syllabe porteuse de la proéminence mélodique /do/ d'une part corrobore la tendance que nous avons constatée chez les auditeurs francophones d'associer systématiquement une proéminence mélodique et un groupement. Nous avons argumenté en faveur des modélisations théoriques rendant compte de ce lien (Gussnehoven 1990, Di Cristo et Hirst 1996). Cette tendance est légitime dans certains contextes (cf. figure 10.9), pourtant dans le cas présent, elle mène à une surprédiction. D'autre part, cette surprédiction résulte directement du choix que nous avons fait de nous positionner dans une perspective d'analyse prosodique non-lexicale (*non-lexical prosodic analysis*, Aylett 2005) qui implique de ne pas prendre en compte la structuration du flux de parole en mots. Si les frontières des mots étaient disponibles, modélisables en termes d'indices acoustiques, la frontière induite par la présence de la proéminence mélodique correspondrait à la frontière de l'unité interpausale (frontière du mot). Il paraît pourtant peu probable qu'une prédiction des frontières de mots soit possible à partir des seuls indices acoustiques (cf. travaux de Kassevitch 2002) ; ou, du moins, une vaste recherche préalable sur de la parole spontanée est nécessaire.

3) En revanche, la frontière après le mot « *больше* » (« *plus* ») est un artefact dû au choix de paramètres mélodiques dans le modèle. Lors de la discussion du travail d'alignement du corpus au niveau de phonèmes, nous avons mentionné, comme une des particularités de la parole spontanée, les cas observées de chutes des voyelles atones qui n'induisent pas pour autant une resyllabation. Nous pouvons même parler d'un léger allongement compensatoire de la consonne dans ce contexte (confirmé par l'indice de z-score de 0,934, la durée de la syllabe étant de 136 ms). Nous associons la surprédiction de la frontière prosodique après /bol'S@/ à l'effet de l'indice d'amplitude totale. D'une part, nous n'avons effectué aucune normalisation pour les valeurs d'amplitude relevées[277] ; de même, l'amplitude totale étant le produit de l'intensité par la durée, une surprédiction de frontières est alors fort envisageable.

[277] Voir Di Cristo 1985 sur la normalisation de l'intensité.

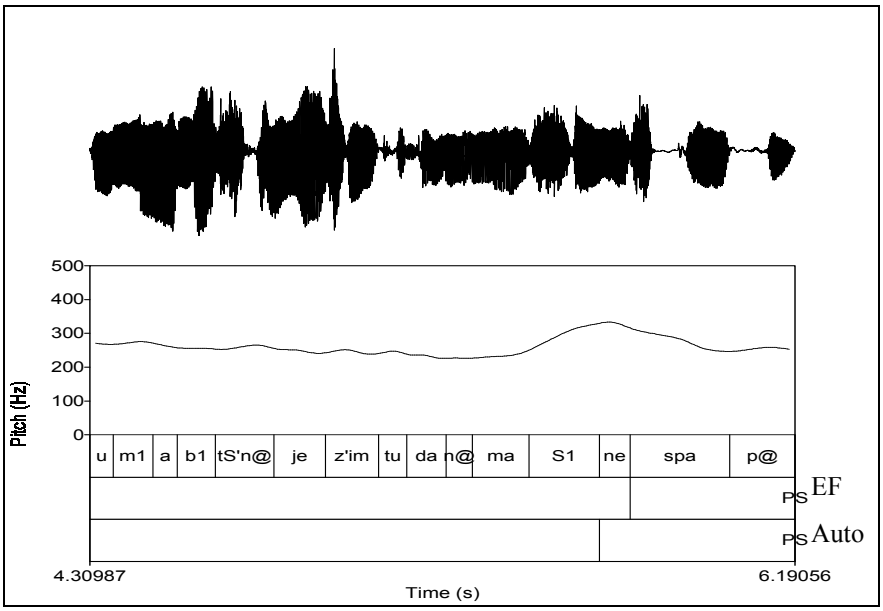


Figure 10.9 : Segmentation en unités intonatives avec une frontière après le mot porteur de l'accent mélodique

La prise en comte de l'amplitude totale est également « responsable » des cas où deux syllabes adjacentes reçoivent chacune l'étiquette de présence d'une frontière prosodique. Comme nous avons choisi un empan pré-frontalier de deux syllabes, l'une d'entre elles, celle qui est porteuse de la proéminence mélodique, ou de la structure complexe (en nombre de segments), exercera une double influence, pour deux positions dans la chaîne syllabique. Une autre interprétation possible de ce résultat réside dans le fait que nous n'avons pas tenu compte de l'influence de la vitesse d'élocution. Nous avons évoqué le débit de parole comme étant un des facteurs régissant le phrasé prosodique (cf. le chapitre V, consacré à la phonétique du phrasé prosodique). Un codage symbolique de l'allongement (cf. Auran et Di Cristo 2003) pourrait également constituer une solution. Il se peut également que, plusieurs syllabes consécutives étant annotées pour la présence de la frontière prosodique, celle-ci se trouve alignée avec la frontière du mot. Une nouvelle perspective de recherche se présente, laquelle concerne la quête d'indices acoustiques aux frontières des mots.

Une autre source de surprédiction, toujours en lien étroit avec l'indice de l'amplitude totale, est illustrée par l'exemple de la figure 10.10 qui reprend l'annotation pour le fragment « *Я получила их в мае* » (*Je les ai eus au mois de mai*). L'unité interpausale ayant une durée de 1,41 seconde, l'expert français atteste une réalisation en une seule unité intonative. Le modèle acoustique prédit la présence d'une frontière supplémentaire : notons, que l'extrait en question est prononcé avec un ralentissement dans le débit de parole qui accompagne l'hésitation de la locutrice (la grille de la figure 10.10 comporte une deuxième ligne où sont présentés des coefficients de z-score, en guise d'illustration) ; il en résulte que cet allongement contribue à la valeur de l'amplitude totale. En effet, le rythme de la parole spontanée est une structure beaucoup moins régulière, les accélérations et les ralentissements, voire les hésitations, s'y enchaînent reflétant ainsi le processus sous-jacent de formulation ainsi qu'une dominance du contenu sur la forme[278]. Il est pertinent de

[278] Nous ne tenons pas compte du style rhétorique.

s'interroger sur les rapports que maintiennent les unités de performance et les unités intonatives, ces dernières faisant partie de l'organisation structurelle du message (cf. *commit & restore model* Clark et Wasow 1998). Dans le travail de Svetozarova *et al.* (1988), toute unité de performance était considérée comme étant une unité intonative : les auteurs reconnaissent une opposition entre les unités structurelles, tels les syntagmes prosodiques, et les unités de performance signalées par des phénomènes de dysfluence, or, ils constatent que cette opposition est nivelée en parole spontanée (p. 147). En revanche, Hansson (2003) prend en considération les unités de performance lors de l'élaboration de la hiérarchie des contraintes gérant le phrasé prosodique et observe que deux contraintes d'alignement sont nécessaires pour rendre compte des phénomènes de dysfluence : la contrainte AlignXP,L (alignement de toute structure XP avec une frontière à sa gauche) ayant pour fonction le marquage de la frontière gauche de l'unité intonative et la contrainte AlignXP,R. La première contrainte s'appliquera à toutes les unités (unités intonatives et unités de performance), dans la mesure où cette contrainte est associée avec le phénomène de réajustement du niveau. Or, la contrainte AlignXP, R ne s'appliquera qu'à des unités de l'organisation structurelle du message (les unités intonatives), étant associée à la présence conjointe des indices acoustiques d'allongement et de ton de frontière. Cependant, l'application de ce critère nous paraît problématique dans le cas d'un « allongement de dysfluence » sans recours aux informations sur la structure sémantico-syntaxique de l'énoncé. De même, Aylett (2005) réfléchit à un modèle acoustique permettant de distinguer les deux ; le modèle par analyse discriminante ayant une performance réduite, l'auteur s'interroge sur le rôle de la forte variabilité des stratégies individuelles. Pour notre part, nous considérons qu'une analyse approfondie des rapports entre la structure prosodique et celle des unités de performance en parole spontanée est de rigueur, en vue d'applications technologiques (synthèse de la parole spontanée, analyse automatique des corpus, cf. également Strangert 2004).

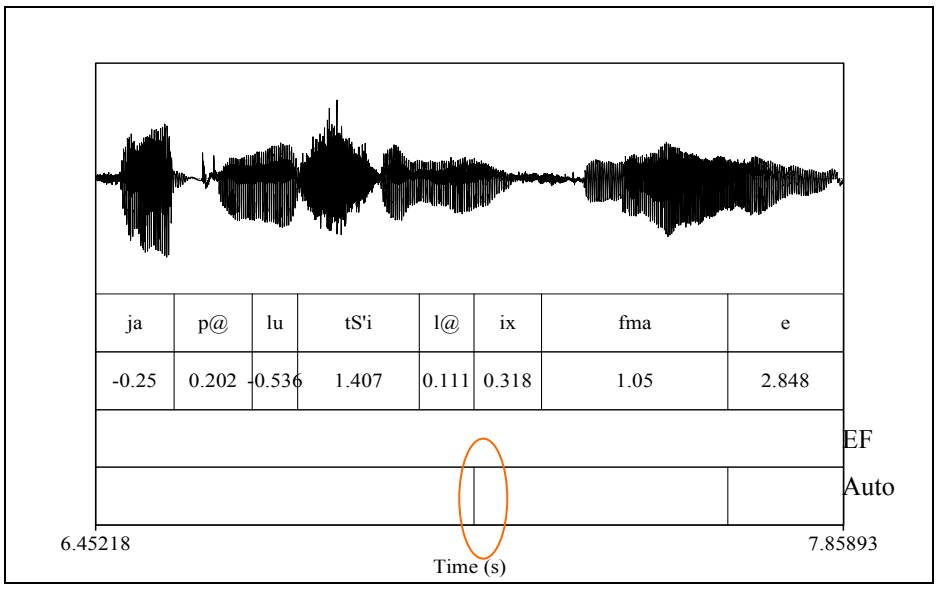


Figure 10.10 : Exemple d'une surprédiction des frontières prosodiques sous l'effet du rythme variable de la parole spontanée

10.3.3. Totalité des données

Nous avons insisté dans la section précédente sur la tendance du modèle acoustique à la surprédiction du phrasé prosodique. Nous voulons ensuite nous assurer que la performance du modèle n'a pas été biaisé par le fait que lors de la recherche d'une modélisation acoustique, nous n'ayons pris en considération qu'une sous-partie d'observation pour la catégorie « absence de frontière »[279]. Il se peut que notre choix empêche le modèle de capturer la multitude des contextes prosodiques (acoustiques) pour la catégorie absence de frontière. Nous avons donc refait des analyses discriminantes sur les données non-équilibrées (corpus de l'expérience). Nous nous sommes concentrée sur les agrégations de trois indices les plus performantes dans la discrimination entre les deux catégories. Nous résumons les résultats obtenus dans les tableaux X.10 et X.11.

Langue	Condition	Pourcentage de classification correcte	Pourcentage Kappa de classification correct	*F-measure* pour la catégorie "absence de frontière"	*F-measure* la catégorie "présence de frontière"
Russe	CondN	81,8	37,96	0,8902	0,4706
	CondD	80	41,3	0,8739	0,5161
Français	CondN	80,56	53,33	0,8627	0,6667
	CondD	79,14	45,75	0,8612	0,5797

Tableau X.10 : Évaluation de la démarche prédictive pour les modèles avec trois paramètres : totalité des positions « absence de frontières » incluse dans l'analyse

Nous constatons tout premièrement une légère augmentation du pourcentage des prédictions correctes. Cependant, elle n'est pas toujours accompagnée de l'augmentation pour les statistiques Kappa : nous constatons même une dégradation pour cette mesure. Le coefficient de Kappa le plus faible caractérise le modèle acoustique du phrasé prosodique observé dans la performance des auditeurs russes sous la condition CondN (présentation normale des stimuli). Ce résultat corrobore les observations que nous avons faites auparavant sur le rôle de la structure syntaxique dans les réponses recueillies. Nous observons également que les valeurs de *f-measure* diffèrent en fonction de la catégorie, celles de la catégorie « absence » étant assez élevées, et presque doubles de celles de la catégorie « présence ». De même, la valeur de f-measure pour la catégorie présence tient surtout à des valeurs élevées de la mesure de précision, c'est-à-dire du faible pourcentage de cas récupérés de façon erronnée. Or, le rappel est inférieur à 50% pour tous les modèles acoustiques sauf celui de la condition CondN chez les auditeurs francophones.

De même, nous avons évalué le pourcentage de la variabilité dans la variable dérivée, expliquée par appartenance à une des deux catégories (tableau X.11).

[279] Nous avons argumenté notre choix de disposer d'un ensemble de données équilibré en ce qui concerne les effectifs des deux catégories, puisque la tendance de ce type d'algorithmes est de privilégier la catégorie la plus présente.

LANGUE	CONDITION	COMBINAISON DE PARAMÈTRES	R^2	p
Russe	CondN	*zscore+max+aampltot*	0.1738	6.03e-10
	CondD	*zscore+bsd+abminmean*	0.1769	5.04e-08
Français	CondN	*bmax+bslope+aamplitot*	0.2742	9.91e-12
		bmax+bamplitot+abamplitot	0.2943	1.31e-12
		bsd+ bamplitot+abamplitot	0.2963	1.07e-12
	CondD	*zscore+asd+aamplitot*	0.2383	6.52e-10
		amin+amax+aamplitot	0.1788	1.29e-07

Tableau X.11 : Lien entre la variable discriminante dérivée et le facteur 'Frontière' : totalité des positions « absence de frontières » incluse dans l'analyse

Nous avons testé en prédiction un des modèles sur la totalité du corpus (*bsd+ bamplitot+abamplitot*). Nous obtenons ainsi un meilleur ratio entre deux catégories d'étiquettes : ce résultat peut être dû à la fois à la diminution de la surprédiction et à l'augmentation des cas de frontières non-prédites.

En effet, pour les résultats obtenus en prédiction (les TextGrids sont contenus dans le dossier 'Prédiction' de CD-room), nous constatons deux cas de figure. D'abord, nous constatons une amélioration de la performance de l'algorithme dans certains contextes. Nous attirons surtout l'attention sur une influence moindre des dysfluences sur la performance du modèle : comme l'exemple de la figure 10.11 l'illustre, pour un fragment « *ну там была буря* » (*bah, là-bas, c'était un orage*), le modèle acoustique fondé sur les données équilibrées prédisait la présence de deux frontières prosodiques (frontière après chaque mot, ligne 2). À l'écoute, nous constatons que ce fragment de discours est réalisé avec une certaine incertitude, résultant du travail de formulation (notons également la présence d'une pause après le verbe). De ce fait, les frontières prédites ne sont pas des frontières structurelles, mais plutôt des ruptures associées au travail de formulation. Le deuxième modèle issu de la totalité des données sur les cas d'absence de frontières permet de pallier cet effet.

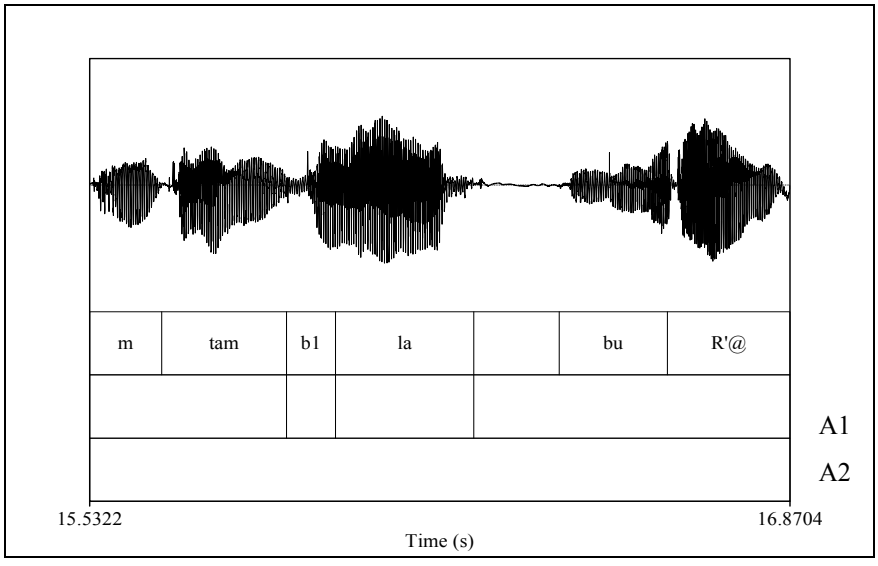


Figure 10.11 : Comparaison des annotations par deux modèles acoustiques : cas de dysfluences. (Légende : A1 : modèle équilibré ; A2 : modèle avec la totalité des positions « absence de frontières » incluse dans l'analyse)

En revanche, nous constatons des absences de prédiction de frontières à l'encontre du premier modèle, obtenu à partir des données équilibrées. C'est le cas notamment de l'extrait de la figure 10.12 : il s'agit d'une frontière après le syntagme nominal en fonction du sujet de l'énoncé, la frontière étant prédite par le modèle équilibré mais non pas par le modèle fondé sur la totalité des données[280].

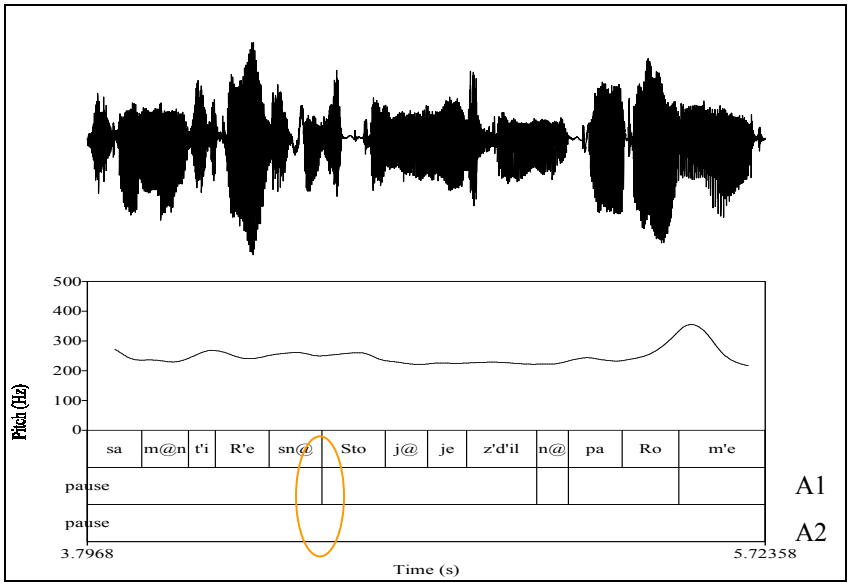


Figure 10.12 : Comparaison des annotations par deux modèles acoustiques

Une source de la surprédiction persistante provient du fait que nous n'avons pas précisé le statut de l'unité que nous prédisons. Nous avons évoqué le statut incertain du syntagme de l'école russe à la fois dans la discussion théorique sur le phrasé (chapitre IV) et dans l'analyse des résultats de l'expérience en perception. S'agit-il d'une unité comparable au syntagme phonologique de la phonologie des domaines (certains travaux le laissent supposer) ou d'un véritable syntagme intonatif ? Parfois les unités prédites sont plus petites que le syntagme intonatif, mais reprennent la structuration au niveau des syntagmes phonologiques. Par exemple, nous pouvons attribuer au fragment « *кстати ты смотришь чемпионат мира по футболу* » (« *à propos, tu regardes le mondial de foot* ») une structure syntaxique représentée sur la figure 10.13.

La figure 10.13 fait apparaître que la structuration prosodique en syntagmes phonologiques issue de la structure syntaxique a la forme de (x), une structure qui correspond au phrasé prédit par le modèle acoustique total :

(x) [кстати]$_{PPh}$ [ты смотришь] $_{PPh}$ [чемпионат мира] $_{PPh}$ [по футболу] $_{PPh}$

[280] Cette frontière étant également absente de l'annotation proposée par l'expert français.

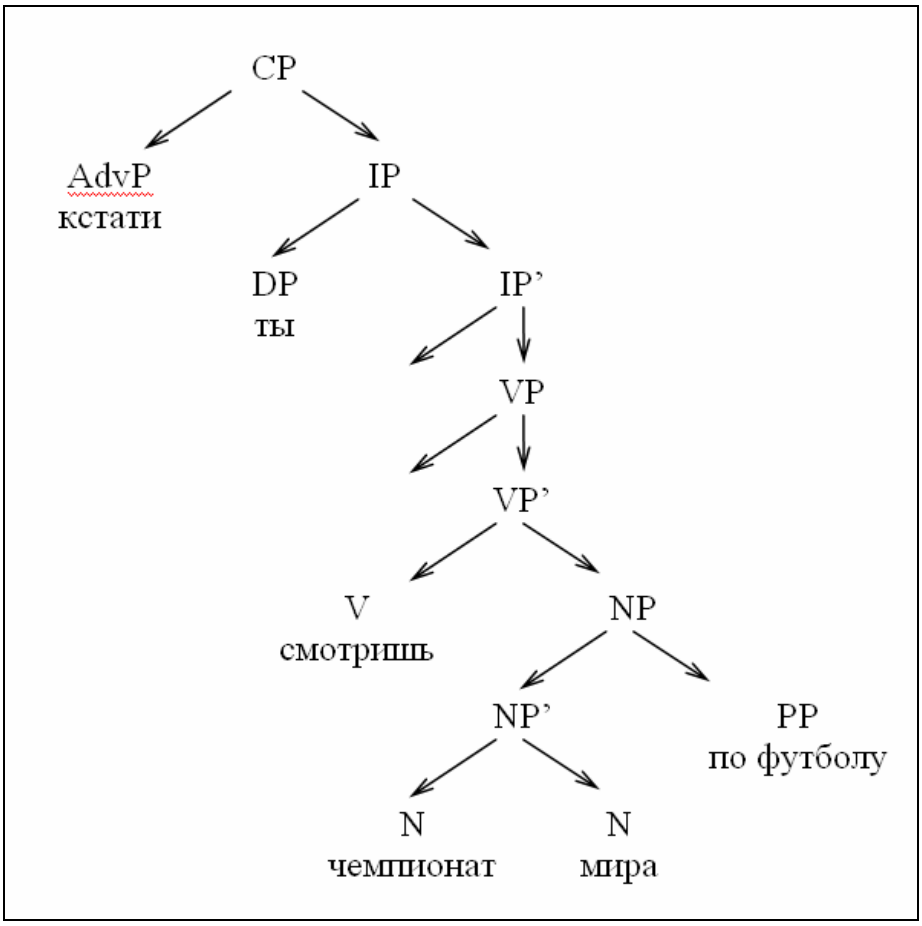


Figure 10.13 : Structure syntaxique de l'énoncé « кстати ты смотришь чемпионат мира по футболу » (« à propos, tu regardes le mondial de foot »)

Si, Gussenhoven (1999), en présentant ToDI, insiste sur le fait que le constituant du syntagme phonologique n'a pas de marquage intonatif en hollandais, d'autres auteurs (Jun et Fougeron, Di Cristo et Hirst, Pierrehumbert), et pour d'autres langues, définissent deux niveaux et plus, contribuant à l'organisation tonale du message. Dans la tradition russe on maintient une distinction entre un syntagme et une phrase (synonyme d'énoncé prosodique) : la question du statut du syntagme par référence à la hiérarchie prosodique est donc toujours ouverte. Par conséquent, afin d'améliorer la performance du modèle acoustique en prédiction, il nous semble nécessaire de clarifier la question des niveaux structurels dotés d'un marquage tonal dans la hiérarchie des constituants prosodiques du russe dans des études expérimentales ultérieures.

10.4. DISCUSSION ET CONCLUSIONS

Le présent chapitre a été consacré à la discussion des résultats d'une modélisation acoustique des frontières prosodiques. Cette modélisation est premièrement fondée sur les résultats de l'étude en perception. À partir du corpus de test, nous avons cherché les indices acoustiques – corrélats du phrasé prosodique perçu. Nous avons construit un classificateur permettant de déterminer pour toute position syllabique si elle est le site

d'une frontière intonative ; ce classificateur a été entraîné sur le corpus de test pour ensuite être évalué sur la totalité de notre corpus.

Le classificateur a été entraîné sur deux types de données : les données équilibrées, contenant les mêmes proportions de cas pour deux niveaux du facteur et les données déséquilibrées, avec toutes les positions de l'absence de frontière prises en compte. Les données équilibrées nous permettent d'obtenir des valeurs de f-measure comparables pour deux catégories ; cependant, ce choix méthodologique aboutit à une forte surprédiction de frontières. Les données déséquilibrées permettent au classificateur d'appréhender une gamme élargie de contextes pour l'absence de frontière. Nous constatons pourtant une meilleure performance pour la catégorie « absence de frontière prosodique », bien que nous observions un abaissement dans la performance de classificateur pour la catégorie « présence de frontière prosodique » (ces observations sont fondées sur des valeurs de f-measure). L'abaissement est surtout dû à une valeur faible de rappel. Rappelons que ce paramètre caractérise et quantifie la proportion de cas de prédiction correcte sur le total des cas pertinents. Plusieurs positions associées avec une frontière prosodique se retrouvent dans un contexte acoustique non saillant. Une autre interprétation possible de ce fait consiste à dire que les contextes acoustiques ne sont pas en rapports bi-univoques avec le facteur de la présence d'une frontière prosodique. Nous pouvons y voir l'influence des facteurs de haut niveau (facteurs rythmiques et/ou discursifs). Dans cette perspective notre étude n'a qu'un caractère exploratoire.

En effet, nous considérons que la modélisation acoustique des frontières nécessite une prise en considération des contextes discursifs. L'annotation du corpus que nous avons effectuée (et dont nous avons détaillé la démarche dans le chapitre VII) ne concerne que la dimension phonologique. En effet, nous y spécifions certain(e)s domaines (unités), ainsi que les détails d'organisation tonale et métrique. Or, une annotation de la structure discursive peut davantage nous renseigner sur les particularités de l'organisation prosodique. Pour revenir aux cas des pro-phrases, question soulevée à plusieurs reprises, l'exemple de la figure 10.14 illustre une situation dans laquelle une frontière prosodique n'est pas prédite après la pro-phrase. S'agit-il d'une erreur de la prédiction, ou d'un effet de l'organisation discursive ? Lors de la discussion des pro-phrases dans les analyses entreprises dans la perspective *top-down*, nous avons présenté une argumentation selon laquelle la présence d'une frontière prosodique après la pro-phrase dépendra fortement du contexte et de l'intention discursive du locuteur (s'il veut étayer ces propos par une argumentation développée, par exemple). L'annotation de l'organisation discursive et la mise en relation systématique de cette structure avec le phrasé prosodique peuvent nous indiquer si la prédiction du classificateur est juste.

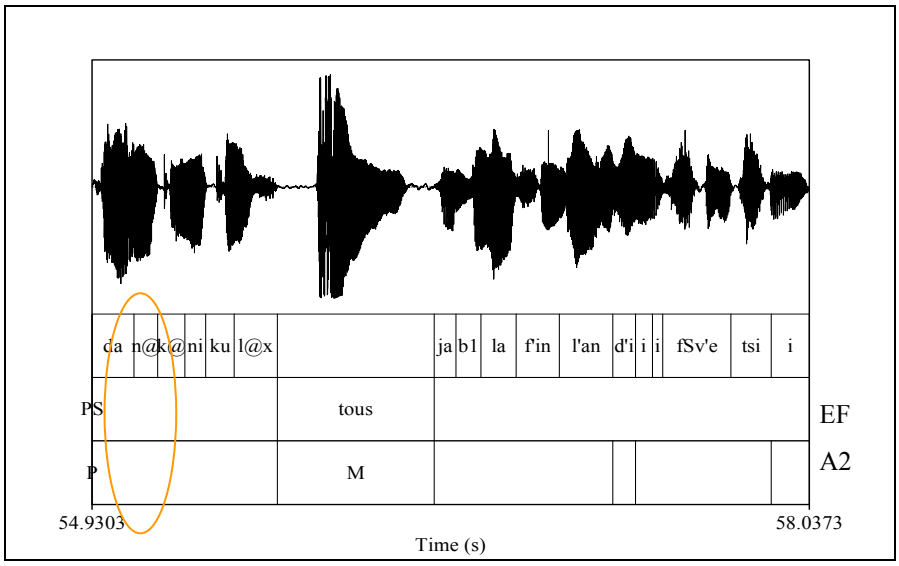


Figure 10.14 : Illustration de l'absence de frontière après une pro-phrase (« Да на каникулах я была в Финляндии и в Швеции », Oui en vacances j'ai visité la Finlande et la Suède) : s'agit-il d'une erreur de prédiction ?

Le vecteur des indices acoustiques que nous avons utilisés en prédiction nécessite lui aussi un amendement. Cela porte surtout sur la mesure d'amplitude totale. Nous avons motivé ce choix par un souci de rendre compte du lien entre les phénomènes du groupement et de la proéminence. Ce lien est explicité dans certaines hiérarchies de la constituance prosodique (cf. l'approche métrique), mais aussi prouvé dans les études psycholinguistiques portant sur le fonctionnement du mécanisme attentionnel et de la mémoire à court et à long terme. Cependant, la prise en compte de cet indice mène à des cas de surprédiction des frontières. Dans ce contexte, la parole spontanée démontre une sensibilité accrue : en effet, le signal acoustique contient plusieurs types de marques de spontanéité (hésitations, allongements et pauses non-structurels) dont les manifestations acoustiques sont similaires à celles du phrasé prosodique (notamment, pour les allongements). L'amplitude totale étant le produit de l'intensité par la durée, la mesure se révèle très sensible à la présence des dysfluences citées. D'autre part, une réflexion sur la normalisation de la mesure de l'intensité en fonction de l'identité des phonèmes est nécessaire (Di Cristo 1985). Rappelons que nous avons déjà souligné le fait que le paramètre d'intensité ainsi que celui de tilt spectrale n'ont pas reçu dans les études prosodiques la même attention que l'organisation tonale et temporelle. La situation commence à changer ces dernières années avec les études de Sluijter et van Heuven (1996), Heldner (2003), Kochansky et al. (2005).

Nous tenons aussi à attirer l'attention sur l'utilisation des statistiques Kappa que nous avons utilisées pour évaluer la performance du classificateur : cette mesure permet d'évaluer l'impact d'indices acoustiques sur l'efficacité de la classification. Nous jugeons cette mesure plus pertinente que le pourcentage des classifications correctes traditionnellement communiquée dans les études en prédiction.

La discussion entreprise nous permet de tracer les perspectives de recherches futures ; nous pouvons dégager deux pistes centrales : l'étude du rôle de l'intensité dans l'organisation prosodique du message (en particulier, dans la perspective de rendre explicite le lien entre

groupement et proéminence), et l'élaboration d'un cadre méthodologique permettant une analyse du phrasé prosodique en prenant en compte toutes les dimensions linguistiques[281].

[281] Cette deuxième piste paraît être moins associée à la problématique de l'annotation automatique du phrasé. Nous la jugeons tout de même importante, surtout à l'étape d'évaluation de la performance du système, mais également pour la réflexion menée lors de l'élaboration du vecteur de paramètres acoustiques.

CHAPITRE XI

ARTICULATION « FORME ~ FONCTION » : LE CAS DE LA PRIMITIVE FONCTIONNELLE « ACCENTUÉ »

11.1. INTRODUCTION

Ce dernier chapitre de notre travail aborde la problématique du lien possible entre les représentations formelle et fonctionnelle de la prosodie de l'oral spontané. Dans les chapitres initiaux, nous avons argumenté en faveur de l'intégration du niveau de représentation formelle dans l'architecture globale pour les faits prosodiques. Nous avons également motivé notre choix de recourir à l'algorithme INTSINT pour modéliser la dimension tonale de l'organisation prosodique. Nous avons tout particulièrement insisté sur la réversibilité entre les représentations proposées pour chaque niveau d'analyse : l'articulation entre le niveau phonétique et le niveau phonologique de surface est assurée par un lien intrinsèque entre les algorithmes MOMEL – INTSINT. Une autre interface a été stipulée pour certaines langues entre le niveau de la représentation phonologique de surface et le niveau phonologique profond : on peut citer à ce sujet l'analyse comparative du français et de l'anglais, la gestion de l'emphase, l'implémentation dans le système de synthèse SYNTAIX (Di Cristo & Di Cristo 2001)

Nous avons également débattu des critiques émises vis-à-vis de l'approche INTSINT. Nous avons insisté sur le fait qu'INTSINT maintient une distinction systématique entre le niveau phonologique de surface et le niveau phonologique profond. Nous avons également présenté des règles et des principes qui gouvernent la linéarisation de la représentation phonologique profonde. Il faut rappeler que l'exemple discuté portait sur de l'anglais et que le même procédé a été aussi appliqué au français. Nous considérons qu'une partie de ces critiques fondent leur légitimité sur le fait qu'INTSINT se limite à exposer des mécanismes dérivationnels ; une grammaire à états finis de l'intonation du français a été proposée (Hirst et al. 1991, cf. figure 11.1), sans pour autant être exploitée dans des recherches prosodiques postérieures.

Sur le schéma de la figure 1.3., qui reprend l'architecture globale de la dimension prosodique (*supra*, page 46, cf. discussion associée), une jonction s'opère au niveau phonologique profond entre la dimension formelle et fonctionnelle dans l'analyse des faits prosodiques. Comme nous l'avons déjà mentionné, Hirst (2005) propose de modéliser la dimension fonctionnelle en termes de primitives fonctionnelles, lesquelles se regroupent en deux pôles autour des concepts d'accent (proéminence) et de frontière.

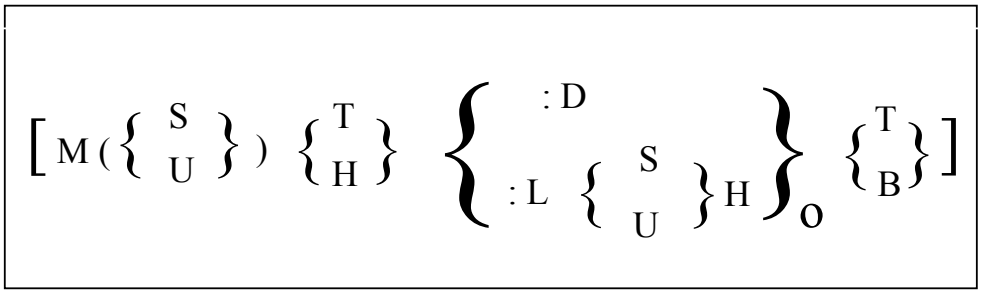


Figure 11.1 : Grammaire INTSINT pour le français (Hirst et al. 1991)

Nous proposons dans ce dernier chapitre une tentative d'analyse orientée vers l'élaboration d'une grammaire INTSINT. Nous réfléchissons sur l'éventualité d'une grammaire probabiliste, destinée à démontrer des régularités en ce qui concerne la façon dont les tons d'INTSINT se combinent. Dans la lignée des réflexions que nous avons menées sur l'interaction entre les représentations formelle et fonctionnelle, nous avons choisi de ne considérer qu'une seule catégorie de primitives prosodiques fonctionnelles, celle de l'accent mélodique. Précisons, pourtant, qu'à cette étape de l'analyse nous n'opérons pas de distinction selon les traits secondaires de la catégorie 'accentué', tels que 'emphatique' et 'nucléaire'.

11.1.1. Sur les approches probabilistes en linguistique

Le terme « *probabiliste* » peut inciter à des confusions : en effet, tout au long de ce travail nous avons pris position en faveur de l'importance des données et d'une approche bidirectionnelle qui intègre les deux démarches ascendante et descendante. Nous avons également manifesté une réticence vis-à-vis des méthodes stochastiques fondées sur des calculs statistiques à partir de données volumineuses.

Cependant, la réflexion que nous cherchons à mener ici est essentiellement fondée sur l'appréhension de la nature probabiliste des représentations et des processus cognitifs. Pierrehumbert (2001) souligne que la tradition générativiste qui prône l'étude de la compétence linguistique chez l'humain et relègue les analyses de la variabilité observée en production au domaine de la performance, déprécie la valeur des analyses probabilistes. Or, les études en psycholinguistique démontrent la nature probabiliste des représentations cognitives dont dispose un humain : les preuves proviennent des études, analysant l'apport des informations sur les fréquences dans l'acquisition du langage et dans la formation des représentations mentales à long terme.

Un des concepts clés de l'approche générativiste est celui de la grammaticalité et de la bonne formation. À partir d'un inventaire de catégories (traits, phonèmes, frontières etc.) et d'un ensemble de principes de leur combinaison, on procède à l'évaluation d'une séquence d'unités pour se prononcer sur son caractère licite dans la langue étudiée, pour émettre donc un jugement de bonne formation / de grammaticalité. Par contre, s'opposant au raisonnement traditionnel de la théorie générative, les approches probabilistes remettent en cause la notion de la grammaticalité des productions linguistiques observées. Les réalisations ou les structures prosodiques atypiques n'y sont pas écartées sous prétexte

d'être non-conformes à la grammaire sous-jacente ; on les considère comme rares, marquées[282], peu fréquentes. La démarche probabiliste est, par conséquent, cohérente avec notre souci de ne pas imposer aux données une structure à partir des seules prémisses théoriques, mais de déduire la structure à partir des données observables. Ce choix sous-tend toutes les analyses qui ont été menées dans le cadre du présent travail.

Les questions auxquelles nous cherchons de répondre portent donc sur les primitives descriptives, formelles et fonctionnelles au sein de l'analyse de l'intonation que nous essayons d'aborder dans une approche probabiliste.

