

UNIVERSITE AIX-MARSEILLE I - Université de Provence

N° attribué par la bibliothèque :

| | | | | | | | | |

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE AIX-MARSEILLE I

Formation doctorale : Langage et Parole

Présentée et soutenue publiquement

par

Caterina PETRONE

le 19 Décembre 2008

**Le rôle de la variabilité phonétique
dans la représentation des contours intonatifs
et de leur sens**

Directrice de Thèse :

Mme Mariapaola D'Imperio

Jury :

Mme Mariapaola D'Imperio, CNRS/Université de Provence (Directrice de Thèse)

M. Carlos Gussenhoven, University of Nijmegen (Rapporteur)

M. Daniel Hirst, CNRS/Université de Provence (Examineur)

M. Robert D. Ladd, University of Edinburgh (Examineur)

Mme Pilar Prieto, Universitat Autònoma de Barcelona (Rapporteur)

Remerciements

Ce projet n'aurait pas pu aboutir, si je n'avais pas rencontré un grand nombre de personnes sur mon parcours.

Je tiens d'abord à remercier ma directrice de thèse, Mariapaola D'Imperio, pour tout ce qu'elle a fait pour moi, à la fois sur le plan professionnel et humain. Je la remercie de m'avoir "initiée" à la prosodie et à la phonologie de laboratoire, et d'avoir été mon guide pendant ce chemin. Son enthousiasme, sa rigueur et sa grande ouverture d'esprit ont suscité en moi une profonde admiration. Merci pour tout ce que tu m'as enseigné, et d'avoir été toujours à l'écoute et prête à dialoguer. Et également merci pour ton amitié et ta générosité : ta maison et ta petite (!?) Camilla ont été ma seconde famille.

Je voudrais exprimer ma reconnaissance aux membres du jury de thèse : Carlos Gusenhoven, Daniel Hirst, Bob Ladd et Pilar Prieto. Je suis très fier du fait qu'ils ont tous accepté d'en faire partie, malgré la surcharge que représente pour eux la lecture de ce manuscrit (en français, en plus!). Un grand merci aussi pour les discussions très enrichissantes que j'ai eues avec eux pendant ces années, et qui m'ont aidé à voir les problématiques liées à la prosodie et à l'intonation selon plusieurs perspectives.

Merci à tous les membres du LPL. En particulier, je remercie Robert "le Parrain" Espesser, d'avoir été mon guru de stats, pour sa patience et le temps qu'il m'a dédié. Merci pour les discours parfois surréalistes, qui embrayaient sur les modèles mixtes et

concluait avec les recettes de cuisine à base de *garum*, les plats typiques de Calabre et même avec les “aventures statistiques” de Bouvard et Pécuchet. Merci aussi pour ton amitié sincère.

Je tiens aussi à remercier Daniel Hirst, pour sa disponibilité et pour toutes les discussions que j’ai pu avoir avec lui, surtout concernant la modélisation de l’intonation. Merci à Noel Nguyen, de m’avoir aidé à plusieurs reprises à travers ses commentaires et orientations, surtout concernant mes études en perception de l’intonation. Un grand merci aussi à Albert Di Cristo, Barbara Hemforth, Alain Marchal, Joel Pynte et Bernard Teston pour les nombreux échanges qui ont instillé en moi l’intérêt pour plusieurs aspects de la linguistique. Merci également à Martine Faraco de m’avoir aidé, tout au début, dans l’apprentissage du français scientifique et à Carine André de m’avoir aidé dans l’utilisation du logiciel PERCEVAL. Merci aussi à l’équipe administrative du LPL, notamment Simone Dufour et Armelle Bonpain pour leur gentillesse et leur efficacité (combien de fois elle m’ont sauvé!). Je remercie aussi mon amie Cristel Portes, pour sa chaleur humaine, qui m’a beaucoup aidé, surtout lors de mon débarquement au LPL.

En dehors du LPL, toute ma gratitude va à Federico Albano Leoni et à Franco Cutugno, respectivement directeur et co-directeur de mon mémoire de “Laurea” à l’Université “Federico II” de Naples. Ils m’ont accordé leur confiance et m’ont incité à poursuivre des études doctorales au LPL et je les en remercie infiniment. Merci à Federico Albano Leoni pour ses cours de linguistique (en particulier sur la phonétique et sur la linguistique historique) qui m’ont poussé à choisir mon sujet de mémoire de première année de maîtrise. Et merci aussi à Franco Cutugno pour son encouragement, et de m’avoir offert la possibilité de mener les expériences de perception décrites dans le chapitre 3 au sein du Département de Physique de l’Université de Naples. Un grand merci à Bob Ladd de m’avoir accueilli au Département de Linguistique de l’Université d’Edimbourg, entre janvier et avril 2005, et pour ces précieuses suggestions qui m’ont aidé dans la structuration des expériences de perception décrites dans le chapitre 3.

J’ai aussi tiré un bénéfice considérable des commentaires et des discussions avec Ajou Chen, Barbara Gili Fivela, Sue Hertz, Kiyoshi Honda, Klaus Kohler, Jean-Marie Marandin, Caroline Menezes, Lisa Selkirk et Marina Vigário.

Mes amis doctorants et postdoctorants ont été très précieux pour moi. Je remercie tout particulièrement ma “soeur” Céline De Looze, pour sa surprenante générosité et amitié. Je ne peux pas oublier sa famille qui m’a accueilli et aidé dans des moments de grandes difficultés. Merci à Leonardo Lancia, pour son humanité, ainsi que pour les discussions stimulantes autour de plusieurs aspects de mon travail. Merci à mes très chers amis Vincent Aubanel, Sil Cho et Emmanuel Makasso pour leur amitié insubstituable, leur ironie et les belles soirées passées ensemble. Je remercie Oliver Niebuhr d’avoir cru en moi et pour m’inspirer par sa créativité de jeune chercheur. Merci aussi pour le soutien et la convivialité (dans les pauses café ou... les pauses pastis) à Francesco Cangemi, Marion Coadou, Alice Foucart, Malie Guénot, Amandine Michelas, Anne Tortel, Cecilou Petitjean et Jean-Philippe Prost.

Je voudrais remercier mes chers camerades de ciné, de tango ou tout simplement de bavardage qui ont rendu mon séjour en France si agréable : Caterina, César, Doris, Géraldine, Erwan, Malek, Moussa, Natalie, Jean-François, Stefano et Tina.

Cette thèse n’aurait pu exister sans l’aide des sujets napolitains, qui ont participé aux expériences de production et perception. Je remercie en particulier Arianna, Diego, Marco et Michela, qui se disputent encore la palme de “napoletano verace”. Merci de leur présence. Merci à Oreste pour sa chère disponibilité et son aide technique (même de la dernière minute). Merci à Domenica et Mariangela. Je remercie ma très chère amie Cinzia, pour avoir été toujours présente pendant ces années, de m’avoir aidée et écoutée au cours d’interminables conversations téléphoniques.

Merci à ma famille. Merci à ma petite soeur Daniela, pour sa chaleur et pour les efforts qu’elle fait pour qu’on se voit régulièrement. Enfin, merci à ma mère et à mon père d’avoir toujours accepté et encouragé mes choix, même si je sais combien mon départ leur a coûté.

Table des matières

1	Introduction	1
1.1	Variabilité phonétique et sens	1
1.2	Forme et fonction de l'intonation dans la Théorie Métrique-Autosegmentale	14
1.3	Représentation phonologique	26
1.3.1	Accents mélodiques et association	26
1.3.2	Tons périphériques et constituance prosodique	34
1.4	Représentation phonétique	40
1.4.1	Cibles acoustiques	40
1.4.2	Cibles perçues	44
1.5	Cibles tonales et facteurs de variabilité	50
2	Variabilité segmentale et réalisation de l'accent mélodique	63
2.1	Problématique	63
2.2	Méthodes	68
2.2.1	Corpus	68
2.2.2	Enregistrement	69
2.2.3	Segmentation acoustique et détection de la f_0	69
2.2.4	Méthodes de détection des cibles tonales	72

2.2.5	Analyse statistique	77
2.2.6	Modélisation de la pente de f_0	79
2.3	Résultats	83
2.3.1	Durée segmentale	83
2.3.2	Détection des cibles tonales	87
2.3.3	Alignement tonal	94
2.3.4	Valeurs mélodiques des cibles tonales	107
2.3.5	Pente de f_0	110
2.4	Discussion	114
3	Effets de structure et d'alignement sur l'identification intonative et lexicale	120
3.1	Problématique	120
3.2	Corpus de base	123
3.3	Tâche préliminaire	124
3.3.1	Préparation des stimuli	124
3.3.2	Déroulement de l'expérience	127
3.3.3	Résultats	128
3.3.4	Discussion	130
3.4	Expérience I	130
3.4.1	Préparation des stimuli	130
3.4.2	Déroulement de l'expérience	132
3.4.3	Résultats	133
3.4.4	Discussion	138
3.5	Expérience II	140
3.5.1	Préparation des stimuli et enregistrement	140
3.5.2	Résultats	141
3.5.3	Discussion	146
4	Variabilité de forme et de pente de f_0 : preuve de l'existence du Syntagme Accentuel en napolitain	151

4.1	Problématique	151
4.2	Expérience I	158
4.2.1	Corpus et enregistrement	161
4.2.2	Résultat	162
4.2.3	Discussion	189
4.3	Expérience II	194
4.3.1	Méthodes	198
4.3.2	Résultats	203
4.3.3	Discussion	224
5	Effets du ton de Syntagme Accentuel sur la perception du contraste intonatif	228
5.1	Problématique	228
5.2	Expérience I	231
5.2.1	Préparation des stimuli	232
5.2.2	Déroulement de l'expérience	234
5.2.3	Analyse statistique	234
5.2.4	Résultats	235
5.2.5	Discussion	237
5.3	Expérience II	239
5.3.1	Préparation des stimuli	242
5.3.2	Échelles sémantiques	244
5.3.3	Déroulement de l'expérience	245
5.3.4	Analyse statistique	246
5.3.5	Résultats	248
5.3.6	Discussion	252
6	Conclusion	255
6.1	Résumé des découvertes	256
6.2	Apports méthodologiques	260
6.3	Variabilité, représentation phonétique et phonologique	264

6.4	Conséquences sur le sens intonatif	269
	References	271
A	Annexes	288

Table des figures

- 1.1 Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *La mamma vuole vedere la nana* (“la maman veut voir la naine”), prononcée comme une affirmation (en haut) et comme une question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de *nana*. Locuteur OM. 4
- 1.2 Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *La mamma vuole vedere la rana* (“la maman veut voir la grenouille”), prononcée comme une affirmation (en haut) et comme une question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de *rana*. Locuteur OM. 5
- 1.3 Spectrogramme, courbe de f_0 et cibles tonales pour l’affirmation *La mamma vuole vedere la Dana* (“la maman veut voir (la) Dana”). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur du prénom *Dana*. Les flèches montrent deux localisations possibles de la cible L. Locuteur AS. 7

1.4	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>È il nonno</i> (“c’est le grand-père”), lue comme affirmation à focalisation étroite (en haut) et question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de <i>nonno</i> . Locuteur CC.	10
1.5	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>È il nono</i> (“c’est le neuvième”), lue comme affirmation à focalisation étroite (en haut) et question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de <i>nono</i> . Locuteur CC.	11
1.6	Analyse du contour intonatif de <i>It’s painted orange!</i> (“c’est peint en orange!”), selon l’approche holistique (en haut); et selon l’approche compositionnelle par mouvements (dans le milieu) et par niveau (en bas). L’encadrement en rouge du contour indique la taille de l’unité d’analyse pour l’approche holistique et par mouvements, et les points rouges figurent la représentation en tons statiques de l’approche par niveaux.	17
1.7	Types différents de transition entre deux tons L et H.	17
1.8	Grammaire de l’intonation de l’anglais selon la Théorie AM. Tiré de Pierrehumbert (1980), p. 29.	20
1.9	Exemples d’accents mélodiques et de leur contribution au sens en anglais. Les phrases sont tirées de Pierrehumbert & Hirschberg (1990).	23
1.10	Représentation schématisée de la structure interne de l’accent mélodique. PA = accent mélodique; T_w = ton faible; T_s = ton fort.	27
1.11	Représentation schématisée de l’association entre la syllabe (σ) et le ton (T).	28
1.12	Représentation schématisée de l’alignement des tons L et H avec la TBU en italien napolitain, dans les questions (à gauche) et dans les affirmations (à droite). Tiré de D’Imperio (2000), p. 25.	31
1.13	Association primaire (ligne continue) et secondaire (ligne tiretée) pour les accents mélodiques $L+H^*$ et L^*+H des affirmations et des questions du napolitain. Adapté de Prieto <i>et al.</i> (2005), p. 25.	32

1.14	Forme d'onde, courbe de f_0 et étiquetage tonal pour les phrases <i>Vedrai MAMMA domani ?</i> ("Verras-tu maman demain ?", panneau en haut) et <i>Vedrai LA BELLA MANO DI MAMMA domani ?</i> ("Verras-tu la jolie main de maman demain ?", en bas). Les lettres majuscules indiquent la taille du constituant focalisé. Dans les deux cas, l'accent de syntagme est suivi par un accent postnucléaire <i>downstepped</i> , étiqueté comme !H*. Tiré de D'Imperio (2002a), p. 48.	36
1.15	Spectrogramme, courbe de f_0 et étiquetage tonal pour la phrase <i>La mamma vuole vedere la nana</i> ("la maman veut voir la naine"), lue comme une affirmation à focalisation étroite (en haut) et une question oui/non (en bas). La position du focus est avancée. Locuteur DD.	39
1.16	Exemple du contour H*H-H% pour la phrase "Can I go now ?" ("Peux-je partir maintenant ?"). Tiré de Pierrehumbert (1980), p. 308.	42
1.17	Résultats de la manipulation d'alignement sur l'identification des questions en napolitain. Le score moyen de réponse (axe des abscisses) est fonction des pas d'alignement temporel (axe des ordonnées). $T1, T2... T7$ = pas d'alignement 1, 2 ...7. Tiré de D'Imperio (2000), p. 165.	48
2.1	Spectrogramme, courbe de f_0 et TextGrid pour la phrase <i>La mamma vuole vedere la rana ?</i> (litt. "la maman veut voir la grenouille ?"), prononcée par le locuteur DD. Dans le TextGrid, <i>la-X</i> dans le premier tier indique l'intervalle de durée correspondant au mot prosodique; le deuxième et le troisième tiers représentent la segmentation en syllabes (<i>s</i>) et segments (<i>v</i> = voyelle; <i>c</i> = consonne).	72
2.2	Courbe de f_0 et lignes de régression estimées par le LSF. Le coude de f_0 est au point d'intersection des deux lignes (<i>elbow</i>). Tiré de D'Imperio (2000), p. 95.	75
2.3	Cible locale calculée pour l'instant X. La cible (<i>target</i>) est définie par sa localisation temporelle (<i>t</i>) et par sa valeur mélodique (<i>h</i>). Tiré de Campione <i>et al.</i> (1997), p. 23.	76

2.4	Courbe de f_0 , courbe quadratique par morceaux estimée (ligne continue) et cibles (cercles) trouvées par MOMEL. Tiré de Campione <i>et al.</i> (1997), p. 24.	76
2.5	Schématisation des 20 points de f_0 extraits entre L1 et H (pour la montée) et entre H et L2 (pour la descente).	81
2.6	Exemple de courbe logistique	82
2.7	Boxplots de la durée de la consonne (s) pour DD (en haut) et AS (en bas). Les résultats sont illustrés séparément pour les affirmations à focalisation étroite (A) et pour les questions oui/non (Q), et selon la classe consonantique.	85
2.8	Boxplots de la durée de la voyelle (s) pour DD (en haut) et AS (en bas). Les résultats sont illustrés séparément pour les affirmations à focalisation étroite (A) et pour les questions oui/non (Q), et selon la classe consonantique.	86
2.9	Exemple de détection des trois cibles par les méthodes manuelle, LSF et MOMEL.	88
2.10	Exemple de détection des trois cibles par les méthodes manuelle, LSF et MOMEL.	89
2.11	Exemple de détection des trois cibles par les méthodes manuelle, LSF et MOMEL. Le deux étiquettes “lsf” et “man” dans le premier <i>tier</i> du TextGrid sont superposées.	90
2.12	Boxplots de la latence de L1 par rapport au début de la consonne (s) comme trouvé par le LSF, la méthode manuelle et MOMEL. Les résultats sont regroupés pour les deux locuteurs. La ligne tiretée marque le début de la consonne.	93
2.13	Valeurs de f_0 et d’alignement tonal de L1, H et L2 pour DD (en haut) et AS (en bas). Les résultats sont présentés séparément en fonction de la modalité intonative et du type de consonne.	99
2.14	Boxplots des latences de L1 par rapport au début de la consonne ($c\theta$) pour DD (en haut) et AS (en bas). Les résultats sont présentés séparément en fonction de la modalité intonative et du type de consonne. La ligne tiretée marque le début de la consonne.	100

2.15	Boxplots des latences de L2 par rapport au début de la consonne ($c\theta$) pour le locuteur DD (en haut). Les résultats sont présentés séparément en fonction de la modalité intonative et du type de consonne. La ligne tiretée marque le début de la consonne.	105
2.16	Valeurs observées de la montée de f_0 (points) et valeurs attendues (lignes) résultantes de la régression logistique pour le locuteur DD.	111
2.17	Valeurs observées de la descente de f_0 (points) et valeurs attendues (lignes) de f_0 résultantes de la régression logistique pour DD (en haut) et AS (en bas).	113
3.1	Spectrogramme, courbe de f_0 et cibles tonales pour les formes de citation <i>nonno</i> et <i>nono</i> , prononcée avec une focalisation large.	125
3.2	Score moyen de <i>nonno</i> pour la série des stimuli créée à partir du mot <i>nonno</i> et <i>nono</i> pour tous les sujets.	129
3.3	Manipulation de l’alignement.	132
3.4	Score moyen des questions pour l’identification intonative pour les stimuli à D8 et à D2 créés à partir d’une question (en haut) et d’une affirmation (en bas).	136
3.5	Score moyen de <i>nonno</i> pour l’identification intonative pour les stimuli à D2 et D8 créés respectivement à partir de <i>nonno</i> (en haut) et <i>nono</i> (en bas). Les résultats sont montrés séparément selon la modalité intonative de la base (A/Q).	137
3.6	Score moyen des réponses “questions” pour l’identification intonative pour les stimuli <i>nonno/D5</i> et <i>nono/D6</i> créés à partir d’une question (en haut) et d’une affirmation (en bas).	142
3.7	Score moyen de <i>nonno</i> pour l’identification intonative pour les stimuli à D5 et D6 créés respectivement à partir de <i>nonno</i> (en haut) et <i>nono</i> (en bas). Les résultats sont montrés séparément selon la modalité intonative de la base (A/Q).	144

4.1	Représentation schématisée de la règle d'interpolation linéaire de Pierrehumbert (1980)	152
4.2	Représentation schématisée de la <i>tone linking rule</i> de Gussenhoven (1984). Le panneau supérieur illustre l'application partielle de la règle (la flèche courbe indique le déplacement du ton L de l'accent H*+L vers la droite). Le panneau inférieur montre le cas d'absence de liaison entre les deux tons bas.	153
4.3	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>La mamma vuole vedere la Dina</i> prononcée comme une question (en haut) ou une affirmation (en bas). Les lignes continues indiquent les frontières des mots. La ligne rouge tiretée schématise l'interpolation linéaire prédite par Pierrehumbert (1980) entre les deux accents montants. La flèche indique le mouvement descendant après le pic prénucléaire, qui est très douce et convexe dans les questions. Locuteur OM.	155
4.4	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>La mamma vuole vedere la nana</i> prononcée comme une question (en haut) ou une affirmation (en bas). Les lignes continues indiquent les frontières des mots. La flèche indique le mouvement descendant après le pic prénucléaire, qui est plus douce et convexe dans les questions. Locuteur AS.	156
4.5	Représentation schématisée de l'hypothèse d'insertion d'un ton différemment spécifié dans les questions (H, en haut) et dans les affirmations (L, en bas).	160
4.6	Représentation schématisée de la modélisation de la courbe de f_0 dans les questions (régression logistique, à gauche) et dans les affirmations (régression linéaire par morceaux, en bas).	161
4.7	Boxplots des latences de la cible prénucléaire L relativement au début de la consonne ($c\theta$) pour OM (en haut) et AS (en bas), dans les questions (Q) et dans les affirmations (A).	165

4.8	Boxplots des latences de la cible prénucléaire H relativement au début de la consonne (<i>cθ</i>) pour OM (en haut) et AS (en bas), dans les questions (Q) et dans les affirmations (A).	166
4.9	Boxplots des valeurs mélodiques de la cible prénucléaire L pour OM (en haut) et AS (en bas), dans les questions (Q) et dans les affirmations (A).	169
4.10	Boxplots des valeurs mélodiques de la cible prénucléaire H pour OM (en haut) et AS (en bas), dans les questions (Q) et dans les affirmations (A).	170
4.11	Boxplots des latences des cibles nucléaires L et H relativement au début de la consonne (<i>cθ</i>) pour OM, dans les questions (Q) et dans les affirmations (A).	173
4.12	Boxplots des valeurs mélodiques des cibles nucléaires L et H pour OM, dans les questions (Q) et dans les affirmations (A).	175
4.13	Exemples de détection manuelle des cibles H et L dans la région suivant le H prénucléaire dans la phrase <i>La mamma vuole vedere la nana</i> (“La maman veut voir la naine”), prononcée comme une question (en haut) et comme une affirmation (en bas). Les flèches représentent de possibles localisations pour les cibles tonales. Locuteur AS.	178
4.14	Exemples de la détection, effectuée par MOMEL, de la cible L dans la région suivant le H prénucléaire dans la phrase <i>La mamma vuole vedere la nana</i> (“La maman veut voir la naine”), répétée deux fois comme affirmation par le locuteur AS. La ligne grise indique la courbe de f_0 . Dans chaque panneau, les points verts représentent les cibles détectées par l’algorithme pour chaque phrase, liées par des interpolations linéaires par le script MOMEL pour PRAAT.	179

4.15	Exemples de la détection, effectuée par le LSF, de la cible H dans la région suivant le H prénucléaire dans les phrases <i>La mamma vuole vedere la nana</i> (“La maman veut voir la naine”) et <i>La mamma vuole vedere la rana</i> (“La maman veut voir la grenouille”), prononcées comme questions par le locuteur OM. Dans chaque panneau, le point vert indique le point d’intersection des deux lignes droites ajustées à la région entre le pic prénucléaire et le début du dernier mot prosodique	180
4.16	Valeurs de f_0 à 20 points de mesure pour OM (en haut) et AS (en bas). Dans chaque graphique, les données sont illustrées séparément pour les questions (Q , à gauche) et les affirmations (A , à droite). Le lissage <i>lowess</i> est indiqué par la ligne continue.	183
4.17	Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression logistique, pour les questions (en haut); et par la régression linéaire par morceaux, pour les affirmations (en bas). Locuteur OM.	187
4.18	Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression logistique, pour les questions (en haut); et par la régression linéaire par morceaux, pour les affirmations (en bas). Locuteur AS.	188
4.19	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>Il benevolo manovale mangiava un panino</i> prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, et la ligne épaisse la fin du constituant en position de Sujet. La ligne rouge tiretée schématise l’interpolation linéaire prédite par Pierrehumbert (1980) entre les deux accents. Locuteur FC.	206
4.20	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>L’agronomo di Benevolo mangiava un panino</i> prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, la ligne épaisse la fin du constituant en position de Sujet. La ligne rouge tiretée schématise l’interpolation linéaire prédite par Pierrehumbert (1980) entre les deux accents. Locuteur FC.	207

4.21	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>L'agronomo di Benevolo, mangiando un panino, si strozzò</i> prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, la ligne épaisse la fin du constituant en position de Sujet. La flèche indique la pente interaccentuelle. Locuteur FC.	208
4.22	Spectrogramme, courbe de f_0 et cibles tonales pour la phrase <i>L'agronomo di Benevolo mangiava un panino</i> prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, la ligne épaisse la fin du constituant en position de Sujet. La flèche indique la pente interaccentuelle. Locuteur DD.	209
4.23	Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression linéaire par morceaux dans les affirmations (A , à gauche) et dans les questions (Q , à droite). Locuteur FC.	215
4.24	Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression linéaire par morceaux dans les affirmations (A , à gauche) et dans les questions (Q , à droite). Locuteur LL.	216
4.25	Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression linéaire par morceaux dans les affirmations (A , à gauche) et dans les questions (Q , à droite). Locuteur MD.	217
4.26	Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression régression linéaire dans les affirmations (A , à gauche) et dans les questions (Q , à droite). Locuteur DD.	218
4.27	Durée moyenne des syllabes de <i>Benevolo</i> . Les résultats, regroupés pour les cinq locuteurs, sont présentés en fonction de la modalité intonative et du niveau de constituance prosodique.	220
4.28	Durée moyenne de la syllabe tonique. Les résultats, séparés par locuteur, sont fonction de la modalité intonative et du niveau de constituance prosodique.	221
4.29	Durée moyenne de la syllabe finale. Les résultats, séparés par locuteur, sont fonction de la modalité intonative et du niveau de constituance prosodique.	222

5.1	Représentation schématisée des trois conditions expérimentales Pren , AP et Nucl	233
5.2	Histogramme de réponses “questions” pour les stimuli dont la base intonative était une question (bleue) ou une affirmation (rouge), pour les conditions Pren , AP et Nucl . Les résultats sont regroupés pour tous les locuteurs	236
5.3	Représentation schématisée du continuum dans le domaine mélodique créé par l’augmentation progressive de la valeur mélodique du ton L_{AP} . L’association des pas du continuum avec les tons de AP est indiquée à droite . . .	243
5.4	Représentation schématisée des trois conditions expérimentales Pren et AP .	244
5.5	Représentation schématisée des deux phrases de contrôle. La ligne continue indique le stimulus représentatif des affirmations, la ligne tiretée celui représentatif des questions.	244
5.6	Score moyen des jugements des auditeurs pour les cinq échelles sémantiques dans la condition Pren . La ligne tiretée signale le pas “zéro” de chaque échelle, correspondante à l’interprétation des stimuli comme “neutres”. Les résultats sont regroupés pour les neuf auditeurs.	249
5.7	Score moyen des jugements des auditeurs pour les cinq échelles sémantiques dans la condition AP . La ligne tiretée signale le pas “zéro” de chaque échelle, correspondante à l’interprétation des stimuli comme “neutres”. Les résultats sont regroupés pour les neuf auditeurs.	250
5.8	Score moyen des jugements des auditeurs pour les cinq échelles sémantiques dans la condition Nucl . La ligne tiretée signale le pas “zéro” de chaque échelle, correspondante à l’interprétation des stimuli comme “neutres”. Les résultats sont regroupés pour les neuf auditeurs.	252
A.1	Quantile-Quantile plot des résidus de la régression logistique effectuée sur les questions (en haut) et de la régression linéaire par morceaux effectuée sur les affirmations (en bas). Locuteur OM.	289
A.2	Quantile-Quantile plot des résidus de la régression logistique effectuée sur les questions (en haut) et de la régression linéaire par morceaux effectuée sur les affirmations (en bas). Locuteur AS.	290

Liste des tableaux

2.1	Phrases du corpus. Les crochets indiquent le domaine de la focalisation. . .	68
2.2	Schématisation des modèles mixtes appliqués aux variables dépendantes . .	79
3.1	Valeurs de p et valeurs de durée de la voyelle accentuée (V1d) et de la coda consonantique (C1d) pour les pas de manipulation D1-D9 (ms). Les colonnes $\Delta\mathbf{V1d}$ et $\Delta\mathbf{Cd}$ indiquent la différence de durée entre chaque pas de manipulation et le successif, respectivement pour la voyelle et pour la consonne.	127
3.2	Valeurs d'alignement de L1, H et L2 par rapport à la fin de la voyelle. . .	132
4.1	Phrases du corpus. Les crochets indiquent les frontières des constituants syntaxiques et prosodiques. <i>N</i> = Nom; <i>V</i> = Verbe; <i>Adj</i> = Adjectif; <i>SN</i> = syntagme nominal; <i>SV</i> = syntagme verbal; <i>P</i> = Phrase; <i>IP</i> = <i>Intonational Phrase</i> (Syntagme intonatif); <i>ip</i> = <i>Intermediate Phrase</i> (Syntagme Intermédiaire); <i>AP</i> = <i>Accentual Phrase</i> (Syntagme Accentuel).	200
4.2	Proportion de ruptures prosodiques dans les questions (en haut) et dans les affirmations (en bas), séparément pour les trois conditions f-IP , f-ip et f-AP	204

4.3	Distribution de différents types d'accents prénucléaires, nucléaires, de tons périphériques à la fin de S et de pauses. Les résultats sont montrés en termes de pourcentage relativement à l'ensemble des phrases du corpus.	205
A.1	Valeurs de AIC des modèles d'alignement temporelle pour L1, H et L2. L'alignement est calculé en fonction de la distance absolue aux unités segmentales (c0 = début de la consonne; v0/v1 = début/fin de la voyelle) . .	288
A.2	Valeurs de AIC des modèles d'alignement temporelle pour L1, H et L2. L'alignement est calculé en fonction de la distance relative aux unités segmentales (syll % = pourcentage d'alignement dans la syllabe; voy % = pourcentage d'alignement dans la voyelle	288
A.3	Valeurs de AIC des modèles d'alignement temporelle pour L1, H et L2. L'alignement est calculé en fonction de la durée des segments (cons + voy = durée de la consonne et de la voyelle; rime = durée de la rime (/voyelle)	291

1.1 Variabilité phonétique et sens

Dans les actes de parole, les variations de fréquence fondamentale (désormais f_0) sont exploitées pour la transmission de différents types d'informations, de type linguistique (par exemple, sur la modalité intonative d'une phrase) ou paralinguistique (sur l'état émotionnel du locuteur). Ces variations, correspondant d'un point de vue physique à des changements dans la vibration des cordes vocales, peuvent être effectuées sur plusieurs dimensions phonétiques. Prenons l'exemple d'une montée de f_0 . La fin de la montée peut être effectuée comme un pic mélodique correspondant à un maximum de f_0 clairement identifiable dans la trace acoustique ou encore comme une configuration à plateaux. Elle peut avoir des valeurs mélodiques différentes, qui peuvent être influencées par l'étendue tonale utilisée par le locuteur. Dans d'autres cas, les propriétés dynamiques de la montée peuvent subir des variations importantes. Par exemple, la pente de la montée de f_0 peut être plus ou moins raide, selon la vitesse de changement du mouvement.

Puisque la f_0 est l'indice le plus robuste pour la transmission du sens intonatif, plusieurs études se sont penchées sur le rôle d'une ou plusieurs parmi ces dimensions pour la

définition des unités primitives de l'intonation. En particulier, dans la Théorie Métrique Autosegmentale de l'intonation (*Autosegmental-Metrical Theory*, Pierrehumbert, 1980), le contour intonatif est représenté comme une séquence de niveaux hauts (*High*, H) ou bas (*Low*, L). Leur réalisation phonétique en tant que "cibles tonales" (*tonal targets*) est définie uniquement par deux coordonnées : la hauteur mélodique (*scaling*, c'est-à-dire la valeur mélodique dans l'empan de f_0 utilisé par un locuteur) et l'alignement tonal (*tonal alignment*, c'est-à-dire la synchronisation temporelle des cibles tonales avec le texte). Les variations sur d'autres dimensions ne sont pas importantes dans la représentation de l'intonation, de sorte que, au niveau formel, tout mouvement entre ces cibles est considéré comme une transition sans aucune importance linguistique.

Les variations de valeur mélodique ou d'alignement tonal peuvent véhiculer des différences de sens intonatif, et la façon dont ces deux dimensions sont exploitées à ce but tend à être différente selon la langue considérée. Prenons par exemple le contraste entre questions oui/non et affirmations. En français, la différence de sens est signalée principalement par la présence/absence d'une montée de f_0 à la fin de l'énoncé. Cette différence peut être décrite comme un glissando qui va d'un point bas à un point haut dans les questions (LH) et d'un point haut à un point bas dans les affirmations (HL). De plus, d'autres indices d'ordre syntaxique peuvent aider à renforcer le contraste intonatif, et en particulier l'ordre des mots. En fait, la position relative du sujet et du verbe peut être inversée dans les questions. Prenons maintenant le cas de l'italien. L'italien est une langue à accent dynamique (la proéminance métrique étant marquée par plusieurs indices, tels que la f_0 , la durée, l'intensité et certaines caractéristiques spectrales, voir Beckman (1986) et, pour l'italien, Bertinetto (1980), *inter alia*) et les syllabes toniques peuvent recevoir un accent mélodique. En ce qui concerne la différence entre questions et affirmations, elle ne peut pas être signalée par des indices morpho-syntaxiques et le seul moyen de distinguer les deux modalités est la différence d'intonation. De plus, dans la variété d'italien parlée à Naples, un indice robuste de ce contraste est l'alignement des cibles tonales de l'accent nucléaire LH et de l'accent de syntagme L- suivant (D'Imperio, 2000).

En italien, l'accent nucléaire est défini comme l'accent le plus à droite dans le constituant focalisé (*the rightmost fully fledged pitch accent in the focused constituent* Grice

et al., 2005, p.380, trad. de D’Imperio, 2005). L’accent de syntagme est le ton périphérique qui marque le fin du syntagme intermédiaire et il est souvent réalisé dans la région suivant immédiatement l’accent nucléaire (D’Imperio, 2000 ; voir aussi les paragraphes suivants pour plus de détails sur le statut phonologique des accents mélodiques et des tons périphériques).

Dans la figure 1.1, la courbe de f_0 et le spectrogramme sont tracés pour la phrase *La mamma vuole vedere la nana* (“la maman veut voir la naine”), lue par un locuteur napolitain. La phrase a été prononcée comme une affirmation à focalisation étroite et comme une question oui/non. Dans les deux cas, l’accent nucléaire LH est réalisé sur le mot paroxyton *nana* et il est précédé par un autre accent LH (défini “prénucléaire”) sur le mot paroxyton *mamma*. Comme on peut le remarquer, dans *nana* la f_0 commence à monter dans la région en correspondance avec le début de la syllabe tonique *na-*, la fin de la montée étant signalée par un pic de f_0 dans la voyelle. Puis, la f_0 descend dans les deux modalités jusqu’à atteindre une valeur minimale après la fin de la voyelle. En termes autosegmentaux, on peut décrire cette montée-descente de f_0 comme un mouvement allant d’un point bas L (le minimum de f_0 au début de la montée) à un point H (le pic maximal de f_0) et à un point bas L (le minimum de f_0 à la fin de la descente). Dans les deux modalités, le premier point L et le point H constituent l’accent nucléaire LH, tandis que le deuxième point bas est l’accent de syntagme L- . Contrairement au français, il n’y a pas de montée finale de f_0 dans les questions, de sorte que les valeurs de f_0 restent basses jusqu’à la fin de l’énoncé dans l’affirmation et encore dans la question. De plus, si l’on observe de plus près la courbe de f_0 , on peut constater que les trois points ne sont pas localisés au même endroit temporel, car l’alignement des cibles L, H et L relativement à la syllabe *na-* est légèrement plus retardé dans la question que dans l’affirmation. Dans l’affirmation, les trois cibles sont réalisées respectivement avant le début de la syllabe, dans la première portion de la voyelle et à la fin de la voyelle. Dans la question, elles sont réalisées respectivement à l’attaque syllabique, au centre de la voyelle tonique et dans la syllabe post-tonique. La différence d’alignement est subtile et en fait dans l’exemple ci-dessous l’emplacement du pic de f_0 dans les questions diffère seulement de 63 ms par rapport à celui des affirmations.

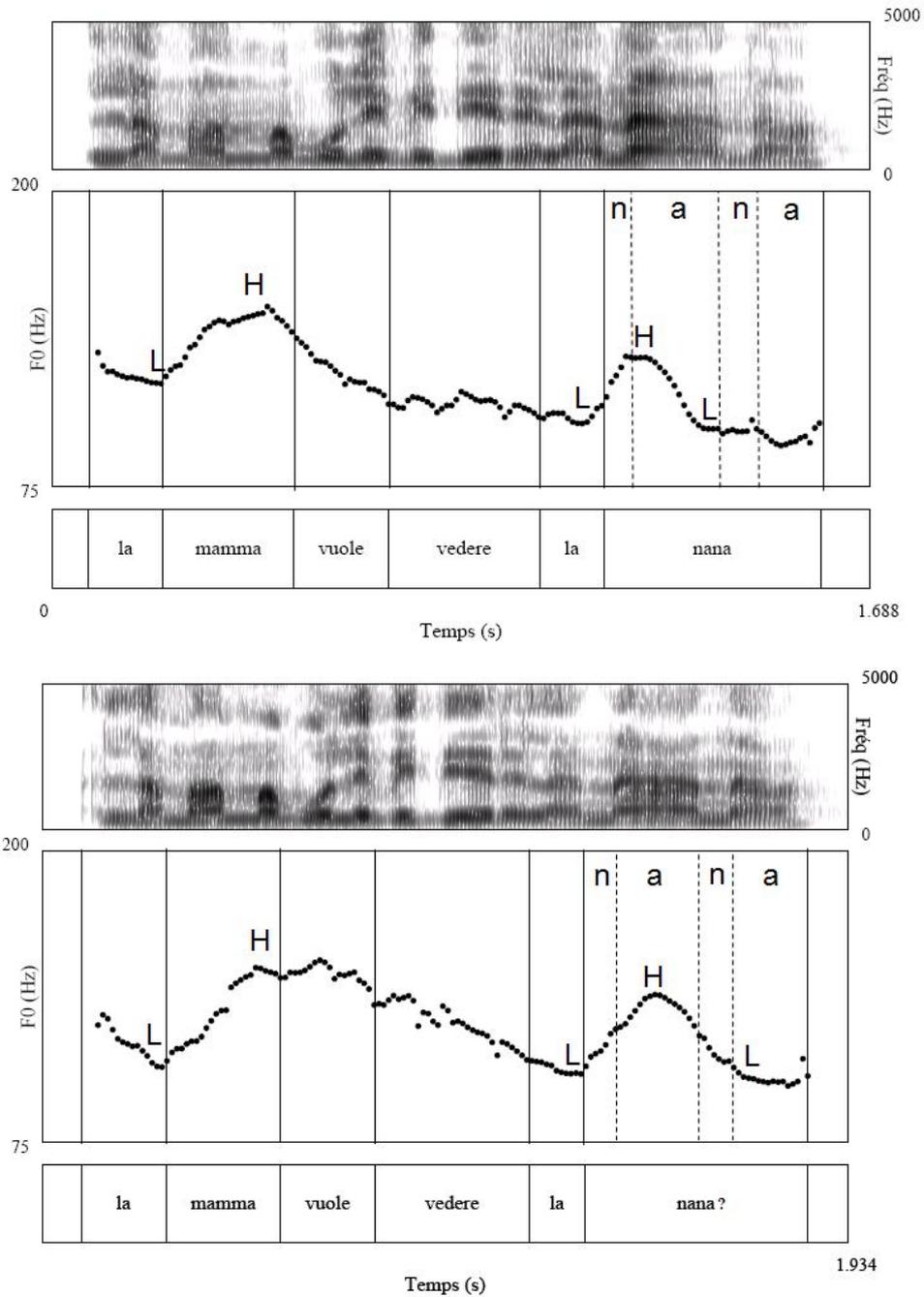


FIGURE 1.1 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *La mamma vuole vedere la nana* (“la maman veut voir la naine”), prononcée comme une affirmation (en haut) et comme une question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de *nana*. Locuteur OM.

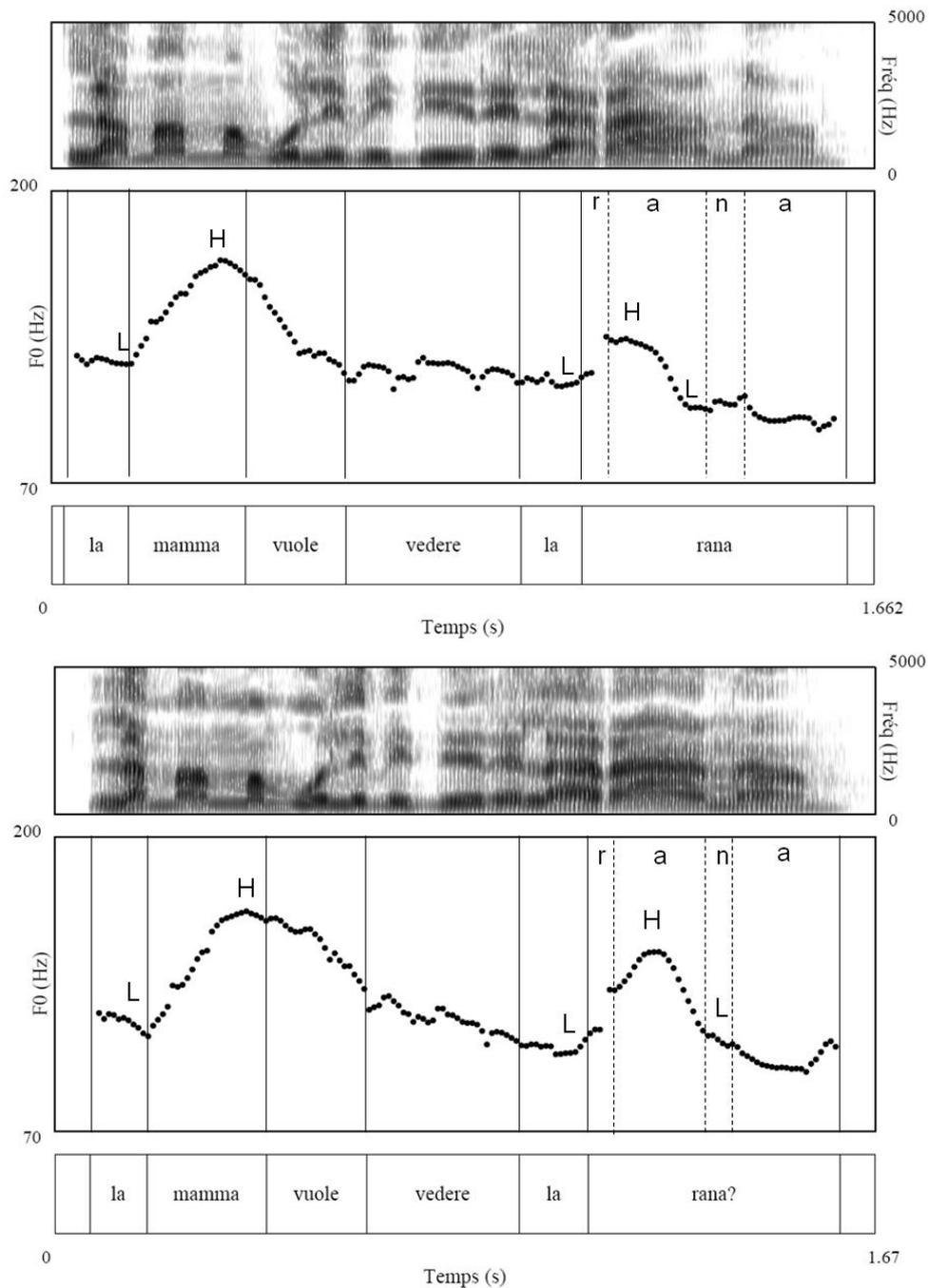


FIGURE 1.2 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *La mamma vuole vedere la rana* (“la maman veut voir la grenouille”), prononcée comme une affirmation (en haut) et comme une question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de *rana*. Locuteur OM.

Jusqu’ici, l’analyse de la courbe de f_0 n’a pas posé de problèmes, et on pourrait même

supposer, à partir des exemples donnés, qu'il semble y avoir une correspondance directe entre les catégories tonales et les contours observés de f_0 : les tons se réalisent en tant que cibles tonales, clairement identifiables comme des points d'inflexion de la courbe de f_0 et en napolitain la différence entre les questions et les affirmations se traduit dans une différence de leur alignement relativement à la syllabe tonique.

Cependant, il suffit de faire de petits changements dans l'identité segmentale de la phrase pour rendre la correspondance entre tons et cibles tonales beaucoup plus difficile à interpréter. Prenons l'exemple dans la fig. 1.3. Dans cette figure, la phrase *La mamma vuole vedere la Dana* ("la maman veut voir (la) Dana") a été prononcée comme une question. Cette fois-ci, l'identité segmentale a été légèrement changée, car l'attaque de la syllabe accentuée est un /d/ et non pas un /n/. La présence de l'obstruante voisée produit une forte baisse de la f_0 sur la consonne, suivie par une augmentation soudaine des valeurs de f_0 à la fin de la constriction consonantique. Puis, la f_0 descend à nouveau au début de la voyelle basse /a/. Par conséquent, deux coudes de f_0 sont créés : l'un est en correspondance de l'attaque syllabique et l'autre à l'attaque vocalique.

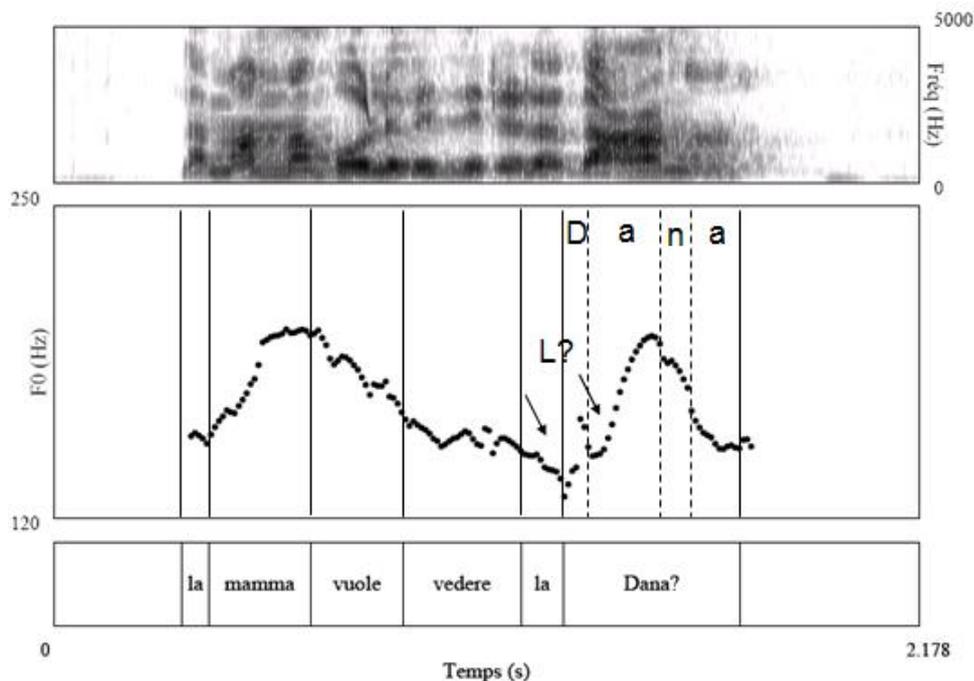


FIGURE 1.3 – Spectrogramme, courbe de f_0 et cibles tonales pour l’affirmation *La mamma vuole vedere la Dana* (“la maman veut voir (la) Dana”). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur du prénom *Dana*. Les flèches montrent deux localisations possibles de la cible L. Locuteur AS.

Le premier problème face à cette courbe de f_0 est d’ordre méthodologique. Comment peut-on localiser de façon objective la cible L ? Si la cible correspond tout simplement au minimum de f_0 , nous pouvons penser qu’elle est alignée avec le début de la syllabe, où la f_0 atteint sa valeur minimale. Mais vu que l’attaque syllabique est une occlusive, cette valeur pourrait tout simplement correspondre au point maximal de constriction consonantique. La cible L est-elle masquée par la perturbation segmentale ? Ou a-t-elle été déplacée dans une autre portion de la syllabe (c’est-à-dire, à l’attaque vocalique) ? D’un point de vue acoustique, un des problèmes lié à la localisation temporelle des cibles tonales peut consister en leur extraction manuelle, après une simple inspection visuelle de la courbe de f_0 . Cependant, le cas illustré dans la fig. 1.3 nous prévient du risque inhérent à cette méthode, car elle nous oblige souvent à adopter des critères subjectifs de localisation. On peut décider de localiser la cible L à la valeur minimale au début de la montée de f_0 , ou bien quelques millisecondes avant ou après la perturbation, ou bien

même encore peut-on négliger les effets segmentaux et prendre la valeur minimale de f_0 dans la portion initiale de l'attaque syllabique. Bien que ces critères puissent être adoptés de façon consistante par le chercheur pour une étude expérimentale spécifique (del Giudice *et al.*, 2007), nous doutons de l'objectivité de la procédure ainsi que de la possibilité de pouvoir comparer les résultats de travaux dont la méthodologie diffère tant.

Il faut noter que le problème méthodologique a des implications très importantes au niveau théorique. En italien napolitain, la différence d'alignement entre les questions et les affirmations est très subtile, car les deux cibles L et H de l'accent montant peuvent être réalisées à l'intérieur de la syllabe accentuée dans les deux modalités intonatives (D'Imperio, 2000 ; voir aussi les figures 1.1, 1.4 et 1.5). Par conséquent, il est crucial de déterminer si et comment la f_0 interagit avec le plan segmental. Par exemple, on pourrait supposer que le plan segmental n'est pas pris en compte dans la production de l'intonation. Dans ce cas, on s'attendra à ce que les cibles tonales soient réalisées de façon stable à des points segmentaux spécifiques, ceci indépendamment de l'identité segmentale de la syllabe accentuée. Une autre possibilité est que le contraste d'alignement entre questions et affirmations est influencé par l'identité des segments synchronisés avec les cibles tonales.

De plus, un problème lié à la distinction entre le plan segmental et le plan suprasegmental est de déterminer l'existence d'une interaction dans la perception des catégories phonologiques. Dans le cadre de la Théorie AM, la représentation acoustique des cibles tonales joue un rôle prédominant dans la définition des catégories phonologiques. L'importance toute particulière accordée à l'acoustique dépend uniquement de raisons méthodologiques, car la courbe de f_0 est la représentation phonétique de l'intonation la plus facilement observable et mesurable (Pierrehumbert, 1980).

Cependant, nous pensons que notre connaissance phonétique des catégories phonologiques ne doit pas seulement inclure l'étude de leurs mécanismes de réalisation acoustique, car leur représentation est aussi influencée par d'autres facteurs perceptifs, cognitifs et articulatoires ('t Hart *et al.*, 1990; Pierrehumbert, 2000). En particulier, dans cette thèse nous nous sommes aussi intéressés à vérifier comment les auditeurs napolitains peuvent accéder à une représentation mentale des cibles tonales même lorsque leur réalisation acoustique est influencée par des facteurs de variabilité, tels que les effets microprosodiques. Nous

savons déjà qu'en italien napolitain, la différence d'alignement des cibles acoustiques entre les questions et les affirmations est exploitée pour l'identification du contraste intonatif. En d'autres termes, lorsque la configuration en montée-descente LHL est déplacée dans la syllabe accentuée, les auditeurs napolitains identifient un nombre plus élevé de questions (D'Imperio & House, 1997; D'Imperio, 2000). Cependant, il n'est pas encore clair si la variabilité systématique d'alignement dû aux effets segmentaux peut influencer l'identification de modalité intonative. Un exemple très intéressant vient des effets de la structure syllabique sur l'alignement acoustique de la configuration LHL. Prenons les exemples illustrés dans les figures 1.4 et 1.5. Deux phrases, *è il nonno* ("c'est le grand-père") et *è il nono* ("c'est le neuvième"), ont été prononcées comme des affirmations à focalisation étroite et des questions oui/non. Les mots qui portent l'accent nucléaire montant, *nonno* et *nono*, sont paroxytons et ils sont en contraste de longueur segmentale et structure syllabique. En fait, en italien la gémination consonantique a une valeur contrastive et influence la structure syllabique : dans *nono*, la syllabe tonique *no-* est ouverte tandis que dans *nonno* la syllabe tonique *non-* est fermée. La structure syllabique a aussi des conséquences sur la durée de la voyelle tonique, qui est plus brève dans les syllabes fermées. Dans les deux phrases, l'accent nucléaire est suivi par une cible L, correspondante à l'accent de syntagme L-. A cause de l'interaction entre la structure syllabique et la modalité intonative, le pic de l'accent est soumis à quatre types d'alignement différents. Dans les affirmations, il est aligné au début de la voyelle tonique dans *nonno* et *nono*, mais la distance entre ce point et la fin de la voyelle est plus importante dans *nono* (205 ms) que dans *nonno* (137 ms). Dans les questions, le pic est aligné à 30 ms de la fin de la voyelle accentuée dans *nono*, et tout de suite après cette frontière dans *nonno*. On peut donc se demander si l'effet de la structure syllabique, systématique et prédictible, peut influencer l'identification du contraste entre questions et affirmations, ou si le plan segmental n'a aucun effet sur le sens intonatif. Il faut aussi remarquer que, puisque l'alignement tonal tend à être plus retardé avec les syllabes ouvertes qu'avec les syllabes fermées dans la même modalité intonative (D'Imperio, 2000; D'Imperio *et al.*, 2007), la différence d'alignement du pic mélodique devient un indice acoustique secondaire de la structure syllabique. Par conséquent, on peut se demander aussi si cette différence d'alignement influence la perception de paires

minimales lexicales en contraste de longueur segmentale structure syllabique.

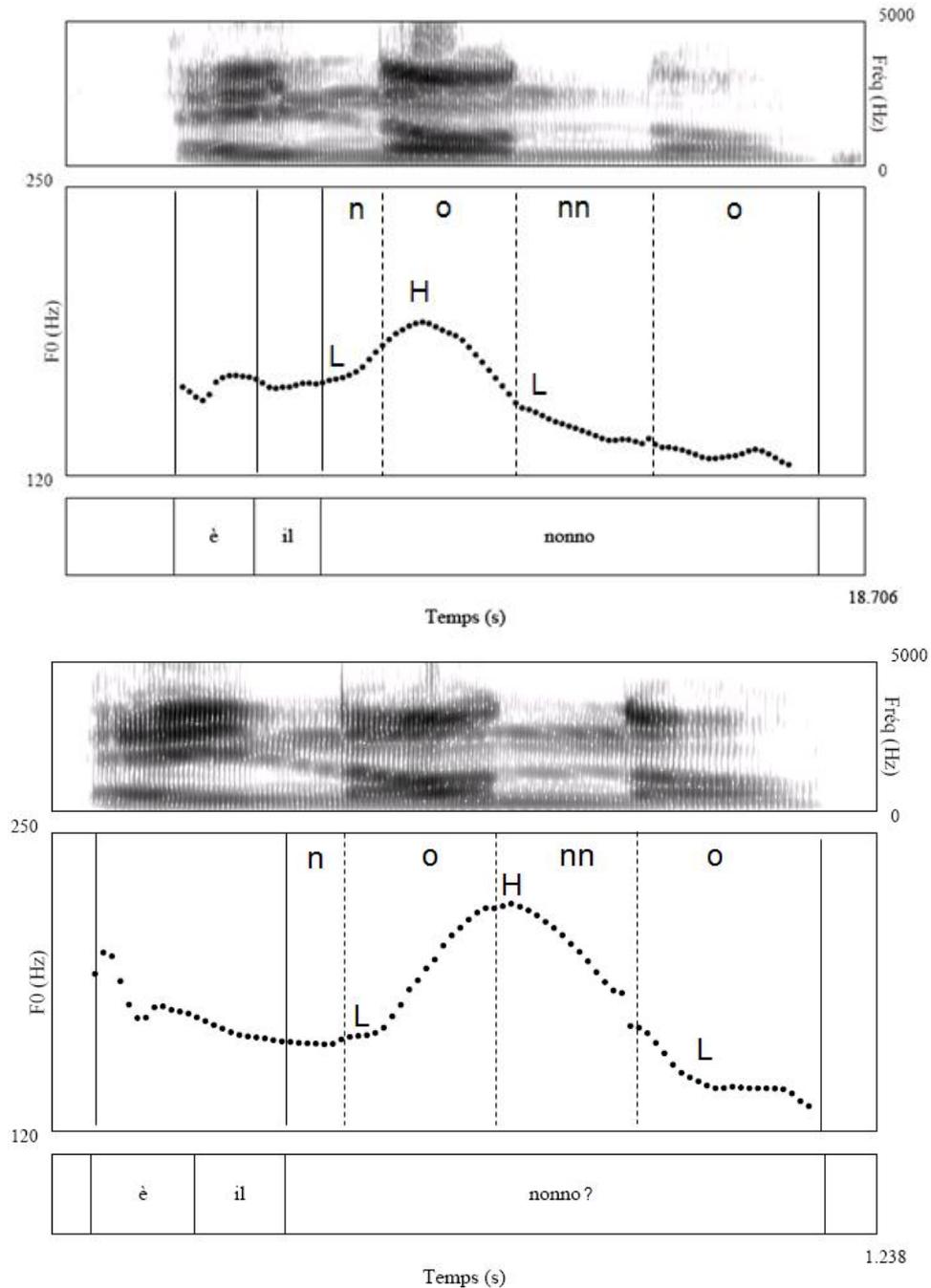


FIGURE 1.4 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *È il nonno* (“c’est le grand-père”), lue comme affirmation à focalisation étroite (en haut) et question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de *nonno*. Locuteur CC.