11.2. MÉTHODOLOGIE

11.2.1. Unité d'analyse

L'analyse se fonde sur l'annotation symbolique des unités porteuses de l'accent mélodique à l'aide des catégories INTSINT. Par conséquent, un problème théorique se pose dès à présent : celui de l'unité d'analyse. Sur quelle unité fonder nos analyses ? En examinant les recherches précédentes, nous pouvons recenser des candidats potentiels qui sont :

a) syllabe, ayant le statut d'unité porteuse du ton dans l'approche autosegmentale (*tone-bearing unit*) ;

b) mot (phonologique, prosodique, ou phonétique de la tradition russe) ;

c) domaine d'association tonale (Gussenhoven 1990, 2005) ;

d) pied accentuel, similaire au pied métrique mais défini par la distribution de proéminences mélodiques (cf. IViE) ;

e) syntagme accentuel (cf. Jun et Fougeron 2000) ;

f) unité tonale de la tradition britannique.

Bien que la syllabe ait le statut d'unité porteuse du ton, il semble impossible d'y limiter l'analyse, non seulement en raison des nombreuses questions soulevées en ce qui concerne la problématique de l'alignement, mais aussi parce qu'un mouvement mélodique couvre souvent plusieurs syllabes adjacentes (notamment, s'il s'agit d'un mot poly-syllabique). D'autres unités citées nous semblent dépendantes des théories desquelles elles émanent, et ancrées dans la tradition d'analyse de l'intonation propre à la langue. Par conséquent, avant de fonder notre analyse sur ces unités, nous considérons comme une étape indispensable de démontrer leur pertinence pour l'analyse de l'intonation du russe dans une étude expérimentale. De ce fait, nous avons décidé de fonder notre analyse sur l'unité du mot phonétique (dans la tradition russe) / groupe clitique de Hayes (1989) et Nespor et Vogel (1986)/ mot prosodique (Astesano 2001).

Sur la totalité de notre corpus, 315 mots phonétiques ont été annotés comme portant un accent mélodique.

[282] Dans le sens de Troubetskoj (1939)

11.2.2. Outils statistiques

Deux concepts mathématiques d'intérêt sont à retenir : celui de la probabilité conditionnelle et celui d'entropie.

Les probabilités conditionnelles nous renseignent sur le lien qui existe entre deux événements, en l'occurrence, deux étiquettes de tons d'INTSINT. La probabilité conditionnelle (notée P(A|B)) quantifie la chance qu'un événement A se produise, étant donné que l'événement B a eu lieu. La formule de Bayes (1) est appliquée :

(1) $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$,

où P(B) est la probabilité de rencontrer la catégorie B, P(A∩B) désignant une probabilité mutuelle, c'est-à-dire, la probabilité que les deux événements aient lieu (la probabilité de retrouver le bigramme dans l'annotation). Les valeurs de la probabilité conditionnée reçoivent l'interprétation 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).

La question se pose de savoir comment évaluer la valeur de probabilité conditionnelle, la magnitude de l'effet de conditionnement ? Pour ce faire, nous avons calculé l'intervalle de confiance, nous permettant d'évaluer l'écart entre la probabilité conditionnelle p(A|B) et la probabilité propre d'événement p(A) : il s'agit da la mesure de ratio entre cet écart et l'erreur standard de l'estimation du nombre des occurrences simultanées de deux événements A et B :

(2) $$z = \frac{p(A|B) - p(A)}{\sigma}$$, où $\sigma \approx \sqrt{p(A \cap B)*(1-p(A \cap B)/N_B}$

Si les deux événements sont indépendants, z est centré sur 0 et possède un écart type de 1 ; par ailleurs, |z| > 2 nous indique la présence d'un lien suffisamment important (à 2 sigma) entre les deux événements, le signe de z indiquant, si l'événement B favorise ou défavorise la présence d'événement A.

Quant à l'entropie, c'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'un système avec **N** états distincts, chacun ayant la probabilité p_i, est définie par (3) :[2]

(3) $$H = \sum_N (p_i * \ln(p_i))$$

L'entropie d'une distribution varie entre 0, valeur correspondant à un système déterministe[283], et lnN pour un système dont les N états sont équiprobables. On appellera entropie normalisée la quantité H/lnN qui varie dans l'intervalle [0,1].

Nous avons entrepris deux types d'analyse : d'abord, nous avons procédé à une analyse par bigrammes, ensuite, toutes les combinaisons de deux, trois et quatre symboles consécutifs ont été analysées. Cependant, nous considérons qu'une réflexion approfondie est nécessaire sur la question du nombre de symboles par unité phonologique pertinente : cette réflexion pourrait être fondée dans un premier temps sur les principes qui sous-tendent le processus de la linéarisation de la représentation phonologique profonde.

11.3. RÉSULTATS : VERS UNE GRAMMAIRE PROBABILISTE DES CONTOURS INTSINT

Dans un premier temps, nous avons analysé les séquences d'étiquettes INTSINT, décrivant les courbes mélodiques associées avec les mots phonétiques proéminents (315 unités au total). La séquence la plus longue (une seule sur 315 unités annotées) comporte 6 symboles. Nous la représentons sur la figure 11.2. La syllabe centrale étant la syllabe métriquement forte, plusieurs points-cibles détectés (et ensuite codés) contribuent à modéliser les deux plateaux de la courbe (signalés par des flèches sur la figure 11.1).

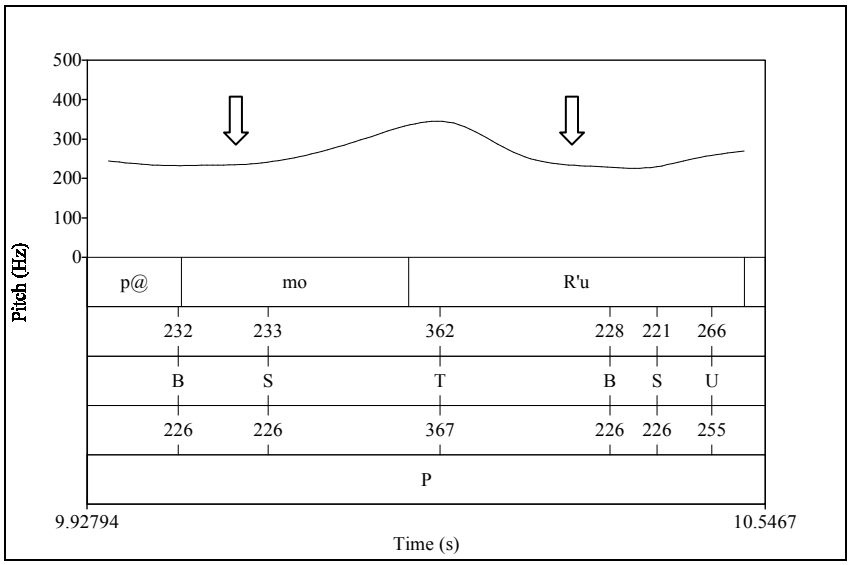


Figure 11.2 : Courbe mélodique correspondant à la séquence la plus longue de symboles INTSINT

Nous résumons dans le tableau XI.1 les informations sur la distribution du nombre d'étiquettes décrivant la courbe mélodique associée avec un mot porteur de la proéminence mélodique.

En nous fondant sur les chiffres du tableau XI.1, nous constatons une certaine stabilité : la majorité des courbes est décrite avec deux ou trois étiquettes (78,41 %). Cependant, les valeurs marginales peuvent être en partie dues à notre choix de l'unité d'analyse. En effet,

[283] Par exemple, H = 0, si un état du système a la probabilité de 1 (l'événement est certain).

comme nous l'avons souligné auparavant, le statut du mot phonétique au sein de la hiérarchie des domaines prosodiques est une question ouverte, qui nécessiterait des recherches approfondies. De ce point de vue, nous pensons que le choix d'une unité davantage associée à la structure rythmique permettrait de limiter la variabilité observée.

NOMBRE d'étiquettes dans la séquence	NOMBRE d'occurrences	FRÉQUENCE	PROBABILITÉ conditionnelle de terminer la séquence
1	5	1,6 %	0,016
2	114	36,19 %	0,3677
3	133	42,22 %	0,6786
4	49	15,56 %	0,7778
5	13	4,13 %	0,9286
6	1	0,3 %	-

Tableau XI.1 : Fréquences des séquences de différentes tailles

INTSINT comportant 8 symboles, nous nous sommes posée la question si toutes les étiquettes ont la même probabilité de ressortir dans l'annotation de la catégorie fonctionnelle « accentué ». L'entropie normalisée de ce système est égale 0.94, proche de l'entropie maximale, ce qui signifie que la distribution des catégories INTSINT est proche d'une distribution équiprobable dans l'annotation de la courbe mélodique associée à la catégorie « accentué ». Nous constatons cependant une faible présence de la catégorie M dans les codages des courbes mélodiques associées à la catégorie « accentué » (41 cas, 3% environ).

Nous avons ensuite analysé les différentes combinaisons d'étiquettes.

Tout d'abord, nous nous sommes posé la question de savoir si n'importe quelle étiquette d'INTSINT pouvait commencer la séquence qui correspond à la partie de la courbe mélodique, réalisée sur le mot porteur de la proéminence mélodique. En quête de réponse à cette question, nous avons examiné les effectifs et les probabilités de retrouver chaque étiquette d'INTSINT en position initiale : les données respectives sont présentées dans le tableau XI.2.

ÉTIQUETTE EN POSITION INITIALE	NOMBRE D'OCCURRENCES	PROBABILITÉ
B	20	0,061
D	47	0,150
H	23	0,079
L	32	0,102
M	39	0,124
S	76	0,242
T	5	0,016
U	73	0,226

Tableau XI.2 : Distribution des probabilités pour des étiquettes initiales dans la séquence

À partir des données, présentées ci-dessus, nous pouvons constater qu'une étiquette S a deux fois plus de chance de correspondre au début d'une courbe conduisant à la proéminence d'un mot phonétique qu'une étiquette L. Les chances qu'une étiquette T commence une telle séquence sont très faibles (au niveau de 1,6%). Si nous tenons compte de l'information qu'une étiquette apparaît au début de la séquence, l'entropie normalisée de notre système diminue[284] (H = 0.86), ce qui indique que cette information pourrait être pertinente dans le système.

Une fois la courbe initiée, dans 98,4% des cas, une autre étiquette suivra toujours alignée avec l'unité proéminente. L'alphabet INTSINT comporte 8 symboles ; nous avons rajouté à cet inventaire un diacritique, « # », désignant qu'une étiquette termine la séquence, associée avec l'unité porteuse de la proéminence mélodique. Les informations sur les probabilités conditionnelles, **σ** et **z** pour chaque bigramme de tons sont résumés dans l'annexe 7[285].

Le tableau XI.3 présente les valeurs d'entropie normalisée qui caractérisent l'état du système, une fois que le premier symbole dans le bigramme est connu. Rappelons que les valeurs de l'entropie normalisée d'un système se rangent dans l'intervalle [0,1], 0 caractérisant un système déterministe. Les meilleures valeurs d'entropie sont observées pour des catégories T et B.

SYMBOLE INTSINT	ENTROPIE NORMALISÉE
B	0,58
D	0,84
H	0,797
L	0,717
M	0,89
S	0,885
T	0,58
U	0,86

Tableau XI.3 : Entropie normalisée du système des bigrammes

Par référence aux probabilités conditionnelles présentées dans l'annexe X, nous apercevons, tout d'abord, que les proportions de cas où le symbole concerné est final dans la séquence sont fort majoritaires pour certaines étiquettes (plus de 50% pour L, 48,9% pour T, ou encore 47,33% pour B) ; pour d'autres catégories elles sont également supérieures aux probabilités d'observer une autre étiquette. Pour d'autres transitions, nous pouvons faire les remarques suivantes, en nous fondant sur les valeurs de z :

- pour l'étiquette T (qui ne peut jamais être suivie par les étiquettes T, U, H), il y a des chances non négligeables d'être suivie par l'étiquette L ; cependant, cette catégorie défavorise la présence des étiquettes M ou S ;

[284] Il s'agit de l'entropie conditionnée par l'événement ***e,*** égal « début de séquence » : l'information mutuelle normalisée apportée par l'événement ***e*** est de I = H(C) – H(C|e) = 0,08.
[285] Nous présentons ces informations en Annexe pour une plus grande lisibilité.

- il est peu probable que l'étiquette M finisse la séquence ; il y a des chances élevées qu'elle soit suivie par un H ;

- les étiquettes L, D, B sont proscrites après un B. Cette étiquette favorise la présence d'un S et défavorise les M et les H.

- la catégorie H favorise la présence d'un L ; cependant, elle est rarement suivie par un M, un H ou un U ;

- du fait de retrouver un L, nous pouvons nous attendre que cette étiquette ne sera pas suivie par un T, un M ou encore un L ;

- après un S, ce sont les étiquettes T et H qui sont privilégiées ; les M, B, et L sont, au contraire, peu probables ;

- après un U, le mouvement mélodique peut être ascendant (T) comme descendant (B) ; les catégories rares sont M, S, U ;

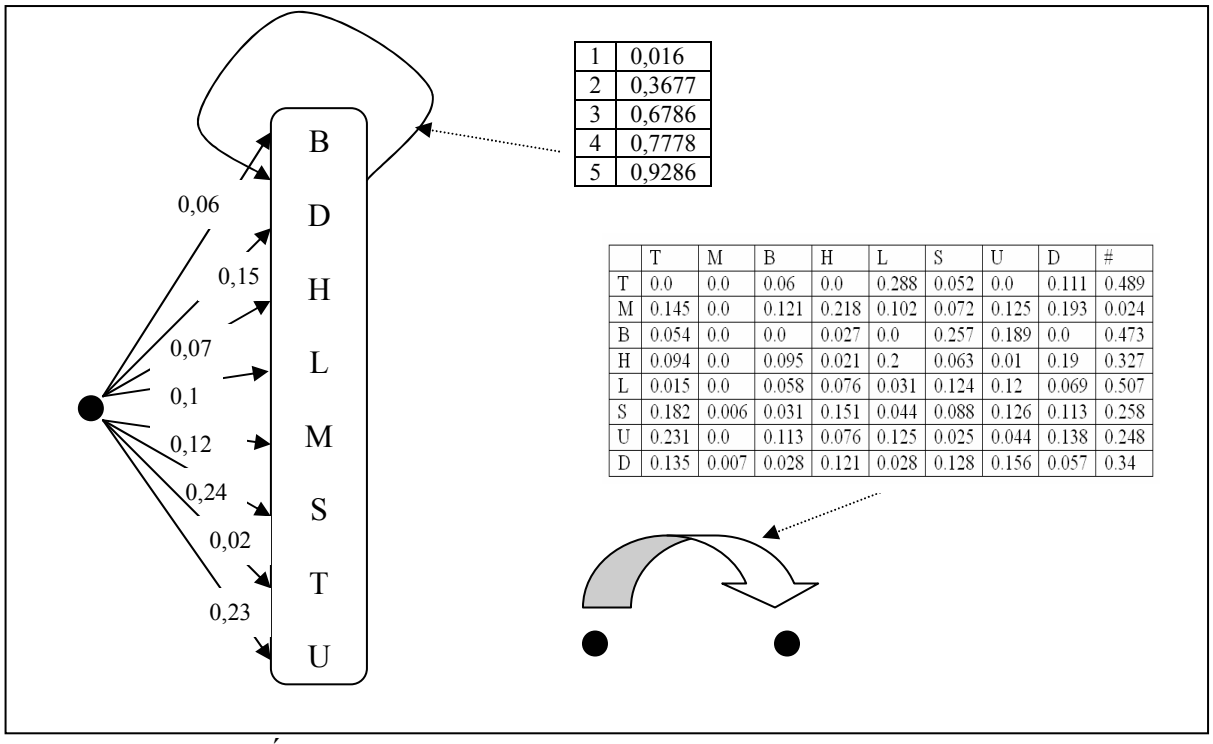
- enfin, un D privilégie la fin de séquence ; dans les configurations étudiées, il est rarement suivi par un M, B, L ou D.

Les outils que nous utilisons permettent également de procéder à une analyse simultanée des bi-, tri- et quadrigrammes. Nous avons retenu les séquences avec une meilleure entropie (tableau XI.4) : il s'agit des séquences dont la présence dans le corpus apporte le plus d'information sur le système. Il serait informatif dans le futur de mener une analyse comparative du comportement de ces séquences dans le corpus « accentué » et dans le reste du corpus. Cette analyse permettra-t-il ensuite une annotation automatique du corpus en termes de distribution des proéminences.

SYMBOLE INTSINT	ENTROPIE NORMALISÉE
T D	0,148
T L	0,223
S D	0,254
U T L	0,292
S D	0,317
L U	0,334
D T	0,388
H B	0,431
B S	0,46
L H	0,464
U B	0,466
H T	0,47
D T	0,474
D U	0,482
D S	0,484

Tableau XI.4 : Les séquences les plus informatives

Les analyses que nous avons menées jusqu'à présent, nous permettent de proposer une ébauche du modèle probabiliste de l'encodage INTSINT pour des unités proéminentes (cf. figure 11.3).



	T	M	B	H	L	S	U	D	#
T	0.0	0.0	0.06	0.0	0.288	0.052	0.0	0.111	0.489
M	0.145	0.0	0.121	0.218	0.102	0.072	0.125	0.193	0.024
B	0.054	0.0	0.0	0.027	0.0	0.257	0.189	0.0	0.473
H	0.094	0.0	0.095	0.021	0.2	0.063	0.01	0.19	0.327
L	0.015	0.0	0.058	0.076	0.031	0.124	0.12	0.069	0.507
S	0.182	0.006	0.031	0.151	0.044	0.088	0.126	0.113	0.258
U	0.231	0.0	0.113	0.076	0.125	0.025	0.044	0.138	0.248
D	0.135	0.007	0.028	0.121	0.028	0.128	0.156	0.057	0.34

Figure 11.3 : Ébauche d'une grammaire probabiliste INTSINT pour la primitive fonctionnelle « accentué »

Contrairement au traitement de Dainora (2001) pour ToBI, le contour que nous cherchons à représenter, directement lié à une primitive fonctionnelle, ne possède pas de structuration interne. En même temps, le modèle que nous avons représenté comporte deux parties : nous sommes amenée à proposer une représentation bi-partite, vu les variations en nombre de points cibles spécifiant la configuration de f0. Remarquons que cette partie est omise dans la représentation de Dainora. En effet, l'auteur établit des probabilités de rencontrer des syntagmes avec *X* accents mélodiques, mais n'inclut pas cette information dans sa grammaire. Probablement, juge-t-elle ces informations comme résultant de la performance. Or, ce raisonnement n'est pas applicable à la représentation INTSINT, vu le statut des points cibles, qui encodent les points de changement de direction dans la courbe de f0. Le panel gauche de la figure11.2 spécifie le processus de la réitération des points cibles, uns fois que la courbe est initiée par une des étiquettes (les probabilités étant également présentées). Notons toutefois que nous ne spécifions pas le processus d'itération en fonction de la nature de l'étiquette initiale de la séquence, faute de disposer de suffisamment d'observations. Le passage d'un point cible à l'autre – dans le cas des structures itérées – est régi par la matrice des probabilités conditionnelles spécifiée à droite. Précisons que ce modèle est stationnaire en ce qui concerne la gestion de passage d'une point cible à l'autre : lors de l'analyse des probabilités conditionnelles, nous ne prenons pas en compte le fait que ce T se retrouve après un L ou après un U. Nous avons fait le choix de procéder par bigrammes, en particulier, parce qu'une partie des courbes n'était spécifiée qu'avec deux points cibles. Nous n'avons pas pris en compte l'analyse des trigrammes comme nous considérons que l'échantillon d'analyse est relativement petit. Remarquons encore une

fois, qu'une homogénéisation des données par rapport au nombre de points cibles par unité, ainsi qu'une réflexion sur l'unité appropriée de l'analyse, sont cruciales pour poursuivre l'application de cette voie méthodologique.

Dans les analyses par bigrammes, nous supposons que les séquences d'étiquettes sont des chaînes Markoviennes du premier ordre : cela signifie que le choix d'une étiquette ne dépend que de la catégorie de l'étiquette précédente, sans faire prendre en considération des étiquettes plus précoces. Il s'agit d'un modèle des plus simples. La grammaire de l'intonation de Pierrehumbert est un autre exemple du modèle Markovien de premier ordre. Or, ce postulat a été remis en cause dans l'étude de Dainora (2001) qui démontre, au moins pour la séquence des étiquettes qui clôturent une unité intonative (c'est-à-dire, comportant un accent nucléaire, un accent de syntagme et un ton de frontière) qu'il s'agit fort probablement d'un modèle du deuxième ordre, le choix du ton de frontière étant dépendant de la catégorie de l'accent nucléaire. À son tour, INTSINT ne propose pas de distinction catégorielle entre les étiquettes. Or, nos données suggèrent que l'entropie normalisée diminue (donc, le système devient plus déterministe), une fois l'analyse menée à partir des trigrammes. Nous pouvons donc supposer que la grammaire INTSINT sera très probablement un système Markovien du deuxième ordre. Cependant, afin de poursuivre des analyses, il nous faut tout d'abord résoudre la question de l'unité optimale d'analyse (de l'unité minimale de l'organisation tonale).

L'analyse que nous avons entreprise jusqu'à présent a été essentiellement guidée par l'annotation symbolique. Or, le problème peut être abordé sous un autre angle. Nous pouvons ainsi essayer de dégager des similarités dans les configurations mélodiques que nous observons pour les mots phonétiques proéminents afin de voir quelles annotations correspondent aux courbes de la même classe. Cette partie de l'analyse est discutée dans la section suivante.

11.4. MODÉLISATION QUANTITATIVE DES DONNÉES

Plusieurs études entreprenant une modélisation quantitative des contours mélodiques ou de leurs parties ont paru récemment. Nous pensons notamment au modèle d'annotation-synthèse de Kochansky et Shih (2005), à l'étude de Klabbers et van Santen (2004) ainsi qu'à celle de Grabe *et al.* (2003), menée dans l'objectif de rendre compte de la variation dialectale dans le domaine de la prosodie. Précisons que ces analyses rejoignent dans leur démarche les méthodes de synthèse à partir d'un lexique de courbes intonatives (cf. chapitre VI). Or, simultanément, avec des objectifs d'application, la démarche d'une modélisation quantitative a valeur d'outil analytique.

À la suite de notre analyse qui vise à informer une grammaire probabiliste d'INTSINT pour la primitive fonctionnelle « accentué », nous nous sommes posé la question des contours prototypes. Nous espérons trouver un ensemble limité de contours en association avec notre unité d'analyse. Une fois cette tâche accomplie, il est légitime de nous interroger sur les variations dans les annotations INTSINT correspondantes.

11.4.1. Méthodologie

Dans notre étude orientée plan d'expression, nous nous sommes inspirée de la méthodologie proposée par Klabbers & van Santen (2004). La méthode proposée cherche à former les agrégations des contours mélodiques, déterminées pour des pieds accentuels trochaïques, afin d'établir les contours mélodiques prototypes. Cette méthode incorpore les présupposés du modèle superpositionnel de l'intonation développé par van Santen (van Santen et Möbius 2000). Le travail est mené à partir d'un corpus segmenté en pieds, les contours mélodiques étant d'une durée variable, suite à la variabilité au niveau du pied en nombre de syllabes et en nombre de phonèmes. Pour rendre compte de la forme de chaque contour, on retient 100 points ; or, sachant qu'il existe un rapport d'équivalence phonologique entre les contours alignés avec le matériau segmental variable, notamment, en référence avec la durée/longueur de l'unité, une normalisation est requise, laquelle porte sur les détails d'alignement du contour. Principalement on tient compte de la position du maximum. Dans ce cas, lors de l'échantillonnage, 50 % des points décrivent le tracé de la courbe de f0 avant le pic et les 50% restants de la courbe après le pic. Précisons que lors de la modélisation ce sont les positions prédites du pic qui sont retenues et non pas les localisations propres à la structure sous analyse. Une deuxième normalisation concerne les variations de la hauteur et du registre mélodiques : une normalisation en « transformé-z » est alors appliquée. Le processus de la formation des agrégations, qui est automatique, s'effectue en appliquant la méthode de classement hiérarchique *hclust* (sous R). Toutes les méthodes de ce type (*clustering analysis*) sont fondées sur des mesures de similarité ou de différenciation entre les agrégations. *Hclust* est une méthode ascendante : au début, chaque observation forme sa propre classe ; ensuite, à chaque étape, l'algorithme qui procède de manière itérative, permet de rassembler les deux agrégations les plus proches (les plus similaires) ; ce processus est répliqué jusqu'à ce que on n'obtienne qu'une seule agrégation. L'avantage de la méthode *hclust* réside dans le fait que l'on ne doit pas spécifier le nombre d'agrégations au préalable. Les différentes méthodes peuvent déterminer le regroupement. Nous avons appliqué la méthode « *ward* » qui est une méthode de variance minimale aboutissant à des agrégations sphériques.

11.4.2. Application de la procédure de normalisation

Sans reprendre à la lettre le modèle de Klabbers et van Santen (2004) discuté ci-dessus et l'approche superpositionnelle défendue par van Santen et Möbius (2000), nous partageons les observations sur l'équivalence phonologique des contours alignés avec le matériau linguistique de longueur variable (ce principe étant un des axiomes de l'approche autosegmentale). Nous avons choisi de suivre la procédure de normalisation proposée par les auteurs ; or, compte tenu du caractère morphologique de l'accentuation en russe, nous sommes obligée de prendre en considération deux paramètres : le nombre de syllabe et la place de l'accent lexical au sein du mot phonétique (qui est notre unité d'analyse).

Dans un premier temps, nous avons procédé à une analyse succincte de la structure rythmique des unités de notre corpus. Nous avons pu constater que la taille moyenne de l'unité est de 2,75 syllabes, avec un maximum de 7 syllabes (une seule observation dans notre corpus, il s'agit du mot « *по телеви́зору* » (*à la télé*). Les données sur la distribution de taille des unités dans le corpus sont résumées dans le tableau XI.5.

NOMBRE DE SYLLABES	NOMBRE D'OCCURRENCES (SUR 315 UNITÉS)	FRÉQUENCE
1	25	7,94%
2	98	31,11%
3	132	41,90%
4	51	16,19%
5	8	2,54%
6	-	
7	1	0,32%

Tableau XI.5 : Fréquences d'unités avec le nombre de syllabes désigné

Nous constatons que la majorité des unités ont de 2 à 3 syllabes (72%), les structures monosyllabiques ainsi que les unités de plus de 4 syllabes sont marginales (tout au moins, pour le corpus étudié).

Le tableau XI.6 ci-après, en plus du nombre de syllabes par unité, prend en compte la structure rythmique proprement dite, c'est-à-dire, la place de la syllabe métriquement forte. La dernière colonne de ce tableau présente la fréquence inter-classe, à savoir, la fréquence d'occurrence de la structure rythmique donnée parmi les structures de même longueur.

STRUCTURE RYTHMIQUE[286]	NOMBRE D'OCCURRENCE (SUR 315 UNITÉS)	FRÉQUENCE INTER-CLASSE
1	25	100%
2_1	52	53,06%
2_2	46	46,94%
3_1	19	14,39%
3_2	84	63,64%
3_3	29	21,97%
4_1	1	1,96%
4_2	15	29,41%
4_3	30	58,82%
4_4	5	9,8%
5_2	1	12,5%
5_3	5	62,5%
5_4	2	25%
7_5	1	100%

Tableau XI.6 : Distirubtion des fréquences de différentes structures rythmiques dans le corpus

Nous constatons alors pour les structures de 3 syllabes et plus une forte tendance accento-centrique : pour les unités de 3 syllabes, les structures avec la syllabe métriquement forte en position centrale sont trois fois plus nombreuses que les structures avec un accent final ;

[286] Dans la désignation « 2_1 », le premier chiffre indique le nombre de syllabes que le mot phonétique proéminent comporte ; le deuxième chiffre indique la position de la syllabe porteuse de l'accent lexical ; il s'agit donc, d'un mot bi-syllabique avec l'accent lexical sur la première syllabe.

le patron similaire est observé pour les unités de 4 syllabes (les unités plus longues étant rares, nous ne pouvons pas étendre les généralisations).

Nous avons ensuite examiné le lien entre la structure rythmique et la durée du mot phonétique moyennant des tests statistiques de modèles linéaires. Le modèle complet, c'est-à-dire, avec l'interaction des facteurs « nombre de syllabes » et « place de la syllabe métriquement forte », révèle l'effet significatif du nombre de syllabes par unité, hors l'effet non-significatif de la place de la syllabe proéminente[287] et de l'interaction de facteurs (tableau XI.7).

FACTEUR	*DF*	*VALEUR F*	*VALEUR p*
nombre de syllabes	*5*	33.3225	< 2e-16
structure rythmique	*3*	3.1042	0.02691
interaction de facteurs	*5*	2.0781	0.06805

Tableau XI.7 : Validation statistique d'influence des facteurs rythmiques sur la durée du mot

Si nous examinons de plus près l'effet du nombre de syllabes, nous constatons tout d'abord une durée similaire entre les structures d'une et de deux syllabes (différence non-significative, t = 1,046, p = 0,296). Ensuite, la durée reste stable pour les structures au-delà de 4 syllabes (notamment pour les structures de 4 et 5 syllabes).

Nous avons ensuite étudié la localisation du maximum de f0 dans l'unité selon sa structure rythmique. Les données sont représentées graphiquement dans la figure 11.4 : nous limitons nos analyses aux structures de quatre syllabes et moins (nous écartons les données sur les structures de 5 et 7 syllabes en raison du nombre limité d'observations).

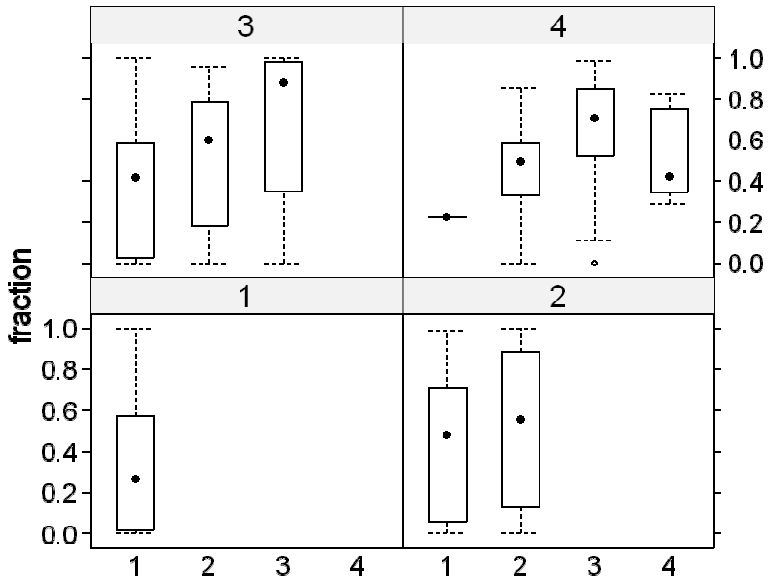


Figure 11.4 : Localisation relative du pic mélodique exprimée en fraction de la durée absolue de l'unité (mot prosodique)

[287] Il se peut que la place de la syllabe proéminente ne soit pas significative, vu le lien intrinsèque avec le facteur du nombre de syllabes. Nous avons testé l'effet de ce facteur sur un sous-ensemble de notre corpus, à savoir, les structures de trois syllabes : l'anova du modèle indique un effet non-significatif de la variable indépendante (F(2, 129) = 1,555, p= 0,215). L'analyse des regresseurs indique pourtant que les structures avec accent initial sont plus longues de trois options.

Nous constatons premièrement la position singulière des structures monosyllabiques, tandis que les structures de 2 syllabes et plus démontrent des patrons de distribution similaires. Pour ces structures, en effet, la fraction dépend surtout de la position de la syllabe porteuse de l'accent lexical. En moyenne, les pics sont localisés à 32% de la durée de l'unité pour les structures monosyllabiques et à 52% pour les autres structures. Le tableau XI.8 présente une analyse plus détaillée de la position du maximum de f0 dans les structures de plus de 2 syllabes. Nous jugeons pourtant que les données sur le maximum dans les structures avec l'accent sur la quatrième syllabe sont propres au corpus d'étude et ne peuvent pas être généralisées à d'autres corpus.

Syllabe porteuse d'accent lexical	*Localisation du maximum de f0 (en % de la durée d'unité)*
1	41%
2	52%
3	66%
4	52%

Tableau XI.8 : Position du maximum de f0 en fraction de la durée absolue de l'unité

Les informations du tableau XI.8 ont été implémentées dans le script *get_shape.praat* qui effectue l'échantillonnage des valeurs de f0 dans les contours mélodiques sous analyse.

11.4.3. Résultats

Avant même d'entreprendre les calculs dont nous avons présenté les détails et les présupposés ci-dessus, il est important de se poser la question du nombre de formes que l'on s'attend à trouver ? Certes, toute réponse exige une vérification expérimentale dans une étude perceptive ; nous insistons sur ce point notamment dans une perspective d'articulation entre la forme et la fonction, et celle de la signification de l'entité prosodique. Cependant, en nous limitant au plan d'expression, nous pouvons déjà procéder à quelques prévisions.

Si l'on prend en considération la direction du mouvement mélodique, quatre alternatives se présentent : contours montants (compte tenu de variants éventuels avec ou sans plateau après le maximum), contours descendants (par exemple, dans la situation du focus large, la descente pouvant être suivie par un plateau), contours montants-descendants et descendants-montants. Nous pouvons également envisager une classe séparée ou une variante avec un pic retardé (*delayed pic*).

Les pronostics sur le nombre d'alternants sont d'autant plus nécessaires que la technique statistique d'analyse discriminante peut nous fournir le nombre voulu d'agrégations. Nous avons examiné, visuellement (les courbes de f0) comme dans une analyse auditive informelle, les issues possibles du processus de classification, comportant de 4 à 10 agrégations. Nous avons choisi de présenter et de discuter une partition avec 6 agrégations ; cependant, ce choix nécessite une vérification postérieure dans une étude en perception, point sur lequel nous débattrons davantage dans la section de discussion.

Pour chaque agrégation nous avons obtenu un patron prototype du groupe que nous présentons dans la figure 11.4. Pour obtenir cette représentation, nous avons procédé de la façon suivante :

a) dans le domaine temporel, chaque contour a été normé selon la procédure décrite dans la section précédente ;

b) dans le domaine fréquentiel, dans un premier temps, nous avons obtenu des contours « transformé-z » (procédant par une soustraction de la f0 moyenne du contour et divisant la différence par écart type) ;

c) pour toute agrégation le contour médian a alors été calculé ;

d) ensuite nous avons recalculé la valeur en Hz, rajoutant la moyenne et l'écart type de l'agrégation.

Les six contours prototypes sont présentés dans la figure 11.5. Le tableau XI.9 détaille la distribution des effectifs selon les agrégations et prend également en compte le paramètre du nombre de syllabes par unités. Nous constatons déjà une répartition plus ou moins paritaire entre les agrégations pour les unités de longueur variable ; sauf, peut-être, les structures mono-syllabiques, pour lesquelles la proportion des unités appartenant à l'agrégation 1 est assez élevée. Ce résultat est probablement dû aux structures avec les pro-phrases (portant une proéminence mélodique de la configuration descendante).

	I	II	III	IV	V	VI
1 syllabe	13 (52 %)	5 (20 %)	4 (16 %)	1 (4 %)	1 (4 %)	1 (4 %)
2 syllabes	17 (17 %)	26 (27 %)	23 (23 %)	10 (10,5 %)	12 (12 %)	10 (10,5 %)
3 syllabes	18 (14 %)	34 (26 %)	21 (16 %)	25 (19 %)	15 (11,5 %)	19 (14,5 %)
4 syllabes	7	7	8	16	4	9
5 syllabes	1	0	1	3	2	1
7 syllabes	0	0	0	0	0	1
Total	*56*	*72*	*57*	*55*	*34*	*41*
Pourcentage	*17,8*	*22,9*	*18,0*	*17,5*	*10,8*	*13,0*

Tableau XI.9 : Distribution des unités de différentes structures rythmiques entre agrégations résultantes

L'examen du tableau nous permet de distinguer déjà 2 groupes de contours : les contours descendants (descendants-montants, agrégations 1, 3, 5) et les contours montants (montants-descendants, agrégations 2, 4, 6). 46,6 % des contours appartiennent au premier groupe, 54,4 % - au deuxième. Nous discutons par la suite de chacun des groupes et de chacune des agrégations.

Pour les contours du premier groupe, la différence majeure réside dans la gestion des rapports entre les composantes-plateaux et les descentes elles-mêmes ; de même, les agrégations se différencient par rapport à la présence d'un mouvement final ascendant.

La première agrégation comprend 17,8% de contours analysés (56 sur 315). Pour les contours de ce groupe, le mouvement descendant est précédé et suivi par un plateau. Le mouvement mélodique n'est pas ample, de 2,5 demi-tons (pour le patron prototype) ; or, comparé à deux autres contours descendants, il est plus abrupt.

18% de contours (57 sur 315) appartiennent à la troisième agrégation (toujours dans le groupe descendant). Le mouvement mélodique commence plus bas que pour les contours du premier groupe, il n'est pas précédé par un plateau, est moins abrupt, et le plateau final exhibe une légère tendance ascendante.

Pour l'agrégation 5 (34 unités sur 315, 10,8%) la tendance ascendante au final est beaucoup plus prononcée. Le mouvement est plutôt descendant-montant, quoique les intervalles mélodiques ne soient pas très larges.