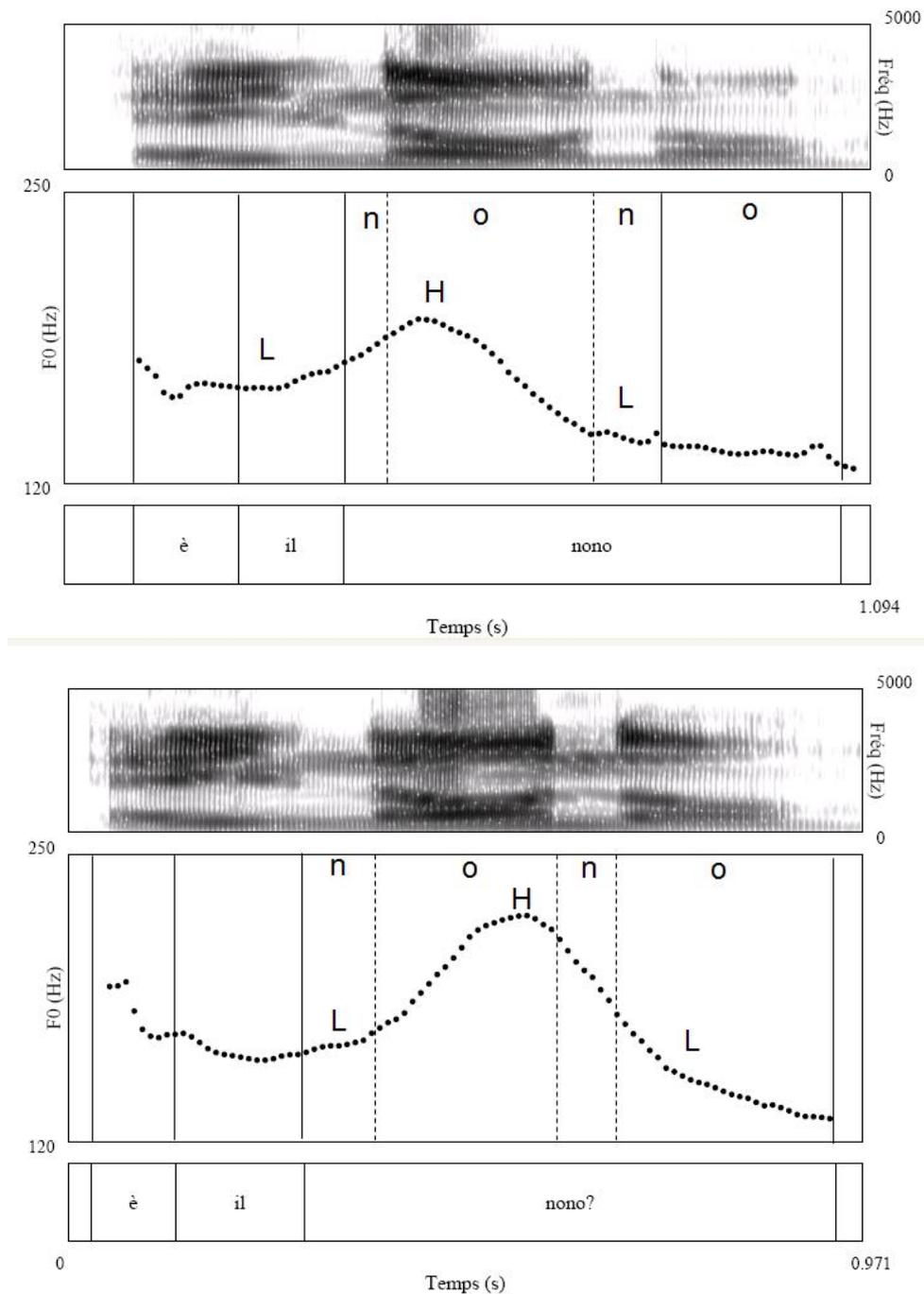


FIGURE 1.5 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *È il nono* (“c’est le neuvième”), lue comme affirmation à focalisation étroite (en haut) et question oui/non (en bas). Les lignes continues indiquent les frontières des mots, les lignes tiretées les frontières des segments à l’intérieur de *nono*. Locuteur CC.

D’autres facteurs de variabilité peuvent être cruciaux dans la production et perception du contraste intonatif. Jusqu’à maintenant, on s’est intéressé à l’accent nucléaire, car nous

savons qu' en napolitain son alignement tonal est l'indice le plus important du contraste entre questions oui/non et affirmations à focalisation étroite. Cependant, si nous portons notre regard à la région qui précède l'accent mélodique, nous découvrons une variabilité systématique de forme et de pente de f_0 de la région qui précède l'accent nucléaire. Dans la fig. 1.1, la question et l'affirmation sont prononcées avec deux accents. L'un est celui analysé jusqu'à maintenant, que nous avons désigné comme nucléaire. L'autre est celui réalisé sur *La mamma*, que nous avons désigné comme prénucléaire. Les deux accents sont montants indépendamment de la modalité intonative. Si l'accent nucléaire était le seul à véhiculer l'information de modalité, nous devrions nous attendre à ce qu'au niveau tonal il n'y ait pas d'autre moyen de distinguer les questions des affirmations. Pourtant, les figures 1.1 et 1.2 nous montrent l'existence d'une variabilité dans la pente de la f_0 entre les deux accents. Dans les affirmations, la f_0 descend rapidement jusqu'à créer un coude dans la région suivant l'accent prénucléaire. Dans les questions, ce mouvement descendant est beaucoup plus doux. Par conséquent, le mouvement de f_0 se réalise parfois sous la forme d'un plateau haut (fig. 1.1) avant de tomber jusqu'à la valeur mélodique du L nucléaire. D'autres fois, la pente de la chute est légèrement moins raide que dans les affirmations, la chute même se réalisant sous une forme convexe (fig. 1.2 ; voir aussi la fig. 1.3). La variabilité de réalisation de la région interaccentuelle nous pose, à nouveau, des problèmes de nature méthodologique et théorique. En premier lieu, comment peut-on apprécier la différence entre les deux modalités de façon objective ? Pour les affirmations, il ne semble pas y avoir trop de problèmes : on pourrait simplement étiqueter manuellement le minimum de f_0 réalisé après le pic prénucléaire comme une cible L. Cette procédure si simple ne peut pas être adoptée pour les questions. Ici, les valeurs de f_0 restent assez élevées même après le premier pic, ce qui pourrait suggérer la présence d'une cible H additionnelle. Mais où peut-on localiser cette cible ? Si une cible H est présente, il est impossible de déterminer sa position manuellement, car elle ne semble pas être réalisée comme un pic accentuel local. Notre interprétation de la variabilité du contour implique, à nouveau, de partir de l'idée "concrète" de cible tonale comme valeur minimale ou maximale de f_0 . Cette fois-ci les différences sont visibles par une subtile variabilité dans la forme et dans la pente du contour, c'est-à-dire par deux dimensions phonétiques normalement

ignorées dans les études dans le cadre de la Théorie AM. Si notre interprétation est correcte, on pourrait se demander la raison de cette variabilité. En d'autres termes, même si un ton est présent quelque part à la droite de la syllabe accentuée "mam-", il faudrait aussi montrer pourquoi ce ton est produit et aussi pourquoi il est produit différemment dans les questions et dans les affirmations.

Les exemples illustrés jusqu'ici nous montrent, sous des angles différents, l'objectif principal de cette thèse : l'exploration de la variabilité phonétique et de sa contribution à la caractérisation acoustique et perceptive de l'intonation. En particulier, l'idée principale que nous défendons est que la variabilité systématique de la f_0 peut être exploitée pour la représentation phonétique et phonologique des contours intonatifs et pour l'interprétation de leur sens. Nous pensons que la correspondance entre catégories intonatives et leur représentation phonétique est très complexe car elle présuppose l'interaction de la f_0 avec d'autres variables phonétiques (notamment, le plan segmental) et la variation de plusieurs paramètres liés à la f_0 même (tel que l'alignement tonal et la forme du contour). De telles variations ont été souvent ignorées ou minimisées dans les études sur l'intonation, de sorte que, à notre avis, cela pourrait ramener à une compréhension partielle des relations entre forme et fonction de l'intonation. Prendre de front la question de la variabilité est aussi un présupposé fondamental pour l'adoption de méthodologies objectives et pertinentes qui puissent nous aider dans l'exacte localisation des cibles tonales même lorsqu'elles ne sont pas clairement visibles en tant que points d'inflexion de la courbe de f_0 .

L'exploration de l'interaction entre le plan segmental et la f_0 a été l'objet des deux premiers chapitres de cette thèse. En particulier, dans le chapitre 2, nous allons vérifier dans quelle mesure les effets segmentaux influencent la réalisation de l'accent nucléaire LH des questions et des affirmations du napolitain, et donc la production du contraste intonatif. Différentes méthodes de localisation des cibles tonales sont comparées, afin de montrer comment la variabilité de choix méthodologiques peut influencer l'interprétation des données. Dans le chapitre 3, nous allons explorer comment la variabilité est prise en compte en perception. En particulier, notre but est de vérifier si l'information segmentale et d'alignement est utilisée activement par les locuteurs respectivement pour l'identification des contrastes intonatifs et lexicaux.

Les chapitres 4 et 5 sont consacrés à l’exploration de la variabilité de forme et de pente dans la région prénucléaire, et de leur rôle dans la réalisation et perception du contraste entre questions et affirmations en italien. En particulier, dans le chap. 4 nous allons fournir des preuves empiriques que la région qui précède l’accent nucléaire est caractérisée par l’insertion d’un ton de périphérie, qui marquerait la fin du syntagme accentuel. Ce ton, dont la spécification tonale est différente dans les questions et dans les affirmations, est responsable de la variabilité de la forme de f_0 . Sur le plan méthodologique, sa localisation acoustique a été effectuée par l’utilisation de méthodes statistiques robustes de modélisation de la courbe. Ces méthodes nous ont permis de quantifier la différence de forme et de pente de la région interaccentuelle là où d’autres procédures de localisation des cibles se sont avérées défailtantes. De plus, dans le chap. 5 nous allons vérifier si la variation de spécification tonale du ton de syntagme accentuel est utilisée dans la perception du sens intonatif. En particulier, les propriétés sémantiques de ce ton semblent similaires à celles de l’accent nucléaire, de sorte que la variabilité de la f_0 dans la région prénucléaire peut être utilisée pour la perception du contraste entre questions et affirmations même lorsque l’information d’alignement de l’accent nucléaire n’est pas disponible.

Dans la section suivante de ce chapitre, nous allons décrire comment le contour intonatif est analysé au niveau formel et fonctionnel par la Théorie Métrique - Autosegmentale (par. 1.2). Vu notre intérêt pour l’accent nucléaire et le ton de syntagme accentuel du napolitain, nous faisons aussi des considérations générales sur le statut phonologique des accents mélodiques et des tons périphériques dans la Théorie AM (par. 1.3). Ensuite, nous allons porter notre attention sur la représentation phonétique des contours intonatifs (par. 1.4) et sur les facteurs de variabilité qui peuvent influencer la réalisation et la perception des cibles tonales (par. 1.5).

1.2 Forme et fonction de l’intonation dans la Théorie Métrique-Autosegmentale

Dans cette thèse, nous adoptons l’analyse de l’intonation développée principalement par Janet Pierrehumbert et ses collègues (Pierrehumbert, 1980; Pierrehumbert & Be-

ckman, 1988; Ladd, 2008, *inter alia*), qui est connue sous le nom de Théorie Métrique - Autosegmentale de l'intonation (*Autosegmental Metrical Theory*, désormais Théorie AM). Un présupposé de ce modèle est que le contour intonatif peut être décomposé en une série d' "événements" intonatifs locaux, les tons bas (L) et hauts (H).

Le "contour intonatif" (*tune*) fait référence, de façon très générique, à la courbe de f_0 d'une phrase, véhiculant une information de sens (Lieberman, 1975). Par exemple, dans la fig. 1.1 la même chaîne de mots a été prononcée avec le même patron rythmique mais avec deux courbes de f_0 différentes, véhiculant en italien napolitain respectivement deux contours intonatifs : celui des affirmations à focalisation étroite et celui des questions oui/non.

Un problème fondamental des théories de l'intonation est de définir de façon explicite les composantes individuelles du contour intonatif et comment elles fonctionnent dans les actes de parole. En particulier, l'idée de "compositionnalité" du contour est cruciale dans la Théorie AM ainsi que dans d'autres descriptions de l'intonation (Pike, 1945; Halliday, 1967; Crystal, 1969; 't Hart *et al.*, 1990, *inter alia*). D'un point de vue strictement formel, cette notion implique la possibilité de décomposer le contour intonatif dans des unités mélodiques plus petites et elle s'oppose donc à la conception "holistique" (Delattre, 1966; Rossi, 1999; Liberman & Sag, 1974; Sag & Liberman, 1975), selon laquelle le contour intonatif ne peut pas faire l'objet d'une analyse en éléments signifiants plus simples. La distinction entre les approches holistique et compositionnelle est fortement liée à la façon dont la relation forme-fonction est conçue par les différentes théories. Dans les approches holistiques, le contour intonatif dans son ensemble a un statut morphémique, chaque contour étant directement associé à une fonction pragmatique. Un exemple typique de l'anglais est le contour "surprise/redondance" décrit par Liberman & Sag (1974). Un autre exemple, cette fois-ci du français est un contour montant de forme concave et de niveaux /2-4+/, qui est associé à la modalité de la question selon Delattre (1966)). Pour les approches compositionnelles, la décomposition du contour intonatif en unités de description plus petites est justifiée par le fait que ces unités (ou leur combinaison, comme dans la tradition structuraliste américaine) ont un statut morphémique, le sens du contour intonatif étant donc le résultat de la contribution de sens véhiculé par les unités qui lui

sont intégrées (voir aussi ci-dessous).

Par conséquent, au niveau phonologique, la même courbe de f_0 peut être décrite de façons différentes. Dans la figure 1.6, la courbe de f_0 est esquissée pour la phrase anglaise *It's painted orange!* (“c’est peint en orange !”). En anglais américain, ce contour est souvent utilisé pour exprimer une idée de “surprise/redondance”¹. En suivant la conception “holistique”, nous pouvons considérer l’intonation comme un seul *gestalt* qui couvre la phrase entière (fig. 1.6, panneau du haut). Dans ce cas, la présence d’une partie du contour est inévitablement liée à celle d’une autre partie du contour (Lieberman & Sag, 1974). Une autre possibilité est d’analyser le contour en unités plus petites, mouvements de f_0 ou tons statiques. En particulier, nous pouvons décrire le contour comme composé par un mouvement montant suivi par un mouvement descendant (fig. 1.6, dans le milieu). Ce type de description est typique des approches “configurationnelles”, telles que celle de l’école britannique (Halliday, 1967; Crystal, 1969), ainsi que de la théorie IPO (’t Hart *et al.*, 1990). La Théorie AM, au contraire, attribue de la pertinence linguistique aux points initiaux et finaux uniquement, négligeant les trajectoires d’un point à l’autre. Le contour intonatif sera donc représenté comme une séquence des niveaux (ou tons statiques) L, H et L (fig. 1.6, en bas). En d’autres termes, selon la Théorie AM les trajectoires entre un ton et l’autre sont de simples transitions sans aucune importance linguistique. Le mouvement entre un point bas et un point haut peut être plus ou moins accéléré, de sorte que la pente de f_0 peut être plus ou moins raide ; il peut avoir une accélération non uniforme, de sorte que la forme de la courbe peut être convexe ou concave ; ou il peut même avoir une durée variable, de sorte que la transition peut être courte ou longue. Le résultat de l’analyse phonologique sera le même dans tous les cas : la courbe de f_0 sera représentée comme une séquence de tons L et H (fig. 1.7).

1. Un contexte typique de sa réalisation est le cas d’un locuteur qui entre dans une salle et exprime sa stupéfaction en voyant le travail de peinture (Lieberman, 1975).

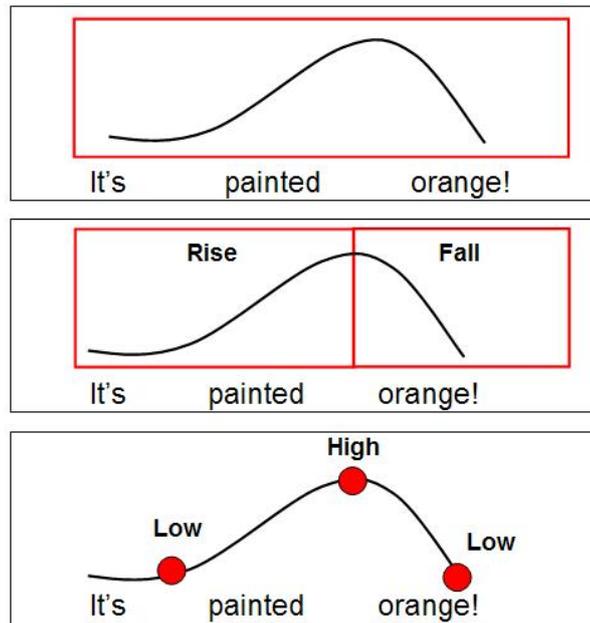


FIGURE 1.6 – Analyse du contour intonatif de *It's painted orange!* (“c’est peint en orange!”), selon l’approche holistique (en haut); et selon l’approche compositionnelle par mouvements (dans le milieu) et par niveau (en bas). L’encadrement en rouge du contour indique la taille de l’unité d’analyse pour l’approche holistique et par mouvements, et les points rouges figurent la représentation en tons statiques de l’approche par niveaux.

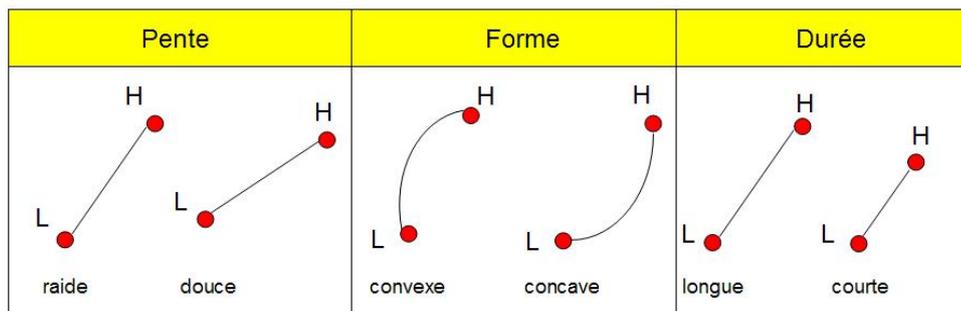


FIGURE 1.7 – Types différents de transition entre deux tons L et H.

L’analyse en termes de niveaux (*tone levels*) n’est pas, en elle même, une nouveauté introduite par la Théorie AM, car elle est déjà typique de la tradition structuraliste américaine (Pike, 1945, *inter alia*) ainsi que des premières descriptions dans le cadre de la phonologie autosegmentale (Leben, 1973; Liberman, 1975; Goldsmith, 1976; Bruce, 1977). Par exemple, dans sa description de l’anglais américain, Pike (1945) soutient l’existence de quatre tons ponctuels en contraste de hauteur mélodique, énumérés de 1 à 4 à partir

du ton dont la valeur mélodique est la plus élevée : le ton extra haut, haut, moyen et bas. Chez Pike (1945), les tons n'ont pas de fonction linguistique directe, mais c'est uniquement leur combinaison qui porte à la création des morphèmes intonatifs. La distinction entre le ton 1 et le ton 2 est analogique à celle, par exemple, des phonèmes /d/ et /t/ : dans les deux cas, les éléments phonologiques sont sélectionnés d'un même répertoire fini (respectivement, celui des tons et celui des consonnes), mais ils ne portent pas de distinction de signifié. Au contraire, la combinaison des tons amène à la création de contours en contraste de signification intonative. Par exemple, en anglais, la combinaison des tons 1 et 2 avec le ton 4 est utilisée pour l'assertion "neutre" (contour /2-4/) ou pour exprimer "insistance" (/1-4/). Il faut noter cependant que l'idée que les mouvements de f_0 puissent être utilisés dans l'intonation n'est pas complètement rejetée par les structuralistes. En premier lieu, les adhérents de cette approche admettent l'importance de la "jonction terminale" (*terminal juncture*) qui marque la fin de la phrase et qui est caractérisée par un mouvement montant, descendant ou plat (Trager & Smith, 1951). Cette jonction semble avoir un "statut" différent relativement aux tons ponctuels à l'intérieur du contour, car elle signale la fin de l'unité intonative (et donc, le *phrasing* prosodique). De plus, Pike (1948) suggère que la justesse de la description phonologique d'une langue dépend de ses caractéristiques spécifiques, de sorte qu'il est possible que dans certaines langues les mouvements tonals - et pas simplement ses points initiaux et terminaux - puissent avoir eux aussi un statut phonologique (voir aussi Anderson (1978) pour une description plus approfondie de ce sujet).

Dans l'approche de Pierrehumbert (1980) - et, même avant, dans le travail de Bruce (1977) -, les répertoires de niveaux est réduit à deux tons, H et L. La réduction du nombre des tons est fondamentale pour la résolution des problèmes inhérents à la tradition par niveaux (Ladd, 2008). Ces problèmes avaient été soulevés initialement par Bolinger (1951, 1958). Comme il l'a remarqué, il est en fait difficile de partitionner le contour intonatif en un nombre élevé de niveaux car la valeur mélodique de chaque niveau dépend des variations de l'étendue tonale du locuteur. Par exemple, il est difficile de dire si une courbe de f_0 qui monte et descend représente 4-1-4 dans une étendue tonale donnée, 4-2-4 dans une étendue plus large ou 4-3-4 dans une autre même plus large. De plus, comme

Ladd (2008) le souligne, un nombre élevé de niveaux amène aussi à une surabondance des combinaisons possibles. En admettant l'existence de deux seuls niveaux, la Théorie AM élude ces impasses, car les différences de hauteur des niveaux H et L sont considérées en termes relatifs, et les variations éventuelles de leur valeur mélodique dans l'étendue tonale du locuteur sont le résultat de simples variations paralinguistiques.

Au moins deux autres observations méritent d'être faites, concernant la distinction entre le modèle AM et l'école structuraliste. En premier lieu, la Théorie AM retient l'idée que la région de f_0 à la fin des unités tonales a un statut différent de celles à l'intérieur de l'unité même. Toutefois, cette distinction ne se traduit pas par l'admission d'un "mouvement tonal" tel que la jonction terminale, mais par l'introduction des tons périphériques, qui marquent la fin des unités prosodiques. En d'autres termes, les propriétés dynamiques de la courbe de f_0 ne "rentrent" pas dans le système phonologique, car tant les mouvements à l'intérieur de l'unité que les mouvements à sa fin ou à son début sont la conséquence de la présence des deux tons H et L. Une autre observation est que dans le modèle de Pierrehumbert (1980), les tons sont considérés comme des unités morphémiques (c'est-à-dire les signes prosodiques les plus petites doués d'un signifiant et d'un signifié). Ces morphèmes ont des domaines phonologiques différents, auxquels correspondent différents domaines d'interprétation (Pierrehumbert & Hirschberg, 1990). Par conséquent, dans la Théorie AM le contour intonatif n'est pas simplement une séquence de tons H et L, mais une séquence des tons organisée de façon *hiérarchique*. Les tons H et L se combinent en accents mélodiques (*pitch accents*) et tons périphériques (*edge tones*). Les accents mélodiques sont associés avec les unités prosodiques les plus petites (telles que les syllabes), tandis que les tons périphériques, tels que l'accent de syntagme (*phrase accents*) et les tons de frontières (*boundary tones*) sont associés avec le début ou la fin de constituants prosodiques (respectivement, le syntagme intermédiaire et le syntagme intonatif).

La notion d' "accent mélodique" (*pitch accent*) et son interprétation en tant qu'unité morphémique sont empruntées à Bolinger (1951, 1958). En particulier, avec "accent mélodique", Bolinger (1951) fait référence au *pitch obtrusion* réalisé en correspondance d'une syllabe tonique, définie donc comme "syllabe accentuée". Dans la Théorie AM, l'accent mélodique a en fait la fonction de rendre prééminente la syllabe qui lui est "associée"

(par. 1.3.1). Chez Pierrehumbert (1980), toutefois, l'accent mélodique n'est pas défini en termes dynamiques, mais il peut être composé par un ou deux tons, ce qui ramène aux accents monotonaux (L^* ou H^*) ou bitonaux (par exemple, L^*+H pour les accents "montants" et H^*+L pour les accents "descendants"). La liaison entre les deux tons de l'accent bitonal est représentée par le signe diacritique $/+ /$. Les tons périphériques, au contraire, sont obligatoirement monotonaux et leur statut est signalé graphiquement par d'autres diacritiques. En particulier, le signe $/\% /$ accompagne le ton de frontière (*boundary tone* : $H\%$ et $L\%$), qui marque la frontière du syntagme intonatif. Le signe $/- /$ signale l'accent de syntagme (*phrase accent* : H^- et L^-), qui marque la frontière du syntagme intermédiaire.

Les accents mélodiques et les tons périphériques sont générés par une grammaire, qui est une des trois composantes principales de l'intonation (les deux autres sont : la représentation métrique du texte et les règles d'association entre contour intonatif et texte). La grammaire de Pierrehumbert (1980) est conçue comme une machine à états finis :

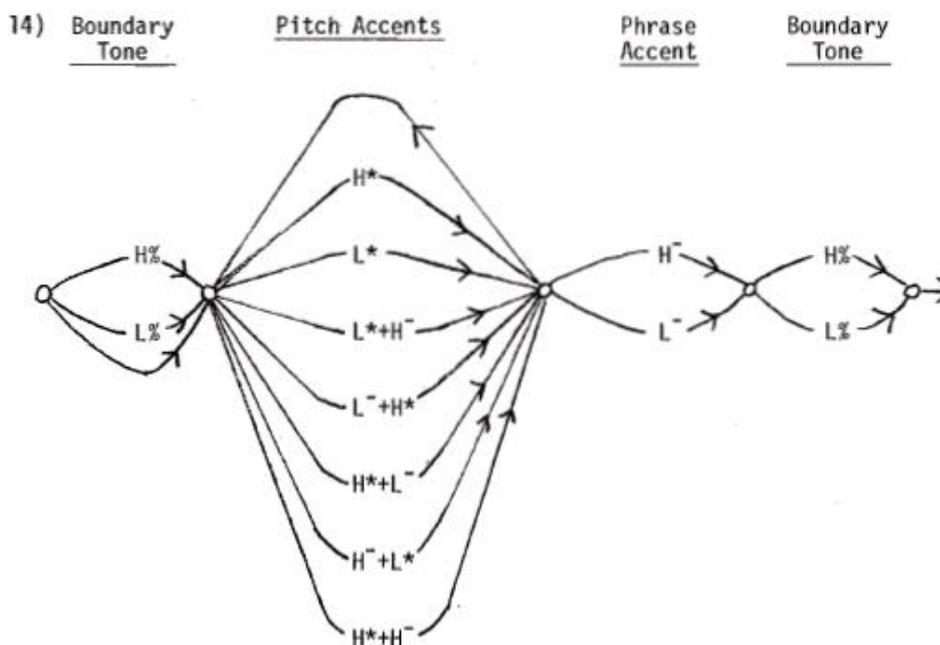


FIGURE 1.8 – Grammaire de l'intonation de l'anglais selon la Théorie AM. Tiré de Pierrehumbert (1980), p. 29.

Dans la fig. 1.8, la séquence des tons représente les états de la machine. Dans chaque état, le répertoire de catégories tonales est fini et discret : pour la langue anglaise, il y

a sept accents mélodiques (réduits à cinq dans la version amendée de Silverman *et al.*, 1992), deux accents de phrase et deux tons de frontière. Les lignes noires représentent les transitions d'un état à l'autre, et les flèches la direction des transitions. Les flèches vers la droite indiquent que la construction du contour s'effectue de gauche à droite, en passant d'un état au suivant : on réalise d'abord un accent mélodique ou plus, ensuite un accent de syntagme et enfin un ton de frontière. Pour les accents mélodiques, la présence de la flèche vers la gauche indique que cet état peut être répété plusieurs fois (le nombre d'accents mélodiques à l'intérieur du même contour intonatif est théoriquement infini). Un contour intonatif bien formé en anglais comprend donc un ou plus accents mélodiques suivis par un accent de syntagme et un ton de frontière. Plusieurs observations peuvent ressortir de ce réseau de transitions. En premier lieu, du point de vue computationnel, la grammaire de Pierrehumbert (1980) est "non-déterministe", car toutes les combinaisons entre les états du système sont possibles. En d'autres termes, tous les types d'accents mélodiques (L+H* ou L*+H, par exemple) peuvent se produire avec tous les types d'accents de phrase (L- ou H-) ou de tons de frontière (L% ou H%). Il faut noter que, bien que Pierrehumbert admette que la longueur infinie du contour est purement théorique, elle tient aussi à préciser qu'il ne semble pas y avoir de contraintes dans la façon dont les accents mélodiques et les tons périphériques s'assemblent (Pierrehumbert, 1980, p. 30) : "[...] *the claim that pitch accents occur in any combination, and with any phrase accent and boundary tone, appears to be better motivated [...] we have not been able to identify any cooccurrence restrictions among the pitch accents, though it is, of course, possible that future research will.*" ("le fait que toutes les combinaisons d'accents mélodiques peuvent se produire, et que les accents mélodiques peuvent se combiner avec tous les types d'accent de syntagme et de ton de frontière, semble être mieux motivé [...] nous n'avons pu identifier aucune contrainte de cooccurrence parmi les accents mélodiques, bien qu'il soit bien évidemment possible que des recherches futures puissent le faire").

La combinaison libre des tons dépend du fait que chaque ton contribue au sens du contour qui l'intègre de façon indépendante (Pierrehumbert & Hirschberg, 1990). En d'autres termes, le formalisme du modèle de Pierrehumbert (1980) est justifié par la compositionnalité même du sens intonatif. L'idée de la compositionnalité du sens intonatif

est à la base de plusieurs théories sémantiques de l'intonation (Gussenhoven, 1984; Pierrehumbert & Hirschberg, 1990; Steedman, 2003, *inter alia*). En particulier, en suivant Gussenhoven (1984), Pierrehumbert & Hirschberg (1990) affirment que la caractérisation du sens intonatif ne fait pas référence aux actes de paroles (Liberman & Sag, 1974; Sag & Liberman, 1975) ou aux attitudes émotionnelles du locuteur (Liberman, 1975), mais à leur interprétation dans le discours. L'interprétation des tons est liée aux savoirs et aux croyances que sont supposés partager les acteurs du discours. Les contours intonatifs sont choisis par les locuteurs “*to specify a particular relationship between the “propositional content” realised in the intonational phrase over which the tune is employed and the mutual beliefs of participants in the current discourse*” (“pour spécifier une relation particulière entre le contenu propositionnel réalisé dans le syntagme intonatif dans lequel le contour est employé et les croyances partagées par les participants au discours en cours”, Pierrehumbert & Hirschberg (1990) p. 285). Comme on l'a déjà dit, le domaine d'interprétation des tons dépend de leur domaine phonologique. Les accents mélodiques sont définis comme des “îlots de saillance”, qui fournissent des informations sur les voix lexicales. L'utilisation d'un accent mélodique plutôt que d'un autre dépend de l'*intention* du locuteur L de modifier ce qu'il croit que l'auditeur A pense être mutuellement partagé par L et A. Par exemple, en anglais, le ton H* indique que l'information relative au mot accentué est nouvelle et donc elle doit être ajoutée aux croyances partagées de l'auditeur. Au contraire, le ton L* est utilisé pour rendre saillant un mot que le locuteur croit ne pas être nouveau pour l'auditeur. Les tons périphériques sont utilisés pour mettre en évidence la relation entre le contenu propositionnel du syntagme en cours et le contenu propositionnel des autres syntagmes. Donc les tons de frontières véhiculent des informations sur la valeur *forward-looking* (“pro-active”) du syntagme intonatif en cours, tandis que les accents de phrases informent davantage sur comment le syntagme intermédiaire en cours doit être interprété en relation avec les syntagmes intermédiaires immédiatement précédents/suivants. L'interprétation de la compositionnalité du sens intonatif donnée par Pierrehumbert & Hirschberg (1990) est très forte : puisque chaque ton a un domaine spécifique qui n' “interfère” pas avec celui des autres tons, le sens du contour intonatif est le résultat de la simple somme de la contribution de sens de chacun de ces tons. Cette

idée est contraire à d'autres études antérieures, tels que Liberman (1975) ou Bolinger (1986). En fait, selon ces derniers, bien que le contour puisse être décomposé en unités plus petites *the whole is always more than the sum of its parts* (“[le contour intonative dans son] ensemble est toujours quelque chose de plus que la somme des ses parties”, Bolinger (1986), p. 246). Il faut noter aussi que chez Pierrehumbert & Hirschberg (1990) le contour nucléaire (l'accent nucléaire et l'accent de syntagme et le ton de frontière, correspondant au *nucleus* de la tradition britannique) constitue implicitement le cœur sémantique des contours intonatifs. Le statut particulier accordé aux contours nucléaires est en effet typique de plusieurs théories sémantiques de l'intonation (voir par exemple Gussenhoven (1984)). En particulier, chez Pierrehumbert & Hirschberg (1990) la combinaison du ton H* avec l'accent de syntagme L- et le ton de frontière L% est décrite comme le patron typique des déclaratives “neutres” en anglais. Au contraire, les combinaisons H* H- H% et L* H- H% sont typiques respectivement des questions avec une montée finale (*high rise questions*) et des affirmations oui/non. Dans les phrases longues qu'elles analysent, les accents prénucléaires sont simplement une ou plusieurs copies de l'accent nucléaire, de sorte qu'il est difficile d'identifier la contribution exacte de la région prénucléaire de f_0 au sens intonatif (fig. 1.9).

Accent	Sens	Utilisation typique	Exemple
H*	élément nouveau	Affirmation déclarative « neutre »	You deliberately deleted my files H* H* H* L- L%
L*	élément déjà partagé	Question oui/non	Do prunes have feet? L* L* H- H%

FIGURE 1.9 – Exemples d'accents mélodiques et de leur contribution au sens en anglais. Les phrases sont tirées de Pierrehumbert & Hirschberg (1990).

Il faut noter toutefois que Pierrehumbert & Hirschberg (1990) admettent la difficulté d'expliquer la compositionnalité de sens de certains phénomènes intonatifs. Par exemple, elles proposent de façon générique l'existence d'une différence de sens entre les accents bitonaux dont le ton-étoile est L* (ex. L*+H et H+L*) vs. ceux dont le ton-étoile est H*

(ex. $L+H^*$ ou $H+L^*$) et elles admettent aussi que des paires minimales en contraste d'association phonologique (ex. $L^*+H/L+H^*$) puissent avoir des significations similaires (voir le par. suivant pour la notion d' "association"). Par exemple, les accents $L^*+H/L+H^*$ véhiculent une idée de "saillance", bien que le degré de saillance de $L+H^*$ soit plus important que celui de L^*+H . Cependant, elles n'arrivent pas à identifier la contribution exacte de chaque ton des accents mélodiques bitonaux au sens intonatif. De plus, elles n'abordent pas la question de comment des phrases avec plusieurs types d'accents mélodiques prénucléaires peuvent être interprétées.

L'interprétation stricte de la compositionnalité intonative a été récemment critiquée par un nombre petit - à vrai dire - de travaux. En particulier, Grabe *et al.* (1998) ont montré qu'en hollandais il y a une interaction entre le choix de l'accent mélodique et celui du ton périphérique au début de l'unité intonative (la "prêtête", selon une terminologie empruntée à l'école britannique et analysée en termes autosegmentales comme %L ou %H), et que cette interaction influence le sens attitudinal des contours intonatifs. Dans leur expérience, il ont fait écouter aux auditeurs une série de phrases dans lesquelles la prêtête était combinée avec plusieurs types d'accent mélodique. Ils ont trouvé que lorsque la spécification tonale de la prêtête et de l'accent mélodique était polaire (par exemple, un ton %L suivi par H^* ; ou un ton %H suivi par L^*), le jugement de locuteurs était plus positif que pour les combinaisons non-polaires, car le locuteur semblait plus gentil, plus sociable et moins irrité. Selon les auteurs, une explication possible de cette interaction entre le choix de la prêtête et de l'accent mélodique est que le sens du contour intonatif n'est pas purement "additionnel". En d'autres termes, en analogie avec les mots composés, on peut supposer que le sens du morphème accentuel peut être modifié par celui de la prêtête. A notre avis, le besoin d'avoir des tons non-polaires/polaires implique aussi qu'au niveau formel la combinaison des tons n'est pas libre. Grabe *et al.* (1998) et Gussenhoven (2004) observent cependant que les résultats peuvent être interprétés différemment : la non-polarité peut être provoquée par l'insertion de la prêtête, tandis que la "polarité" peut signaler tout simplement l'absence de la prêtête. Dans ce dernier cas, la polarité est en réalité l'effet de la propagation à gauche de l'accent mélodique.

La compositionnalité du contour a été mise fortement en question par Dainora (2002,

2006). Dans ses travaux, Dainora analyse la composition des contours intonatifs utilisés dans un corpus de parole semi-spontanée de l'anglais américain (*The Boston University Radio News Corpus*). Une découverte très importante est que la combinaison des tons n'est pas complètement libre, car le choix d'un type de ton influence celui des tons successifs. Par exemple, la présence d'un particulier type d'accent nucléaire et/ou d'accent de syntagme est un bon indice de la spécification des tons de frontière successifs. Ainsi, la probabilité d'avoir un ton H% est très élevée lorsque dans le contour intonatif un accent nucléaire L* ou un accent de syntagme L- sont réalisés, tandis qu'elle est beaucoup plus basse lorsque le ton H% est précédé par un accent mélodique L*+H ou par un accent de syntagme H- . Dainora affirme donc que la grammaire de Pierrehumbert doit encoder les informations relatives aux probabilités de transitions entre un état et l'autre de la machine. En particulier, elle propose que les contours intonatifs contenant un accent nucléaire, un accent de syntagme et un ton de frontière sont représentables par une chaîne de Markov de deuxième degré, car la probabilité d'obtenir un type particulier de ton de frontière dépend des deux états précédents de la machine (c'est-à-dire, de l'accent nucléaire et de l'accent de syntagme).

Dans cette thèse, nous fournir des preuves qui suggèrent qu'en napolitain la compositionnalité du contour n'est pas libre, mais elle serait influencée par le contraste phonologique entre questions et affirmations. En particulier, les questions et les affirmations se distingueraient par une différence de catégorie tonale de l'accent mélodique nucléaire (L*+H/L+H*) et du ton périphérique qui le précède (H_{AP}/L_{AP}). Nous allons définir ce ton comme un ton de syntagme accentuel (*Accentual Phrase*). Au niveau phonétique, cette différence se reflète respectivement dans une variabilité systématique de l'alignement tonal de l'accent nucléaire et dans une différence dans la forme et pente du contour de la région prénucléaire. La différence de choix phonologique fait penser à une combinaison "déterministe" du contour intonatif, car nous avons trouvé que le L*+H est toujours précédé par le ton H_{AP}, et L+H* est toujours précédé par L_{AP}. Cette différence de combinaison semble relever même d'un point de vue perceptif, suggérant que l'interrelation entre les tons à l'intérieur d'un contour intonatif ne peut pas être capturée par une approche strictement compositionnelle du sens intonatif.

1.3 Représentation phonologique

Dans cette section nous voulons porter notre attention sur le statut phonologique des accents mélodiques et des tons périphériques. En particulier nous sommes intéressés à explorer la notion d’association entre les accents mélodiques et le texte, car elle est fondamentale pour la distinction des accents nucléaires des questions (L^*+H) et des affirmations ($L+H^*$) en napolitain et pour leur représentation phonétique. De plus, nous abordons brièvement la relation entre les tons périphériques et les niveaux de constituance, car dans cette thèse nous explorons l’existence d’un troisième niveau de constituance en italien napolitain, le syntagme accentuel, spécifié différemment dans les questions (H_{AP}) et dans les affirmations (L_{AP}).

1.3.1 Accents mélodiques et association

Nous avons vu que selon la grammaire de Pierrehumbert, les accents mélodiques peuvent être monotonaux ou bitonaux. Dans la fig. 1.8, les accents mélodiques bitonaux sont caractérisés par le $/+/-$. De plus, les accents sont caractérisés par deux autres signes diacritiques, l’étoile $/*$ (présent dans les accents monotonaux et bitonaux) et le tiret haut $/^-$ (dans les accents bitonaux). Dans la terminologie métrique-autosegmentale, le ton qui porte le signe $/*$ est défini ton-étoile (*starred tone*), tandis que le ton qui porte le tiret est défini *leading* ou *trailing* selon sa position (avant ou après) relativement au ton-étoile. Le tiret haut $/^-$ a été supprimé dans les versions successives de la Théorie AM pour des raisons de simplicité typographique. Au contraire, la notation de l’étoile est fondamentale pour plusieurs raisons phonologiques.

En premier lieu, l’étoile indique le relation de prééminence entre les deux tons de l’accent bitonal, le ton *starred* étant métriquement plus fort (et, donc, plus prééminent) que le ton *unstarred* (voir fig. 1.10). Comme Pierrehumbert (1980) le remarque, la relation entre les tons *starred/unstarred* à l’intérieur de l’accent mélodique est la même que celle entre les syllabes fortes/faibles dans le pied accentuel. Cette idée est fortement inspirée des études de la phonologie métrique (Lieberman, 1975; Selkirk, 1978) et elle a été plus

récemment reprise par Grice (1995)². Selon Pierrehumbert (1980), cette différence entre les deux tons se traduit, au niveau perceptif, par le fait que les tons *starred* sont plus saillants que ceux *unstarred*.

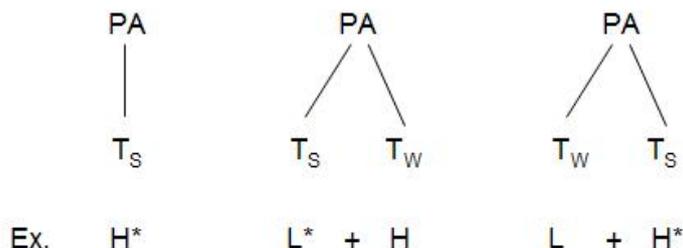


FIGURE 1.10 – Représentation schématisée de la structure interne de l’accent mélodique. *PA* = accent mélodique ; *Tw* = ton faible ; *Ts* = ton fort.

De plus, la présence de l’étoile nous donne une information sur comment les tons des accents mélodiques sont associés au texte. Nous rappelons que selon la phonologie auto-segmentale, les tons font partie d’une structure indépendante du texte, représentable dans une couche (*tier*) séparée : les tons sont des segments autonomes ou “autosegments”. L’association est donc le mécanisme par lequel deux segments de deux couches indépendantes (notamment, la couche textuelle et celle tonale) sont liés. Dans les premières descriptions autosegmentales des langues africaines (Leben, 1973; Goldsmith, 1976, *inter alia*), l’unité à laquelle le ton est associé (*Tone Bearing Unit*, TBU ; Goldsmith (1976)) est la voyelle : chaque voyelle doit être associée à un ton et chaque ton doit être associé à une voyelle. L’association est représentée par une ligne droite liant les deux autosegments et le ton associé au TBU est marqué par l’étoile.

Dans la Théorie AM, l’association de l’accent mélodique au TBU est aussi indiquée par l’étoile. En suivant les développements de la phonologie métrique (Lieberman & Prince, 1977), la syllabe acquiert un rôle crucial pour l’association du contour avec le texte car elle devient l’unité fondamentale pour cette association (fig. 1.11). Nous rappelons de plus

2. Selon Grice (1995), la structure interne de l’accent mélodique est plus similaire à celle du mot prosodique qu’à celle du pied accentuel. Selon la chercheuse, les tons *leading* et *trailing* des accents bitonaux ont un statut métrique différent, qui reflète la différence métrique entre les syllabes prétoniques et posttoniques dans les mots prosodiques : en anglais, le ton *leading* est dominé par le nœud *SUPERTONE* τ_w , tandis que le ton *trailing* est dominé par τ_s .

que selon la théorie métrique (voir en particulier Liberman, 1975), le texte est structuré de façon hiérarchique en positions métriques “fortes” et “faibles”, son organisation rythmique étant représentable sous la forme d’une grille métrique. L’association est donc fonction de cette organisation, car le ton “fort” (celui *starred*) des accents mélodiques monotonaux et bitonaux peut être associé uniquement à une syllabe métriquement proéminente. Par conséquent, bien que les accents bitonaux L^*+H et $L+H^*$ de l’anglais soient composés par les mêmes tons L et H , la relation de proéminence entre ces deux tons ainsi que leur association avec le texte sont différentes. Une opposition similaire a été trouvée aussi en italien napolitain, où les questions oui/non et les affirmations à focalisation étroite sont caractérisées par une différence de catégorie tonale de l’accent nucléaire, qui est L^*+H dans les questions et $L+H^*$ dans les affirmations (D’Imperio, 2000). Dans les deux langues, le ton L est le plus fort dans L^*+H et il est le plus faible dans $L+H^*$. De plus, il est le ton associé dans L^*+H tandis qu’il est le ton *leading* dans $L+H^*$.

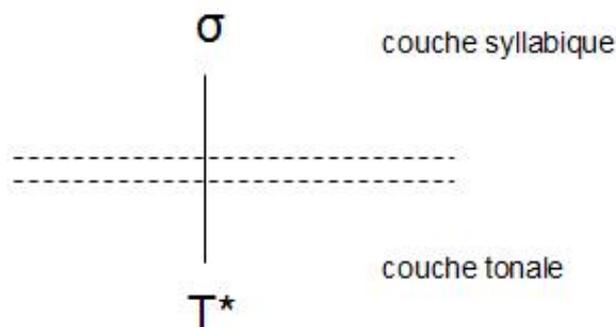


FIGURE 1.11 – Représentation schématisée de l’association entre la syllabe (σ) et le ton (T).

Plusieurs études dans le cadre de la Théorie AM ont souligné l’importance de l’association des tons sur le plan phonétique, et notamment sur l’alignement tonal, c’est-à-dire sur leur synchronisation temporelle avec le texte. Cette idée, introduite par Pierrehumbert (1980), est fortement inspirée du travail de Bruce (1977) sur l’accent lexical et l’intonation de la variété de suédois parlée à Stoccolme. Bien que chez Bruce l’association n’ait pas de définition explicite, dans son travail l’alignement temporel des tons relativement à la

syllabe tonique est fondamentale pour la distinction entre l'Accent Lexical I (*Accent I*) et II (*Accent II*). Dans le suédois de Stocolme, le ton a une valeur lexicale et la syllabe tonique des mots non-monosyllabiques peut porter l'Accent I ou II (les mots monosyllabiques portent uniquement l'Accent I). Bruce montre que lorsque les deux accents ne sont pas proches temporellement d'autres événements tonals de type intonatif (la présence de l'accent de syntagme et des jonctions terminales), ils sont très similaires : les deux accents sont réalisés sous forme d'un mouvement descendant de f_0 en correspondance avec la syllabe accentuée. Bruce remarque toutefois que le pic de f_0 au début de la montée se produit dans la syllabe tonique, dans l'Accent II ; et dans la syllabe prétonique, dans l'Accent I. Il propose donc que l'alignement retardé/avancé du pic est le trait phonétique crucial pour la distinction des deux accents.

En suivant Pierrehumbert (1980), la distinction d'alignement relatif entre deux accents peut être interprétée comme le reflet d'une différence d'association des tons de deux accents bitonaux. Par conséquent, en termes autosegmentaux, les deux Accents I et II seront composés des mêmes tons, H (correspondant au début de la descente de f_0) et L (correspondant à la fin de la descente), associés différemment à la syllabe tonique : l'Accent I sera donc analysé comme H^*+L et l'Accent II $H+L^*$ (cette analyse a été adoptée par Bruce (1983)). Chez Pierrehumbert (1980) et Pierrehumbert & Beckman (1988) la relation entre association et alignement est fondamentale pour la description des catégories intonatives de l'anglais. Ainsi, en anglais, l'alignement de l'accent montant relativement à la syllabe accentuée est plus retardé dans L^*+H que dans $L+H^*$. L'alignement temporel a, dans ce cas, une valeur phonologique, car indique la présence de deux catégories tonales séparées (voir aussi Pierrehumbert & Steele (1989) pour une preuve empirique de la valeur catégorielle de ce contraste phonétique). Il faut noter cependant que Pierrehumbert (1980) ne fait aucune prédiction phonétique sur la coordination temporelle exacte entre le ton et la TBU, mais elle indique de façon très générique que dans les accents bitonaux “*it is the stronger one which lines up with the syllable*” (“c'est le (ton) le plus fort qui est aligné avec la syllabe”, Pierrehumbert (1980), p. 77). De même, Pierrehumbert & Beckman (1988), remarquent tout simplement que l'association du ton-étoile avec le texte “*retains influences on its phonetic realization*” (“a une quelque influence sur sa réalisation

phonétique”, p. 121). Cette position est motivée par le fait que, selon les deux chercheuses, la signification de la notation */** est purement phonologique, car elle indique l’existence de catégories accentuelles en contraste pour l’alignement relatif de l’accent bitonal avec la TBU (voir aussi leurs commentaires à la présentation de Arvaniti, Ladd et Mennen à la conférence LabPhon V, et mentionnés dans Arvaniti *et al.* (2000)). De plus, la réalisation du ton *unstarred* ne dépend pas de quelque relation directe entre celui-ci et le texte : il est réalisé à une distance fixe du ton *starred*, et cela indépendamment de l’identité segmentale et de la structure syllabique du texte (Pierrehumbert, 1980).

L’alignement des tons *starred/unstarred* et son interprétation phonologique ont été l’objet de travaux successifs dans le cadre AM. D’un côté, la notation */** a été de plus en plus interprétée comme une indication de l’alignement du ton *starred* avec la syllabe tonique (et, vice-versa, l’alignement tonal a été interprété comme une indication de l’association phonologique). Par exemple, Grice *et al.* (2007) expliquent les différences d’association des accents mélodiques de l’allemand en termes de différences dans l’alignement phonétique des tons relativement à la TBU. Donc, en allemand, l’accent H+L* et H+!H* sont caractérisés par le fait que le ton *starred* est, dans les deux cas, réalisé dans la syllabe accentuée, tandis que le ton *leading* est aligné avec la syllabe atone précédente. Des présupposés similaires sont aussi à la base de la transcription de l’intonation d’autres langues (voir Prieto *et al.*, 2005 pour un bref état de l’art).

De l’autre côté, plusieurs études ont montré que la relation entre association et alignement peut varier selon plusieurs paramètres, tels que la langue (ou la variété de langue) considérée (Arvaniti *et al.*, 2000; D’Imperio, 2000; Prieto *et al.*, 2005). Par exemple, en italien napolitain, la différence entre les accents nucléaires L*+H et L+H* ramène, comme en anglais, à un contraste d’alignement. Cependant, D’Imperio (2000) a montré que dans les deux accents, les tons LH se réalisent à l’intérieur de la même syllabe, les tons de l’accent L*+H étant légèrement plus retardés des tons de l’accent L+H*. Si on garde l’idée qu’il y a une correspondance directe entre alignement et association, il serait donc difficile d’établir quel est le ton (L ou H) associé à la TBU (fig.1.12). Dans une autre langue, le grec, la situation est inverse, car les accents prénucléaires LH sont caractérisés par le fait que les deux tons son réalisé au delà des frontières de la syllabe accentuée

(Arvaniti *et al.*, 2000). Dans d'autres cas, l'alignement semble dépendre de l'interaction avec d'autres facteurs, d'ordre phonologique (tels que la structure interne de l'accent mélodique, Grice (1995)) et phonétique (tels que l'allongement final ou le voisinage aux frontières prosodiques, Silverman & Pierrehumbert (1990) *inter alia*).

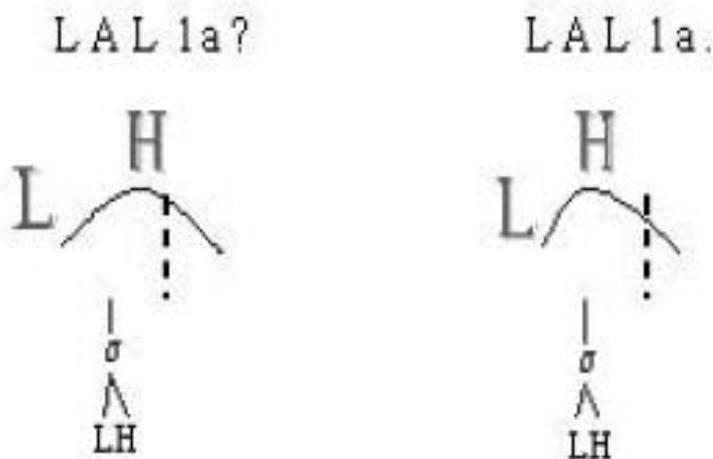


FIGURE 1.12 – Représentation schématisée de l'alignement des tons L et H avec la TBU en italien napolitain, dans les questions (à gauche) et dans les affirmations (à droite). Tiré de D'Imperio (2000), p. 25.

La variabilité d'alignement a été différemment interprétée par les chercheurs. Par exemple, Prieto *et al.* (2005) ont affirmé que l'alignement a une valeur phonologique lorsque des différences d'alignement phonétique se reflètent sur le choix des catégories tonales, et cette information doit donc être encodée dans la spécification des catégories tonales mêmes. Les auteurs proposent que l'alignement “contrastif” peut être le résultat du mécanisme phonologique d'association secondaire. La notion d'association secondaire de l'accent mélodique s'inspire à des découvertes relatives à l'alignement des tons périphériques effectuées par Pierrehumbert & Beckman (1988) pour le japonais. Dans Pierrehumbert (1980), l'association avec la TBU est typique uniquement des accents mélodiques. Les tons périphériques ne bénéficient pas de ce mécanisme, car ils sont associés au début ou à la fin des constituants prosodiques. Pierrehumbert & Beckman (1988) trouvent cependant qu'en japonais les tons périphériques peuvent chercher pas seulement une association avec les bornes prosodiques, mais aussi avec des unités métriques. En fait, le ton H qui marque le syntagme accentuel (le constituant prosodique le plus petit en japonais) a

une association primaire avec la frontière droite de ce syntagme, et une association secondaire simultanée avec la première ou la deuxième mora sonorante du constituant même (la mora est en fait la TBU minimale dans cette langue). Au contraire, le ton de frontière L% cherche une association secondaire avec la première mora du syntagme accentuel suivant, mais uniquement lorsque cette position n'est pas occupée par un autre ton. La notion d'association secondaire des tons périphériques a été reprise par plusieurs chercheurs pour expliquer l'alignement des tons périphériques dans d'autres langues (Gussenhoven, 2000; Grice *et al.*, 2000).

En suivant ce raisonnement, Prieto *et al.* (2005) suggèrent que l'accent mélodique peut présenter une association primaire à la TBU et une association secondaire à des bornes métriques pour des contrastes spécifiques d'alignement. Dans le cas de l'italien napolitain, cela signifie que les accents mélodiques des questions (L*+H) et des affirmations (L+H*) sont associés à la syllabe tonique, et la différence de catégorie phonologique est représentée par le fait que le ton *starred* a une association secondaire avec la première mora de la syllabe. Cette différence d'association résulte dans la différence d'alignement relatif des tons à l'intérieur de la syllabe (fig. 1.13).

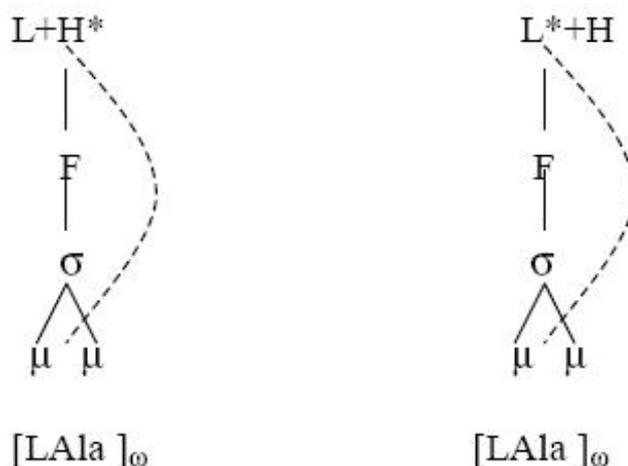


FIGURE 1.13 – Association primaire (ligne continue) et secondaire (ligne tiretée) pour les accents mélodiques L+H* et L*+H des affirmations et des questions du napolitain. Adapté de Prieto *et al.* (2005), p. 25.

Selon une perspective différente, soutenue par Ladd et ses collègues (Atterer & Ladd, 2004; Ladd, 2008), la variabilité d'alignement est preuve du fait que l'alignement phoné-

tique ne reflète pas l'association phonologique. Comme Ladd (Ladd, 2008, p. 8) l'écrit :

“[...] the simple fact of (phonological) association entails no detailed predictions about (phonetic) alignment. Obviously, if a H tone is associated with a given prominent syllable, we may expect to find a peak of F0 somewhere in the general vicinity of that syllable, not several syllables earlier or later, but the peak may be early in the syllable or late, and indeed it may be outside the temporal limits of the syllable altogether. Association should be kept as a phonological concept, and detailed differences of alignment should generally not be given a direct phonological representation.

(“La notion d'association (phonologique) n'implique pas en soi-même aucune prédiction précise de l'alignement (phonétique). Naturellement, si un ton H est associé avec une syllabe proéminente spécifique, nous nous attendrions à ce que le pic sera quelque part près de la syllabe, et non beaucoup de syllabes avant ou après ; cependant le pic peut être avancé dans la syllabe ou retardé, et de fait il peut être réalisé même au delà des bornes temporels de la syllabe. L'association devrait être regardée comme un concept phonologique, et de différences dans les détails d'alignement ne devraient pas avoir une représentation phonologique directe”).

Nous sommes d'accord avec Ladd sur le fait que, au moins dans certains cas, les différences d'alignement phonétique sont le simple résultat de l'implémentation phonétique. Par exemple, Atterer & Ladd (2004) ont montré que les détails d'alignement la même catégorie tonale L+H* sont réalisés différemment selon la langue (anglais et allemand) et les variétés de la langue (allemand du nord/sud), de sorte que une seule catégorie est réalisé dans un continuum phonétique graduel d'alignement. Cependant, nous ne sommes pas d'accord avec la généralisation de l'interprétation de ces découvertes. Atterer & Ladd (2004) comparent cette variabilité d'alignement à celle trouvée pour la réalisation du VOT (*Voice Onset Time*) dans des langues différentes. Selon les chercheurs, l'association est, comme le VOT, une entité phonologique discrète, qui se manifeste par de nombreuses variations d'alignement. En d'autres termes, il n'y a pas de valeur absolue d'alignement, qui corresponde à une catégorie phonologique spécifique (Ladd, 2008). Nous pensons au

contraire que dans une Théorie dans laquelle les primitifs mélodiques sont spécifiés par une seule paire de coordonnées phonétiques (l’alignement tonal et la valeur mélodique), la grammaticalisation de l’alignement phonétique est parfois nécessaire pour distinguer des accents bitonaux composées par la même séquence des tons et qui diffèrent pour la façon dont les tons sont alignés avec la TBU. Cela est le cas du napolitain, où l’alignement avancé/retardé des tons L et H est utilisé pour signaler deux catégories tonales montants (L*+H vs. L+H*). Il faut noter en fait qu’en napolitain, la catégorialité de ce contraste est justifiée par un contraste de sens véhiculé par les deux accents (questions oui/no vs. affirmations à focalisation étroite).

1.3.2 Tons périphériques et constituance prosodique

Outre les accents mélodiques, la grammaire de Pierrehumbert génère les tons périphériques, tels que les tons de frontières (L% et H%) et les accents de phrase (L- et H-). Les tons de frontières marquent la fin d’unités prosodiques larges, les syntagmes intonatifs (*intonational phrases*). Dans Pierrehumbert (1980), l’accent de syntagme est défini comme l’accent qui suit immédiatement l’accent nucléaire. Cette définition est inspirée de celle d’ “accent d’enoncé” (*sentence accent*) donnée par Bruce dans son travail de thèse sur le suédois (Bruce, 1977). En particulier, Bruce (1977) a trouvé que dans le suédois de Stoccolme, le contour intonatif des affirmations est caractérisé par un ou plusieurs accents de mot (c’est-à-dire, des accents spécifiés au niveau lexical) suivi immédiatement par une montée de f_0 et par une descente à la fin de l’unité intonative. Cette montée de f_0 est utilisée comme la preuve empirique de la présence du *sentence accent*, spécifié au niveau intonatif et distinct de la descente finale (*terminal juncture fall*).

Beckman & Pierrehumbert (1986) ont successivement reformulé la définition de l’accent de syntagme adoptée par Pierrehumbert (1980) en admettant qu’il marque la fin d’un constituant prosodique de rang inférieur au syntagme intonative, le syntagme intermédiaire. Les constituants prosodiques sont structurés de façon hiérarchique, et puisque le syntagme intermédiaire est de rang inférieur au syntagme intonatif, il ne peut pas le dominer. Cela est en accord avec le principe de l’étagement strict (*Strict Layer Hypothesis*, Selkirk (1984)). Il faut noter cependant que, bien que l’accent de syntagme soit généra-

lement accepté pour la description de l'intonation du suédois, son statut dans d'autres langues est assez controversé. Par exemple, Ladd (1983) critique l'adoption de l'accent de syntagme pour l'anglais car chez Pierrehumbert (1980) l'existence de ce ton semble être motivée par de raisons théoriques internes (notamment, le besoin d'admettre un accent descendant H^*+L , responsable de l'abaissement local de la hauteur des tons ou *downstep*). De l'autre côté, d'autres études ont rejeté la notion d'accent de syntagme, telles que Gussenhoven (1988) pour le hollandais et l'anglais et Féry (1993) pour l'allemand. Plus récemment, l'existence de l'accent de syntagme a été empiriquement prouvée par Grice *et al.* (2000) pour plusieurs langues de l'Europe de l'Est.

En italien, les tons de frontières ont été aussi attestés par des études dans le cadre métrique-autosegmental (voir Grice *et al.* (2005) pour un résumé). Ce ton marque la frontière droite de syntagmes intonatifs et il peut être haut ou bas. En particulier, un ton bas $L\%$ marque la fin du syntagme intonatif des affirmations et des questions oui/non du napolitain (D'Imperio, 2000; Grice *et al.*, 2005). L'accent de syntagme en napolitain a aussi été confirmée par D'Imperio (1997, 2001, 2002a). En italien napolitain, l'accent montant nucléaire est suivi par un mouvement descendant (fig. 1.1). De plus, dans les questions, l'accent nucléaire en position non-finale de syntagme intonatif peut être suivi par d'autres accents post-focaux. Ce phénomène est le même que celui observé en suédois (Bruce, 1977), langue qui admette aussi, comme mentionné précédemment, l'accent de syntagme. A partir de la constatation de cette similarité, D'Imperio (2001) a montré que les questions du napolitain sont caractérisées par un accent de syntagme, analysé comme un ton complexe HL- et ancré à la dernière syllabe tonique du constituant focalisé. Lorsque la syllabe qui porte l'accent nucléaire est la dernière syllabe tonique du constituant, le ton complexe HL- suit immédiatement l'accent nucléaire et la conséquence de ce voisinage est que le ton H de l'accent nucléaire L^*+H et le ton H de l'accent HL- sont réalisés comme un seul pic accentuel (fig. 1.14, en haut). Lorsque la syllabe accentuée n'est pas en position finale de constituant, le ton HL- est visible dans la trace acoustique comme un mouvement séparé de l'accent nucléaire : la fin de la montée nucléaire est alignée avec la syllabe accentuée, tandis que le début de l'accent de syntagme est ancré à la dernière syllabe tonique. Le plateau haut entre ce deux points est le résultat d'une simple

interpolation (fig. 1.14, en bas). La présence d'un ton complexe HL- est postulée aussi pour les affirmations (D'Imperio, 2000).

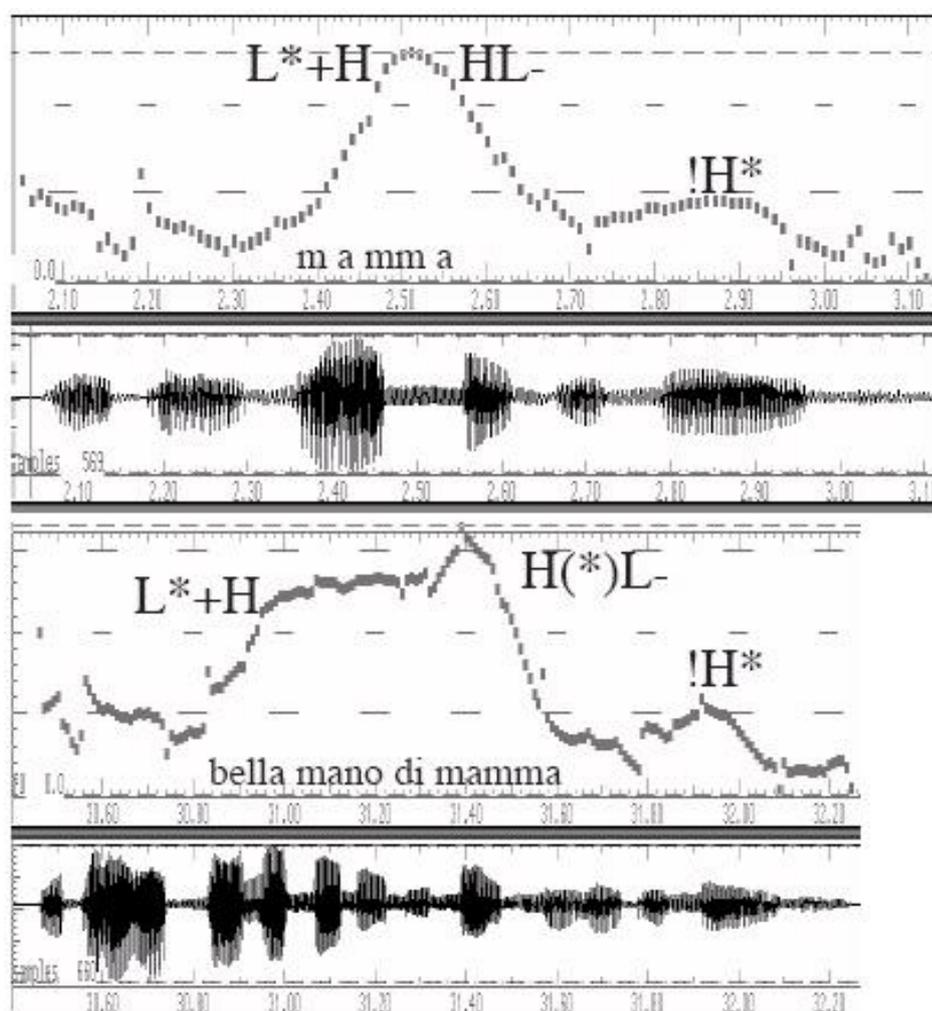


FIGURE 1.14 – Forme d’onde, courbe de f_0 et étiquetage tonal pour les phrases *Vedrai MAMMA domani?* (“Verras-tu maman demain?”, panneau en haut) et *Vedrai LA BELLA MANO DI MAMMA domani?* (“Verras-tu la jolie main de maman demain?”, en bas). Les lettres majuscules indiquent la taille du constituant focalisé. Dans les deux cas, l’accent de syntagme est suivi par un accent postnucléaire *downstepped*, étiqueté comme !H*. Tiré de D’Imperio (2002a), p. 48.

Au-delà du débat sur le statut phonologique de l’accent de syntagme, il faut spécifier que le type et le nombre des niveaux prosodiques (et donc, de tons périphériques) varient aussi selon la langue considérée. Nous avons vu que dans la description de l’anglais effectuée par Pierrehumbert (1980), il y a deux niveaux prosodiques. Cependant, d’autres

études ont montré l'existence d'un troisième niveau de constituance, de rang inférieur au syntagme intermédiaire et signalé par un événement tonal différent de l'accent de syntagme : le syntagme accentuel (*Accentual Phrase*). Par exemple, en japonais le syntagme accentuel est le domaine de l'accent culminatif, de sorte que chaque mot prosodique est signalé par la présence de tons périphériques à ses frontières gauche et droite (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988). Des preuves empiriques pour l'existence du syntagme accentuel ont été fournies pour le coréen (Jun, 1993) et même pour une langue romane, le français (Jun & Fougeron, 2000, 2002). En particulier, selon Jun & Fougeron (2002), le syntagme accentuel (appelé également *unité rythmique* par Di Cristo & Hirst (1993)) est composé d'au moins un mot prosodique associé à l'accent final, et il est marqué au niveau tonal par un ton de périphérie complexe montant (/LHi/). L'existence du syntagme accentuel dans les langues à accent dynamique (c'est-à-dire, dans les langues dans lesquelles la proéminence accentuelle est signalée par plusieurs indices outre la f_0 , tels que l'allongement de la durée segmentale et l'augmentation de l'intensité) est plus problématique. Beckman & Pierrehumbert (1986) abordent explicitement cette question pour la langue anglaise. Bien que certains travaux dans le cadre de la phonologie métrique puissent suggérer l'existence du syntagme accentuel en anglais, Beckman & Pierrehumbert (1986) admettent qu'aucune preuve empirique de la présence de marques tonales n'a été trouvée pour ce constituant.

L'existence du syntagme accentuel dans les langues à accent dynamique a été attestée par Hellmuth (2004, 2007) pour l'arabe du Caire. En arabe du Caire, chaque mot prosodique est accentué. Hellmuth (2004) a également suggéré que ce dialecte est caractérisé par deux niveaux de constituance prosodique au dessous du mot prosodique, le syntagme intermédiaire et le syntagme accentuel³. Le syntagme accentuel est composé minimalement par deux mots prosodiques, et il est signalé au niveau tonal par l'abaissement des valeurs de f_0 à la fin du constituant (*final lowering*) ainsi que par la remise à niveau au début du syntagme accentuel successif (*pitch reset*).

3. Plus précisément, Hellmuth (2004) appelle ces constituants, respectivement, Syntagme Phonologique Majeur (*Major Phonological Phrase, MaP*) et Syntagme Phonologique Mineur (*Minor Phonological Phrase, MiP*). Cette dénomination est due au fait que le travail de Hellmuth s'inscrit dans des récents développements de la Théorie de la Phonologie Prosodique (Selkirk, 2000), où le MiP et le MaP sont équivalents au syntagme intermédiaire et au syntagme accentuel de la Théorie AM.

L'italien napolitain présente aussi une situation très intéressante, qui peut être comparée à celle du japonais, du français ou de l'arabe du Caire. Les phrases illustrées dans la fig. 1.1 ont été lues comme un seul syntagme intonatif, contenant un seul syntagme intermédiaire. La fin du syntagme intermédiaire est signalée par le mouvement descendant qui suit l'accent nucléaire montant LH réalisé sur *nana*. De plus, la région entre l'accent prénucléaire et l'accent nucléaire semble être caractérisée par la présence d'un ton additionnel. En particulier, dans les affirmations, l'accent prénucléaire montant sur *mamma* est suivi par une chute de la f_0 jusqu'à la création d'un coude bas dans le verbe *vuole*. Dans les questions, la région suivante le premier accent mélodique offre à une forme parabolique, dont le point d'inflexion semble lui aussi être réalisé à la fin du mot prosodique. Ce patron intonatif est très différent de celui illustré dans la fig. 1.15, où la phrase a été prononcée avec une focalisation avancée sur *La mamma*. Dans cette figure, le mot *mamma* porte un accent nucléaire qui est spécifié comme L+H* dans les affirmations et comme L*+H dans les questions. Il faut noter en fait que l'alignement du pic est, dans cette figure, beaucoup plus avancé que l'alignement de l'accent prénucléaire montré dans la 1.1 : le pic nucléaire est réalisé à l'intérieur (dans les affirmations) et à la fin (dans les questions) de la voyelle tonique de *mamma*, tandis que le pic prénucléaire est réalisé dans la syllabe post-tonique suivante dans les deux modalités intonatives. Comme on s'y attendait, l'accent de syntagme est réalisé comme un mouvement descendant raide et son alignement est différent selon la modalité intonative : la fin de la descente est en fait réalisée dans la syllabe post-tonique dans les affirmations et après la fin de *La mamma* dans les questions. Cela suggère que le mouvement tonal après l'accent prénucléaire dans les figg. 1.1 et 1.2 peut être attribué à la présence d'un ton différent de l'accent de syntagme. Cette hypothèse sera mieux explorée dans le chapitre 4.

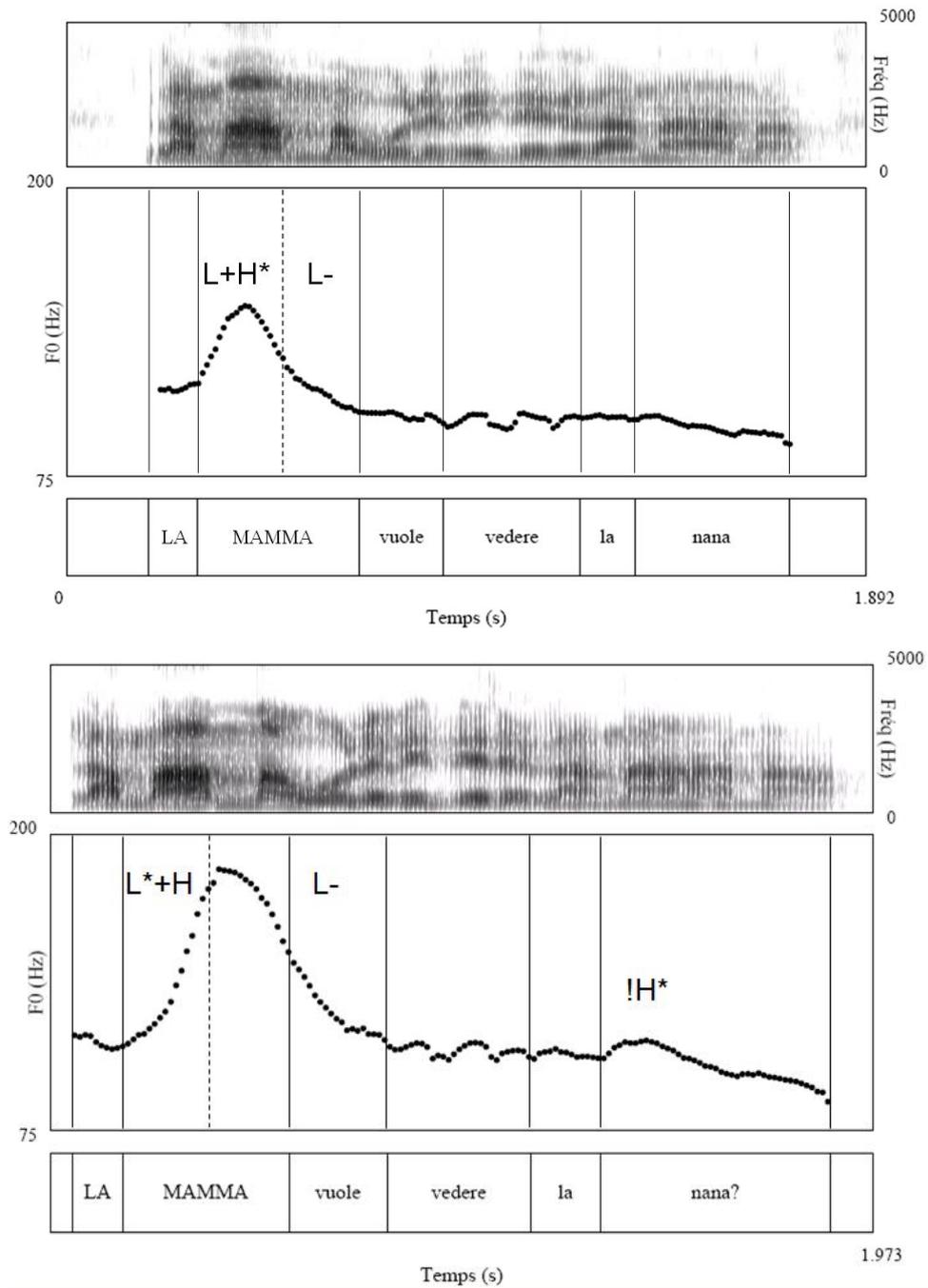


FIGURE 1.15 – Spectrogramme, courbe de f_0 et étiquetage tonal pour la phrase *La mamma vuole vedere la nana* (“la maman veut voir la naine”), lue comme une affirmation à focalisation étroite (en haut) et une question oui/non (en bas). La position du focus est avancée. Locuteur DD.

1.4 Représentation phonétique

Dans cette section nous explorons comment les représentations phonologiques sont représentées au niveau phonétique. Bien que la Théorie AM ait traditionnellement placé l'accent sur la représentation acoustique de l'intonation, nous pensons qu'aborder la question de la variabilité phonétique des catégories tonales ne puisse pas être effectuée sans prendre en compte comment l'auditeur y fait face. Dans cette thèse, le problème de la variabilité est abordé d'un point de vue acoustique et perceptif. Par conséquent, nous allons définir ci-dessous les notions de cible tonale "acoustique" et de cible tonale "perçue".

1.4.1 Cibles acoustiques

Un présupposé important dans la Théorie AM est qu'il y a une correspondance directe entre le contour intonatif et sa représentation physique. En particulier, selon Pierrehumbert (1980), la représentation phonétique la plus facilement observable de l'intonation est la f_0 . Cette correspondance directe contraste avec l'idée que le *mapping* entre représentation phonologique et signal acoustique est dû à l'intervention d'un niveau intermédiaire de représentation, tel que la représentation phonologique de surface dans la Théorie générativiste standard (Chomsky & Halle, 1968). Prenons l'exemple de l'alphabet IPA. En suivant la Théorie générativiste standard, on pourrait considérer l'alphabet comme un matrice universelle, capable de décrire la réalisation des phonèmes et des allophones de toutes les langues du monde par un ensemble fini de traits. Cependant, il a été largement prouvé que l'implémentation phonétique des catégories phonologiques est graduelle et dépend de la langue considérée (voir Pierrehumbert *et al.* (2000) pour un résumé), de sorte que les phonèmes ne sont pas vraiment réalisés de la même façon parmi les langues : *[...] there is no known case of two corresponding phonemes in two languages having fully comparable denotations...therefore phonological inventories only exhibit strong analogies* ("il n'y a pas de cas connus de deux phonèmes correspondants dans deux langues qui peuvent être dénotés de la même façon [...] par conséquent, les inventaires phonologiques montrent uniquement des fortes analogies", Pierrehumbert *et al.* (2000), p. 9).

De même, selon la Théorie AM, il n'y a pas de module intermédiaire universel entre

les catégories tonales abstraites et le contour de la f_0 . Les détails de réalisation des tons changent de langue à langue et la correspondance entre le contour intonatif et la f_0 est le résultat d'un *mapping* complexe défini par l'application de règles phonologiques et phonétiques différentes selon la langue considérée (telles que la règle du *downstep* ou les règles d'interpolation). Pour cette raison, la Théorie AM s'oppose à d'autres descriptions de l'intonation, tels que INTSINT (Hirst *et al.*, 2000) dans lesquelles la représentation phonologique des contours intonatifs se base non sur la f_0 , mais sur un niveau de représentation intermédiaire "universel" (on parle, d'ailleurs d'INTSINT comme d'un "alphabet intonatif", Di Cristo (2004)).

Dans la Théorie AM, les tons (accents mélodiques et tons périphériques) sont implémentés dans la courbe de f_0 comme cibles tonales (*tonal targets*), définies selon deux dimensions phonétiques : l'alignement tonal et la valeur mélodique. Les règles d'interpolation phonétique définissent comment la courbe de f_0 entre les cibles tonales est construite. Dans les études sur la prosodie, la notion de "cible" avait été déjà employée par Bruce (1977). Dans son travail de thèse, Bruce identifie les tons avec les valeurs maximales et minimales de f_0 dans la courbe de f_0 , qui correspondent à des points d'inflexion de la courbe (*turning points*) bien visibles. Par conséquent, on pourrait dire que chez Bruce (1977), le pic accentuel correspond à la cible H et le minimum de f_0 réalisé au début de la montée ou à la fin de la descente correspond à la cible L. La correspondance entre tons, cibles tonales et points d'inflexion de la courbe est beaucoup plus complexe chez Pierrehumbert (1980). En premier lieu, Pierrehumbert (1980) montre que dans plusieurs cas, les tons ne semblent pas être réalisés comme des point d'inflexion dans la courbe de f_0 . Par exemple, dans une séquence de tons H* H- L% (fig. 1.16), il est difficile d'établir la localisation exacte des cibles H* H- et H% car leur présence est signalée plutôt par une seule montée de f_0 de la syllabe accentuée "go" jusqu'à la fin de la phrase. Dans d'autres cas, le ton n'est pas réalisé dans la courbe, mais sa présence est uniquement sous-jacente (comme pour le ton flottant L de l'accent H*+L, dont la présence est indispensable pour expliquer le *downstep* en anglais). Au contraire, il y a des cas dans lesquels les points d'inflexion ne coïncident pas avec des cibles phonologiques, mais ils résultent de l'application des règles d'interpolation. Pierrehumbert (1980) affirme que pour l'anglais il y a deux

types d'interpolation : l'interpolation linéaire et l'interpolation *dipping*. Lorsque deux tons successifs ont une spécification tonale différente (par exemple, un L suivi par un H ou un H suivi par un L) ou sont tous les deux bas, la f_0 entre les deux cibles correspondantes sera linéaire. Cependant, deux pics mélodiques successifs seront liés par une interpolation *sagging*, qui implique un mouvement descendant de la f_0 après le premier pic, suivi par une montée jusqu'au pic successif. Cette transition provoque la réalisation d'un coude de f_0 entre les deux pics, qui n'est pas le reflet d'une cible phonologique, mais uniquement le sous-produit de la règle d'interpolation⁴.

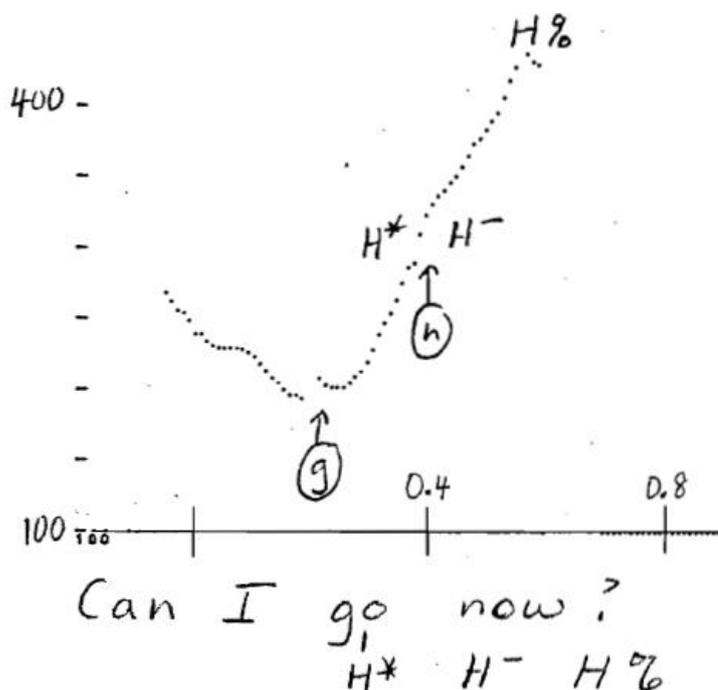


FIGURE 1.16 – Exemple du contour H*H-H% pour la phrase "Can I go now?" ("Peux-je partir maintenant?"). Tiré de Pierrehumbert (1980), p. 308.

Il faut noter cependant que les études successives dans le cadre de la Théorie AM (et en particulier, les études sur l'alignement tonal) ont souvent adopté la notion concrète de cible tonale de Bruce (1977). Ce choix dépend surtout de raisons méthodologiques : il est motivé par l'intérêt croissant pour la définition de la localisation temporelle exacte des cibles tonales en relation avec les frontières segmentales. La recherche des cibles acoustiques réside donc dans l'identification visuelle des cibles L et H avec les points correspon-

4. L'interpolation *dipping* a été récemment mise en discussion par (Ladd & Schepman, 2003), qui ont montré qu'en anglais, le minimum de f_0 entre deux accents H* correspond de fait à une cible tonale L.

dants aux valeurs minimales et maximales de la courbe de f_0 (voir la section 1.4). D'autres études ont souligné la difficulté d'appliquer cette méthode de façon objective et cohérente. Par exemple, dans plusieurs cas, les cibles sont masquées par la présence de perturbations microprosodiques, de sorte qu'il est impossible de localiser la cible sans adopter des procédures arbitraires de détection (Silverman & Pierrehumbert, 1990; D'Imperio, 2000, *inter alia*). Un cas typique est représenté dans la fig. 1.3, où la détection de la cible L peut être biaisée par la présence de l'occlusive sonore /d/ à l'attaque syllabique. Dans d'autres cas, les cibles H et L ne sont pas réalisées comme des points d'inflexion, mais comme des configurations à plateau (D'Imperio, 2000; Knight, 2003; Knight & Nolan, 2006). Une situation similaire est observée dans la fig. 1.1. Dans cette figure, le pic H de l'accent montant prénucléaire et le L de l'accent nucléaire ne sont pas liés par une interpolation linéaire. Dans la question, les valeurs de f_0 après le pic prénucléaire restent assez élevées, ce qui suggère la présence d'une cible additionnelle haute. La figure 1.2 montre une situation 'plus complexe : ici, la chute de la f_0 est clairement visible dans la question ainsi que dans l'affirmation, bien que la pente de la chute soit beaucoup plus douce dans la question, donc assumant la forme d'une parabole descendante. Si une cible additionnelle H est présente, il est pratiquement impossible de l'identifier avec un pic maximal de f_0 car sa présence semble être signalée uniquement par la forme convexe de la courbe de f_0 après le pic prénucléaire.

Une partie des problèmes concernant la représentation phonétique des catégories tonales pourrait être résolue par l'utilisation de méthodologies fiables d'un point de vue statistique pour la modélisation de la courbe de f_0 . Ces procédures doivent être basées, à notre avis, sur des hypothèses phonologiques sur la représentation des contours intonatifs. Une méthode "prometteuse" de modélisation de la courbe est l'algorithme MOMEL, développé par le laboratoire d'Aix-en-Provence (Hirst, 1983; Hirst & Espesser, 1993; Hirst *et al.*, 2000). Cet algorithme se base sur l'idée que le contour intonatif est représentable comme une séquence des tons statiques et que la courbe de f_0 peut être modélisée par une séquence de paraboles, les points d'inflexion des paraboles correspondant à la localisation temporelle des cibles tonales. Une autre méthode, encore peu connue et peu testée, est la méthode du moindre carré (*Least Square Fitting*, D'Imperio, 2000; Welby, 2003),

basée sur l’hypothèse que la courbe de f_0 est représentable par une séquence des cibles liées par des interpolations linéaires. La recherche de méthodologies objectives peut être aussi améliorée par des questions théoriques spécifiques. Par exemple, dans son travail de thèse, D’Imperio (2000) se pose la question de vérifier si, en napolitain, la pente de f_0 a une pertinence linguistique dans la définition des cibles tonales. Cette question amène l’auteur à utiliser des techniques non-linéaires de modélisation de la pente (la régression logistique), capables de capturer les propriétés dynamiques de la courbe de f_0 .

1.4.2 Cibles perçues

A part le développement méthodologique d’extraction des cibles acoustiques, la connaissance phonétique des catégories phonologiques pourrait être approfondie par des études en perception. Traditionnellement, les théories phonologiques (segmentales et suprasegmentales) ont fait prévaloir la production sur la perception de la parole pour la description des systèmes phonémiques. Cela est principalement dû au fait que, bien que les réalisations des catégories phonologiques soient relativement faciles à observer et mesurer, explorer les mécanismes de perception est une tâche beaucoup plus ardue. Ce manque d’intérêt constitue un “paradoxe” dans la mesure où la perception joue un rôle fondamental dans l’acquisition du langage et dans la construction des représentations cognitives (Pierrehumbert, 2000; Nguyen, 2004).

L’idée que la perception puisse jouer un rôle crucial dans la définition des catégories intonatives est à la base du travail de Kohler (1987) sur l’allemand. Il faut noter que les études de Kohler ne s’inscrivent pas dans le cadre métrique-autosegmentale, mais dans celui du modèle KIM (*Kiel Intonation Model*) développé par Kohler et ses collaborateurs (Kohler, 1991; Niebuhr, 2003, *inter alia*). Dans le papier de 1987, Kohler manipule la position du pic de f_0 de la configuration en montée-descente relativement à la voyelle accentuée. La voyelle était celle de la syllabe *-lo-* de *gelogen*, qui était inséré dans la phrase “*Sie hat ja gelogen*” (“elle jasait”). La manipulation d’alignement du pic est créée sur un continuum temporel de la syllabe prétonique à la syllabe post-tonique en pas de 30 ms. Ce continuum doit en fait couvrir l’espace sémantique de trois catégories pragmatique “établi”/“nouveau”/“emphatique”. L’expérience de perception se base sur le paradigme complet

de perception catégorielle déjà utilisé dans la perception segmentale (Repp, 1984), et comprend donc un test d'identification et un test de discrimination. Kohler (1987) montre que le continuum de f_0 est partitionné dans trois types d'alignement, *early/medial/late peak*, et que ce contraste est catégoriel. Les trois type d'alignement correspondent à trois catégories pragmatiques : lorsque le pic est aligné avant le début de la voyelle (*early peak*), l'accent véhicule le sens d' "établi", tandis que l'alignement du pic à l'intérieur de la voyelle *medial peak* véhicule l'idée de "nouveau". Finalement, la valeur pragmatique "emphatique" est véhiculée uniquement lorsque le pic est aligné avec la consonne suivante *late peak*. Le contraste entre *early/medial/late peak* est catégoriel, bien que Kohler trouve que la fonction de réponse est plus raide pour le contraste *early/medial* que pour *medial/late*. Ce travail de Kohler a influencé beaucoup de la littérature suivante sur l'intonation, et le paradigme de perception catégorielle a été employé pour montrer l'existence de catégories intonatives en contraste d'alignement et même de *scaling* dans plusieurs langues (voir Varnell Bosch (2006) pour un état de l'art).

La notion de cible tonale "perçue" est explicitement adoptée par D'Imperio (2000, 2005). Le travail de D'Imperio est inspiré au cadre de la Théorie AM, et la notion de cible "perçue" est donc complémentaire à celle déjà fournie pour les cibles acoustiques. Selon la chercheuse, les cibles "perçues" peuvent être définies comme "la représentation mentale d'une cible tonale en tenant compte des contraintes auditives. Si les cibles tonales sont définies au niveau acoustique à travers une paire de coordonnées de f_0 et de temps, une cible perçue sera alors définie par la traduction de ces valeurs au niveau perceptif " (D'Imperio, 2005, p. 92).

Dans son travail de thèse, D'Imperio (2000) étudie le rôle de la perception dans la définition des cibles tonales en italien napolitain. Dans une expérience de production préliminaire avec deux locuteurs, elle observe comment l'alignement tonal varie en fonction de plusieurs facteurs : la modalité intonative, la structure syllabique, et l'identité de la consonne initiale de la syllabe postaccentuelle. Les mots cibles sont au nombre de quatre : deux ayant une syllabe fermée, dont la coda consonantique est suivie par une occlusive nasale ou vélaire (*mammo/mango* "maman"⁵/"mangue" et les deux autres ayant une syl-

5. En italien, la durée consonantique est phonologique. *Mammo* est en réalité masculin, et avec ce

labe ouverte, dont la voyelle est suivie par une occlusive nasale ou vélaire (*nano/mago* “nain”/”magicien”). Les mots sont paroxytons et ils sont insérés en position finale d’énoncé dans deux phrases brèves. Les phrases sont lues comme affirmations à focalisation étroite ou comme questions oui/non. Le focus est retardé, de sorte que les mots cibles portent un accent nucléaire sur la syllabe tonique. Nous portons notre attention, dans ce paragraphe, sur l’effet de la modalité intonative sur la réalisation de l’accent nucléaire montant LH et de l’accent de syntagme suivant L-. D’Imperio (2000) observe que les trois cibles tonales sont plus retardées dans les questions oui/non que dans les affirmations. Elle remarque cependant que la modalité intonative provoque la variation d’autres paramètres acoustiques, tels que la hauteur des cibles et la pente de f_0 . En particulier, elle note que la valeur mélodique de la cible L de l’accent nucléaire est plus élevée dans les affirmations, tandis que la cible L de l’accent de syntagme est plus élevée dans les questions. Les résultats de l’analyse de la pente montrent une certaine variabilité inter-locuteurs : chez un locuteur, la pente de la montée nucléaire est plus raide dans les affirmations que dans les questions tandis que la situation est inversée pour le deuxième locuteur. Quels sont les indices pertinents pour la définition des cibles tonales perceptives ? En accord avec la définition métrique-autosegmentale de “cible tonale”, D’Imperio (2000) émet l’hypothèse que l’alignement acoustique est déterminant dans la définition des cibles “perçues” et que cette information est employée par les Napolitains pour l’identification du contraste entre questions et affirmations.

De plus, elle relève qu’en acoustique la cible H est parfois réalisée non comme un pic de f_0 , mais comme un plateau haut. Par conséquent, on peut se demander où localiser la cible dans le plateau, vu que tous ses points (le début du plateau, le milieu, etc.) sont tous de bons candidats pour sa localisation. D’Imperio (2000) affirme que cette ambiguïté peut être résolue uniquement en perception.

Dans une étude de perception très sophistiquée, elle teste donc l’effet perceptif de plusieurs facteurs phonétiques sur la perception de la modalité intonative. Elle manipule donc l’alignement tonal, la hauteur mélodique, la pente, et la forme de la configuration en montée-descente de f_0 constituée par l’accent nucléaire et l’accent de syntagme suivant.

terme on peut désigner un homme qui joue le rôle d’une maman.

Une phrase interrogative résynthétisée *Vedrai il nono ?* (“Verras-tu le neuvième?”) est employée comme base pour la création des stimuli, dans laquelle un seul accent est réalisé, à savoir un accent nucléaire sur *nono*. En ce qui concerne l’alignement, elle manipule la localisation temporelle des cibles de l’accent nucléaire L*+H et de l’accent de syntagme L-⁶ en 7 pas de 15 ms relativement au début de la voyelle accentuée. Dans cette série de stimuli, la cible H est réalisée comme un pic de f_0 , liée aux cibles L par des interpolations linéaires. De plus, pour tester l’effet de la forme, elle ajoute à son corpus deux autres séries de stimuli. Une série inclue 4 stimuli modifiés dans l’alignement, où la cible H est réalisée comme un plateau de 45 ms. L’autre série inclue également 4 stimuli, modifiés dans l’alignement, où la forme du contour en montée-descente est asymétrique. L’asymétrie est obtenue par des variations simultanées des valeurs mélodiques de la configuration LHL. Donc, par exemple, dans un stimulus le premier L a une valeur mélodique très basse, tandis que le H et le deuxième L ont une valeur très élevée ; dans le deuxième, les deux L ont des valeurs plus élevée, et le H a une valeur plus basse, etc.

En suivant le paradigme de perception catégorielle, la pertinence perceptive de ces facteurs est explorée par un test d’identification : trente locuteurs doivent écouter chaque stimulus et puis l’identifier comme une question oui/non ou une affirmation à focalisation étroite (*two-forced choiced task*). Les résultats ont montré que l’alignement avancé/retardé par rapport à la voyelle accentuée est employé systématiquement par les auditeurs napolitains. Pour les stimuli à la gauche du continuum d’alignement, l’identification des questions est très basse ; lorsque l’alignement est décalé vers la droite, l’identification des questions augmente (fig. 1.17). Il faut préciser que la contribution de chaque cible à la perception du contraste intonatif est indépendante. En fait, dans deux autres séries de stimuli, D’Imperio (2000) modifie respectivement l’alignement de la cible L de l’accent nucléaire et de la cible L de l’accent de syntagme. Elle observe que la manipulation de ces cibles influence l’identification des questions et des affirmations, ceci indépendamment de la position du pic. Au contraire, ni la manipulation de hauteur ni celle de pente ne semblent avoir un impact important pour l’identification de la modalité intonative.

6. Pour souci de simplicité, nous adoptons la transcription L- au lieu de HL- pour l’accent de syntagme. En fait, dans les expériences décrites dans cette thèse, le constituant focalisé est toujours court, ce qui donne une configuration en montée-descente de l’accent nucléaire et de l’accent de syntagme.

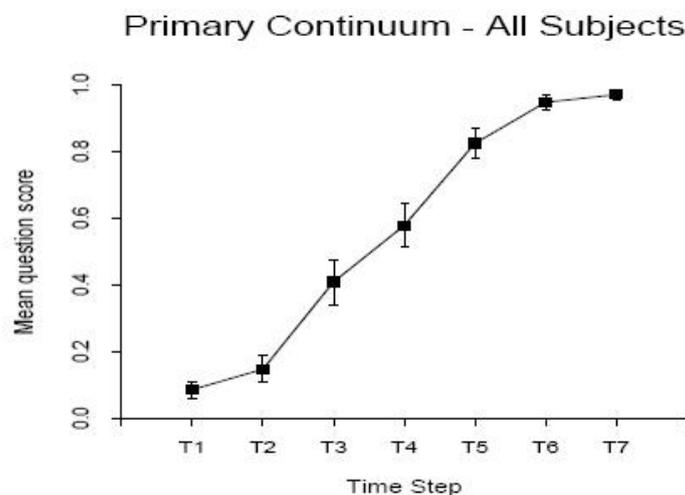


FIGURE 1.17 – Résultats de la manipulation d’alignement sur l’identification des questions en napolitain. Le score moyen de réponse (axe des abscisses) est fonction des pas d’alignement temporel (axe des ordonnées). $T1, T2... T7$ = pas d’alignement 1, 2 ...7. Tiré de D’Imperio (2000), p. 165.

En ce qui concerne la manipulation de forme, la présence du pic/plateau influence aussi la perception de la modalité intonative. Ceci est visible lorsque l’on compare les résultats d’identification des stimuli présentant le pic avec ceux des stimuli présentant un plateau. Lorsque les stimuli présentant un pic ont le pic aligné avec le début du plateau, le pourcentage de réponses “questions” est beaucoup plus élevé dans les stimuli ayant un plateau. Au contraire, l’écart entre les deux séries est beaucoup plus petit lorsque les stimuli présentant le pic sont alignés avec la fin du plateau. Les résultats pour les stimuli “asymétriques” sont plus difficiles à interpréter car l’identification des réponses ne semble pas être dû uniquement aux valeurs mélodiques des trois cibles et à l’alignement. La chercheuse émet l’hypothèse que pour ces stimuli, la cible perçue est le résultat d’une pondération globale entre les valeurs mélodiques atteintes par la montée de f_0 et des celles atteintes par la descente. Cependant, cette hypothèse n’est pas explorée.

Les résultats de cette expérience montrent que l’alignement des tons L et H de l’accent nucléaire et de l’accent de syntagme L- est l’indice le plus robuste de l’identification intonative. De plus, en napolitain, la position temporelle de la cible perceptive dépend aussi de la forme du contour, car dans les stimuli ayant un plateau, la cible perceptive H est

alignée à la fin du plateau (pour des résultats similaires en anglais et en italien de Pise, voir respectivement Knight & Nolan, 2006, et Gili Fivela & D’Imperio, 2007). D’Imperio (2000) établit donc une relation très étroite entre variabilité d’alignement tonal, représentation phonologique et sens : les résultats des expériences de production et de perception montrent que la variabilité systématique d’alignement de l’accent nucléaire est phonologique, ce qui confirme l’existence de deux catégories montantes en napolitain, L^*+H et $L+H^*$ (voir aussi D’Imperio & House (1997); D’Imperio (1999, 2002b) pour d’autres preuves empiriques). La grammaticalisation de l’alignement est utilisée pour véhiculer deux significations différentes, respectivement celle des questions et des affirmations.

Deux observations découlent des découvertes de D’Imperio (2000). En premier lieu, la distinction entre les deux catégories tonales L^*+H et $L+H^*$ - et donc celle de modalité intonative - repose principalement sur une variabilité d’alignement subtile mais systématique, qui peut même se réaliser à l’intérieur de la syllabe accentuée (D’Imperio, 2000). Cependant, nous nous demandons si cette variabilité “contrastive” peut être influencée par d’autres facteurs de variabilité, tels que l’identité segmentale ou la structure de la syllabe accentuée, et de quelles façons. Quelles sont les conséquences de ces facteurs sur la représentation acoustique de l’intonation ? Et comment l’auditeur fait-il face à de possibles effets segmentaux lorsqu’il doit interpréter le sens d’un contour ?

De plus, si la notion stricte de compositionnalité du contour est vraie et si le “nucleus” est réellement le cœur sémantique des contours intonatifs (par. 1.2), nous nous attendrions à ce que l’alignement de l’accent nucléaire et de l’accent de syntagme soit la seule source d’information de modalité intonative. Cependant, nous avons observé au cours de ce chapitre l’existence d’une variabilité dans la région prénucléaire. Nous avons parlé de cette variabilité en termes de “variabilité de forme et de pente” du contour, afin de souligner l’absence de points d’inflexion très visibles dans la courbe (en particulier dans les questions, où il n’y a pas de pic maximal de f_0). Cette variabilité est-elle systématique ? Influence-t-elle la représentation phonétique et phonologique du contour prénucléaire ? A-t-elle un impact sur le sens ?

1.5 Cibles tonales et facteurs de variabilité

Un problème “classique” qui se pose aux linguistes face au signal acoustique est son énorme variabilité. Dans les études segmentales, le rôle de cette variabilité dans la détermination des contrastes linguistiques a donné lieu à des hypothèses différentes. Pendant longtemps, la variabilité a été considérée comme inutile voire même nocive à la communication verbale. Par conséquent, la plupart des travaux se sont concentrés sur la recherche des invariants articulatoires (Lieberman *et al.*, 1967), acoustiques (Blumstein, 1986) ou perceptifs (Lahiri & Marslen-Wilson, 1991). Au contraire, d’autres travaux ont mis en évidence que la variabilité systématique du signal peut être utile aux locuteurs (Elman & McClelland, 1986) et que cette variabilité ne dépend pas seulement de contraintes physiologiques, mais aussi de la langue considérée (Pierrehumbert *et al.*, 2000).

Sur le plan intonatif, plusieurs études ont été menées à propos des conséquences de la variabilité sur la production et perception des cibles tonales, et en particulier sur leur alignement tonal. En acoustique, les découvertes dans la littérature sont assez problématiques. Une hypothèse importante, qui a été testée dans plusieurs langues, est celle de l’ancrage segmental (*segmental anchoring hypothesis* : Arvaniti *et al.*, 1998; Ladd *et al.*, 1999, *inter alia*). Selon l’interprétation stricte de cette hypothèse, les cibles tonales des accents mélodiques L et H sont synchronisées de façon systématique à des frontières segmentales, et ceci indépendamment d’effets possibles de variabilité phonétique tels que l’identité segmentale dans la syllabe accentuée ou le débit d’élocution.

Un des premiers travaux qui ont confirmé cette hypothèse est Arvaniti *et al.* (1998). Arvaniti *et al.* (1998) rendent compte de trois expériences de production où ils ont trouvé qu’en grec, le début et la fin des accents montants prénucléaires sont alignés respectivement avec le début de la syllabe accentuée et la voyelle post-tonique. Dans la première expérience, ils modifient l’identité segmentale dans la syllabe post-tonique, l’attaque consonantique étant une fricative, une nasale ou une occlusive et ils trouvent que le pic accentuel est aligné environ 10 ms après le début de la voyelle post-tonique, ceci indépendamment des variations de durée intrinsèque parmi les trois classes consonantiques. En ce qui concerne la cible L, ils nous informent que dans une étude précédente (Arvaniti &

Ladd, 1995), ils avaient trouvé que le L était aligné 5 ms avant l'attaque de la syllabe accentuée, mais ils ne donnent pas d'informations similaires sur l'alignement absolu du L dans cette nouvelle expérience. Cela est dû probablement au fait que dans cette nouvelle expérience, les chercheurs n'ont pas contrôlé de façon explicite l'identité segmentale de l'attaque de la syllabe accentuée (qui pouvait être composée même d'occlusives sourdes), de sorte qu'ils ont décidé de prendre la valeur de f_0 au début de l'attaque consonantique comme correspondante à la cible L même lorsque la cible n'était pas visible dans le tracé acoustique. Leur choix était justifié par les résultats obtenus par (Arvaniti & Ladd, 1995). Dans la deuxième expérience, Arvaniti *et al.* (1998) confirment que l'alignement de L et H est indépendant des variations de durée segmentale (notamment, de la durée de la syllabe accentuée). Plus précisément, ils trouvent que la distance entre les deux cibles est corrélée avec l'intervalle temporel entre le début de la syllabe accentuée et celui de la voyelle post-tonique, suggérant que la distance entre L et H n'est pas fixe (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988, voir par. 1.2) : les cibles L et H sont alignées indépendamment l'une de l'autre avec la chaîne segmentale. Puisque dans les expériences précédentes les mots cibles étaient des proparoxytons, dans la troisième expérience, ils utilisent des paroxytons et des oxytons pour vérifier l'existence d'effets possibles de voisinage à la frontière de mot sur l'alignement du pic. De plus, ils modifient aussi le nombre de syllabes entre l'accent prénucléaire et l'accent suivant pour vérifier si l'alignement de l'accent prénucléaire est influencé par le voisinage avec d'autres accents (*tonal crowding*; Silverman & Pierrehumbert, 1990; Prieto *et al.*, 1995, voir ci dessous). Leur résultats dans l'ensemble montrent que la durée de la montée est variable et que les cibles tonales L et H sont alignées avec des points spécifiques de la chaîne segmentale.

Des résultats similaires ont été obtenus par Ladd *et al.* (1999). Ladd *et al.* (1999) ont examiné les effets de vitesse d'élocution sur l'alignement tonal des accents prénucléaires montants en anglais par deux expériences de production. L'hypothèse de l'ancrage segmental est testée en fonction de deux hypothèses "concurrentes" : la durée de la montée et l'inclinaison de la pente sont les dimensions phonétiques fondamentales pour la définition des cibles tonales. Dans la première expérience, les auteurs ont demandé à trois locuteurs de lire des paragraphes à trois vitesses d'élocution différentes (rapide/normale/douce), et

ils ont montré que la durée de la montée et la pente de f_0 est influencée par la vitesse d'élocution : lorsque le débit est accéléré, la durée de la montée est plus courte et l'inclinaison de la pente plus raide. Dans la deuxième expérience, ils s'adressent explicitement aux effets de la vitesse d'élocution sur les points d'inflexions minimaux et maximaux au début et à la fin de la montée, points qui sont considérés comme correspondants aux cibles L et H de l'accent prénucléaire. Leurs résultats montrent que le début de la montée est ancré de façon stable avec le début de la syllabe. En ce qui concerne le H, ils trouvent une certaine variabilité d'alignement relativement à la fin de la voyelle accentuée (où le H était réalisé). Par conséquent, ils émettent l'hypothèse que l'alignement du pic accentuel n'est pas calculé en termes absolus, mais comme proportion de l'intervalle de durée entre la fin de la voyelle et le début de la voyelle atone dans la syllabe suivante. Ils trouvent que, lorsque l'alignement de H est mesuré en termes relatifs, les effets de vitesse d'élocution tendent à disparaître. Les auteurs trouvent aussi que l'excursion de f_0 (la différence entre la valeur mélodique de L et de H) est assez stable pour les trois conditions de débit, confirmant que le *scaling* des cibles est indépendant de la durée de la montée. La stabilité d'alignement et de *scaling* de ces points d'inflexion sont donc considérées comme une preuve empirique que les unités primitives de l'intonation sont les tons statiques, tandis que la variabilité de durée de la montée ainsi que celle de l'inclinaison de la pente sont interprétées comme une preuve que les mouvements tonals ne sont pas importants pour la définition des primitifs de l'intonation.

Des études acoustiques dans plusieurs langues et dans de cadres théoriques différents semblent confirmer la stabilité d'alignement de la cible L relativement à la syllabe accentuée. Par exemple, Caspers & van Heuven (1993) ont trouvé que l'alignement du début de la montée en hollandais est synchronisé au début de la syllabe accentuée, indépendamment des contraintes de pression temporelle telles que la vitesse d'élocution et la durée segmentale. Comme pour le grec, Prieto *et al.* (1995) ont remarqué qu'en espagnol mexicain l'alignement du début de la montée de l'accent H* est réalisé quelques millisecondes après l'attaque syllabique. En anglais, Ladd & Schepman (2003) ont trouvé que dans des paires minimales composées d'un prénom suivi d'un nom de famille (ex. *Norman Elson/Norma Nelson*) et dans lesquelles deux pics H* sont réalisés, le minimum de f_0 entre les deux pics

est aligné avec le début de la syllabe accentuée du deuxième mot. En d'autres termes, il est aligné avec la voyelle /e/ dans *Norman Elson* et avec la consonne /n/ dans *Norma Nelson*, donc constituant aussi un indice prosodique secondaire de la frontière de mot. En chinois mandarin, le début et la fin des mouvements tonals sont mis en phase avec le début et la fin des syllabes avec un angle de 0 degré (Xu, 2002).

Bien que les résultats pour la cible L semblent très clairs, l'alignement de la cible H semble être soumis à plusieurs sources de variation phonétique et phonologique. Parmi ces facteurs, nous retrouvons : l'environnement tonal (voisinage temporel de deux tons ou voisinage aux frontières prosodiques : Silverman & Pierrehumbert, 1990; Prieto, 2002; D'Imperio, 2002a), l'identité segmentale dans la syllabe accentuée (type d'attaque ou de coda syllabique : Silverman & Pierrehumbert, 1990; van Santen & Hirschberg, 1994; Prieto *et al.*, 1995; Welby & Loevenbruck, 2005) et la structure syllabique (Ladd *et al.*, 2000; D'Imperio, 2000; D'Imperio *et al.*, 2007; Petrone, 2005; Prieto & Torreira, à paraître).

L'étude de Silverman & Pierrehumbert (1990) considère l'effet de plusieurs facteurs sur l'alignement du pic de l'accent prénucléaire H* en anglais américain : la vitesse d'élocution, le voisinage à la frontière de mot et la collision à la syllabe tonique suivante (*stress clash*). Ils trouvent que l'alignement du pic accentuel est plus retardé lorsque la vitesse d'élocution est plus douce que lorsqu'elle est plus accélérée. De plus, il y a une corrélation négative entre l'alignement retardé du pic et le contexte prosodique à droite, car le pic est plus retardé lorsque la syllabe accentuée est loin de la frontière de mot et de la syllabe tonique suivante. Ces effets sont expliqués en termes d'ajustement aux variations de durée de la rime syllabique, et le degré de retard de l'alignement dépend de la source d'allongement de la rime. En fait, l'alignement retardé du pic est plus important lorsque la rime est allongée par effet d'une diminution de la vitesse d'élocution que lorsqu'elle est allongée par effet du voisinage avec la frontière de mot ou avec d'autres syllabes toniques. L'alignement proportionnel du pic (le rapport entre l'alignement du pic et la durée de la rime) en fonction du contexte prosodique et de la vitesse d'élocution est le résultat de l'équation linéaire suivante :

$$\text{peak proportion} = a + b.\mathbf{wb} + c.\mathbf{sc} + d.\mathbf{fast} + e.\mathbf{slow}$$

(1.1)

Dans l'équation (1.1), l'alignement proportionnel de H* (*peak proportion*) est fonction du *stress clash* (**sc**), du voisinage à la frontière de mot (**wb**) et de la vitesse d'élocution (**fast/slow**; le niveau de référence est la vitesse "normale"). Les coefficients *a*, *b*, *c*, *d*, *e* sont obtenus par la régression multiple. Dans Silverman & Pierrehumbert (1990), la variance expliquée par le modèle est de 65% environ.