Pour le groupe des contours ascendants, nous constatons une variation pour la position du pic mélodique ainsi que pour la localisation du début du mouvement ascendant. Les empans mélodiques parcourus sont comparables entre les trois agrégations. La deuxième agrégation est la plus grande, elle contient 22,9% des contours (72 sur 315). Il s'agit d'un mouvement mélodique montant, précédé d'un plateau, avec un pic mélodique localisé au-delà de 80% de la durée de l'unité ; nous n'y observons aucune partie descendante. Pour les deux autres agrégations, le pic, réalisé plus tôt dans la structure, est suivi d'un mouvement descendant. Par référence au groupe des contours ascendants, se pose la question de l'influence éventuelle de la structure rythmique de l'unité sur le patron mélodique associé (cf. le débat sur les mécanismes de troncation et de compression). Plus particulièrement, on se demande si pour les structures de plus d'une syllabe, le choix entre les structures 2, 4 et 6 est guidé par la structure rythmique. Nous avons cherché à répondre à cette question en analysant la distribution des agrégations pour les structures tri-syllabiques (le fait de disposer de suffisamment de structures tri-syllabiques dans le corpus analysé a motivé ce choix restrictif). Nous présentons les données dans le tableau XI.10.

	AGRÉGATION2	AGRÉGATION4	AGRÉGATION6
Accent initial	0	7	3
Accent médian	19	17	14
Accent final	15	1	2

Tableau XI.10 : Distribution des structures prosodiques de trois syllabes entre les agrégations des contours ascendants

L'analyse du tableau XI.10 montre que le patron d'agrégation 2 (structure avec un pic retardé) n'est jamais associé avec les structures qui contiennent un accent lexical en position initiale ; or, cette agrégation est dominante pour les unités avec un accent lexical final. Cependant, la distribution est relativement équilibrée pour les structures avec un accent médian (ce résultat étant confirmé par le test des proportions, $t(2) = 0.948$, $p = 0.623$). De même, les effectifs que nous observons pour les structures à accents initial et final, peuvent être dues au choix de l'unité d'analyse, ainsi qu'au fait que nous ne tenons pas compte du facteur structurel de la position au sein du syntagme intonatif.

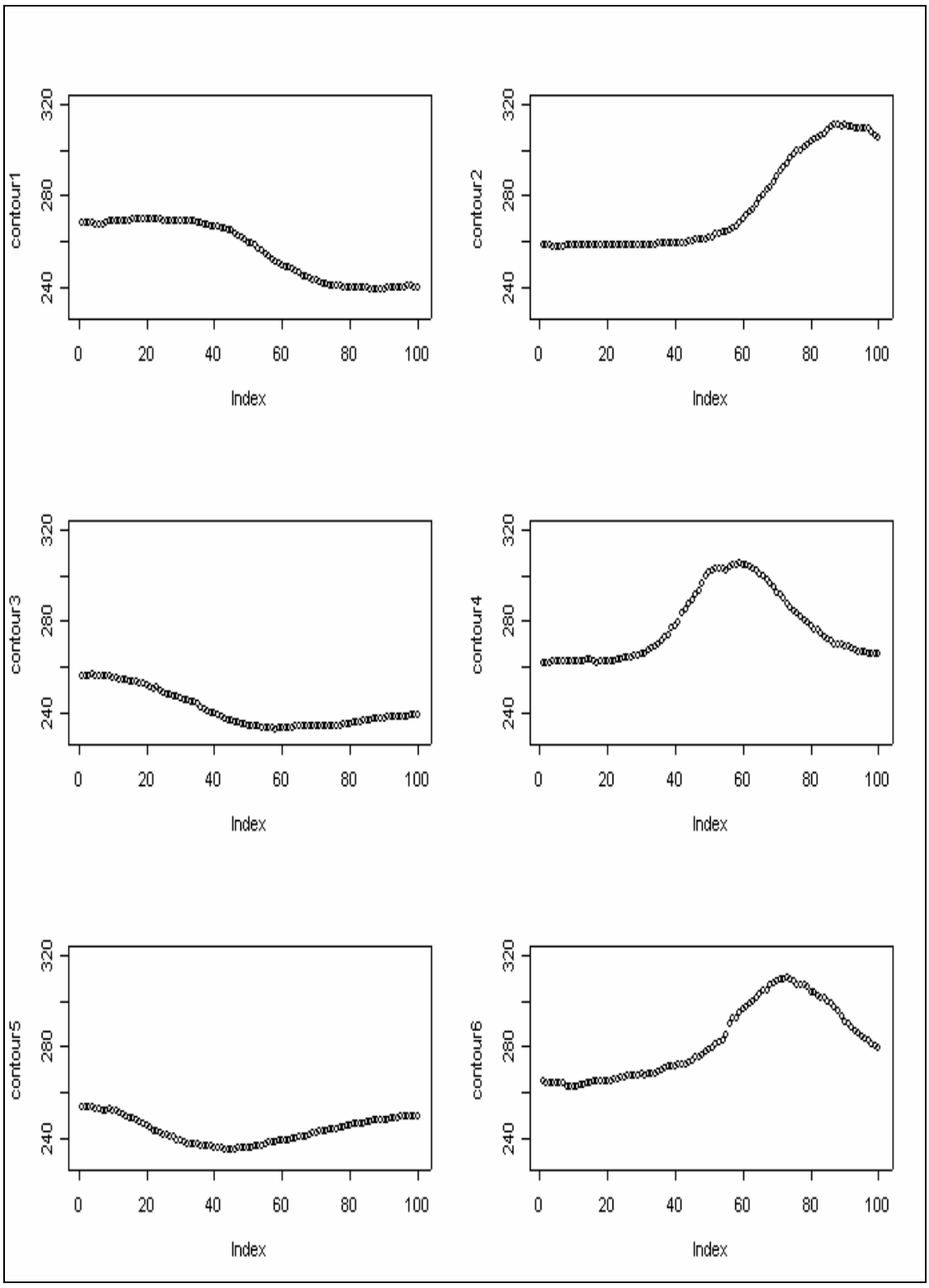


Figure 11.5 : Contours prototypes des agrégations obtenues

Dans un deuxième temps, nous avons cherché à examiner la variabilité interclasse. À cette fin, nous avons représenté tous les contours appartenant à chacune des agrégations (figure 11.6). Nous constatons tout premièrement une variabilité considérable du niveau mélodique, mais également de l'empan entre les points d'inflection de la courbe. Deux conséquences importantes peuvent être tirées de l'examen de la figure 11.6 :

a) nous pouvons nous poser la question de savoir sur quel plan la différence de l'empan mélodique peut être catégorielle. S'agit-il d'une signification distincte véhiculée ou du degré d'expression d'une signification ?

b) nous pouvons également réfléchir sur le fait que les différences d'empan vont certainement induire une variabilité dans les étiquettes INTSINT associées avec les contours. La question se pose de savoir comment gérer cette variabilité en vue d'une grammaire probabiliste. En effet, nous avons examiné trois contours appartenant à l'agrégation 6 (figure 11.7). Suite aux choix méthodologiques que nous avons faits lors de la réalisation de l'annotation MOMEL - INTSINT[288], à savoir, le fait de traiter séparément toute intervention de la locutrice afin de tenir compte de la variation du registre mélodique, il arrive que les points qui ont la même valeur sur l'échelle des fréquences reçoivent des étiquettes différentes. C'est le cas notamment de l'annotation du sommet mélodique pour les panels 1 et 3 qui ont une même valeur fréquentielle de 334-336 Hz, bien que l'un soit annoté « H » et l'autre « T ».

Dans un autre cas, les deux points différant de 1,2 demi-ton ont une même annotation (en l'occurrence, « T », panels 2 et 3 de la figure). Remarquons également les différences en nombre d'étiquettes pour l'encodage du plateau initial du contour. Notons, pourtant, que les différences sont atténuées pour des valeurs prédites de f0. Or, comme nous l'avons déjà dit, une étude perceptive est nécessaire pour s'assurer de l'équivalence entre le contour observé et le contour prédit. Dans le contexte des données présentées, nous pouvons alors réfléchir sur le besoin d'une méta-annotation, permettant de gérer la variabilité attestée. Cette démarche nous paraît également avantageuse pour l'augmentation de la force prédictive de la grammaire probabiliste, mais également pour la gestion de la variabilité dans la parole synthétisée.

11.5. RÉSUMÉ ET DISCUSSION

Une description formelle de l'organisation tonale cherche à proposer une analyse symbolique des parties du contour mélodique qui actualisent certaines catégories intonatives (primitives fonctionnelles). Simultanément, dès le début du présent travail nous avons défendu la position selon laquelle une étude fondée sur le corpus pourrait aboutir à un modèle formel de l'intonation, voire un modèle enrichi par la prise en compte systématique des facteurs discursifs. Dans ce dernier chapitre, nous avons entrepris une telle étude pour la catégorie 'accentué' et ceci dans une double perspective : d'abord, nous avons effectué des analyses en nous fondant sur l'annotation symbolique du niveau tonal ; ensuite, dans la deuxième partie, l'étude a été menée à partir de l'implémentation phonétique des courbes mélodiques associées avec des unités porteuses de la proéminence mélodique. Entre les unités de domaine d'association tonale, de pied accentuel et de mot phonétique (mot prosodique), nous avons choisi de travailler avec l'unité du mot phonétique, du moins à cette étape initiale.

Nous interprétons les points cibles détectés par l'algorithme MOMEL comme indiquant des moments d'une manipulation volontaire de la courbe mélodique de la part du locuteur ;

[288] Nous avons discuté de ces choix dans le chapitre V.

cette manipulation contribue à la gestion de la structure informationnelle et du plan de contenu des unités intonatives[289]. Notre objectif, dans la quête d'une grammaire probabiliste des catégories INTSINT, est d'acquérir des informations sur les probabilités de différentes configurations dans leur association avec la primitive fonctionnelle désignée ('accentué'). Vu le caractère (semi)-automatique de l'annotation MOMEL-INTSINT et sa qualité objective, une grammaire probabiliste pourra être appliquée par la suite dans une annotation automatique du corpus en termes de primitives fonctionnelles. Parallèlement, une variabilité inhérente au concept de la probabilité, assurera une intonation non-monotone et non-répétitive-à-outrance en synthèse. Certes, la taille de corpus d'analyse dans le présent chapitre – 315 unités porteuses de la proéminence mélodique, avec une éventualité forte des sub-catégories [± nucléaire], [± emphatique][290] – ne permet pas de considérer les résultats obtenus comme définitifs. De surcroît, la question de l'alignement des points cibles INTSINT avec la chaîne segmental/syllabique nécessite d'être clarifiée. Dans le contexte actuel, nous avons limité nos analyses aux bi-grammes. Pourtant, une normalisation de l'alignement s'impose[291] : nous avons observé les cas où une même configuration – plateau précédant une excursion mélodique – reçoit des annotations divergentes en nombre de symboles.

Nous avons également constaté que la subcatégorisation des étiquettes INTSINT est parfois trop fractionnée dans le cadre d'une grammaire probabiliste (rappelons que l'algorithme automatique de l'annotation n'est plus fondé sur le modèle INTSINT configurationnel). Les distinctions entre les catégories U, H, T traduisent selon nous des valeurs pragmatiques rajoutées, à la fois dans le contexte local (excursion plus importante de la courbe) ou globale (en présentant notre corpus, nous avons spécifié que l'unité du traitement lors de l'annotation INTSINT était une intervention de la locutrice). Notons également que nous cherchons une grammaire probabiliste pour une catégorie fonctionnelle. Or, comme le remarque Calhoun (2005) il n'existe aucune association statistiquement significative entre l'étiquette d'accent mélodique ToBI et le statut « topic ».

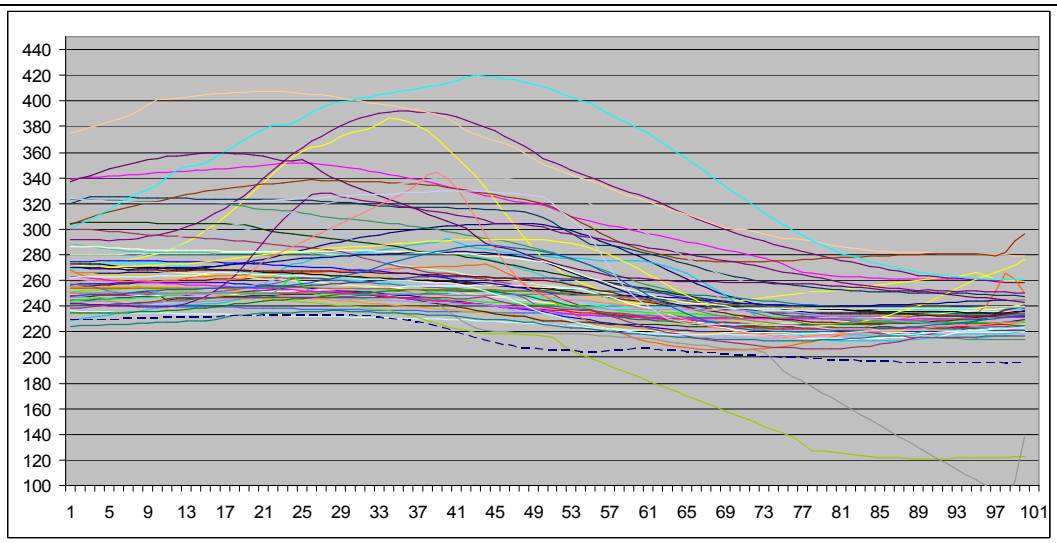
Dans la même lignée, l'analyse menée à partir de la forme phonétique des contours, nous a permis de révéler 6 classes majeures de courbes, avec une importante variabilité interne aux agrégations. Nous constatons à la fois une variation pour le point de départ de la courbe ainsi que pour la hauteur du pic. Dans le cadre de l'approche associant la distribution des proéminences mélodiques avec la gestion du focus (*focus-to-accent*), il nous paraît nécessaire de tenir compte des recherches récentes qui pointent la complexité de la problématique du domaine d'implémentation de la catégorie « focus ». Il a été démontré que l'implémentation de cette catégorie pragmatique nécessite non seulement un accompagnement par une excursion substantielle dans la courbe de f0, mais aussi un réajustement dans le registre des parties pré- et post-focales (Xu et al. 2004). L'étude antérieure de Rump et Collier (1996) analyse l'impact de la manipulation de hauteur du pic

[289] Se pose alors la question du statut des corrections manuelles des points cibles, surtout dans le cas des points cibles rajoutés.

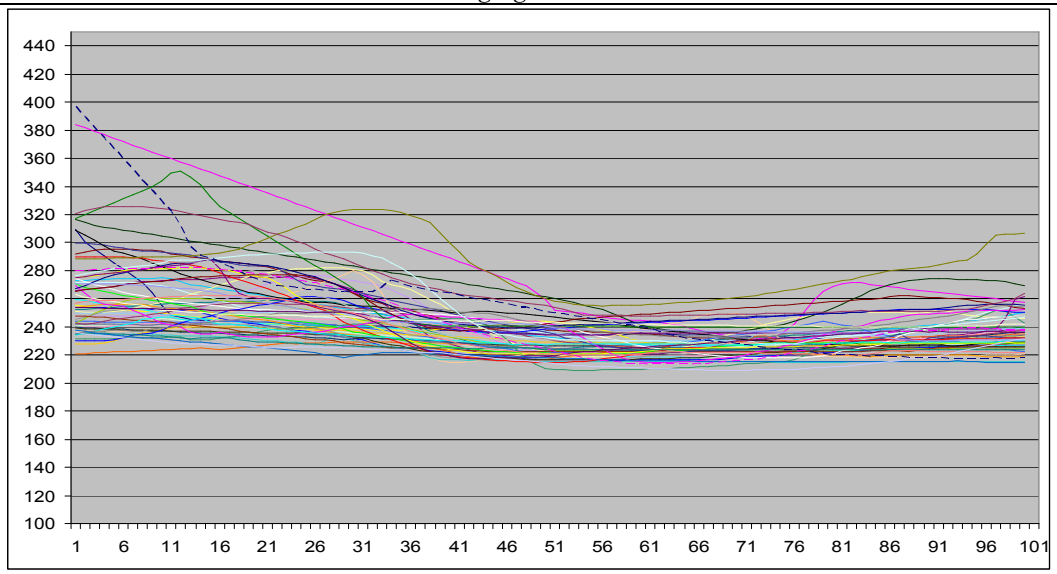
[290] Plusieurs paramètres sont à contrôler dans les études postérieures : distinctions dans les spécifications secondaires (accent nucléaire, emphatique), mais aussi position dans la structure du syntagme (« avant une frontière prosodique majeure/mineure »).

[291] Nous n'envisageons pas d'imposer l'association d'un point cible par mot, mais de mener une réflexion méthodique, probablement, en référence à un des niveaux du phrasé prosodique. Nous pensons notamment au concept d'unité tonale.

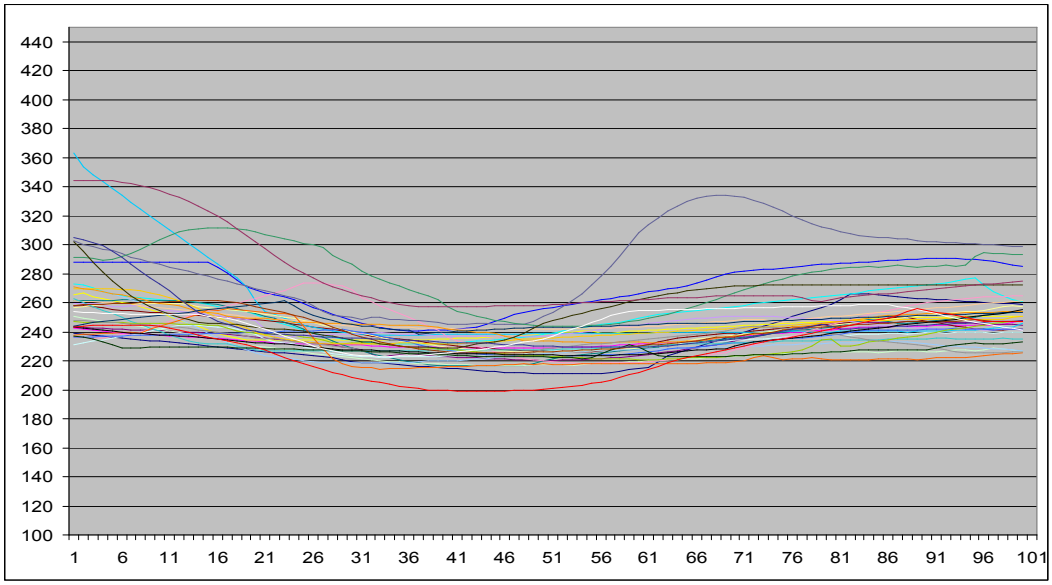
mélodique sur la perception de la structure informationnelle du message en termes de distribution du focus. Il est démontré que pour la perception d'un focus étroit au début de l'énoncé, le pic mélodique correspondant doit être très saillant (une excursion importante), suivi d'un pic très faible plus tard dans le syntagme. Or, si le focus est associé à la fin du syntagme, il n'y a pas de contrainte sur la hauteur du premier pic, le deuxième pic devant simplement être plus saillant. De ce fait, nous pouvons conclure que la variabilité interne aux agrégations des courbes mélodiques à valeur proéminente ont une double source : a) dans les effets paralinguistiques et b) dans les effets de la structure hiérarchique. Leur apport respectif nécessite une investigation approfondie dans une étude expérimentale en perception. Dans ce contexte, la taille de l'unité d'analyse doit également être sujette aux investigations postérieures.



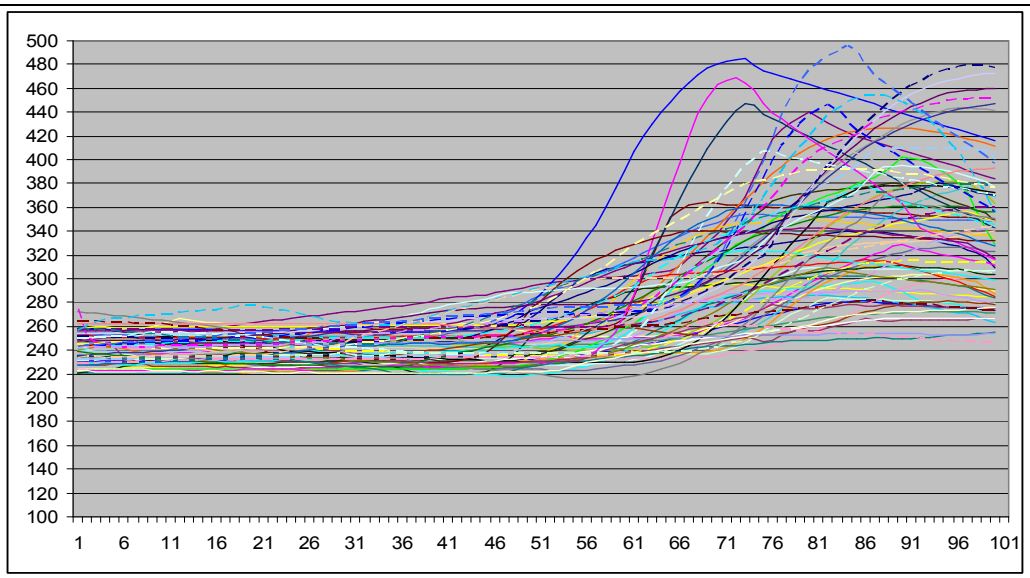
Agrégation1



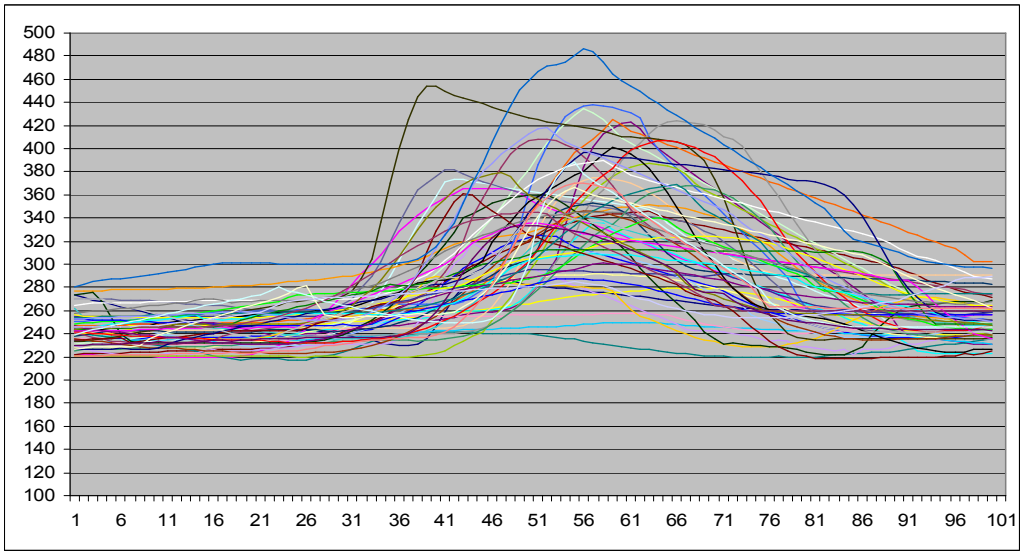
Agrégation 3



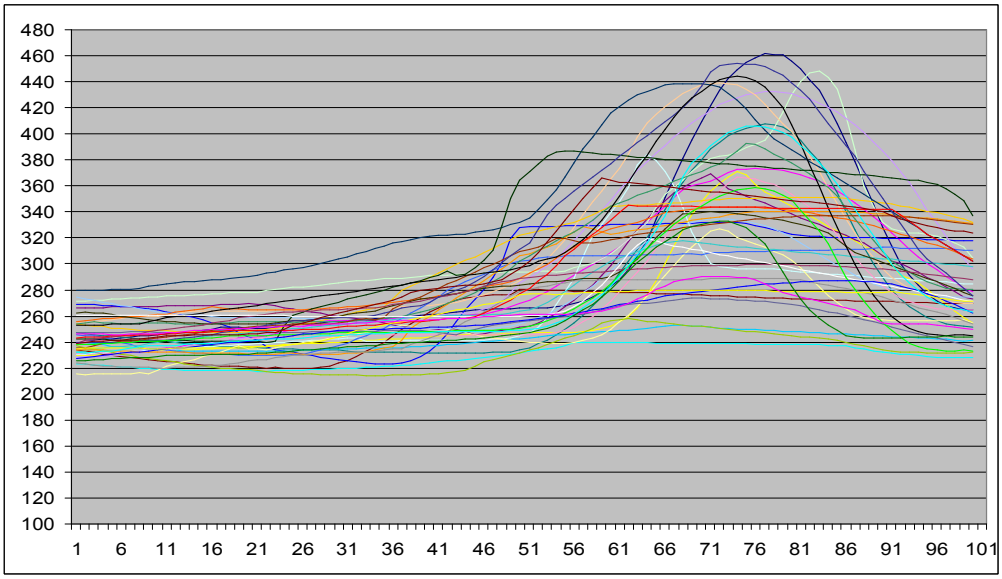
Agrégation 5



Agrégation2



Agrégation4



Agrégation6

Figure 11.6 : Variabilité interne de chaque agrégation

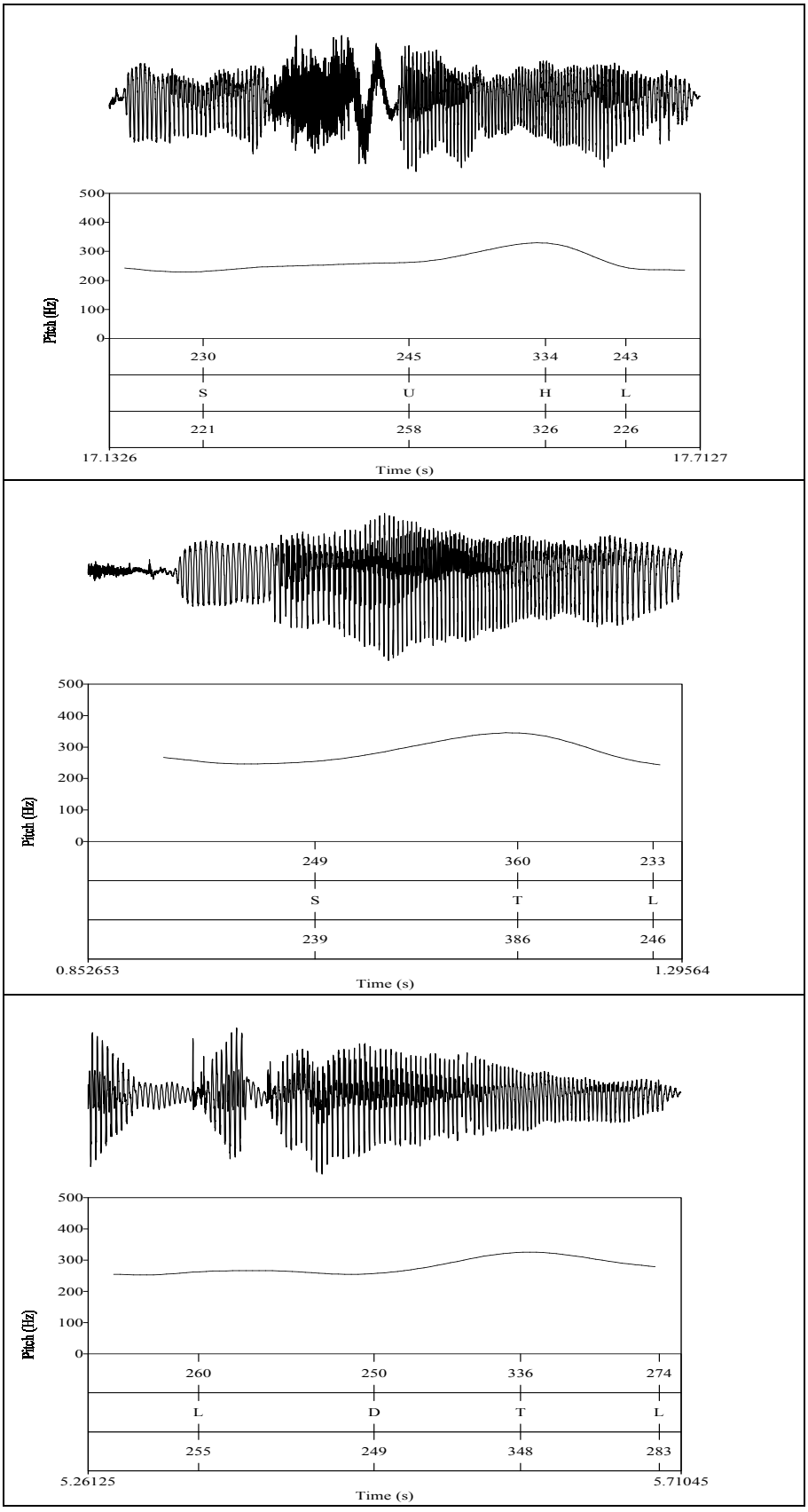


Figure 11.7 : Illustration de la variabilité dans l'encodage INTSINT

CONCLUSIONS ET PERSPECTIVES D'AVENIR

L'objectif de ce travail était de contribuer à la représentation formelle de l'intonation du parler russe spontané, notamment en vue d'une application à un système de synthèse de la parole qui utiliserait la méthodologie d'une double sélection d'unités. Nous sommes persuadée qu'une entreprise de formalisation du système prosodique se doit de conjuguer l'étude de la dimension tonale et celle du phrasé prosodique. Nous nous sommes donc fixée deux objectifs complémentaires :

a) décrire et modéliser le phrasé prosodique dans ces interactions avec d'autres niveaux de la description linguistique, avant d'entreprendre sa modélisation acoustique ;

b) déterminer des formalismes qui permettront d'appréhender le lien entre les représentations formelle et fonctionnelle des phénomènes prosodiques.

La réalisation de ces objectifs a nécessité l'élaboration d'un parcours progressif qui constitue le fil conducteur de notre thèse. Dans cette optique, il nous a fallu :

a) débattre certaines questions fondamentales qui concernent les choix des cadres théoriques et méthodologiques retenus dans la présente étude : celui de la phonologie intonative et celui de la phonologie des domaines ;

b) développer une méthodologie conséquente, compte tenu de nos convictions sur l'importance des données perceptives et sur le rôle de la dimension de la variabilité pour toute modélisation en vue d'applications technologiques ;

c) mettre en œuvre des recherches expérimentales dans une triple perspective : servir d'heuristique de vérification des hypothèses de départ, aboutir à une modélisation rigoureuse des phénomènes observés et permettre la formulation de nouvelles hypothèses.

Les résultats de cette approche s'organisent pareillement autour de deux axes qui se rapportent respectivement à la problématique du phrasé prosodique et à celle de l'articulation « forme ~ fonction ».

Le phrasé prosodique

Le phrasé prosodique est conçu par la phonologie des domaines comme une structure hiérarchique de constituants qui consolide sous la forme d'une architecture multiniveaux la modélisation des interfaces entre la phonologie, d'une part, et la syntaxe, la sémantique, la pragmatique, d'autre part. Nous avons particulièrement insisté sur la nécessité plurielle de poursuivre un travail de réflexion, de réaliser des analyses empiriques et de recueillir des informations destinées à établir la hiérarchie prosodique pour la langue russe, mais aussi de recueillir des informations sur l'emplacement des frontières prosodiques sous l'influence des contraintes linguistiques. Précisons également que le phrasé prosodique fournit un cadre qui assure la distinction entre les différentes forces de frontières, cette distinction se révélant importante en synthèse vocale (Collier et Sanderman 1996).

Nous nous sommes surtout intéressée au constituant du syntagme intonatif qui a fait ainsi l'objet d'une étude en perception et d'une modélisation acoustique. Précisons que certaines approches (Gussenhoven 1990, par exemple) excluent ce constituant de la hiérarchie prosodique ; d'autres auteurs encore attachent moins d'importance à la problématique de la structuration de la parole en termes d'unités intonatives dans le contexte d'une influence prépondérante exercée par des facteurs discursifs et par la structure informationnelle. Les résultats de notre étude révèlent l'intérêt d'appliquer une démarche trans-linguistique lorsque l'on se situe dans la perspective formelle. En effet, nos données suggèrent que dans le domaine acoustique des indices similaires participent à l'actualisation du phrasé prosodique, au moins en russe et en français (cette observation est notamment confirmée lors de l'observation des stratégies mises en oeuvre par des auditeurs dans des conditions desémentisées).

En adoptant une démarche descendante, nous avons analysé le rôle des contraintes syntaxiques et discursives sur la perception du phrasé prosodique dont nous avons tenté de rendre compte à l'aide du formalisme de la théorie de l'optimalité. Nous avons pu constater que les contraintes ancrées syntaxe assurent environ 65 % de prédictions correctes de la structuration du flux de discours en termes de syntagmes intonatifs. Ce résultat incite à deux observations : premièrement, la démarche prédictive fondée sur des contraintes ancrées syntaxe peut être améliorée, notamment à travers l'implémentation des contraintes discursives (mais aussi rythmiques, omises dans la présente étude). En second lieu, le niveau de la performance du modèle syntaxique étant non négligeable, nous y voyons une sollicitation à l'affinement du principe de parallélisme de Pechkovskij (1928) ou encore de l'approche de compétition (Blache et Di Crisot 2002). En effet, les approches citées considèrent que les désaccords structurels sont optionnels ou encore qu'ils reposent sur le degré d'ambiguïté des niveaux concernés. Cependant, il est difficile de supposer que 65% des contextes soient ambigus.

En parallèle avec la démarche descendante, nous défendons la position qui consiste à affirmer que toute structure phonologique imposée aux données observables doit être validée et interprétée du point de vue phonétique. D'autant plus que le phrasé prosodique paraît fournir un cadre de gestion des dimensions phonétique et acoustique de la parole. Dans cette perspective nous avons entrepris des analyses visant une modélisation acoustique des frontières prosodiques perçues. Sur la base des indices acoustiques retenus, nous avons développé un classificateur entraîné à partir des annotations issues de

l'expérience de perception, et nous l'avons appliqué à l'annotation de la totalité de notre corpus en termes de frontières intonatives. Le classificateur permet d'appréhender l'impact des indices globaux et locaux sur la prédiction du phrasé. La démarche prédictive révèle notamment un lien entre les phénomènes de groupement et de proéminence (d'autres analyses menées corroborent à cette hypothèse). Ce lien nous paraît également plausible dans une perspective psycholinguistique, notamment au regard de la gestion des ressources attentionnelles, dans la mesure où la perception d'une rupture prosodique peut initier le processus de re-analyse du matériau linguistique stocké dans la mémoire de travail (Fodor *et al.* 1974). L'indice de « l'amplitude totale » que nous avons retenu pour rendre tangible le lien entre les deux catégories de phénomènes mentionnés, à savoir le groupement et la proéminence, se révèle peu approprié pour l'analyse de la parole spontanée. En effet, le signal acoustique contient plusieurs types de marques de spontanéité, telles que des hésitations, des allongements et des pauses non-structurelles, qui influent sur les tendances constatées à propos du lien de cet indice et l'organisation structurelle du message. Ce résultat incite à une réflexion sur le rôle de l'indice d'intensité dans le signalement des phénomènes de groupement et en particulier de proéminence (Kochansky *et al*. 2005) et sur son intégration dans le modèle global des frontières prosodiques, modélisées en termes de faisceaux d'indices acoustiques.

Articulation « forme ~ fonction »

Dans une seconde partie de la thèse, nous avons entrepris une modélisation probabiliste afin d'appréhender des rapports entre une représentation formelle et une représentation fonctionnelle : le concept de probabilité atténue la contrainte de la grammaticalité des productions langagières et il permet de ce fait de capturer la variabilité propre au discours humain. À cette étape de notre recherche nous nous sommes résolument placée dans une perspective ascendante : premièrement, la grammaire probabiliste est fondée sur l'annotation symbolique semi-automatique des enregistrements ; ensuite, nous cherchons à maîtriser la variabilité dans les annotations INTSINT à travers la méthodologie de construction des agrégations à partir des données acoustiques. Les analyses conduites dans cette voie souffrent de la taille du corpus d'étude, notamment dans la quête de la grammaire probabiliste, mais, selon notre opinion, elles indiquent une méthodologie intéressante à tester sur des données plus volumineuses. Les résultats obtenus pointent vers la nécessité de maîtriser la dimension qui se rapporte à l'alignement des points cibles définis par MOMEL avec la chaîne syllabique : dans cette perspective, l'adoption du concept de l'unité tonale (dont nous avons discuté dans le chapitre IV) nous paraît être une piste fructueuse.

L'étude des agrégations de contours mélodiques qui se ressemblent du point de vue acoustique révèle une grande variabilité intraclasse : nous avons constaté des variations à la fois pour le point de départ de la courbe et pour la hauteur du pic. Cette variabilité peut provenir de deux sources : a) des effets paralinguistiques et b) des effets de la structure hiérarchique (notamment dans le contexte des études récentes sur le domaine du focus, Xu *et al.* 2004). Dans le prolongement de cette analyse, les aspects perceptifs associés à la variabilité constatée ainsi que les classes d'équivalence obtenues nécessitent une validation empirique. Nous envisageons également une étude en production, en proposant aux sujets une tâche d'imitation (protocole proposé par Pierrehumbert et Steele 1989).

Dans sa globalité, nos résultats démontrent l'intérêt qu'il y a à mener une étude de la formalisation de l'intonation et de l'organisation prosodique en général à partir des corpus oraux. Cependant, une démarche complémentaire s'impose : le corpus de l'étude nécessite d'être doté d'une annotation appropriée, effectuée par plusieurs juges/experts, qui portera essentiellement sur la fonction prosodique. Cependant, le schéma d'annotation doit être soumis à une évaluation de la concordance des annotations : l'objectif de cette démarche consiste à assurer une reproductibilité suffisante des annotations. En même temps, la méthodologie des annotations pose le problème des jugements métalinguistiques, de leur forte dépendance des présupposés théoriques sous-jacents, partagés ou non par les participants. Dans cette perspective, l'établissement d'un protocole objectif d'analyse et d'annotation est prioritaire. De même, tout au long de ce travail nous défendons les principes d'une démarche conjuguée qui tiendra compte des données perceptives mais qui modélisera également des phénomènes prosodiques moyennant des faisceaux d'indices acoustiques.

Au terme de cette étude nous avons également constaté que plusieurs points nécessitent des recherches expérimentales approfondies si l'on souhaite avancer de façon significative dans l'élaboration d'une description formelle de l'intonation du parler russe spontané. Nos études futures porteront essentiellement sur la question du statut du constituant du pied et sur la problématique du nombre de niveaux de la hiérarchie prosodique qui reçoivent un marquage intonatif ; nous poursuivrons le travail sur l'élaboration d'un vecteur d'indices acoustiques associés au phrasé prosodique. En même temps, des réflexions d'ordre général et des analyses empiriques associées s'imposent : leurs sujets éventuels seront la quête d'un indice de la proéminence perceptive en parole spontanée (un indice composite ?) ou encore le cadre d'exploration du fonctionnement et du rôle de l'indice d'intensité. Nous projetons également d'entreprendre des études ciblées sur l'application du formalisme de la théorie de l'optimalité à un échantillon élargi du corpus. Ces propositions conduisent à envisager l'élaboration d'un projet de recherche pluridisciplinaire et celle d'un cadre de travail exploitable dans le paradigme fécond du modèle d'analyse par synthèse.

RÉFÉRENCES BIBLIOGRAPHIQUES

Abercrombie, D. (1964). Syllable quantity and enclitics in English. In Abercrombie, D., Fry, P., MacCarthy, N. et Trim, J. (eds): *In Honor of Daniel Jones*. London: Longman, 216-222.

Abercrombie, D. (1973). A phonetician's view of verse structure. In W.E; Jones & J. Laver (eds.), *Phonetics in Linguistics: a book of readings.* London: Longman.

Allen, J., Hunnicutt, M.S., and Klatt, D. (1987), *From Text to Speech. The MITalk System*. Cambridge University Press, Cambridge, England.

Altenberg, B. (1987). *Prosodic Patterns in Spocken English*. Lund University Press, Sweden.

Anttila, A. (1997). Deriving Variation from Grammar. In Hinskens, F., van Hout, R. & Wetzels, W.L. (eds.), *Variation, Change, and Phonological Theory*. Amsterdam: John Benjamins, 35-68.

Arvaniti, A., D. R. Ladd & I. Mennen (2000). What is a starred tone? Evidence from Greek. In Michael Broe & Janet Pierrehumbert (eds), *Papers in Laboratory Phonology V*, Cambridge: Cambridge University Press, 119-131.

Astesano, C. (2001) Rythme et accentuation en français. Invariance et variabilité stylistique. Coll. Langue et Parole. Paris, France : L'Harmattan, 2001, 336 p.. Coll. Langue et Parole. Paris, France : L'Harmattan, 2001, 336 p

Astruc, L.(2003). Sentence external elements in Catalan. Catal Journal of linguistics, **2**, 15-31.

Aubergé, V. (1992). Developing A Structured Lexicon For Synthesis Of Prosody. In Bailly, G., Benoit, Ch., Sawallis, Th. R. (eds.), *Talking Machines. Theories, Models, And Designs,* 307-321.

Auran, C. (2004). *Prosodie et anaphore dans le discours en anglais et en français: cohésion et attribution référentielle*. Thèse de doctorat en Sciences du langage, phonétique, Université d'Aix-Marseille 1, Université de Provence, Aix-en-Provence.

Auran C. & Di Cristo A. (2003). Towards Automatic Annotation of Temporal Features in Discourse: the Case of Syllabic Duration in Spontaneous French. *Proceedings of the 15th ICPhS*, August 3-9 2003, Barcelona.

Aust, H., Oerder, M., Seide, F. & Steinbiss, V. (1995). The Philips automatic train timetable information system. *Speech Communication,* **17**, 249-262.

Austin, L.J. (1962). *How to do things with words*. Oxford: Oxford University Press.

Avanessov, R.I. (1954). Аванесов, Р.И. О слогоразделе и строении слога в русском языке. *Вопросы языкознания*, **6**.

Avesani, C., Hirschberg, J. & Prieto, P. (1995) The intonational disambiguation of potentially ambiguous utterances in English, Italian and Spanish. In *Proceedings of the XIIIth ICPhS,* Stockholm, 174-177.

Aylett, M. (2005). Extracting the acoustic features of interruption points using non-lexical prosodic analysis. *Proceedings of the DiSS-2005conference*, 17-20.

Bachenko, J. & Fitzpatrick, E. (1990). A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, **16(3),** 155-170.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, **42**, 1-22.

Barry, W.J. (1995). Phonetics and phonology of speaking styles. *Proceedings of the 13th ICPhS*, **2**, 4-10.

Beckman, M.E. (1986). *Stress and Non-Stress Accent*. Coll. Netherlands Phonetic Archives, 7. Dordrecht, Pays-Bas : Foris Publications.

Beckman, M.E. (1996). The parsing of prosody. *Language and Cognitive Processes*, **11**, 17-67.

Beckman, M.E. (1997). A typology of spontaneous speech. In Y. Sagisaka, W. N. Campbell & N. Higuchi (eds.) *Computing Prosody*, New York: Springer-Verlag, 7-26.

Beckman, M.E. & Ayers E.G. (1997). *Guidelines for ToBI labelling*. Retrieved 2002-01-11, from http://www.ling.ohiostate.

Beckman, M.E. & Edwards, J. (1990). Lengthenings and Shortenings and the Nature of Prosodic Constituency. In Kingston, J., Beckman, M.E. (eds), *Papers In Laboratory Phonology I. Between The Grammar And Physics Of Speech*, 1990, 152-178

Beckman, M.E. & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, Cambridge : Cambridge University Press, 7-33.

Beckman, M.E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (ed.) *Prosodic Typology – The Phonology of Intonation and Phrasing.*

Beckman, M.E. & Pierrehumbert, J.B. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook,* **3**, 255-309.

Berg, R. van den, Gussenhoven, C. & Rietveld, A. (1992). Downstep in Dutch: Implications for a model. In Docherty D.J. & D. R. Ladd (eds), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press. 335-359.

Blaauw, E. (1994). The Contribution of Prosodic Boundary Markers to the Perceptual Difference between Read and Spontaneous Speech. *Speech Communication*, **14**, 359-375.

Blache, Ph. & Di Cristo, A. (2002). Variabilité et dépendances des composants linguistiques. *Actes de la conférence Traitement Automatique du Langage Naturel (TALN-2002)*.

Black, A. & Hunt, A. (1996). Generating FO contours from ToBI labels using linear regression. Proceedings of ICSLP-1996, Philadelphia, Penn., **3**, 1385-1388.

Black, A. & Taylor, P. (1994). Assigning intonation elements and prosodic phrasing for English speech synthesis from high level linguistic input. *Proceedings ICSLP-94*, Yokohama, Japan, **2**, 715-718.

Black, A. & Taylor, P. (1997). Automatically clustering similar units for unit selection in speech synthesis. *Proceedings of Eurospeech 97*, Rhodes, Greece, **2**, 601-604.

Blanche-Benvéniste, Cl. & Jeanjean, (1987). *Le français parlé. Transcription et édition*. Paris, Publication du Trésor de la Langue Française.

Blevins, J. (1995). The Syllable in Phonological Theory. In Goldsmith, J.A. (éd.), *The Handbook Of Phonological Theory*, 206-244.

Boersma, P. & Weenink, D. (2005). *Praat: a system for doing phonetics by computer*, http://www.fon.hum.uva.nl/praat/

Bolinger, D.L. (1958). A theory of pitch accent in English. *Word*, **14**, 109-149.

Bolinger, D.L. (1961). Generality, gradience, and the all-or-none. The Hague-Mouton.

Bolinger, D.L. (1972). Accent is predictable (if you're a mind-reader). *Language*, **48**, 633-644.

Bolinger, D.L. (1981). *Two Kinds of Vowels, Two Kinds of Rhythm*. Bloomington, USA: IULC Publications.

Bolinger, D.L. (1989). *Intonation and its uses: Melody in grammar and discourse*. London: Edward Arnold.

Bondarko *et al.* (1993). Бондарко, Л.В., Кузнецов, В.И., Степанова, С.Б. Теоретические и практические проблемы транскрипции текста. В книге Т.М. Николева (отв. ред.), *Вопросы фонетики I*, 8-20.

Bondarko *et al.* (2000). Л.В. Бондарко (отв. ред.), *Фонология речевой деятельности*. Изд-во СпбГУ, 2000.

Bondarko, L.V. (1981). Бондарко, Л.В. Фонетическое описание языка и фонологическое описание речи. Л., 1981.

Bondarko, L.V. (1998). Бондарко, Л. В., Фонетика современного русского языка. Издательство СПбГУ, 1998.

Bondarko, L.V. (2000). Бондарко, Л.В. Фология речевой деятельности как предмет фонетических исследований. В книге Л.В. Бондарко (отв. ред.), *Фонология речевой деятельности*. СПбГУ, 6-46.

Bondarko, Verbitskaja, Gordina *et al.* (1974). Бондарко, Л.В., Вербицкая, Л.А., Гордина, М.В., Зиндер, Л.Р., Касевич, В.Б. Стили произношения и типы произнесения. *Вопросы языкознания*, **2**.

Bouzon, C. (2004). Rythme et structuration prosodique en anglais britannique comtemporain. Thèse de Doctorat en phonétique, Université de Provence, Aix-Marseille I, Aix-en-Provence, France

Boyanus, S.C. (1936). The main types of Russian intonation. In *Proceedings of the 2nd ICPhS*, 110-113.

Boyanus, S.C. (1955). *Russian Pronunciation*. London

Braunschweiler, N. (2003). ProsAlign - The Automatic Prosodic Aligner**.** *Proceedings of the 15th ICPhS*, August, 3-9, 2003, Barcelona.

Brown, G., Anderson, A., Shillcock, R.C. & Yule, G. (1984). *Teaching Talk*. Cambridge: Cambridge University Press.

Brown G. & Yule G. (1983). *Discourse analysis*, Cambridge University Press.

Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective*. Travaux de l'Institut de Linguistique de Lund, **12**, Institut de Linguistique de Lund, Lund, Suède.

Bruce, G. (1998). *Allmän och svensk prosodi.* Praktisk lingvistik 16. Lund: Department of Linguistics and Phonetics, Lund University.

Bruce, G., Granström, B., Gustafson, K., & House, D. (1992). Interaction of F0 and duration in the perception of prosodic phrasing in Swedish. *Nordic Prosody VI. Papers from a Symposium, Stockholm,* August 12-14 1992, 7-22.

Bryzgounova, E. (1963). Брызгунова Е. А Практическая фонетика и интонация русского языка. // М., 1963. 307 с.

Bryzgounova E. (1977). Е.А. Брызгунова. *Звуки и интонация русской речи. (Les sons et l'intonation du parler russe.)* Moscou.

Buhmann, J., Caspers, J., van Heuven, V., Hoekstra, H., Martens, J-P., Swerts, M., (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by no-expert transcribers in the spoken Dutch corpus. *Proceedings of LREC 2002*, Paris: ELRA, 779-785.

Butcher, A. (1981). *Aspects of the speech pause: phonetic correlates and communicative functions.* Unpublished doctoral dissertation.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, **57**, 3-16.

Byrd, D. & Saltzman, E.(1998). Intergestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, **26**, 173-199.

Cahn, J. (1999). A Computational Memory And Processing Model For Prosody. PhD thesis, MIT.

Calhoun, S. (2005). Overloaded ToBI and what to do about it: An Argument for Function-Based Phonological Intonation Categories. Présenté lors de la conférence *Linguistics Postgraduate Conference*, University of Edinburgh, May 2004 (disponible sur le site http://homepages.inf.ed.ac.uk/s0199920/research.html).

Cambier-Langeveld, T. & Turk, A. E. (1999). A comparison of accentual lengthening in Dutch and English. *Journal of Phonetics*, **27**, 255-280.

Campbell, N. (1992). *Multi-level Timing in Speech*, Ph.D. Thesis, Sussex University.

Campione, E. (2001). *Etiquetage prosodique semi-automatique de corpus oraux : algorithmes et méthodologie.* Thèse de doctorat. Aix-en-Provence: Université de Provence.

Carlson, R., Granström, B., Heldner, M., House, D., Megyesi, B., Strangert, E., & Swerts, M. (2002). Boundaries and groupings - the structuring of speech in different communicative situations: a description of the GROG project. *TMH-QPSR, 43 - Fonetik 2002.*

Carlson, R., Hirschberg, J. & Swerts, M. (2005). Cues to upcoming Swedish prosodic boundaries: subjective judgment studies and acoustic correlates. *Speech Communication*, **46(3/4),** 326-333.

Carr, Ph. (1999). *Phonology.* Coll. Modern Linguistics Series. Londres, Grande Bretagne : Macmillan Press Ltd.

Chalonova, K. (1999). Шалонова, К. Б., *Автоматическое транскрибирование как способ отображения вариативности речи. Моделирование произносительных особенностей на материале русской речи*. Дис. … канд. филол. наук : 10.02.19 СПб., 1999.

Cho, H.-S. & Hirst, D. (2006). The contribution of silent pauses to the perception of prosodic boundaries in Korean read speech. *Proceedings of Speech Prosody-2006*.

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Chomsky, N. (1995). Bare phrase structure. In G. Webelbuth (ed.), *Government and binding theory and the minimalist program*, Blackwell publishers, 383–439.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. Coll. Studies in Language. New York, NY, USA : Harper & Row

Clark, H. H. & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, **37**, 201-242.

Clark, R.A.J. (1999). Using prosodic structure to improve pitch range variation in text to speech synthesis. *Proceedings of the. XIVth ICPhS,* **1**, 69-72.

Clements, G.N. & Keyser, S.J. (1983). From CV Phonology : A Generative Theory of the Syllable. In : Goldsmith, J.A. (éd.) : *Phonological Theory. The Essential Readings*. Malden, MA, USA : Blackwell Publ., 185-200

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20,** 37-46.

Coleman, J. S. (1990). YorkTalk: "Synthesis-by-rule" without segments or rules. In P*roceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France, 25 sepetmbre 1990.

Cooper, W.E. & Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, USA: Harvard University Press.

Couper-Kuhlen, E. (1986). *An Introduction to English Prosody*. Tübingen, RFA : Niemeyer.

Couper-Kuhlen, E. & Selting, M. (eds) (1996). *Prosody in conversation*, Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, **24(1)**, 87-185.

Crystal, T.H. & House, A.S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, **88**, 101-112

Cutler, A. (1984). Stress and Accent in Language Production and Understanding. In Gibbon, D., Richter, H. (eds), *Intonation, Accent and Rhythm*. Studies in Discourse Phonology, 1984, 77-90

Cutler, A. & Clifton, C. E. (1999). Comprehending spoken language: A blueprint of the listener. In C. Brown & P. Hagoort (Eds.), *Neurocognition of Language*. Oxford: Oxford Univ. Press, 123-166.

Cutler, A., Dahan, D. & Donselaar, W. (1997), « Prosody in the Comprehension of Spoken Language : A Literature Review », *Language and Speech,* **40(2)**, 141-201.

Cutler, A. & Darwin, C.J. (1981). Phoneme monitoring reaction time and preceding prosody: Effects of stop closure duration and fundemental frequency. *Perception & Psychologysics*, **28**, 217-224.

Cutler, A. & Foss, D.J. (1977). On the Role of Sentence Stress in Sentence Processing, *Language and Speech*, **20**, 1-10.

Cutler, A. & Ladd, D.R. (1983). *Prosody : Models and Measurements*. Coll. Springer Series in Language and Communication, **14**. Berlin, Allemagne : Springer.

Cutler, A., Mehler, J., Norris, D. & J. Segui (1986) The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language,* **25,** 385-400.

D'Imperio, M. (2002). Tonal alignment, scaling and slope in Italian question and statement tunes. *TIPA - Travaux Interdisciplinaires du Laboratoire Parole et Langage*, **21**, 25-44.

D'Imperio, M. (2005). La Phonologie de Laboratoire : finalités et quelques applications. In Nguyen, N., Wauquier-Gravelines, S., Durand, J. (eds.), *Phonologie et phonétique: Forme et substance*. Paris: Hermès, 241-264.

D'Imperio, M. & House, D. (1997). Perception of questions and statements in Neapolitan Italian. In *Proceedings of Eurospeech '97*, **1**, 251-254.

Dainora, A. (2001). *An empirically based probabilistic model of intonation in American English*. University of Chicago Ph.D. Dissertation.

Danieli, M., Garrido, J.M., Moneglia, M., Panizza, A., Quazza, S., Swerts, M. (2004). Evaluation of consensus on the annotation of prosodic breaks in the Romance corpus of spontaneous speech "C-ORAL-ROM". In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, 2004, 1513-1516

Dankovičová, J. (1997). The domain of articulation rate variation in Czech. *Journal of Phonetics,* **25**, 287-312.

de Pijper, J.R. (1983). *Modelling British English Intonation.* Coll. Netherlands Phonetic Archives, 3. Dordrecht, Pays-Bas : Foris Publications.

de Pijper, J.R. (1997). High quality message-to-speech generation in a practical application. In: J. van Santen, R. Sproat, J. Olive, and J. Hirschberg (eds.), *Progress in speech synthesis*, New York: Springer-Verlag, 575-586.

de Pijper, J.R. & Sanderman, A.A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America,* **96**, 2037-2047.

Delais-Roussarie, E. (1995). *Pour une approche parallèle de la structure prosodique. Etude de l'organisation prosodique et rythmique de la phrase française.* Thèse de Doctorat en Sciences du langage, Université de Toulouse-Le Mirail, Toulouse

Delais-Roussarie, E. (2000). Vers une nouvelle approche de la structure prosodique, *Langue Française,* **126**, 92-112.

Delais-Roussarie, E. (2003). Quelques outils d'aide à la transcription et à l'annotation de données audio pour constituer des corpus oraux. In E.Delais-Roussarie & J. Durand (eds.), *Corpus et variation en phonologie du français. Méthodes et analyses*. Coll. Interlangues linguistique et didactique. Toulouse, France : Presses universitaires du Mirail.

Delais-Roussarie, E. (2005). Interface Phonologie / Syntaxe : des domaines phonologiques à l'organisation de la Grammaire. In *Phonologie et Phonétique: Approches actuelles*, J. Durand, N. Nguyen, V. Rey et S. Wauquier-Gravelines (eds). Editions Hermès.

Delais-Roussarie, E. & Fougeron, C. (2004). Les notions de Domaine et de constituant en phonétique et en Phonologie. *Journées d'Etudes Linguistiques* (JEL 2004), Nantes, Mai 2004.

Delattre, P. (1966). Les dix intonations de base du français. *The French Review*, **9(1),** 1-14.

Di Cristo, A. (1978). *De la Microprosodie à l'Intonosyntaxe*, Thèse du doctorat d'état, Université de Provence.

Di Cristo, A. (1985). De la microprosodie à l'intonosyntaxe. Aix-en-Provence, France : Publications de l'Université de Provence.

Di Cristo, A. (1999). Vers une modélisation de l'accentuation du français : 1ière et 2ème parties. *French Language Studies*, **9 & 10**, 1999, 143-179, 27-44.

Di Cristo, A. (2000). « Interpréter la prosodie ». *Actes des XXIIIe Journées d'Etude sur la Parole*. Aussois (13-23 juin 2000), 13-29.

Di Cristo, A. (2000b).La problématique de la prosodie dans l'étude de la parole dite spontanée. Revue Parole, **15-16**, 189-250.

Di Cristo, A. (2004). La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonction. *Travaux Interdisciplinaires du* Laboratoire Parole et Langage *d'Aix-en-Provence*, **67**, 167-212.

Di Cristo, A. & Di Cristo, P. (2001). SYNTAIX : une approche métrique-autosegmentale de la prosodie. *Traitement Automatique des Langues* (n° spécial sur la synthèse), **42(1),** 69-111.

Di Cristo, A. & Hirst, D. (1986). Modelling French micromelody: analysis and synthesis. *Phonetica*, **43 (1),** 11-30

Di Cristo, A. & Hirst, D. (1996). Vers une typologie des unités intonatives du français. *Actes des XXIèmes Journées d'Etude sur la Parole* (Avignon, juin 1996), 219-222.

Dik, S. (1997). *The theory of Functional Grammar*. Mouton-De Gruyter.

Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M. (1996). Glottalization of word-initial vowels as fonction of prosodic structure. *Journal of Phonetics*, **24**, 223-244.

Dixon, R. & Maxey, H. (1968). Terminal analog synthesis of continuous speech using the diphone method of segment assembly. *IEEE Transactions on Audio and Electro-acoustics AU-16*, 40-50.

Duez, D. (1985). Perception of Silent Pauses In Continuous Speech. *Language and Speech*, oct-déc. 1985, 377-390

Duez, D. (1993). Acoustic Correlates of Subjective Pauses. *Journal of Psycholinguistic Research*, **22(1)**, 21-39.

Dupoux, E., Pallier, C., Sebastián-Gallés, N. & J. Mehler (1997) A destressing "deafness" in French? *Journal of Memory and Language,* **36** , 406-421.

Durand, J. & Lyche, C. (2001). Des règles aux contraintes en phonologie générative. *Revue québecoise de linguistique,* **30**, 91-154.

Durand, J., Laks, B. & Lyche, C. (2002). Phonologie du français contemporain, usages, variétés et structure. *Bulletin PFC*, **1**, Toulouse-Le Mirail.

Dusterhoff, K. E., Black, A. W. & Taylor, P. (1999). Using decision trees within the tilt intonation model to predict F0 contours. *Proceedings of Eurospeech-1999*, 1627-1630.

Eefting, W. (1991). The effect of "information value"' and "accentuation" on the duration of Dutch words, syllables, and segments. *Journal of the Acoustical Society of America*, **89(1)**, 412-424.

Eskénazi, M. (1994). Trends in Speaking Styles Research. *Europseech '93*, 501-508.

Eskenazi, M. & Lacheret-Dujour, A. (1991). Exploration of individual strategies in continuous speech. *Speech Communication*, **10**, 249-264.

Fant, G. (1960). Acoustic theory of speech production. Mouton, La Haye.

Fant, G., Kruckenberg, A. & Ferreira, J. B. (2003). Individual vairations in pausing. A study of read speech. *PHONUM 9*, University of Umeå, 193-196.

Fernandez-Cruz, R. (2000). *Analyse acoustique et phonologique du portugais parlé par les communautés noires de l'Amazonie (Brésil)*. Thèse de Doctorat en phonétique, Université de Provence, Aix-Marseille I, Aix-en-Provence, France, 2000

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen*. **7**, 179-188.

Fletcher, J., Harrington, J. & J. Hayek (1994). Phonemic vowel length and prosody in Australian English. *Proceeding of the 5th Australian International Conference on Speech Science and Technology*, Perth, 656-661.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The Psychology of Language*. York: McGraw Hill.

Fonagy, I. (1980). *L'accent en français contemporain.* Coll. Studia Phonetica, 15. Ottawa, Canada : Marcel Didier, Inc.

Fougeron, I. (1989). *Prosodie et organisation du message, Analyse de la phrase assertive en russe contemporain,* Paris, Klincksieck, Collection linguistique, Société de linguistique de Paris LXXVI.

Fougeron, I. & Ventsov, A. (2002). И. Фужерон, А.В. Венцов. Сегментация французского звучашчего текста при его восприятии. *Вопросы фонетики*, 4, 128-147.

Fox, A. (2000). *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Coll. Oxford Linguistics. Oxford, Grande Bretagne : Oxford University Press.

Fraisse, P. (1974). Psychologie du rythme. Paris, France : PUF, Presses Universitaires de France.

Frazier, L., Clifton, Ch. Jr. & Carlson, K. (2004). Don't break, or do: prosodic boundary preferences. *Lingua*, **114(1)**, 3-27.

Frota, S., Vigário, M., Martins, F. (2002). Language discrimination and rhythm classes: evidence from Portuguese. *Proceedings of Speech Prosody-2002*, 319-322.

Fry, D.B. (1955). Duration and Intensity as Physical Correlates of linguistic Stress. *Journal of the Acoustical Society of America*, **27**, 765-768.

Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech*, **1**, 26-152.

Fudge, E.C. (1969). Syllables. In Goldsmith, J.A. (éd.) : *Phonological Theory. The Essential Readings*. Malden, MA, USA : Blackwell Publ., pp.370-391.

Fujisaki, H. (1983), Dynamic characteristics of voice fundamental frequency in speech and singing. In *The production of speech*, Heidelberg: Springer-Verlag, 39-55.

Garde, P. (1968). *L'accent*. Coll. SUP, Le Linguiste, 5. Paris, France : PUF, Presses Universitaires de France.

Gårding, E. (1983). A Generative Model of Intonation. In: Cutler, A. and D. R. Ladd (eds), *Prosody: Models and Measurements*, 11-25. Germany: Springer-Verlag.

Gasparov, M. (1984). Гаспаров М.Л. Очерк истории русского стиха. М.: Наука, 1984.

Gee, J. & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.

Giot, J. (1977). Études comparatives de syllabes accentuées et pré-toniques du français sur les plans articulatoires et acoustiques. *Travaux de l'Institut de Phonétique de Strasbourg*, **9**, 89-189.

Goldman-Eisler, F. (1968). *Psycholinguistics : Experiments in Spontaneous Speech,*

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis*, **2**, 23-68.

Grabe, E. (1998). Pitch accent realisation in English and German. *Journal of Phonetics,* **26**, 129-144.

Grabe, E., Kochanski, G. & Coleman, J. (2003). Quantitative modelling of intonational Variation. *Proceedings of Speech Analysis and Recognition in Technology, Linguistics and Medicine 2003* (http://www.phon.ox.ac.uk/~esther/Grabe_Kochanski_Coleman.pdf)

Grabe, E. & Post, B. (2002). Intonational Variation in English. *Proceedings of the Speech Prosody 2002 Conference*, April, 11-13th 2002, Aix-en-Provence: Laboratoire Parole et Langage, 343-346.

Grabe, E., Post, B., Nolan, F. & Farrar, K. (2000). Pitch accent realisation in four varieties of British English. *Journal of Phonetics* 28.

Grabe, E., Warren, P. & Nolan, F. (1994). Resolving category ambiguities - evidence from stress shift. *Speech Communication,* **15**, 101-114.

Grammaire russe (1980) : Chvedova, N.Ju. *et al.* (eds.) (1980). *Русская грамматика.* Moskva: ANSSSR

Grammont, M. (1933). *Traité de phonétique*. Paris: Delagrave

Greenberg, S. (2002). From here to utility - Melding phonetic insight with speech technology. In Barry, W. & W. Domelen (eds.), *Integrating Phonetic Knowledge with Speech Technology*, Dordrecht: Kluwer (disponible sur le site http://www.icsi.berkeley.edu/~steveng/)

Grice, M. (1995). Leading tones and downstep in English. *Phonology*, **12,** 183-233.

Grice, M., Ladd, D.R. & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology,* **17,** 143-185.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J. & Batliner, A. (1996). Consistency in Transcription and Labelling of German Intonation with GToBI. *Proceedings of ICSLP-96,* Philadelphia, 1716-1719.

Grönnum, N. (1992). *The Groundworks of Danish. Intonation*. Museum Tusculanum Press. University of Copenhagen.

Grover C., Hitzeman, J. & Moens, M. (1995). Algorithms for analysing the temporal structure of discourse. In *Sixth International Conference of the European Chapter of the Association for Computational Linguistics*: ACL.

Gumpertz, J.J. (1971). *Language in Social Groups*. Stanford: Stanford University Press.

Gussenhoven, C. (1983). A three-dimensional scaling of nine English tones. *Journal of Semantics,* **2**, 186-204.

Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Coll. Publications in Language Sciences, 16. Dordrecht, Pays-Bas : Foris Publications.

Gussenhoven, C. (1988). Adequacy in intonation analysis: The case of Dutch. In H. van der Hulst & N. Smith (eds). *Autosegmental studies on pitch accent*. Dordrecht: Foris. 95-121.

Gussenhoven, C. (1990). Tonal association domains and the prosodic hierarchy in English. In S. M. Ramsaran (ed.), *Studies in the pronunciation of English. A commemorative volume in honour of A.C. Gimson*. London: Routledge. 27-37.

Gussenhoven, C. (1991). The English Rhythm Rule as an accent deletion rule. *Phonology,* **8,** 1-35.

Gussenhoven, C. (2002a). Intonation and interpretation: Phonetics and Phonology. In *Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody*. Aix-en-Provence, 47-57.

Gussenhoven, C. (2002b). Phonology of intonation. State-of-the-Article. *GLOT International,* **6 (9/10),** 271-284.

Gussenhoven, C. (2005). Transcription of Dutch Intonation. In Jun, Sun-Ah (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 118-145

Gussenhoven, C. & Driessen, W. (2004). Explaining two correlations between vowel quality and tone: The duration connection. In B. Bel & I. Marlien (eds) *Speech Prosody 2004*. Nara, Japan. 179-182.

Gussenhoven, C., B.H. Repp, A. Rietveld, W.H. Rump & J. Terken (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America,* **102**, 3009-3022.

Gussenhoven, C. & Rietveld, T., (1988). Fundamental frequency declination in Dutch : testing three hypotheses. *Journal of Phonetics,* **16** , 355-369.

Gussenhoven, C. & Rietveld, T. (1992). Intonation contours, prosodic structure, and preboundary lenghthening. *Journal of Phonetics* 20, 283-303.

Gussenhoven, C. & Rietveld, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech,* **43**, 183-203.

Gvozdev, A. (1949). Гвоздев А.Н. О фонологических средствах русского языка. М., 1949.

't Hart J., Collier, R. & Cohen A. (1990). *A Perceptual Study of Intonation. An Experimental Phonetic Approach to Speech Melody*. Coll. Cambridge Studies in Speech Sciences and Communication. Cambridge, Grande Bretagne : Cambridge University Press

't Hart, J. & Collier, R. (1975). The role of intonation in speech perception. In Cohen, A. et Noteboom, S.G. (eds): *Structure and Process in Speech Perception*, Symposium for Scandinavian Students of Phonetics, Umeä, Suède, 11, 4-6 août, Coll. Communication and Cybernetics n°11. Berlin, Allemagne, 107-123

't Hart, J. (1984). A phonetic approach to intonation : from pitch contours to intonation patterns. In Gibbon, D. & Richter, H. (eds.), *Intonation, Accent and Rythm :Studies in Discourse Phonology,* Berlin, New-York : de Gruyter, 193-202.

Halle, M. & Vergnaud, J.-R. (1987). An Essay on Stress. Coll. Current Studies in Linguistics, 15. Cambridge, MA, USA : The MIT Press

Halliday, (1967a). *Intonation and grammar in British English*. The Hague:Mouton.

Halliday, M.A.K. (1970). *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.

Hallyday M.A.K. (1967b). Notes on transitivity and theme in English. *Journal of Linguistics,* **3**, 199-244.

Hansson, P. (1998). Pausing in Spontaneous Speech. *Proceedings FONETIK 98,* The Swedish Phonetics Conference, May 27-29 1998, 158-161.

Hansson, P. (2001). Constraints on prosodic phrasing in spontaneous speech. In: Svantesson, Jan-Olof (eds.), *Working Papers* 48, 87-100. Lund: Department of Linguistics and Phonetics, Lund University.

Hansson, P. (2003). *Prosodic phrasing in spontaneous Swedish*. Doctoral dissertation. Travaux de l'institut de linguistique de Lund 43. Lund: Department of Linguistics and Phonetics, Lund University.

Harrington, J., Fletcher, J. & C. Roberts (1995). Coarticulation and the accented/unaccented distinction: evidence from jaw movement data. *Journal of Phonetics*, **23**, 305-322.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (eds), *Rhythm and meter*, Academic Press, Orlando, 201-260.

Hayes, B. (1995). *Metrical Stress Theory. Principles and Case Studies*. Chicago, IL.

Heldner, M. (2003**).** On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics,* **31(1)**, 39-62.

Heldner, M., Edlund, J., & Björkenstam, T. (2004). Automatically extracted F0 features as acoustic correlates of prosodic boundaries. *Proceedings of Fonetik 2004*, Stockholm. (disponible sur le site http://www.speech.kth.se/grog/)

Heldner, M. & Megyesi, B. (2003). Exploring the prosody-syntax interface in conversations. *Proceedings of the 15th ICPhS,* August, 3-9, 2003, Barcelona, 2501-2504.

Hermes, D.J., Beaugendre, F. & House, D. (1997). Temporal alignment of accentuation boundaries in Dutch. *Proceedings of the ESCA Tutorial and Research Workshop on Intonation: Theory, Models and Applications, Athens, Greece*, 177-180.

Hirschberg, J. (1990). Accent and Discourse Context: Assigning Pitch Accent in Synthetic Speech. *Proceedings of American Assocation for Artificial Intelligence*, 1990.

Hirschberg, J. (1993). Pitch Accent in Context: Predicting Intonational Prominence from Text. *Journal of Artificial Intelligence*, **63**, 305-340.

Hirschberg, J. & Prieto, P. (1996). Training Intonational Phrasing Rules Automatically for English and Spanish text-to-speech. *Speech Communication*, **18**, 281-290.

Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.W., Cassidy, K.W., Druss, B. & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition,* **26**, 269-86.

Hirst, D. (1977). *Intonative Features. A Syntactic Approach to English Intonation*. Coll. Janua Linguarum, Series Minor, 139. La Haye, Pays-Bas : Mouton & Co.

Hirst, D. (1981). The transcription of english intonation. In P. Leon & M. Rossi (éds.), *Problèmes de prosodie, vol. 1 : Approches théoriques*. Coll. Studia Phonetica, 17. Ottawa, Canada : Didier, 1981, 29-40.

Hirst, D. (1991). Intonation models. In *Proceedings of 12th International Congress of Phonetic Sciences*, vol. 1, pp. 305-310. Aix-en-Provence.

Hirst, D. (2004). Lexical and Non-lexical Tone and Prosodic Typology. In *Proceedings of International Symposium on Tonal Aspects of Language*, 28-30 mars 2004, Beijing, China), 81-88.

Hirst, D. (2005). Form and function in the representation of speech prosody. *Speech Communication (Special Issue)*, **46(3-4),** 334-347.

Hirst, D. & Di Cristo, A. (1993). Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix* 15, 9-24.

Hirst, D. & Di Cristo, A. (1998). *Intonation Systems. A Survey of Twenty Languages*. Cambridge, Grande Bretagne : Cambridge University Press.

Hirst, D. & Di Cristo, A. (1998). A survey of intonation systems. In Hirst & Di Cristo (eds). I*ntonation Systems : A Survey of Twenty Languages.,*1-44.

Hirst, D., Di Cristo, A. & Espesser, R. (2000). Levels of description and levels of representation in the analysis of intonation. In M. Horne (ed.), *Prosody: Theory and Experiment*, Dordrecht : Kluwer Academic Press, 51-87.

Hirst, D., Di Cristo, A., Espesser R. & Delran-Bado, S. (1991). Constituants prosodiques et macro-segmentation du signal: méthodologie et critères d'évaluation. In Haton, J.P. & Rossi, M. (eds.), *Prépublications des Actes du Séminaire Prosodie*, GRECO PRC-HM, Aix-en-Provence, 21-22 mars 1991, 1-8.

Hockett, C. (1955). *A Manual of Phonology*. The University of Chicago Press.

Holm, B. & Bailly, G. (2002). Learning the hidden structure of intonation: implementing various functions of prosody. *Proceeding of Speech Prosody-2002*, 399-402

Horne, M. (1990). Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics*, **28**, 959-981.

Horne, M., Filipsson, M., Johansson, Ch., Ljungqvist, M., Lindstrom, A. (1993). Improving the prosody in TTS systems: morphological and lexical-semantic methods for tracking 'new' vs. 'given' information. In *Prosody-1993*, 208-211.

Horne, M., Strangert, E. & Heldner, M. (1995). Prosodic Boundary Strength in Swedish: Final Lengthening and Silence interval Duration. *Proceedings of ICPhS-9*, 170-173.

House, J. & Youd, N. (1992). Evaluating the prosody of synthesized utterances within a dialogue system. *Proceedings of ICSLP-1992*, Banff, Alberta, 1175-1178.

Howitt, A.W. (2000). *Automatic Syllable Detection of Vowel Landmarks*. PhD Thesis, MIT.

Igarashi, Y. (2002). Так называемая «нейтрализация интонации» - Фонологическое описание русской интонации. In *Bulletin of the Japanese Association of Russian Scholars*, **34**, 15-21.

Inkelas, S. & Zec, D.(1995). Syntax-phonology interface. In J. Goldsmith, ed., *The Handbook of phonological theory*, Blackwell Publishers Inc, Cambridge (MA), 535-549.

Isard, A. & Carletta, J. (1995). Replicability of Transaction and Action Coding in the Map Task Corpus. *AAAI 1995 Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, March 27-29 1995, Stanford University, Palo Alto, CA, USA (disponible sur le site http://www.ltg.ed.ac.uk/~amyi/publications.html).