Silverman & Pierrehumbert (1990) s'interrogent aussi sur les raisons de ces effets. Ils supposent que leurs données peuvent être expliquées par la combinaison de trois mécanismes : le changement du profil de sonorité (*sonority profile*), le chevauchement gestuel (*gestural overlap*) et la répulsion tonale (*tonal repulsion*). Selon la première hypothèse, la montée de f_0 est coordonnée au profil de sonorité de la syllabe, profile qui est défini par un geste d'ouverture (qui commence au début de la syllabe et se termine au pic de sonorité) et un geste de clôture (qui commence au pic de sonorité et se termine à la fin de la syllabe). Lorsque la syllabe n'est pas allongée, la montée de f_0 est réalisée pendant la durée de la syllabe, et le pic accentuel est réalisé à la fin de la phase de clôture. En présence de facteurs d'allongement à la droite de la syllabe, la phase de clôture est allongée et l'alignement du pic de f_0 est donc avancé. La deuxième hypothèse est celle du chevauchement gestuel. Selon cette hypothèse, déjà bien connue dans la littérature segmentale (Lindblom, 1963), le voisinage de deux gestes articulatoires concurrents est responsable des phénomènes de variabilité acoustique tels que l'*undershoot* tonal, la troncation et la répulsion tonale. En fait, lorsque deux gestes concurrents (tels que ceux responsables d'une montée et d'une chute de f_0) sont immédiatement contigus, les valeurs de f_0 des cibles correspondantes peuvent être inférieures à celles prévues, car les mouvements tonals n'arrivent pas à atteindre cette valeur (*undershoot* tonale). Dans d'autres cas, un geste articulatoire "gagne" sur l'autre, de sorte que le mouvement de f_0 du geste "gagnant" est complètement réalisé, et sa cible peut être atteinte. La valeur mélodique de l'autre cible sera cependant inférieure à celle prévue, par effet d'une troncation du mouvement de f_0

correspondant. Une troisième possibilité pour éviter le chevauchement gestuel est qu'un des deux gestes articulatoires est déplacé dans le temps (répulsion tonale). Dans ce cas, la valeur mélodique des cibles tonales est complètement atteinte, mais l'effet de la répulsion sera visible dans une variation d'alignement tonal.

Les résultats de Silverman & Pierrehumbert (1990) pour les accents prénucléaires sont similaires à ceux qui avaient été déjà trouvés par Steele (1986) pour les accents nucléaires, ce qui confirme, selon les deux auteurs, que les accents nucléaires et prénucléaires font partie du même inventaire phonologique. Cette idée justifie donc la grammaire à états finis de Pierrehumbert (1980), où il n'y a pas de distinction entre les accents prénucléaires et nucléaires. Il faut noter cependant que Silverman & Pierrehumbert (1990) trouvent aussi que, dans les mêmes conditions expérimentales, l'alignement du pic des accents prénucléaires est plus avancé que celui des accents nucléaires. Ils émettent l'hypothèse que l'alignement avancé du pic nucléaire est dû à une répulsion temporelle, répulsion qui évite un chevauchement avec le geste articulatoire de l'accent de syntagme suivant.

L'*undershoot* tonal, la troncation et la répulsion tonale ont été attestés aussi en italien (D'Imperio, 2000; Grice *et al.*, 2005). Un effet de chevauchement gestuel est aussi à la base de la différence de réalisation du contour intonatif dans la fig. 1.14, où la cible H de l'accent nucléaire L*+H et le début de l'accent de syntagme HL- sont réalisés comme un seul pic de f_0 lorsque le constituant focalisé est de petite taille, mais ils sont séparés par un plateau haut dans le constituant long⁷.

L'effet de l'identité segmentale sur l'alignement tonal est aussi l'objet d'étude de van Santen & Hirschberg (1994) et de Prieto *et al.* (1995). Selon van Santen & Hirschberg (1994), en anglais américain, l'alignement du pic de l'accent prénucléaire H* est fonction de la classe phonétique d'appartenance des segments ainsi que de la durée de l'attaque syllabique et de la rime. Le modèle de van Santen et collègues (van Santen & M obius, 1999; van Santen, 2002, *inter alia*) se base sur l'idée que les contours mélodiques peuvent être modélisés par une fonction de déformation temporelle (*time warping*), pour laquelle

7. Dans la fig. 1.14 (panneau en bas), la valeur initiale de l'accent HL- est plus élevée que celle du pic accentuel. En suivant l'hypothèse du chevauchement gestuel, on peut supposer que lorsque les deux mouvements sont très proches, la valeur maximale de HL- n'est pas atteinte. Une explication alternative est que la valeur mélodique du pic accentuel dans la fig. 1.14 (panneau en haut) est intermédiaire entre celle du début du plateau accentuel et de HL- dans la fig. 1.14 (en bas).

la localisation de chaque point du contour est le résultat de la combinaison pondérée de ces facteurs. Il faut noter que le choix de regarder le pic nucléaire n'a pas, pour le modèle de van Santen, une valeur théorique, dans la mesure où l'unité d'analyse de base est la courbe accentuelle définie comme “*a local pitch excursion that corresponds to an accented syllable*” (“une excursion mélodique locale qui correspond à la syllabe accentuée” : van Santen & M obius (1999), p. 270). Par conséquent, chaque point de la courbe (le début, la fin, le point du milieu, etc.) est soumis à cette déformation temporelle. La déformation est non-linéaire, de sorte que le poids des facteurs de *warping* dépend de la position temporelle du point de f_0 . Par exemple, le poids de l'attaque syllabique est plus fort pour la détermination temporelle du point initial de l'excursion tonale que pour celle du point final. Dans van Santen & Hirschberg (1994), les auteurs rendent compte des résultats de la modélisation de l'alignement du pic accentuel (c'est-à-dire, du point de f_0 à la fin de la courbe accentuelle). Dans ce papier, ils analysent 2000 phrases déclaratives lues par un locuteur féminin. Les phrases sont du type *Now I know C₀.V.X* (“Maintenant, je connais C₀.V.X ”), où C₀.V.X est le mot cible, susceptible d'avoir un accent H* ; C₀ indique la position de l'attaque syllabique (qui peut être occupée par 0 ou plusieurs consonnes) ; V est la position du nucleus vocalique (composé d'une voyelle ou d'une diphtongue), C₀ est la coda consonantique et X les syllabes atones suivantes (si présentes). Les positions C₀ et C₀ sont composées de consonnes sourdes/voisées et de consonnes non sonorantes/sonorantes. L'alignement du pic est donc mesuré de façon différente : comme distance absolue de l'attaque syllabique ou vocalique ; et comme proportion relative à la durée de la syllabe accentuée ou de la voyelle. Les auteurs rendent compte d'une forte variabilité inter-classe lorsque le pic est calculé selon ces mesures. Au contraire, ils trouvent que la localisation temporelle du pic (T_{peak}) peut être modélisée par l'équation linéaire suivante :

$$\mathbf{T}_{peak}(D_{C0}, D_{s-rhyme} ; C_0, v, C_c) = \alpha_{C0,Cc} \times \mathbf{D}_{C0} + \beta_{C0,Cc} \times \mathbf{D}_{s-rhyme} + \mu_{C0,Cc} \quad (1.2)$$

Dans l'équation (1.2), l'alignement de H^* est le résultat de la combinaison pondérée de la durée de l'attaque syllabique (\mathbf{D}_{C0}), de la rime ($\mathbf{D}_{s-rhyme}$) et de la classe consonantique de l'attaque et de la coda. Les coefficients α , β et μ sont obtenus par l'analyse de régression multiple. Il faut noter aussi que la notion de "rime" employée par van Santen & Hirschberg (1994) est différente de celle standard, utilisée par Silverman & Pierrehumbert (1990), car elle comprend tous les segments sonorants de l'attaque syllabique, la voyelle et tous les segments sonorants de la coda. Cette analyse de régression explique environ 80% de la variance.

Le modèle de van Santen & Hirschberg (1994) a été employé successivement par Prieto *et al.* (1995) pour expliquer l'alignement de l'accent prénucléaire H^* en espagnol mexicain. Dans cette étude, les mots cibles sont composés par des syllabes toniques ouvertes, dont la composition segmentale est contrôlée (l'attaque syllabique étant toujours une consonne sonorante). Plusieurs facteurs prosodiques sont manipulés, tels que la distance de l'accent par rapport aux frontières prosodiques (frontière droite de syntagme intonatif et intermédiaire ; position de la syllabe accentuée à l'intérieur du syntagme intermédiaire), la distance par rapport à la frontière de mot ainsi que le voisinage à une autre syllabe tonique. Prieto *et al.* (1995) ont montré que l'alignement du pic accentuel est fonction de la durée des segments à l'intérieur de la syllabe tonique, et que le modèle est amélioré lorsque les facteurs prosodiques sont inclus (la variance expliquée est entre 66% et 85% selon la condition prosodique).

La réalisation des cibles tonales peut également dépendre de la longueur phonologique et de la structure syllabique. La plupart des études à ce sujet cherchent à vérifier la validité de l'hypothèse de l'ancrage segmental. Parmi ces travaux, nous rappelons d'abord l'étude de Ladd *et al.* (2000). Ladd *et al.* (2000) ont montré qu'en hollandais, l'alignement du pic des accents prénucléaires est plus avancé dans les syllabes qui contiennent des voyelles longues (ex. dans /e :/ de *MElige*) que dans celles qui contiennent des voyelles courtes (dans /ε/ de *RENnende*). Cet effet est indépendant de la durée acoustique réelle des voyelles : on observe également une différence d'alignement dans les voyelles hautes (/i/ vs. /I/), qui, en hollandais, ont normalement la même durée et diffèrent uniquement pour la qualité de la voyelle. La variabilité d'alignement de la cible H peut être expliquée, selon

les chercheurs, en termes d' "ancrage" de la cible H à la frontière droite de la syllabe, puisque en hollandais les voyelles longues sont toujours en syllabe ouverte tandis que les voyelles brèves peuvent être en syllabe fermée ou ouverte (la consonne suivante étant ambisyllabique). La cible H est donc alignée avec la frontière droite de la voyelle tonique quand la syllabe contient une voyelle longue d'un point de vue phonologique, car la syllabe est ouverte; et avec la consonne suivante quand la syllabe contient une voyelle brève, car la syllabe est considérée comme fermée. Cette hypothèse est proposée comme une version révisée de l'hypothèse d'ancrage segmental qui tient en compte des contraintes phonologiques spécifiques à la langue considérée pour mesurer l'alignement tonal (*revised structural explanation*, "explication structurelle révisée").

Au contraire, d'autres études dans plusieurs langues ont montré que le ton H n'est pas synchronisé avec la fin de la syllabe, ceci même lorsque les effets de structure syllabique sont pris en compte (Gili Fivela & Savino (2003) pour les variétés d'italien parlé à Pisa et Bari, Schepman *et al.* (2006) pour le hollandais, Welby & Loevenbruck (2005) pour le français, Prieto & Torreira (à paraître) pour l'espagnol castillan, *inter alia*).

L'alignement tonal est influencé par la structure syllabique aussi en italien napolitain (D'Imperio, 2000). En italien, contrairement au hollandais, la durée de la voyelle n'a pas de valeur contrastive, tandis que la durée consonantique est phonologique. Cependant, il a été trouvé dans plusieurs études que la durée de la voyelle varie selon la structure de la syllabe et sa position à l'intérieur du mot : la voyelle est plus longue dans les syllabes ouvertes et pénultièmes que dans les autres positions (Marotta, 1985; Farnetani & Kori, 1991; D'Imperio & Rosenthal, 1999). En particulier, D'Imperio & Rosenthal (1999) rendent compte des résultats d'une étude menée en production, dans laquelle ils trouvent que la durée de la voyelle tonique est plus longue dans les syllabes ouvertes et pénultièmes que dans les syllabes fermées et pénultièmes ou encore dans les syllabes (ouvertes et fermées) ante-pénultièmes. Les chercheurs interprètent leurs données en termes d'interactions de contraintes qui s'inspirent de la Théorie de l'Optimalité (Prince & Smolensky, 1993). Selon D'Imperio & Rosenthal (1999), l'allongement de la voyelle a en fait une raison phonologique, car seulement les voyelles dans les syllabes toniques pénultièmes et ouvertes peuvent être bimoraïques (le pied trochaïque est en fait le pied métrique typique de

l'italien). Dans le cas des syllabes fermées, la première mora est occupée par la voyelle et la deuxième par la coda consonantique suivante ; par conséquent, la voyelle n'est pas allongée. La voyelle tonique n'est pas allongée lorsque le mot est proparoxyton, car dans ce cas, la deuxième mora est occupée par la syllabe suivante (le pied devient donc dissyllabique).

Les effets de la structure syllabique sur la durée des voyelles sont aussi relevés par D'Imperio (2000). En particulier, l'étude de production de D'Imperio (2000) ne se limite pas à observer les effets de modalité intonative sur la réalisation de l'accent nucléaire, mais aussi à vérifier l'existence de l'impact de la structure syllabique et de l'identité segmentale de l'attaque consonantique de la syllabe post-tonique sur l'alignement tonal. Pour cela, elle utilise des mots cibles ayant la syllabe tonique fermée ou ouverte. En accord avec l'étude de D'Imperio & Rosenthal (1999), la durée de la voyelle est plus longue dans les syllabes ouvertes que dans celles ayant une coda consonantique. Au niveau tonal, cette différence est accompagnée d'une variabilité systématique d'alignement du pic accentuel : l'alignement de la cible H relativement à la fin de la voyelle accentuée est plus retardé dans les syllabes fermées que dans les syllabes ouvertes.

La variabilité d'alignement du pic accentuel a eu des interprétations différentes. Par exemple, il a été proposé que l'alignement des pics ne soit pas synchronisé aux frontières segmentales, car il "chercherait" un ancrage dans une région temporelle plus large autour de la frontière même. Donc, il a été suggéré de substituer la notion de *anchoring* avec celle de *anchorage* (une région temporelle plus large qui se trouve autour de la frontière même, Welby & Loevenbruck (2005)) ou de considérer l'hypothèse de l'ancrage segmental comme une simple régularité phonétique, qui n'implique pas de synchronisation avec les bornes métriques ou prosodiques (Ladd, 2008). D'autres chercheurs ont expliqué la différence de comportement des cibles L et H par une différence de coordination articulaire des gestes tonals en position d'attaque et de coda (Prieto & Torreira, à paraître) ou même par l'idée que la nature de l'ancrage segmental des cibles tonales est articulaire (D'Imperio *et al.*, à paraître).

Une question que nous nous posons au regard de la variabilité acoustique et de savoir si elle est prise en compte pour la détermination des catégories phonologiques et de leur sens. En particulier, un problème encore irrésolu est de savoir s'il existe une interaction

entre la f_0 et le plan segmental pour la perception de l'intonation. Les réponses de la littérature sont contradictoires. Par exemple, selon 't Hart *et al.* (1990) et Hirst *et al.* (2000), les perturbations segmentales sur la f_0 sont le résultat de mécanismes physiologiques et universels, et donc elles ne sont pas pertinentes dans la perception de l'intonation. Par conséquent, dans leurs techniques de modélisation de la f_0 , la courbe de l'intonation est obtenue par la correction drastique des effets microprosodiques. De l'autre côté, certains chercheurs ont montré que les effets segmentaux sont pris en compte par les auditeurs, et que les auditeurs exploitent cette information selon leur langue maternelle. Par exemple Silverman (1987) a montré que la perception de la proéminence accentuelle des syllabes dépend partiellement des différences intrinsèques de f_0 des voyelles accentuées et, de l'autre côté, que la perception des contrastes segmentaux tels que /ba/ vs. /pa/ dépendent du type du contour dans lequel ils sont insérés (contour montant vs. descendant). A cause de cette interaction "bi-directionnelle", en synthèse de la parole, il implémente le modèle de l'intonation de Pierrehumbert (1980) en tenant compte des effets de la microprosodie sur la forme de la courbe de f_0 . L'idée que la relation entre segments et f_0 puisse dépendre de la langue considérée est soutenue par Kohler (1990). Kohler (1990) examine les effets de la f_0 sur la perception du voisement en anglais et allemand. Dans son étude, il crée un continuum de f_0 dans lequel il varie la durée de la voyelle accentuée et la direction du contour de f_0 dans le même segment. Il demande donc à des auditeurs anglais et allemands de juger le voisement de la consonne occlusive apicale qui suit la voyelle. Ses résultats montrent que la pourcentage d'identification de la consonne voisée augmente lorsque la configuration de f_0 est plate, pour les Anglais et pour les Allemands. De plus, il note une différence significative entre les deux groupes d'auditeurs, car la frontière catégorielle entre /t/ et /d/ est plus avancée pour les Anglais.

Il faut aussi noter que d'autres études ont montré que la f_0 n'est souvent pas le seul paramètre acoustique à véhiculer le sens intonatif. Par exemple, Scherer *et al.* (1984) ont trouvé que l'interprétation des contours intonatifs est beaucoup plus ardue quand la parole est filtrée par un filtre à bande basse, qui élimine l'information segmentale (Scherer *et al.*, 1984). Niebuhr (à paraître) a montré que la perception des accents montants ayant un pic avancé/retardé en allemand est influencée par l'aspiration de la consonne /t/

synchronisée, dans la mesure où sa présence modifie la signification associée aux deux accents (respectivement, “établi” et “nouveau”).

L’identité des segments de la syllabe accentuée joue aussi un rôle pour l’identification intonative en hollandais (Rietveld & Gussenhoven, 1995). L’étude de Rietveld & Gussenhoven (1995) est très intéressante, car les résultats semblent en désaccord avec la littérature sur le napolitain. Le but principal des deux chercheurs est celui d’améliorer la synthèse de l’intonation en hollandais. Ils remarquent qu’en synthèse de la parole, la localisation des cibles tonales dépend uniquement de la durée de la voyelle. Cependant, lorsque l’identité segmentale du texte est changée, le même contour intonatif est perçu de façon différente. Ils émettent donc l’hypothèse que l’identité segmentale de la syllabe accentuée puisse influencer l’identification intonative. Cette question est abordée par l’exploration des effets segmentaux sur l’identification des deux contours mélodiques, le “chapeau plat” (*flat hat*) et le “chapeau plat abaissé” (*downstepped flat hat*). Les deux contours se présentent de la façon suivante : le chapeau plat (non-abaissé) est réalisé comme une montée de f_0 suivie d’un plateau et d’une descente (analysable, en termes autosegmentaux, comme %L H* H*L L%), tandis que le chapeau plat abaissé présente une descente de f_0 plus retardée dans le temps (en termes autosegmentaux, due à l’abaissement du deuxième pic : %L H* !H*L L%). La distinction entre les deux contours est considérée comme catégorielle car elle ramène à une différence de sens claire, mais, comme il est admis par les deux auteurs, difficile à définir par des étiquettes appropriées : le chapeau plat abaissé semble véhiculer une idée de “finalité, “contribution définitive au discours” plus forte que le chapeau plat non-abaissé. Les stimuli pour l’expérience de perception sont donc obtenus par la manipulation de l’alignement temporel de la descente de f_0 , l’alignement étant déplacé en 10 pas de 20 ms. De plus, l’identité segmentale de la syllabe accentuée est manipulée : l’attaque syllabique est composée d’aucune ou de plusieurs consonnes ; l’attaque et la coda sont voisées et non-voisées. Les auditeurs doivent identifier chaque stimulus comme “calme, avec une mélodie basse” (*quiet, low-pitched*) ou comme “plus emphatique, avec une mélodie haute” (*more emphatic, high-pitched*). Il faut remarquer ici que le choix des étiquettes est assez problématique, car il oblige les auditeurs à porter leur attention sur un paramètre acoustique spécifique (la f_0) plutôt que sur le sens linguistique véhiculé par

les stimuli. Cela reflète probablement la difficulté à attribuer des étiquettes linguistiques précises au sens des deux contours. Les réponses des auditeurs sont calculées en termes de proportion de réponses “chapeau plat abaissé” et successivement évaluées en termes de position du “Point de Egalité Subjective” (*Point of Subjective Equality*, PSE). L’hypothèse nulle est que le PSE, défini comme “le moment psychologique d’occurrence d’une syllabe” (“*the psychological moment of occurrence of a syllable*”, p. 378) est fixe indépendamment de l’identité segmentale de la syllabe. Les résultats montrent cependant un effet très fort de la durée de l’attaque, du voisement de la coda et de la durée de la portion voisée de l’attaque (la variance expliquée est de 89%). En particulier, la localisation perceptive de H* est déplacée vers la gauche lorsque l’attaque contient des consonnes sonorantes; de façon analogue, la présence de consonnes sonorantes dans la coda “attire” la cible vers la droite. Les résultats sont, selon les auteurs, en accord avec les découvertes en acoustique de van Santen & Hirschberg (1994) : la durée et la classe consonantique de l’attaque et de la coda syllabique influencent l’alignement tonal.

Variabilité segmentale et réalisation de l'accent mélodique

2.1 Problématique

L'idée que la microprosodie influence la réalisation des cibles acoustiques est assez controversée dans les études sur l'intonation. Dans ces dernières années, plusieurs travaux dans le cadre de la Théorie Autosegmentale-Métrique ont souligné la stabilité d'alignement des cibles tonales, indépendamment du matériel segmental qui leur est associé. En particulier, selon l'hypothèse de l'ancrage segmental, l'alignement du début et de la fin des accents mélodiques montants et descendants est défini en fonction de frontières spécifiques dans la chaîne segmentale (Arvaniti *et al.* 1998; Ladd *et al.* 1999, *inter alia*). Les études qui soutiennent cette hypothèse rendent compte d'une coordination temporelle très étroite entre les cibles acoustiques et les frontières segmentales. Par exemple, Arvaniti & Ladd (1995) ont trouvé qu'en grec le début de la montée des accents prénucléaires est aligné à environ 5 ms avant l'attaque syllabique. De même, Arvaniti *et al.* (1998) affirment que la fin de la montée est alignée à environ 10 ms de l'attaque de la voyelle dans la syllabe post-accentuelle, ceci même indépendamment de l'identité segmentale dans la syllabe accentuée. De résultats similaires ont été trouvés aussi par Ladd *et al.* (1999). Dans leur

étude, Ladd *et al.* (1999) ont examiné l’alignement des accents prénucléaires en anglais dans de différentes conditions de vitesse d’élocution. Ils ont découvert que le début et la fin de l’accent prénucléaire montant sont ancrés respectivement au début de la syllabe tonique et au début de la voyelle post-tonique, indépendamment de variations de durée dues à une élocution rapide/normale/douce. Au contraire, leur étude montre que lorsque la vitesse d’élocution augmente, la durée segmentale diminue et l’ajustement des cibles tonales aux frontières segmentales provoque une augmentation de l’inclinaison de la pente et un raccourci du temps de la montée.

Il faut aussi noter que l’hypothèse de l’ancrage segmental a des conséquences importantes au niveau phonologique. En fait, la stabilité des points d’inflexion de f_0 a été utilisée pour soutenir l’idée, inhérente à la Théorie Métrique-Autosegmentale, que les primitifs de l’intonation sont les tons statiques, dont la réalisation en tant que cibles tonales est définie uniquement selon deux dimensions phonétiques, leur alignement et leur valeur mélodique. Au contraire, la variabilité de réalisation de la pente de f_0 et du temps de la montée a été considérée comme une preuve empirique contre tous modèles qui posent comme unité de description de l’intonation les propriétés dynamiques de la courbe de f_0 (tels que l’école Britannique ; voir Crystal (1969), *inter alia*).

L’hypothèse de l’ancrage segmental est soutenue dans plusieurs langues et par des études dans des cadres théoriques différents (Caspers & van Heuven (1993) pour le hollandais, Prieto *et al.* (1995) pour l’espagnol mexicain, Xu (1998) pour le chinois mandarin, *inter alia*). Cependant, dans ces travaux il apparaît souvent une dichotomie entre le comportement de la cible L et de la cible H. En fait, l’alignement de la cible L semble être synchronisé de façon très stable au début de l’attaque vocalique, tandis que la cible H est soumise à plusieurs facteurs de variabilité d’ordre phonologique et phonétique.

Par exemple, Silverman & Pierrehumbert (1990) ont trouvé que l’alignement du pic mélodique est fonction de la durée de la rime syllabique et que le degré d’allongement de la rime varie selon le facteur considéré (vitesse d’élocution ou effets du contexte prosodique à droite, tels que le *stress clash* ou le voisinage à la frontière de mot). Selon van Santen & Hirschberg (1994), l’alignement du pic dépend de la durée et de la classe consonantique de l’attaque syllabique et de la rime. La rime, comme définie par van Santen

& Hirschberg (1994), comprend tous les segments sonorants de l'attaque syllabique, la voyelle et tous les segments sonorants de la coda. Prieto *et al.* (1995) ont trouvé qu'en espagnol mexicain l'alignement du pic est le résultat de la combinaison pondérée de la durée de l'attaque consonantique et de la voyelle accentuée ainsi que de plusieurs facteurs prosodiques (voisinage aux frontières prosodiques, *stress clash*, voisinage à la frontière de mot). Les résultats de van Santen & Hirschberg (1994) et de Prieto *et al.* (1995) sont naturellement contraires à celui de Silverman & Pierrehumbert (1990), dans la mesure où l'attaque syllabique acquiert un poids important dans la prédiction de l'alignement du pic. D'autres chercheurs ont proposé de substituer la notion d'ancrage avec celle d'"anchorage", c'est à dire d'une région dans laquelle les cibles tonales peuvent se réaliser (Welby & Loevenbruck, 2005).

Dans ce chapitre, nous allons examiner les effets de la variabilité segmentale sur la réalisation du contraste entre les accents nucléaires des questions (L*+H) et des affirmations (L+H*) dans l'italien de Naples. En particulier, nous avons manipulé la modalité intonative ainsi que la classe consonantique de la consonne à l'attaque de la syllabe accentuée, où on suppose que la cible L de l'accent nucléaire LH est attendue. En suivant l'hypothèse de l'ancrage segmental strict, nous supposons que l'alignement de la cible L ne sera pas modifié par l'identité segmentale de l'attaque syllabique. En particulier, notre hypothèse est que le L sera aligné en relation avec la frontière gauche de la syllabe, et que son alignement dépendra uniquement de la modalité intonative (D'Imperio, 2000).

L'hypothèse alternative est que l'alignement est influencé par l'identité segmentale dans la syllabe tonique. Bien que les études de Silverman & Pierrehumbert (1990) ou de van Santen & Hirschberg (1994) se soient intéressées uniquement aux variations d'alignement de la cible H, nous supposons que la cible L variera avec la durée de la consonne ou avec la classe consonantique (Welby & Loevenbruck, 2005; van Santen & Hirschberg, 1994) de façon similaire au pic accentuel.

Il faut noter que, bien que l'hypothèse de l'ancrage segmental s'adresse surtout à l'alignement tonal, elle implique également une stabilité de l'excursion de la f_0 (c'est-à-dire de la différence mélodique entre le L et le H). Si le L et le H sont des cibles indépendantes, les variations de durée segmentale dues à des contraintes de pression temporelle ne vont

pas influencer leur valeur mélodique. Dans notre étude, nous n'avons pas manipulé de façon explicite la vitesse d'élocution. Cependant, nous savons que la durée intrinsèque des segments dépend de leur classe phonologique d'appartenance (Di Cristo, 1985). Si l'hypothèse de l'ancrage segmental est confirmée, les valeurs de f_0 des cibles tonales ne devront pas être influencées par des variations de la durée segmentale.

La pente de f_0 de l'accent montant et descendant de la configuration LHL a été aussi examinée. En particulier, si l'hypothèse de l'ancrage segmental est vraie, la pente sera variable, et ses valeurs seront prévisibles à partir des valeurs d'alignement et de f_0 des cibles tonales. Par conséquent, la pente de la montée de f_0 sera plus raide avec l'obstruante (qui provoque une forte baisse locale des valeurs de f_0 , Di Cristo (1985), *inter alia*) que avec la consonne nasale (qui normalement ne devrait pas influencer, selon la littérature, la courbe de l'intonation, voir à nouveau Di Cristo (1985)), et dans les questions plus que dans les affirmations. La pente de la descente sera modifiée uniquement par la modalité intonative.

Les effets microprosodiques sur la réalisation des cibles tonales posent aussi des problèmes méthodologiques. Dans la plupart des travaux sur l'alignement tonal, les cibles L et H sont identifiées manuellement avec les valeurs minimales et maximales dans le contour de f_0 , visibles sous forme de coudes ou pics dans le tracé acoustique (Prieto *et al.*, 1995; Arvaniti *et al.*, 1998; Ladd *et al.*, 1999, 2000, *inter alia*). Cependant, la localisation des cibles peut être très difficile, de sorte que des critères acoustiques arbitraires doivent être adoptés. La détection de L dans les accents mélodiques montants (ou descendants) est particulièrement ardue. En fait, la transition entre le début et la fin de la montée (ou de la descente) est souvent très graduelle, et plusieurs valeurs minimales de f_0 peuvent être considérées comme des candidats possibles pour la localisation du L. De plus, la présence de perturbations segmentales peut masquer les cibles tonales (par exemple, en correspondance avec les consonnes sourdes) ou créer des coudes "additionnels" (comme les baisses locales des valeurs de f_0 en correspondance avec le début des consonnes nasales ou encore la baisse drastique de f_0 en présence d'une obstruante sourde). Les annotateurs peuvent décider de localiser le L quelques millisecondes avant ou après la perturbation, ou ils peuvent négliger la perturbation et identifier le L avec le minimum de f_0 au dé-

but de l'attaque consonantique. Dans d'autres travaux, de procédures automatiques ou semi-automatiques de modélisation de la f_0 ont été adoptées, car elles seraient plus objectives que la détection manuelle (Campione *et al.*, 1997; Hirst *et al.*, 2000; D'Imperio, 2000; Welby, 2003). Ces procédures estiment les cibles tonales par la correction/réduction automatique des perturbations segmentales et/ou par la stylisation du contour de f_0 . Par conséquent, elles devraient être moins sensibles aux erreurs de détection que la méthode manuelle. Bien que plusieurs méthodes de détection (semi-)automatique aient été proposées, il est encore très difficile de déterminer la méthode la "meilleure", c'est-à-dire la plus précise, fiable et indépendante de présupposés théoriques. Par exemple, la fiabilité de la méthode manuelle pour la localisation de la cible L a été soutenue par del Giudice *et al.* (2007) pour l'anglais américain. Dans cette étude, il a été montré que, lorsque les annotateurs humains adoptent les mêmes critères acoustiques de localisation de la cible L, ils sont plus cohérents que les procédures automatiques, et cela surtout lorsqu'ils savent où le L est attendu (par exemple, autour du début ou de la fin de la syllabe). Mais la cohérence est-elle un critère suffisant pour préférer une méthode à l'autre ?

Dans ce chapitre, nous allons comparer trois méthodes de localisation des cibles tonales, et nous allons discuter les possibles conséquences de choix méthodologiques sur l'interprétation phonologique des données. Cela est particulièrement intéressant pour l'italien napolitain, car dans cette variété l'alignement tonal de l'accent nucléaire LH est exploité pour signaler le contraste entre questions oui/non et affirmations. De plus, dans une étude perceptive, D'Imperio (2000) a montré que lorsque l'alignement du L est manipulé de façon indépendante, car son alignement avancé vs. retardé est employé pour l'identification des affirmations vs. questions. Ce résultat est une preuve de la valeur phonologique de l'alignement du L. Si la méthodologie n'a pas d'importance, on s'attendra à ce que les méthodes manuelles, automatiques et semi-automatiques soient capables de "capturer" les différences d'alignement du L. Si, au contraire, l'utilisation de méthodes différentes de détection influence les résultats d'alignement, l'interprétation des données pourrait être conditionnée par nos choix méthodologiques.

2.2 Méthodes

2.2.1 Corpus

Le corpus était constitué d’une série de phrases dans lesquelles six mots réels paroxytons étaient insérés en position finale dans la même phrase porteuse (*La mamma vuole vedere la X*, “La maman veut voir la X” ; X = mot-cible). Les phrases étaient manipulées dans l’identité segmentale et dans la modalité intonative (Tab.2.1). Le mode d’articulation de la consonne à l’attaque syllabique variait car la consonne pouvait être nasale (/n/), liquide (/r/) ou obstruante (/d/). Chaque consonne était suivie par une voyelle haute (/i/) ou basse (/a/).

Les phrases ont été lues comme affirmations à focalisation étroite et comme questions oui/non. Dans les deux modalités intonatives, l’accent nucléaire est réalisé comme un accent montant LH sur le mot-cible suivi par un accent de syntagme L-. Notamment, un accent nucléaire L+H* est attendu dans les affirmations, et un accent nucléaire L*+H est attendu dans les questions (D’Imperio, 2000, 2002a). De plus, un accent prénucléaire est attendu sur *La mamma*. L’ensemble des phrases utilisées dans le corpus est illustré dans la Table 2.1.

Int	Cons	Voy. /i/	Voy. /a/
A	Nas	La mamma vuole vedere [la Nina]	La mamma vuole vedere [la nana]
A	Liq	La mamma vuole vedere [la Rina]	La mamma vuole vedere [la rana]
A	Obs	La mamma vuole vedere [la Dina]	La mamma vuole vedere [la Dana]
Q	Nas	La mamma vuole vedere [la Nina] ?	La mamma vuole vedere [la nana] ?
Q	Liq	La mamma vuole vedere [la Rina] ?	La mamma vuole vedere [la rana] ?
Q	Obs	La mamma vuole vedere [la Dina] ?	La mamma vuole vedere [la Dana] ?

TABLE 2.1 – Phrases du corpus. Les crochets indiquent le domaine de la focalisation.

2.2.2 Enregistrement

Deux locuteurs ont été enregistrés, un homme (DD) et une femme (AS). Les sujets, âgés de 28 ans, sont nés et ont grandi à Naples. Les deux locuteurs étaient “naïfs” quant au but de notre étude, bien qu’habitues à être enregistrés pour des études en production de la parole.

Les phrases, transcrites sur des papiers A4, ont été présentées en ordre aléatoire parmi d’autres phrases de remplissage¹. Pour chaque locuteur, sept répétitions par énoncé ont été enregistrées. Nous avons demandé aux locuteurs de lire les phrases de la façon la plus naturelle possible et à une vitesse d’élocution jugée “normale” par eux mêmes. Nous n’avons pas donné d’instructions explicites à propos de l’intonation que les sujets devaient utiliser. Toutefois, afin d’obtenir l’intonation appropriée, les mots focalisés étaient soulignés et les phrases étaient insérées dans des dialogues très courts. Les enregistrements ont été faits à l’aide d’un lecteur MD et avec un microphone sur pied placé à une distance constante du sujet. Les signaux audio ont ensuite été numérisés et échantillonnés à 24kHz. Chaque phrase a été sauvegardée dans des fichiers audio séparés (.wav).

A partir d’un corpus de 168 phrases, 145 phrases ont été analysées. Les autres ont été éliminées pour plusieurs raisons (présence d’une région de voix craquée sur le mots-cible, intonation inappropriée). Chez DD, les affirmations à focalisation étroite retardée présentent un patron intonatif légèrement différent de celui attendu, car elles étaient réalisées avec une focalisation double, un focus additionnel étant produit sur *La mamma*. Cela ne devrait pas influencer l’objet de notre étude, car nous savons que le domaine temporel de la focalisation est assez étroit (voir Dohen (2006) pour un état de l’art).

2.2.3 Segmentation acoustique et détection de la f_0

La segmentation acoustique, la détection de la f_0 et des cibles tonales ont été effectuées grâce à l’utilisation du logiciel PRAAT (Boersma & Weenink, 2007).

Des critères traditionnels ont été adoptés (Peterson & Lehiste, 1960) afin de rendre

1. Les phrases de remplissage, répétées aussi sept fois, étaient constituées par la même phrase porteuse utilisée dans cette expérience et suivie par 5 mots propres paroxytons : Mina, Lina, Rina, Pina, Tina. Les phrases de remplissage et les phrases dans la Table 2.1 seront utilisées dans l’Expérience I décrite dans le chapitre 4.

la segmentation la plus homogène possible dans le corpus. Les mesures ont été obtenues à l'aide de l'inspection visuelle du spectrogramme et de la forme d'onde. Chaque phrase a été segmentée en mots, syllabes et segments. Toutes les étiquettes correspondant à des passages par zéro du signal. Le début de chaque phrase a été placé avant le premier pic d'amplitude dépassant le bruit de fond (Kvale, 1993) tandis que sa fin a été fixée à la dernière pulsation glottale de f_0 ². Les frontières gauche et droite des consonnes nasales ont été placées à la baisse du premier formant (F1) qui se présente à la transition entre la voyelle et la consonne nasale. Les frontières de la nasale étaient également signalées par une diminution soudaine de l'énergie du deuxième (F2) et du troisième (F3) formant. La réalisation de la liquide /r/ intervocalique était assez variable. Dans la majorité des cas, /r/ était produit comme un *tap* apical, composée de 3 régions distinctes : une première région vocalique, fusionnée avec la voyelle précédente (/a/ de "la") ; une région caractérisée d'un bruit de friction plus fort et une deuxième région vocalique, fusionnée avec la voyelle suivante (/i/ de "Rina" ou /a/ de "rana"). Le début et la fin du /r/ étaient signalés par un changement de la structure et de l'amplitude de la période vocalique, correspondant sur le spectrogramme à une variation soudaine de la courbe d'intensité jusqu'à la région plus fricativisée, où la courbe atteint son minimum. Les transitions formantiques (en particulier de F1 et de F2) ont été utilisées comme critère de segmentation secondaire pour la détermination des frontières consonantiques. Le début de la consonne obstruante /d/ était marqué surtout par la diminution soudaine de l'énergie de F2, accompagnée par une baisse soudaine de l'amplitude de la période et de l'intensité. La frontière entre la voyelle et la consonne a été placée à la fin de la dernière période vocalique complète identifiée dans la forme d'onde. Cette consonne était parfois réalisée sans une complète occlusion, de sorte qu'elle était caractérisée, au niveau acoustique, par un faible bruit de friction. Dans ce cas, l'obstruante était signalée par une baisse graduelle de F2 et de la courbe d'intensité, dont le point initial a été utilisé pour marquer le début de la consonne. La frontière entre la consonne et la voyelle suivante a été marquée à l'apparition de la première période vocalique complète. Les informations concernant les décisions de

2. Ce critère a été préféré à celui basé sur l'amplitude du signal, car la fin de l'énoncé était souvent caractérisée par une région de voix soufflée, ce qui rendait la localisation du dernier pic d'amplitude assez difficile.

segmentation pour chaque fichier son ont été stockées dans des fichiers de texte “TextGrid”. Un exemple de segmentation est donné dans la fig. 2.1³.

La f_0 a été calculée grâce à l’utilisation d’un script, qui repose sur un algorithme développé par Daniel Hirst et Céline De Looze (Laboratoire Parola et Langage, Aix-en-Provence). Dans cet algorithme, les mesures de f_0 sont prises pour chaque fichier son à intervalles de 10 ms. Dans un premier temps, les valeurs de f_0 pour chaque son de la parole sont repérées entre un seuil minimal (*floor*, “plancher”) de 75 Hz et un seuil maximal (*ceiling*, “plafond”) de 600 Hz. Ensuite, les valeurs correspondantes au 75ème et au 25ème percentile sont extraites : le minimum ($\min f_0$) et maximum ($\max f_0$) sont calculés, respectivement, à partir de la multiplication du 25ème percentile par la valeur constante de 1.5 et du 75ème percentile par 0.75. Les valeurs ainsi obtenues sont utilisées respectivement comme plancher et plafond pour une nouvelle détection de la f_0 sur le même fichier son. Les informations relatives aux valeurs fréquentielles et temporelles de la f_0 ont été contrôlées et donc stockées dans des objets “Pitch”.

3. Dans l’exemple, nous montrons aussi le type de segmentation effectué pour le mot prosodique *La mamma*, car il sera l’objet de nos attentions dans le chapitre 4.

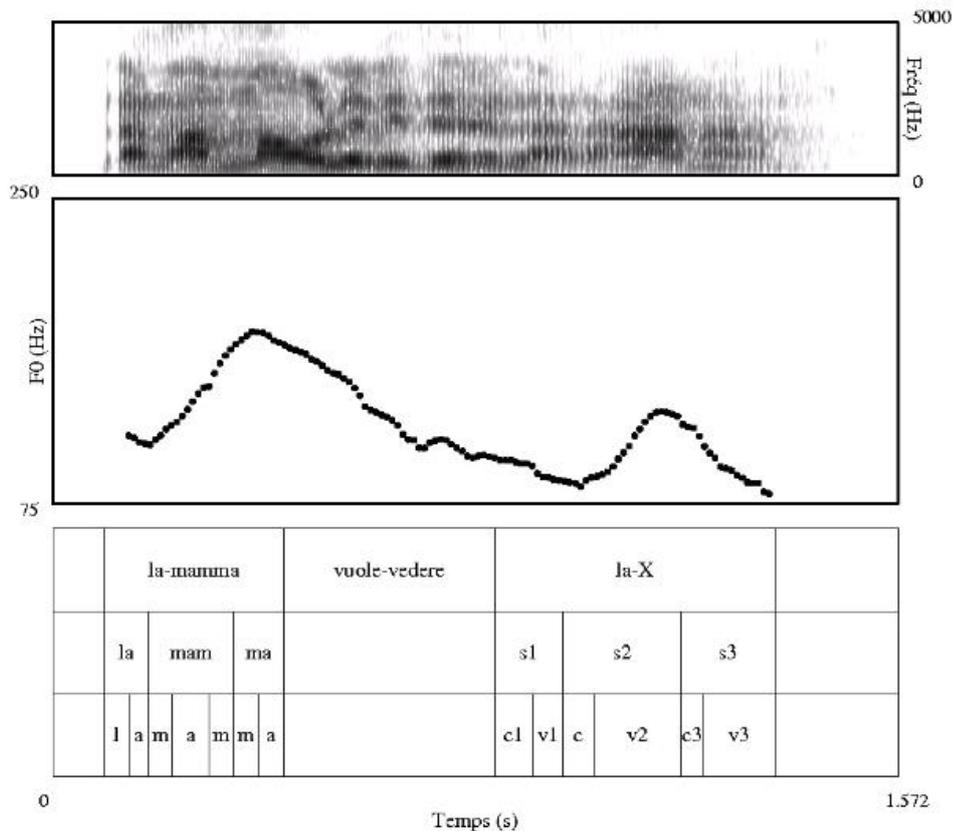


FIGURE 2.1 – Spectrogramme, courbe de f_0 et TextGrid pour la phrase *La mamma vuole vedere la rana?* (litt. “la maman veut voir la grenouille?”), prononcée par le locuteur DD. Dans le TextGrid, *la-X* dans le premier tier indique l’intervall de durée correspondant au mot prosodique; le deuxième et le troisième tiers représentent la segmentation en syllabes (*s*) et segments (*v* = voyelle; *c* = consonne).

2.2.4 Méthodes de détection des cibles tonales

La localisation des cibles tonales a été effectuée pour les questions oui/non ainsi que pour les affirmations à focalisation retardée. A l’intérieur de la configuration montante-descendante de f_0 , nous avons mesuré la valeur de f_0 (en Hz) pour les points correspondants à la cible basse L (L1) et à la cible haute H de l’accent nucléaire ainsi qu’à l’accent de syntagme L- (L2) suivant.

La localisation de L1 a été particulièrement difficile parce que la région près du début de la syllabe accentuée (où on s’attend à la réalisation de L1, D’Imperio (2000)) était souvent caractérisée par des perturbations segmentales dues au type de consonne en position d’attaque. Par exemple, dans la fig. 1.3, le début de la consonne /d/ de “Dana” est

marqué par une chute de la f_0 , suivie d'une montée soudaine jusqu'à la frontière avec la voyelle basse /a/. A la fin de la constriction consonantique, les valeurs de f_0 diminuent sensiblement et un creux de f_0 est créé dans la première portion de la voyelle.

Dans des cas problématiques comme celui qu'on vient de décrire, l'utilisation d'une méthode de détection des cibles qui soit subjective ou peu robuste pourrait biaiser fortement les résultats relatifs à l'alignement tonal. Nous savons que les choix de localisation peuvent être très différents. La plupart des chercheurs s'appuient sur la détection manuelle. Dans ces cas, les critères de localisation de la cible, bien que consistants, peuvent différer selon l'étude spécifique : on peut localiser la cible basse au début de la montée de f_0 , au minimum local de f_0 ou, en cas de perturbations segmentales, quelques millisecondes avant ou après la perturbation. Dans d'autres cas, des procédures algorithmiques d'extraction de l'information tonale sont employées. Cependant, il y a encore très peu d'études qui ont comparé ces méthodes de façon systématique (del Giudice *et al.*, 2007).

Dans notre expérience, nous allons comparer les résultats de la méthode manuelle avec ceux provenant de deux procédures automatiques pour la détection de L1 : la méthode des moindres carrés (*Least Square Fitting*, LSF) déjà utilisée, entre autres, par D'Imperio (2000) et Welby (2003) ; et l'algorithme MOMEL, développé par Daniel Hirst et collègues (Hirst & Espesser 1993; Hirst *et al.* 2000; Campione *et al.* 1997). Nous avons décidé de limiter la comparaison à L1. En fait, à cause de la variabilité de la consonne d'attaque syllabique (autour du quel L1 est réalisé), L1 était la cible la plus difficile à localiser, et donc aussi celle pour laquelle les différences entre les trois méthodes auraient pu mieux ressortir.

L'étiquetage manuel de L1 a été effectué par l'auteur. Le protocole suivi était basé sur des critères purement acoustiques inspirés aux critères de détection tonale déjà employés dans la littérature (Prieto *et al.*, 1995; Arvaniti *et al.*, 1998; Ladd *et al.*, 1999, 2000; del Giudice *et al.*, 2007, *inter alia*) :

- Localisation de L1 dans la région du début de la syllabe accentuée, où la cible est attendue ;
- Étiquetage du minimum de f_0 avant le début de la montée comme L1 ;
- Si ce minimum est dû à une perturbation microprosodique locale, prendre la valeur

de f_0 à 10 ms avant le minimum (car nous supposons que ce type de perturbation ait des effets strictement locaux) ;

- Si un plateau bas de f_0 est présent, prendre le point final du plateau, qui est placé avant le début de la montée de f_0 ;⁴

Pour chaque fichier, la forme d’onde, le spectrogramme et le contour de f_0 de l’énoncé étaient visibles. La décision concernant la localisation de la cible L1 a été stockée dans un niveau d’annotation qui allait s’ajouter aux niveaux de segmentation déjà existants (fig. 2.1) dans les TextGrids correspondants.

La méthode des moindres carrés (LSF) a été conçue par Mary Beckman et utilisée, entre autres, par D’Imperio (2000) en italien napolitain.⁵ Nous précisons qu’ici nous n’utilisons pas la version R du script de Mary Beckman, mais une version équivalente de l’algorithme adapté à l’environnement Cygwin par Robert Espesser (Laboratoire Parole et Langage, Aix-en-Provence). Dans cet algorithme, la position du creux est à l’intersection des deux lignes droites, qui représentent le résultat de deux régressions linéaires ajustées dans une région temporelle pré-spécifiée du contour de la f_0 de chaque phrase (une pour les points de f_0 à gauche du coude et l’autre pour les points à sa droite). Le point d’intersection sélectionné en tant que “coude” est celui du modèle dont l’écart total est le plus petit (fig. 2.2). Le succès de l’estimation dépend aussi de la région du contour sélectionnée, dans laquelle l’algorithme cherche le coude. Dans notre expérience, nous avons choisi la portion de f_0 du début du mot prosodique jusqu’au pic de f_0 . Ce choix est motivé par des présupposés théoriques, car plusieurs études ont montré qu’en napolitain la cible L tend être réalisée autour de l’attaque consonantique et ceci, indépendamment de la durée du mot dans lequel il est réalisé (D’Imperio, 2000; Petrone & Ladd, 2007). De plus, la fenêtre d’analyse était suffisamment large (environ 300 ms) pour minimiser l’impact de

4. Nous considérons le plateau bas comme l’ensemble des points de f_0 pour lequel la valeur de chaque point ne diffère pas de +/- 2 Hz relativement au point précédent. Ce critère a été déjà employé par D’Imperio (2000) pour les plateaux hauts en napolitain.

5. D’Imperio (2000) appelle cette technique *two-line regression method* (“méthode des deux lignes de régression”). Bien que cette appellation soit plus appropriée, dans ce travail nous accueillons le nom “LSF” introduit par del Giudice *et al.* (2007). En fait, il faut noter que l’expression “méthode des moindres carrés” désigne, en général, une méthode d’ajustement des paramètres d’un modèle (intercepte, pente) pour trouver la fonction qui reproduit le mieux les données expérimentales, c’est-à-dire qui minimise la somme quadratique des écarts entre données et le modèle. Cette méthode est sous-jacente non seulement à la *two-line regression method*, mais aussi à d’autres algorithmes d’approximation du contour intonatif, tels que MOMEL.

variations microprosodiques sur l'ajustement du modèle.

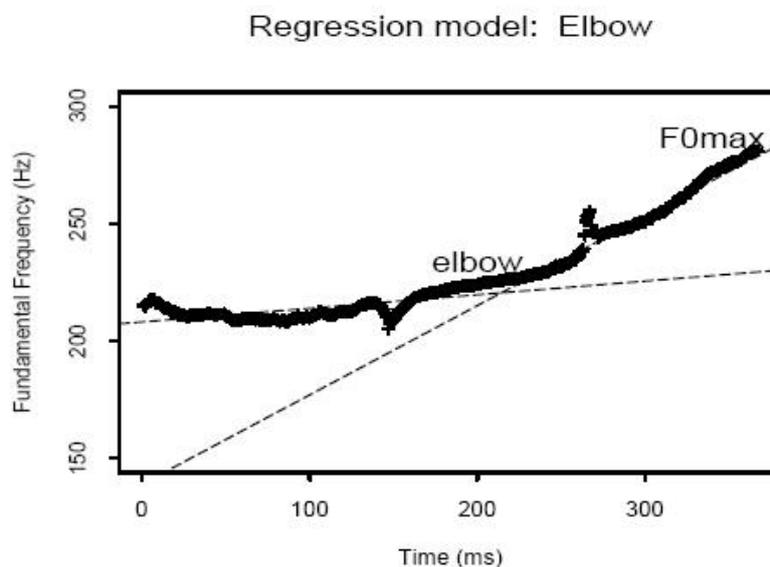


FIGURE 2.2 – Courbe de f_0 et lignes de régression estimées par le LSF. Le coude de f_0 est au point d'intersection des deux lignes (*elbow*). Tiré de D'Imperio (2000), p. 95.

L'algorithme MOMEL repose sur l'hypothèse selon laquelle le contour de f_0 est le résultat de la superposition de deux phénomènes distincts et indépendants, la macroprosodie (qui caractérise le choix intonatif d'élocution) et la microprosodie (qui correspond aux variations mélodiques à court terme propres aux segments). Le contour intonatif est obtenu par un lissage de la courbe de f_0 qui vise à filtrer de façon drastique les effets segmentaux. Le contour est représenté comme une séquence de cibles tonales liées par des arcs de parabole. Dans une première étape, l'algorithme supprime tous les points de la courbe initiale dont la valeur mélodique est située à plus de 5% au-dessus de celle des points avoisinants, car ils seraient dûs à des simples perturbations segmentales telles que le début du voisement. Une fenêtre glissante de 300 ms parcourt le signal acoustique de gauche à droite. La courbe de f_0 extraite sur chaque fenêtre est estimée par un polynôme de second degré. Dans cette étape, les valeurs de f_0 situées à plus de 5% au-dessous du polynôme sont aussi considérées comme le résultat d'effets segmentaux et donc supprimées. Dans chaque fenêtre, les points qui correspondent aux extrémités minimales ou maximales de la parabole (c'est-à-dire aux passages par zéro de la tangente) sont extraits

comme candidats possibles pour la localisation des cibles (fig. 2.3). Pour ce faire, l'espace temporel est partitionné de manière à isoler chacun des lieux où se regroupent les points d'inflexion. Sur chaque partition sont alors calculés la moyenne des candidats, et l'écart type des réalisations (écart type en temps et en fréquence). Les valeurs éloignées des valeurs moyennes d'une distance supérieure à l'écart type (soit en temps, soit en fréquence) sont supprimées, et la moyenne est alors recalculée sur les points restants. La cible finale correspond donc à cette moyenne en temps et en fréquence (fig. 2.4). Nous renvoyons à Hirst & Espesser (1993) pour plus de détails.

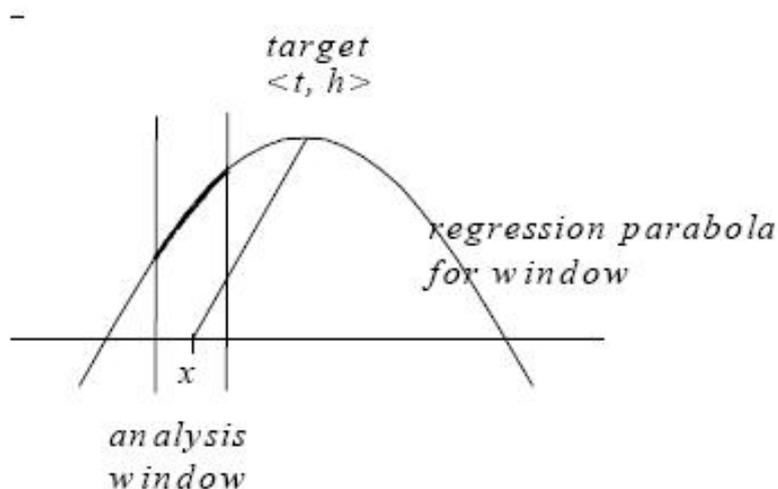


FIGURE 2.3 – Cible locale calculée pour l'instant X . La cible (*target*) est définie par sa localisation temporelle (t) et par sa valeur mélodique (h). Tiré de Campione *et al.* (1997), p. 23.

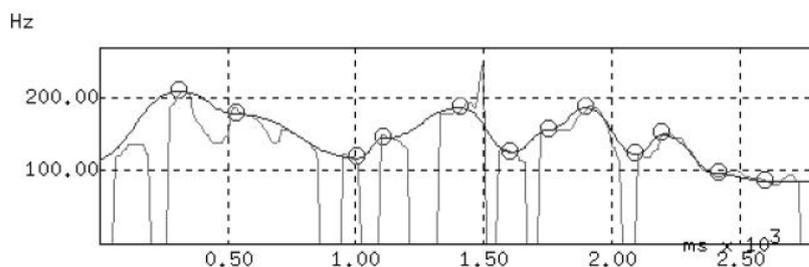


FIGURE 2.4 – Courbe de f_0 , courbe quadratique par morceaux estimée (ligne continue) et cibles (cercles) trouvées par MOMEL. Tiré de Campione *et al.* (1997), p. 24.

Bien que nous ne sachions pas où la cible L1 va être réalisée exactement, nous nous

attendons à ce que l’alignement de cette cible soit plus retardé dans les questions que dans les affirmations. En fait, D’Imperio (2000) a montré que dans l’italien de Naples, lorsque la cible L “perçue” a un alignement acoustique avancé relativement à la syllabe, les auditeurs identifient un nombre élevé d’“affirmations” ; lorsque la cible “perçue” montre un alignement plus retardé, les auditeurs identifient un nombre élevé de “questions”. Puisque nous supposons qu’il y a une correspondance entre les cibles acoustiques et les cibles perçues, nous employons les découvertes de D’Imperio (2000) comme critère de fiabilité pour la localisation de L1 dans nos données (voir aussi le par. 2.1).

2.2.5 Analyse statistique

L’analyse statistique, faite à l’aide de l’environnement de travail R (R Development Core Team, 2008), a inclus une série de régressions multiples à effets mixtes et d’ANOVAs effectués séparément pour chacun des locuteurs.

Le modèle à effets mixtes, ou plus simplement “modèle mixte” (Pinheiro & Bates, 2000) incorpore des facteurs à effets fixes et à effets aléatoires. Généralement, un facteur est à effets “fixes” si l’objectif de l’expérimentateur est de comparer les niveaux de la variable indépendante inclus dans son étude et il est à effets “aléatoires” si ces niveaux sont considérés comme échantillonnés au hasard dans une population plus large que les données elles-mêmes. Le facteur fixe est donc reproductible tandis que l’aléatoire ne l’est pas, car la réplication de l’expérience impliquerait la sélection de nouveaux niveaux du facteur (Baayen, à paraître; Baayen *et al.*, à paraître). En pratique, toutefois, la distinction entre les deux types de facteurs n’est pas souvent si stricte, car le fait de considérer un facteur à effets fixes ou aléatoires dépend aussi du type d’inférence que l’expérimentateur veut faire pour les niveaux du facteur considéré (Karpinski, 2007).

Dans nos modèles statistiques, les facteurs fixes étaient deux : la classe d’appartenance de la consonne en position d’attaque syllabique (nasale vs. liquide vs. obstruante), et la modalité intonative (affirmations à focalisation étroite vs. questions oui/non). Leur effets ont été testés sur la durée segmentale et sur certaines propriétés de l’accent nucléaire réalisé dans le mot-cible (alignement tonal, valeur mélodique et pente de f_0). Nous désignons les trois niveaux de la classe consonantique, respectivement : **Nas/Liq/Obs** ; et les deux

niveaux de la modalité intonative, respectivement : **A/Q**.

De plus, on se souviendra que dans notre corpus la voyelle accentuée du mot-cible pouvait être haute ou basse. Ici cependant nous n’avons pas cherché à tester les effets de la hauteur vocalique, car cela n’était pas le but de notre expérience. Or, nous avons supposé, en accord avec Clark (1973), que les mots utilisés dans notre expérience avaient été choisis au hasard en tant que échantillons représentatifs de l’ensemble de la population des mots. Par conséquent, un facteur aléatoire “mot” à six niveaux, correspondant aux six mots du corpus, a été introduit dans notre modèle. Ceci nous a permis de doubler le nombre d’observations pour chaque classe consonantique (car pour chaque classe, nous avons regroupé les résultats obtenus pour les mots présentant la voyelle haute et basse) tout en neutralisant des effets possibles de hauteur vocalique sur les variables dépendantes. Le terme aléatoire a été enlevé du modèle si la variance estimée de cette composante était près de zéro (ce qui signifie que la variance de la variable dépendante ne dépend pas de la variabilité parmi les mots utilisés).

Il faut noter également que, en général, les analyses de régression offrent beaucoup plus d’avantages que l’ANOVA “omnibus” lorsque la variable dépendante est continue et la variable indépendante est continue ou catégorielle (mais incluant plus de deux niveaux). Dans le premier cas, la régression nous restitue la relation d’ordre entre la variable dépendante et celle indépendante (par exemple, une corrélation positive ou négative ; linéaire ou quadratique). Dans le deuxième, elle nous permet d’avoir des informations sur comment chaque niveau d’un facteur influence la variable dépendante, et cela grâce à une comparaison par paire entre le niveau de référence et les autres niveaux d’un même facteur⁶. Dans notre modèle, le niveau de référence de la classe consonantique était **Nas**, qui a été donc comparé à **Liq** (contraste 1) et à **Obs** (contraste 2). Ce choix est dû au fait que les consonnes nasales provoquent le moins de perturbations microprosodiques et donc qu’elles peuvent être considérées comme notre groupe de contrôle. L’ensemble des facteurs des modèles mixtes est schématisé dans la Tab.2.2. Vue la complexité des analyses de régression multiple menées sur le jeu des données, nous avons choisi un α de *pMCMC*

6. Nous rappelons en fait que l’ANOVA est un “cas” spécial de régression, de sorte que la somme des carrés totale de l’ANOVA corresponde à celle de la régression avec les facteurs catégoriels.

<.01⁷.

Modèles mixtes		
Variable dépendante	Effets fixes	Effets aléatoires
Durée	Modalité Intonative Consonne	Mots
Alignement	Modalité Intonative Consonne	Mots
Valeur mélodique	Modalité Intonative Consonne	Mots
Pente	Modalité Intonative Consonne	Mots

TABLE 2.2 – Schématisation des modèles mixtes appliqués aux variables dépendantes

2.2.6 Modélisation de la pente de f_0

Comme on se souviendra, une hypothèse contraire à celle de l’ancrage segmental est que l’unité primitive de l’intonation est la pente du contour accentuel, dont la réalisation est supposée être invariable. Si l’hypothèse de l’ancrage segmental est vraie, nous nous attendons à ce qu’en napolitain les cibles tonales soient fixées à des ancrs segmentales spécifiques et que l’excursion tonale soit stable, indépendamment de l’identité segmentale ou de possibles changements de vitesse d’élocution entre les questions et les affirmations. Si la syllabe accentuée est soumise à ces facteurs de variabilité, nous nous attendons donc à ce que ces variations influencent uniquement les propriétés dynamiques de la courbe. Pour tester l’hypothèse de la “pente-invariable”, nous avons donc calculé le degré d’inclinaison

7. La version courante du paquet R utilisé pour effectuer l’analyse de régression linéaire à effets mixtes (*lme4*) ne restitue pas une valeur de p pour les t-tests, car il n’est pas encore clair comment on peut calculer les degrés de liberté dans les modèles qui incluent des effets aléatoires. Une alternative valide aux valeurs “classiques” de p dans les modèles mixtes est d’utiliser ceux calculés par un échantillonnage MONTE CARLO effectué selon une chaîne de Markov ($pMCMC$ = Monte Carlo Markov Chain). Ici, nous écrivons “ $pMCMC$ ” pour les valeurs de p obtenues pour les modèles à effets mixtes ; et tout simplement p pour les modèles sans effets aléatoires (soit lorsque nous avons décidé d’utiliser *a priori* la régression “classique” ou encore lorsque l’effet aléatoire a été enlevé du modèle mixte car sa variance était presque nulle).

de la pente montante entre L1 et H. De plus, nous avons calculé la pente entre H et L2, pour vérifier l'existence d'effets à longue distance de l'attaque segmentale sur la pente descendante de la configuration en montée-descente.

Il faut noter que la façon dont nous avons calculé la pente est différente par rapport à la méthode employée par la plupart des chercheurs dans le cadre de la Théorie Métrique-Autosegmentale. En fait, la pente de f_0 est souvent calculée comme le rapport entre l'excursion de f_0 et le temps de la montée (ou de la descente), car on fait l'hypothèse que l'interpolation entre deux cibles de spécification tonale différente est linéaire. Cependant, certains chercheurs dans le même cadre de travail ont utilisé des techniques plus sophistiquées afin d'obtenir les mesures de pente. Par exemple, D'Imperio (2000) a remarqué qu'en napolitain, la qualité d'ajustement obtenue par la modélisation logistique est meilleure que celle obtenue par la modélisation linéaire, et donc elle utilise une régression logistique pour modéliser la région entre les cibles L et H de l'accent nucléaire (voir ci-dessous pour les détails mathématiques).

Dans notre étude, nous avons examiné la pente de f_0 par l'utilisation d'un modèle logistique à effets mixtes (Pinheiro & Bates, 2000). Nous avons préféré utiliser une technique de modélisation similaire à celle utilisée par D'Imperio (2000) déjà pour le napolitain pour pouvoir comparer nos données avec les siennes. En fait, D'Imperio (2000) avait trouvé un effet pas clair sur la pente de la montée, car elle était plus raide dans les questions d'un locuteur, et plus douce dans les questions de l'autre. Dans notre étude, nous pouvons donc vérifier si les résultats de D'Imperio (2000) peuvent être répliqués, et s'il existe une interaction entre la modalité intonative et la classe consonantique sur la pente. Il faut noter cependant que notre modélisation est différente de celle de D'Imperio (2000). Dans D'Imperio (2000), une courbe logistique était ajustée à chaque observation, et, pour chacune de ces observations, l'information concernant un seul point de la courbe (le point d'inflexion) était extraite. Dans notre modèle, la courbe dans son ensemble est modélisée, ce qui nous permet de contrôler de plus près les effets de la modalité intonative et de la classe consonantique sur la courbe de f_0 . Par exemple, on pourrait supposer que la classe consonantique provoque des variations de la valeur mélodique de L1, et que, pour atteindre la cible H successive, la f_0 soit soumise à une accélération de la vitesse de la

montée avant ou après le point d'inflexion.

En d'autres termes, notre but était de vérifier si les propriétés dynamiques du contour de f_0 dans son ensemble sont directement influencées par la modalité intonative et la classe consonantique, ou si leurs variations peuvent être entièrement prévisibles par les ajustements mélodiques ou d'alignement des cibles tonales. Pour cela, nous avons extrait les valeurs de f_0 de 20 points équidistants dans la région de f_0 incluse entre les cibles L1 et H, pour la pente montante (fig. 2.5, à gauche) et dans la région incluse entre H et L2, pour la pente descendante (fig. 2.5, à droite).

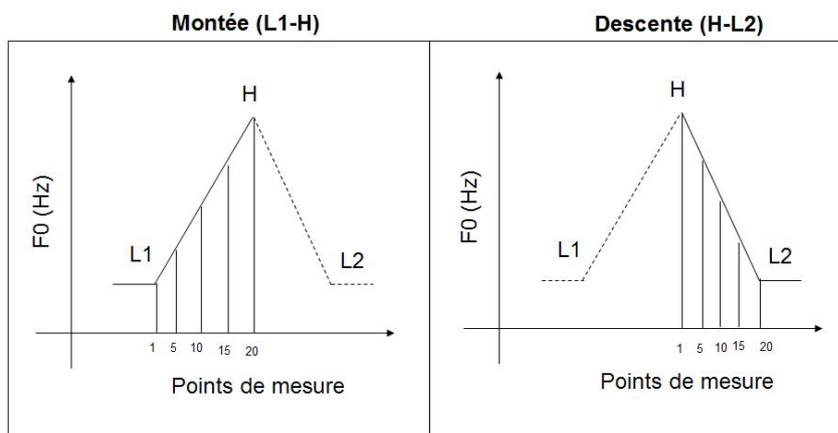


FIGURE 2.5 – Schématisation des 20 points de f_0 extraits entre L1 et H (pour la montée) et entre H et L2 (pour la descente).

Ces deux portions de f_0 ont été modélisées par l'utilisation de la fonction logistique suivante :

$$f(\alpha, \beta, \lambda, \theta, x) = \alpha + \frac{\beta - \alpha}{1 + \exp[(x - \lambda)/\theta]} \quad (2.1)$$

où α représente l'asymptote horizontale à droite (les valeurs maximales de f_0 , ou *topline*), β l'asymptote horizontale à gauche (les valeurs minimales de f_0 , ou *baseline*),

λ le point d'inflexion de la courbe (à mi-chemin entre α et β) et θ la différence entre la valeur correspondante à 0.75α et λ (et donc, la vitesse de la montée). La représentation de la courbe logistique est illustrée dans la fig. 2.6.

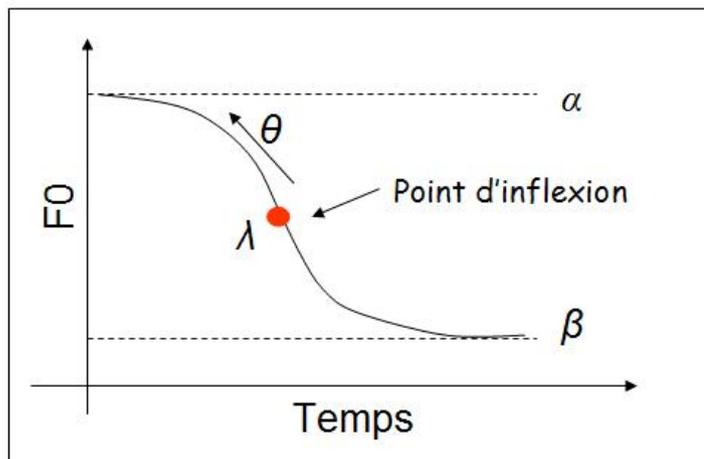


FIGURE 2.6 – Exemple de courbe logistique

De façon similaire aux modèles linéaires (par. 2.2.5), dans les modèles non linéaires à effets mixtes, les effets aléatoires représentent des déviations de la moyenne de la population qui ne sont pas expliquées par les facteurs de traitement. Pour vérifier si un ou plus des quatre paramètres de la courbe logistique sont influencés par la classe consonantique et la modalité intonative, nous avons adopté la procédure suivante.

D'abord, nous avons utilisé un modèle logistique à effets aléatoires, dans lequel la modélisation de la courbe ne prenait pas en compte l'existence d'effets possibles des deux variables prédictives. Ensuite, les deux variables ont été ajoutées progressivement dans le modèle, pour chacun de ces quatre paramètres. La validité de l'inclusion des facteurs fixes a été évaluée par une série de tests de vraisemblance (ANOVA). En statistique, le test de vraisemblance est souvent utilisé pour comparer des modèles deux à deux, et il s'applique uniquement à des modèles emboîtés (un des deux modèles en question comprend un sous-ensemble des paramètres de l'autre). Le facteur fixe est retenu uniquement lorsqu'il modifie de façon significative l'ajustement du modèle aux données. Sinon, il est supprimé et le modèle aléatoire est préféré.

2.3 Résultats

2.3.1 Durée segmentale

Nous avons d'abord vérifié l'existence de différences de durée entre les questions et les affirmations ainsi que parmi les trois classes consonantiques, car plusieurs études ont suggéré que les variations d'alignement tonal peuvent dépendre de variations de la durée segmentale (voir aussi le par. 1.5 dans l'introduction). Les graphiques dans la fig. 2.7 montrent les valeurs de durée des consonnes pour les locuteurs DD (panneau en haut) et AS (en bas). Dans chaque panneau, les résultats de durée (axe des ordonnées) sont fonction de la classe consonantique d'appartenance de l'attaque consonantique (axe des abscisses), et ils sont illustrés séparément pour les deux modalités intonatives, **A** et **Q**. Comme l'on peut remarquer, chez DD, il y a une tendance linéaire à allonger la durée des consonnes de la nasale, à la liquide et jusqu'à l'obstruante dans les affirmations. L'effet de la classe consonantique est inversé dans les questions pour l'obstruante, car sa durée est plus raccourcie que celle de la consonne nasale.

L'ANOVA omnibus du modèle de régression nous montre que chez DD l'effet de la modalité est significatif [$F(2,64) = 19.3$, $pMCMC < .01$], tandis que l'effet de la consonne ne l'est pas [$F(2,64) = 0.22$, $pMCMC > .01$]. L'interaction entre les deux facteurs expérimentaux est juste au dessous du seuil de significativité [$F(2,64) = 3.49$, $pMCMC = .03$]. L'analyse de régression à effets mixtes nous donne d'autres détails. En particulier, chez DD l'effet de modalité intonative est significatif uniquement sur l'obstruante, qui tend à être plus longue dans les affirmations que dans les questions (**A** = 0.066 s, **Q** = 0.057 s; $t = -4.89$, $pMCMC < .01$). Chez AS, l'ANOVA omnibus nous montre qu'il n'y a pas d'effets significatifs sur la durée de la consonne, bien que l'interaction entre modalité intonative et classe consonantique soit juste au dessous du niveau de significativité [$F(2,69) = 3.66$, $pMCMC = .03$]. En fait, l'analyse de régression nous révèle que la durée de la nasale est plus longue dans les affirmations que dans les questions ($t = 2.89$, $pMCMC < .01$), tandis que la durée des consonnes liquides et des obstruantes ne diffère pas selon la modalité intonative.

La durée des voyelles est influencée par la modalité intonative de façon beaucoup

plus nette. Chez DD, l'ANOVA omnibus du modèle nous confirme un effet significatif de la modalité intonative [$F(1,64) = 135.69$, $pMCMC < .01$], tandis que l'effet de la classe consonantique [$F(2,64) = 0.02$, $pMCMC < .01$] et les interactions ne sont pas significatifs [$F(2,64) = 1.75$, $pMCMC < .01$]. En fait, l'analyse de régression nous montre que le contraste entre **A** et **Q** est significatif pour les voyelle précédées par la nasale ($t = -6.9$, $pMCMC < .01$), la liquide ($t = -5.66$, $pMCMC < .01$) et l'obstruante ($t = -7.6$, $pMCMC < .01$), tandis que l'effet de la classe consonantique et les interactions ne sont pas significatifs ($pMCMC > .01$). Cela signifie que les voyelles sont globalement plus allongées dans les affirmations que dans les questions, et la séparation selon le type de consonne effectuée par le modèle de régression n'a pas de pertinence statistique.

Chez AS, on retrouve la même situation. L'ANOVA nous montre que l'effet de la modalité est significative [$F(1,64) = 20.24$, $pMCMC < .01$], mais non l'effet de la classe consonantique [$F(2,64) = 0.04$, $pMCMC > .01$] ni les interactions [$F(2,64) = 0.76$, $pMCMC > .01$]. Les contrastes par paires dans l'analyse de régression nous révèlent que la durée des voyelles est plus allongée dans les affirmations, soit s'elles sont précédées par la nasale ($t = -3.8$, $pMCMC < .01$), que par la liquide ($t = -3.2$, $pMCMC < .01$) ou l'obstruante ($t = -2.9$, $pMCMC < .01$), mais dans aucun des trois cas l'effet de la classe consonantique est significatif ($pMCMC > .01$). Globalement, les résultats de durée confirment que la vitesse d'élocution est un indice secondaire de la modalité intonative, car les questions tendent à être prononcées plus rapidement des affirmations (D'Imperio, 2000; Haan, 2001; van Heuven & van Zanten, 2005). De plus, l'effet de la modalité intonative s'applique de façon consistante sur la durée des voyelles, tandis qu'il dépend de la classe consonantique et du locuteur pour les consonnes en position d'attaque consonantique.

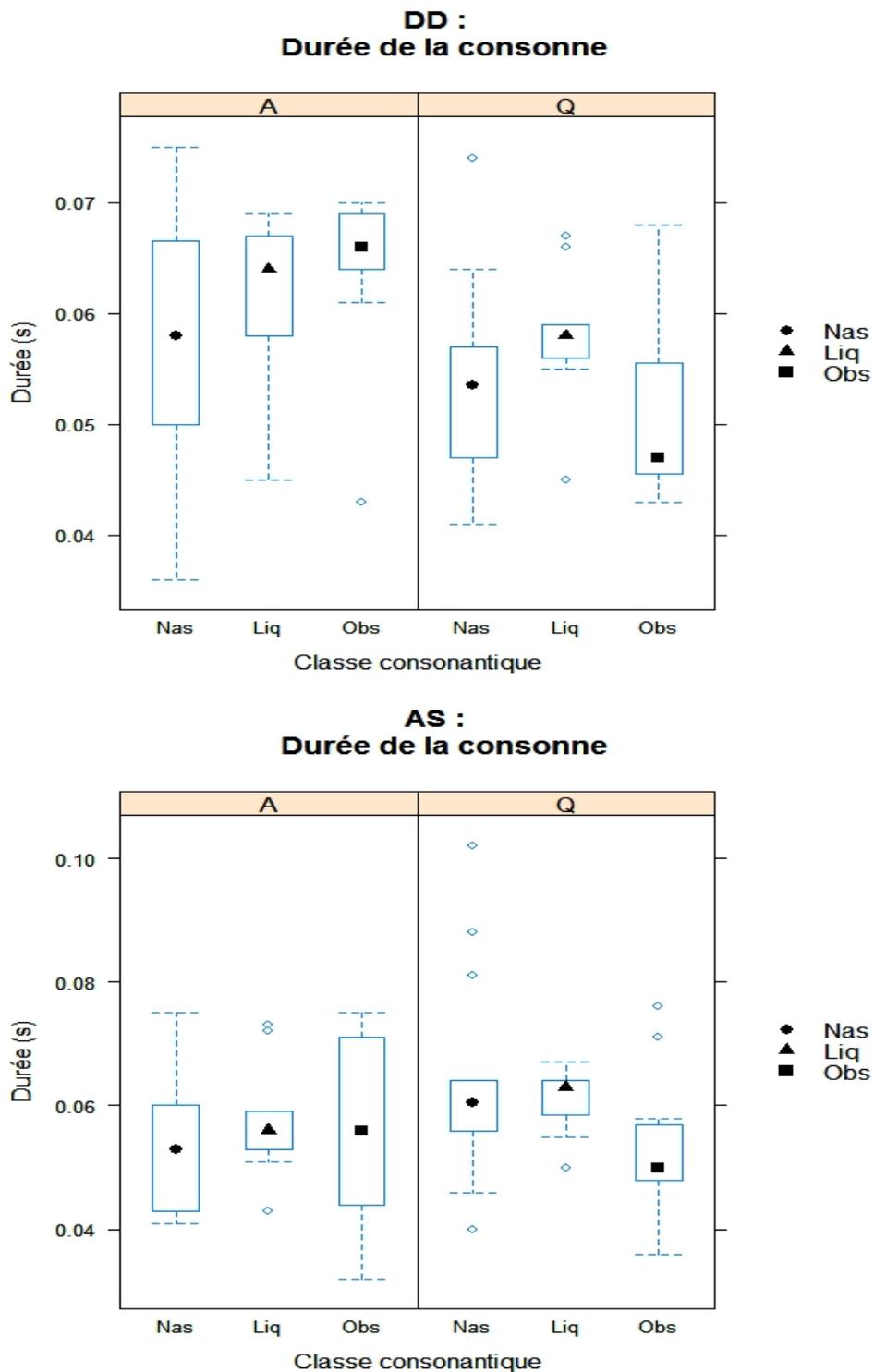


FIGURE 2.7 – Boxplots de la durée de la consonne (s) pour DD (en haut) et AS (en bas). Les résultats sont illustrés séparément pour les affirmations à focalisation étroite (A) et pour les questions oui/non (Q), et selon la classe consonantique.

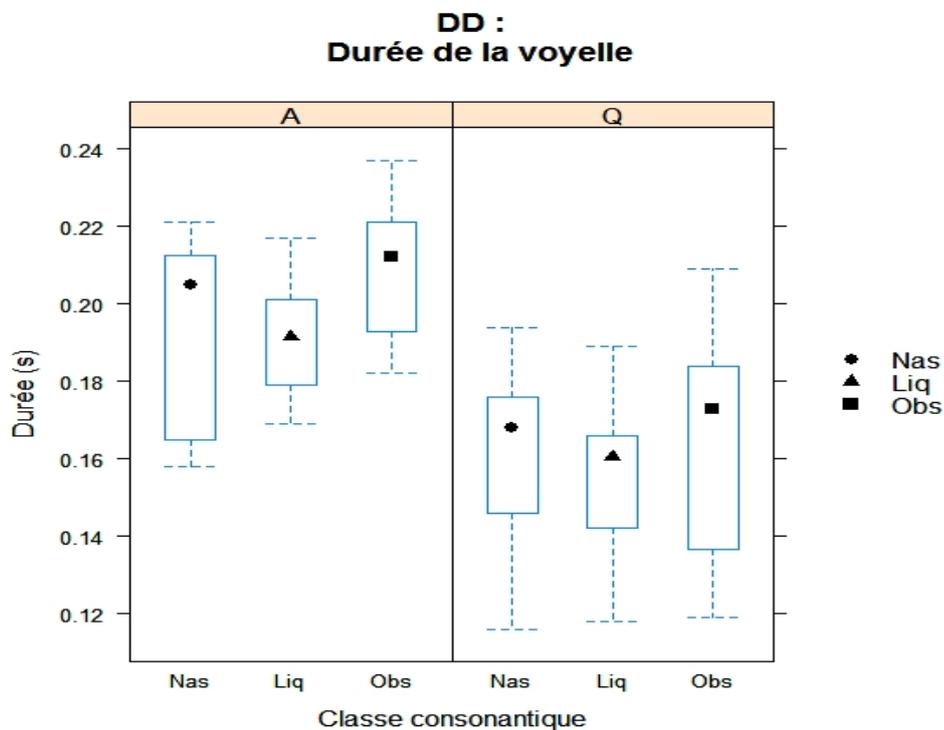
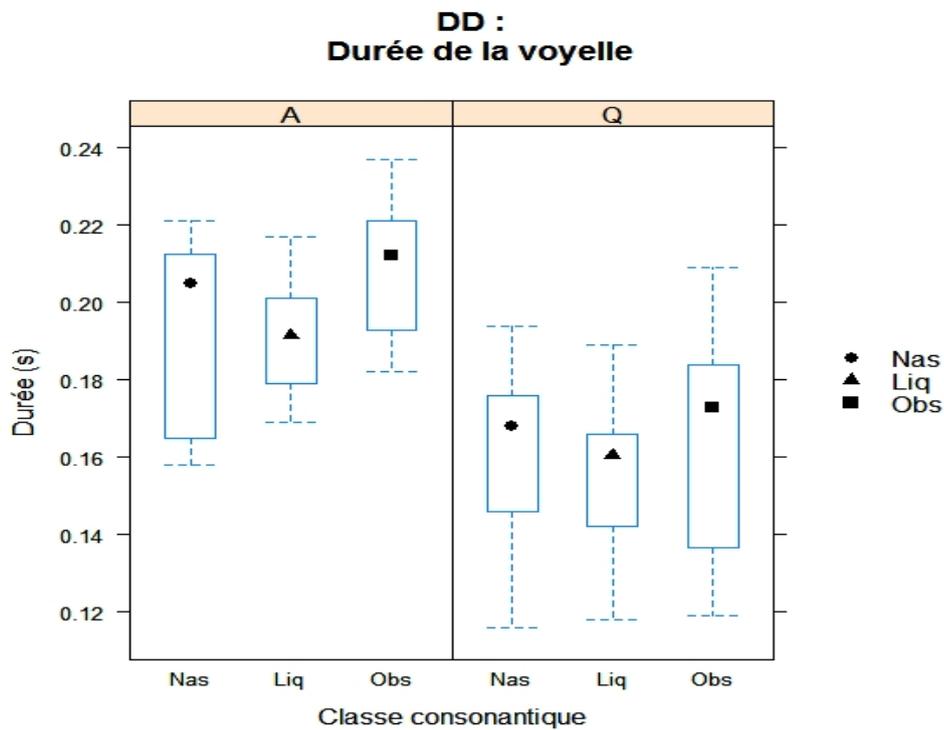


FIGURE 2.8 – Boxplots de la durée de la voyelle (s) pour DD (en haut) et AS (en bas). Les résultats sont illustrés séparément pour les affirmations à focalisation étroite (**A**) et pour les questions oui/non (**Q**), et selon la classe consonantique.

2.3.2 Détection des cibles tonales

La détection des cibles tonales a été effectuée selon la procédure manuelle, le LSF et MOMEL. Nous allons d’abord commenter quelques exemples de détection. Dans les figures 2.9, 2.10 et 2.11, la dernière portion de la phrase porteuse est illustrée, qui contient le verbe *vedere* et le dernier mot prosodique (c’est-à-dire, le mot-cible). Nous avons fait un “zoom” sur cette région pour examiner de plus près les différences de localisation parmi les trois procédures. Les résultats de la localisation sont indiqués par un cercle (sur le point de f_0 identifié comme L1) et par une flèche. Les étiquettes indiquent : *MOM*= MOMEL ; *LSF*= Least Square Fitting ; *man*= détection manuelle. Dans le TextGrid, le premier *tier* contient l’information d’alignement détectée par les trois méthodes, et le deuxième les segments (pour rendre la lecture plus simple, nous avons fait une transcription orthographique dans les exemples). Comme on peut le remarquer, les trois procédures donnent des résultats très différents.

Dans la fig. 2.9, par exemple, la cible L1 est détectée au début de la montée par la méthode manuelle, début qui coïncide aussi avec le début de l’attaque syllabique /d/. LSF et MOMEL localisent L1 dans une région temporelle plus avancée. En particulier, L1 est localisé par LSF dans le plateau de la première voyelle du mot prosodique *La Dana*, donc avant le début de la syllabe accentuée. Une possible explication est que la détection du LSF a été influencée par la une forme convexe dans la région entre le début de la syllabe accentuée et le pic maximal de f_0 . Puisque l’algorithme se base sur une fonction linéaire, il est probable qu’il n’a pas réussi à apprécier cette subtile variation de forme. Il faut noter que nous ne faisons pas ici des hypothèses précises sur la raison de cette forme “convexe” (qui peut être dû à la présence de l’occlusive sonore en position d’attaque ou encore à des propriétés typiques de l’accent nucléaire des affirmations), mais tout simplement sur comment certains paramètres acoustiques peuvent avoir influencé la détection. MOMEL tend à placer la cible encore plus avant que le LSF, notamment sur le creux provoqué par la liquide /l/. En général, en fait, la localisation de la cible basse par MOMEL était beaucoup plus avancée par rapport à celle du LSF et à la détection manuelle. Cela peut être dû à une différence intrinsèque dans la notion de “cible”, qui est définie dans MOMEL comme l’extrémité d’une parabole. En d’autres termes, la cible

est le résultat de l'application d'une régression polynomiale quadratique, et non d'une régression linéaire, comme dans le LSF.

Dans d'autres cas, la présence des perturbations segmentales a créé des problèmes pour la détection manuelle. Par exemple, dans la fig. 2.10 le bruit de friction de la consonne /r/ provoque une forte baisse de la f_0 en correspondance avec cette consonne. La valeur minimale de f_0 atteinte dans l'attaque consonantique a été donc considérée comme le résultat d'une perturbation segmentale, et la cible a été localisée quelques millisecondes avant la perturbation. Pour la même courbe de f_0 , le LSF localise la cible après la perturbation et MOMEL, au contraire, au début du mot prosodique. Il est clair que sans aucune preuve articulatoire ou perceptive, il serait impossible d'identifier la localisation de la cible L1, disons, avec le point trouvé par le LSF ou par la détection manuelle. Dans plusieurs cas, cependant, les résultats de la détection manuelle et du LSF sont très similaires (fig. 2.11).

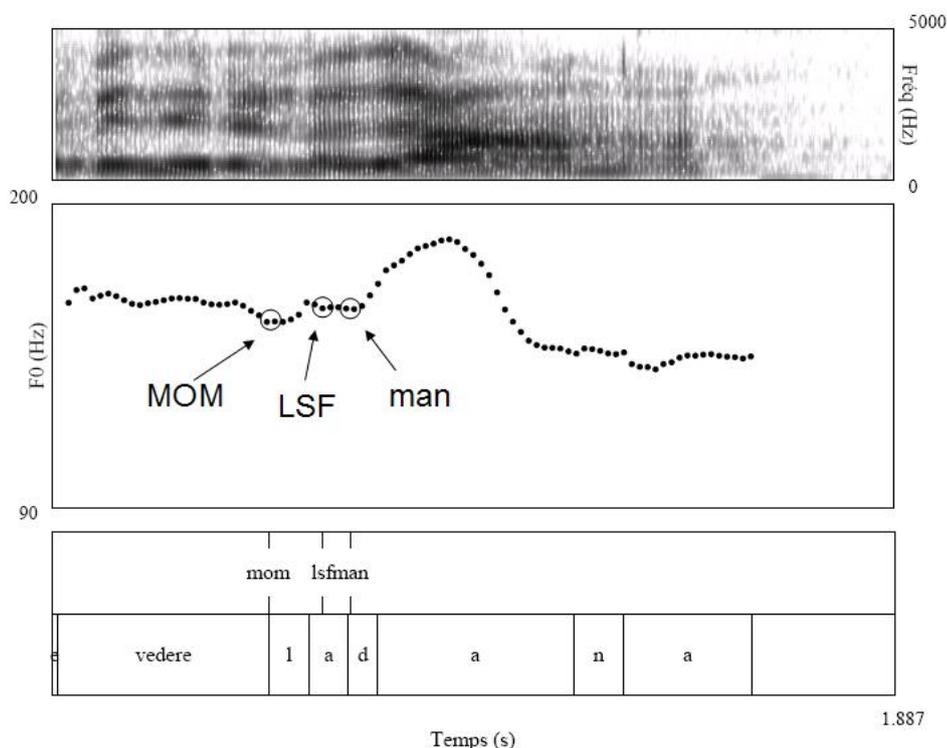


FIGURE 2.9 – Exemple de détection des trois cibles par les méthodes manuelle, LSF et MOMEL.

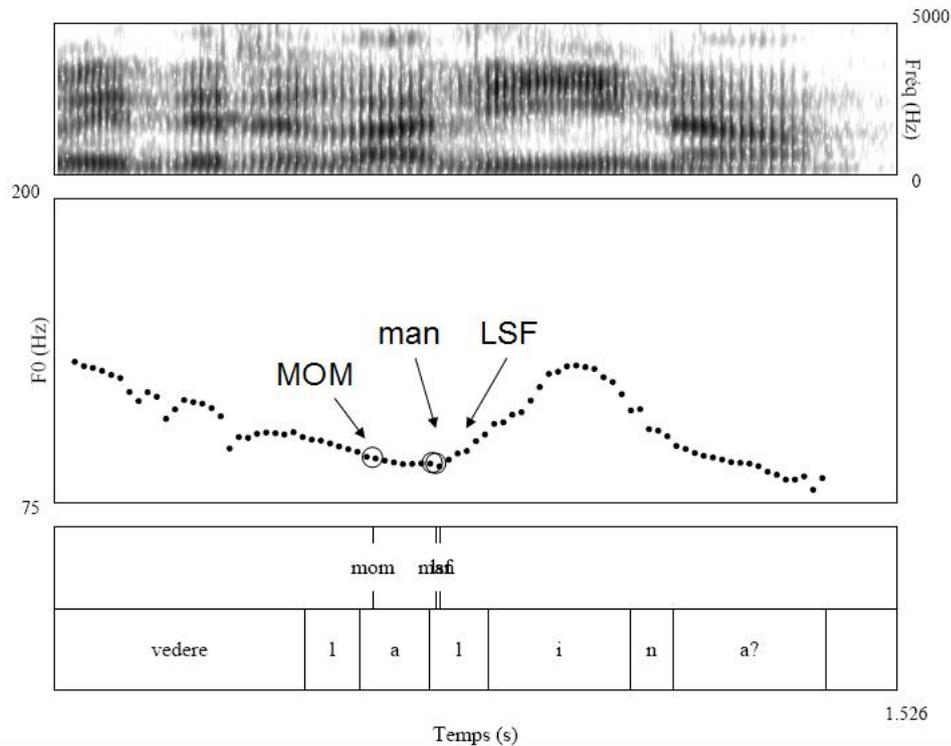


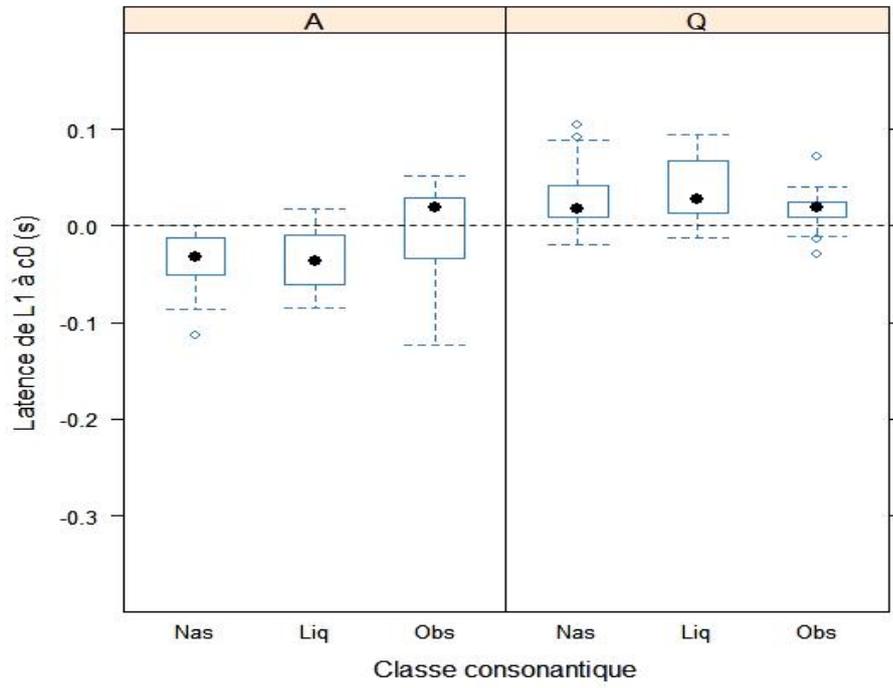
FIGURE 2.11 – Exemple de détection des trois cibles par les méthodes manuelle, LSF et MOMEL. Le deux étiquettes “lsf” et “man” dans le premier *tier* du TextGrid sont superposées.

Dans la fig. 2.12, les valeurs d’alignement de L1 avec l’attaque consonantique $c\theta$ (ordonnées) en fonction des classes consonantiques (abscisses) sont montrées séparément pour le LSF, la méthode manuelle et MOMEL. Les résultats pour les affirmations et les questions sont illustrés respectivement dans les panneaux à gauche et à droite de chaque graphique.

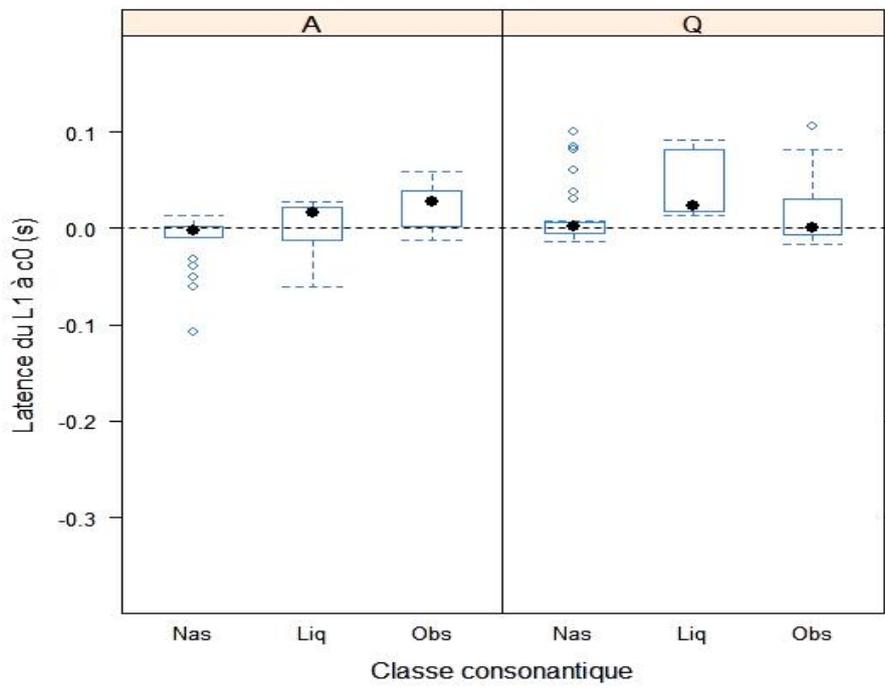
Comme on peut le remarquer, l’alignement de L1 est plus retardé dans les questions que dans les affirmations, surtout lorsque la cible est détectée par le LSF et MOMEL. L’alignement de L1 est ancré de façon stable à l’attaque de la syllabe lorsque L1 est détecté par la méthode manuelle, et ceci indépendamment de la modalité intonative. De plus, fig. 2.12 la longueur des boxes est très courte pour la détection manuelle et le LSF, la différence entre le premier et troisième quartile étant très petite. Cela est en ligne avec del Giudice *et al.* (2007), qui ont montré que les annotateurs humains sont très cohérents, et que le LSF peut obtenir des résultats similaires à la méthode manuelle. Bien que les résultats de la détection manuelle soient en accord avec l’hypothèse de l’ancrage

segmental, ceux obtenus par le LSF et MOMEL nous semblent plus fiables. En fait, la différence d’alignement entre les deux modalités intonatives reflète les résultats perceptifs trouvés par D’Imperio (2000). Nous rappelons que D’Imperio (2000) a montré que les cibles “perçues” de la configuration LHL de l’accent nucléaire sont systématiquement plus retardées dans les questions que dans les affirmations.

LSF



Détection manuelle



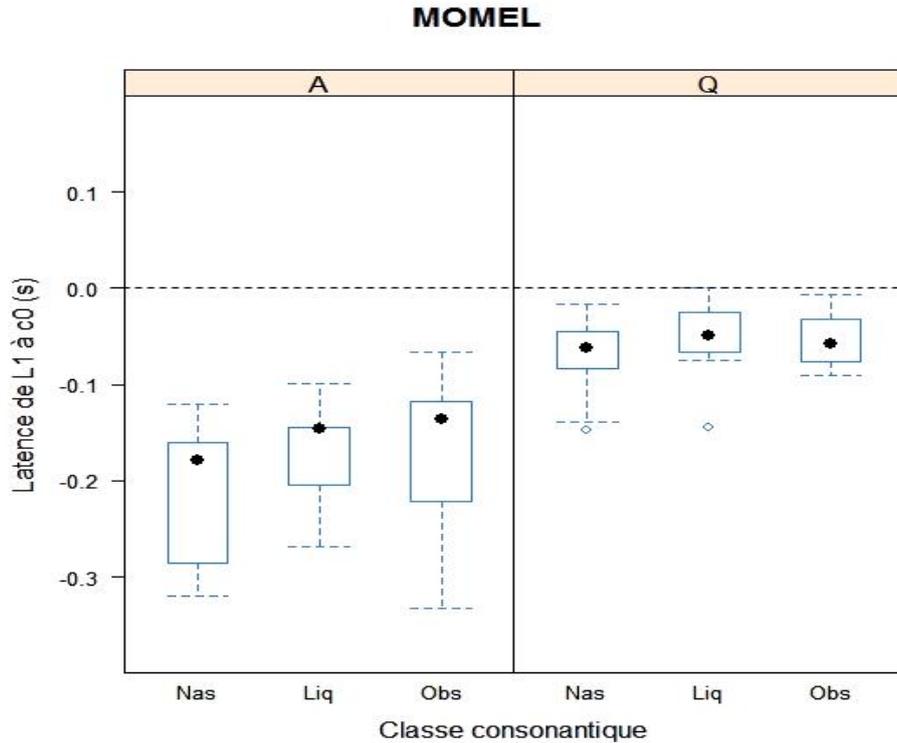


FIGURE 2.12 – Boxplots de la latence de L1 par rapport au début de la consonne (s) comme trouvé par le LSF, la méthode manuelle et MOMEL. Les résultats sont regroupés pour les deux locuteurs. La ligne tiretée marque le début de la consonne.

Cependant, la localisation temporelle exacte de L1 est différente entre le LSF et MOMEL. Le LSF tend à localiser L1 avant ou en relation avec l’attaque syllabique dans les affirmations, et juste après cette frontière dans les questions. Il faut noter cependant que les résultats de la détection varient en fonction de la classe consonantique, car la différence d’alignement entre questions et affirmations est plus forte en présence des consonnes nasales et liquides. Au contraire, le L semble être aligné avec l’attaque obstruant, indépendamment de la modalité intonative. L1 est détectée à une distance bien plus large par MOMEL, car sa valeur médiane est de 57 ms avant l’attaque syllabique avant $c0$ dans les questions, et de 168 ms dans les affirmations. De plus, la fig. 2.12 nous indique que dans les affirmations, la localisation de L1 effectuée par MOMEL est beaucoup plus variable parmi les répétitions du même mot, indépendamment du type de consonne à l’attaque syllabique. En fait, la distance entre le premier et le troisième quartile est ici beaucoup plus large. L’incohérence des mesures d’alignement tonal obtenues par MOMEL pourrait être due à l’absence d’un vrai point d’inflexion dans les affirmations, qui pourrait refléter

l'absence d'un ton phonologique bas dans les affirmations (Hirst, c.p.). Cette explication est contraire aux résultats obtenus par des investigations acoustiques menées par D'Imperio (1999), qui a montré que le coude de f_0 dans les constituants focalisés longs des affirmations représente une cible réelle L, donc confirmant l'analyse LH de cet accent. Nous pensons, comme Hirst, que la cible L des affirmations a des caractéristiques acoustiques différentes du L des questions. Pendant notre expérience d'analyse acoustique, nous avons souvent remarqué que le L des questions oui/non est signalé par une accélération importante du mouvement de descente de la f_0 . Le L a une forme parabolique, concave dans les questions (et cela pourrait aussi expliquer pourquoi MOMEL marche mieux dans cette modalité intonative), et sa valeur mélodique est souvent beaucoup plus basse que dans les affirmations (voir aussi D'Imperio (2000)). Au contraire, dans les affirmations, le L coïncide tout simplement avec le début de la montée, souvent difficilement visible comme un point d'inflexion⁸.

Nous avons décidé de garder l'algorithme LSF pour la localisation des cibles tonales. Pour un souci de cohérence, cette procédure automatique a été employée également pour la détection des cibles tonales H et L2. Notamment, L- a été localisée au point d'intersection de deux lignes droites ajustées dans la région de f_0 entre le pic de f_0 et la fin de l'énoncé, et la cible H au point d'intersection de deux lignes entre L1 et L2.

2.3.3 Alignement tonal

Dans ce paragraphe, nous allons rendre compte de l'effet de modalité intonative et de classe consonantique sur l'alignement temporel des cibles de l'accent nucléaire et de l'accent de syntagme suivant (L1, H, et L2). La localisation temporelle a été calculée séparément pour les trois cibles et selon trois types de mesures. En premier lieu, l'alignement a été mesuré en terme de distance absolue des frontières segmentales (le début de la consonne, le début de la voyelle et la fin de la voyelle), pour vérifier si l'alignement temporel des cibles est défini par rapport à des points d'ancrage spécifique (Arvaniti *et al.*,

8. Bien que ces différences de forme du L ne soient pas prises en compte par le modèle Métrique-Autosegmentale, nous pensons qu'il pourrait être très intéressant de voir si et comment elles influencent la perception de cette cible, ou si la cible L des questions et des affirmations est déterminée par les mêmes commandes articulatoires.

1998, *inter alia*). De plus, nous avons calculé l’alignement comme proportionnel à la durée des segments intra-syllabiques, car nous avons considéré l’hypothèse selon laquelle l’alignement tonal varie selon la durée de la voyelle accentuée (Silverman & Pierrehumbert, 1990) ou selon la durée de la consonne et de la voyelle de la syllabe accentuée (van Santen & Hirschberg, 1994; Prieto *et al.*, 1995). Afin d’établir l’existence d’effets de la durée intra-syllabique sur l’alignement des cibles, nous avons sélectionné le début de la consonne comme point de référence pour mesurer l’alignement (Prieto *et al.*, 1995). Finalement, l’alignement des cibles a été calculé en terme de pourcentage dans la durée de la voyelle et de la syllabe accentuée et ce pourcentage a été utilisé directement en tant que variable dépendante dans le modèle de régression.

La première question qui nous abordons maintenant est de déterminer quel modèle représente au mieux l’alignement des cibles. En statistique, l’adéquation de différents modèles aux données observées peut être évaluée à l’aide de différentes méthodes, telles que l’Akaike’s Information Criterion (AIC) ou le test du rapport de vraisemblance (*likelihood ratio test*, Venables & Ripley, 1999). L’AIC est un critère d’information similaire au R^2 mais plus fiable, car il permet de comparer plusieurs modèles différents entre eux, en pondérant la bonté d’ajustement de chaque modèle avec le nombre des paramètres qu’il comporte. Le modèle qui présente la valeur absolue de AIC la plus faible est considéré comme le plus parcimonieux et donc il est retenu en vertu du rasoir de Ockham. Pour le test de vraisemblance, nous renvoyons au par. 2.2.5.

Dans notre étude, nous avons utilisé les deux méthodes de comparaison. En premier lieu, nous avons comparé la valeur AIC des modèles représentant l’alignement des cibles L1, H et L2 en termes absolus, proportionnels ou de pourcentage. Comme on le remarquera par les Table A.1 et A.3, globalement, les modèles les plus performants sont ceux pour lesquels l’alignement a été mesuré en terme de distance absolue des frontières segmentales et en fonction de la durée des segments intra-syllabiques. Au contraire, lorsque l’alignement est introduit dans le modèle en pourcentage, les valeurs de AIC sont beaucoup plus élevées, ce qui signifie que ce type de mesure n’est pas approprié pour représenter l’alignement tonal⁹ introduction de la variable dépendante en pourcentage est, en réalité,

9. L

inappropriée dans les modèles de régression classiques car cela provoque une violation de l'assomption de normalité du modèle. Dans ce cas, il faudrait plutôt utiliser un modèle linéaire généralisé..

Ensuite, nous avons effectué une série de tests de vraisemblance pour vérifier quel est le modèle le meilleur pour exprimer l'alignement. Nous avons comparé deux groupes de modèles : (1) le modèle de van Santen & Hirschberg (1994), dans lequel les variables comparés étaient quatre (modalité intonative, durée de l'attaque consonantique, durée de la voyelle accentuée, classe consonantique, **ModvSH**), avec un modèle présentant deux covariées (modalité intonative et classe consonantique, **Mod1a**) et avec un modèle qui considère uniquement la modalité intonative comme variable indépendante (**Mod1b**); (2) le modèle de Silverman & Pierrehumbert (1990), dans lequel les régresseurs étaient deux (durée de la voyelle et modalité intonative, **ModSP**) avec un modèle présentant la modalité intonative comme seule variable indépendante (**Mod1c**). A cause du nombre élevé des variables pris en compte, la variance du facteur mot était proche de zéro et donc l'effet aléatoire n'a pas été considéré. Nous avons décidé de tester le pouvoir explicatif de ces modèles sur l'alignement de la cible H de la configuration en montée-descente, car les modèles de van Santen & Hirschberg (1994) et de Silverman & Pierrehumbert (1990) ont été utilisés pour modéliser la variabilité d'alignement du pic accentuel, mais non l'alignement de la cible basse (mais pour van Santen, ce choix n'a pas de valeurs théoriques, car chaque point de la courbe accentuelle peut être modélisé par une fonction de déformation temporelle, voir par. 1.4). Il faut aussi noter que dans nos modèles la modalité intonative est aussi prise en compte, car Silverman & Pierrehumbert (1990) et Prieto *et al.* (1995) avaient déjà montré que l'insertion des facteurs prosodiques dans le modèle d'alignement améliore le pouvoir explicatif du modèle de façon significative.

Pour le premier groupe de modèles, la variable dépendante était l'alignement du pic relativement à l'attaque syllabique, car cette mesure est celle typiquement utilisée dans le modèle de van Santen & Hirschberg (1994) (Prieto *et al.*, 1995; voir aussi Prieto & Torreira (à paraître) pour une discussion sur ce sujet). Pour le deuxième, la variable dépendante était l'alignement du pic relativement à l'attaque vocalique, car ce type d'alignement est celui typiquement utilisé comme base pour obtenir la mesure proportionnelle

de l'alignement du pic relativement à la rime (Silverman & Pierrehumbert, 1990).

Les résultats de tests de vraisemblance nous montrent une différence de qualité d'ajustement des deux modèles selon le locuteur. Pour DD, **ModvSH** n'est pas meilleur de **Mod1a** ($p > .01$) ni de **Mod1b** ($p > .01$). En fait, la variance expliquée par les trois modèles était très similaire (entre 78% et 79%). De façon similaire, **ModSP** n'était pas meilleur de **Mod1c** (variance expliquée autour de 80% pour les deux modèles). Pour le locuteur AS, au contraire **ModvSH** est meilleure de **Mod1a** ($p < .01$) et de **Mod1b** ($p < .01$). Donc, l'insertion des informations relatives à la durée intra-segmentale et à la classe consonantique permettent une amélioration significative du modèle de 2% relativement aux deux modèles plus simples (la variance expliquée par **ModvSH** étant de 91%). Le modèle de Silverman & Pierrehumbert (1990) est aussi meilleur que celui plus simple ($p < .01$), et la variance expliquée par ce modèle est très similaire à celle du modèle de van Santen & Hirschberg (1994) (90%).

Ces résultats peuvent indiquer que, chez DD, l'alignement des cibles tonales est calculé relativement aux frontières segmentales. En d'autres termes, l'alignement est insensible aux variations de la durée intra-syllabique, car les cibles sont alignées de façon stable à certains ancrages (tels que l'attaque syllabique ou la fin de la voyelle), et la seule source de variabilité à laquelle elles sont soumises est la modalité intonative. Pour AS, l'interprétation des résultats est plus problématique, mais nous ne pouvons pas exclure que même chez ce locuteur, l'hypothèse de l'ancrage segmental puisse ramener à une modélisation de l'alignement meilleure ou comparable à celle obtenue par les modèles de Silverman & Pierrehumbert (1990) et de van Santen & Hirschberg (1994). Par conséquent, nous allons détailler ici les résultats d'alignement des cibles, exprimés comme distance absolue de frontières segmentales et en fonction de la modalité intonative et de la classe consonantique.

Les valeurs moyennes d'alignement de L1, H et L2 avec l'attaque consonantique ($c\theta$) sont consignées dans la fig. 2.13 pour DD (en haut) et pour AS (en bas). Les résultats pour les affirmations et les questions sont montrés respectivement dans les panneaux à gauche et à droite de chaque graphique. Dans chaque panneau, les cibles tonales sont définies par leur latence relativement au début de l'attaque consonantique (axe des abscisses) et

par leur valeur mélodique (axe des ordonnées). Les résultats d'alignement des cibles sont représentés en fonction de la classe d'appartenance de la consonne par trois symboles différents (**Liq** = points, **Nas** = triangles, **Obs** = carrés).

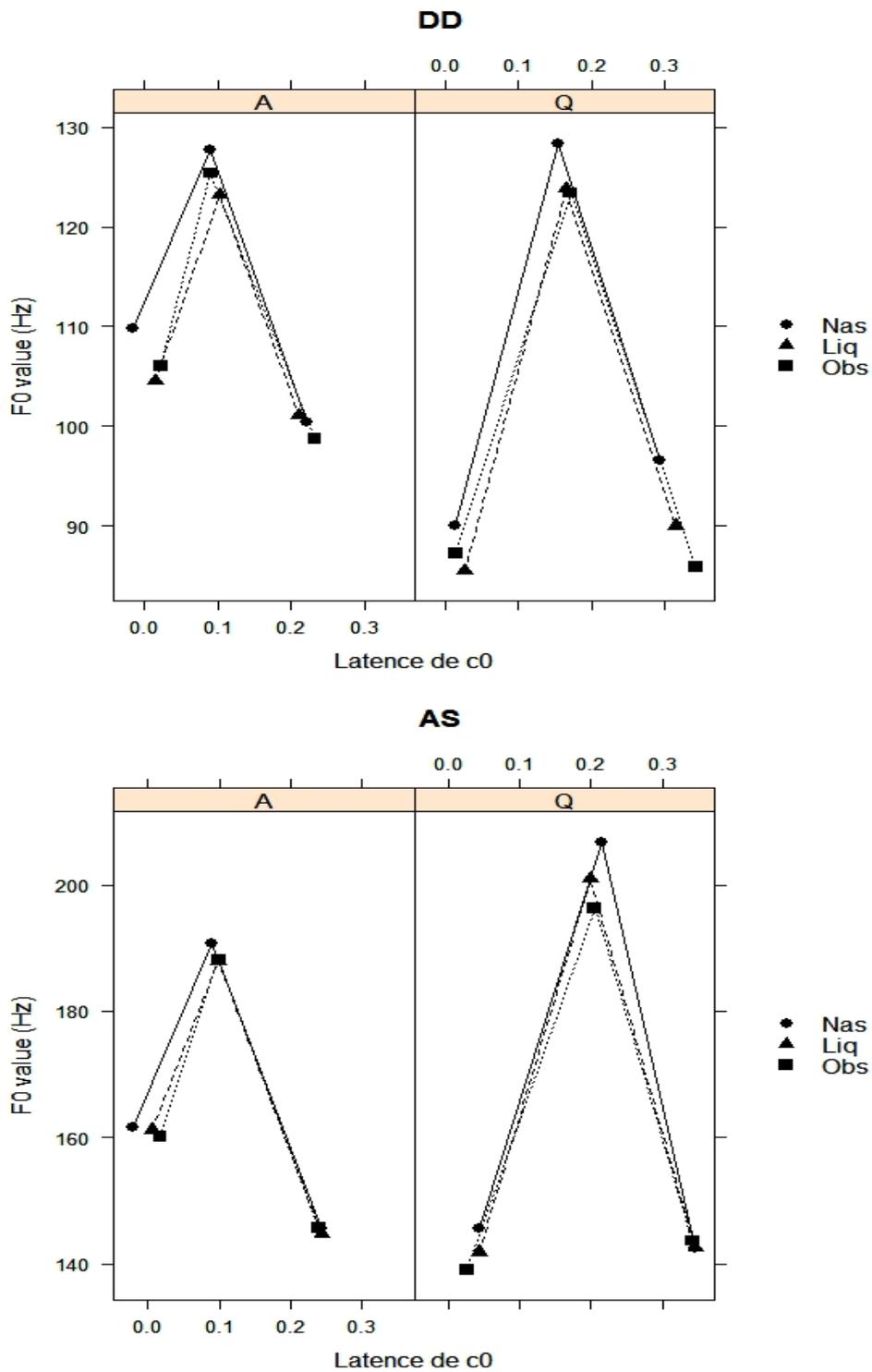


FIGURE 2.13 – Valeurs de f_0 et d'alignement tonal de L1, H et L2 pour DD (en haut) et AS (en bas). Les résultats sont présentés séparément en fonction de la modalité intonative et du type de consonne.