Jackendoff, R. (1972). *Semantic interpretation in Generative Grammar*. MA:MIT Press, Cambridge.

Jasinskaja, K. (2001). *Prenuclear Pitch Accents in Russian Declaratives and Interrogatives*. Workshop on Tone and Intonation in Europe. June 28-30, 2001. Vitoria-Gasteiz, the Basque Country.

Jassem, W. (1951). *Intonation in Conversational English (Educated Southern British)*. Polish Academy of Science, Warszawa.

Jassem, W. (1952). Stress in Modern English. *Bulletin de la société polonaise de linguistique, fascicule XI.* Poland : Cracovie.

Jensen, J.T. (2004). *Principles of Generative Phonology. An introduction*. Coll. Current Issues in Linguistic Theory, 250. Amsterdam, Pays-Bas : John Benjamins Publishing Company.

Johnson, M.E. (1996). Synthesis of English Intonation Using Explicit Models of Reading and Spontaneous Speech. In Proceedings ICSLP-1996.

Jun, S.-A. (2003). The effect of phrase length and speech rate on prosodic phrasing. *Proceedings of the ICPhS-2003* , Barcelona , Spain . 2-9 août 2003.

Jun, S.-A. & Fougeron, C. (2000). « A phonological model of French intonation », in Botinis, A. (ed.) *Intonation : Analysis, Modeling and Technology*, Dordrecht : Kluwer Academic Publishers, 209-242.

Kahn, D. (1976). *Syllable-Based Generalizations in English Phonology*. Bloomington.

Karaali, O., Corrigan, G. & Gerson, I. (1996). Speech synthesis with neural networks. In *Invited Paper*, *World Congress on Neural Networks*.

Karcevskij, S. (1931). Sur la phonologie de la phrase. In *Travaux du Cercle Linguistique de Prague*, IV.

Kassevitch, V. (2002). Касевич, В.Б. О фонетическом слове. *Вопросы фонетики*, 4.

Kassevitch, V. *et al.* (1990). Касевич, В.Б., Шабельникова, Е.М., Рыбин, В.В. Ударение и тон в языке и речевой деятельности. Л., 1990.

Keating, P. A. (1995a). Segmental phonology and non-segmental phonetics. *Proceedings of the 13th International Congress of Phonetic Sciences*, Stockholm, **3**, 26-32.

Kerbrat-Orecchioni, C. (2001). *Oui, Non, Si :* un trio célèbre et méconnu. Revue électronique *Marges Linguistiques*, **2**, novembre 2001.

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry,* **10(4)**, 421-441

Klabbers, E. (2000). *Segmental and Prosodic Improvements to Speech Generation*, PhD Thesis, Eindhoven University of Technology.

Klabbers, E., Odijk, J., de Pijper, J. & Theune, M. (1996). GoalGetter: From teletext to speech. In: *IPO Annual Progress Report*, **31**, Eindhoven, the Netherlands, 66-75.

Klabbers, E. & van Santen, J. (2004). Clustering of foot-based pitch contours in expressive speech, *Proceedings ISCA Speech Synthesis Workshop 5, Pittsburgh, PA.*

Klabbers, E., van Santen, J. & Wouters, J. (2002). Prosodic factors for predicting local pitch shape. *IEEE 2002 Workshop on Speech Synthesis, Santa Monica, CA,* 11-13 septembre 2002.

Klatt, D.H. (1975) Vowel lengthening is syntactically determined in a connected discourse, *Journal of Phonetics*, **3**, 129-140.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, pp.1208-1221.

Klatt, D.H. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America,* **82 (3),** 737-793.

Knowles, G. (1991). Prosodic labelling: the problem of tone group boundaries. In Johansson, S. & Stenström, A.-B. (eds.). *English computer corpora: selected papers and research guide*. Berlin: Mouton de Gruyter, pp. 149-63.

Kochansky, G., Grabe, E., Coleman, J. & Rosner, B. (2005). Loudness predicts prominence; fundamental frequency lends little. *The Journal of the Acoustical Society of America,* **118(2)**, 1038-1054.

Kochansky, G. & Shih, C. (2003). Prosody modeling with soft templates. *Speech Communication*, **39,** 311–352.

Kodzasov, S. (1996). С.В. Кодзасов. Комбинаторная модель фразовой просодии. Законы фразовой акцентуации. *Просодический строй русской речи*. Москва, 1996.

Kodzasov, S. (1999a). Russian. In H. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe*. Berlin, New York, 852-876.

Kodzasov, S. (1999b). Кодзасов С. В. Уровни, единицы и процессы в интонации. *Проблемы фонетики*, 3. М., Наука. 1999.

Koiso, H., Horuichi, Y., Tutiya, S., Ichikawa, A. & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech,* **41**, 295-321.

Konopczynski, G. (1999). L'acquisition du système prosodique de la langue maternelle et ses implications pour l'apprentissage d'une L2. In *Proceedings of Eurocall'99*, Besançon.

Kouznetsov, V.I. (2000). Кузнецов, В.И. Фонемный состав слова в связной речи. В книге Л.В. Бондарко (отв. ред.), *Фология речевой деятельности*. СПбГУ, 47-73.

Kozhevnikov, V. & Chistovich, L.(1965). *Speech : Articulation and Perception*. Joint Publication Research Service, Washington: U.S. Department of Commerce, 30-543.

Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage.

Krivnova, O. (1999). Automatic synthesis of Russian Speech. In Proceedings of ICPhS-1999, San Francisco, 507-510.

Krivnova, O. & Tchardine, I. (1999). О.Ф.Кривнова, И.С.Чардин. Паузирование при автоматическом синтезе речи. Из кн.: *Теория и практика речевых исследований (АРСО-99)*. Материалы конференции. М., 1999.

Kuosmanen, A. & de Silva, V. (2003). Why don't Russians answer my questions? Finnish students' problems in producing Russian interrogative intonation. *Proceedings of the 15th ICPhS,* 3-9 August 2003, Barcelona, 523-526.

Kurylowicz, J. (1948). Contribution à la théorie de la syllabe. *Bulletin de la Société Polonaise de Linguistique,* **8**, pp.80-104.

Ladd, D.R. (1983). Phonological Features of Intonational Peaks. *Language,* **59(4)**, 721-759.

Ladd, D.R. (1986). Intonational Phrasing: The Case for Recursive Prosodic Structure. P*honology Yearbook,* **3,** 311-340.

Ladd, D.R. (1992). An Introduction to Intonational Phonology (Prosody). In Docherty G. J., Ladd D. R. (eds), *Papers In Laboratory Phonology II. Gesture, Segment, Prosody.,* 321-334.

Ladd, D.R. (1996). *Intonational Phonology*. Coll. Cambridge Studies in Linguistics, 79. Cambridge, Grande Bretagne : Cambridge University Press.

Ladd, D.R. & Campbell, N. (1991). Theories of prosodic structure: Evidence from syllable duration. *Proceedings of the XIIth ICPhS*, Aix-en-Provence, **2**, 290–293.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge : Cambridge University Press.

Laver, J. (1994). *Principles of Phonetics*. Coll. Cambridge Textbooks in Linguistics. Cambridge, Grande Bretagne : Cambridge University Press.

Lea, W.A. (1980). *Prosodic Aids to Speech Recognition*. Prentice Hall.

Leben, W.R. (1973). Suprasegmental Phonology. Bloomington: Indiana University Linguistics Club.

Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. *American Studies in Uralic Linguistics*, vol. 1. Bloomington: Indiana University.

Lehiste, I. (1963). Review of K. Hadding-Koch, Acoustico-phonetic studies in the intonation of Southern Swedish, *Glerup Language,* **39**, 352-360.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa,* **7**, 107-122.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, **5**, pp.253-263.

Lehiste, I. (1979). Perception of sentence and paragraph boundaries. In Lindblom, B. & Oehman, S. (eds), Frontiers of Speech Communication Research, Academic Press, London, 191-201.

Lehiste, I., Olive, J. P. & Streeter, L. A. (1976) Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, **60(5),** 1199-1202.

Lehiste, I. & Wang, W. S-Y. (1977). Perception of sentence boundaries with and without semantic information. In Dressler, W. U., Pfeiffer, O. E., *Phonologica 1976*: Innsbrucker Beiträge zur Sprachwissenschaft. Institut für Sprachwissenschaft der Universität Innsbruck. Innsbruck, 227-283.

Levelt, W.J.M. (1989). *Speaking. From Intention to Articulation*. Cambridge, MA.

Liberman, M. (1975). *The Intonational System of English.* Ph.D. dissertation. MIT, Cambridge/MA.

Liberman, M. & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (eds.), *Language Sound Structure*, Cambridge, MA: M.I.T. Press, 157-233

Liberman, M. & Prince, A.S. (1977). On Stress and Linguistic Rhythm. In Goldsmith, J.A. (éd.): *Phonological Theory. The Essential Readings*. USA : Blackwell Publ., 392-404.

Lickley, R., Schepman, A. & D.R. Ladd (2005). Alignment of "phrase accent" lows in Dutch falling-rising questions: Theoretical and methodological implications. *Language and Speech*, **48**, 157-183.

Lieberman, P. (1960). Some Acoustic Correlates of Word Stress in American English, *Journal of Acoustical Society of America,* **32(4)**, 451–454.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of Acoustical Society of America*, **35(11)**, 1773-17781.

Lindblom, B. (1990). Expaining phonetic variation : a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, Dordrecht: Kluwer Academic Publishers, 403-439.

Lindblom, B. (1995). Approche intégrée de la production et de la perception. In. H. Meloni (ed.), *Ecole thématique : Fondements et perspectives en traitement automatique de la parole*, *AUPELF UREF,* 1–8.

Lindblom, B. & Rapp, K. (1973) *Some temporal regularities of spoken Swedish*. Papers from the Institute of Linguistics, University of Stockholm, 21.

Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. Journal of Phonetics, **31(3-4),** 321-339.

Malfrère, F., Deroo, O., Dutoit, T. & Ris, C. (2003). Phonetic alignement : speechsynthesis-based versus Viterbi-based. *Speech Communication*, **40(4)**, 503-517.

Malfrère, F., Dutoit, Th. & Mertens, P. (1998). Fully Automatic Prosody Generator for Text-to-Speech Synthesis. In *Proceedings ICSLP-1998*, Sidney, Australia, 1395-1398.

Marsi, E. (2001). *Intonation in Spoken Language Generation.* PhD thesis. Utrecht, LOT, 2001.

Masanobu, A. (1996). Speaking Styles: Statistical Analysis and Synthesis by a Text-to-Speech System. In van Santen et al. (eds.), Progress in Speech Synthesis, Springer-Verlag, 495-510.

Mayo, C. (1996). *Prosodic transcription of Glasgow English*. Masters thesis, University of Edinburgh.

Mayo, C., Aylett, M. & Ladd, D.R. (1997). Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI. In *Proceedings of an ESCA Workshop, Intonation: Theory, Models and Applications*, 231–234.

McCarthy, J. & Prince, A. (1993). Generalized alignment. *Yearbook of morphology,* 79-153.

Mehler, J. (1968). La grammaire générative, a-t-elle une réalité psychologique? *Psychologie française* 13 : 137-156.

Mehler, J., Dupoux, E. & Segui, J. (1990). Constraining Models of Lexical Access: The Onset of Word Recognition. In G. Altmann (ed.) *Cognitive models of speech processing: Psycholinguistic and computational perspective*. Cambridge, MA : MIT Press, 263-280.

Mertens, P., Auchlin, A., Goldman, J.-Ph., Grobet, A., Gaudinat, A. (2001). Intonation du discours et synthèse de la parole : premiers résultats d'une approche par balises. *Cahiers de Linguistque Française,* **23**, 189-209.

Meynadier, Y. (2001). La syllabe phonétique et phonologique : une introduction. *Travaux Interdisciplinaires du Laboratoire PArole et Langage d'Aix-en-Provence*, **20**, 91-148.

Meynadier, Y. (2003). *Interaction entre structure prosodique et coarticulation des consonnes linguopalatales en français*. Thèse de Doctorat, Université d'Aix-Marseille 1, Université de Provence, Aix-en-Provence.

Miller, J., Grosjean, F. & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, **41**, 215-225.

Möbius, B. (1993). Perceptual evaluation of rule-generated intonation contours for German interrogatives. *Working Papers (Dept. of Linguistics and Phonetics, Univ. Lund) (Proceedings of the ESCA Workshop on Prosody, Lund, 27-29 Sept. 1993)*, **41**, 216-219.

Möbius, B. (2000). Corpus-based speech synthesis: methods and challenges. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung* (Univ. Stuttgart), AIMS 6 (4), 87-116.

Monaghan, A.I.C. & Ladd, D.R. (1990). Speaker-dependent and speaker-independent parameters in intonation. In *SCST-1990*, 167-174.

Morel, M.-A., Danon-Boileau, L. (1998). *Grammaire de l'intonation. L'exemple du français.* Coll. Bibliothèque des Faits de Langue. Gap, France : Editions Ophrys

Nakatani, L.H., O'Connor, K.D. & Aston, C.H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica*, **38**, 84-106.

Navarro, H. (1997). *La entonación colloquial. Functión demarcativa y unidades de habla*, *Quaderns de Filologia*, *ANEJO XXI*, Université de Valence.

Nespor, M. & Vogel, I. (1982). Prosodic domains of external sandhi rules. In H.v.d.Hulst, Norval Smith (eds.), *The Structure of Phonological Representations*. Foris. Dordrecht. 225-265.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*, Foris Publication, Dordrecht.

Nesterenko, I. (2001). Contribution à l'étude expérimentale du statut cognitive de la syllabe. *Actes des Journées Prosodie*, Grenoble, 10-11 octobre 2001, pp. 71-74.

Nesterenko, I. (2002). French Listeners Counting Syllables in Read French and Russian: Implications for the Cognitive Reality of Syllable. *Proceedings of SP2002*, 531-534.

Nikolaeva, T. (1969). Николаева Т. М. *Интонация сложного предложения в славянских языках*. М., 1969.

Nikolaeva, T. (1977). Николаева Т. М. *Фразовая интонация славянских языков*. М., 1977.

Nikolaeva, T. (1982). Николаева Т. М. *Семантика акцентного выделения*. Москва, 1982.

Nikolaeva, T. *et al.* (1996). Николаева, Т.М. (отв. ред.), *Просодический строй русской речи*. М., И-т русского языка РАН, 1996.

Nolan, F., and Grabe, E. (1997). Can ToBI transcribe intonational variation in English? In *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications*, Athens, Greece.

O'Connor, J. D. et Arnold, G. F. (1973). *Intonation of Colloquial English*. *Second edition*. London: Longman.

O'Shaughnessy, D. (1995). Timing patterns in fluent and disfluent spontaneous speech. *Proceedings ICASSP'95*. Detroit, USA, 600-603.

Odé, C. (1989). *Russian Intonation: A Perceptual Description*. Rodopi, Amsterdam-Atlanta.

Odé, C. (1991). Combinations of types of pitch accent in a corpus of Russian speech. *Proceedings of the 12thICPhS*, Aix-en-Provence, **4**, 206-209.

Odé, C. (2003). Toward ToRI, a manual transcription system of Russian intonation. *Proceedings 25,* Institute of Phonetic Sciences, University of Amsterdam, 1-14.

Odé, C. (2005). Neutralization or Truncation? The perception of two Russian pitch accents on utterance-final syllables. In *Speech Communication*, **47/1-2**, 71-79.

Ostendorf, M. (2000). Prosodic Boundary Detection. In Horne, M. (éd.), *Prosody : Theory and Experiment*. Coll. Text, Speech and Language Technbology, 14. Dordrecht, Pays-Bas : Kluwer Akademic Publishers, 263-279.

Ostendorf, M. (2003). Prosody models for conversational speech recognition. *Proceedings of the 2nd Plenary Meeting and Symposium on Prosody and Speech Processing*, University of Tokyo, February 2003.

Ostendorf, M. & Veilleux, N. (1994). A hierarchical stochastic model for automatic prediction of prosodic boundary location. *Computational Linguistics*, 20(1).

Palmer, H.E. (1922*). English Intonation with Systematic Exercises*. Cambridge: Heffer.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., Friederici, A. D. (2005). Prosody-driven sentence processing: an event-related brain potential study. *Journal of Cognitive Neurosciences,* **7(3),** 407-421.

Pannekamp, A., Toepel, U., Hahne, A. & Friederici, AD (2003). The brain's response to Hummed Sentences. In *Proceedings of ICPhS-2003*, Barcelone, 877-879.

Pasdeloup, V. (1988). Essai d'analyse du système accentuel du français : distribution de l'accent secondaire. *Actes des 17èmes Journées d'Etude sur la Parole*, Nancy, 20-23 Septembre 1988, 65-70.

Pasdeloup, V. (1992). A Prosodic Model for French Text-to-Speech Synthesis : A Psycholinguistic Approach. In Bailly, G., Benoit, Ch., Sawallis, Th. R. (eds.), *Talking Machines. Theories, Models, And Designs*, 335-348.

Passonneau, R.J. & Litman, D.J. (1993). Intention-Based Segmentation: Human Reliability and Correlation with Linguistic Cues. *Proceedings of the 31st Annual Meeting of the ACL*, Columbus, Ohio, June 1993. Association for Computational Linguistics, 148-155.

Patterson, D. (2000). *A linguistic approach to pitch range modelling*. Thèse de doctorat, Université d'Édimbourg, Écosse.

Peperkamp, S. (1999). Prosodic words. *GLOT International,* **4(4)**, 15-16.

Peperkamp, S**.** (2002). *L'acquisition de la phonologie et ses conséquences pour la théorie linguistique.* Unpublished thesis for 'Habilitation à Diriger des Recherches'. École des Hautes Études en Sciences Sociales.

Peshkovskij, A. (1928). А.М. Пешковский, *Интонация и грамматика*. Избранные труды. М., 1959

Peshkovskij, A. (1930). Пешковский А.М. Ритмика «Стихотворений в прозе» Тургенева. А.М. Пешковский, *Вопросы методики родного языка, лингвистики и стилистики*. М.-Л.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*, PhD dissertation, MIT.

Pierrehumbert, J. (1981). *Synthesizing intonation*. Journal of the Acoustic Society of America, **70**, 985-995.

Pierrehumbert, J. (2001). Stochastic phonology. *GLOT*, **6**, 1-13.

Pierrehumbert, J. & Beckman, M.E. (1988). *Japanese Tone Structure*. Coll. Linguistic Inquiry Monographs, 15. Cambridge, MA, USA : The MIT Press.

Pierrehumbert, J. & Hirschberg, J. (1990). The Meaning of Intonational Countours in the Interpretation of Discourse. In Cohen P. et al. (eds). *Intentions in Communication*, Cambridge: MIT Press, 271-311.

Pierrehumbert, J. & Steele, S.A. (1989). Categories of tonal alignment in English. *Phonetica*, **46,** 181-196.

Pierrehumbert, J. & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty & R. Ladd (eds), *Papers in Laboratory Phonology II: Gesture, Segment, and Prosody*, Cambridge: Cambridge University Press, 90-117.

Pike, K.L. (1945). *The Intonation of American English*. USA: The University of Michigan Press.

Pitrelli, J. F., Beckman, M. E., & Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of ICSLP-94*, **1**, 123-126.

Polivanov, E.D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague,* **4**, 79-96.

Portes, C. & Di Cristo, A. (2003). Pitch range in spontaneous speech: semi-automatic approach versus subjective judgement. *Proceedings of the 15th ICPhS*, 3-9 août, Barcelone, Espagne, 583-586.

Post, B. (2000). *Tonal and phrasal structures in French intonation*, The Hague : Holland Academic graphics.

Potebnia, A.A. (1973). Потебня, А.А. *Ударение.* Киев.

Prevost, S. & Steedman, M. (1994). Specifying intonation from context for text-to-speech synthesis. *Speech Communication*, **15**, 139-154.

Price, P., Ostendorf, M., S. Shattuck-Hufnagel, & C. Fong. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, **90**, 2956-2970.

Prieto, P. (sous presse). Phonological phrasing in Spanish In Colina, S. & Martínez-Gil, F., *Optimality-Theoretic Advances in Spanish Phonology*, John Benjamins.

Prieto, P., D'Imperio, M. & Gili Fivela, B. (sous presse). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. Submitted for the *Special Issue on Intonation in Language Varieties*, ed. by Paul Warren.

Prince, A., & Smolensky, P. (1993). *Optimality theory : constraints interaction in generative grammar*, Ms. Rutgers University, Newbrunswick & University of Colorado, Boulder, RuCCS-TR-2, 1993

Pulgram, E. (1970). *Syllable, Word, Nexus, Cursus*. The Hague: Mouton.

Pynte, J. & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, 11, 165-191

Quené, H. & Kager, R. (1989). Automatic accentuation and prosodic phrasing for Dutch text-to-speech conversion. *Proceedings of Eurospeech-1989*, **2**, 243-246.

Ramus, F. & Mehler, J. (1999). Language identification with suprasegmental cues : a study based on speech resynthesis. *Journal of the Acoustical Society of America,* **105**, 512-521.

Repp, B.H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition,* **44***,* 241-281.

Rietveld, T. & van Lieshout, P. (1995). Articulation and accentuation: preliminary results. *Proceedings of the Department of Language and Speech of University of Nijmegen*, **19**, 99-104.

Ross, K. & Ostendorf, M. (1996). Prediction of Abstract Prosodic Labels for Speech Synthesis. *Computer, Speech and Language*, **10(3),** 155-185.

Rossi M. (1999) *L'intonation, le système du français : description et modélisation*. Paris : Ophrys.

Rossi, M. & Chafcouloff, M. (1972). Recherches sur le seuil de fréquence fondamentale pour la parole. *TIPA* , **1**, 177-184.

Rump, H. & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, **39**, 1-17.

Sag, I. & Liberman, M. (1975). The Intonational Disambiguation of Indirect Speech Acts. *Chicago Linguistic Society (CLS 10),* 487-498.

Sagisaka, Y. (1988), Speech synthesis by rule using an optimal selection of non-uniform synthesis units. *Proceedings ICASSP-88*.

Sagisaka, Y., Takumi, Y. & Kokenawa, Y. (2005). Generation and perception of F_0 markedness for communicative speech synthesis. *Speech Communication* (Special Issue), **46 (3-4)**, 376-384.

Sanderman, A. A. & Collier, R. (1996). Prosodic rules for the implementation of phrase boundaries in synthetic speech. *Journal of the Acoustical Society of America,* **100(5),** 3390-3397.

Schaffer, D. (1984). The role of intonation as a cue to topic management in conversation. *Journal of Phonetics,* **12**, 327-344.

Scherba, L.V. (1963). Щерба Л.В. *Фонетика французского языка*. 7-е изд. М.,1963

Selkirk, E. (1972). *The phrase phonology of English and French,* PhD Dissertation MIT, Garland, New-York, 1972

Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. *Nordic Prosody* II, TAPIR, Trondheim, 111-140.

Selkirk, E. (1984). *Phonology and Syntax : The Relation Between Sound and Structure*. Coll. Current Studies in Linguistics Series, 10. Cambridge, MA, USA : The MIT Press.

Selkirk, E. (1986). On derived Domains in Sentence Phonology. *Phonology,* **3**, 371-405.

Selkirk, E. (1995a). Sentence prosody: intonation, stress and phrasing. In J. Goldsmith (ed.): *Handbook of Phonological theory*. Cambridge: Basil Blackwell, 550-569.

Selkirk, E. (1995b). The prosodic structure of function words. In J. Beckman, S. Urbanczyk & L. Walsh (eds.), *Optimality Theory Occasional Papers: UMOP 18*, UMASS, Amherst, 439-470.

Selkirk, E. (2000). The Interaction of Constraints on Prosodic Phrasing. In Horne, M. (ed.), *Prosody : Theory and Experiment.* Coll. Text, Speech and Language Technbology, 14. Dordrecht, Pays-Bas : Kluwer Akademic Publishers, 231-261.

Selkirk, E. & K. Tateishi. (1991). Syntax and Downstep in Japanese. In Georgopoulos, C. and R. Ishihara. (Eds.). *Interdisciplinary Approaches to Language: Essays in Honor of S. –Y. Kuroda*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Shattuck-Hufnagel, S. (1992). Stress shift as pitch accent placement: within-word early accent placement in American English. In *Proceedings of ICSLP-1992*, 747-750.

Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics*, **22**, 357-388.

Shattuck-Hufnagel, S. & Turk, A.E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research,* **25(2)**, 193-247.

Shih, Ch. & Kochansky, G. (2001). *Synthesis of Prosodic Styles*. 4th ISCA Tutorial and Research Workshop on Speech Synthesis, Scotland, 2001 (disponible sur le site https://netfiles.uiuc.edu/cls/www/paper.html)

Siegel, S. & Castellan, N.J.Jr. (1988). *Nonparametric statistics for the behavioral sciences.* Boston, MA: McGraw-Hill.

Simon, A.-C. (2004), *La structuration prosodique du discours en français*, Berne, Peter Lang.

Skreline, P.A. (1999). Скрелин, П.А., *Сегментация и транскрипция*. СПб, 1999

Skreline, P.A. (2001). Sub-Allophone Synthesis of Russian Vowels. *100 лет экспериментальной фонетике в России*. СПб, 2001

Sluijter, A.M.C. & van Heuven, V.J. (1993). Perceptual cues of linguistic stress: intensity revisited. In *Prosody-1993*, *Working Papers 41, Proceedings of the ESCA Workshop on Prosody,* Lund University, Sweden, 246-249.

Sluijter, A.M.C. & van Heuven, V.J. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. *Phonetica*, **52**, 71-89.

Sluijter, A.M.C. & van Heuven, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of Acoustical Society of America,* **100 (4)**, 2471-2485.

Sluijter, A.M.C., van Heuven, V.J. & Pacilly, J.J.A. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of Acoustical Society of America,* **101 (1)**, 503–513.

Sproat, R. (1992). *Morphology and Computation*. MIT Press: Cambridge.

Sproat, R., Hirschberg, J. & Yarowsky, D. (1992). A Corpus-based Synthesizer. *Proceedings ICSLP*, Banff, Alberta, October 1992, 563-566.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry,* **31**, 649-689.

Steinhauer, K., Alter, K. & Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience,* **2***,* 191-196.

Stepanova, S.B. (1988). Степанова С.Б. Фонетические свойства русской речи: Реализация и транскрипция. Автореф. канд. дисс., Л., 1988.

Stockwell, R.P. (1960). The role of intonation in generative grammar of English. *Language*, **36**, 360-367.

Straka, G. (1963). La division des sons du langage entre voyelles et consonnes peut-elle être justifiée ? *Travaux de Linguistique et de Littérature de Strasbourg*, **I**, 17-99.

Strangert, E. (1990). Perceived pauses, silence intervals and syntactic boundaries. *Reports from the Department of Phonetics*. University of Umeå.

Strangert, E. (1992). Prosodic cues to the perception of syntactic boundaries. *Proceeding of ICSLP-92*, 1283-1286.

Strangert, E. (2004). Speech chunks in conversation: Syntactic and prosodic aspects. *Proceedings Speech Prosody 2004,* Nara, Japan.

Streefkerk, B. M., Pols L.C.W. & ten Bosch L.F.M. (1997). Prominence in read aloud sentences, as marked by listeners and classified automatically. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, **21**, 101-116.

Svetozarova N.D. (1973). The inner structure of intonation contours in Russian. In *Symposium in auditory analysis and perception of speech*. Leningrad.

Svetozarova N.D. (2000). Светозарова Н.Д. Роль фразовой интонации в речевой деятельности и возможности её моделирования. В книге Л.В. Бондарко (отв. ред.), *Фонология речевой деятельности*, издательство СПбГУ, 102-136.

Svetozarova, N.D. (1982). Светозарова Н.Д. *Интонационная система русского языка*. Л., 1982.

Svetozarova, N.D. (1998). Intonation du russe. In Hirst, D. & Di Cristo, A. (eds.), *Intonation Systems*, 261-274.

Svetozarova *et al*. (1988). Фонетика спонтанной речи. Отв. ред. Н.Д. Светозарова. Изд-во СПбГУ, 1988.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America,* **101 (1),** 514-521.

Swinney, D. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior,* **14**, 645-660.

Syrdal, A., J. Hirschberg, McGory, J. & Beckman, M. (2001). Automatic ToBI prediction and alignment to speed manual labeling of prosody. *Speech Communication,* **33,**135-351

Syrdal, A.K. & McGory, J. (2000). Inter-trinscriber reliability of ToBI prosodic labelling. In *Proceedings of ICSLP 2000*.

Takeda, K., Abe, K., & Sagisaka, Y. (1992). On the Basic Scheme and Algorithms in Non-Uniform Unit Speech Synthesis. In Bailly, G., Benoit, Ch., Sawallis, Th. R. (eds.), *Talking Machines. Theories, Models, And Designs*, 1992, 93-105

Tao, J. (2004). Acoustic and linguistic information based Chinese prosodic boundary labelling. *TAL-2004*, 181-184.

Taylor, P. (1995). Using neural networks to locate pitch accents. *Proceedings of. Eurospeech-1995,* Madrid.

Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America*, **107(3),** 1697-1714.

Taylor, P. & Black, A. (1999). Speech Synthesis by Phonological Structure Matching. *Proceedings of Eurospeech 1999,* Budapest, Hungary.

Tcheremissina N.V. (1969). Черемисина Н.В. Ритмико-интонационная структура предложения в русской художественной речи. Черемисина Н.В. (отв. ред.), *Синтаксис и интонация*, Башкирск. гос. ун-т., Уфа, **1,** 96-248.

Tcheremissina N.V. (1989). Черемисина, Н.В. *Русская интонация: поэзия, проза, разговорная речь*. М., Русская речь.

Terken, J. & Collier, R. (1992). Syntactic Influences on Prosody. In Tohkura, Y., Vatikiotis-Bateson, E., Sagisaka, Y. (eds), *Speech Perception, Production And Linguistic Structure*, 427-438

Terken, J. & Nooteboom, S. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information. *Language and Cognitive Processes,* **2(3/4),** 145-163.

Toepel, U. & Alter, K. (2002). Cerebral strategies in the segmentation and interpretation of speech. *Proceedings of Speech Prosody 2002*, 71-76.

Tomson, A.I. (1910). Томсон А.И. *Общее языкознание*. Одесса.

Torsoujeva, I.G. (1974). И.Г. Торсуева. *Теория интонации.* Москва, Изд - во МГПИИЯ.

Traber, C. (1992). F0 Generation with a Database of Natural F0 Patterns and with a Neural Network. In Bailly, G., Benoit, Ch., Sawallis, Th. R. (eds.), *Talking Machines. Theories, Models, And Designs*, 287-304.

Traber, C. (1995). *SVOX: The Implementation of a Textto -Speech System for German*. PhD thesis, No. 11064, Computer Engineering and Networks Laboratory, ETH Zurich, March 1995.

Trager, G.L. & Smith, H.L.J. (1957). An outline of English structure. *Studies in Linguistics: Occasional Papers 3*. American Council of Learned Societies, Washington D.C. (5th printing).

Troubetzkoy, N.S. (1939). *Grundzüge der Phonologie*. Prague. Travaux du Cercle Linguistique de Prague. Traduction française de Cantineau, J. (1949). *Principes de phonologie*. Paris : Klincksieck.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry,* **30**, 219-255.

Truckenbrodt, H. (2005). A short report on intonation phrase boundaries in German. *Linguistische Berichte,* **203**, 273-296.

Tseplitis, L.K. (1974). Цеплитис Л.К. *Анализ речевой интонации*. Рига, 1974, 67-82.

Turk, A.E. & Sawusch, J.R. (1996). The processing of duration and intensity cues to prominence. *Journal of Acoustical Society of America,* **99 (6)**, 3782–3790.

Turk, A.E. & Shattuck-Hufnagel, S. (2000). Word-Boundary-Related Duration Patterns in English. *Journal of Phonetics*, **28**, 397-440.

Turk, A.E. & White, L. (1999). Structural effects on accentual lengthening in English. *Journal of Phonetics*, **27**, 171-206

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, **58(2)**, 434-445.

Vaissière, J. (1983). Language-independant prosodic features. In Cutler & Ladd (eds.), *Prosody, models and measurements, Springer Series in Language and Communication,* **14**, Berlin, Heidelberg, New-York, Tokyo : Springer-Verlag, 53-66.

Vaissière, J. (1997). Langues, prosodie et syntaxe. *Traitement Automatique des Langues,* **38(1),** 53-82

Vaissière, J. (2002). Cross-linguistic prosodic transcription. In N.B. Volskaja, P.A. Skrelin, N.D. Svetozarova (eds.), *Problems and methods in experimental phonetics* (*Проблемы и методы экспериментально-фонетических исследований*). St.-Petersburg State University, 147-164.

Vaissière, J. (2005). Perception of intonation. In D. B. Pisoni & R. E. Remez (eds.), *Handbook of Speech Perception*, Oxford, Blackwell, 236-263.

Vaissière, J. & Michaud, A. (*à paraître*). Prosodic constituents in French: a data-driven approach. À paraître dans . Fónagy, Y. Kawaguchi and T. Moriguchi (eds.), *Prosody and syntax*, "Usage-based linguistic informatics" series, Amsterdam:John Benjamins.

Van Rijsbergen, C.J. (1979). *Information Retrieval,* 2nd edition, Glasgow, University of Glasgow**.**

van Santen, J.P.H. (1992). Contextual effects on vowel duration. *Speech Communication*, **11**, 513-546.

van Santen, J., Black, L., Cohen, G., Klabbers, E., Mishra, T. de Villiers, J. & Niu, X. (2003). Applications of computer generated expressive speech for communication disorders. *Proceedings Eeurospeach-2003, Geneva, Switzerland*, 1657-1660.

van Santen, J. & Hirshberg, J. (1994). Segmental effects on the timing and height of pitch contours. In *Proceedings of ICSLP-1994*, 719-722.

van Santen, J., Kain, A., Klabbers, E. & Mishra, T. (2005). Synthesis of prosody using multi-level unit sequences. *Speech Communication*, **46 (3-4)**, 365-375.

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 Technbology, 15. Dordrecht, Pays-Bas : Kluwer Akademic Publishers, 269-288.

van Santen, J. & Sproat,R. (1998). Introduction. In Sproat, R. (ed.), *Multilingual Text-to-Speech synthesis. The Bell Labs Approach*. Kluwer Academic Publishers, 1-6.

van Santen, J., Sproat, R., Olive, J., and Hirschberg, J. (eds.) (1997) *Progress in Speech Synthesis*, New York: Springer-Verlag.

Venditti, J. (1997). Japanese ToBI labeling guidelines. In *OSU Working Papers in Linguistics*, **50**, 127-162.

Veronis J., Di Cristo, Ph., Courtois, F. & Lagrue, B. (1997). Un modèle stochastique des contours intonatifs en français. *TIPA*, **17**, 137- 154.

Volskaya N.B. (2001b). Вольская Н.Б. Особенности интонационного и синтагматического членения вопросительных высказываний разной длины. *Материалы международной конференции 100 лет экспериментальной фонетике в России*.

Volskaya, E.G. (2001a). Вольская Е. Г. О нетрадиционных способах интонационного оформления синтагм. *Материалы международной конференции 100 лет экспериментальной фонетике в России*.

Vsevolodskij-Gerngoss, V.N. (1922). Всеволодский-Гернгросс В.Н. *Теория русской речевой интонации.*

Wang, M.Q. & Hirschberg, J. (1992). Automatic classification of intonational phrase boundaries. *Computer Speech and Language*, **6(2)**, 175-196.

Ward, G. & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, **61**, 747-776.

Waterworth, J. (1983). Effect of intonation form and pause durations of automatic telephone number announcements on subjective preference and memory performance. *Applied Ergonomics,* **14(1)**, 39-42.

Watson, D. & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, **19**, 713-755

Wewers, M. E. & N. K. Lowe. (1990). A Critical Review of Visual Analogue Scales in the Measurement of Clinical Phenomena. *Research in Nursing & Health,* **13**, 227-236.

Wightman, C. (2002). ToBI or not toBI? In *Speech Prosody-2002*, 25-29.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & P.J. Price. (1992). Segmental Durations in the Vicinity of Prosodic Boundaries. *Journal of the Acoustic Society of America*, **91**, 1707-1717

Wightman, C., Syrdal, A., Stemmer, G., Conkie, A. & Beutnagel, M. (2000). Perceptually based automatic prosody labeling and prosodically enriched unit selection improve concatenative speech synthesis. In *Proceedings ICSLP-2000*, **2**, 71-74.

Wightman, C.W., Syrdal, A.K., Stemmer, G., Conkie, A., Beutnagel, M. (2000). Perceptually based automatic prosody labeling and prosodically enriched unit

selection improve concatenative text-to-speech synthesis. *Proceedings of ICSLP-2000*, **2**, 71-74.

Wingfield, A. & Nolan, K.A. (1980). Spontaneous segmentation in normal and time-compressed speech. *Perception and Psychophysics*, **28(2)**, 97-102.

Xu, Y., Ching, X. X. & Xuejing, S. (2004). On the temporal domain of focus. *Proceedings of the conference Speech Prosody-2004*, Nara, Japan.

Yokoyama, O. (1987). Towards a Unified Framework of Russian Intonation. In *Proceedings of the Eleventh International Congress of Phonetic Sciences*, Tallin

Yokoyama, O. (2003). Neutral and non-neutral intonation in Russian: a reinterpretation of the IK system. *Вопросы языкознания*, **5.**

Yoon, T., Chavarria, S., Cole, J., & Hasegawa-Johnson, M. (2004) Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. In *Interspeech--2004*, 2729-2732.

Yule, G. (1980). Speaker's topic and major paratones. *Lingua,* **52**, 33-47.

Zalizniak, A.A.(1977). Зализняк А.А., *Грамматический словарь русского языка*. Москва.

Zec, D. & Inkelas, S. (1990). Prosodically constrained syntax. In S. Inkelas & D. Zec (eds.), *The phonology-syntax connection*, The University of Chicago Press, Chicago, 365-378.

Zinder, L.R. (1979). Зиндер, Л.Р. *Общая фонетика*. М., 1979.

Zlatooustova, L., Potapova, R., Trounine, V. (1986). Phonétique générale et appliquée. Moscou, URSS : Editions de L'Université de Moscou, 1986 (en russe).

LES ANNEXES

1. Annexe I : Symboles SAMPA et symboles IPA correspondants pour le russe

Consonnes :

p	p	tS’	t͡ʃʲ	x’	xʲ
p’	pʲ	f	f	m	m
b	b	f’	fʲ	m’	mʲ
b’	bʲ	v	v	n	n
t	t	v’	vʲ	n’	nʲ
t’	tʲ	s	s	l	l
d	d	s’	sʲ	l’	lʲ
d’	dʲ	z	z	r	r
k	k	z’	zʲ	r’	rʲ
k’	kʲ	S	ʃ	j	j
g	g	S’:	ʃʲː		
g’	gʲ	Z	ʒ		
ts	t͡s	x	x		

Voyelles :

i	i
e	e
a	a
o	o
u	u
1	ɨ
@	ə

2. Annexe II : Composition phonémique du corpus

PHONÈME (SAMPA)	DURÉE MOYENNE (en secondes)	ÉCART TYPE (en secondes)	EFFECTIFS
@	0,054	0,035	364
l	0,056	0,035	145
a	0,103	0,055	651
b	0,073	0,017	80
b'	0,069	0,018	15
d	0,065	0,026	156
d'	0,069	0,025	60
e	0,092	0,044	199
f	0,059	0,033	39
f'	0,058	0,014	5
g	0,064	0,017	62
g'	0,068	0,015	11
i	0,054	0,029	438
j	0,063	0,021	91
k	0,085	0,025	133
k'	0,101	0,033	26
l	0,050	0,015	104
l'	0,054	0,016	84
m	0,067	0,028	193
m'	0,064	0,017	28
n	0,057	0,023	276
n'	0,059	0,025	113
o	0,096	0,043	283
p	0,069	0,022	125
p'	0,077	0,022	21
R	0,040	0,009	130
R'	0,042	0,012	49
S	0,093	0,030	101
s	0,079	0,035	144
s'	0,099	0,035	68
S':	0,106	0,028	45
t	0,079	0,033	277
t'	0,089	0,036	85
ts	0,125	0,040	39
tS'	0,098	0,025	65
u	0,066	0,043	264
v	0,054	0,026	104
v'	0,044	0,014	31
x	0,089	0,033	50
x'	0,140	0,024	3
Z	0,064	0,027	52
z	0,063	0,020	73
z'	0,057	0,016	26

3. Annexe III : Accord entre les auditeurs

a. Pourcentage d'accord

	CondR	CondD	CondN		CondR	CondD	CondN
NB	0,877	0,902	0,956	AC	0,904	0,907	0,909
PA	0,874	0,906	0,947	CB	0,881	0,892	0,892
SB	0,881	0,902	0,955	CP	0,885	0,901	0,910
SO	0,874	0,900	0,951	CA	0,893	0,906	0,911
OG	0,870	0,901	0,952	LC	0,876	0,868	0,880
OB	0,879	0,904	0,949	MF	0,873	0,899	0,893
EA	0,869	0,915	0,938	RB	0,901	0,907	0,892
Moyenne	*0,875*	*0,904*	*0,950*	*Moyenne*	*0,888*	*0,897*	*0,898*
Écart type	*0,0048*	*0,0048*	*0,0048*	*Écart type*	*0,0048*	*0,0048*	*0,0048*

Tableau 1 : Stabilité individuelle : pourcentage d'accord pour chaque annotateur avec les six autres. Panel gauche : auditeurs russes ; panel droit : auditeurs francophones.

b. Coefficients kappa.

	CondR		CondD		CondN	
	Méthode Cohen	Méthode Siegel	Méthode Cohen	Méthode Siegel	Méthode Cohen	Méthode Siegel
NB	0,4531	0,4520	0,5138	0,5134	0,7844	0,7842
PA	0,4001	0,3974	0,5751	0,5744	0,7595	0,7594
SB	0,4588	0,4571	0,5255	0,5252	0,7928	0,7927
SO	0,4698	0,4684	0,4903	0,4894	0,7622	0,7619
OG	0,4876	0,4846	0,5354	0,5595	0,7749	0,7748
OB	0,4558	0,4546	0,5457	0,5455	0,7585	0,7584
EA	0,4737	0,4714	0,5960	0,5958	0,7262	0,7258

Tableau 2 : Stabilité individuelle : coefficients de Kappa, mesurant l'accord de chaque annotateur avec les six autres. Auditeurs russes.

	CondR		CondD		CondN	
	Méthode Cohen	Méthode Siegel	Méthode Cohen	Méthode Siegel	Méthode Cohen	Méthode Siegel
AC	0,5126	0,5099	0,5294	0,5270	0,5841	0,5819
CB	0,4835	0,4806	0,3660	0,3534	0,4179	0,4062
CP	0,5150	0,4470	0,5314	0,5283	0,5869	0,5847
CA	0,4326	0,4269	0,5470	0,5443	0,6120	0,6096
LC	0,4998	0,4753	0,4912	0,4797	0,5482	0,5401
MF	0,4085	0,4062	0,4930	0,4889	0,4922	0,4895
RB	0,4840	0,4799	0,5405	0,5382	0,5616	0,5577

Tableau 3 : Stabilité individuelle : coefficients de Kappa, mesurant l'accord de chaque annotateur avec les six autres. Auditeurs francophones.

4. Annexe IV. Étude de la dimension localisante

a. Matrices originales

AUDITEURS RUSSES

CondR \ CondD

	0	1	2	3	4	5	6	7
0	247	11	12	3	3	4	4	3
1	28	4	5	0	2	0	1	0
2	12	2	4	2	0	1	2	1
3	9	2	1	4	4	0	1	0
4	12	0	5	0	0	1	0	2
5	2	2	1	1	0	0	1	0
6	1	0	1	0	3	2	2	2
7	4	0	1	0	0	0	1	3

CondR \ CondN

	0	1	2	3	4	5	6	7
0	244	13	4	4	5	3	5	9
1	35	1	1	0	0	0	0	3
2	14	2	1	1	0	1	1	4
3	16	1	1	0	0	0	0	3
4	13	1	0	0	0	0	0	6
5	6	1	0	0	0	0	0	0
6	5	1	0	1	0	2	0	2
7	4	0	0	1	0	0	2	2

CondD \ CondN

	0	1	2	3	4	5	6	7
0	273	12	6	4	3	3	4	10
1	18	1	0	0	0	0	1	1
2	24	2	0	2	1	0	1	0
3	5	1	1	0	1	1	0	1
4	5	2	0	1	0	0	0	4
5	4	2	0	0	0	0	0	2
6	5	0	0	0	0	2	0	5
7	3	0	0	0	0	0	2	6

AUDITEURS FRANÇAIS

CondR \ CondD

	0	1	2	3	4	5	6	7
0	237	16	17	4	11	6	7	2
1	26	5	3	1	0	0	1	0
2	15	2	2	1	1	2	1	0
3	7	3	0	1	1	0	1	1
4	9	3	1	2	2	0	1	0
5	5	1	0	0	0	0	0	1
6	4	4	1	0	0	0	2	2
7	4	0	0	0	0	1	1	1

CondR \ CondN

	0	1	2	3	4	5	6	7
0	225	22	7	13	8	9	10	6
1	24	4	3	2	1	0	2	0
2	19	2	0	0	1	1	0	1
3	7	0	1	5	0	1	0	0
4	16	0	0	1	0	0	0	1
5	7	0	0	0	0	0	0	0
6	6	2	0	0	1	0	3	1
7	4	1	0	0	0	1	0	1

CondD \ CondN

	0	1	2	3	4	5	6	7
0	256	18	7	9	4	6	6	1
1	17	5	3	5	1	1	2	0
2	14	5	1	1	0	2	1	0
3	6	2	0	0	0	0	0	1
4	9	0	0	2	2	1	1	0
5	2	1	0	1	0	0	3	2
6	1	0	0	3	4	0	2	4
7	3	0	0	0	0	2	0	2

ANALYSE TRANS-LANGUE

CondR

Russe \ Français	0	1	2	3	4	5	6	7
0	220	23	15	6	11	3	5	4
1	31	4	0	2	2	0	1	0
2	14	3	1	3	1	1	1	0
3	9	4	2	2	3	0	0	1
4	10	1	2	1	0	1	5	0
5	4	1	0	0	1	1	0	0
6	5	0	3	0	0	1	1	1
7	7	0	1	0	0	0	0	1

CondD

Russe \ Français	0	1	2	3	4	5	6	7
0	259	23	12	2	5	6	6	2
1	14	3	1	1	2	0	0	0
2	14	4	5	3	2	1	1	0
3	5	2	2	0	1	0	0	0
4	5	2	2	0	1	0	2	0
5	3	0	1	0	3	0	1	0
6	3	0	0	2	1	1	2	3
7	4	0	1	1	0	1	2	2

CondN

Russe \ Français	0	1	2	3	4	5	6	7
0	273	25	8	11	10	5	5	0
1	15	2	0	1	0	1	1	0
2	3	1	1	1	0	0	1	0
3	4	1	0	1	0	0	0	1
4	2	0	0	1	0	0	2	0
5	1	0	0	1	0	2	1	1
6	0	2	2	0	0	1	1	2
7	10	0	0	5	1	3	4	6

b. Matrices dérivées avec trois catégories de frontières : « absence de frontière », « une frontière optionnelle », « une frontière stable (consensuelle) ».

Auditeurs russes

CondD / CondR

	A	O	S
A	247	26	14
O	49	24	12
S	19	11	17

CondN / CondR

	A	O	S
A	244	21	22
O	65	8	12
S	28	5	14

CondN / CondD

	A	O	S
A	273	22	20
O	47	7	7
S	17	5	21

Auditeurs français

CondD / CondR

	A	O	S
A	237	37	26
O	48	18	8
S	22	12	11

CondN / CondR

	A	O	S
A	225	42	33
O	50	17	7
S	33	4	8

CondN / CondD

	A	O	S
A	256	34	17
O	37	22	8
S	15	7	23

ANALYSE TRANS-LANGUE

CondR

Français / Russe

	A	O	S
A	220	44	23
O	54	21	10
S	26	9	12

CondD

Français / Russe

	A	O	S
A	259	37	19
O	33	21	7
S	15	9	19

CondN

Français / Russe

	A	O	S
A	273	44	20
O	22	8	4
S	13	11	24

5. Annexe V : Détails des analyses par des modèles linéaires mixtes.

Nous présentons ici les résultats des tests statistiques (moyennant les modèles linéaires mixtes) qui portaient sur la problématique de différence des durées entre les syntagmes terminaux versus non terminaux.
Nous voulons tester si la durée des syntagmes non-terminaux diffère systématiquement de celle des syntagmes terminaux, et ceci en fonction de la condition expérimentale. Deux facteurs fixes sont alors en jeu : un facteur à trois niveaux 'Condition' et un facteur 'iutype' (type d'unité intonative) à deux niveaux. Afin de tester plus facilement nos hypothèses, nous avons transformé les (3x2) combinaisons des facteurs 'Condition' et 'iutype' en un unique facteur 'conduit' à 6 niveaux.

```
{       > factor(paste( lengthru2$condition, lengthru2$iutype)) -> lengthru2$condiut
        > levels(lengthru2$condiut)
        [1] "1 f"  "1 nf" "2 f"  "2 nf" "3 f"  "3 nf"
        > factor(paste( lengthru2$condition, lengthru2$iutype,sep="")) ->
lengthru2$condiut
        > levels(lengthru2$condiut)
        [1] "1f"  "1nf" "2f"  "2nf" "3f"  "3nf"
        > cbind(c(1,-1,0,0,0,0),c(0,0,1,-1,0,0), c(0,0,0,0,1,-1) ) -> bid
        > bid
             [,1] [,2] [,3]
        [1,]    1    0    0
        [2,]   -1    0    0
        [3,]    0    1    0
        [4,]    0   -1    0
        [5,]    0    0    1
        [6,]    0    0   -1
        > contrasts(lengthru2$condiut) <- bid        }
```

Nous sommes donc amenée à tester six hypothèses. Dans ces tests, 3 jeux de contrastes ont été nécessaires, pour des questions d'orthogonalité de contrastes

Un premier jeu de contrastes permet de tester les hypothèses qui comparent en fonction de la condition expérimentale les durées des syntagmes terminaux versus non-terminaux).
Selon la syntaxe de R, notre modèle mixte s'écrit alors :
length ~ conduit, random= ~1|sujet/stimulus
(le terme de random reprend des facteurs de répétition, avec le facteur stimulus emboîté dans le facteur 'sujet').
Pour le facteur fixe, nous obtenons les résultats suivants :

{ *Fixed effects: length ~ condiut*

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.9499744	0.02204100	1447	43.10034	0.0000
condiut1	0.1538518	0.01916526	1447	8.02764	0.0000
condiut2	0.0176677	0.01954265	1447	0.90406	0.3661
condiut3	0.0717217	0.01944588	1447	3.68827	0.0002
condiut4	-0.0010065	0.02745853	1447	-0.03666	0.9708
condiut5	0.0225791	0.02719342	1447	0.83031	0.4065

}

Un second jeu de contraste (contraste de type "traitement", avec CondR pris comme niveau de référence) permet de tester les hypothèses sur la stabilité de la durée des syntagmes non-terminaux : a) entre les conditions CondR et CondD ; b) entre les conditions CondR et CondN.
Selon la syntaxe de R, le modèle testé s'écrit :
*length ~ condition*iutype, random= ~1|sujet/stimulus*

{ *Fixed effects: length ~ condition * iutype2*

	Value	Std.Error	DF	t-value	p-value
(Intercept)	*0.7496304*	*0.04031675*	*1447*	*18.593524*	*0.0000*
condition2	*0.1583737*	*0.03101761*	*1447*	*5.105931*	*0.0000*
condition3	*0.1067233*	*0.03076890*	*1447*	*3.468545*	*0.0005*
iutype2f	*0.3411598*	*0.03835463*	*1447*	*8.894877*	*0.0000*
condition2:iutype2f	*-0.2745005*	*0.05462115*	*1447*	*-5.025536*	*0.0000*
condition3:iutype2f	*-0.1698026*	*0.05448030*	*1447*	*-3.116770*	*0.0019*

}

Enfin, un troisième jeu de contraste (contraste de type 'traitement', avec CondD pris comme niveau de référence) permet d'obtenir le test de l'hypothèse, qui compare les durées des syntagmes non-terminaux entre les conditions CondD et CondN.

{ *Fixed effects: length ~ condition2 * iutype2*

	Value	Std.Error	DF	t-value	p-value
(Intercept)	*0.9080042*	*0.04112526*	*1447*	*22.078991*	*0.0000*
condition21	*-0.1583737*	*0.03101761*	*1447*	*-5.105931*	*0.0000*
condition23	*-0.0516504*	*0.03168944*	*1447*	*-1.629894*	*0.1033*
iutype2f	*0.0666592*	*0.03920363*	*1447*	*1.700333*	*0.0893*
condition21:iutype2f	*0.2745005*	*0.05462115*	*1447*	*5.025536*	*0.0000*
condition23:iutype2f	*0.1046980*	*0.05500544*	*1447*	*1.903411*	*0.0572*

}

Ces 3 jeux de contrastes sont simultanés, le seuil de significativité doit être corrigé en conséquence. Nous avons choisi la correction de Bonferroni, la plus simple: le seuil devient alors environ égal à 0.017 (=0.05/3).

Les mêmes traitements ont été appliqués aux données des auditeurs francophones. Nous présentons ci-après les sommaires simplifiés des modèles testés.

Premier jeu de contraste :

{ *Fixed effects: length ~ condiut*

	Value	Std.Error	DF	t-value	p-value
(Intercept)	*0.9846325*	*0.05463066*	*1431*	*18.023442*	*0.0000*
condiut1	*0.1646758*	*0.01743499*	*1431*	*9.445134*	*0.0000*
condiut2	*0.1103294*	*0.01751912*	*1431*	*6.297658*	*0.0000*
condiut3	*0.0959729*	*0.01734743*	*1431*	*5.532401*	*0.0000*
condiut4	*-0.0662965*	*0.02451284*	*1431*	*-2.704561*	*0.0069*
condiut5	*-0.0133663*	*0.02460105*	*1431*	*-0.543320*	*0.5870*

}

Deuxième jeu de contraste :

{ *Fixed effects: length ~ condition * iutype*

	Value	*Std.Error*	*DF*	*t-value*	*p-value*
(Intercept)	*0.8054077*	*0.04631164*	*1431*	*17.391043*	*0.0000*
condition2	*0.0691056*	*0.02843126*	*1431*	*2.430620*	*0.0152*
condition3	*0.0196405*	*0.02807989*	*1431*	*0.699449*	*0.4844*
iutypef	*0.3516176*	*0.03479810*	*1431*	*10.104506*	*0.0000*
condition2:iutypef	*-0.1020286*	*0.04910855*	*1431*	*-2.077615*	*0.0379*
condition3:iutypef	*-0.1330670*	*0.04890604*	*1431*	*-2.720871*	*0.0066*

}

Troisième jeu de contraste :

{ *Fixed effects: length ~ condition2 * iutype*

	Value	*Std.Error*	*DF*	*t-value*	*p-value*
(Intercept)	*0.9035928*	*0.05749352*	*1431*	*15.716430*	*0.0000*
condition21	*-0.0759191*	*0.02858390*	*1431*	*-2.656008*	*0.0080*
condition23	*-0.0519400*	*0.02817481*	*1431*	*-1.843490*	*0.0655*
iutypef	*0.2206589*	*0.03503825*	*1431*	*6.297658*	*0.0000*
condition21:iutypef	*0.1086927*	*0.04924099*	*1431*	*2.207362*	*0.0274*
condition23:iutypef	*-0.0287130*	*0.04900315*	*1431*	*-0.585942*	*0.5580*

}

6. Annexe VI : Prédictions des contraintes de la théorie de l'optimalité

a. Stimuli

1. [[Главное]$_{NP}$ [не куда поехать]$_{IP}$]$_{CP}$ [а [главное]$_{NP}$ [с кем поехать]$_{IP}$]$_{CP}$
2. [Небольшой]$_{Adj}$ [[[корабль]$_{NP}$ очень качало]$_{VP}$]$_{CP}$ [я испугалась]$_{CP}$
3. [Ну там была]$_{VP}$ [буря]$_{NP}$
4. [Там тоже]$_{AdvP}$ [наверно]$_{AdvP}$ [много молодёжи]$_{NP}$
5. [В кино]$_{PP}$ ой [ты]$_{DP}$ [знаешь [[я]$_{DP}$ [вообще-то не очень люблю [ходить в кино]$_{IP}$]$_{VP}$]$_{CP}$]$_{VP}$
6. [А на чём]$_{PP}$ [мы]$_{DP}$ [поедем]$_{VP}$ [[на электричке]$_{PP}$ или [на твоей машине]$_{PP}$]$_{PP}$
7. [Ну да] [[в принципе]$_{PP}$ [[у многих]$_{PP}$ есть]$_{VP}$ [машины]$_{NP}$]$_{CP}$
8. [Очень удобно]$_{CP}$ [не нужно [никаких пересадок]$_{NP}$ [ни на метро]$_{PP}$]$_{CP}$
9. [А раньше]$_{AdvP}$ [ты]$_{DP}$ [[на метро]$_{PP}$ ездила]$_{VP}$
10. [[Наверно]$_{AdvP}$ [очень много денег]$_{NP}$ [[на бензин]$_{PP}$ уходит]$_{VP}$]$_{CP}$ [да]$_{Pro}$
11. [Ну я не знаю]$_{CP}$ [[[мне]$_{DP}$ сказали [что мне позвонят]$_{CP}$ [если я понадоблюсь в школе]$_{CP}$]$_{VP}$]$_{CP}$ [но пока-что не звонят]$_{CP}$
12. [Да]$_{Pro}$ [[с детьми]$_{PP}$ [очень здорово работать]$_{VP}$]$_{CP}$ [мне очень нравится]$_{CP}$
13. [[Ты]$_{DP}$ [очень хорошо водишь]$_{VP}$]$_{CP}$ [мне нравится]$_{CP}$
14. [[В автошколе]$_{PP}$ [я]$_{DP}$ [сдала с третьего раза]$_{VP}$]$_{CP}$ [а [в ГАИ]$_{PP}$ [я]$_{DP}$ [сдала с первого раза]$_{VP}$]$_{CP}$.
15. [Им приходилось [[по нескольку раз]$_{PP}$ [по четыре]$_{PP}$ [по пять раз]$_{PP}$ ходить и пересдавать [в ГАИ]$_{PP}$]$_{IP}$]$_{IP}$
16. [Да]$_{Pro}$ [[и в принципе]$_{PP}$ [за тебя всё решает]$_{VP}$ [инструктор]$_{NP}$ [когда [ты]$_{DP}$ [с ним едешь]$_{VP}$]$_{CP}$]$_{CP}$
17. [Ничего себе]$_{CP}$ [нет]$_{Pro}$ [[у нас]$_{DP}$ [было поменьше]$_{VP}$]$_{CP}$ [[[у нас]$_{DP}$ [сорок вопросов]$_{NP}$]$_{VP}$]$_{CP}$
18. [[Я тоже]$_{DP}$ [так думаю]$_{VP}$]$_{CP}$ [конечно [лучше учиться одной]$_{VP}$]$_{CP}$
19. [[Пока-что]$_{Adv}$ нет]$_{Pro}$ [но [у меня]$_{PP}$ [очень многие]$_{NP}$ [сейчас пошли [[на права]$_{PP}$ сдавать]$_{VP}$]$_{IP}$]$_{CP}$
20. [Да]$_{Pro}$ [[у меня]$_{DP}$ была]$_{VP}$ [куча разных животных]$_{NP}$ [в детстве]$_{PP}$]$_{CP}$ [[[у меня]$_{DP}$ были]$_{VP}$ [хомяки]$_{NP}$]$_{CP}$
21. [Ну вот [с собакой]$_{PP}$ [не очень удобно]$_{AdvP}$]$_{CP}$ [[с ней]$_{PP}$ [нужно гулять]$_{VP}$ [никуда не уехать надолго]$_{VP}$]$_{CP}$
22. [Ну [[у дедушки]$_{PP}$ [там]$_{Adv}$ [своих забот]$_{NP}$ [хватает]$_{V}$]$_{VP}$]$_{CP}$ [[у него тоже]$_{PP}$ [там]$_{Adv}$ [кошки всякие]$_{NP}$ [собака]$_{NP}$ [есть]$_{V}$]$_{CP}$
23. [Нет]$_{Pro}$ [[сейчас]$_{Adv}$ [ничего]$_{DP}$ [не смотрю]$_{V}$]$_{VP}$ [потому что]$_{CP}$
24. [Я]$_{DP}$ [думаю [что [[мне]$_{DP}$ [стоит]$_{V}$ [[ещё]$_{Adv}$ [этим]$_{DP}$ [позаниматься]$_{V}$]$_{IP}$]$_{VP}$]$_{CP}$]$_{VP}$
25. [Нет]$_{Pro}$ [я]$_{DP}$ [к сожалению]$_{PP}$ [не попала [на эти выступления]$_{PP}$]$_{VP}$

a. Matrices d'effectifs par condition

Auditeurs russes, totalité de données

CONDITION CONDR

	ALIGNXP,R	WRAPXP
frontières prédites	62	47
frontières non-prédites	55	70
TOTAL	117	117

CONDITION CONDD

	ALIGNXP,R	WRAPXP
frontières prédites	55	44
frontières non-prédites	52	63
TOTAL	107	107

CONDITION CONDN

	ALIGNXP,R	WRAPXP
frontières prédites	60	50
frontières non-prédites	21	31
TOTAL	81	81

Auditeurs russes, frontières consensuelles

CONDITION CONDR

	ALIGNXP,R	WRAPXP
frontières prédites	31	29
frontières non-prédites	16	18
TOTAL	47	47

CONDITION CONDD

	ALIGNXP,R	WRAPXP
frontières prédites	33	30
frontières non-prédites	12	15
TOTAL	45	45

CONDITION CONDN

	ALIGNXP,R	WRAPXP
frontières prédites	44	38
frontières non-prédites	5	11
TOTAL	49	49

Auditeurs francophones, totalité de données

CONDITION CONDR

	ALIGNXP,R	WRAPXP
frontières prédites	59	49
frontières non-prédites	52	62
TOTAL	111	111

CONDITION CONDD

	ALIGNXP,R	WRAPXP
frontières prédites	61	47
frontières non-prédites	48	62
TOTAL	109	109

CONDITION CONDN

	ALIGNXP,R	WRAPXP
frontières prédites	64	49
frontières non-prédites	47	62
TOTAL	111	111

Auditeurs francophones, frontières consensuelles

CONDITION CONDR

	ALIGNXP,R	WRAPXP
frontières prédites	30	27
frontières non-prédites	18	21
TOTAL	48	48

CONDITION CONDD

	ALIGNXP,R	WRAPXP
frontières prédites	29	22
frontières non-prédites	16	23
TOTAL	45	45

CONDITION CONDN

	ALIGNXP,R	WRAPXP
frontières prédites	31	27
frontières non-prédites	17	21
TOTAL	48	48

Annexe 7 : Grammaire probabiliste INTSINT

7.1. Analyse des bigrammes

Number of categories : 9

Category 0 : #
Category 1 : Top
Category 2 : Mid
Category 3 : Bottom
Category 4 : Higher
Category 5 : Lower
Category 6 : Same
Category 7 : Upset
Category 8 : Downstepped

Entropy : 2.0653682 Maximum Entropy : 2.1972246 Normalized Entropy : 0.93998957

---------- PATTERNS MODEL ----------
Number of patterns : 10
Model Entropy : 1.7571504 Mutual Information of the model : 0.30821776
Normalized Model Entropy : 0.7997136 Normalized Mutual Information of the model : 0.14027594

Pattern 0 : <|
Number of occurrences : 1210 Pattern probability : 0.0
Mutual information : 0.0 Entropy : 2.0653682
Normalized Mutual information : 0.0 Normalized Entropy : 0.93998957
Contribution to Mutual information : 0.0 Contribution to Entropy : 0.0
Contribution to Normalized Mutual information : 0.0Contribution to Normalized Entropy : 0.0

#	1	0.25950414	0.25950414	0.0	0.012602038	0.0
Top	2	0.09256198	0.09256198	0.0	0.008331671	0.0
Mid	3	0.033884298	0.033884298	0.0	0.0052014105	0.0
Bottom	4	0.061157025	0.061157025	0.0	0.0068885367	0.0
Higher	5	0.07933884	0.07933884	0.0	0.0077696266	0.0
Lower	6	0.10247934	0.10247934	0.0	0.008718619	0.0
Same	7	0.13057852	0.13057852	0.0	0.00968631	0.0
Upset	8	0.12644628	0.12644628	0.0	0.009554438	0.0
Downstepped	9	0.11404958	0.11404958	0.0	0.009138162	0.0

Pattern 1 : <#|
Number of occurrences : 314 Pattern probability : 0.25950414
Mutual information : 0.4534335 Entropy : 1.8933698
Normalized Mutual information : 0.20636648 Normalized Entropy : 0.8617097
Contribution to Mutual information : 0.117667876 Contribution to Entropy : 0.4913373
Contribution to Normalized Mutual information : 0.05355296Contribution to Normalized Entropy : 0.22361724

#	1	0.0	0.25950414	-0.25950414	0.0031847134	-81.4843
Top	2	0.01590866	0.09256198	-0.076653324	0.007061057	-10.855786
Mid	3	0.12408756	0.033884298	0.090203255	0.018605	4.8483343
Bottom	4	0.061099775	0.061157025	-5.725026E-5	0.013516515	-0.0042356
Higher	5	0.079264574	0.07933884	-7.426739E-5	0.015245517	-0.0048714
Lower	7	0.24181165	0.13057852	0.11123313	0.02416364	4.603327
Upset	8	0.22590297	0.12644628	0.0994567	0.023599014	4.214443
Downstepped	9	0.14954141	0.11404958	0.035491824	0.020125305	1.7635422

Pattern 2 : <Top|
Number of occurrences : 112 Pattern probability : 0.09256198
Mutual information : 0.5546767 Entropy : 1.2757356
Normalized Mutual information : 0.25244424 Normalized Entropy : 0.5806123
Contribution to Mutual information : 0.051341977 Contribution to Entropy : 0.11808462
Contribution to Normalized Mutual information : 0.02336674Contribution to Normalized Entropy : 0.053742625

#	1	0.48855752	0.25950414	0.22905338	0.047233183	4.8494167
Top	2	0.0	0.09256198	-0.09256198	0.008928572	-10.366942
Mid	3	0.0	0.033884298	-0.033884298	0.008928572	-3.7950413
Bottom	4	0.05975745	0.061157025	-0.0013995767	0.022397881	-0.062487
Higher	5	0.0	0.07933884	-0.07933884	0.008928572	-8.88595
Lower	6	0.28789997	0.10247934	0.18542063	0.04278409	4.3338685
Same	7	0.05234545	0.13057852	-0.07823306	0.021045318	-3.717362
Upset	8	0.0	0.12644628	-0.12644628	0.008928572	-14.161984
Downstepped	9	0.11143956	0.11404958	-0.0026100203	0.029734058	-0.0877788

Pattern 3 : <Mid|
Number of occurrences : 41 Pattern probability : 0.033884298
Mutual information : 0.36641496 Entropy : 1.9574394
Normalized Mutual information : 0.16676264 Normalized Entropy : 0.8908691
Contribution to Mutual information : 0.012415714 Contribution to Entropy : 0.06632646
Contribution to Normalized Mutual information : 0.005650635 Contribution to Normalized Entropy : 0.030186472

#	1	0.024162788	0.25950414	-0.23534136	0.023981165	-9.813591
Top	2	0.14497672	0.09256198	0.052414738	0.054985248	0.95325094
Mid	3	0.0	0.033884298	-0.033884298	0.024390243	-1.3892562
Bottom	4	0.12081394	0.061157025	0.059656918	0.050898734	1.1720707
Higher	5	0.21746509	0.07933884	0.13812625	0.06442504	2.143984
Lower	6	0.101523645	0.10247934	-9.556934E-4	0.047167733	-0.0202616
Same	7	0.07248836	0.13057852	-0.058090158	0.040495045	-1.4345003
Upset	8	0.12526707	0.12644628	-0.0011792034	0.051696874	-0.02281
Downstepped	9	0.1933023	0.11404958	0.07925272	0.06167114	1.285086

Pattern 4 : <Bottom|
Number of occurrences : 74 Pattern probability : 0.061157025
Mutual information : 0.47556612 Entropy : 1.2735273
Normalized Mutual information : 0.21643947 Normalized Entropy : 0.57960725
Contribution to Mutual information : 0.02908421 Contribution to Entropy : 0.077885136
Contribution to Normalized Mutual information : 0.013236794 Contribution to Normalized Entropy : 0.035447054

#	1	0.47297296	0.25950414	0.21346882	0.058038842	3.6780338
Top	2	0.054054055	0.09256198	-0.038507927	0.02628642	-1.4649361
Mid	3	0.0	0.033884298	-0.033884298	0.013513514	-2.507438
Bottom	4	0.0	0.061157025	-0.061157025	0.013513514	-4.52562
Higher	5	0.027027028	0.07933884	-0.052311815	0.01885097	-2.77502
Lower	6	0.0	0.10247934	-0.10247934	0.013513514	-7.5834713
Same	7	0.25675675	0.13057852	0.12617823	0.05078209	2.4846997
Upset	8	0.1891892	0.12644628	0.06274292	0.045529418	1.3780743
Downstepped	9	0.0	0.11404958	-0.11404958	0.013513514	-8.43967

Pattern 5 : <Higher|
Number of occurrences : 96 Pattern probability : 0.07933884
Mutual information : 0.24863575 Entropy : 1.75197
Normalized Mutual information : 0.11315901 Normalized Entropy : 0.7973559
Contribution to Mutual information : 0.019726472 Contribution to Entropy : 0.13899927
Contribution to Normalized Mutual information : 0.008977904 Contribution to Normalized Entropy : 0.06326129

#	1	0.326708	0.25950414	0.06720385	0.04786808	1.4039387
Top	2	0.09364875	0.09256198	0.001086764	0.029734707	0.03654867
Mid	3	0.0	0.033884298	-0.033884298	0.010416667	-3.2528925
Bottom	4	0.09485071	0.061157025	0.033693686	0.029905068	1.1266881
Higher	5	0.021077937	0.07933884	-0.058260903	0.014660625	-3.9739714
Lower	6	0.20024039	0.10247934	0.09776105	0.04084322	2.3935688
Same	7	0.06323381	0.13057852	-0.06734471	0.024840176	-2.7111204
Upset	8	0.010538968	0.12644628	-0.11590731	0.010422281	-11.121108
Downstepped	9	0.18970142	0.11404958	0.07565184	0.040014945	1.8905896

Pattern 6 : <Lower|
Number of occurrences : 124 Pattern probability : 0.10247934
Mutual information : 0.22085974 Entropy : 1.5747312
Normalized Mutual information : 0.10051759 Normalized Entropy : 0.7166911
Contribution to Mutual information : 0.02263356 Contribution to Entropy : 0.16137742
Contribution to Normalized Mutual information : 0.010300977 Contribution to Normalized Entropy : 0.07344603

#	1	0.5065523	0.25950414	0.24704814	0.04489747	5.502496
Top	2	0.015350069	0.09256198	-0.07721192	0.011040411	-6.9935727
Mid	3	0.0	0.033884298	-0.033884298	0.008064516	-4.201653
Bottom	4	0.058203403	0.061157025	-0.0029536225	0.021025285	-0.1404796
Higher	5	0.07550711	0.07933884	-0.0038317293	0.023726564	-0.1614953
Lower	6	0.030700138	0.10247934	-0.0717792	0.015491319	-4.633511
Same	7	0.12427214	0.13057852	-0.00630638	0.029625155	-0.2128725
Upset	8	0.12033946	0.12644628	-0.006106816	0.029218018	-0.2090086
Downstepped	9	0.06907531	0.11404958	-0.044974275	0.02277235	-1.974951

Pattern 7 : <Same|
Number of occurrences : 158 Pattern probability : 0.13057852
Mutual information : 0.11425485 Entropy : 1.9447712
Normalized Mutual information : 0.05199962 Normalized Entropy : 0.88510346
Contribution to Mutual information : 0.014919229 Contribution to Entropy : 0.25394535
Contribution to Normalized Mutual information : 0.0067900335 Contribution to Normalized Entropy : 0.11557551

#	1	0.25787205	0.25950414	-0.0016320944	0.034802705	-0.0468956
Top	2	0.18238994	0.09256198	0.08982796	0.030721683	2.923927
Mid	3	0.006289309	0.033884298	-0.02759499	0.0062893084	-4.3876033
Bottom	4	0.031446543	0.061157025	-0.029710483	0.013884163	-2.1398828
Higher	5	0.1509434	0.07933884	0.07160456	0.02848045	2.5141652
Lower	6	0.04402516	0.10247934	-0.05845418	0.01632094	-3.5815449
Same	7	0.08805032	0.13057852	-0.042528197	0.022543553	-1.8864905
Upset	8	0.12565102	0.12644628	-7.952601E-4	0.026369195	-0.0301587
Downstepped	9	0.113332294	0.11404958	-7.1728975E-4	0.025219051	-0.0284424

Pattern 8 : <Upset|
Number of occurrences : 153 Pattern probability : 0.12644628
Mutual information : 0.22980063 Entropy : 1.8886449
Normalized Mutual information : 0.10458677 Normalized Entropy : 0.85955936
Contribution to Mutual information : 0.029057434 Contribution to Entropy : 0.23881212
Contribution to Normalized Mutual information : 0.013224608 Contribution to Normalized Entropy : 0.10868808

#	1	0.24839462	0.25950414	-0.011109516	0.03493175	-0.3180349
Top	2	0.23147719	0.09256198	0.13891521	0.034098618	4.073925
Mid	3	0.0	0.033884298	-0.033884298	0.006535948	-5.1842976
Bottom	4	0.11261052	0.061157025	0.051453494	0.025556471	2.0133255
Higher	5	0.0759423	0.07933884	-0.0033965409	0.021416362	-0.1585956
Lower	6	0.1251228	0.10247934	0.022643462	0.026748292	0.8465386
Same	7	0.02502456	0.13057852	-0.105553955	0.012627994	-8.358727
Upset	8	0.04379298	0.12644628	-0.0826533	0.016543696	-4.9960604
Downstepped	9	0.13763508	0.11404958	0.023585498	0.027852513	0.8467997

Pattern 9 : <Downstepped|
Number of occurrences : 138 Pattern probability : 0.11404958
Mutual information : 0.114423096 Entropy : 1.8446614
Normalized Mutual information : 0.052076194 Normalized Entropy : 0.83954155
Contribution to Mutual information : 0.013049906 Contribution to Entropy : 0.21038286
Contribution to Normalized Mutual information : 0.005939268 Contribution to Normalized Entropy : 0.09574936