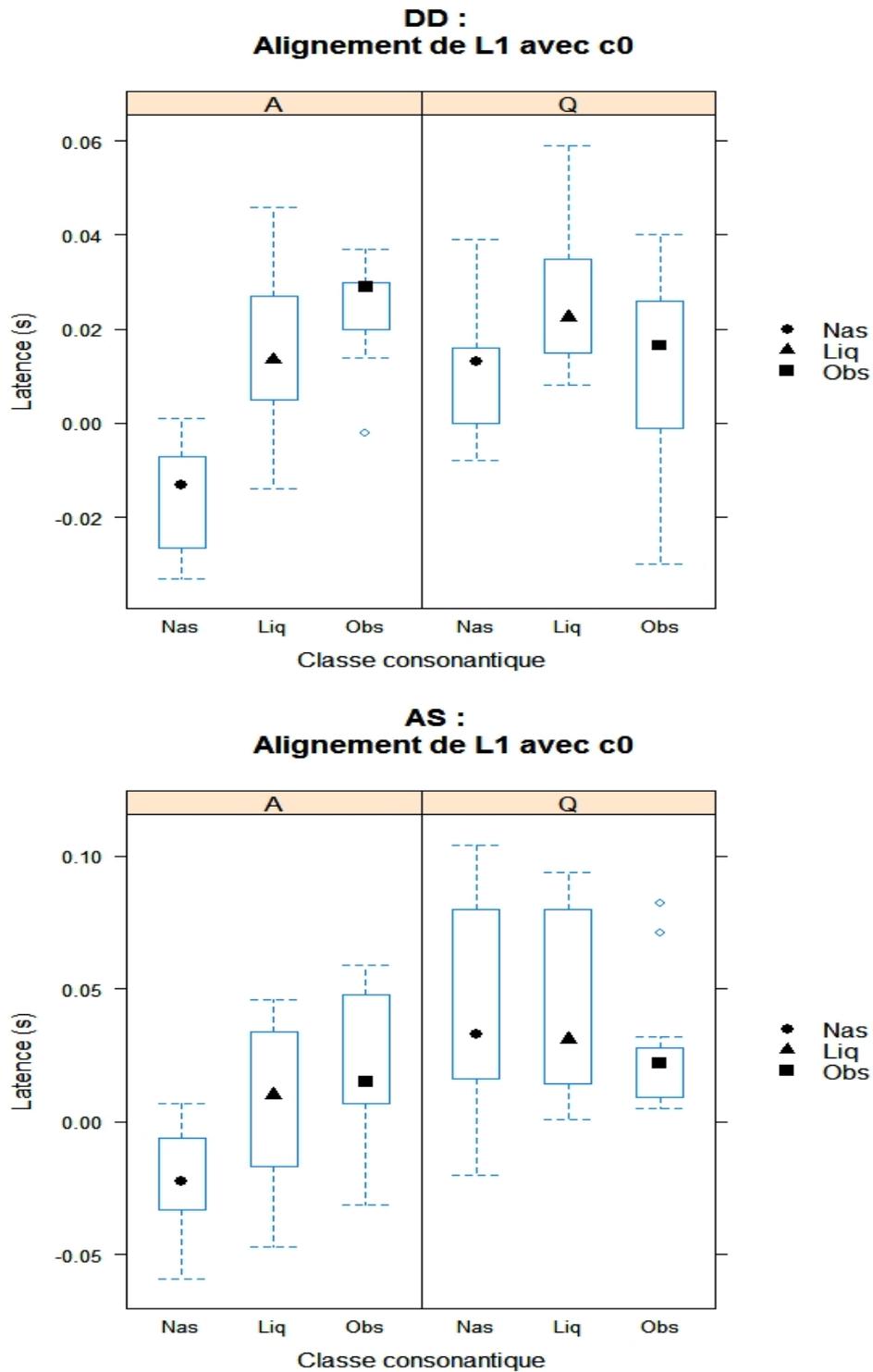


FIGURE 2.14 – Boxplots des latences de L1 par rapport au début de la consonne (c_0) pour DD (en haut) et AS (en bas). Les résultats sont présentés séparément en fonction de la modalité intonative et du type de consonne. La ligne tiretée marque le début de la consonne.

Globalement, l'alignement des cibles tonales est plus retardé dans les questions que dans les affirmations. En particulier, la cible L1 se produit en relation avec le début de l'attaque consonantique dans les affirmations chez les deux locuteurs mais elle était alignée à 17 ms et à 38 ms après cette frontière dans les questions, respectivement chez AS et DD. La cible H était réalisée 35 ms après l'attaque vocalique dans les affirmations, et dans la portion finale de la voyelle accentuée dans les questions. En particulier, dans les questions, le pic accentuel se produisait à 51 ms et 28 ms avant la fin de la voyelle respectivement chez DD et AS. Finalement, L2 tendait à se produire à la fin de la voyelle accentuée dans les affirmations chez AS et 38 ms avant cette frontière chez DD. Dans les questions, L2 était beaucoup plus retardée, son alignement étant réalisé dans la syllabe post-accentuelle chez les deux locuteurs. Ce résultat n'est pas surprenant, car il confirme les découvertes déjà mentionnées dans la littérature sur l'italien napolitain (D'Imperio, 2000, 2002a) : les cibles tonales tendent à être plus retardées dans les questions que dans les affirmations.

La fig. 2.13 nous montre aussi que, contrairement à nos attentes, les effets de modalité intonative ne sont pas les mêmes en fonction des classes consonantiques. De plus, la cible L1 n'est pas ancrée au même endroit temporel mais son alignement semble soumis à l'identité segmentale de l'attaque consonantique. Nous voulons arrêter notre attention, donc, sur les effets de la classe consonantique sur l'alignement des trois cibles, et sur son interaction avec la modalité intonative.

La fig. 2.14 montre l'alignement de L1 relativement au début de la consonne (*c0*). Cette mesure a été sélectionnée car plusieurs études ont montré que le début de l'attaque syllabique (correspondant ici à *c0*) est l'ancre "préférée" pour l'alignement du début des accents montants. De plus, en napolitain l'alignement de la cible L de l'accent nucléaire LH semblent être assez stable lorsqu'elle est mesurée relativement à la frontière gauche de la syllabe (D'Imperio, 2000; Petrone & Ladd, 2007).

La fig. 2.14 nous montre que bien que L1 tende à être aligné de façon différente dans les affirmations et dans les questions, il y a une forte variabilité à l'intérieur et parmi les classes consonantiques. Notamment, dans les affirmations de DD et AS, la cible L1 tend à être alignée juste avant le début de la consonne dans la nasale, tandis qu'elle est réalisée en relation avec ou juste après *c0* respectivement dans les liquides et les obstruantes. Dans

les questions, comme on s'y attendait, L1 est réalisé après *c0* chez les deux locuteurs, et cela est indépendant de la classe consonantique.

L'écart d'alignement de L1 entre questions et affirmations est donc très important en présence de la consonne nasale : L1 est alignée 29 ms plus tard dans les questions que dans les affirmations chez DD, et 64 ms plus tard chez AS. Les résultats d'alignement avec la liquide sont incohérents entre les deux locuteurs, car L1 est aligné 8 ms plus tard dans les questions que dans les affirmations chez DD, mais il a les mêmes valeurs de latence chez AS. Finalement, l'écart d'alignement de L1 entre questions et affirmations en présence de l'obstruante est de 11 ms chez DD et de 7 ms chez AS.

Les résultats de l'ANOVA omnibus du modèle de régression montrent un effet significatif de la modalité intonative chez DD ($[F(2,64) = 7.64; p < .01]$) et AS ($[F(2,69) = 30.93; p > .01]$). De plus, nous avons trouvé un effet principal de classe consonantique uniquement chez DD ($[F(2,64) = 13.16; p < .01]$), mais pas chez AS ($[F(2,69) = 1.54; p > .01]$). Cependant, l'interaction entre classe consonantique et modalité intonative est significative pour DD ($[F(2,64) = 9.87; p < .01]$) et AS ($[F(2,64) = 5.3; p < .01]$). Ces résultats suggèrent que, bien que l'alignement de L1 tende à être plus retardé dans les questions, il y a aussi des différences d'alignement parmi les classes consonantiques et entre les deux locuteurs. L'ANOVA omnibus nous restitue des informations partielles sur les types de contrastes entre les classes consonantiques et l'interaction entre ce facteur et la modalité. Nous allons donc détailler les résultats de l'analyse par paire de la régression. L'analyse de régression nous révèle que l'effet de modalité est significatif pour les nasales, pour DD ($t = 4.7, p < .01$) ainsi que pour AS ($t = 5.57, p < .01$); pour les liquides, il est significatif uniquement pour AS ($t = 3.24, p < .01$), et non pas pour DD ($t = 1.34, p > .01$); enfin, il n'y a pas de différences significatives de modalité intonative en présence des obstruantes, ni chez DD ($t = -1.73, p > .01$) ni chez AS ($t = 0.53, p > .01$).

La fig. 2.14 permet aussi de constater l'existence de différences d'alignement parmi les classes consonantiques et dans la même modalité intonative. Dans les affirmations, l'alignement de L1 est de plus en plus avancé avec la consonne nasale, la liquide et l'obstruante. Le contraste d'alignement de L1 avec la consonne nasale/liquide est significative chez DD ($t = 4.89, p < .01$), mais elle n'atteint pas le seuil de significativité chez AS ($t = 2.33, p$

= .02). La différence d’alignement en présence de la nasale/l’obstruante est significative chez les deux locuteurs (DD : $t = 5.8, p < .01$; AS : $t = 3.06, p < .01$). Dans les questions, l’alignement de L1 ne varie pas en fonction de la classe consonantique. En effet, le contraste d’alignement en présence de la consonne nasale/liquide n’est pas significatif ni pour DD ($t = -2.12, p > .01$) ni pour AS ($t = 0.20, p > .01$). De même, le contraste en présence de la consonne liquide vs obstruante n’est pas non plus significatif (DD : $t = -0.10, p > .01$; AS : $t = -1.45, p > .01$). Cela signifie non seulement que l’alignement de L1 varie en fonction de la classe consonantique, mais également que cette variation n’est pas prévisible, car elle dépend aussi de la modalité intonative.

Des résultats similaires ont été trouvés lorsque l’alignement de L1 a été mesuré par rapport à l’attaque vocalique. Plus précisément, pour tous les deux locuteurs L1 est plus retardé dans les questions que dans les affirmations en présence de la consonne nasale (DD : $t = 5.2, p < .01$; AS : $t = 4.7, p < .01$), tandis qu’il n’y a pas de différences significatives en présence de l’obstruante (DD : $t = 0.24, p > .01$; AS : $t = 0.82, p > .01$). L’effet de modalité intonative en présence de la liquide est significatif, à nouveau, uniquement chez AS ($t = 3.09, p < .01$). En ce qui concerne les contrastes parmi les classes consonantiques, dans les affirmations, l’alignement est avancé dans la nasale plus que dans la liquide ($t = 4.2, p < .01$) chez DD ; et dans la nasale plus que dans l’obstruante chez DD ($t = 4.85, p < .01$) et AS ($t = 4.78, p < .01$). Dans les questions, l’effet de la classe consonantique, à nouveau, n’est pas significatif.

L’alignement de L1 avec la fin de la voyelle était beaucoup plus stable car les variations dues à l’identité segmentale de l’attaque consonantique n’étaient pas prises en compte. La stabilité d’alignement pourrait cependant être due à une différence d’ “échelle” de la mesure utilisée : puisque la distance temporelle de L1 est plus large avec la fin de la voyelle qu’avec le début consonantique, la significativité de l’identité segmentale pourrait être noyée. En effet, le seul facteur significatif était la modalité intonative, pour les deux locuteurs. En particulier, la différence de modalité était significative dans les trois classes consonantiques chez DD (**Nas** : $t = 5.72, p < .01$; **Liq** : $t = 4.02, p < .01$; **Obst** : $t = 3.52, p < .01$) et pour la nasale et la liquide chez AS (**Nas** : $t = 5.62, p < .01$; **Liq** : $t = 3.81, p < .01$). Encore une fois, chez AS l’alignement de L1 avec l’obstruante n’était pas

significativement différent entre les deux modalités intonatives.

Pour résumer jusqu'ici, l'alignement de L1 est fortement par la modalité intonative, surtout lorsque les consonnes sont nasales ou liquides, tandis que les effets de la classe consonantique varient selon la mesure d'alignement considérée (latence de L1 relativement à l'attaque syllabique/attaque de la voyelle/fin de la voyelle) et selon la modalité intonative (car, généralement, elles sont plus forts dans les questions que dans les affirmations). Les effets de la classe consonantique disparaissent uniquement lorsque l'alignement de L1 est calculé relativement à la fin de la syllabe, ce qui indique que la frontière droite de la syllabe est l'ancre la plus stable pour L1. Cela est en désaccord avec D'Imperio (2000), qui avait trouvé que l'alignement des cibles tonales de l'accent nucléaire est calculé relativement à l'attaque syllabique. On retournera sur ce point dans la discussion (par. 2.4).

En ce qui concerne la cible H, nous avons trouvé que son alignement n'est pas influencé par la classe consonantique ni chez DD ni chez AS, indépendamment du type de mesure utilisée (attaque consonantique, attaque vocalique ou fin de la voyelle), comme on peut aussi le remarquer dans la figure 2.13. Cependant, la situation est partialement différente pour L2. En particulier, dans les questions du locuteur DD, l'alignement de la cible L2 est de plus en plus retardé dans la nasale, dans la liquide et dans l'obstruante (2.15). La différence d'alignement avec la nasale vs l'obstruante est toujours significative, indépendamment du type de mesure (alignement avec l'attaque syllabique ou avec le début/fin de la voyelle).

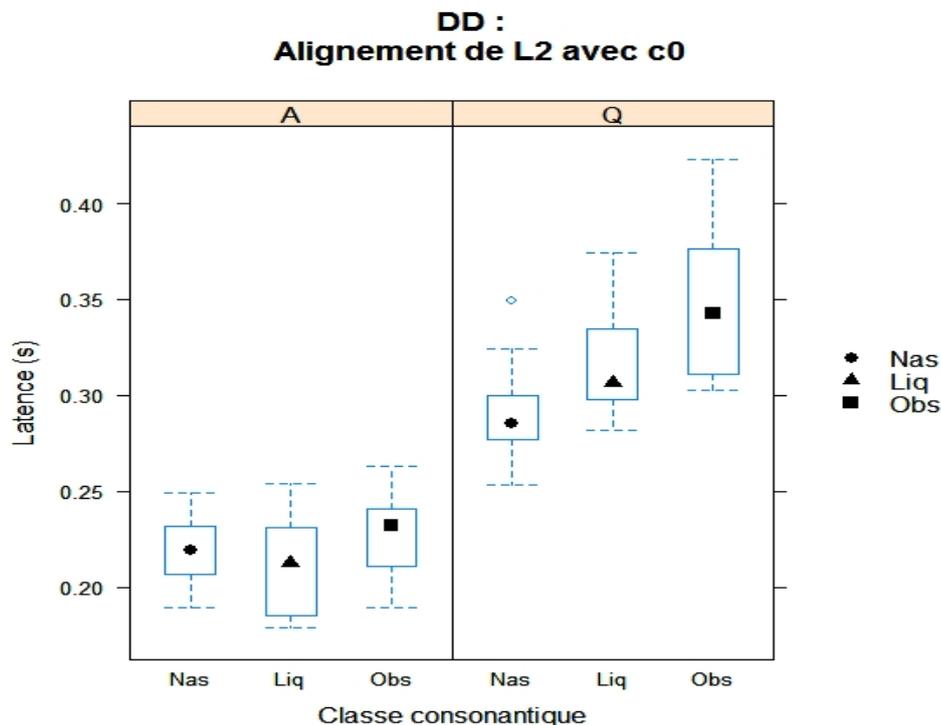


FIGURE 2.15 – Boxplots des latences de L2 par rapport au début de la consonne ($c0$) pour le locuteur DD (en haut). Les résultats sont présentés séparément en fonction de la modalité intonative et du type de consonne. La ligne tiretée marque le début de la consonne.

Pour être cohérents, nous allons rendre compte ici uniquement des résultats statistiques relatifs à l’alignement de H et L2 avec le début de la consonne. Pour la cible H, L’ANOVA omnibus confirme qu’il n’y a pas d’effets significatifs de la classe consonantique ni pour DD ($[F(2,64) = 0.06 ; p > .01]$) ni pour AS ($[F(2,69) = 0.30 ; p > .01]$). De plus, l’interaction avec la modalité intonative n’est pas significative (DD : $[F(2,64) = 3.04 ; p > .01]$; AS : $[F(2,69) = 0.02 ; p > .01]$). Les comparaisons par paires de l’analyse de régression confirment le test ANOVA, et donc ne seront pas détaillées ici. Pour L2, l’ANOVA montre chez le locuteur DD un effet principal de la classe consonantique $[F(2,64) = 9.58 ; p > .01]$, tandis que l’interaction n’est pas significative $[F(2,64) = 4.07 ; p = .02]$. L’analyse de régression nous révèle que l’effet de la classe consonantique est important uniquement dans les questions, lorsque la consonne à l’attaque consonantique est une obstruante : L2 est significativement plus retardée dans la condition **Obs** que dans **Nas** ($t = 5.04, p < .01$). Aucun effet de la classe consonantique n’a été trouvé, au contraire, dans les

affirmations (**Nas** vs **Obs** : $t = 0.45$, $p > .01$; **Nas** vs **Liq** : $t = - 0.51$, $p > .01$). Chez AS, l'ANOVA nous montre que L2 n'est pas influencée par la classe consonantique, car ni son effet principal [$F(2,69) = 0.3$; $p > .01$] ni son interaction avec la modalité intonative [$F(2,69) = 0.02$; $p > .01$] sont significatifs. Ces résultats sont confirmés par les comparaisons par paires dans le modèle de régression.

Finalement, nous avons aussi considéré l'hypothèse que l'alignement du ton *unstarred* des accents bitonaux LH est à une distance fixe de celui *starred*. Cette hypothèse est en accord Pierrehumbert (1980), selon laquelle les tons *starred* sont les seuls à être alignés avec les texte. Si cela est vrai, nous nous attendons à ce que le temps de de la montée de f_0 soit le même dans les questions et dans les affirmations, et cela indépendamment de la classe consonantique. Cependant, nous avons trouvé que le temps de montée est plus long dans les questions que dans les affirmations pour DD ($[F(1,64) = 129.6$; $p < .01]$) et AS ($[F(1,69) = 394.90$; $p < .01]$). L'ANOVA omnibus et les analyses de régression nous révèlent aussi que la classe consonantique n'est pas significative ni pour DD ni pour AS. La différence de temps de la montée ne peut pas être expliquée par de simples variations de vitesse d'élocution. En fait, dans ce cas on aurait dû s'attendre à ce que le temps de la montée soit plus long dans les affirmations, car la durée segmentale est plus importante dans cette modalité intonative (voir par. 2.3.1). Le graphique 2.13 nous donne une explication possible pour cette différence entre les deux modalités. En fait, on peut remarquer que dans les questions de DD et AS, l'excursion tonale (la différence en Hz) entre L1 et H est plus importante que dans les affirmations. Par conséquent, le mouvement de f_0 allant de L1 à H pourrait être réalisé dans un intervalle temporel plus important dans cette modalité. En ce qui concerne le temps de la descente, D'Imperio (2000) avait déjà testé l'hypothèse que le mouvement entre H et L2 ait une durée fixe, ce qui suggère que les affirmations et les questions sont caractérisées par le même accent de syntagme HL-. Ses résultats étaient mixtes : le temps de descente variait selon la modalité intonative chez un locuteur, mais il était stable chez l'autre. Par conséquent, nous avons également testé si le temps de descente est fixe ou s'il dépend de la modalité intonative/de la classe consonantique. Nos résultats sont très similaires à ceux obtenus par D'Imperio (2000). Chez DD, le temps de descente est plus important dans les questions, et en fait

la modalité intonative est significative pour ce locuteur ($[F(1,64) = 22.53; p < .01]$). De plus, l'effet de la classe consonantique ainsi que l'interaction entre classe consonantique et modalité ne sont pas non plus significatifs. La comparaison par paire de l'analyse de régression nous restitue des résultats similaires. Une hypothèse possible, suggérée par le graphique 2.13, est que le temps de descente est plus long dans les questions car l'excursion tonale entre H et L2 est plus important chez ce locuteur.

Chez AS, au contraire, nous n'avons pas trouvé de différences de modalité intonative dans le temps de la descente ($[F(1,69) = 2.28; p > .01]$). La classe consonantique ($[F(2,69) = 0.84; p > .01]$) et les interactions ($[F(2,69) = 2.03; p > .01]$) n'étaient pas non plus significatifs. Puisque le graphique 2.13 nous suggère l'existence d'une différence d'excursion tonale entre H et L2 selon la modalité intonative, il est possible que pour ce locuteur l'absence de différences dans le temps de la descente dépende d'une variation de l'inclinaison de la pente de f_0 . Si cela est vrai, on s'attendra que la pente soit plus raide dans les questions, car l'excursion tonale est plus élevée dans cette modalité intonative. Les analyses de *scaling* et de pente sont détaillées dans les deux paragraphes suivants.

2.3.4 Valeurs mélodiques des cibles tonales

Les valeurs moyennes de f_0 pour les trois cibles tonales L1, H et L2 sont montrées dans la fig. 2.13. Pour les deux locuteurs, l'excursion des mouvements montants-descendants de f_0 est plus importante dans les questions que dans les affirmations. Chez DD, cela se réalise surtout par une baisse des valeurs de L1 et L2 dans les questions par rapport aux valeurs correspondantes dans les affirmations. En particulier, les valeurs mélodiques moyennes de L1 et L2 sont respectivement de 109 Hz et de 100 Hz dans les affirmations, et de 87 Hz et 91 Hz dans les questions. La valeur mélodique de H était de 145 Hz dans les deux modalités intonatives. Chez AS, L1 et H sont respectivement plus bas et plus élevés dans les questions que dans les affirmations. Spécifiquement, L1 et H étaient à 161 Hz et à 189 Hz dans les affirmations; et à 143 Hz et à 201 dans les questions. La valeur de L2 était d'environ 143 Hz dans les affirmations ainsi que dans les questions. De plus, la classe consonantique semble jouer un rôle non seulement dans la détermination de la valeur de L1, mais aussi de H et de L2. En particulier, l'accent montant LH dans son

ensemble semble être plus élevé lorsque l'attaque consonantique est nasale, qu'en présence de la liquide ou de l'obstruante, surtout chez le locuteur DD. La valeur de L2 semble plus stable, bien que la fig. 2.13 suggère que dans les questions de DD, il y a un effet de la classe consonantique même sur cette cible.

Les analyses statistiques confirment nos observations. L'ANOVA omnibus du modèle de régression, effectué sur L1 montre un effet significatif de la modalité intonative pour DD [$F(1,64) = 304.71$; $p < .01$] et AS [$F(1,69) = 230.56$, $p < .01$]. Cela confirme les résultats déjà trouvés par D'Imperio (2000). Cependant, l'ANOVA suggère aussi que la classe consonantique peut influencer la valeur de L1, car l'effet principal de la classe est significatif chez DD [$F(1,64) = 0.04$; $p < .01$] et il est près du niveau de significativité chez AS [$F(1,64) = 3.51$; $p = .03$]. Les interactions ne sont pas significatives.

L'analyse de régression va mieux nous éclaircir sur l'existence de contrastes parmi les niveaux. En premier lieu, la différence de hauteur de L1 liée à la manipulation de la modalité intonative est significative pour les deux locuteurs et pour les trois classes de consonnes. L1 est donc plus bas dans les questions que dans les affirmations en présence de la nasale (DD : $t = -10.41$, $p < .01$; AS : $t = -7.8$, $p < .01$), de la liquide (DD : $t = -9.93$, $p < .01$; AS : $t = -8.83$, $p < .01$) et de l'obstruante (DD : $t = -2.74$, $p < .01$; AS : $t = -9.28$, $p < .01$). Cependant, la grandeur de l'effet dépend de la modalité intonative et du locuteur. La cible L1 est significativement plus élevée avec la nasale qu'avec la liquide, indépendamment de la modalité intonative chez DD ($t = 3.5$, $p < .01$). Chez AS, il n'y a pas de différences liées à la classe consonantique dans les affirmations, mais L1 est plus bas avec l'obstruante qu'avec la nasale dans les questions ($t = -3.2$, $p < .01$). Cela explique la raison pour laquelle l'effet de la classe consonantique dans l'ANOVA omnibus a un p de .03, car la baisse du valeur mélodique de L1 est plus importante uniquement avec les obstruantes.

L'ANOVA effectuée sur H nous montre aussi que la manipulation de la modalité intonative est significative pour AS [$F(1,69) = 52.65$, $p < .01$] mais pas pour DD. Le résultat obtenu pour AS est similaire à d'autres découvertes dans de nombreuses langues, selon lesquelles l'augmentation de la fréquence fondamentale, et particulièrement celle du

pic accentuel, est employée pour signaler les questions (Gussenhoven, 2004)¹⁰. L'absence d'effet de modalité sur le pic accentuel dans les énoncés de DD n'est pas surprenant, car il est peut-être le résultat d'un rehaussement de l'accent mélodique dans les affirmations suite à la double focalisation de ces énoncés (par. 2.2.2). La classe consonantique est elle aussi significative uniquement pour AS [$F(1,69) = 52.65, p < .01$]. Il n'y a pas d'interaction entre les deux facteurs. L'analyse de régression montre que l'effet de modalité chez AS n'est pas le même selon les classes consonantiques : le H est plus élevé dans les questions que dans les affirmations seulement lorsque l'attaque consonantique est une consonne nasale ($t = 5.84, p < .01$) ou une liquide ($t = 4.42, p < .01$), tandis qu'il n'y a pas d'effets avec l'obstruante ($t = 2.23, p = .02$). Les différences inter-classes, à leur tour, sont différentes selon la modalité intonative : dans les affirmations la valeur du H est indépendante de la classe, mais dans les questions, H est plus élevé en présence d'une nasale que d'une obstruante ($t = 3.72, p < .01$).

En ce qui concerne la valeur mélodique de L2, les deux facteurs expérimentaux sont significatifs uniquement chez le locuteur DD (fig. 2.15) : L2 est plus bas dans les questions que dans les affirmations [$F(1,64) = 304.71, p < .01$] et l'effet principale de la classe consonantique est significatif [$F(2,64) = 6.09, p < .01$]. Aucun effet significatif n'a été trouvé chez AS. L'effet de modalité chez DD est en réalité significatif seulement lorsque l'attaque consonantique est une consonne liquide ($t = -6.46, p < .01$) ou obstruante ($t = -6.58, p < .01$), mais ils ne sont pas significatifs avec la consonne nasale ($t = -2.05, p < .01$). De plus, l'effet de la classe consonantique est significatif uniquement dans les questions, car L1 tend à être plus basse avec un attaque consonantique nasal que liquide ($t = -4.11, p < .01$) ou obstruant ($t = -6.35, p < .01$).

Pour résumer jusqu'ici, les valeurs de L1 sont plus basses dans les questions que dans les affirmations chez les deux locuteurs, ce qui confirme les résultats de D'Imperio (2000). De plus, les valeurs mélodiques de L1 sont plus élevées en présence d'une consonne nasale en position d'attaque syllabique. Les effets segmentaux sur H et L2, bien que petits et incohérents parmi les locuteurs, soutiennent l'hypothèse que les effets microprosodiques peuvent s'appliquer sur de longues distances.

10. D'Imperio (2000) a trouvé qu'en napolitain le pic de f_0 peut être plus élevé dans les affirmations à focalisation étroite que dans les questions lorsque les affirmations sont lues sous emphase.

Nous allons maintenant détailler les résultats de l'analyse de l'excursion tonale entre L1 et H et entre H et L2. En fait, dans le paragraphe précédent, nous avons émis l'hypothèse que le temps de la montée est plus important dans les questions des deux locuteurs, car la différence mélodique entre L1 et H est plus importante dans cette modalité intonative. Nos analyses nous confirment, en fait, un effet significatif de la modalité sur l'excursion tonale pour DD [$F(1,64) = 6.77, p < .01$] et pour AS [$F(1,69) = 394.9, p < .01$]. Un effet significatif de modalité a été trouvé aussi pour l'excursion tonale entre H et L2, chez DD [$F(1,64) = 22.53, p < .01$] et chez AS [$F(1,69) = 66.77, p < .01$]. Les effets de la classe consonantique de l'attaque et les interactions ne sont pas significatifs, ni dans l'ANOVA ni dans la régression.

2.3.5 Pente de f_0

La modélisation logistique de la montée était très similaire entre DD et AS, et donc ses prédictions sont illustrées uniquement pour DD dans la fig. 2.16. Pour la descente, les résultats sont consignés dans la figure 2.17 pour DD (graphique en haut) et AS (graphique en bas). Chaque graphique montre la courbe logistique ajustée aux valeurs de f_0 (axe y) correspondantes aux 20 points de mesure (axe x) entre L1 et H (pour la montée) et entre H et L2 (pour la descente). Les résultats sont montrés séparément pour les affirmations (panneaux à gauche) et les questions (à droite), et pour les trois classes consonantiques.

Pour la montée de f_0 de L1 à H, les deux facteurs expérimentaux ont été introduits dans le modèle uniquement pour expliquer la variation des deux asymptotes (*baseline* et *topline*). En particulier, chez DD et AS, la valeur de f_0 de la *baseline* est plus basse dans les questions que dans les affirmations. Cela était attendu, car nous avons vu dans le par. 2.3.4 que la hauteur de L1 diminue dans les questions chez les deux locuteurs. De plus, chez AS la modalité influence aussi la *topline*, car chez ce locuteur le pic accentuel des questions était plus élevé que celui des affirmations. Les valeurs de f_0 de la *baseline* et de la *topline* sont plus basses avec la liquide et l'obstruante qu'avec la nasale chez les deux locuteurs. La modalité intonative et la classe des consonnes ne jouent aucun rôle sur la variation du point d'inflexion et de la vitesse de la montée, car le test de vraisemblance nous montre que l'inclusion de ces variables, faite progressivement et pour

chacun des deux paramètres, n'améliore pas l'ajustement du modèle à la courbe ($p > .01$). Ces résultats montrent donc que les valeurs de pente pour l'accent montant LH dépendent uniquement des ajustements mélodiques des cibles tonales, tout en confirmant l'existence d'effets à longue distance de la classe consonantique.

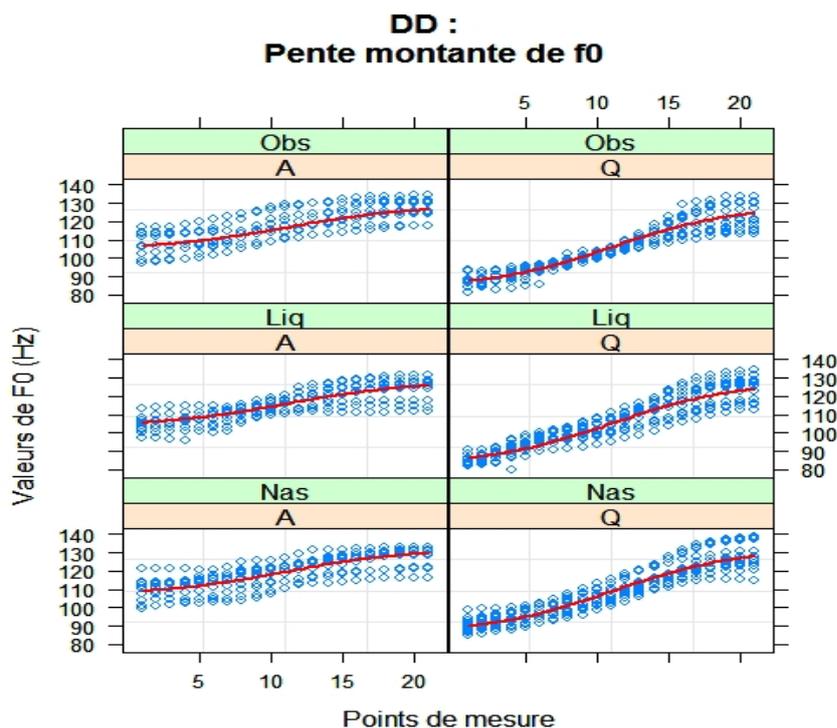


FIGURE 2.16 – Valeurs observées de la montée de f_0 (points) et valeurs attendues (lignes) résultantes de la régression logistique pour le locuteur DD.

Dans la descente de f_0 de H à L2, l'inclusion de la modalité intonative et de la classe consonantique dans le modèle semble influencer non seulement les valeurs minimales et maximales de f_0 , mais aussi les propriétés dynamiques du contour de f_0 . Chez DD (fig. 2.17, en haut), la vitesse de la descente de f_0 est influencée elle aussi par la classe de la consonne ($\log - \text{lik} = 29.7$, $p < .01$) : la descente est plus douce lorsque l'attaque syllabique est une nasale (**Nas/Liq** : $t = 5.53$, $p < .01$; **Nas/Obs** : $t = 4.01$, $p < .01$). De plus, il y a une corrélation significative entre la *baseline* et la modalité intonative et entre la *baseline* et le type de consonne : les valeurs minimales de f_0 augmentent dans les questions et en présence d'une attaque nasale dans la syllabe accentuée. La différence de pente est donc tout à fait explicable si l'on prend en compte le fait que, chez ce locuteur,

L2 était beaucoup plus élevé avec la consonne nasale qu'avec la liquide et l'obstruante. Chez AS, la modalité intonative influence la position du point d'inflexion ($log - lik = 128.1, p < .01$; voir fig. 2.17, en bas). Par conséquent, la pente est plus douce dans les affirmations que dans les questions ($t = 3.34, p < .01$). On se souviendra, en fait, que pour ce locuteur le pic accentuel était beaucoup plus élevé dans les questions que dans les affirmations, tandis que la valeur de L2 ainsi que le temps de la descente étaient les mêmes dans les deux modalités. En conclusion, les valeurs de pente de la montée et de la descente sont entièrement prévisibles à partir des valeurs mélodiques des cibles tonales.

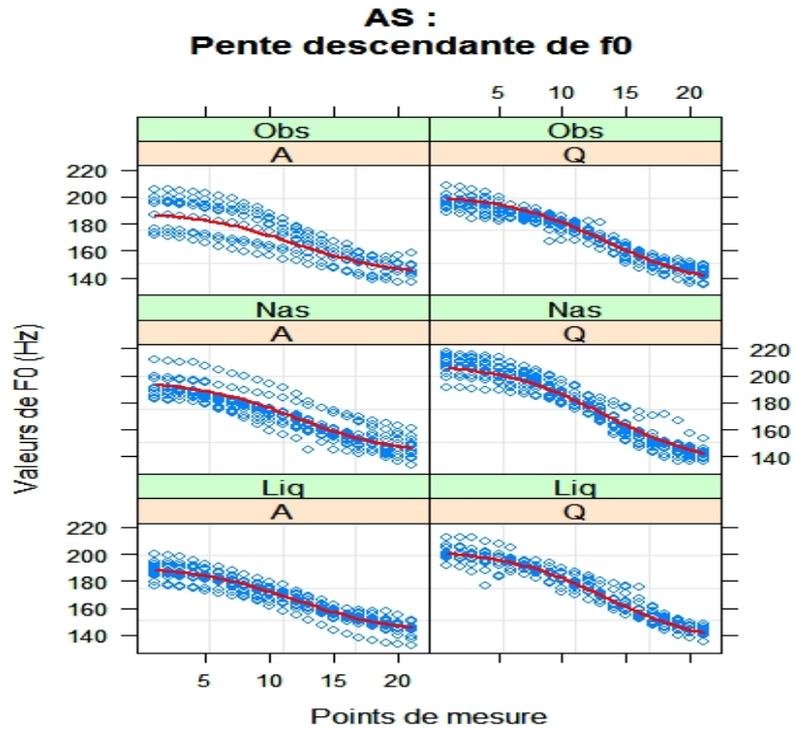
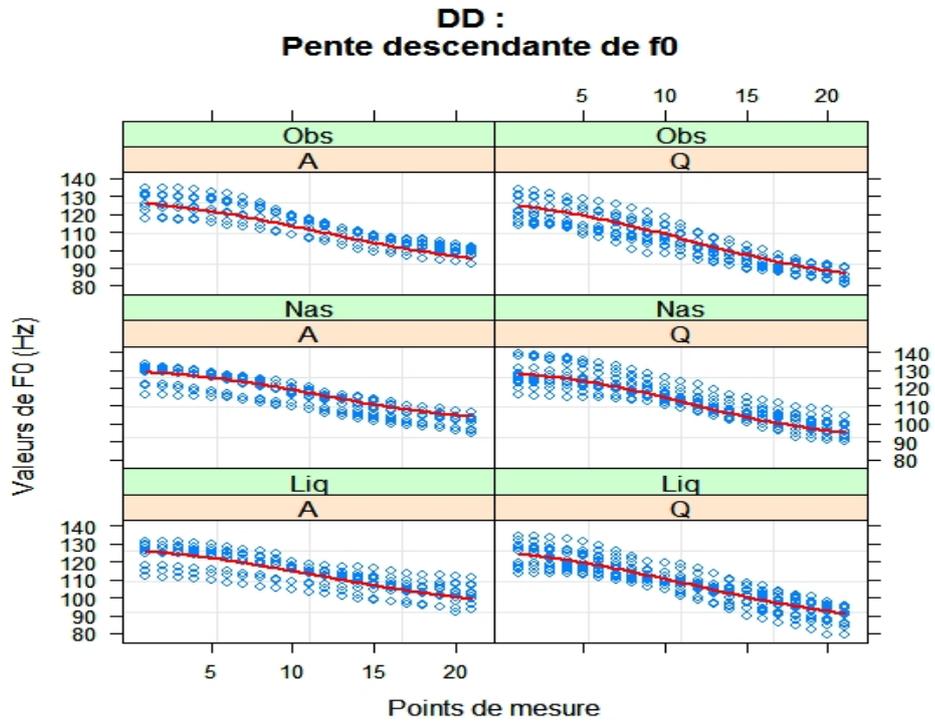


FIGURE 2.17 – Valeurs observées de la descente de f_0 (points) et valeurs attendues (lignes) de f_0 résultantes de la régression logistique pour DD (en haut) et AS (en bas).

2.4 Discussion

Les premières questions que nous voulons aborder dans cette section sont d'ordre méthodologique, et concernent : (1) la détection des cibles acoustiques et (2) les mesures choisies pour exprimer l'alignement.

Nous avons comparé dans ce chapitre trois procédures pour la détection des cibles (la méthode manuelle, LSF et MOMEL), et nous avons utilisé une combinaison des deux premières méthodes pour la localisation temporelle des cibles L1, H et L2. Les trois méthodes présentent en fait des avantages et des inconvénients. La méthode manuelle ainsi que le LSF sont très cohérents, mais la cohérence ne semble pas un critère suffisant pour juger la qualité des méthodes de détection. En fait, l'utilisation de la méthode manuelle a eu une conséquence importante au niveau phonologique : l' "effacement" de la différence d'alignement de L1 entre les questions et les affirmations. Cette différence a été déjà mise en évidence par D'Imperio (2000), qui a montré qu'en napolitain l'alignement de la cible L1 est employé systématiquement pour identifier le contraste entre affirmations à focalisation étroite et les questions oui/non. S'il y a une correspondance entre les cibles perçues et les cibles acoustiques, nous nous demandons la raison pour laquelle dans notre expérience la méthode manuelle n'a pas pu capturer la différence d'alignement de L1.

La méthode LSF semble plus fiable de ce point de vue, car ses résultats sont similaires à l'étude perceptive de D'Imperio (2000). Cependant, il semble y avoir des incohérences parmi les classes consonantiques. En fait, de façon similaire à la méthode manuelle, il n'y a pas de différences d'alignement de L1 entre questions et affirmations, lorsque la consonne en position d'attaque consonantique est une obstruante. A ce point de nos recherches, il est encore difficile d'établir si cette incohérence du LSF peut être attribuée à un erreur de détection. Une hypothèse alternative est que les perturbations microprosodiques en correspondance avec les obstruantes sont autant fortes qu'elles arrivent à effacer l'information phonologique concernant le contraste d'alignement du L. Il faut aussi noter que dans le LSF, la localisation exacte de la cible varie considérablement en fonction de la taille de la région choisie pour l'ajustement des deux lignes de régression (del Giudice *et al.*, 2007). Par conséquent, il est important de spécifier de façon explicite les frontières

de cette région, et de garder les mêmes critères pendant l'analyse acoustique. Par exemple, pour la localisation de L1 nous avons choisi une région entre le début du mot prosodique et le pic maximale de f_0 . Ce choix était guidé par des présupposés "phonologiques", car nous avons supposé que les cibles tonales des accents mélodiques tendent à être réalisées autour de la TBU (voir D'Imperio (2000) et Petrone & Ladd (2007) pour des preuves empiriques pour le napolitain).

En ce qui concerne MOMEL, il a capturé la différence d'alignement de L1 entre les questions et les affirmations, indépendamment de la classe consonantique de l'attaque. Cependant, la localisation de L1 était beaucoup plus variable parmi les modalités intonatives, parmi les mots et parmi les répétitions du même mot. Bien que nous soyons d'accord avec la "philosophie" de MOMEL (modéliser la courbe même en présence de fortes perturbations microprosodiques, utiliser une équation polynomiale pour la modélisation, qui permet aussi un ajustement meilleur à la courbe réelle), nous croyons que l'effacement drastique des perturbations microprosodiques, peut être très dangereux pour la représentation phonologique de l'intonation d'une langue spécifique. L'interaction entre micro- et macroprosodie a été en fait attestée dans plusieurs langues (Silverman (1987) pour l'anglais ; Kohler (1990) et Niebuhr (à paraître) pour l'allemand, *inter alia*). Cependant, pour nos données nous ne connaissons pas les raisons d'autant de variabilité de localisation de MOMEL (est-elle due à l'effacement de l'information prosodique, ou est-elle la conséquence de la modélisation polynomiale ?).

Nous supportons l'idée que les études sur l'intonation peuvent et doivent être basées sur des méthodes robustes de modélisation de la courbe de f_0 , et nous croyons que le développement de ces techniques doit se baser non seulement sur des critères purement acoustiques, mais aussi sur la connaissance d'autres contraintes phonologiques (de nature articulatoire ou perceptive) qui peuvent nous aider dans la représentation acoustique de l'intonation.

La deuxième question méthodologique concerne les mesures d'alignement tonal. Nous avons utilisé plusieurs méthodes, testées sur l'alignement du pic accentuel. En fait, dans la littérature on cherche souvent à modéliser la variabilité d'alignement de la cible H. En particulier, dans ce chapitre nous avons représenté l'alignement du pic en fonction de la

durée de la rime syllabique (Silverman & Pierrehumbert, 1990) ; de la durée intrasyllabique et de la classe consonantique de l'attaque syllabique (van Santen & Hirschberg, 1994, Prieto *et al.*, 1995) ; des frontières segmentales (Arvaniti *et al.*, 1998; D'Imperio, 2000; Atterer & Ladd, 2004). Les modèles de Silverman & Pierrehumbert (1990) et de van Santen & Hirschberg (1994) se sont avérés défaillants, dans la mesure où les résultats parmi les locuteurs était très différents. Chez AS, les deux modèles ont expliqué une pourcentage élevée de la variance (autour du 80%-90%) ; cependant, les deux procédures n'ont pas marché pour DD, où un modèle avec moins de paramètres et donc plus simple, a eu le même pouvoir explicatif des modèles de Silverman & Pierrehumbert (1990) et van Santen & Hirschberg (1994). Cela pourrait être dû au fait que dans notre expérience la durée segmentale était fonction essentiellement d'un facteur prosodique, la modalité intonative, tandis que la classe consonantique de l'attaque syllabique semble jouer un rôle mineur pour la durée des consonnes mêmes et des voyelles suivantes. L'existence d'un effet de la modalité intonative sur la durée segmentale n'est pas surprenant, car plusieurs études ont montré que les questions tendent à être prononcées plus rapidement que les affirmations, y compris en napolitain (D'Imperio, 2000; Haan, 2001; van Heuven & van Zanten, 2005). Par conséquent, nous avons représenté l'alignement des cibles en fonction de la distance absolue des frontières segmentales (l'attaque syllabique, le début et la fin de la voyelle).

En ce qui concerne l'alignement de L1, une découverte importante est que son alignement est plus stable lorsqu'il est mesuré par rapport à la fin de la voyelle (correspondant à la fin de la syllabe accentuée) qu'à la frontière gauche de la syllabe. En fait, lorsque l'alignement de L1 est calculé par rapport au début de l'attaque consonantique et au début de la voyelle accentuée, il est soumis à plusieurs facteurs de variabilité (la classe consonantique de l'attaque syllabique ainsi que la modalité intonative). Au contraire, lorsqu'il est calculé par rapport à la fin de la voyelle, l'effet de modalité est le seul à être significatif. Cette découverte est contraire à l'hypothèse stricte de l'ancrage segmental (Arvaniti *et al.* (1998), *inter alia*). De plus, elle est contraire à l'observation de Atterer & Ladd (2004), selon laquelle les cibles tonales sont alignées avec les frontières segmentales les plus proches (dans le cas de L1, l'attaque syllabique ou le début de la voyelle accentuée). En particulier, selon l'hypothèse de l'ancrage segmental, la cible L des accents mélodiques montants doit

être alignée de façon stable à l'attaque consonantique, et ceci indépendamment de son identité segmentale. En napolitain, cependant, la cible L des accents nucléaires LH varie en fonction de la modalité intonative : elle est plus retardée dans les questions oui/non que dans les affirmations à focalisation étroite (D'Imperio, 2000). Cela a été confirmé par notre étude. De plus, l'alignement de L1 est influencé par la classe consonantique. Les effets de la classe consonantique sont petits mais significatifs, et ils dépendent du locuteur et de la modalité intonative. En premier lieu, la différence entre les deux modalités est plus forte si l'attaque consonantique est une consonne nasale ou liquide. Aucun effet de modalité n'a été trouvé en présence des obstruantes. Comme on l'a déjà supposé dans le par. 2.3.2, cela pourrait indiquer que la microprosodie peut effacer complètement l'information macroprosodique d'alignement ou encore que LSF est plus sensible aux variations microprosodiques provoquées par l'obstruante¹¹. De plus, nous avons trouvé que dans les affirmations, l'alignement de L1 dépend de la classe consonantique, tandis qu'aucun effet n'a été trouvé dans les questions. Ces effets nous préviennent que l'alignement tonal est influencé par le plan segmental, et que les effets segmentaux ne sont pas entièrement prédictibles. De plus, notre découverte est contraire à celle de D'Imperio (2000). Dans son expérience de production, D'Imperio (2000) avait trouvé que les cibles L1, H et L2 des affirmations à focalisation étroite et des questions oui/non sont alignées relativement à l'attaque de la voyelle accentuée, donc à la frontière gauche de la syllabe. Nous rappelons que dans son expérience, les facteurs expérimentaux étaient trois : la modalité intonative (questions/affirmations), la structure syllabique (syllabe fermée/ouverte) et la classe consonantique de la consonne dans la syllabe post-tonique (nasale/vélaire). Ses résultats montrent que lorsque l'alignement est calculé relativement au début de la voyelle accentuée, l'effet de structure syllabique et celui de la classe consonantique de la consonne dans la syllabe post-tonique disparaissent. Au contraire, dans notre expérience nous avons trouvé que L1 est alignée relativement à la frontière droite de la syllabe. À part L1, dans notre expérience, la modalité intonative influence également l'alignement de la cible H et de L2. Aucun effet de la classe consonantique n'a été trouvé, ce qui suggère que l'effet

11. Le pourcentage de localisation du LSF corrigée manuellement est de 8%, ce qui exclut la possibilité que l'absence d'un contraste d'alignement dans les obstruantes est dû à l'utilisation combinée de la méthode LSF et de celle manuelle.

de l'identité segmentale de l'attaque syllabique est local. Nous supportons l'idée que les études sur l'intonation peuvent et doivent être basées sur des méthodes robustes de modélisation de la courbe de f_0 , et nous croyons que le développement de ces techniques doit se baser non seulement sur des critères purement acoustiques, mais aussi sur la connaissance d'autres contraintes phonologiques (de nature articulatoire ou perceptive) qui peuvent nous aider dans la représentation acoustique de l'intonation.

En ce qui concerne la valeur mélodique des cibles, la modalité intonative influence le *scaling* de L1, H et L2, mais les variations de hauteur dépendent du locuteur. Chez AS, la cible L1 est plus basse dans les questions que dans les affirmations, tandis que la cible H est plus élevée dans les questions. L2 ne varie pas en fonction de la modalité. Le patron observé pour l'accent montant n'est pas surprenant, car il a été déjà trouvé chez un des deux locuteurs napolitains dans l'expérience de production menée par D'Imperio (2000). De plus, notre résultat est cohérent avec l'idée que les questions sont caractérisées, d'un point de vue typologique, par une augmentation de la valeur mélodique et/ou par un alignement retardé du pic accentuel (Gussenhoven (2004), voir le chap. 5 pour plus de détails sur l'hypothèse du "Code de Fréquence"). Il faut noter que chez AS, les questions sont caractérisées par une augmentation du temps de la montée de f_0 ainsi que une excursion de f_0 très importante, ce qui est justifié par le fait que les cibles L1 et H ont des valeurs extrêmes comparées à celles des mêmes cibles dans les affirmations. Chez DD, L1 est, à nouveau, plus bas dans les questions que dans les affirmations. Cependant, chez ce locuteur les valeurs du pic accentuel sont les mêmes indépendamment de la modalité intonative. Nous avons suggéré dans le par. 2.3.3 que ce résultat peut être dû au fait que DD a prononcé les affirmations à focalisation étroite avec une double focalisation. Par conséquent, le pic mélodique des accents nucléaire peut être soumis à un *reset* de son valeur mélodique. Cependant, l'excursion de f_0 est plus élevée dans les questions que dans les affirmations. Cet effet, similaire à celui trouvé chez AS, pourrait suggérer que l'augmentation de l'excursion de f_0 est un indice acoustique secondaire de la modalité intonative. Il faut noter aussi que chez ce locuteur L2 est plus bas dans les questions, et cela se reflète dans une augmentation du temps de descente. La classe consonantique a aussi un effet sur L1, car cette cible est plus élevée en présence des consonnes nasales qu'en

présence des autres classes consonantiques. Cela est vrai dans les affirmations du locuteur AS et dans les questions et dans les affirmations de DD. De plus, la classe consonantique a un effet à longue distance sur H (chez AS) et sur L2 (chez DD), qui est cohérent avec les résultats trouvés pour L1. En fait, dans ces deux cas, la valeur mélodique de la cible est plus élevée avec les consonnes nasales qu'avec les consonnes liquides et obstruantes. Ces effets, complètement inattendus, suggèrent que les variations microprosodiques sur la f_0 ne sont pas strictement locales. Finalement, les variations de l'inclinaison de la pente sont tout à fait explicables en termes de variations de *scaling* des cibles L1, H, L2, pour DD et AS .

Effets de structure et d'alignement sur l'identification intonative et lexicale

3.1 Problématique

Dans le chapitre précédent, nous avons examiné comment différents facteurs de variabilité phonétique et phonologique peuvent influencer la réalisation de l'accent mélodique dans les questions et les affirmations en italien napolitain. Dans ce chapitre nous allons explorer l'interaction entre alignement tonal et structure syllabique sur l'identification intonative et lexicale en napolitain.

En effet, en italien napolitain, l'alignement du pic de l'accent nucléaire ne varie pas seulement en fonction du contraste intonatif, mais aussi en fonction de la durée et de la structure de la syllabe accentuée. Nous savons déjà que, bien que la durée de la voyelle ne soit pas contrastive en italien, les voyelles toniques en syllabe ouverte sont plus longues qu'en syllabe fermée, et cela est vrai surtout lorsque le mot est paroxyton (D'Imperio & Rosenthal, 1999). De plus, D'Imperio (2000) a montré que les cibles tonales de l'accent nucléaire sont plus retardées en syllabe fermée qu'en syllabe ouverte (voir aussi Petrone (2005) pour un contraste similaire lié à un phénomène de sandhi externe). En particulier, le

pic accentuel des accents nucléaires LH est localisé différemment dans des paires lexicales minimales en contraste de longueur consonantique et de structure syllabique, telles que *nonno* (“grand-père”) vs *nono* (“neuvième”) ou *palla* (“ballon”) vs *pala* (“pâle”). Dans ces paires, le premier mot contient une consonne géminée (représentée dans l’orthographe par la double lettre), tandis que le deuxième mot contient une consonne simple (représentée par une seule lettre). En effet, la longueur consonantique est phonologique en italien. En outre, la voyelle à l’intérieur de la syllabe fermée dans le mot avec la consonne simple est plus courte que celle à l’intérieur de la syllabe fermée dans le mot avec la géminée. L’interaction du contraste intonatif et structural se traduit par quatre types d’alignement possibles (D’Imperio, 2000). Dans les affirmations, le pic H de l’accent nucléaire LH est aligné avec le début ou le centre de la voyelle en présence, respectivement, d’une syllabe fermée ou d’une syllabe ouverte. Dans les questions, il est localisé vers la fin de la voyelle si la syllabe est ouverte ou dans la coda consonantique si la syllabe est fermée (voir figures 1.4 et 1.5).

Les effets de la structure syllabique sur l’identification intonative ne sont pas encore clairs. D’Imperio (2000), en manipulant l’alignement du contour montant-descendant avec des syllabes fermées et ouvertes, a mis en évidence qu’en napolitain la structure syllabique ne provoque pas un déplacement de la frontière catégorielle entre les questions et les affirmations. Elle en conclut que l’alignement perceptif des cibles L et H de l’accent nucléaire est relatif à la frontière gauche de la syllabe accentuée. Cependant, elle rend compte aussi d’un effet de structure syllabique sur la forme de la fonction des réponses, effet qui semble d’interprétation difficile : le score d’identification des questions augmente plus rapidement dans le mot ayant la syllabe ouverte (*nono*) que dans le mot ayant une syllabe fermée (*nonno*) avant le *cross-over point*, tandis que cette tendance est inversée après le *cross-over point*. Il faut noter aussi que dans cette étude la durée segmentale des stimuli n’a pas été manipulée, celle-ci correspondant à la durée des productions naturelles. Par conséquent, un problème potentiel avec ces résultats est que nous ne savons pas si la différence perceptive de durée des stimuli employés était suffisamment important pour provoquer un déplacement de la frontière catégorielle entre les deux modalités.

En napolitain, l’alignement tonal semble influencer aussi l’identification lexicale. D’Im-

perio *et al.* (2007) ont montré que les auditeurs napolitains exploitent l'information d'alignement pour l'identification de la paire minimale *nonno/nono*, mais uniquement lorsque la durée segmentale est ambiguë. Dans des stimuli jugés comme questions, l'alignement du pic accentuel avec la coda consonantique provoque une augmentation de réponses "*nonno*". En effet, cet alignement est plus cohérent avec celui trouvé dans les syllabes fermées. Dans les stimuli jugés comme affirmations, les résultats de l'identification étaient moins clairs : contrairement aux attentes, le score de réponses "*nonno*" étaient plus bas lorsque le pic était plus retardé dans la voyelle. Il faut noter que dans cette étude, la base intonative de l'énoncé naturel était une question. Par conséquent, une explication de ce résultat est que l'identification lexicale pour les stimuli jugés comme affirmations a pu être biaisée par certaines propriétés du stimulus de base employé pour les manipulations de durée et d'alignement. En effet, l'existence de différences dans les caractéristiques spectrales des questions et des affirmations a déjà été suggérée par D'Imperio (2000).

Dans ce chapitre, nous allons rendre compte de deux expériences de perception dont le but est de mieux évaluer l'impact de la structure syllabique sur l'identification intonative et, dans le même temps, l'impact de l'alignement sur l'identification lexicale. Dans l'Expérience I, nous avons modifié l'alignement tonal dans des stimuli qui contenaient une syllabe ouverte ou fermée. Dans l'Expérience II, la même modification d'alignement a été appliquée à une série de stimuli contenant une syllabe ambiguë par durée segmentale et par structure syllabique. Pour l'identification intonative, en accord avec D'Imperio (2000) notre hypothèse nulle est que la structure syllabique et la durée segmentale n'influencent pas la perception du contraste entre affirmations et questions. Si cela est vrai, le score d'identification sera le même dans l'Expérience I et II, indépendamment de la modification de durée segmentale et de structure syllabique. Au contraire, si la durée segmentale joue un rôle dans l'identification intonative, ses effets seront visibles uniquement dans l'Expérience I, où les stimuli utilisés étaient clairement jugés par les auditeurs comme ayant une syllabe fermée ou ouverte. Aucun effet ne sera attendu dans l'Expérience II, où les stimuli sont ambigus d'un point de vue perceptif. De plus, dans l'Expérience I, l'identification lexicale ne sera pas influencée par la manipulation d'alignement, car nous savons qu'en italien la gémination est signalée principalement par la durée de la coda consonantique

et de la voyelle tonique précédente (Esposito & di Benedetto, 1999). Par contre, puisque dans l'Expérience II la structure syllabique est ambiguë, nous supposons que les auditeurs napolitains vont utiliser les détails d'alignement comme indice secondaire pour l'identification lexicale. En accord avec D'Imperio *et al.* (2007) nous nous attendons à ce que le score de réponses “*nonno*” soit de plus en plus élevé en fonction de l'alignement retardé du pic de f_0 dans les stimuli jugés comme questions et dans ceux jugés comme affirmations. De plus, si l'identification lexicale dépend aussi de la base intonative de l'énoncé, l'identification de *nonno* dans les stimuli identifiés comme affirmations sera plus cohérent avec le patron d'alignement trouvé dans les productions naturelles, surtout lorsque la base de l'énoncé est elle aussi une affirmation.

Une tâche préliminaire a été effectuée pour trouver (1) des stimuli clairement jugés par les auditeurs comme ayant une syllabe ouverte ou fermée, qui ont été employés comme base lexicale pour les modifications d'alignement dans l'Expérience I ; (2) des stimuli ambigus d'un point de vue perceptif en termes de durée segmentale et de structure syllabique, qui ont été employés dans l'Expérience II.

3.2 Corpus de base

Un locuteur masculin d'italien napolitain (CC) a produit deux mots en contraste de longueur consonantique et de structure syllabique *nonno* (“grand-père”) vs *nono* (“neuvième”) dans trois modalités intonative différentes (affirmations à focalisation large et étroite ; questions oui/non). Les mots ont été réalisés sous forme de citation ou insérés en position finale dans la phrase porteuse courte : “E' il [MOT-CIBLE]”. Le mot *nonno* (/ˈnɔnno/) contient une consonne géminée, et sa syllabe tonique est fermée (non-); le mot *nono* (/ˈnono/), contient une consonne simple et sa syllabe tonique est ouverte (no-). Lorsque ces mots ont été prononcés sous forme de citation, ils portaient un accent nucléaire H+L* sur leur syllabe tonique, qui est typique des affirmations à focalisation large en napolitain (D'Imperio, 2002a; Grice *et al.*, 2005). Au contraire, les phrases courtes ont été lues comme affirmations à focalisation étroite sur le mot-cible ainsi que comme questions oui/non. Par conséquent, dans ces phrases le mot-cible pouvait porter un accent

nucléaire $L+H^*$ vs L^*+H . Tous les énoncés étaient prononcés avec les tons de frontières $L-$ $L\%$.

La fig. 3.1 illustre les deux réalisations des mots-cible sous forme de citation. Dans les deux stimuli, le H et le L de l'accent $H+L^*$ sont réalisés entre les frontières de la syllabe accentuée, mais les détails d'alignement varient entre les deux mots, car les deux cibles sont plus retardées dans *nonno* que dans *nono*.

Les figures 1.4 et 1.5, dans l'introduction de cette thèse, nous donnent un exemple de l'interaction entre le contraste de structure syllabique (*nonno/nono*) et celui de modalité intonative ($L+H^*/L^*+H$) sur l'alignement du pic de f_0 .

Comme on s'y attendait, le pic des affirmations à focalisation étroite ($L+H^*$) est globalement plus avancé que celui des questions oui/non (L^*+H) : il est aligné au début de la voyelle tonique dans les affirmations et autour sa fin dans les questions. De plus, l'alignement précis de H dépend de la structure de la syllabe accentuée. Dans les affirmations, le H est respectivement 205 ms et 137 ms avant la fin de la voyelle tonique dans *nono* et dans *nonno*. Dans les questions, le H se produit 30 ms avant la fin de la voyelle tonique dans *nono*, et tout de suite après cette frontière dans *nonno*.

Les deux mots *nonno* et *nono*, prononcés sous forme de citation, ont constitué la base des manipulations perceptives dans l'expérience préliminaire décrite dans le par. 5.3, tandis que les phrases courtes prononcées comme affirmations à focalisation étroite et les questions oui/non (fig. 1.4 et 1.5) ont été employées comme base pour les Expériences de perception I et II. Les manipulations effectuées sur les six énoncés concernaient la durée segmentale et l'alignement tonal. Ces manipulations, obtenues grâce à l'utilisation de PRAAT (Boersma & Weenink, 2007), sont détaillées dans les sections suivantes.

3.3 Tâche préliminaire

3.3.1 Préparation des stimuli

La manipulation de durée s'appuie sur deux présuppositions. En premier lieu, nous parlons du principe qu'en italien le seul indice segmental pour la gémation consonantique est la durée. Pour les consonnes nasales, cette affirmation semble être soutenue par une

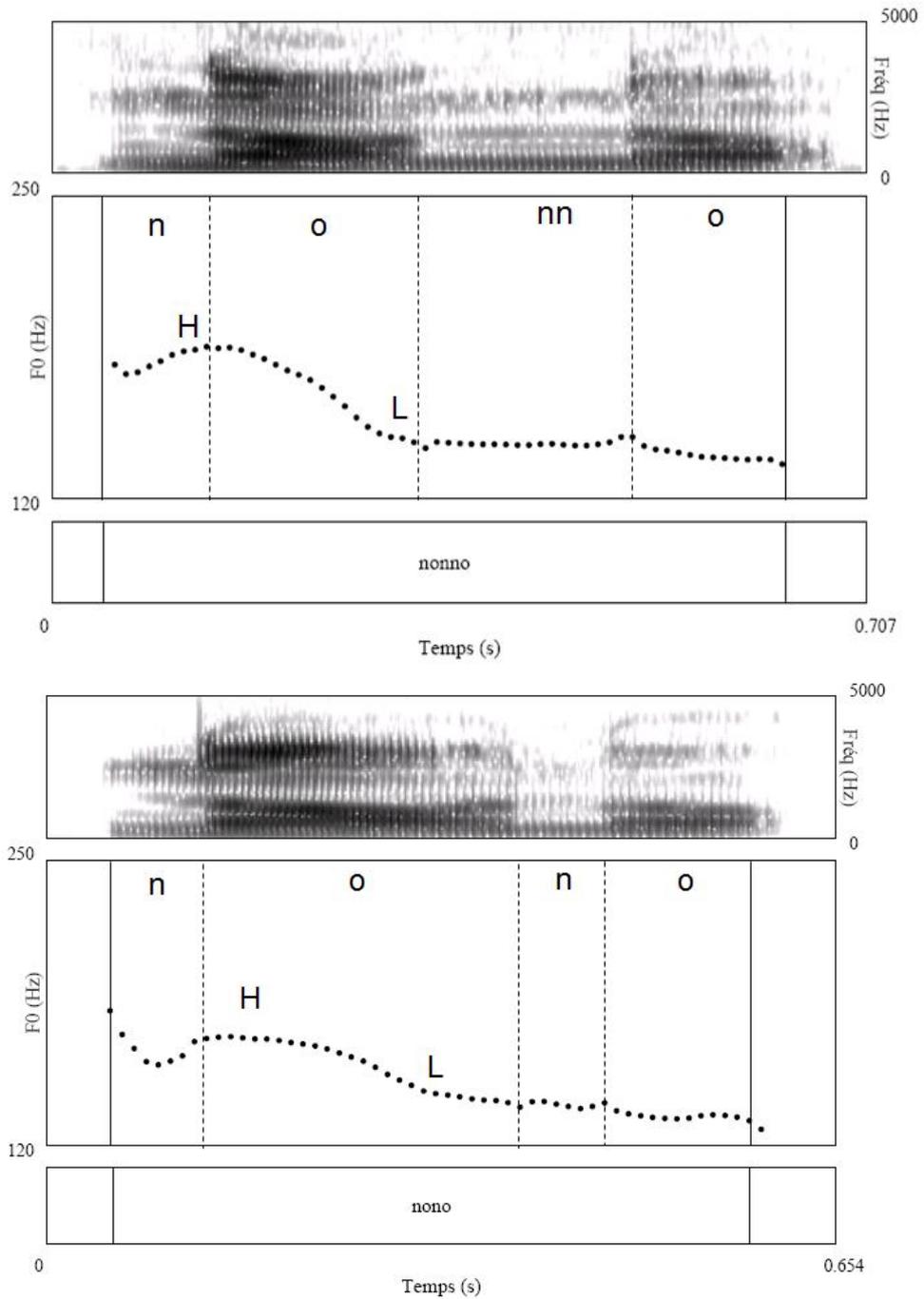


FIGURE 3.1 – Spectrogramme, courbe de f_0 et cibles tonales pour les formes de citation *nonno* et *nono*, prononcée avec une focalisation large.

étude préliminaire effectuée par Mattei & di Benedetto (2000). Une autre pré-supposition est que la variation compensatoire des valeurs de durée entre la coda consonantique et la voyelle tonique précédente est linéaire. Quelques preuves pour l'existence d'une relation linéaire de durée entre les deux segments ont été données pour d'autres langues, telle que le japonais (Hirata & Forbes, 2007), qui montre des phénomènes de covariation compensatoire similaire à l'italien. De plus, Mattei, dans un travail non encore publié sur la perception de la gémination en italien, montre qu'en italien une telle covariation peut être décrite par un simple modèle de régression linéaire.

Dans l'expérience préliminaire, une série de 18 stimuli a été créée à partir des deux stimuli de base illustrés dans la fig. 3.1. En particulier, la durée de la voyelle accentuée a été modifiée en neuf étapes (ou "pas") grâce à la méthode déjà utilisée par Mattei. D'abord, nous avons calculé les valeurs moyennes de durée de la coda consonantique (C) et de la voyelle précédente (V) à travers les trois modalités intonatives, séparément pour les deux mots-cible (*nonno* : $V = 180$ ms, $C = 175$ ms ; *nono* : $V = 240$ ms, $C = 85$ ms). Les valeurs 25 et 75 ont été affectées au paramètre \mathbf{p} (représentant l'intercept dans l'équation de régression linéaire) avec les valeurs indiquées ci-dessus respectivement pour *nonno* et *nono*. Nous avons donc estimé les droites qui passent par les points (25, V), (75, C) pour obtenir les valeurs de durée pour le mot avec la gémignée, et les points (75, V) et (25, C) pour le mot avec la consonne simple. De cette façon, il nous a été possible de calculer la durée de V et de C pour chaque valeur de \mathbf{p} . En d'autres termes, lorsque la valeur de \mathbf{p} était fixée à 25, nous avons obtenu la durée pour la consonne simple, tandis que \mathbf{p} était égal à 75, nous avons obtenu les valeurs de la consonne gémignée. Au contraire, en variant les valeurs de \mathbf{p} entre celles de 25 et de 75, nous avons trouvé des valeurs de V et C intermédiaires entre celles du mot avec la consonne simple et la consonne gémignée. Comme on peut le remarquer dans la Table 3.1, les valeurs moyennes de V et C pour le mot contenant la consonne gémignée et la consonne simple correspondent respectivement aux stimuli en D2 et D9 du continuum.

p	Durée	V1d	Δ V1d	Cd	Δ Cd
17	D1	250		70	15
25	D2	240	-10	85	15
33	D3	230	-10	100	15
42	D4	220	-10	115	15
50	D5	210	-10	130	15
58	D6	200	-10	145	15
67	D7	190	-10	160	15
75	D8	180	-10	175	15
83	D9	170	-10	190	15

TABLE 3.1 – Valeurs de **p** et valeurs de durée de la voyelle accentuée (**V1d**) et de la coda consonantique (**C1d**) pour les pas de manipulation D1-D9 (ms). Les colonnes Δ **V1d** et Δ **Cd** indiquent la différence de durée entre chaque pas de manipulation et le successif, respectivement pour la voyelle et pour la consonne.

Ensuite, nous avons stylisé le contour de f_0 par une ligne droite qui allait d’un point initial de 170 Hz (dont la localisation temporelle était au début de chaque mot) à un point final de 130 Hz (situé à la fin de chaque mot). De telles valeurs ont été choisies car intermédiaires entre celles trouvées dans les stimuli naturels *nonno* et *nono* pour ce locuteur. La différence mélodique entre les deux points nous a servi pour reproduire le phénomène de déclinaison de la f_0 , tandis que l’aplatissement du contour de f_0 a permis d’exclure des influences possibles d’alignement de l’accent mélodique sur l’identification lexicale. Tous les stimuli ont donc été resynthétisés à l’aide de PSOLA (*Pitch Synchronous Overlap and Add*, Carpentier & Moulines, 1990). Les différences d’amplitude ont été neutralisées par la normalisation de l’intensité à une valeur moyenne mesurée pour ces stimuli.

3.3.2 Déroulement de l’expérience

L’ensemble des stimuli a été reproduit sur un ordinateur portable à l’aide de PERCEVAL, un logiciel utilisé pour effectuer des expériences auditives et visuelles et développé au Laboratoire Parole et Langage par André *et al.* (2007). La tâche a été précédée par une

phase d'entraînement dans laquelle les auditeurs devaient identifier dix essais sélectionnés sur l'ensemble des stimuli en tant que représentatifs des deux catégories lexicales *nonno* et *nono*. Tous les stimuli ont été présentés de façon binaurale par des casques professionnels (Senneheizer HD 497) dans un bureau à l'Université "Federico II" de Naples. Dans le même temps, les deux mots étaient présentés visuellement sur l'écran de l'ordinateur dans le même ordre (1. *nonno*, 2. *nono*) pour éviter des incertitudes dans les réponses dues à des changements d'ordre. Les stimuli ont été écoutés 7 fois dans 7 blocs randomisés. Ils étaient séparés par une pause de 4 secondes, et une pause de 60 secondes était insérée entre un bloc et le suivant. Nous avons demandé à sept sujets napolitains d'écouter les stimuli et d'identifier un des deux mots de la paire minimale. Plus précisément, ils devaient indiquer leur choix en pressant sur le clavier la flèche droite pour *nonno*, et la flèche gauche pour *nono*. Tous les auditeurs étaient droitiers. L'expérience durait moins de 10 minutes.

3.3.3 Résultats

Les résultats de l'identification lexicale pour les stimuli créés à partir de la base avec la syllabe fermée et de la base avec la syllabe ouverte sont illustrés dans la fig.3.2. Le score moyen de réponses "*nonno*" (axe y), regroupé pour tous les auditeurs, est représenté en fonction des 9 pas de durée (axe x). Les deux lignes à l'intérieur du graphique montrent les réponses en fonction de la base du stimulus (*nonno* = cercles; *nono* = triangles). Comme on s'y attendait, le pourcentage de réponses augmente de *nono* à *nonno* en fonction de la manipulation de durée : aux emplacements avancés à l'intérieur du continuum (D1 - D4) la réponse "*nono*" est dominante, tandis qu'à l'opposé du continuum (D9 - D10) la réponse "*nonno*" atteint le score maximal. Ces résultats sont en accord avec la littérature sur les effets de durée sur l'identification de la gemination (Esposito & di Benedetto, 1999). De plus, la fig. 3.2 suggère un effet du stimulus de base, car le pourcentage d'identification de *nonno* à travers les pas de durée est plus basse dans les stimuli créés à partir de la base *nono* que dans ceux créés à partir de *nonno*.

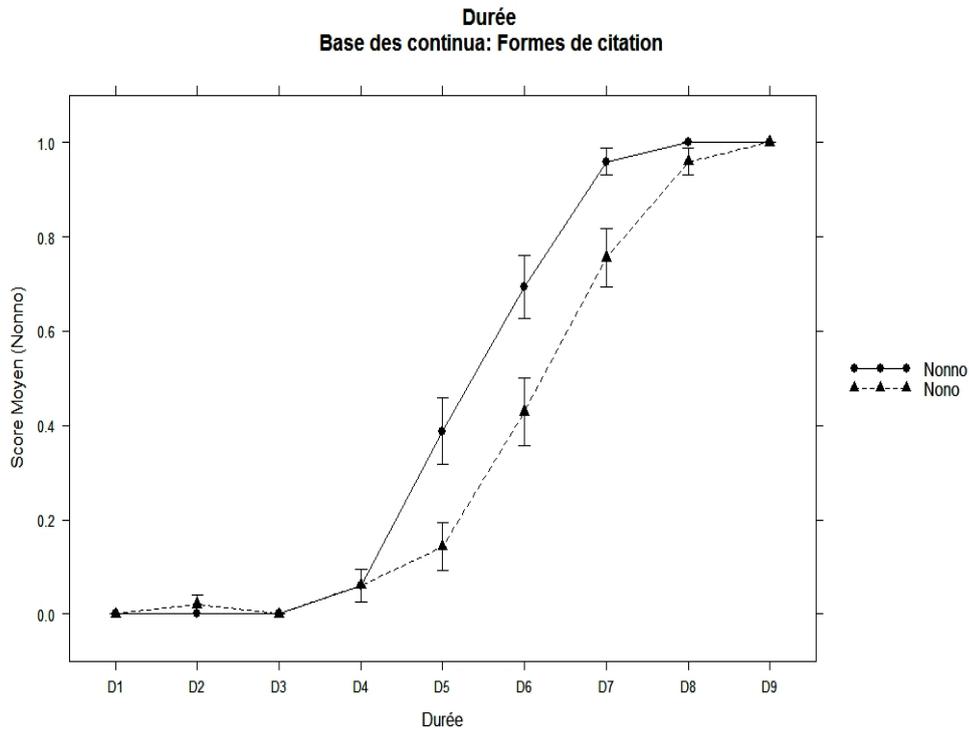


FIGURE 3.2 – Score moyen de nonno pour la série des stimuli créée à partir du mot *nonno* et *nono* pour tous les sujets.

Afin d'évaluer l'effet de la durée sur l'identification lexicale, nous avons ajusté un modèle de régression logistique aux données. En perception, le modèle logistique basé sur la distribution binomiale du score d'identification (modèle de régression généralisé) est souvent utilisé pour modéliser des jeux de données dans lesquels la variable dépendante est binaire (Baayen, à paraître). En particulier, nous avons d'abord utilisé un modèle de régression logistique à effets mixtes, dans lequel les facteurs de traitement étaient la durée et la base lexicale du stimulus, tandis qu'un intercept aléatoire a été introduit pour chaque auditeur. Puisque la variance parmi les auditeurs était proche de zéro, le facteur aléatoire a été enlevé. Par conséquent, le modèle logistique avait comme variables indépendantes uniquement la durée et la base lexicale du stimulus. Les résultats révèlent un effet significatif de la durée [$z= 13.71, p < .01$] et de la base lexicale [$z= -4.42, p < .01$]. En particulier, le calcul du *crossover point* a montré que la frontière catégorielle est située autour de D5 pour la base *nonno* (frontière moyenne = 5.04) et de D6 pour *nono* (6.09).

3.3.4 Discussion

La tâche préliminaire confirme des résultats déjà connus dans la littérature sur la production et perception de la gémiation en italien : la durée de la consonne et de la voyelle précédente sont les indices principaux pour la distinction de paires minimales lexicales en contraste de longueur consonantique et de structure syllabique (Esposito & di Benedetto, 1999). De plus, nous avons trouvé un effet de la base lexicale du stimulus, qui suggère l'existence d'autres indices segmentaux que la durée pour signaler la gémiation. La nature de ces indices va cependant au delà des buts de cette expérience, qui a été effectuée pour trouver des stimuli utilisables comme base lexicale pour les manipulations d'alignement dans les Expériences de perception I et II. Notamment, dans l'Expérience I nous allons tester l'effet de l'alignement tonal sur l'identification intonative et lexicale dans des syllabes clairement perçues comme fermées/ouvertes, tandis que dans l'Expérience II la base lexicale sera ambiguë entre *nonno* et *nono*. Pour l'Expérience I, nous avons modifié la durée de la syllabe accentuée de *nono* et *nonno* insérés respectivement dans une affirmation à focalisation étroite et dans une question oui/non. Cette modification a été faite afin d'obtenir des stimuli dont la valeur segmentale correspondait à celle à D2 et D8, respectivement pour *nono* (*nono*/D2) et *nonno* (*nonno*/D8). En fait, la tâche préliminaire nous a montré que dans ces pas d'alignement, les stimuli étaient perçus clairement comme l'un des mots de la paire minimale lexicale. De plus, les stimuli étaient très ambigus entre les deux catégories lexicales à D5 pour la base *nonno* (*nonno*/D5) et à D6 pour la base *nono* (*nono*/D6). Les valeurs de durée de ces stimuli seront donc utilisées pour les manipulations de durée dans l'Expérience II.

3.4 Expérience I

3.4.1 Préparation des stimuli

Comme il a déjà été dit, les énoncés dans les figures 1.4 et 1.5 ont été la base de la manipulation de la durée et de l'alignement tonal dans l'Expérience I et II. Dans l'Expérience I, nous avons d'abord modifié la durée de la coda consonantique et de la

voyelle accentuée dans les deux mots *nonno* et *nono* afin de créer des stimuli similaires en durée à ceux obtenus pour les pas de durée D2 (*nono/D2*) et D8 (*nonno/D8*) de l'Expérience préliminaire. Ensuite, les valeurs temporelles des cibles nucléaires L et H ainsi que de l'accent de phrase L- ont été modifiées. Pour les quatre stimuli obtenus par la modification de durée (2 pas de durée x 2 bases lexicales), nous avons effectué une stylisation linéaire du contour de f_0 dans laquelle 5 points ont été interpolés : un point au début de l'énoncé ; trois points, respectivement pour L (L1), le sommet H (H) et le deuxième L (L2) dans la configuration montante-descendante du contour ; et un point à la fin de l'énoncé. La hauteur mélodique des 5 points a été fixée à une valeur moyenne mesurée pour les énoncés réalisés comme questions et comme affirmations à focalisation étroite par le locuteur CC. La valeur de f_0 au début de la phrase et celle de la cible L1 ont été fixées à 160 Hz pour éviter des effets éventuels de la valeur mélodique de la tête du contour. Les valeurs de H et de L2 ont été fixées respectivement à 200 et à 145 Hz, et la valeur finale de f_0 à 130 Hz. Les valeurs fixes de f_0 nous ont permis de contrôler des effets éventuels de hauteur de la f_0 .

Finalement, les trois cibles tonales L1, H et L2 ont été déplacées simultanément à droite et à gauche dans le temps afin d'obtenir le même nombre de questions et d'affirmations. Tous les déplacements ont été effectués par rapport à la fin de la voyelle. En particulier, l'alignement du contour montant-descendant a été déplacé en quatorze pas de 15 ms (A1-A14, fig. 3.3). En A7 (c'est à dire, à la moitié du continuum) le pic de f_0 est de 45 ms avant la fin de la syllabe, et donc son alignement est plus similaire à celui trouvé pour les questions contenant une syllabe ouverte dans le stimulus naturel. Par conséquent, nous supposons que la région segmentale autour de A7 pourrait être aussi la frontière entre les affirmations et les questions. Nous nous sommes aussi assurés que les manipulations d'alignement ne produisent pas de stimuli qui pourraient sembler artificiels aux auditeurs.

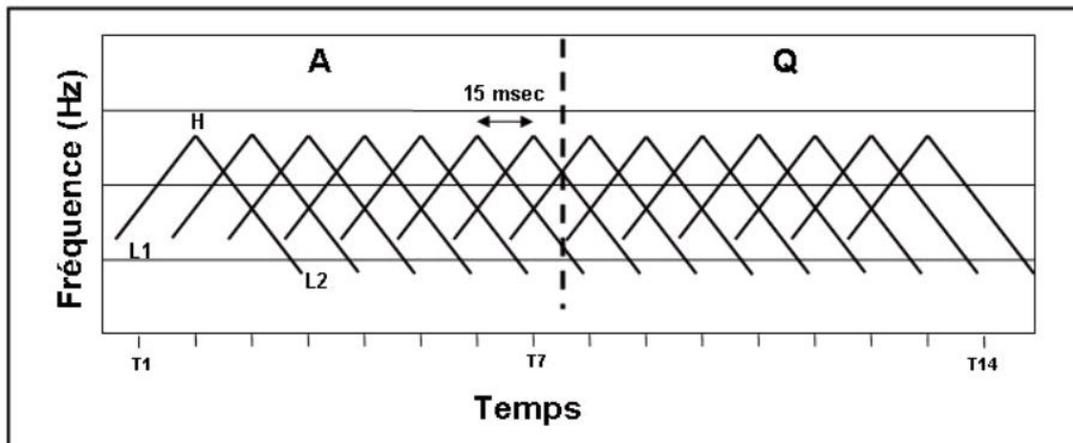


FIGURE 3.3 – Manipulation de l'alignement.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
L1	-305	-290	-275	-260	-245	-230	-215	-200	-185	-170	-155	-140	-125	-110
H	-135	-120	-105	-90	-75	-60	-45	-30	-15	0	15	30	45	60
L2	35	50	65	80	95	110	125	140	155	170	185	200	215	230

TABLE 3.2 – Valeurs d'alignement de L1, H et L2 par rapport à la fin de la voyelle.

Les 56 stimuli, créés par la combinaison des deux bases intonatives (A vs.Q), des deux bases lexicales (*nonno*/D8 vs.*nono*/D2), et des quatorze pas d'alignement (A1-A14), ont été resynthésés à l'aide de PSOLA. Ce corpus a été la base de l'Expérience I. Les valeurs de l'alignement temporel sont présentées dans la Table 3.2.

3.4.2 Déroulement de l'expérience

Les stimuli, reproduits sur un ordinateur portable grâce au logiciel PERCEVAL, ont été écoutés deux fois dans deux blocs randomisés, chaque stimulus étant présenté une fois par bloc. Un groupe de vingt auditeurs napolitains a participé à l'expérience. Deux sous-groupes de dix personnes ont été établis selon l'ordre des tâches à accomplir. Le premier sous-groupe devait écouter le premier bloc de stimuli et les identifier comme étant une question ou une affirmation. Ensuite, les même stimuli devaient être identifiés comme *nonno* ou *nono*. Le deuxième devait accomplir les même tâches, mais dans un ordre

contraire : les auditeurs devaient d’abord identifier les stimuli quant au contraste lexical et ensuite quant au contraste intonatif. Une telle procédure expérimentale a été préférée à un jugement forcé à quatre choix (*four-alternative forced-choice task*, c’est à dire une tâche dans la quelle les auditeurs auraient dû identifier les stimuli quant au contraste intonatif et lexical simultanément). Une telle procédure nous a permis donc d’éviter des incertitudes dans la décision dues à la complexité de la tâche. Chaque bloc était précédé de quatorze essais d’entraînement, qui ont été choisis car représentatifs du contraste d’alignement entre questions et affirmations pour chacune de deux bases lexicales et pour chacun de deux pas de durée. Après l’écoute de chaque stimulus, les auditeurs devaient faire leur choix en pressant sur la flèche droite (pour *nonno* ou pour les questions) ou sur la flèche gauche (pour *nono* ou pour les affirmations) du clavier. Tous les auditeurs étaient droitiers, et ils avaient entre 20 et 30 ans. Aucun des auditeurs n’avait de problème connu d’audition. Les étiquettes *1. domanda, 2. affermazione* (“question/ affirmation”) et “*1. nonno, 2. nono*” (“grand-père/ neuvième”) étaient aussi visualisées sur l’écran de l’ordinateur. L’expérience durait environ 30 minutes.

3.4.3 Résultats

La figure 3.4 rend compte des résultats de l’identification du contraste intonatif, regroupés pour les vingt auditeurs. Les résultats sont tracés pour les continua dont la base intonative était une question (en haut) ou une affirmation (en bas). Dans chaque graphique, le score moyen des questions (axe des ordonnées) est fonction des pas d’alignement (axe des abscisses), séparément pour les bases lexicales *nono/D2* (points) et *nonno/D8* (triangles).

Nos données confirment des résultats déjà connus dans la littérature : l’identification des affirmations/questions dépend de la localisation temporelle du contour montant-descendant de f_0 (D’Imperio, 2000). Plus particulièrement, l’identification des questions augmente lorsque L1, H et L2 sont déplacés vers la droite du continuum, et elle diminue lorsqu’ils sont déplacés vers la gauche.

Le score de réponses correctes de 0.8 (ou 80%, en pourcentage) est utilisé normalement comme un critère adéquat pour la réussite d’un test. Dans notre expérience, nous

avons adopté ce critère pour l'identification des questions, et parallèlement, celui de 0.2 pour l'identification des affirmations. Par conséquent, nous avons trouvé que les auditeurs tendaient à identifier comme affirmations les stimuli en A1 - A2 (le score moyen pour chaque pas étant inférieur à 0.2 (ou 20%, en pourcentage) pour tous les quatre continua) tandis que les stimuli entre A9 et à A14 étaient clairement jugés comme des questions. En correspondance avec ces pas d'alignement, la fig. 3.4 nous montre aussi qu'il n'y a pas de différence dans le score d'identification en fonction de la structure syllabique. Cependant, les résultats obtenus pour les stimuli en A3 - A8 nous montrent une situation différente de celle attendue. Nous rappelons que l'hypothèse nulle est que le score moyen des questions est indépendant de la structure de la syllabe accentuée. Cependant, en A3 - A8 le score d'identification varie selon la base lexicale. En particulier, en A3 le score d'identification "question" est déjà autour de 0.4 pour *nono*/D2, donc proche du niveau de hasard. Ce score est aussi indépendant de la base intonative. En correspondance avec le même pas d'alignement, le score d'identification est beaucoup moins élevé pour *nonno*/D8. Les fig. 3.4 nous montre aussi qu'à partir d'A3, la fonction de réponse est plus raide dans les stimuli avec la syllabe ouverte que dans ceux avec la syllabe fermée. Par conséquent, en A8 (où le pic de f_0 était placé 30 ms avant la fin de la voyelle), le score d'identification a déjà atteint 0.8 pour les séries créées à partir de *nono*/D2, indépendamment de la base intonative de l'énoncé. Pour *nonno*/D8 il atteint le score de 0.8 dans les questions, mais il est encore autour de 0.65 pour les affirmations.

Nous avons effectué une régression logistique à effets mixtes, dans laquelle la manipulation de l'alignement tonal, de la base lexicale et de la base intonative constituaient les effets fixes, et les auditeurs le facteur aléatoire. Les résultats montrent un effet significatif de l'alignement tonal ($z = 10.6$, $p < .01$) ainsi que de la base lexicale ($z = 4.52$, $p < .01$). La manipulation de la base intonative, au contraire, n'est pas significative ($z = 0.12$, $p = .9$). Cela confirme que le déplacement à droite du sommet H porte les auditeurs à la catégorisation des stimuli en tant que questions, et que le score d'identification est plus élevé lorsque la syllabe accentuée est ouverte.

En ce qui concerne l'identification lexicale, notre hypothèse est que l'identification des stimuli ne sera pas conditionnée ni par la manipulation de l'alignement ni par la

manipulation de la base intonative des stimuli, car la durée segmentale et la structure syllabique ne sont pas ambiguës entre celles de *nonno* et de *nono*. La figure 3.5 confirme cette hypothèse. Chaque panneau de la figure illustre les résultats pour le continuum créé à partir de la base lexicale avec la syllabe fermée (*nonno*/D8, en haut) et ouverte (*nono*/D2, en bas). Le score moyen des réponses “*nonno*” (axe des ordonnées) est fonction des pas d’alignement (axe des abscisses). Les lignes représentent le score obtenu pour les énoncés dont la base intonative était une affirmation (points) ou une question (triangles). Comme on peut le remarquer, dans la série à base lexicale *nonno*/D8, le score des réponses “*nonno*” était de 1 indépendamment de la modalité intonative, tandis que dans la série *nono*/D2 la réponse “*nono*” était dominante. La régression logistique confirme que le contraste phonologique entre ces la paire lexicale minimale dépend de la durée segmentale, tandis que ni l’alignement tonal ($z = -0.18, p > .01$) ni la base intonative ($z = -0.43, p > .01$) n’influencent l’identification lexicale de façon significative.

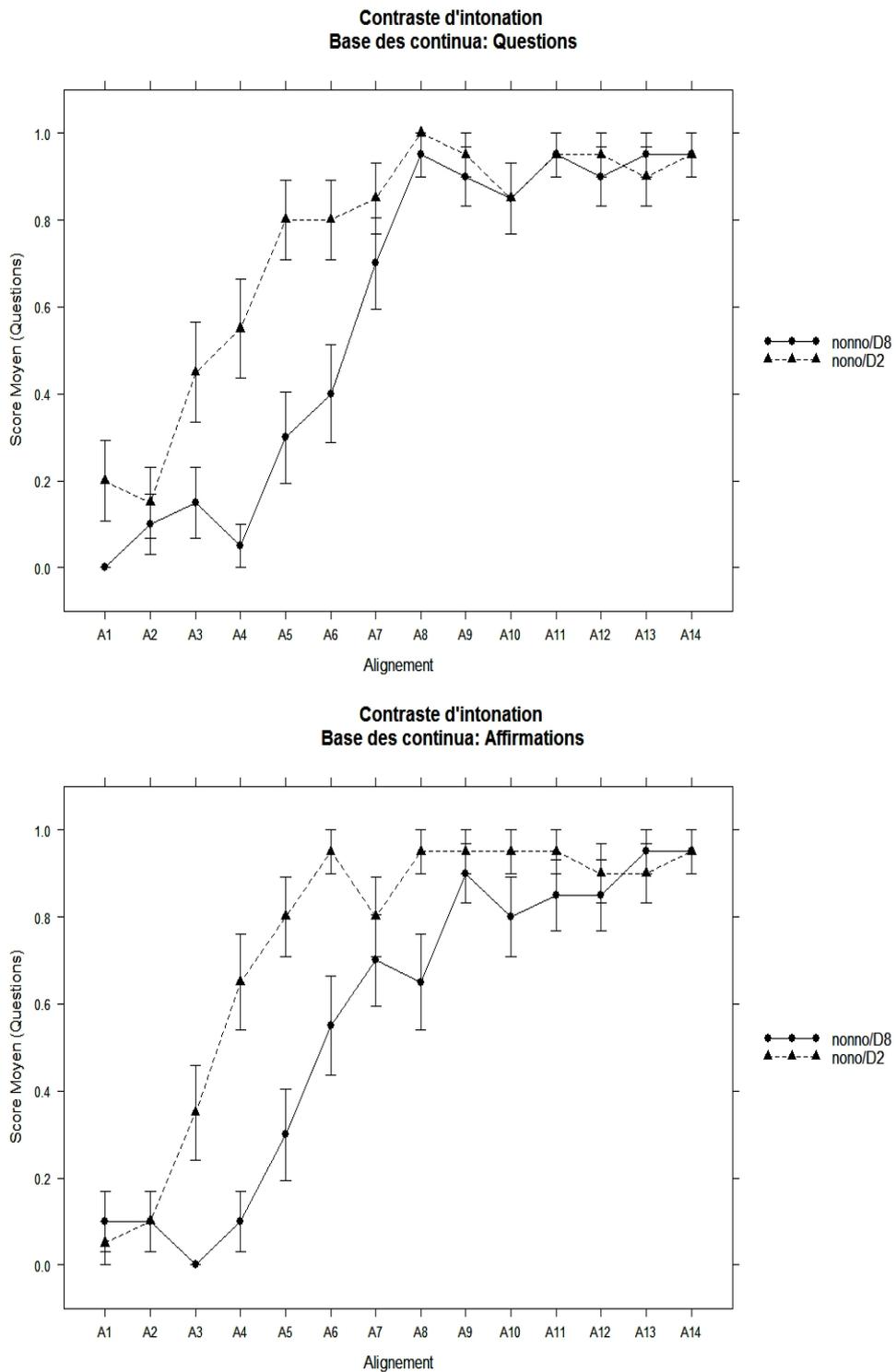


FIGURE 3.4 – Score moyen des questions pour l'identification intonative pour les stimuli à D8 et à D2 créés à partir d'une question (en haut) et d'une affirmation (en bas).

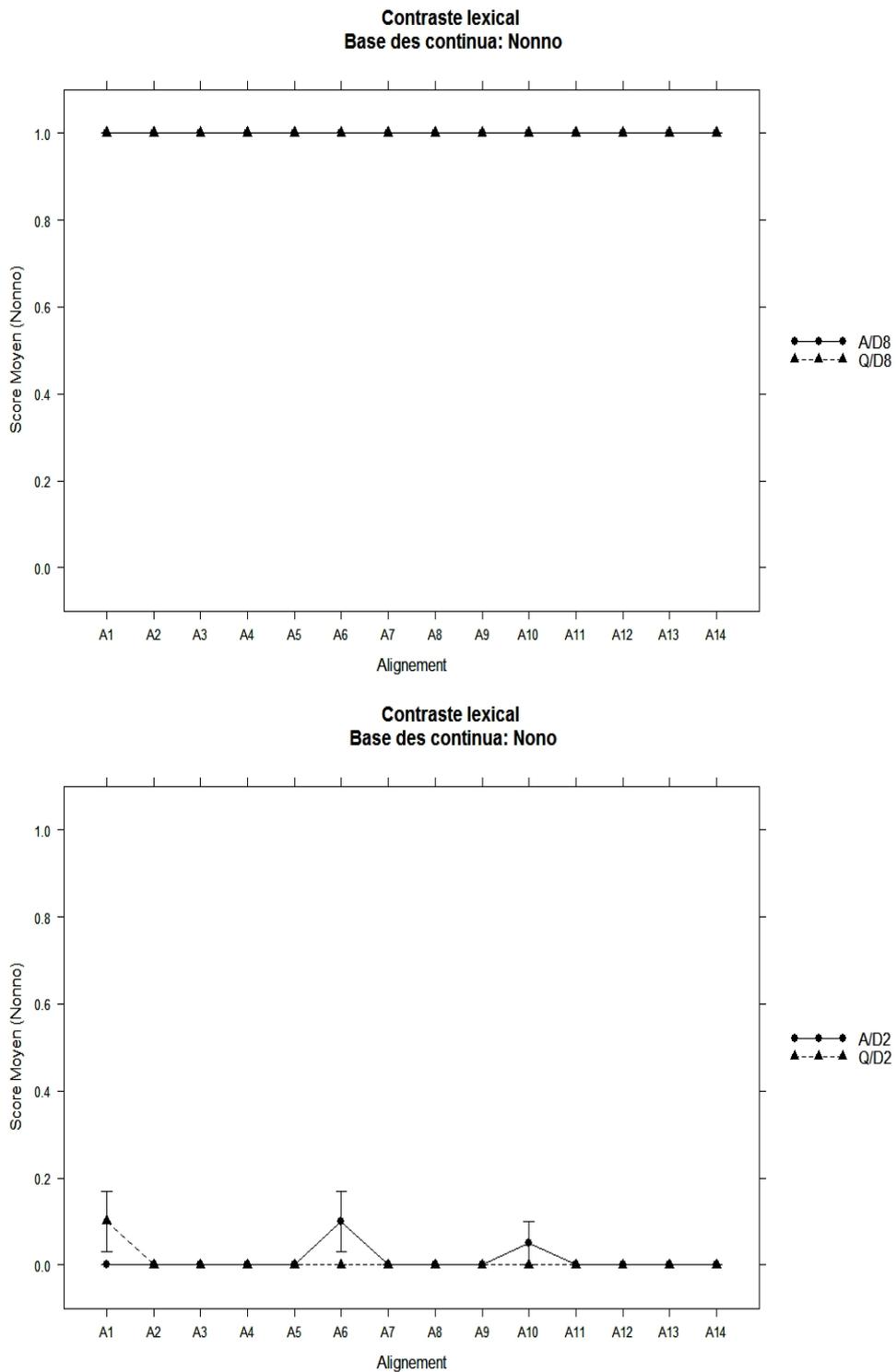


FIGURE 3.5 – Score moyen de *nonno* pour l'identification intonative pour les stimuli à D2 et D8 créés respectivement à partir de *nonno* (en haut) et *nono* (en bas). Les résultats sont montrés séparément selon la modalité intonative de la base (A/Q).

3.4.4 Discussion

Les résultats de l'Expérience I soutiennent l'hypothèse que la structure syllabique influence elle aussi l'identification de la modalité intonative. En fait, nous avons trouvé que le score des réponses "questions" était plus élevé dans *nono*/D2 que dans *nonno*/D8, et cela était indépendant de la base intonative de l'énoncé.

Ce résultat n'est pas en accord avec celui obtenu par D'Imperio (2000) car dans son étude la localisation de la frontière catégorielle entre questions et affirmations n'est pas influencée par la présence/absence d'une coda consonantique. Une hypothèse possible qui expliquerait cette divergence est que les auditeurs sont sensibles au contraste d'alignement entre la syllabe fermée/ouverte uniquement lorsque la durée segmentale de la voyelle accentuée et celle de la coda sont perceptivement très différentes entre les deux structures syllabiques. Par conséquent, on pourrait imaginer que dans notre expérience, les effets de structure syllabique sont plus forts car la durée a été légèrement modifiée afin d'obtenir des stimuli représentatifs d'un point de vue perceptif du contraste syllabique. Dans l'expérience de perception menée par D'Imperio (2000), au contraire, les stimuli lexicaux de base étaient, eux, naturels, et avaient été choisis par l'auteur comme représentatifs du contraste syllabique. Cependant, il est possible que ces stimuli étaient moins différents au niveau de la durée segmentale, ce qui aurait pu provoquer tout simplement une différence dans la forme de la fonction des réponses (voir par. 3.1). Ces considérations sont toutefois purement spéculatives.

Il faudra noter que dans nos données l'effet de la structure syllabique n'est pas le même à travers les pas d'alignement, car il est moins fort aux extrémités qu'au centre des continua des stimuli. En effet, en A1 et A2, où le pic était localisé dans la portion initiale de la voyelle pour *nono*/D2 et dans la portion centrale pour *nonno*/D8, aucune ambiguïté n'était créée dans la perception du contraste intonatif : puisque cet alignement est typique des affirmations à focalisation étroite, les auditeurs ont catégorisé les stimuli comme affirmations. De même, de A9 à A14 le pic se trouve dans la toute dernière portion de la voyelle ou dans la coda consonantique, ce qui est exploité comme un indice pour la catégorisation des stimuli comme questions. Au contraire, dans les stimuli A3 - A8 le pic de f_0 était situé dans la deuxième portion de la voyelle, ce qui aurait pu provoquer une

ambiguïté entre les affirmations et les questions. A ce moment là, l'indice d'alignement relativement à la frontière droite de la syllabe aurait été exploité activement. A A3, le pic de f_0 était placé à 105 ms avant la fin de la voyelle, donc 17 ms plus tard par rapport à la production naturelle de *nono* dans l'affirmation. Au contraire, ce type d'alignement était plus avancé par rapport à la production naturelle de *nonno* dans l'affirmation, où le pic de f_0 était aligné à 143 ms avant la fin de la voyelle. Par conséquent, la différence de score obtenue en A3 pourrait être due au fait que l'alignement tonal était ambigu pour les stimuli avec la syllabe ouverte, car plus similaire à celui trouvé dans les questions. Cela n'est pas le cas dans les stimuli avec la syllabe fermée, car sur le plan intonatif ils étaient clairement identifiés comme étant des affirmations. Ensuite, le score des réponses "questions" augmente de A4 à A8 mais la fonction des réponses est plus raide pour *nono*/D2 que pour *nonno*/D8. Dans ces pas d'alignement, le pic accentuel était de plus en plus retardé dans la voyelle, ce qui a dû provoquer une augmentation de réponses "question" plus importante dans les stimuli avec la syllabe ouverte que dans les stimuli avec la syllabe fermée. En effet, nous savons que le pic accentuel des questions peut être placé dans le milieu ou dans la portion finale de la voyelle accentuée dans les syllabes ouvertes, et à la fin de la voyelle ou dans la coda consonantique des syllabes fermées. En A8, le pic de f_0 était situé à 35 ms avant la fin de la voyelle tonique, ce qui est cohérent avec l'alignement trouvé dans les questions ayant une syllabe ouverte (le pic accentuel étant 38 ms avant la fin de la voyelle dans *nono*) mais pas avec celui trouvé dans les questions ayant une syllabe fermée (le pic étant 43 ms après la fin de la voyelle dans *nonno*). En effet, dans ce pas le score d'identification atteint 0.8 pour *nono*, mais il est encore assez bas pour *nonno* (et notamment lorsque la base intonative des stimuli est une affirmation).

Si cette hypothèse est vraie, on devrait s'attendre à ce que lorsque la durée segmentale et la structure syllabique sont ambiguës entre celles des syllabes ouvertes et des syllabes fermées, il n'y aura pas de différences dans l'identification du contraste intonatif entre *nono* et *nonno*. Cette hypothèse sera testée dans l'Expérience II.

En ce qui concerne le contraste lexical, l'Expérience I nous confirme que lorsque la durée et la structure syllabique ne sont pas ambiguës, les auditeurs exploitent cette information pour l'identification du contraste entre *nono* et *nonno*, car en italien la durée

segmentale est l'indice principal de la gémation consonantique. Cependant, nous nous attendons à ce que lorsque la durée n'est plus accessible, les auditeurs vont exploiter d'autres indices secondaires, et notamment l'information d'alignement tonal. Cette hypothèse sera aussi testée dans l'Expérience II.

3.5 Expérience II

3.5.1 Préparation des stimuli et enregistrement

Dans cette expérience, nous allons tester les effets de la manipulation de l'alignement et de la structure syllabique sur l'identification intonative et lexicale lorsque la durée segmentale est ambiguë entre *nonno* et *nono*. Pour cela, les quatre stimuli dans les figures 1.4 et 1.5 ont été la base de la manipulation de la durée et de l'alignement tonal. En premier lieu, la durée de la coda consonantique et de la voyelle accentuée dans *nonno* et *nono* ont été modifiées afin de créer des stimuli similaires en durée à ceux produits en correspondance avec les pas D5 et D6 de la série avec la base *nonno* (*nonno*/D5) et *nono* (*nono*/D6). En fait, la tâche préliminaire nous a montré que la durée segmentale de ces stimuli est ambiguë. De plus, nous avons gardé les mêmes modifications de l'alignement et de la hauteur mélodique utilisées dans l'Expérience I. Par conséquent, 56 stimuli ont été créés par la combinaison des deux bases intonatives (A/Q), des deux bases lexicales (*nonno*/D5 / (*nono*/D6) et des quatorze pas d'alignement (A1-A14).

Le déroulement de l'Expérience II était le même que celui décrit pour l'Expérience I. Vingt deux auditeurs napolitains, qui n'avaient pas participé à l'Expérience I, ont été séparés en deux sous-groupes de onze personnes. Le premier groupe devait d'abord identifier les stimuli par rapport au contraste lexical et ensuite par rapport au contraste intonatif. Le deuxième devait accomplir les mêmes tâches, mais dans l'ordre inverse. Tous les auditeurs étaient droitiers. Les étiquettes *1. domanda*, *2. affermazione* ("question/ affirmation") et "*1. nonno*, *2. nono*" ("grand-père/ neuvième") étaient présentés visuellement sur l'écran de l'ordinateur. L'expérience durait environ 30 minutes.

3.5.2 Résultats

Dans la figure 3.6 les résultats de l'identification du contraste intonatif sont montrés séparément pour les stimuli dont la base intonative était une question (en haut) et une affirmation (en bas). Chaque graphique illustre le score moyen de réponses "questions" regroupés pour tous les auditeurs (ordonnées) et en fonction des pas d'alignement (abscisses), et pour les stimuli *nonno* en D5 (points) et *nono* en D6 (triangles). Comme on peut le remarquer, la manipulation de l'alignement provoque encore une fois un déplacement de la frontière catégorielle entre les questions et les affirmations, de sorte qu'à l'extrémité gauche du continuum, le score moyen des questions est globalement très bas (0.03) et à l'extrémité droite très élevé (0.93), indépendamment de la base intonative de l'énoncé. De plus, le score d'identification est très similaire pour les deux bases lexicales *nonno* et de *nono*. Cela était attendu, car nous avons supposé que lorsque les stimuli étaient ambigus quant à leur durée segmentale, les effets de la structure syllabique sur l'identification intonative doivent être nuls. En fait, la régression logistique à effets mixtes nous révèle un effet de la manipulation de l'alignement [$z= 11.14, p <.01$] tandis que ni l'effet de la base lexicale [$z= -0.20, p >.01$] ni celui de la base intonative [$z= 0.10, p >.01$] ne sont significatifs. Les interactions ne sont pas non plus significatives.

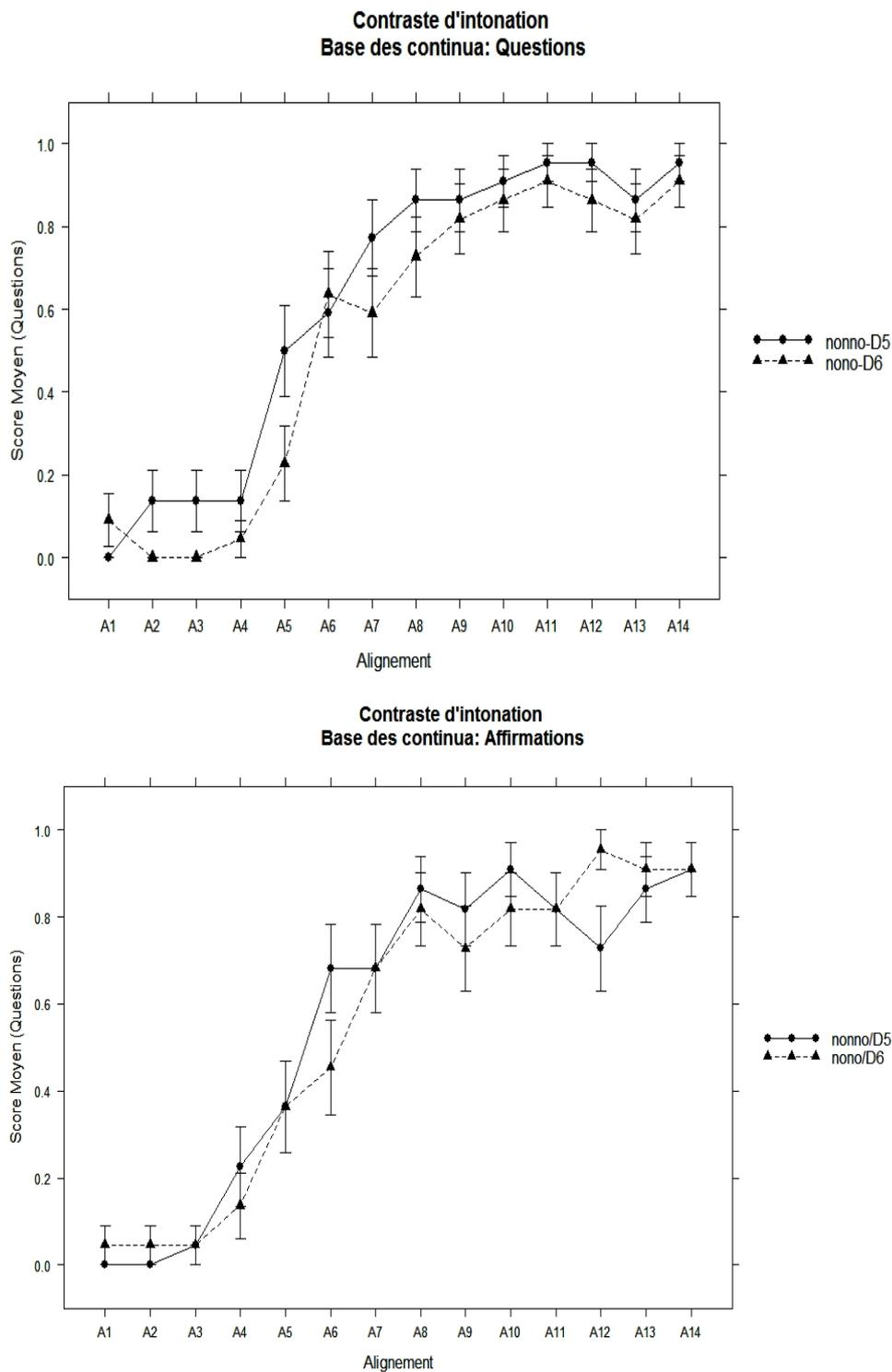


FIGURE 3.6 – Score moyen des réponses “questions” pour l’identification intonative pour les stimuli *nonno/D5* et *nono/D6* créés à partir d’une question (en haut) et d’une affirmation (en bas).

Nous allons rendre compte maintenant des résultats pour le contraste lexical. La figure 3.7 montre le score moyen de réponses “*nonno*” à travers les pas d’alignement, respectivement pour la série de stimuli créée à partir de la base *nonno*/D5 (panneau du haut) et de la base *nono*/D6 (panneau du bas). Dans chaque panneau, les résultats sont illustrés séparément selon la modalité intonative de l’énoncé naturel.

En premier lieu, on peut remarquer que le score d’identification ne varie pas radicalement en fonction de l’alignement, dans les quatre séries de continua. Cela était attendu, car l’indice principal pour le contraste *nonno*/*nono* (la durée) était ambigu d’un point de vue perceptif.

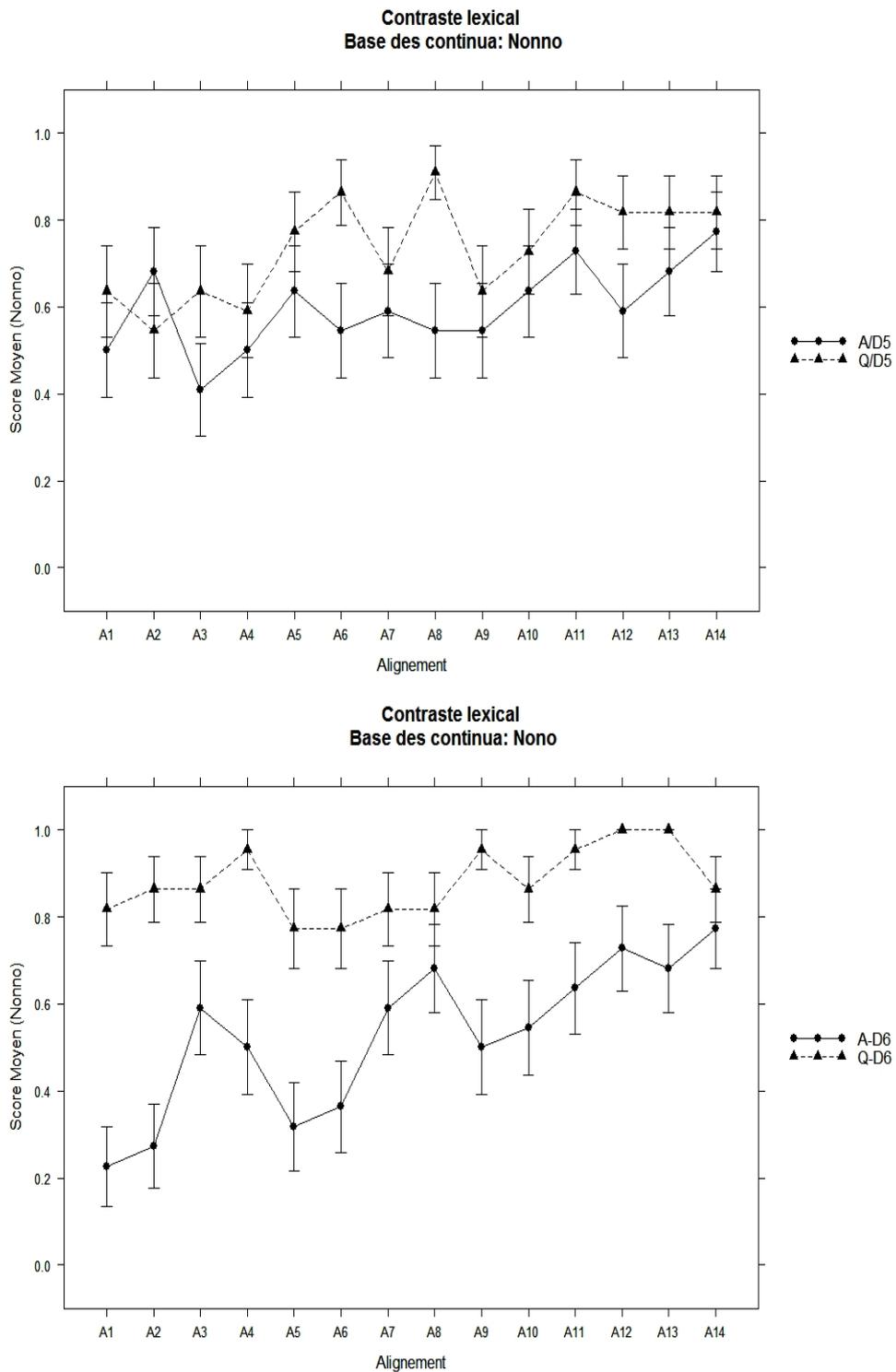


FIGURE 3.7 – Score moyen de *nonno* pour l'identification intonative pour les stimuli à D5 et D6 créés respectivement à partir de *nonno* (en haut) et *nono* (en bas). Les résultats sont montrés séparément selon la modalité intonative de la base (A/Q).