#	1	0.34037766	0.25950414	0.08087352	0.040335614	2.0050151
Top	2	0.13473281	0.09256198	0.04217083	0.029065128	1.4509081
Mid	3	0.007091201	0.033884298	-0.026793096	0.0071429075	-3.751007
Bottom	4	0.028364804	0.061157025	-0.03279222	0.014131946	-2.3204322
Higher	5	0.12055042	0.07933884	0.041211575	0.027717259	1.4868561
Lower	6	0.028364804	0.10247934	-0.07411453	0.014131946	-5.2444677
Same	7	0.12778227	0.13057852	-0.0027962476	0.028418962	-0.0983937
Upset	8	0.15600643	0.12644628	0.029560149	0.030888816	0.9569855
Downstepped	9	0.056729607	0.11404958	-0.057319976	0.019691711	-2.9108682

7.2. Analyse mixte avec les bi-, tri- et quadrigrammes

Number of categories : 9

Category 0 : EOW
Category 1 : Top
Category 2 : Mid
Category 3 : Bottom
Category 4 : Higher
Category 5 : Lower
Category 6 : Same
Category 7 : Upset
Category 8 : Downstepped

Entropy : 2.0653682 Maximum Entropy : 2.1972246 Normalized Entropy : 0.93998957

---------- PATTERNS MODEL ----------
Number of patterns : 56
Model Entropy : 1.5404831 Mutual Information of the model : 0.52488506
Normalized Model Entropy : 0.70110404 Normalized Mutual Information of the model : 0.23888548

Pattern 0 : <|
Number of occurrences : 1210 Pattern probability : 0.0
Mutual information : 0.0 Entropy : 2.0653682
Normalized Mutual information : 0.0 Normalized Entropy : **0.93998957**
Contribution to Mutual information : 0.0 Contribution to Entropy : 0.0
Contribution to Normalized Mutual information : 0.0Contribution to Normalized Entropy : 0.0

EOW	1	0.25950414	0.25950414	0.0	0.012602038	0.0
Top	20	0.09256198	0.09256198	0.0	0.008331671	0.0
Mid	23	0.033884298	0.033884298	0.0	0.0052014105	0.0
Bottom	25	0.061157025	0.061157025	0.0	0.0068885367	0.0
Higher	28	0.07933884	0.07933884	0.0	0.0077696266	0.0
Lower	33	0.10247934	0.10247934	0.0	0.008718619	0.0
Same	38	0.13057852	0.13057852	0.0	0.00968631	0.0
Upset	44	0.12644628	0.12644628	0.0	0.009554438	0.0
Downstepped	51	0.11404958	0.11404958	0.0	0.009138162	0.0

Pattern 1 : <EOW|
Number of occurrences : 314 Pattern probability : 0.25950414
Mutual information : 0.4534335 Entropy : 1.8933698
Normalized Mutual information : 0.20636648 Normalized Entropy : **0.8617097**
Contribution to Mutual information : 0.117667876 Contribution to Entropy : 0.4913373
Contribution to Normalized Mutual information : 0.05355296Contribution to Normalized Entropy : 0.22361724

EOW	1	0.0	0.25950414	-0.25950414	0.0031847134	-81.4843
Top	20	0.01590866	0.09256198	-0.076653324	0.007061057	-10.855786
Mid	2	0.12408756	0.033884298	0.090203255	0.018605	4.8483343
Bottom	3	0.061099775	0.061157025	-5.725026E-5	0.013516515	-0.0042355787
Higher	5	0.079264574	0.07933884	-7.426739E-5	0.015245517	-0.0048714248
Lower	7	0.102383405	0.10247934	-9.5933676E-5	0.01710785	-0.005607582
Same	8	0.24181165	0.13057852	0.11123313	0.02416364	4.603327
Upset	13	0.22590297	0.12644628	0.0994567	0.023599014	4.214443
Downstepped	16	0.14954141	0.11404958	0.035491824	0.020125305	1.7635422

Pattern 2 : <EOW|Mid|
Number of occurrences : 39 Pattern probability : 0.032231405
Mutual information : 0.45872924 Entropy : 1.8892643
Normalized Mutual information : 0.20877667 Normalized Entropy : **0.8598412**
Contribution to Mutual information : 0.014785488 Contribution to Entropy : 0.060893644
Contribution to Normalized Mutual information : 0.0067291656 Contribution to Normalized Entropy : 0.027713891

EOW	1	0.0	0.25950414	-0.25950414	0.025641026	-10.120662
Top	20	0.14856651	0.09256198	0.05600453	0.056951284	0.9833761
Mid	23	0.0	0.033884298	-0.033884298	0.025641026	-1.3214877
Bottom	25	0.12380544	0.061157025	0.062648416	0.05273972	1.1878792
Higher	24	0.22284979	0.07933884	0.14351094	0.06663866	2.1535687
Lower	33	0.10403749	0.10247934	0.0015581548	0.048888594	0.031871542
Same	38	0.07428326	0.13057852	-0.05629526	0.04199061	-1.3406631
Upset	44	0.12836884	0.12644628	0.0019225627	0.053562876	0.035893567
Downstepped	51	0.1980887	0.11404958	0.08403912	0.06382057	1.3168031

Pattern 3 : <EOW|Bottom|
Number of occurrences : 20 Pattern probability : 0.016528925
Mutual information : 0.9198724 Entropy : 1.1238272
Normalized Mutual information : 0.41865197 Normalized Entropy : **0.5114758**
Contribution to Mutual information : 0.015204502 Contribution to Entropy : 0.018575655
Contribution to Normalized Mutual information : 0.006919867 Contribution to Normalized Entropy : 0.008454145

EOW	1	0.053701013	0.25950414	-0.20580313	0.05040695	-4.0828323
Top	20	0.05805515	0.09256198	-0.03450683	0.052289937	-0.6599134
Mid	23	0.0	0.033884298	-0.033884298	0.05	-0.677686
Bottom	25	0.0	0.061157025	-0.061157025	0.05	-1.2231405
Higher	28	0.029027576	0.07933884	-0.050311267	0.037539963	-1.3402056
Lower	33	0.0	0.10247934	-0.10247934	0.05	-2.0495868
Same	4	0.53701013	0.13057852	0.40643162	0.111496694	3.6452346
Upset	27	0.3222061	0.12644628	0.19575983	0.10449625	1.8733671
Downstepped	51	0.0	0.11404958	-0.11404958	0.05	-2.2809916

Pattern 4 : <EOW|Bottom|Same|
Number of occurrences : 10 Pattern probability : 0.008264462
Mutual information : 0.6760306 Entropy : 1.0163456
Normalized Mutual information : 0.30767477 Normalized Entropy : **0.4625588**
Contribution to Mutual information : 0.0055870293 Contribution to Entropy : 0.00839955
Contribution to Normalized Mutual information : 0.0025427665 Contribution to Normalized Entropy : 0.0038227998

EOW	1	0.6365109	0.25950414	0.37700677	0.15210679	2.4785664
Top	39	0.18381424	0.09256198	0.09125226	0.12248534	0.74500555
Mid	23	0.0	0.033884298	-0.033884298	0.1	-0.338843
Bottom	25	0.0	0.061157025	-0.061157025	0.1	-0.61157024
Higher	40	0.053042583	0.07933884	-0.026296258	0.07087247	-0.3710363
Lower	33	0.0	0.10247934	-0.10247934	0.1	-1.0247934
Same	41	0.0	0.13057852	-0.13057852	0.1	-1.3057852
Upset	42	0.12663223	0.12644628	1.8595159E-4	0.10516488	0.001768191
Downstepped	43	0.0	0.11404958	-0.11404958	0.1	-1.1404958

Pattern 5 : <EOW|Higher|
Number of occurrences : 23 Pattern probability : 0.019008264
Mutual information : 0.8780444 Entropy : 1.4708109
Normalized Mutual information : 0.39961523 Normalized Entropy : **0.66939485**
Contribution to Mutual information : 0.0166901 Contribution to Entropy : 0.027957562
Contribution to Normalized Mutual information : 0.0075959917 Contribution to Normalized Entropy : 0.0127240345

EOW	1	0.0	0.25950414	-0.25950414	0.04347826	-5.968595
Top	29	0.09591645	0.09256198	0.0033544675	0.061402623	0.054630686
Mid	23	0.0	0.033884298	-0.033884298	0.04347826	-0.77933884
Bottom	30	0.17812435	0.061157025	0.11696733	0.07978124	1.4661006
Higher	28	0.021588339	0.07933884	-0.0577505	0.030304456	-1.9056768
Lower	6	0.44531086	0.10247934	0.34283152	0.10363169	3.3081727
Same	38	0.064765014	0.13057852	-0.065813504	0.05131764	-1.2824733
Upset	44	0.0	0.12644628	-0.12644628	0.04347826	-2.9082644
Downstepped	32	0.19429503	0.11404958	0.08024545	0.08250019	0.9726699

Pattern 6 : <EOW|Higher|Lower|
Number of occurrences : 10 Pattern probability : 0.008264462
Mutual information : 0.4196182 Entropy : 1.3776207
Normalized Mutual information : 0.19097647 Normalized Entropy : **0.6269822**
Contribution to Mutual information : 0.0034679188 Contribution to Entropy : 0.011385295
Contribution to Normalized Mutual information : 0.0015783178 Contribution to Normalized Entropy : 0.0051816707

EOW	1	0.5015657	0.25950414	0.24206156	0.1581131	1.5309392
Top	20	0.0	0.09256198	-0.09256198	0.1	-0.92561984
Mid	23	0.0	0.033884298	-0.033884298	0.1	-0.338843
Bottom	25	0.0	0.061157025	-0.061157025	0.1	-0.61157024
Higher	34	0.15634039	0.07933884	0.07700155	0.114846885	0.6704714
Lower	33	0.09989029	0.10247934	-0.002589047	0.09482206	-0.02730427
Same	35	0.12304879	0.13057852	-0.007529728	0.10387867	-0.0724858
Upset	36	0.119154826	0.12644628	-0.007291451	0.1024485	-0.071171865
Downstepped	37	0.0	0.11404958	-0.11404958	0.1	-1.1404958

Pattern 7 : <EOW|Lower|
Number of occurrences : 32 Pattern probability : 0.026446281
Mutual information : 0.37219113 Entropy : 1.8230588
Normalized Mutual information : 0.16939148 Normalized Entropy : **0.8297098**
Contribution to Mutual information : 0.009843071 Contribution to Entropy : 0.048213128
Contribution to Normalized Mutual information : 0.0044797747 Contribution to Normalized Entropy : 0.02194274

EOW	1	0.09531796	0.25950414	-0.16418618	0.051911116	-3.1628327
Top	20	0.015606797	0.09256198	-0.076955184	0.021911202	-3.5121388
Mid	23	0.0	0.033884298	-0.033884298	0.03125	-1.0842975
Bottom	25	0.12709062	0.061157025	0.06593359	0.0588798	1.1197999
Higher	34	0.15886328	0.07933884	0.079524435	0.064620465	1.2306385
Lower	33	0.031213595	0.10247934	-0.07126574	0.0307405	-2.3183014
Same	35	0.25418124	0.13057852	0.12360272	0.076968566	1.6058857
Upset	36	0.06354531	0.12644628	-0.06290097	0.04312312	-1.4586368
Downstepped	37	0.25418124	0.11404958	0.14013165	0.076968566	1.8206348

Pattern 8 : <EOW|Same|
Number of occurrences : 76 Pattern probability : 0.062809914
Mutual information : 0.5577429 Entropy : 1.7235606
Normalized Mutual information : 0.25383973 Normalized Entropy : **0.7844262**
Contribution to Mutual information : 0.035031784 Contribution to Entropy : 0.10825669
Contribution to Normalized Mutual information : 0.015943652 Contribution to Normalized Entropy : 0.049269743

EOW	1	0.0	0.25950414	-0.25950414	0.013157895	-19.722315
Top	39	0.30575567	0.09256198	0.21319368	0.052848984	4.0340166
Mid	23	0.006643063	0.033884298	-0.027241236	0.009318158	-2.9234571
Bottom	25	0.03321531	0.061157025	-0.027941715	0.020555463	-1.3593328
Higher	9	0.15943348	0.07933884	0.08009464	0.041992236	1.907368
Lower	33	0.013897985	0.10247934	-0.08858135	0.013428578	-6.5964804
Same	10	0.13897985	0.13057852	0.008401334	0.039680377	0.21172516
Upset	11	0.22236776	0.12644628	0.09592149	0.047699764	2.0109425
Downstepped	12	0.11970688	0.11404958	0.005657293	0.03723629	0.15192956

Pattern 9 : <EOW|Same|Higher|
Number of occurrences : 14 Pattern probability : 0.011570248
Mutual information : 0.67908454 Entropy : 1.2539325
Normalized Mutual information : 0.3090647 Normalized Entropy : **0.57068926**
Contribution to Mutual information : 0.0078571765 Contribution to Entropy : 0.01450831
Contribution to Normalized Mutual information : 0.003575955 Contribution to Normalized Entropy : 0.006603016

EOW	1	0.43319172	0.25950414	0.17368758	0.13243237	1.311519
Top	29	0.2887945	0.09256198	0.1962325	0.12112338	1.6201042
Mid	23	0.0	0.033884298	-0.033884298	0.071428575	-0.47438017
Bottom	30	0.092671275	0.061157025	0.03151425	0.07749807	0.4066456
Higher	28	0.0	0.07933884	-0.07933884	0.071428575	-1.1107438
Lower	31	0.0	0.10247934	-0.10247934	0.071428575	-1.4347107
Same	38	0.0	0.13057852	-0.13057852	0.071428575	-1.8280993
Upset	44	0.0	0.12644628	-0.12644628	0.071428575	-1.7702479
Downstepped	32	0.18534255	0.11404958	0.07129297	0.103851095	0.6864922

Pattern 10 : <EOW|Same|Same|
Number of occurrences : 10 Pattern probability : 0.008264462
Mutual information : 0.922235 Entropy : 1.4374679
Normalized Mutual information : 0.41972724 Normalized Entropy : **0.65421987**
Contribution to Mutual information : 0.007621777 Contribution to Entropy : 0.0118799
Contribution to Normalized Mutual information : 0.0034688201 Contribution to Normalized Entropy : 0.005406775

EOW	1	0.0	0.25950414	-0.25950414	0.1	-2.5950413
Top	39	0.2556228	0.09256198	0.16306081	0.13794194	1.1820974
Mid	23	0.0	0.033884298	-0.033884298	0.1	-0.338843
Bottom	25	0.0	0.061157025	-0.061157025	0.1	-0.61157024
Higher	40	0.40043375	0.07933884	0.3210949	0.15494727	2.072285
Lower	33	0.061702058	0.10247934	-0.04077728	0.076088704	-0.53591764
Same	41	0.123404115	0.13057852	-0.0071744025	0.10400747	-0.06897969
Upset	42	0.0	0.12644628	-0.12644628	0.1	-1.2644627
Downstepped	43	0.15883726	0.11404958	0.044787675	0.11558892	0.3874738

Pattern 11 : <EOW|Same|Upset|
Number of occurrences : 16 Pattern probability : 0.0132231405
Mutual information : 0.5051315 Entropy : 1.758797
Normalized Mutual information : 0.22989523 Normalized Entropy : **0.800463**
Contribution to Mutual information : 0.0066794246 Contribution to Entropy : 0.02325682
Contribution to Normalized Mutual information : 0.003039937 Contribution to Normalized Entropy : 0.010584635

EOW	1	0.14267258	0.25950414	-0.116831556	0.08743466	-1.3362156
Top	45	0.26470724	0.09256198	0.17214525	0.110294305	1.5607809
Mid	23	0.0	0.033884298	-0.033884298	0.0625	-0.54214877
Bottom	47	0.12877649	0.061157025	0.067619465	0.083738096	0.8075114
Higher	48	0.16336523	0.07933884	0.08402639	0.09242464	0.9091341
Lower	49	0.14308499	0.10247934	0.04060565	0.087539874	0.46385318
Same	38	0.0	0.13057852	-0.13057852	0.0625	-2.0892563
Upset	44	0.0	0.12644628	-0.12644628	0.0625	-2.0231404
Downstepped	50	0.1573935	0.11404958	0.043343917	0.09104284	0.47608265

Pattern 12 : <EOW|Same|Downstepped|
Number of occurrences : 11 Pattern probability : 0.009090909
Mutual information : 0.88271445 Entropy : 0.6958374
Normalized Mutual information : 0.40174064 Normalized Entropy : **0.31668922**
Contribution to Mutual information : 0.008024677 Contribution to Entropy : 0.006325794
Contribution to Normalized Mutual information : 0.0036521878 Contribution to Normalized Entropy : 0.0028789928

EOW	1	0.7671383	0.25950414	0.50763416	0.12743524	3.9834678
Top	52	0.0	0.09256198	-0.09256198	0.09090909	-1.0181818
Mid	23	0.0	0.033884298	-0.033884298	0.09090909	-0.37272727
Bottom	25	0.0855651	0.061157025	0.024408072	0.084338985	0.2894044
Higher	53	0.0	0.07933884	-0.07933884	0.09090909	-0.8727273
Lower	33	0.0	0.10247934	-0.10247934	0.09090909	-1.1272727
Same	54	0.0	0.13057852	-0.13057852	0.09090909	-1.4363637
Upset	55	0.14729661	0.12644628	0.02085033	0.10685603	0.19512545
Downstepped	51	0.0	0.11404958	-0.11404958	0.09090909	-1.2545455

Pattern 13 : <EOW|Upset|
Number of occurrences : 71 Pattern probability : 0.058677685
Mutual information : 0.5185114 Entropy : 1.8151543
Normalized Mutual information : 0.23598471 Normalized Entropy : **0.82611233**
Contribution to Mutual information : 0.03042505 Contribution to Entropy : 0.10650905
Contribution to Normalized Mutual information : 0.013847036 Contribution to Normalized Entropy : 0.048474357

EOW	1	0.014660258	0.25950414	-0.24484389	0.01426378	-17.165428
Top	45	0.24093959	0.09256198	0.1483776	0.050753158	2.9235146
Mid	23	0.0	0.033884298	-0.033884298	0.014084507	-2.405785
Bottom	47	0.16126284	0.061157025	0.100105815	0.04364665	2.293551
Higher	48	0.043980774	0.07933884	-0.035358068	0.024335237	-1.4529576
Lower	14	0.20524362	0.10247934	0.10276428	0.047931682	2.143974
Same	38	0.02604752	0.13057852	-0.104531	0.01890266	-5.5299625
Upset	44	0.10262181	0.12644628	-0.023824468	0.036014587	-0.66152275
Downstepped	15	0.20524362	0.11404958	0.09119403	0.047931682	1.9025836

Pattern 14 : <EOW|Upset|Lower|
Number of occurrences : 14 Pattern probability : 0.011570248
Mutual information : 0.56244147 Entropy : 1.1380904
Normalized Mutual information : 0.25597814 Normalized Entropy : **0.5179672**
Contribution to Mutual information : 0.006507587 Contribution to Entropy : 0.013167988
Contribution to Normalized Mutual information : 0.0029617306 Contribution to Normalized Entropy : 0.005993009

EOW	1	0.55890393	0.25950414	0.2993998	0.13270007	2.2562144
Top	20	0.0	0.09256198	-0.09256198	0.071428575	-1.2958678
Mid	23	0.0	0.033884298	-0.033884298	0.071428575	-0.47438017
Bottom	25	0.0	0.061157025	-0.061157025	0.071428575	-0.8561984
Higher	34	0.08331069	0.07933884	0.0039718524	0.07385798	0.05377689
Lower	33	0.0	0.10247934	-0.10247934	0.071428575	-1.4347107
Same	35	0.13711554	0.13057852	0.00653702	0.09192966	0.07110894
Upset	36	0.22066979	0.12644628	0.094223514	0.110832766	0.85014135
Downstepped	37	0.0	0.11404958	-0.11404958	0.071428575	-1.5966942

Pattern 15 : <EOW|Upset|Downstepped|
Number of occurrences : 14 Pattern probability : 0.011570248
Mutual information : 0.38005275 Entropy : 1.6520779
Normalized Mutual information : 0.17296946 Normalized Entropy : **0.75189304**
Contribution to Mutual information : 0.0043973043 Contribution to Entropy : 0.01911495
Contribution to Normalized Mutual information : 0.0020012993 Contribution to Normalized Entropy : 0.0086995885

EOW	1	0.16144161	0.25950414	-0.09806253	0.098335505	-0.99722403
Top	52	0.1768242	0.09256198	0.084262215	0.10196547	0.8263799
Mid	23	0.0	0.033884298	-0.033884298	0.071428575	-0.47438017
Bottom	25	0.0	0.061157025	-0.061157025	0.071428575	-0.8561984
Higher	53	0.059654836	0.07933884	-0.019684006	0.06329982	-0.31096464
Lower	33	0.0	0.10247934	-0.10247934	0.071428575	-1.4347107
Same	54	0.32288322	0.13057852	0.1923047	0.12496561	1.538861
Upset	55	0.20474382	0.12644628	0.07829754	0.107843734	0.7260277
Downstepped	51	0.074452296	0.11404958	-0.039597288	0.07015755	-0.5644052

Pattern 16 : <EOW|Downstepped|
Number of occurrences : 47 Pattern probability : 0.038842976
Mutual information : 0.51658815 Entropy : 1.8024913
Normalized Mutual information : 0.23510939 Normalized Entropy : **0.82034916**
Contribution to Mutual information : 0.020065822 Contribution to Entropy : 0.07001413
Contribution to Normalized Mutual information : 0.009132349 Contribution to Normalized Entropy : 0.031864803

EOW	1	0.0	0.25950414	-0.25950414	0.021276595	-12.196694
Top	17	0.21091503	0.09256198	0.11835305	0.059506815	1.988899
Mid	23	0.0	0.033884298	-0.033884298	0.021276595	-1.592562
Bottom	25	0.070305005	0.061157025	0.009147979	0.037291896	0.24530743
Higher	18	0.25778502	0.07933884	0.17844617	0.06380354	2.7968068
Lower	33	0.031242274	0.10247934	-0.071237065	0.025376398	-2.8072174
Same	54	0.14074515	0.13057852	0.01016663	0.050725784	0.20042332
Upset	19	0.17183252	0.12644628	0.04538624	0.055025354	0.82482415
Downstepped	51	0.11717501	0.11404958	0.0031254292	0.04691437	0.06661987

Pattern 17 : <EOW|Downstepped|Top|
Number of occurrences : 9 Pattern probability : 0.0074380166
Mutual information : 0.8618053 Entropy : 0.85329795
Normalized Mutual information : 0.3922245 Normalized Entropy : **0.38835263**
Contribution to Mutual information : 0.0064101224 Contribution to Entropy : 0.006346844
Contribution to Normalized Mutual information : 0.0029173724 Contribution to Normalized Entropy : 0.002888573

EOW	1	0.58947366	0.25950414	0.32996953	0.16397645	2.0122983
Top	20	0.0	0.09256198	-0.09256198	0.11111111	-0.8330578
Mid	23	0.0	0.033884298	-0.033884298	0.11111111	-0.30495867
Bottom	25	0.0	0.061157025	-0.061157025	0.11111111	-0.55041325
Higher	28	0.0	0.07933884	-0.07933884	0.11111111	-0.7140496
Lower	21	0.34736842	0.10247934	0.24488908	0.15871133	1.5429842
Same	38	0.06315789	0.13057852	-0.067420624	0.08108225	-0.83150905
Upset	44	0.0	0.12644628	-0.12644628	0.11111111	-1.1380165
Downstepped	22	0.0	0.11404958	-0.11404958	0.11111111	-1.0264462

Pattern 18 : <EOW|Downstepped|Higher|
Number of occurrences : 11 Pattern probability : 0.009090909
Mutual information : 0.5038022 Entropy : 1.5512633
Normalized Mutual information : 0.22929025 Normalized Entropy : **0.70601034**
Contribution to Mutual information : 0.0045800195 Contribution to Entropy : 0.014102394
Contribution to Normalized Mutual information : 0.0020844566 Contribution to Normalized Entropy : 0.0064182756

EOW	1	0.2163987	0.25950414	-0.04310544	0.1241591	-0.34717906
Top	29	0.2196195	0.09256198	0.12705752	0.12482233	1.017907
Mid	23	0.0	0.033884298	-0.033884298	0.09090909	-0.37272727
Bottom	30	0.0	0.061157025	-0.061157025	0.09090909	-0.6727273
Higher	28	0.0	0.07933884	-0.07933884	0.09090909	-0.8727273
Lower	31	0.24920125	0.10247934	0.14672191	0.13041891	1.1250049
Same	38	0.07869514	0.13057852	-0.051883377	0.081185654	-0.63907075
Upset	44	0.0	0.12644628	-0.12644628	0.09090909	-1.3909091
Downstepped	32	0.2360854	0.11404958	0.12203582	0.12804441	0.95307416

Pattern 19 : <EOW|Downstepped|Upset|
Number of occurrences : 9 Pattern probability : 0.0074380166
Mutual information : 1.2836907 Entropy : 1.0608569
Normalized Mutual information : 0.5842328 Normalized Entropy : **0.4828168**
Contribution to Mutual information : 0.009548113 Contribution to Entropy : 0.0078906715
Contribution to Normalized Mutual information : 0.0043455334 Contribution to Normalized Entropy : 0.0035911994

EOW	1	0.0	0.25950414	-0.25950414	0.11111111	-2.3355372
Top	45	0.44444445	0.09256198	0.35188246	0.16563468	2.1244493
Mid	23	0.0	0.033884298	-0.033884298	0.11111111	-0.30495867
Bottom	47	0.0	0.061157025	-0.061157025	0.11111111	-0.55041325
Higher	48	0.22222222	0.07933884	0.14288339	0.1385799	1.0310541
Lower	49	0.0	0.10247934	-0.10247934	0.11111111	-0.92231405
Same	38	0.0	0.13057852	-0.13057852	0.11111111	-1.1752067
Upset	44	0.0	0.12644628	-0.12644628	0.11111111	-1.1380165
Downstepped	50	0.33333334	0.11404958	0.21928376	0.15713483	1.3955134

Pattern 20 : <Top|
Number of occurrences : 112 Pattern probability : 0.014049587
Mutual information : 0.5546767 Entropy : 1.2757356
Normalized Mutual information : 0.25244424 Normalized Entropy : **0.5806123**
Contribution to Mutual information : 0.0077929785 Contribution to Entropy : 0.017923558
Contribution to Normalized Mutual information : 0.0035467374 Contribution to Normalized Entropy : 0.008157363

EOW	1	0.48855752	0.25950414	0.22905338	0.047233183	4.8494167
Top	20	0.0	0.09256198	-0.09256198	0.008928572	-10.366942
Mid	23	0.0	0.033884298	-0.033884298	0.008928572	-3.7950413
Bottom	25	0.05975745	0.061157025	-0.0013995767	0.022397881	-0.06248701
Higher	28	0.0	0.07933884	-0.07933884	0.008928572	-8.88595
Lower	21	0.28789997	0.10247934	0.18542063	0.04278409	4.3338685
Same	38	0.05234545	0.13057852	-0.07823306	0.021045318	-3.717362
Upset	44	0.0	0.12644628	-0.12644628	0.008928572	-14.1619835
Downstepped	22	0.11143956	0.11404958	-0.0026100203	0.029734058	-0.087778814

Pattern 21 : <Top|Lower|
Number of occurrences : 33 Pattern probability : 0.01983471
Mutual information : 0.9917611 Entropy : 0.49094862
Normalized Mutual information : 0.4513699 Normalized Entropy : **0.22344033**
Contribution to Mutual information : 0.019671293 Contribution to Entropy : 0.009737823
Contribution to Normalized Mutual information : 0.008952791 Contribution to Normalized Entropy : 0.004431874

EOW	1	0.8805539	0.25950414	0.62104976	0.056455597	11.000676
Top	20	0.0	0.09256198	-0.09256198	0.030303031	-3.0545454
Mid	23	0.0	0.033884298	-0.033884298	0.030303031	-1.1181818
Bottom	25	0.05832037	0.061157025	-0.0028366558	0.040794797	-0.06953474
Higher	34	0.0	0.07933884	-0.07933884	0.030303031	-2.6181817
Lower	33	0.030761832	0.10247934	-0.07171751	0.030058296	-2.3859472
Same	35	0.030363929	0.13057852	-0.100214586	0.029869393	-3.3550928
Upset	36	0.0	0.12644628	-0.12644628	0.030303031	-4.172727
Downstepped	37	0.0	0.11404958	-0.11404958	0.030303031	-3.7636364

Pattern 22 : <Top|Downstepped|
Number of occurrences : 10 Pattern probability : 0.008264462
Mutual information : 1.1168107 Entropy : 0.325083
Normalized Mutual information : 0.5082824 Normalized Entropy : **0.14795165**
Contribution to Mutual information : 0.00922984 Contribution to Entropy : 0.002686636
Contribution to Normalized Mutual information : 0.004200681 Contribution to Normalized Entropy : 0.0012227407

EOW	1	0.9	0.25950414	0.64049584	0.09486834	6.751418
Top	52	0.0	0.09256198	-0.09256198	0.1	-0.92561984
Mid	23	0.0	0.033884298	-0.033884298	0.1	-0.338843
Bottom	25	0.0	0.061157025	-0.061157025	0.1	-0.61157024
Higher	53	0.0	0.07933884	-0.07933884	0.1	-0.7933884
Lower	33	0.1	0.10247934	-0.0024793372	0.094868325	-0.02613451
Same	54	0.0	0.13057852	-0.13057852	0.1	-1.3057852
Upset	55	0.0	0.12644628	-0.12644628	0.1	-1.2644627
Downstepped	51	0.0	0.11404958	-0.11404958	0.1	-1.1404958

Pattern 23 : <Mid|
Number of occurrences : 41 Pattern probability : 0.0016528926
Mutual information : 0.36641496 Entropy : 1.9574394
Normalized Mutual information : 0.16676264 Normalized Entropy : **0.8908691**
Contribution to Mutual information : 6.0564454E-4 Contribution to Entropy : 0.003235437
Contribution to Normalized Mutual information : 2.756407E-4 Contribution to Normalized Entropy : 0.0014725107

EOW	1	0.024162788	0.25950414	-0.23534136	0.023981165	-9.813591
Top	20	0.14497672	0.09256198	0.052414738	0.054985248	0.95325094
Mid	23	0.0	0.033884298	-0.033884298	0.024390243	-1.3892562
Bottom	25	0.12081394	0.061157025	0.059656918	0.050898734	1.1720707
Higher	24	0.21746509	0.07933884	0.13812625	0.06442504	2.143984
Lower	33	0.101523645	0.10247934	-9.556934E-4	0.047167733	-0.020261593
Same	38	0.07248836	0.13057852	-0.058090158	0.040495045	-1.4345003
Upset	44	0.12526707	0.12644628	-0.0011792034	0.051696874	-0.022809956
Downstepped	51	0.1933023	0.11404958	0.07925272	0.06167114	1.285086

Pattern 24 : <Mid|Higher|
Number of occurrences : 9 Pattern probability : 0.0074380166
Mutual information : 0.58204734 Entropy : 1.261157
Normalized Mutual information : 0.26490116 Normalized Entropy : **0.5739773**
Contribution to Mutual information : 0.0043292777 Contribution to Entropy : 0.009380507
Contribution to Normalized Mutual information : 0.001970339 Contribution to Normalized Entropy : 0.004269253

EOW	1	0.4189189	0.25950414	0.15941477	0.16446069	0.9693184
Top	29	0.0	0.09256198	-0.09256198	0.11111111	-0.8330578
Mid	23	0.0	0.033884298	-0.033884298	0.11111111	-0.30495867
Bottom	30	0.0	0.061157025	-0.061157025	0.11111111	-0.55041325
Higher	28	0.0	0.07933884	-0.07933884	0.11111111	-0.7140496
Lower	31	0.25675675	0.10247934	0.15427741	0.14561468	1.0594908
Same	38	0.081081085	0.13057852	-0.049497433	0.09098653	-0.5440084
Upset	44	0.0	0.12644628	-0.12644628	0.11111111	-1.1380165
Downstepped	32	0.24324325	0.11404958	0.12919366	0.1430136	0.9033663

Pattern 25 : <Bottom|
Number of occurrences : 74 Pattern probability : 0.02231405
Mutual information : 0.47556612 Entropy : 1.2735273
Normalized Mutual information : 0.21643947 Normalized Entropy : **0.57960725**
Contribution to Mutual information : 0.010611806 Contribution to Entropy : 0.02841755
Contribution to Normalized Mutual information : 0.004829641 Contribution to Normalized Entropy : 0.012933384

EOW	1	0.47297296	0.25950414	0.21346882	0.058038842	3.6780338
Top	20	0.054054055	0.09256198	-0.038507927	0.02628642	-1.4649361
Mid	23	0.0	0.033884298	-0.033884298	0.013513514	-2.507438
Bottom	25	0.0	0.061157025	-0.061157025	0.013513514	-4.52562
Higher	28	0.027027028	0.07933884	-0.052311815	0.01885097	-2.77502
Lower	33	0.0	0.10247934	-0.10247934	0.013513514	-7.5834713
Same	26	0.25675675	0.13057852	0.12617823	0.05078209	2.4846997
Upset	27	0.1891892	0.12644628	0.06274292	0.045529418	1.3780743
Downstepped	51	0.0	0.11404958	-0.11404958	0.013513514	-8.43967

Pattern 26 : <Bottom|Same|
Number of occurrences : 19 Pattern probability : 0.0074380166
Mutual information : 0.504427 Entropy : 1.2159387
Normalized Mutual information : 0.22957462 Normalized Entropy : **0.55339754**
Contribution to Mutual information : 0.0037519366 Contribution to Entropy : 0.009044172
Contribution to Normalized Mutual information : 0.0017075799 Contribution to Normalized Entropy : 0.00411618

EOW	1	0.5846319	0.25950414	0.32512778	0.11305272	2.8758953
Top	39	0.1688324	0.09256198	0.076270424	0.08594001	0.88748443
Mid	23	0.0	0.033884298	-0.033884298	0.05263158	-0.6438017
Bottom	25	0.0	0.061157025	-0.061157025	0.05263158	-1.1619835
Higher	40	0.04871933	0.07933884	-0.03061951	0.04938877	-0.61996907
Lower	33	0.0	0.10247934	-0.10247934	0.05263158	-1.9471074
Same	41	0.0815053	0.13057852	-0.04907322	0.06277034	-0.78178996
Upset	42	0.116311036	0.12644628	-0.010135241	0.07355012	-0.13780047
Downstepped	43	0.0	0.11404958	-0.11404958	0.05263158	-2.1669421

Pattern 27 : <Bottom|Upset|
Number of occurrences : 14 Pattern probability : 0.011570248
Mutual information : 0.57801664 Entropy : 1.1634672
Normalized Mutual information : 0.2630667 Normalized Entropy : **0.5295167**
Contribution to Mutual information : 0.0066877957 Contribution to Entropy : 0.013461604
Contribution to Normalized Mutual information : 0.003043747 Contribution to Normalized Entropy : 0.00612664

EOW	1	0.628129	0.25950414	0.36862487	0.12916848	2.8538299
Top	45	0.069792114	0.09256198	-0.022769868	0.06809719	-0.33437312
Mid	23	0.0	0.033884298	-0.033884298	0.071428575	-0.47438017
Bottom	47	0.11003056	0.061157025	0.048873536	0.08363348	0.58437765
Higher	48	0.0	0.07933884	-0.07933884	0.071428575	-1.1107438
Lower	49	0.12225618	0.10247934	0.019776843	0.08754983	0.22589242
Same	38	0.069792114	0.13057852	-0.060786404	0.06809719	-0.8926419
Upset	44	0.0	0.12644628	-0.12644628	0.071428575	-1.7702479
Downstepped	50	0.0	0.11404958	-0.11404958	0.071428575	-1.5966942

Pattern 28 : <Higher|
Number of occurrences : 96 Pattern probability : 0.0033057851
Mutual information : 0.24863575 Entropy : 1.75197
Normalized Mutual information : 0.11315901 Normalized Entropy : **0.7973559**
Contribution to Mutual information : 8.2193635E-4 Contribution to Entropy : 0.0057916366
Contribution to Normalized Mutual information : 3.7407936E-4 Contribution to Normalized Entropy : 0.0026358874

EOW	1	0.326708	0.25950414	0.06720385	0.04786808	1.4039387
Top	29	0.09364875	0.09256198	0.001086764	0.029734707	0.03654867
Mid	23	0.0	0.033884298	-0.033884298	0.010416667	-3.2528925
Bottom	30	0.09485071	0.061157025	0.033693686	0.029905068	1.1266881
Higher	28	0.021077937	0.07933884	-0.058260903	0.014660625	-3.9739714
Lower	31	0.20024039	0.10247934	0.09776105	0.04084322	2.3935688
Same	38	0.06323381	0.13057852	-0.06734471	0.024840176	-2.7111204
Upset	44	0.010538968	0.12644628	-0.11590731	0.010422281	-11.121108
Downstepped	32	0.18970142	0.11404958	0.07565184	0.040014945	1.8905896

Pattern 29 : <Higher|Top|
Number of occurrences : 10 Pattern probability : 0.008264462
Mutual information : 0.7322606 Entropy : 1.0313339
Normalized Mutual information : 0.33326614 Normalized Entropy : **0.4693803**
Contribution to Mutual information : 0.00605174 Contribution to Entropy : 0.00852342
Contribution to Normalized Mutual information : 0.0027542657 Contribution to Normalized Entropy : 0.0038791758

EOW	1	0.5003367	0.25950414	0.24083257	0.15811384	1.5231594
Top	20	0.0	0.09256198	-0.09256198	0.1	-0.92561984
Mid	23	0.0	0.033884298	-0.033884298	0.1	-0.338843
Bottom	25	0.0	0.061157025	-0.061157025	0.1	-0.61157024
Higher	28	0.0	0.07933884	-0.07933884	0.1	-0.7933884
Lower	21	0.2948413	0.10247934	0.19236195	0.14419082	1.334079
Same	38	0.20482203	0.13057852	0.074243516	0.12762052	0.5817522
Upset	44	0.0	0.12644628	-0.12644628	0.1	-1.2644627
Downstepped	22	0.0	0.11404958	-0.11404958	0.1	-1.1404958

Pattern 30 : <Higher|Bottom|
Number of occurrences : 9 Pattern probability : 0.0074380166
Mutual information : 0.6831857 Entropy : 0.94654036
Normalized Mutual information : 0.3109312 Normalized Entropy : **0.43078908**
Contribution to Mutual information : 0.0050815465 Contribution to Entropy : 0.007040383
Contribution to Normalized Mutual information : 0.0023127114 Contribution to Normalized Entropy : 0.0032042162

EOW	1	0.5991903	0.25950414	0.33968616	0.16335417	2.0794458
Top	20	0.0	0.09256198	-0.09256198	0.11111111	-0.8330578
Mid	23	0.0	0.033884298	-0.033884298	0.11111111	-0.30495867
Bottom	25	0.0	0.061157025	-0.061157025	0.11111111	-0.55041325
Higher	28	0.0	0.07933884	-0.07933884	0.11111111	-0.7140496
Lower	33	0.0	0.10247934	-0.10247934	0.11111111	-0.92231405
Same	26	0.23076923	0.13057852	0.100190714	0.14044169	0.71339726
Upset	27	0.17004049	0.12644628	0.04359421	0.12522279	0.3481332
Downstepped	51	0.0	0.11404958	-0.11404958	0.11111111	-1.0264462

Pattern 31 : <Higher|Lower|
Number of occurrences : 19 Pattern probability : 0.0074380166
Mutual information : 0.3517419 Entropy : 1.4376444
Normalized Mutual information : 0.16008464 Normalized Entropy : **0.65430015**
Contribution to Mutual information : 0.002616262 Contribution to Entropy : 0.010693222
Contribution to Normalized Mutual information : 0.0011907122 Contribution to Normalized Entropy : 0.004866695

EOW	1	0.50211656	0.25950414	0.24261242	0.11470684	2.115065
Top	20	0.0	0.09256198	-0.09256198	0.05263158	-1.7586777
Mid	23	0.0	0.033884298	-0.033884298	0.05263158	-0.6438017
Bottom	25	0.0	0.061157025	-0.061157025	0.05263158	-1.1619835
Higher	34	0.1565121	0.07933884	0.077173255	0.08335595	0.9258278
Lower	33	0.030431308	0.10247934	-0.07204803	0.039406948	-1.8283079
Same	35	0.12318393	0.13057852	-0.0073945895	0.075397074	-0.098075286
Upset	36	0.11928569	0.12644628	-0.007160589	0.07435923	-0.09629725
Downstepped	37	0.06847044	0.11404958	-0.045579143	0.057939276	-0.7866709

Pattern 32 : <Higher|Downstepped|
Number of occurrences : 18 Pattern probability : 0.014876033
Mutual information : 0.27393046 Entropy : 1.5463248
Normalized Mutual information : 0.124671124 Normalized Entropy : **0.70376277**
Contribution to Mutual information : 0.0040749987 Contribution to Entropy : 0.02300318
Contribution to Normalized Mutual information : 0.0018546118 Contribution to Normalized Entropy : 0.010469198

EOW	1	0.48841694	0.25950414	0.2289128	0.1178195	1.9429109
Top	52	0.13161159	0.09256198	0.039049603	0.07968339	0.4900595
Mid	23	0.0	0.033884298	-0.033884298	0.055555556	-0.60991734
Bottom	25	0.0	0.061157025	-0.061157025	0.055555556	-1.1008265
Higher	53	0.11775773	0.07933884	0.03841889	0.07597179	0.5056994
Lower	33	0.027707702	0.10247934	-0.074771635	0.038686763	-1.9327446
Same	54	0.12482205	0.13057852	-0.0057564676	0.07790361	-0.073892176
Upset	55	0.05426855	0.12644628	-0.07217772	0.053397603	-1.3517034
Downstepped	51	0.055415403	0.11404958	-0.05863418	0.05392615	-1.0873051

Pattern 33 : <Lower|
Number of occurrences : 124 Pattern probability : 0.016528925
Mutual information : 0.22085974 Entropy : 1.5747312
Normalized Mutual information : 0.10051759 Normalized Entropy : **0.7166911**
Contribution to Mutual information : 0.003650574 Contribution to Entropy : 0.026028614
Contribution to Normalized Mutual information : 0.0016614478 Contribution to Normalized Entropy : 0.011846133

EOW	1	0.5065523	0.25950414	0.24704814	0.04489747	5.502496
Top	20	0.015350069	0.09256198	-0.07721192	0.011040411	-6.9935727
Mid	23	0.0	0.033884298	-0.033884298	0.008064516	-4.201653
Bottom	25	0.058203403	0.061157025	-0.0029536225	0.021025285	-0.14047955
Higher	34	0.07550711	0.07933884	-0.0038317293	0.023726564	-0.16149533
Lower	33	0.030700138	0.10247934	-0.0717792	0.015491319	-4.633511
Same	35	0.12427214	0.13057852	-0.00630638	0.029625155	-0.21287248
Upset	36	0.12033946	0.12644628	-0.006106816	0.029218018	-0.20900857
Downstepped	37	0.06907531	0.11404958	-0.044974275	0.02277235	-1.974951

Pattern 34 : <Lower|Higher|
Number of occurrences : 10 Pattern probability : 0.008264462
Mutual information : 0.6408529 Entropy : 1.020761
Normalized Mutual information : 0.29166472 Normalized Entropy : **0.46456835**
Contribution to Mutual information : 0.005296305 Contribution to Entropy : 0.008436041
Contribution to Normalized Mutual information : 0.002410452 Contribution to Normalized Entropy : 0.0038394076

EOW	1	0.6338582	0.25950414	0.37435406	0.15234236	2.457321
Top	29	0.09893338	0.09256198	0.006371394	0.09441693	0.06748147
Mid	23	0.0	0.033884298	-0.033884298	0.1	-0.338843
Bottom	30	0.0	0.061157025	-0.061157025	0.1	-0.61157024
Higher	28	0.0	0.07933884	-0.07933884	0.1	-0.7933884
Lower	31	0.0	0.10247934	-0.10247934	0.1	-1.0247934
Same	38	0.06680211	0.13057852	-0.06377641	0.07895542	-0.8077522
Upset	44	0.0	0.12644628	-0.12644628	0.1	-1.2644627
Downstepped	32	0.20040633	0.11404958	0.086356744	0.12658738	0.68219084

Pattern 35 : <Lower|Same|
Number of occurrences : 15 Pattern probability : 0.012396694
Mutual information : 0.24346255 Entropy : 1.6704427
Normalized Mutual information : 0.11080458 Normalized Entropy : **0.7602512**
Contribution to Mutual information : 0.0030181308 Contribution to Entropy : 0.020707967
Contribution to Normalized Mutual information : 0.0013736105 Contribution to Normalized Entropy : 0.009424602

EOW	1	0.44029456	0.25950414	0.18079042	0.12817572	1.4104888
Top	39	0.05503682	0.09256198	-0.037525162	0.058882803	-0.6372856
Mid	23	0.0	0.033884298	-0.033884298	0.06666667	-0.5082645
Bottom	25	0.11007364	0.061157025	0.048916616	0.080811486	0.6053176
Higher	40	0.12461166	0.07933884	0.04527282	0.08527742	0.5308887
Lower	33	0.0	0.10247934	-0.10247934	0.06666667	-1.5371901
Same	41	0.07269014	0.13057852	-0.05788838	0.06703545	-0.86354876
Upset	42	0.103731476	0.12644628	-0.022714801	0.07872791	-0.28852287
Downstepped	43	0.09356173	0.11404958	-0.020487852	0.07519217	-0.27247322

Pattern 36 : <Lower|Upset|
Number of occurrences : 12 Pattern probability : 0.009917355
Mutual information : 0.8901304 Entropy : 0.7346988
Normalized Mutual information : 0.40511578 Normalized Entropy : **0.33437583**
Contribution to Mutual information : 0.008827739 Contribution to Entropy : 0.0072862688
Contribution to Normalized Mutual information : 0.004017677 Contribution to Normalized Entropy : 0.0033161237

EOW	1	0.7498159	0.25950414	0.49031174	0.12503067	3.921532
Top	45	0.0	0.09256198	-0.09256198	0.083333336	-1.1107438
Mid	23	0.0	0.033884298	-0.033884298	0.083333336	-0.40661156
Bottom	47	0.11258287	0.061157025	0.051425844	0.09124507	0.5636014
Higher	48	0.0	0.07933884	-0.07933884	0.083333336	-0.95206606
Lower	49	0.0	0.10247934	-0.10247934	0.083333336	-1.2297521
Same	38	0.0	0.13057852	-0.13057852	0.083333336	-1.5669422
Upset	44	0.0	0.12644628	-0.12644628	0.083333336	-1.5173553
Downstepped	50	0.13760129	0.11404958	0.023551702	0.09944311	0.23683594

Pattern 37 : <Lower|Downstepped|
Number of occurrences : 9 Pattern probability : 0.0074380166
Mutual information : 0.6837016 Entropy : 1.2433846
Normalized Mutual information : 0.311166 Normalized Entropy : **0.5658887**
Contribution to Mutual information : 0.0050853835 Contribution to Entropy : 0.009248315
Contribution to Normalized Mutual information : 0.0023144577 Contribution to Normalized Entropy : 0.0042090896

EOW	1	0.32725307	0.25950414	0.067748934	0.1564035	0.43316764
Top	52	0.12953766	0.09256198	0.036975674	0.111931354	0.3303424
Mid	23	0.0	0.033884298	-0.033884298	0.11111111	-0.30495867
Bottom	25	0.0	0.061157025	-0.061157025	0.11111111	-0.55041325
Higher	53	0.11590212	0.07933884	0.036563277	0.1067025	0.3426656
Lower	33	0.0	0.10247934	-0.10247934	0.11111111	-0.92231405
Same	54	0.0	0.13057852	-0.13057852	0.11111111	-1.1752067
Upset	55	0.42730716	0.12644628	0.30086088	0.16489585	1.824551
Downstepped	51	0.0	0.11404958	-0.11404958	0.11111111	-1.0264462

Pattern 38 : <Same|
Number of occurrences : 158 Pattern probability : 0.015702479
Mutual information : 0.11425485 Entropy : 1.9447712
Normalized Mutual information : 0.05199962 Normalized Entropy : **0.88510346**
Contribution to Mutual information : 0.0017940843 Contribution to Entropy : 0.030537728
Contribution to Normalized Mutual information : 8.165229E-4 Contribution to Normalized Entropy : 0.013898319

EOW	1	0.25787205	0.25950414	-0.0016320944	0.034802705	-0.04689562
Top	39	0.18238994	0.09256198	0.08982796	0.030721683	2.923927
Mid	23	0.006289309	0.033884298	-0.02759499	0.0062893084	-4.3876033
Bottom	25	0.031446543	0.061157025	-0.029710483	0.013884163	-2.1398828
Higher	40	0.1509434	0.07933884	0.07160456	0.02848045	2.5141652
Lower	33	0.04402516	0.10247934	-0.05845418	0.01632094	-3.5815449
Same	41	0.08805032	0.13057852	-0.042528197	0.022543553	-1.8864905
Upset	42	0.12565102	0.12644628	-7.952601E-4	0.026369195	-0.030158678
Downstepped	43	0.113332294	0.11404958	-7.1728975E-4	0.025219051	-0.028442377

Pattern 39 : <Same|Top|
Number of occurrences : 29 Pattern probability : 0.023966942
Mutual information : 0.62759215 Entropy : 1.2518921
Normalized Mutual information : 0.28562948 Normalized Entropy : **0.5697606**
Contribution to Mutual information : 0.015041465 Contribution to Entropy : 0.030004025
Contribution to Normalized Mutual information : 0.0068456656 Contribution to Normalized Entropy : 0.013655419

EOW	1	0.48796377	0.25950414	0.22845963	0.09282076	2.4612987
Top	20	0.0	0.09256198	-0.09256198	0.03448276	-2.6842976
Mid	23	0.0	0.033884298	-0.033884298	0.03448276	-0.9826446
Bottom	25	0.13776341	0.061157025	0.076606385	0.06400018	1.1969714
Higher	28	0.0	0.07933884	-0.07933884	0.03448276	-2.3008263
Lower	21	0.2875501	0.10247934	0.18507075	0.08404946	2.2019267
Same	38	0.052281834	0.13057852	-0.078296684	0.041334815	-1.8942066
Upset	44	0.0	0.12644628	-0.12644628	0.03448276	-3.6669421
Downstepped	22	0.034440853	0.11404958	-0.07960873	0.033863153	-2.3508954

Pattern 40 : <Same|Higher|
Number of occurrences : 24 Pattern probability : 0.008264462
Mutual information : 0.65102446 Entropy : 1.1472995
Normalized Mutual information : 0.296294 Normalized Entropy : **0.5221585**
Contribution to Mutual information : 0.0053803674 Contribution to Entropy : 0.009481814
Contribution to Normalized Mutual information : 0.0024487106 Contribution to Normalized Entropy : 0.0043153595

EOW	1	0.56385106	0.25950414	0.30434692	0.10122645	3.006595
Top	23	0.0	0.033884298	-0.033884298	0.041666668	-0.8132231
Bottom	30	0.09168287	0.061157025	0.030525848	0.05890569	0.5182156
Higher	28	0.0	0.07933884	-0.07933884	0.041666668	-1.9041321
Lower	31	0.0	0.10247934	-0.10247934	0.041666668	-2.4595041
Same	38	0.0	0.13057852	-0.13057852	0.041666668	-3.1338844
Upset	44	0.0	0.12644628	-0.12644628	0.041666668	-3.0347106
Downstepped	32	0.18336575	0.11404958	0.06931616	0.078989126	0.8775406

Pattern 41 : <Same|Same|
Number of occurrences : 14 Pattern probability : 0.0033057851
Mutual information : 0.4568235 Entropy : 1.6345649
Normalized Mutual information : 0.20790933 Normalized Entropy : **0.74392253**
Contribution to Mutual information : 0.0015101603 Contribution to Entropy : 0.00540352
Contribution to Normalized Mutual information : 6.8730355E-4 Contribution to Normalized Entropy : 0.0024592478

EOW	1	0.2654687	0.25950414	0.0059645474	0.118017875	0.050539356
Top	39	0.18776296	0.09256198	0.09520098	0.1043716	0.9121349
Mid	23	0.0	0.033884298	-0.033884298	0.071428575	-0.47438017
Bottom	25	0.0	0.061157025	-0.061157025	0.071428575	-0.8561984
Higher	40	0.29413113	0.07933884	0.21479228	0.1217779	1.7638035
Lower	33	0.045322094	0.10247934	-0.057157245	0.05559291	-1.0281391
Same	41	0.09064419	0.13057852	-0.03993433	0.07673136	-0.5204434
Upset	42	0.0	0.12644628	-0.12644628	0.071428575	-1.7702479
Downstepped	43	0.116670944	0.11404958	0.0026213601	0.08579828	0.030552596

Pattern 42 : <Same|Upset|
Number of occurrences : 21 Pattern probability : 0.004132231
Mutual information : 0.4075957 Entropy : 1.7430003
Normalized Mutual information : 0.1855048 Normalized Entropy : **0.79327357**
Contribution to Mutual information : 0.0016842796 Contribution to Entropy : 0.00720248
Contribution to Normalized Mutual information : 7.665487E-4 Contribution to Normalized Entropy : 0.0032779898

EOW	1	0.24886814	0.25950414	-0.010636002	0.09434809	-0.1127315
Top	45	0.23191845	0.09256198	0.13935646	0.092100434	1.5130924
Mid	23	0.0	0.033884298	-0.033884298	0.04761905	-0.71157026
Bottom	47	0.11282519	0.061157025	0.051668167	0.06903956	0.74838495
Higher	48	0.14312948	0.07933884	0.06379064	0.07642096	0.834727
Lower	49	0.12536132	0.10247934	0.022881985	0.072258085	0.31667024
Same	38	0.0	0.13057852	-0.13057852	0.04761905	-2.7421489
Upset	44	0.0	0.12644628	-0.12644628	0.04761905	-2.655372
Downstepped	50	0.13789746	0.11404958	0.023847878	0.07523985	0.31695807

Pattern 43 : <Same|Downstepped|
Number of occurrences : 16 Pattern probability : 0.004132231
Mutual information : 0.94440126 Entropy : 0.55822
Normalized Mutual information : 0.42981553 Normalized Entropy **: 0.25405687**
Contribution to Mutual information : 0.0039024844 Contribution to Entropy : 0.0023066942
Contribution to Normalized Mutual information : 0.0017760972 Contribution to Normalized Entropy : 0.0010498217

EOW	1	0.8150501	0.25950414	0.555546	0.09706423	5.7234883
Top	52	0.0	0.09256198	-0.09256198	0.0625	-1.4809917
Mid	23	0.0	0.033884298	-0.033884298	0.0625	-0.54214877
Bottom	25	0.028453829	0.061157025	-0.0327032	0.04156637	-0.78677064
Higher	53	0.0	0.07933884	-0.07933884	0.0625	-1.2694215
Lower	33	0.0	0.10247934	-0.10247934	0.0625	-1.6396694
Same	54	0.0	0.13057852	-0.13057852	0.0625	-2.0892563
Upset	55	0.15649606	0.12644628	0.030049786	0.09083125	0.33083093
Downstepped	51	0.0	0.11404958	-0.11404958	0.0625	-1.8247933

Pattern 44 : <Upset|
Number of occurrences : 153 Pattern probability : 0.0107438015
Mutual information : 0.22980063 Entropy : 1.8886449
Normalized Mutual information : 0.10458677 Normalized Entropy : **0.85955936**
Contribution to Mutual information : 0.0024689324 Contribution to Entropy : 0.020291226
Contribution to Normalized Mutual information : 0.0011236595 Contribution to Normalized Entropy : 0.009234935

EOW	1	0.24839462	0.25950414	-0.011109516	0.03493175	-0.31803492
Top	45	0.23147719	0.09256198	0.13891521	0.034098618	4.073925
Mid	23	0.0	0.033884298	-0.033884298	0.006535948	-5.1842976
Bottom	47	0.11261052	0.061157025	0.051453494	0.025556471	2.0133255
Higher	48	0.0759423	0.07933884	-0.0033965409	0.021416362	-0.1585956
Lower	49	0.1251228	0.10247934	0.022643462	0.026748292	0.8465386
Same	38	0.02502456	0.13057852	-0.105553955	0.012627994	-8.358727
Upset	44	0.04379298	0.12644628	-0.0826533	0.016543696	-4.9960604
Downstepped	50	0.13763508	0.11404958	0.023585498	0.027852513	0.8467997

Pattern 45 : <Upset|Top|
Number of occurrences : 37 Pattern probability : 0.030578513
Mutual information : 0.65301824 Entropy : 1.1161333
Normalized Mutual information : 0.2972014 Normalized Entropy : **0.50797415**
Contribution to Mutual information : 0.019968327 Contribution to Entropy : 0.034129698
Contribution to Normalized Mutual information : 0.009087977 Contribution to Normalized Entropy : 0.015533094

EOW	1	0.51960796	0.25950414	0.26010382	0.082136266	3.1667354
Top	20	0.0	0.09256198	-0.09256198	0.027027028	-3.4247932
Mid	23	0.0	0.033884298	-0.033884298	0.027027028	-1.253719
Bottom	25	0.0	0.061157025	-0.061157025	0.027027028	-2.26281
Higher	28	0.0	0.07933884	-0.07933884	0.027027028	-2.935537
Lower	46	0.30619755	0.10247934	0.20371822	0.075773604	2.6885116
Same	38	0.055672284	0.13057852	-0.07490623	0.037694678	-1.9871832
Upset	44	0.0	0.12644628	-0.12644628	0.027027028	-4.678512
Downstepped	22	0.118522145	0.11404958	0.004472561	0.053137925	0.08416891

Pattern 46 : <Upset|Top|Lower|
Number of occurrences : 9 Pattern probability : 0.0074380166
Mutual information : 0.86962664 Entropy : 0.64216816
Normalized Mutual information : 0.39578414 Normalized Entropy **: 0.29226333**
Contribution to Mutual information : 0.0064682974 Contribution to Entropy : 0.0047764573
Contribution to Normalized Mutual information : 0.002943849 Contribution to Normalized Entropy : 0.0021738594

EOW	1	0.7984883	0.25950414	0.5389842	0.13370976	4.031001
Top	20	0.0	0.09256198	-0.09256198	0.11111111	-0.8330578
Mid	23	0.0	0.033884298	-0.033884298	0.11111111	-0.30495867
Bottom	25	0.0	0.061157025	-0.061157025	0.11111111	-0.55041325
Higher	34	0.0	0.07933884	-0.07933884	0.11111111	-0.7140496
Lower	33	0.10075582	0.10247934	-0.0017235205	0.10033504	-0.017177653
Same	35	0.10075582	0.13057852	-0.0298227	0.10033504	-0.29723117
Upset	36	0.0	0.12644628	-0.12644628	0.11111111	-1.1380165
Downstepped	37	0.0	0.11404958	-0.11404958	0.11111111	-1.0264462

Pattern 47 : <Upset|Bottom|
Number of occurrences : 18 Pattern probability : 0.014876033
Mutual information : 0.6653993 Entropy : 1.023503
Normalized Mutual information : 0.30283627 Normalized Entropy : **0.46581626**
Contribution to Mutual information : 0.009898502 Contribution to Entropy : 0.015225664
Contribution to Normalized Mutual information : 0.0045050024 Contribution to Normalized Entropy : 0.006929498

EOW	1	0.5147059	0.25950414	0.25520176	0.11780015	2.166396
Top	20	0.0	0.09256198	-0.09256198	0.055555556	-1.6661156
Mid	23	0.0	0.033884298	-0.033884298	0.055555556	-0.60991734
Bottom	25	0.0	0.061157025	-0.061157025	0.055555556	-1.1008265
Higher	28	0.0	0.07933884	-0.07933884	0.055555556	-1.4280992
Lower	33	0.0	0.10247934	-0.10247934	0.055555556	-1.8446281
Same	26	0.27941176	0.13057852	0.14883325	0.105762005	1.4072468
Upset	27	0.20588236	0.12644628	0.07943608	0.09530501	0.83349323
Downstepped	51	0.0	0.11404958	-0.11404958	0.055555556	-2.0528924

Pattern 48 : <Upset|Higher|
Number of occurrences : 9 Pattern probability : 0.0074380166
Mutual information : 1.0241047 Entropy : 1.3182484
Normalized Mutual information : 0.46609014 Normalized Entropy : **0.5999607**
Contribution to Mutual information : 0.007617308 Contribution to Entropy : 0.009805153
Contribution to Normalized Mutual information : 0.0034667861 Contribution to Normalized Entropy : 0.0044625173

EOW	1	0.0	0.25950414	-0.25950414	0.11111111	-2.3355372
Top	29	0.0	0.09256198	-0.09256198	0.11111111	-0.8330578
Mid	23	0.0	0.033884298	-0.033884298	0.11111111	-0.30495867
Bottom	30	0.25633913	0.061157025	0.19518211	0.14553708	1.3411161
Higher	28	0.0	0.07933884	-0.07933884	0.11111111	-0.7140496
Lower	31	0.23098251	0.10247934	0.12850317	0.14048709	0.9146974
Same	38	0.0	0.13057852	-0.13057852	0.11111111	-1.1752067
Upset	44	0.12816957	0.12644628	0.0017232895	0.111426175	0.015465751
Downstepped	32	0.38450873	0.11404958	0.27045915	0.16215965	1.6678572

Pattern 49 : <Upset|Lower|
Number of occurrences : 20 Pattern probability : 0.0049586776
Mutual information : 0.56244165 Entropy : 1.1380904
Normalized Mutual information : 0.25597823 Normalized Entropy : **0.5179672**
Contribution to Mutual information : 0.0027889668 Contribution to Entropy : 0.005643423
Contribution to Normalized Mutual information : 0.0012693134 Contribution to Normalized Entropy : 0.0025684326

EOW	1	0.558904	0.25950414	0.29939985	0.11102484	2.6966925
Top	20	0.0	0.09256198	-0.09256198	0.05	-1.8512397
Mid	23	0.0	0.033884298	-0.033884298	0.05	-0.677686
Bottom	25	0.0	0.061157025	-0.061157025	0.05	-1.2231405
Higher	34	0.0833107	0.07933884	0.00397186	0.061794024	0.064275794
Lower	33	0.0	0.10247934	-0.10247934	0.05	-2.0495868
Same	35	0.13711555	0.13057852	0.006537035	0.07691387	0.08499163
Upset	36	0.22066982	0.12644628	0.094223544	0.09272935	1.0161135
Downstepped	37	0.0	0.11404958	-0.11404958	0.05	-2.2809916

Pattern 50 : <Upset|Downstepped|
Number of occurrences : 22 Pattern probability : 0.0066115703
Mutual information : 0.27497798 Entropy : 1.5842024
Normalized Mutual information : 0.12514786 Normalized Entropy : **0.72100157**
Contribution to Mutual information : 0.0018180362 Contribution to Entropy : 0.010474065
Contribution to Normalized Mutual information : 8.2742394E-4 Contribution to Normalized Entropy : 0.0047669522

EOW	1	0.39529005	0.25950414	0.13578591	0.10423659	1.3026704
Top	52	0.15646896	0.09256198	0.063906975	0.077455685	0.8250779
Mid	23	0.0	0.033884298	-0.033884298	0.045454547	-0.74545455
Bottom	25	0.0	0.061157025	-0.061157025	0.045454547	-1.3454546
Higher	53	0.052787628	0.07933884	-0.026551213	0.04767365	-0.55693686
Lower	33	0.0	0.10247934	-0.10247934	0.045454547	-2.2545455
Same	54	0.1483971	0.13057852	0.017818585	0.0757914	0.23510034
Upset	55	0.18117459	0.12644628	0.054728314	0.08211698	0.6664677
Downstepped	51	0.06588167	0.11404958	-0.048167914	0.052889798	-0.9107222

Pattern 51 : <Downstepped|
Number of occurrences : 138 Pattern probability : 0.0132231405
Mutual information : 0.114423096 Entropy : 1.8446614
Normalized Mutual information : 0.052076194 Normalized Entropy : **0.83954155**
Contribution to Mutual information : 0.0015130327 Contribution to Entropy : 0.024392216
Contribution to Normalized Mutual information : 6.8861083E-4 Contribution to Normalized Entropy : 0.011101375

EOW	1	0.34037766	0.25950414	0.08087352	0.040335614	2.0050151
Top	52	0.13473281	0.09256198	0.04217083	0.029065128	1.4509081
Mid	23	0.007091201	0.033884298	-0.026793096	0.0071429075	-3.751007
Bottom	25	0.028364804	0.061157025	-0.03279222	0.014131946	-2.3204322
Higher	53	0.12055042	0.07933884	0.041211575	0.027717259	1.4868561
Lower	33	0.028364804	0.10247934	-0.07411453	0.014131946	-5.2444677
Same	54	0.12778227	0.13057852	-0.0027962476	0.028418962	-0.09839373
Upset	55	0.15600643	0.12644628	0.029560149	0.030888816	0.9569855
Downstepped	51	0.056729607	0.11404958	-0.057319976	0.019691711	-2.9108682

Pattern 52 : <Downstepped|Top|
Number of occurrences : 19 Pattern probability : 0.008264462
Mutual information : 0.74529815 Entropy : 1.0423839
Normalized Mutual information : 0.3391998 Normalized Entropy : **0.47440934**
Contribution to Mutual information : 0.0061594886 Contribution to Entropy : 0.008614743
Contribution to Normalized Mutual information : 0.002803304 Contribution to Normalized Entropy : 0.0039207386

EOW	1	0.5498304	0.25950414	0.29032624	0.11413679	2.5436692
Top	20	0.0	0.09256198	-0.09256198	0.05263158	-1.7586777
Mid	23	0.0	0.033884298	-0.033884298	0.05263158	-0.6438017
Bottom	25	0.06725198	0.061157025	0.006094955	0.05745898	0.10607489
Higher	28	0.0	0.07933884	-0.07933884	0.05263158	-1.507438
Lower	21	0.3240072	0.10247934	0.22152787	0.10736719	2.0632734
Same	38	0.058910403	0.13057852	-0.07166812	0.054017525	-1.3267568
Upset	44	0.0	0.12644628	-0.12644628	0.05263158	-2.4024792
Downstepped	22	0.0	0.11404958	-0.11404958	0.05263158	-2.1669421

Pattern 53 : <Downstepped|Higher|
Number of occurrences : 17 Pattern probability : 0.0049586776
Mutual information : 0.43560553 Entropy : 1.506628
Normalized Mutual information : 0.19825262 Normalized Entropy : **0.68569595**
Contribution to Mutual information : 0.0021600274 Contribution to Entropy : 0.0074708825
Contribution to Normalized Mutual information : 9.830708E-4 Contribution to Normalized Entropy : 0.003400145

EOW	1	0.34161818	0.25950414	0.08211404	0.11502307	0.7138919
Top	29	0.1845243	0.09256198	0.091962315	0.09408228	0.9774669
Mid	23	0.0	0.033884298	-0.033884298	0.05882353	-0.57603306
Bottom	30	0.0	0.061157025	-0.061157025	0.05882353	-1.0396694
Higher	28	0.0	0.07933884	-0.07933884	0.05882353	-1.3487602
Lower	31	0.2093789	0.10247934	0.10689956	0.09867933	1.0833025
Same	38	0.066119656	0.13057852	-0.06445886	0.060267955	-1.0695379
Upset	44	0.0	0.12644628	-0.12644628	0.05882353	-2.1495867
Downstepped	32	0.19835895	0.11404958	0.08430937	0.09671446	0.8717349

Pattern 54 : <Downstepped|Same|
Number of occurrences : 15 Pattern probability : 0.012396694
Mutual information : 0.64401966 Entropy : 1.0611299
Normalized Mutual information : 0.293106 Normalized Entropy : **0.48294103**
Contribution to Mutual information : 0.007983714 Contribution to Entropy : 0.013154503
Contribution to Normalized Mutual information : 0.003633545 Contribution to Normalized Entropy : 0.005986872

EOW	1	0.6561211	0.25950414	0.39661694	0.122644804	3.2338667
Top	39	0.0	0.09256198	-0.09256198	0.06666667	-1.3884298
Mid	23	0.0	0.033884298	-0.033884298	0.06666667	-0.5082645
Bottom	25	0.034387898	0.061157025	-0.026769128	0.04704988	-0.5689521
Higher	40	0.16506189	0.07933884	0.08572305	0.09585282	0.89431953
Lower	33	0.048143055	0.10247934	-0.054336283	0.055272236	-0.9830665
Same	41	0.09628611	0.13057852	-0.034292407	0.076164335	-0.45024234
Upset	42	0.0	0.12644628	-0.12644628	0.06666667	-1.8966942
Downstepped	43	0.0	0.11404958	-0.11404958	0.06666667	-1.7107438

Pattern 55 : <Downstepped|Upset|
Number of occurrences : 22 Pattern probability : 0.0107438015
Mutual information : 0.6499421 Entropy : 1.1901242
Normalized Mutual information : 0.2958014 Normalized Entropy : **0.54164886**
Contribution to Mutual information : 0.006982849 Contribution to Entropy : 0.012786457
Contribution to Normalized Mutual information : 0.0031780314 Contribution to Normalized Entropy : 0.0058193672

EOW	1	0.5052751	0.25950414	0.24577093	0.10659443	2.305664
Top	45	0.25731122	0.09256198	0.16474923	0.09320119	1.7676731
Mid	23	0.0	0.033884298	-0.033884298	0.045454547	-0.74545455
Bottom	47	0.0	0.061157025	-0.061157025	0.045454547	-1.3454546
Higher	48	0.08441785	0.07933884	0.0050790086	0.05927267	0.08568888
Lower	49	0.0	0.10247934	-0.10247934	0.045454547	-2.2545455
Same	38	0.0	0.13057852	-0.13057852	0.045454547	-2.8727274
Upset	44	0.0	0.12644628	-0.12644628	0.045454547	-2.7818182
Downstepped	50	0.15299587	0.11404958	0.038946286	0.07674874	0.5074518

Résumé.

L'objectif de notre thèse est de recueillir des données (à partir d'un corpus oral) et de développer une approche théorique et méthodologique adaptée à la description formelle de l'intonation du parler russe spontané. Cette description s'inscrit dans la perspective d'une application dans un système de synthèse vocale fondé sur une « double sélection d'unités » (modèle développé par van Santen *et al.* 2005).

Dans ce travail, nous défendons l'idée qu'une entreprise de formalisation d'un système prosodique se doit de conjuguer l'étude de la dimension tonale et celle du phrasé prosodique. Deux objectifs complémentaires découlent de cette prise de position, qui se rapportent respectivement :

a) à la nécessité de décrire et de modéliser le phrasé prosodique dans ses interactions avec d'autres niveaux de la description linguistique, avant d'entreprendre la modélisation acoustique de ce phrasé ;

b) à la détermination des formalismes qui permettront d'appréhender le lien entre les représentations formelle et fonctionnelle des phénomènes prosodiques.

Les études empiriques conduites dans cette thèse se fondent sur deux types de données : l'analyse d'un corpus de parole spontanée et les données issues d'une étude perceptive. Dans l'expérience de perception, une tâche d'annotation prosodique (en unités intonatives) a été proposée aux auditeurs experts qui ont été soumis à trois conditions expérimentales. Cette démarche a été adoptée afin d'obtenir des précisions sur l'articulation de l'interface « prosodie ~ syntaxe ».

Dans une étape ultérieure de notre travail, nous analysons :

i) l'accord entre les auditeurs dans la tâche d'annotation ;

ii) les propriétés rythmico-temporelles des unités intonatives perçues ;

iii) par référence au phrasé prosodique perçu, le rôle, la force prédictive et la valeur hiérarchique des contraintes AlignXP,R et Wrap-XP, issues de la théorie de l'optimalité. À partir des données recueillies, nous proposons une modélisation acoustique des frontières et nous procédons à l'évaluation statistique de la qualité de la démarche prédictive.

Nous nous appuyons sur le formalisme INTSINT (Hirst *et al.* 2000) dans l'annotation de l'organisation tonale. Cette annotation étant (semi)-automatique, nous évaluons sa performance dans le corpus de parole spontanée en langue russe. Dans un deuxième temps, en nous fondant sur l'annotation symbolique obtenue, nous proposons une méthodologie d'analyse objective de l'articulation « forme ~ fonction », en privilégiant dans cette démarche l'approche probabiliste.

Au cours de notre thèse, nous avons essayé de construire une réflexion épistémologique sur les méthodologies à mettre en place afin de s'assurer de l'objectivité des démarches et des conclusions. Au terme de ce travail, nous estimons avoir démontré l'intérêt qu'il y a à mener une étude de la formalisation de l'intonation et de l'organisation prosodique en général à partir des corpus oraux.

Mots clés : phrasé prosodique, syntagme intonatif, constituance prosodique, analyse formelle de la prosodie, interface « prosodie ~ syntaxe », articulation « forme ~ fonction », approche probabiliste.

Abstract

This study aims at collecting evidence from a study of a speech corpus and at developing theoretical and methodological approaches to enable a formal description of the intonation of Russian spontaneous speech. With reference to an application in a TTS system, the resulting description is expected to come within the context of a double unit selection model, as proposed recently by van Santen *et al.* (2005).

In our study we take the position that the task of formalisation of the prosodic system encompasses the work on both tonal and phrasing dimensions. Hence, two complementary issues should be addressed that relate to:

a) describing and modelling prosodic phrasing, with special attention paid to its interactions with other linguistic levels as well as to its acoustic correlates;

b) conceiving a set of formalisms intended to capture the "form-function" articulation for the prosodic phenomena.

The study is based on two types of data: a corpus of spontaneous Russian speech and the data obtained from a perceptual study. In the perceptual experiment, the task of prosodic annotation (in terms of intonational units) was proposed to trained judges under three experimental conditions. This protocol was applied in order to test and specify the modalities of a "syntax ~ prosody" interface. Subsequently we analyse

i) inter-judges agreement;

ii) the rythmico-temporal properties of the perceived units;

iii) the role, predictive strength and hierarchy of the Optimality Theory constraints AlignXP, R and WrapXP. On the basis of the perceptual data we propose an acoustic model of the prosodic boundaries and we evaluate statistically the quality of the resulting classifier.

The tonal dimension is modelled using the INTSINT formalism (Hirst *et al.* 2000). Given the semi-automatic character of this annotation, we evaluate the algorithm's performance on the corpus of Russian spontaneous speech. Next, we propose an objective method designed to capture the "form ~ function" association: the approach is based on the symbolic coding of the f0 curve and is probabilistic in nature.

In this study we discuss the epistemological issues in relation to the methodology adopted, our goal being to prove the objectivity of the approaches and conclusions. In conclusion, we seek to demonstrate the interest of the study of the formalisation of prosodic analyses grounded on speech corpora.