Pour les stimuli dont la base lexicale était *nonno*, l’alignement ne semble pas avoir influencé fortement l’identification lexicale. Contrairement aux attentes, le score des réponses “*nonno*” est déjà assez élevé en A1 - A4, que ce soit dans le continuum créé à partir d’une affirmation (A1 : 0.50), ou dans le continuum créé à partir d’une question (A1 : 0.63). Cela signifie probablement que la manipulation de l’alignement à la droite du continuum n’a pas été utilisée par les auditeurs comme un indice de la présence d’une consonne simple. Cependant, le score moyen tend à augmenter d’A1 à A14. En fait, en A14, le score moyen est autour de 0.80 dans le continuum créé aussi bien à partir d’une affirmation que dans celui créé à partir d’une question.

Pour les stimuli dont la base lexicale était *nono*, les auditeurs ont répondu de façon différente selon la modalité intonative de la base. Dans les stimuli créés à partir d’une affirmation, le score moyen de réponses “*nonno*” augmente dans les stimuli de A1 à A4 (c’est à dire, dans les stimuli identifiés comme des affirmations). En particulier, le score en A1 et A2 est de 0.22 et 0.23, indiquant que pour ces pas d’alignement, les stimuli ont été identifiés comme *nono*. Ensuite, la courbe d’identification monte autour du niveau de hasard (à 0.58 en A3 et à 0.50 en A4). Dans les pas A9 - A14 (c’est à dire, dans les stimuli identifiés comme des questions), le score des réponses augmente progressivement. En A1, le pic de f_0 du stimulus était aligné à 15 ms avant la fin de la voyelle, et le score d’identification est de 0.50 (niveau de hasard). Cependant, ce score augmente de façon linéaire et graduelle dans les pas successifs jusqu’à A14 (0.77). Ces résultats semblent supporter les résultats de D’Imperio *et al.* (2007), qui avaient trouvé que les réponses “*nonno*” tendent à augmenter lorsque l’alignement du pic de f_0 est retardé dans la coda consonantique. Nous allons maintenant examiner les stimuli créés à partir de *nono* mais dont la base intonative est une question. Dans ces stimuli, le score d’identification ne semble pas être influencé par la manipulation de l’alignement, et en effet les valeurs aux extrêmes du continuum sont assez similaires (A1 : 0.81 ; A14 : 0.86). Il est important de remarquer que ce score est beaucoup plus élevé que celui trouvé pour les stimuli de durées correspondantes (D6) employés dans la tâche préliminaire, tâche dans laquelle l’identification de *nonno* était proche du niveau de hasard. Cela pourrait indiquer que l’identification de *nonno* dépende de la modalité intonative dans laquelle le mot a été prononcé, car dans la

tâche préliminaire les mots cibles étaient sous forme de citation (et donc prononcés comme des affirmations à focalisation large), tandis qu'ici le mot était inséré dans des questions oui/non. De plus, les résultats pour ce continuum sont globalement plus élevés que ceux trouvés pour les continua dont la base était “*nonno*”. Cela est très surprenant, et pourrait suggérer une interaction entre la durée (qui était plus longue dans les stimuli en *nonno*/D6 que en *nonno*/D5) et la modalité intonative.

L'analyse de régression logistique à effets mixtes sur les quatre continua révèle que l'effet d'alignement est significatif [$z = 3.26, p < .01$]. Il n'y a pas d'effets principaux de la base lexicale et intonative ($p = .01$), mais leur interaction est significative [$z = 5.05, p < .01$]. Compte tenu des différences de réponses entre la base lexicale et la base intonative, nous avons effectué une analyse de régression logistique séparément pour les deux bases lexicales. L'analyse confirme que dans les stimuli en *nonno*/D5 l'effet global d'alignement est très petit et il n'atteint pas le niveau de signification [$z = 2.40, p = .016$]. De plus, ni la base intonative ni l'interaction ne sont significatives. Au contraire, dans les stimuli en *nonno*/D6, l'effet de l'alignement [$z = -3.19, p < .01$] et de modalité intonative [$z = 5.17, p < .01$] sont significatifs. L'interaction entre alignement et modalité intonative n'est pas significative. Cela suggère que l'alignement a été exploité pour l'identification lexicale principalement pour les stimuli créés à partir de *nonno*/D6, bien que cet effet soit plus fort lorsque la modalité intonative de la base est une question.

3.5.3 Discussion

Les résultats de l'Expérience II soutiennent l'hypothèse selon laquelle la structure syllabique ne joue pas de rôle dans l'identification intonative lorsque les stimuli sont ambigus quant à la durée segmentale. En fait, nous n'avons pas trouvé d'effets de la base lexicale (*nonno/nono*) dans l'identification du contraste entre questions et affirmations, et cela indépendamment de la modalité intonative de l'énoncé.

En ce qui concerne l'identification lexicale, l'Expérience II suggère que les Napolitains peuvent employer les détails d'alignement dus à la durée segmentale et à la structure syllabique, mais uniquement lorsque les stimuli sont ambigus. Cependant, les résultats de l'identification lexicale sont difficiles à interpréter. Nous avons émis l'hypothèse que

le score d'identification *nonno* devait augmenter lorsque le pic de f_0 était retardé dans la syllabe accentuée, soit dans les stimuli jugés comme affirmations, soit dans ceux jugés comme questions. En d'autres termes, nous nous attendions à ce que l'alignement au début/dans le centre de la voyelle (pour les affirmations) et à la fin de la voyelle/dans la coda consonantique (pour les questions) influence la perception de la syllabe comme ouverte/fermée. Dans nos données, cette hypothèse semble être confirmée par la série des stimuli en *nono*/D6 qui avait été créée à partir d'une affirmation.

Pour les stimuli jugés comme affirmations (A1 - A4), le score d'identification est très bas en A1 et A2 (autour de 20%), indiquant que l'alignement très avancé dans la voyelle est exploité comme indice de la présence d'une syllabe ouverte. Le score atteint le niveau de hasard en A3 et A4. En correspondance avec ces pas d'alignement les auditeurs ne sont plus capables d'identifier les stimuli comme un des deux mots de la paire minimale lexicale.

Les résultats en A5 - A8 sont plus difficiles à interpréter. Nous savons que le pic de f_0 est aligné avec le centre de la voyelle pour les mots ayant une consonne géminée et qui sont insérés dans les affirmations; et dans la partie finale de la voyelle pour les mots ayant une consonne simple et qui sont insérés dans les questions. Dans les stimuli à A5 - A8, la distance du pic de f_0 par rapport à la fin de la voyelle varie entre 75 ms (A5) et 30 ms (A8). Par conséquent, une explication possible est que l'identification lexicale peut avoir été biaisée par une ambiguïté dans l'identification intonative. Dans l'Expérience II, nous avons établi que les stimuli étaient bien identifiés comme des questions uniquement lorsque leur score d'identification était de 0.80 ou plus, et comme des affirmations lorsque le score était de 0.20 ou moins. La fig. 3.6 nous montre que pour les stimuli en A5 - A8 le score des réponses "questions" était entre 0.2 et 0.8. Cela signifie que les stimuli en A5 - A8 étaient ambigus d'un point de vue perceptif entre les questions et les affirmations. Cette hypothèse semble être confirmée par D'Imperio *et al.* (2007), qui avaient trouvé que lorsque l'alignement des stimuli est ambigu entre celui des questions et des affirmations, les résultats de l'identification lexicale sont très bruités.

Dans les stimuli à A9 - A14, qui avaient été clairement jugés comme des questions, des différences d'alignement tonal à la fin de la voyelle et à l'intérieur de la coda consonantique

ont provoqué des augmentations du score de réponses “*nonno*”. Encore une fois, ces résultats répliquent ceux obtenus par D’Imperio *et al.* (2007). Dans leur travail, D’Imperio *et al.* (2007) ont trouvé que dans des stimuli identifiés comme des questions, la manipulation de l’alignement vers la droite du continuum (c’est-à-dire, dans la coda consonantique) influence le score des réponses “*nonno*”. En particulier, le déplacement du pic dans la coda consonantique conduit les auditeurs à catégoriser ces stimuli comme des mots ayant une consonne géminée. Ces résultats suggèrent donc que les différences d’alignement provoquent un déplacement graduel et linéaire de la frontière catégorielle entre *nono* et *nonno*.

Nous allons maintenant détailler les résultats de l’identification lexicale pour les autres continua. Pour la série dont la base lexicale était *nonno*, le déplacement à droite du pic ne permettait pas d’influencer la perception du contraste lexical entre les deux mots, car déjà en A1 le score d’identification était de 50%. Cela peut indiquer que la manipulation de l’alignement n’a pas influencé l’identification lexicale, car la durée segmentale n’était pas vraiment ambiguë. Cependant, il faut remarquer une faible tendance graduelle à l’augmentation du score de A1 à A14. En plus, l’augmentation progressive de réponses “*nonno*” en fonction des pas d’alignement est moins marqué dans les stimuli jugés comme affirmations (A1 - A4) que comme questions (A9 - A14). En particulier, dans les stimuli jugés comme questions (A9-A14), le score d’identification augmente de façon linéaire, de façon similaire à ce qui a été déjà trouvé par D’Imperio *et al.* (2007).

Finalement, nous avons trouvé une interaction significative entre la base lexicale et la base intonative. En effet, le score de réponses “*nonno*” était plus élevé lorsque la base lexicale était “*nono*” et la base intonative était une question. Cela est très surprenant, car nous nous attendions à ce que l’identification soit meilleure dans la série à base *nonno*, car plus cohérent avec les propriétés phonétiques du stimulus original. Cependant, il faut se rappeler que dans l’Expérience II, nous avons employé des stimuli de durée différente selon la base lexicale, car la tâche préliminaire nous avait montré que les stimuli les plus ambigus par durée étaient à D5 pour la série à base *nonno* et à D6 pour la série à base *nono*. En d’autres termes, les valeurs absolues de durée segmentale pour les stimuli créés à partir de *nono* étaient plus élevées que pour *nonno*. De plus, la valeur de la coda

consonantique à *nono*/D6 (145 ms) était plus similaire à celle trouvée dans la question naturelle avec la syllabe fermée (156 ms) qu’avec la syllabe ouverte (79 ms). L’interaction entre la durée et la f_0 aurait pu contribuer à biaiser les résultats de l’identification vers “*nonno*”. Par conséquent, une étude plus contrôlée est souhaitée dans le futur dans laquelle les effets de la modalité intonative et de la structure syllabique sur la perception de la durée segmentale pourraient être testés.

Pour résumer, les Expériences I et II présentées dans ce chapitre montrent que la structure syllabique et l’alignement tonal sont exploités pour l’identification intonative et lexicale. Dans le par. 3.1 nous avons supposé que, si la structure syllabique jouait un rôle dans la perception du contraste intonatif, ses effets devaient être significatifs uniquement lorsque les stimuli sont clairement identifiés comme *nonno* ou *nono* ; au contraire, aucun effet était attendu lorsque la structure syllabique était ambiguë entre une syllabe fermée et une syllabe ouverte. Cela est confirmé par nos données. En particulier, l’Expérience I a montré l’existence de ces effets pour les stimuli *nono* et *nonno* lorsque leurs durées étaient similaires à celles trouvées dans les stimuli naturels, à D2 et D8. Ces effets étaient très forts surtout lorsque l’identification intonative est ambiguë (A3 - A8). Contrairement à D’Imperio (2000), nos résultats suggèrent donc que l’alignement tonal à la frontière droite de la syllabe peut être exploité pour l’identification du contraste intonatif entre affirmations et questions. De plus, l’Expérience II nous a montré que ces effets disparaissent lorsque la structure syllabique est ambiguë. Du point de vue perceptif, cela ne soutient pas, “l’explication structurelle révisée” (*revised structural explanation*, voir l’introduction) affirmée par Ladd *et al.* (2000) car les auditeurs napolitains n’ont pas été capables de calculer les différences d’alignement dues à la structure syllabique, lorsque la durée segmentale est ambiguë.

En ce qui concerne les effets d’alignement sur l’identification lexicale, l’Expérience I et II suggèrent que les Napolitains peuvent employer les détails d’alignement (dûs aux effets de durée segmentale et de structure syllabique), mais uniquement lorsque l’information de durée segmentale n’est pas disponible. Nos résultats montrent donc que l’interaction entre l’information segmentale et suprasegmentale est pertinente au niveau perceptif même pour l’identification des mots en contraste de structure syllabique et de durée segmentale, car

le contraste phonologique peut s'appuyer aussi sur des indices secondaires, différents que la durée segmentale.

Variabilité de forme et de pente de f_0 : preuve de l'existence du
Syntagme Accentuel en napolitain

4.1 Problématique

Dans les chapitres précédents, nous avons analysé les caractéristiques acoustiques et perceptives de la configuration en montée-descente constituée par l'accent nucléaire LH et par l'accent de phrase L-. Nous nous sommes focalisée sur l'interaction entre la variabilité segmentale et l'alignement tonal, car nous avons supposé, en ligne avec des études précédentes, que l'alignement tonal de l'accent nucléaire était un indice robuste pour le contraste entre les questions oui/non et les affirmations à focalisation étroite, en napolitain. En particulier, nous avons montré une interaction importante entre le niveau segmental et suprasegmental dans la réalisation et l'identification du contraste intonatif, ce qui nous a conduit à rejeter l'interprétation stricte de l'hypothèse de l'ancrage segmental pour l'italien napolitain et, plus généralement, à rejeter l'idée que l'information véhiculée par la f_0 est indépendante de celle fournie par le plan segmental. D'un point de vue méthodologique, l'exploration acoustique des effets segmentaux sur l'accent mélodique (chapitre 2) nous a fait éloigner de la notion "concrète" de cibles tonales comme

minima et maxima de f_0 , et donc à prendre avec précaution les résultats d’alignement obtenus par une simple détection manuelle des cibles.

Dans ce chapitre et dans celui qui suit, nous allons rendre compte d’une différence établie entre les deux modalités, relative à la forme et à la pente de f_0 , et qui caractérise la région entre l’accent prénucléaire et nucléaire. Les données acoustiques présentées dans ce chapitre suggèrent que cette différence est due à l’insertion d’un ton périphérique qui est différemment spécifié dans les questions et dans les affirmations. Dans le chapitre 5, nous allons aussi présenter les résultats de deux expériences de perception, qui montrent que cette différence de spécification tonale est utilisée activement par les auditeurs napolitains dans l’identification de la modalité intonative.

Selon la Théorie AM le contour intonatif est représenté par une séquence des tons H et L liés entre eux grâce à l’application de règles d’interpolations (Pierrehumbert, 1980). En particulier, dans une séquence de deux accents montants LH, le pic H du premier accent est obligatoirement lié au ton bas L du deuxième par une ligne droite, car la Théorie prédit que la transition entre deux tons spécifiés différemment est linéaire (fig.4.1).

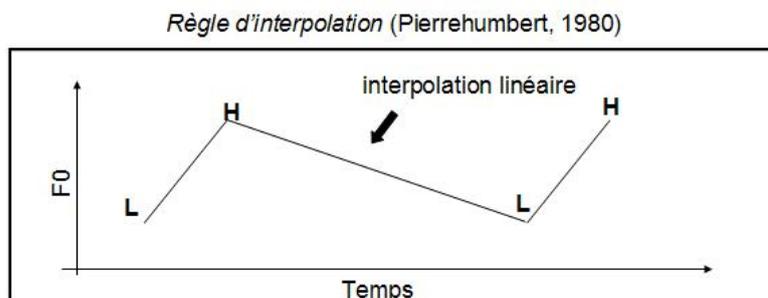


FIGURE 4.1 – Représentation schématisée de la règle d’interpolation linéaire de Pierrehumbert (1980)

D’autres descriptions autosegmentales ont proposé des règles d’interpolations différentes pour expliquer comment les tons se combinent dans le contour intonatif. Dans Gussenhoven (1984), la transition linéaire entre deux tons est le résultat de l’application partielle de la *tone linking rule* (“règle de liaison des tons”). Par exemple, lorsqu’un accent prénucléaire descendant H^*+L se combine avec un accent nucléaire LH, l’application partielle de cette règle (ou *partial linking*) oblige le ton *trailing* L du premier accent à se

déplacer dans une position adjacente au ton L de l'accent suivant. Au niveau phonétique, ce déplacement provoque une interpolation linéaire entre le pic de l'accent H^*+L et le ton L de l'accent nucléaire L^*+H . Cependant, si la règle n'est pas appliquée, les deux tons L seront implémentés et leur réalisation sera visible dans la trace acoustique comme deux minima de f_0 séparés (fig.4.2).

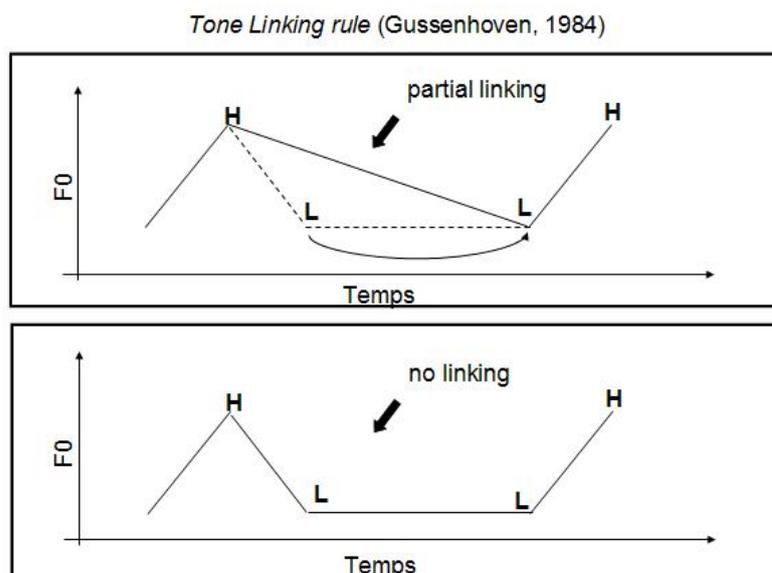


FIGURE 4.2 – Représentation schématisée de la *tone linking rule* de Gussenhoven (1984). Le panneau supérieur illustre l'application partielle de la règle (la flèche courbe indique le déplacement du ton L de l'accent H^*+L vers la droite). Le panneau inférieur montre le cas d'absence de liaison entre les deux tons bas.

Les prédictions des deux auteurs ne sont pas appropriées pour décrire les données du napolitain. La fig. 4.3 illustre le spectrogramme et la courbe de f_0 pour l'énoncé *La mamma vuole vedere la Dina*, lue comme une question oui/non (panneaux en haut) et comme une affirmation à focalisation étroite (panneaux en bas). Dans les deux modalités intonatives, un accent prénucléaire montant (LH^*) a été réalisé sur la syllabe tonique *mam-* de *La mammadans* les deux modalités et un accent nucléaire montant (L^*+H dans les questions et $L+H^*$ dans les affirmations) a été réalisé sur la syllabe tonique *Di-* de *la Dina*. Si les deux accents étaient liés par une interpolation linéaire (Pierrehumbert, 1980), les affirmations et les questions auraient dû se distinguer uniquement par la catégorie tonale de l'accent nucléaire. Au contraire, les deux modalités présentent une différence supplémentaire dans

la région interaccentuelle de f_0 qui suit immédiatement l'accent prénucléaire. Dans les questions, la pente de f_0 descend doucement du ton H de l'accent prénucléaire jusqu'à la fin de *La mamma*, de sorte que le contour de f_0 dans la région qui suit immédiatement l'accent prénucléaire se réalise sous la forme d'une parabole convexe. Après cette région, la pente devient plus raide afin d'atteindre la valeur basse du L* de l'accent nucléaire. Dans les affirmations, la chute de la f_0 est beaucoup plus rapide, de sorte que la région suivant l'accent prénucléaire est caractérisée par la réalisation d'un coude bas de f_0 dans la région située autour de la fin du mot prosodique *La mamma*.

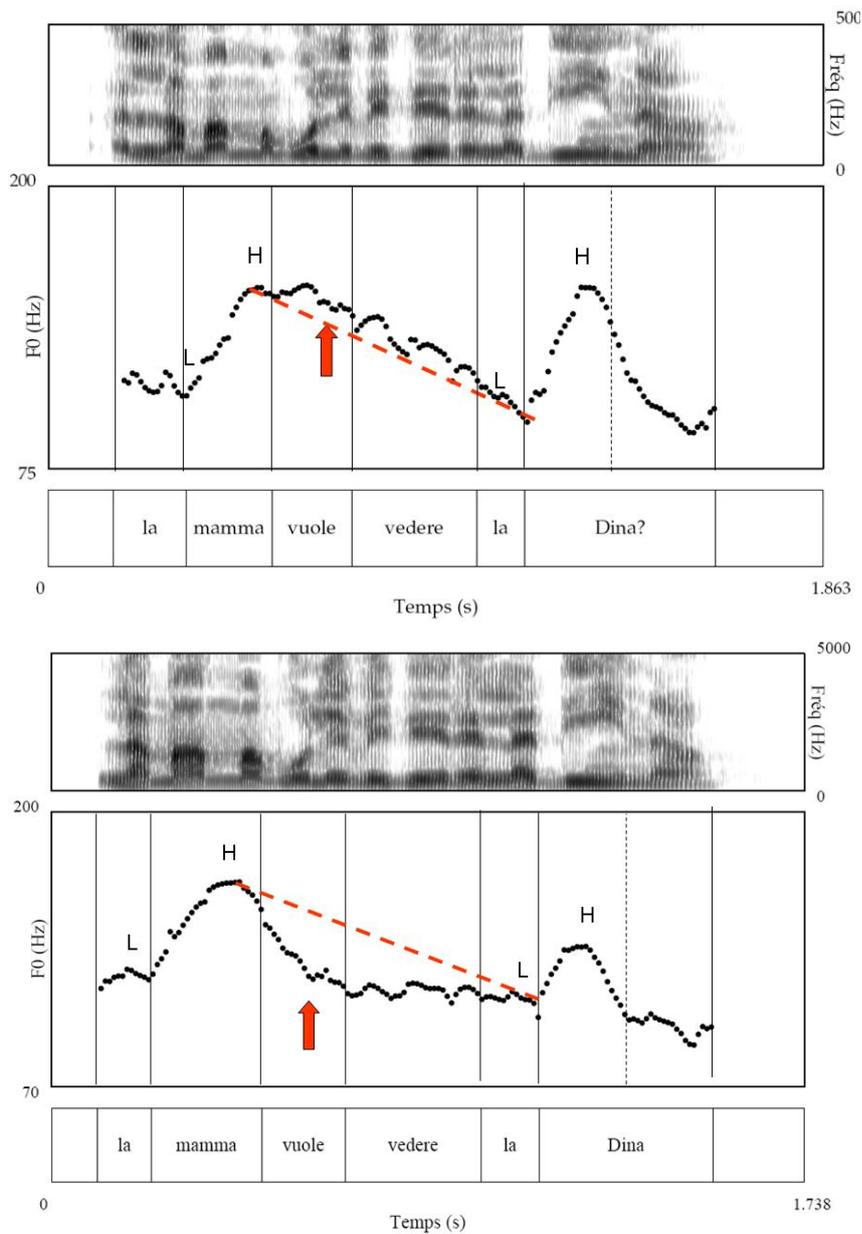


FIGURE 4.3 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *La mamma vuole vedere la Dina* prononcée comme une question (en haut) ou une affirmation (en bas). Les lignes continues indiquent les frontières des mots. La ligne rouge tiretée schématise l'interpolation linéaire prédite par Pierrehumbert (1980) entre les deux accents montants. La flèche indique le mouvement descendant après le pic pré-nucléaire, qui est très douce et convexe dans les questions. Locuteur OM.

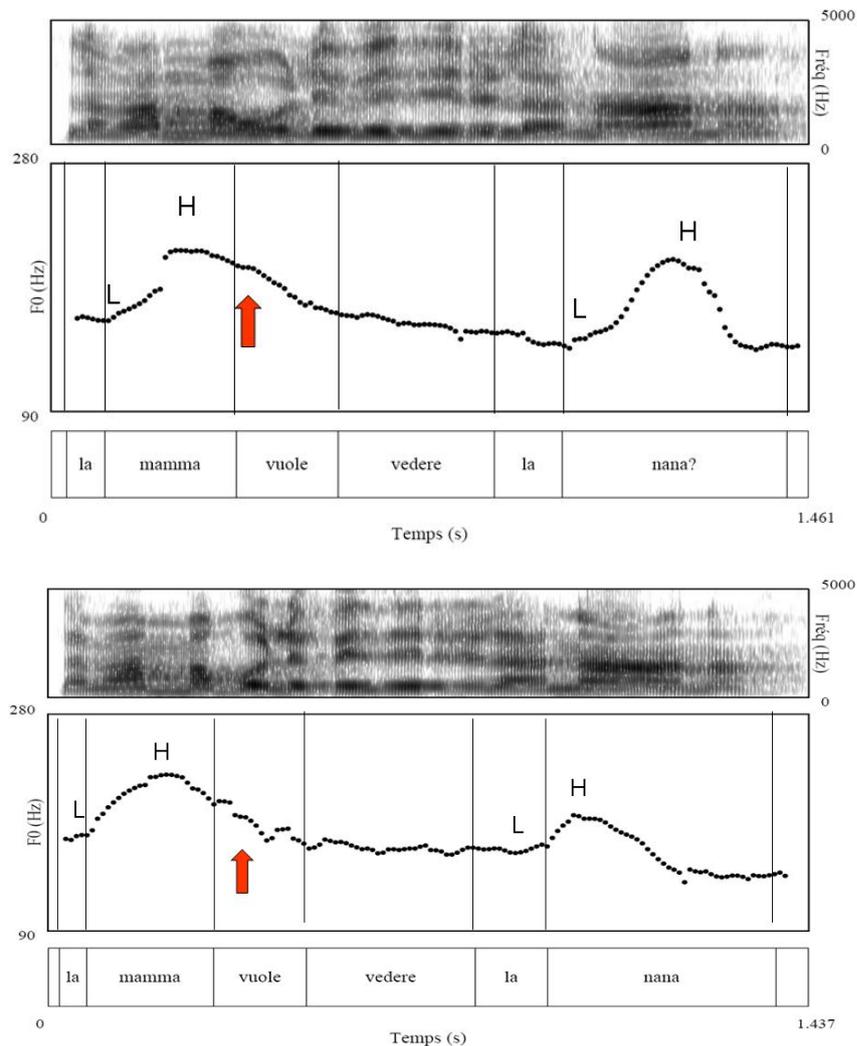


FIGURE 4.4 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *La mamma vuole vedere la nana* prononcée comme une question (en haut) ou une affirmation (en bas). Les lignes continues indiquent les frontières des mots. La flèche indique le mouvement descendant après le pic prénucléaire, qui est plus douce et convexe dans les questions. Locuteur AS.

A première vue, la différence de pente interaccentuelle semble être en accord avec la prédiction faite par Gussenhoven (1984). En fait, on pourrait supposer que, dans les questions, les deux accents sont liés par la *partial linking*, règle qui n'est pas appliquée dans les affirmations. Cette hypothèse ne prédit pas les résultats pour deux raisons. En premier lieu, elle implique une réanalyse de l'accent prénucléaire comme H^*+L . Cependant, suite aux résultats d'une étude préliminaire sur l'accent prénucléaire en italien (Gili Fivela &

D’Imperio, 2003), nous pensons plutôt que l’accent prénucléaire dans la fig. 4.3 est (LH*)¹. Une autre objection, plus importante d’ailleurs, est que le *partial linking* ne rend pas compte de la forme parabolique trouvée dans la région postaccentuelle des questions. En effet, dans cette modalité intonative, l’interpolation entre le H prénucléaire et le L nucléaire n’est pas linéaire. Au contraire, le mouvement de f_0 suivant le pic prénucléaire est souvent caractérisée par un plateau haut suivi, à son tour, par un mouvement descendant (fig. 4.3) ou bien par une chute très douce juste après le pic prénucléaire. car elle prédit que la transition entre le H prénucléaire et le L nucléaire est linéaire.

Nous avons donc émis l’hypothèse que la variabilité de forme et de pente de f_0 entre l’accent prénucléaire (LH*) et nucléaire (L*+H ou L+H*) est due à l’insertion d’un ton dans la région interaccentuelle des questions et des affirmations. En particulier, nous supposons que ce ton, spécifié différemment dans les questions et dans les affirmations, est utilisé pour marquer la fin d’un constituant prosodique inférieur au syntagme intermédiaire. Un candidat possible serait le syntagme accentuel (*Accentual Phrase, AP*), dont l’existence a déjà été attestée dans d’autres langues telles que le japonais (Beckman & Pierrehumbert, 1986 ; Pierrehumbert & Beckman, 1988), le coréen (Jun, 1993), le français (Jun & Fougeron, 2000, 2002) et l’arabe du Caire (Hellmuth, 2004, 2007). Dans les questions, un ton H_{AP} serait responsable de l’inclinaison douce de la pente de f_0 dans la région suivant immédiatement l’accent prénucléaire ; dans les affirmations, un ton L_{AP} serait responsable de la chute soudaine de la f_0 après l’accent prénucléaire.

Notre hypothèse a été testée dans deux expériences de production. La première (Expérience I) vise à déterminer si la variabilité de pente et de forme de f_0 décrite ci-dessus dépend d’une différence phonologique entre les deux modalités intonatives. En particulier, si un ton haut ou bas est inséré après l’accent prénucléaire, la différence de forme de la région interaccentuelle sera systématique et indépendante de facteurs phonétiques de variabilité, tels que la vitesse d’élocution et l’environnement tonal. Dans la deuxième

1. Gili Fivela & D’Imperio (2003) transcrivent cet accent comme (L)H* pour indiquer que la cible L est sujette à un certain degré de variabilité phonétique. Nous nous orientons ici vers la transcription de l’accent prénucléaire comme (LH*). Cette transcription est utilisée pour distinguer l’accent prénucléaire de l’accent nucléaire des affirmations à focalisation étroite, L+H*. Ce choix est adopté sur la base du fait que le pic nucléaire semble avoir une association secondaire avec la première mora de la syllabe accentuée, tandis que celui prénucléaire ne semble pas avoir d’association secondaire avec des points d’ancrage dans la structure métrique (Prieto *et al.*, 2005)

expérience (Expérience II), nous allons mieux examiner l'existence du syntagme accentuel en italien, dont le domaine est inférieur à celui du syntagme intermédiaire. Si cette hypothèse est soutenue, nous nous attendons à ce que la fin de l'*AP* soit signalée par un ton de frontière différent de l'accent de phrase. De plus, ce constituant sera caractérisé par un degré d'allongement final et par la perception d'une rupture prosodique inférieure à ceux trouvés pour le syntagme intermédiaire.

Il faut noter que la recherche d'une cible tonale dans la région interaccentuelle pose problème notamment pour son identification dans la trace acoustique. Cela est vrai surtout pour les questions, où la f_0 de la région qui suit l'accent prénucléaire ne semble pas montrer aucun pic maximal de f_0 . Comment peut-on apprécier la différence entre les questions et les affirmations ? Nous soutenons dans cette thèse réside dans le fait que les cibles tonales ne sont pas toujours identifiables avec des points minimaux ou maximaux de la courbe de f_0 . Par exemple, dans le chapitre 2 nous avons vu que la localisation exacte des cibles peut être masquée par des perturbations microprosodiques. Dans d'autres cas, les cibles ne sont pas réalisées comme des points d'inflexion, comme dans les configurations à plateau (D'Imperio, 2000; Knight, 2003; Knight & Nolan, 2006). De même, nous pensons que, si un ton H_{AP} est réalisé dans les questions du napolitain, sa présence sera signalée uniquement par la forme parabolique et l'inclinaison douce de la pente de f_0 . Cela signifie aussi que la détection des cibles tonales doit être effectuée par le biais de techniques de modélisation sophistiquées, qui soient capables d'identifier les différences dans les propriétés dynamiques entre les deux modalités intonatives. La recherche de ces techniques fera donc également l'objet de ce chapitre.

4.2 Expérience I

Comme on l'a mentionné plus tôt, le but de l'Expérience I est d'examiner si la nature de la variabilité de forme et de pente de f_0 dépend de la présence de deux tons dont la spécification tonale serait différente. Puisque le contour de f_0 entre l'accent prénucléaire et l'accent nucléaire n'est pas le résultat d'une simple interpolation entre les deux accents, il y a au moins deux autres prédictions possibles.

Une hypothèse prédit que dans les deux modalités, la région entre l’accent prénucléaire et l’accent nucléaire est caractérisée par l’insertion du même ton L. En fait, dans nos données la présence d’un ton bas après l’accent prénucléaire semble être signalé par le pente de f_0 très raide après le pic prénucléaire, dont la fin est réalisée dans le mot inaccentué *vuole* dans la fig. 4.3. De même, nous pouvons supposer qu’un ton L caractérise aussi la région interaccentuelle des questions, mais que, dans cette modalité intonative, le mouvement de f_0 n’atteint pas sa cible à cause de contraintes de pression temporelle telles que le *tonal undershoot* (Silverman & Pierrehumbert, 1990). Par exemple, plusieurs travaux dans des langues différentes ont montré que la vitesse d’élocution des questions est plus élevée que celle des affirmations (van Heuven & van Zanten, 2005; Haan, 2001). Cette différence a été trouvée aussi par D’Imperio (2000) pour le napolitain. Par conséquent, si une cible basse est réalisée après le pic prénucléaire dans les questions et si les questions sont prononcées plus rapidement que les affirmations, la cible L sera *undershot*. Une autre source du *tonal undershoot* peut être l’environnement tonal. Les phénomènes de pression phonétique dus au voisinage de tons précédents ou suivants ou à la présence de frontières prosodiques ont été observés dans plusieurs langues (Silverman & Pierrehumbert, 1990; Prieto, 2002; Ladd, 2008), y compris l’italien (Grice *et al.*, 2005; D’Imperio, 2002a). Par conséquent, nous pouvons supposer que l’absence d’une cible tonale basse dans la région interaccentuelle des questions est due à des ajustements temporels ou mélodiques de l’accent prénucléaire et de l’accent nucléaire.

L’hypothèse alternative serait que la différence de forme et de pente est due à l’insertion d’un ton différemment spécifié dans les questions et dans les affirmations. Par conséquent, nous nous attendons à ce que la variabilité trouvée dans la région interaccentuelle de f_0 soit indépendante de la pression phonétique liée à la vitesse d’élocution ou à la réalisation des accents mélodiques qui la délimitent. En particulier, dans les questions l’accent prénucléaire sera suivi par un ton H, tandis que dans les affirmations l’accent prénucléaire sera suivi par un ton L, et la forme du contour sera concave (fig. 4.5). Cette hypothèse repose sur l’idée que, bien que la cible H ne soit pas réalisée comme une valeur maximale locale, sa présence empêche une chute soudaine de la f_0 de sorte que la forme du contour de f_0 sera convexe. Si la cible H correspond à un point du contour, ce point sera

identifiable avec le point d'inflexion de la courbe convexe. Par conséquent, nous supposons que la valeur mélodique de ce point sera plus élevée que celle trouvée dans les affirmations au même endroit temporel.

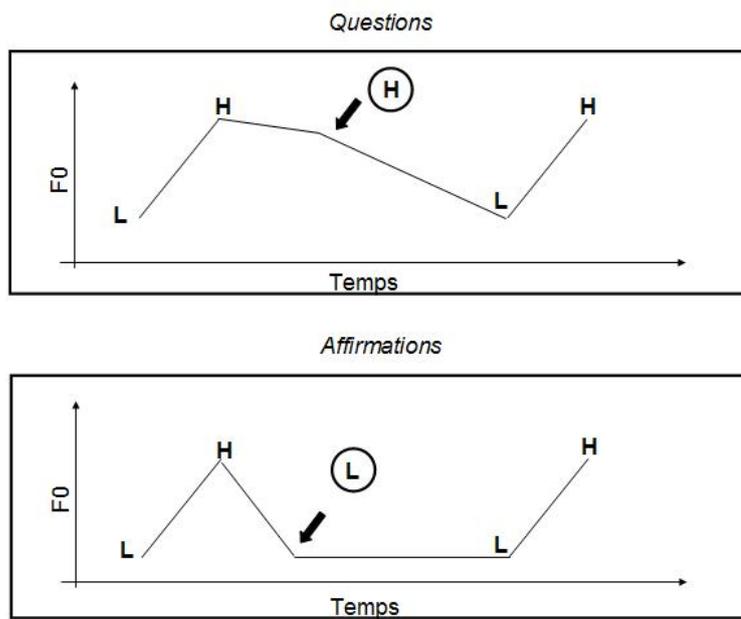


FIGURE 4.5 – Représentation schématisée de l’hypothèse d’insertion d’un ton différemment spécifié dans les questions (H, en haut) et dans les affirmations (L, en bas).

Sur le plan méthodologique, nous supposons aussi que cette différence ne sera pas capturée par les méthodes de détection des cibles déjà utilisées dans le chapitre 2. En particulier, notre hypothèse est que la méthode manuelle et les méthodes automatiques LSF et MOMEL ne seront pas capables de localiser une cible tonale dans les questions, car, dans cette modalité intonative, le ton haut n’est pas toujours réalisé comme un pic maximal de f_0 . En fait, les trois procédures présupposent que les cibles tonales se réalisent de façon “concrète” comme des points d’inflexion minimaux et maximaux de la courbe de f_0 . Au contraire, nous allons utiliser des techniques de modélisation de la courbe de f_0 pour quantifier les différences de forme de la f_0 dans les deux modalités intonatives. En particulier, la forme convexe de la f_0 dans les questions sera mieux prise en compte par un modèle de régression non-linéaire (la régression logistique) tandis que la forme concave des affirmations sera mieux décrite grâce à la régression linéaire par morceaux

(*piecewise linear regression*, voir chap. 2 pour une description plus détaillée de la régression logistique). La fig. 4.6 illustre les deux types de modélisation.

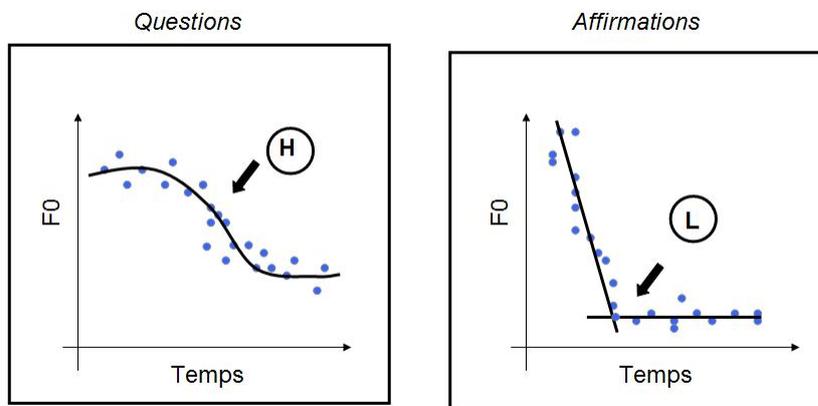


FIGURE 4.6 – Représentation schématisée de la modélisation de la courbe de f_0 dans les questions (régression logistique, à gauche) et dans les affirmations (régression linéaire par morceaux, en bas).

4.2.1 Corpus et enregistrement

Pour cette étude, nous avons utilisé le même corpus de six phrases décrit dans le chapitre 2. Dans ce corpus, six mots cibles étaient insérés dans la même phrase porteuse, avec une focalisation étroite sur le dernier syntagme nominal : [La mamma]_{NP} [vuole vedere]_{VP} [la X]_{NP} (“la maman veut voir la X”). A ces six phrases, nous avons ajouté les cinq phrases de remplissage utilisées dans cette expérience, qui étaient constituées par cinq mots propres paroxytons (Mina, Lina, Rina, Pina, Tina) en position finale de phrase et insérés dans la même phrase porteuse que les phrases cibles.

Les phrases ont été lues par deux locuteurs napolitains (AS et OM) comme des questions oui/non et comme des affirmations à focalisation étroite sur le NP en position d’Objet. Chaque phrase a été répétée sept fois. En particulier, nous avons utilisé à nouveau les phrases du locuteur AS enregistrées pour l’expérience décrite dans le chapitre 2 car, cette fois-ci, c’est sur la région de f_0 qui précède l’accent nucléaire que se portera notre attention. En fait, comme on se souviendra, AS avait lu ces phrases avec un accent pré-nucléaire montant sur *La mamma* suivi par un accent nucléaire montant sur le mot cible. Au contraire, les phrases enregistrées pour DD (voir chapitre 2) ne seront pas analysées

ici, car les productions de ce locuteur étaient caractérisées par une double focalisation (le mot *La mamma* étant lui aussi associé à un accent nucléaire). Par conséquent, un autre locuteur, OM, a été enregistré. De façon similaire à AS, ce locuteur a produit les phrases avec un accent prénucléaire montant sur *La mamma* suivi par un accent nucléaire (L^*+H pour les questions et $L+H^*$ pour les affirmations) sur le mot cible.

Sur un corpus de 308 phrases (2 modalités intonatives X 11 mots X 7 répétitions X 2 locuteurs), un total de 244 phrases a été analysé. L'analyse acoustique a été effectuée à l'aide de PRAAT. Nous avons marqué les frontières des mots de chaque phrase ainsi que leurs syllabes et segments. La localisation temporelle des cibles tonales des accents prénucléaires et nucléaires a été aussi effectuée. Pour cela, nous avons combiné la procédure LSF et la méthode manuelle. En particulier, le L et le H de l'accent nucléaire ont été trouvés principalement à l'aide de l'algorithme LSF, selon la procédure déjà décrite dans le chapitre 2. Le début de la montée prénucléaire a été localisée par le LSF au point d'intersection entre deux lignes droites ajustées dans la région entre le début de la phrase et le pic maximal de f_0 dans le mot *La mamma*. Cependant, les résultats obtenus pour la détection de cette cible n'étaient pas satisfaisants. Cela est vrai surtout pour AS : chez ce locuteur, le début de la montée est souvent placé juste après le début de la phrase, de sorte que la ligne à la droite du point d'intersection ne peut pas être ajustée à une portion de f_0 suffisamment large. Dans ces cas, la localisation de cette cible a été effectuée manuellement. De plus, le pic de l'accent prénucléaire a été localisé au point maximal de f_0 dans le mot prosodique. Même dans ce cas, la méthode manuelle a été préférée au LSF car nous avons remarqué que, à cause de la variabilité de réalisation dans la région qui suit immédiatement le pic prénucléaire, la qualité d'ajustement de l'algorithme était très faible.

4.2.2 Résultat

Durée

Avant d'analyser la pente interaccentuelle, nous avons mesuré la durée segmentale ainsi que l'alignement et la valeur mélodique des cibles des accents montants prénucléaires et

nucléaires. En fait, si les différences de pente sont authentiques, nous supposons qu'elles seront indépendantes des contraintes de pression temporelles dues aux variations de la vitesse d'élocution entre les deux modalités ainsi qu'aux ajustements temporels ou mélodiques de l'accent prénucléaire et de l'accent nucléaire.

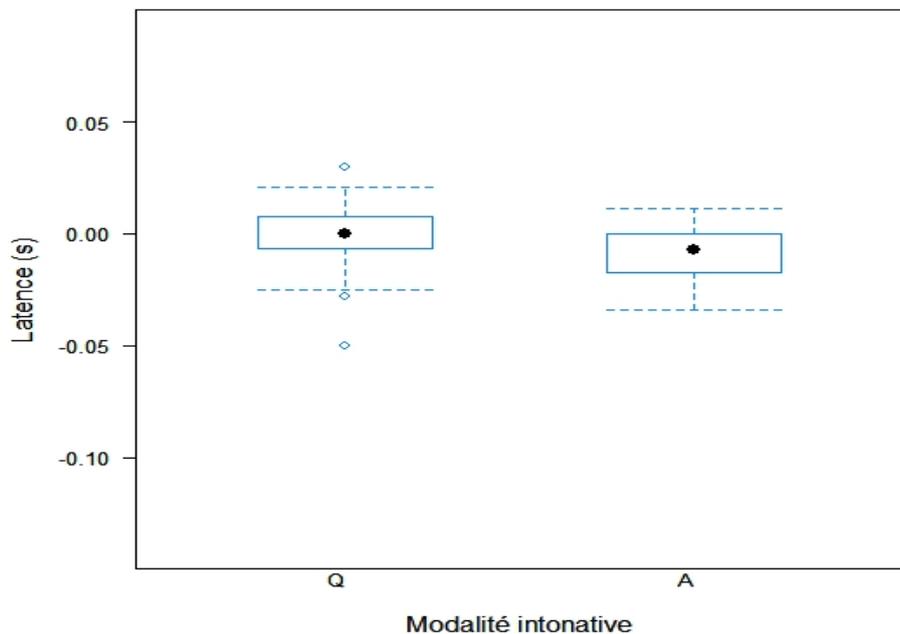
La durée de la phrase entière ainsi que celle du premier mot prosodique a été mesurée. Les résultats de nos analyses sont mixtes. Pour les deux locuteurs, la durée moyenne de la phrase était plus longue dans les questions (OM : 1.507 s ; AS : 1.439 s) que dans les affirmations (OM : 1.443 s ; AS : 1.358 s). Une analyse ANOVA à un facteur effectuée séparément locuteur par locuteur, a montré que l'effet de modalité est significatif pour OM [$F(1,105) = 29.7, p < 0.1$] ainsi que pour AS [$F(1, 135) = 51.1, p < .01$]. Ces résultats sont très surprenants, car ils sont en désaccord avec la tendance inter-langues à prononcer plus rapidement les questions que les affirmations. Cependant, lorsqu'on regarde la durée de *La mamma*, on remarquera que pour OM le mot prosodique tend à être prononcé plus rapidement dans les questions (0.389 s) que dans les affirmations (0.408 s), cet effet étant significatif [$F(1,105) = 7.91, p < 0.1$]. Les données pour ce locuteur pourraient donc indiquer que la distribution de la vitesse d'élocution n'est pas uniforme à l'intérieur des phrases. En fait, d'autres analyses effectuées sur OM ont montré que l'allongement de durée dans les questions était dû surtout à l'allongement du mot prosodique en position finale de phrase (Q : 0.596 s ; A : 0.539) [$F(1, 95) = 69.9, p < .01$]. Pour AS, la durée de *La mamma* est plus importante dans les questions (Q : 0.354 s ; A : 0.315) [$F(1, 135) = 51.1, p < .01$]. Globalement, les résultats de durée ne semblent pas conforter l'hypothèse que les différences de forme et de pente de f_0 soient dues simplement à la vitesse d'élocution.

Alignement tonal et valeur mélodique de l'accent prénucléaire

L'alignement des cibles prénucléaires, réalisé sur la syllabe accentuée *mam-* de *mamma*, a été calculé relativement à : l'attaque consonantique ($c0$), le début de la voyelle ($v0$), la fin de la voyelle ($v1$) et à la fin de la syllabe ($s1$). De plus, nous avons calculé l'alignement du pic de l'accent prénucléaire en fonction de la durée de la syllabe et de la voyelle accentuée. Puisque nous n'avons pas encore des critères objectifs pour la segmentation syllabique en présence des consonnes géminées, la fin de la syllabe accentuée a été placée à la moitié de

la gémignée. Les figures 4.7 et 4.8 illustrent les valeurs d'alignement pour les cibles L et H de l'accent prénucléaire respectivement pour OM et AS. Dans chaque graphique, l'alignement temporel des cibles (abscisses) est fonction de la modalité intonative (ordonnées). Pour un souci de cohérence avec l'illustration des résultats dans le chapitre 2, les valeurs de latence sont établies par rapport au début de la syllabe ($c\theta$).

OM :
Alignement du L prénucléaire avec c0



AS :
Alignement du L prénucléaire avec c0

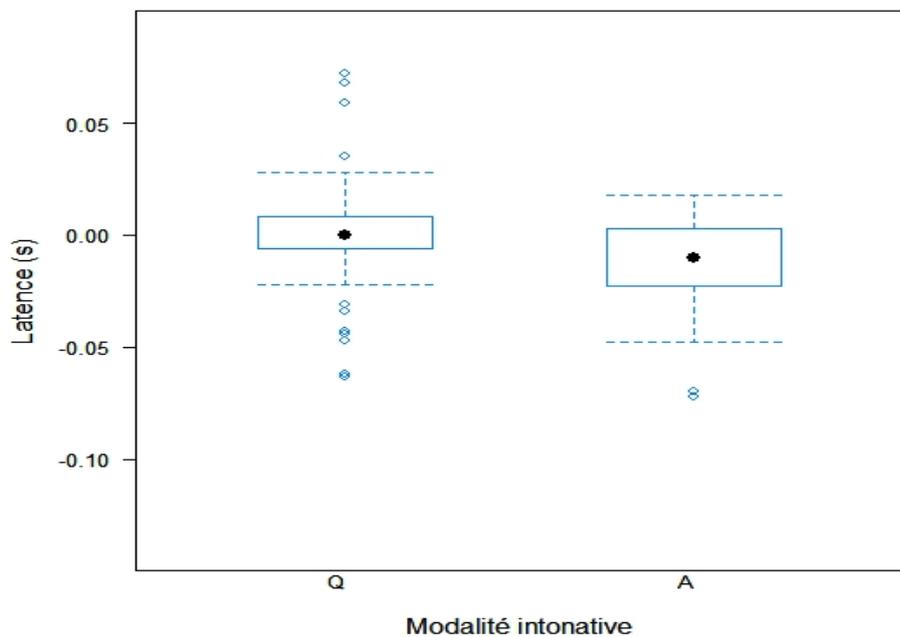
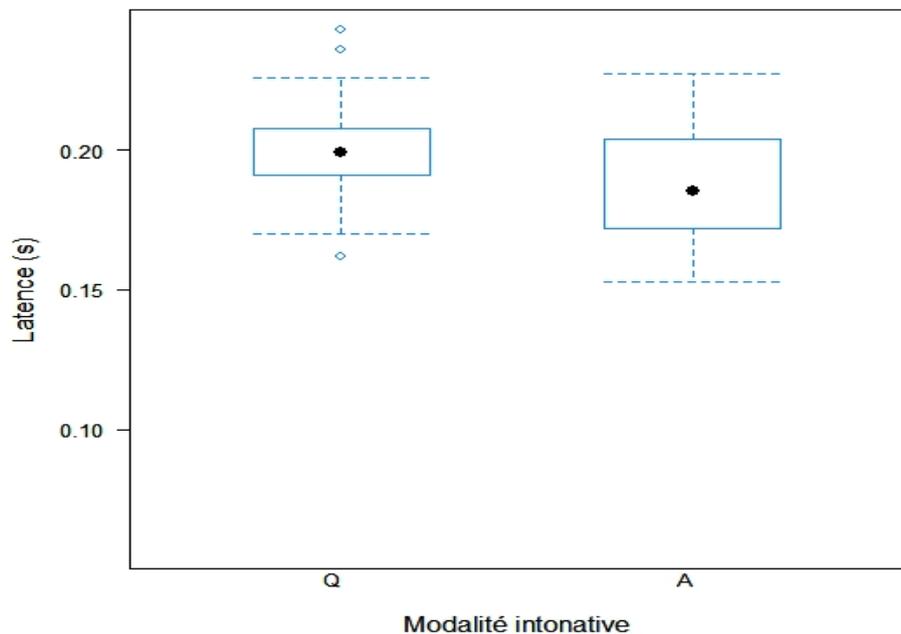


FIGURE 4.7 – Boxplots des latences de la cible prénucléaire L relativement au début de la consonne (*c0*) pour OM (en haut) et AS (en bas), dans les questions (Q) et dans les affirmations (A).

OM :
Alignement du H prénucléaire avec c0



AS :
Alignement du H prénucléaire avec c0

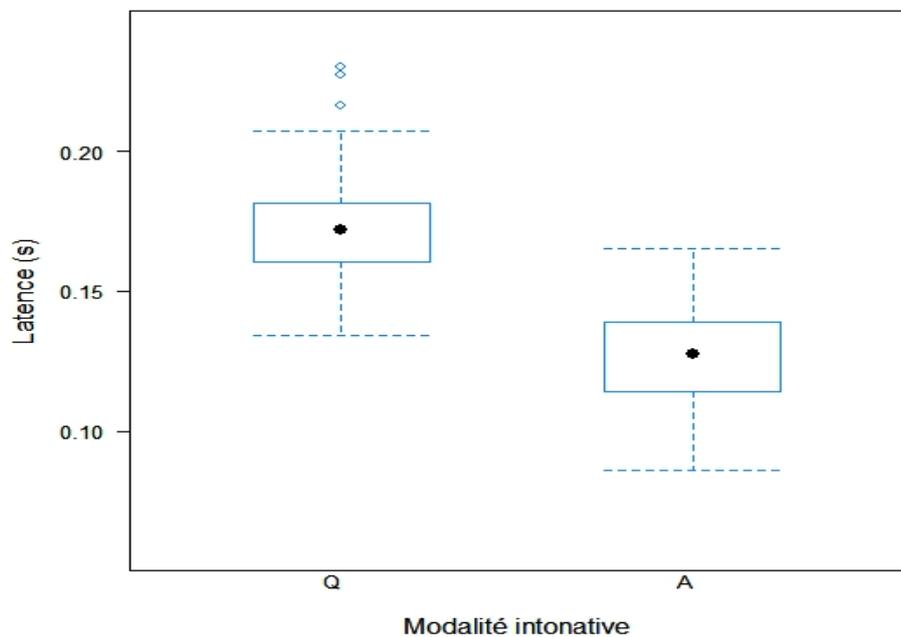


FIGURE 4.8 – Boxplots des latences de la cible prénucléaire H relativement au début de la consonne (*c0*) pour OM (en haut) et AS (en bas), dans les questions (**Q**) et dans les affirmations (**A**).

Les cibles L et H étaient réalisées respectivement en correspondance avec le début de la syllabe tonique et avec la portion finale de la gémée [m :] de *mamma* dans les deux modalités intonatives et pour les deux locuteurs. La fig. 4.7 nous montre que le début de la montée se produisait systématiquement à l’attaque syllabique pour OM, suggérant la présence d’un ton phonologique bas. Pour AS, la cible basse était réalisée en moyenne 12 ms avant l’attaque syllabique dans les affirmations, alors qu’elle était ancrée à l’attaque syllabique dans les questions. Le pic H (fig. 4.8) était aligné avec la coda consonantique, ce qui est cohérent avec l’alignement retardé typique des accents prénucléaires. Cependant, l’exacte localisation du pic accentuel est fonction de la modalité intonative : chez OM, il est réalisé à la fin de la consonne gémée dans les questions, et en son centre dans les affirmations ; chez AS, la cible H était au centre de la gémée dans les questions, et après la fin de la voyelle tonique dans les affirmations. L’analyse statistique révèle que chez OM l’effet de modalité intonative sur l’alignement de L n’est pas significatif, si ce n’est lorsque l’alignement est mesuré en relation avec le début de la syllabe *c0* [$F(1,105) = 4.12, p = .04$]. Cependant, lorsque l’alignement de L est calculé relativement à *v0* [$F(1,105) = 25.4, p < .01$], *v1* [$F(1,105) = 16.05, p < .01$] et *s1* [$F(1,105) = 17.07, p < .01$], cet effet est très significatif. Chez AS, au contraire, la différence d’alignement ressort uniquement lorsque la latence de L est mesurée par rapport à *s0* [$F(1,135) = 11.83, p < .01$], tandis qu’aucun effet n’a été trouvé sur l’alignement de L avec *v0*, *v1* et *c1* ($p > .01$). En ce qui concerne le pic prénucléaire, chez OM il n’y a pas de différence d’alignement avec *c0*, mais l’effet est significatif pour l’alignement avec *v0*, *v1* et *s1*. Aucun “ancrage” segmental semble exister pour H chez AS car son alignement est systématiquement plus retardé dans les questions que dans les affirmations, indépendamment de la frontière choisie pour mesurer l’alignement : *c0*, *v0*, *v1* et *s1* ($p > .01$).

De plus, nous avons trouvé un effet de modalité intonative lorsque l’alignement du pic accentuel était mesuré en fonction de la durée de la syllabe accentuée. Cet effet était proche du seuil de significativité chez OM ($[F(1,105) = 6.10, p = .01]$) et en dessous de ce seuil chez AS ($[F(1,135) = 99.52, p < .01]$). L’effet de modalité était significatif chez les deux locuteurs lorsque l’alignement du pic est mesuré en fonction de la durée de la voyelle tonique (OM : $F(1,105) = 12.33, p < .01$; AS : $[F(1,135) = 55.27, p <$

.01]). Ces résultats montrent que chez OM, l'alignement de L et H ne varie pas avec la modalité intonative, mais cela uniquement lorsqu'il est mesuré par rapport à $c\theta$. Cela suggère que les cibles tonales sont relativement alignées à la frontière gauche de la syllabe. Pour AS, l'alignement de L est stable uniquement lorsqu'il est mesuré à d'autres frontières segmentales, telles que le début et la fin de la voyelle, ainsi que la fin de la syllabe. De plus, chez AS la différence d'alignement de H entre les deux modalités est authentique, car elle a été trouvée indépendamment de la frontière ou de l'unité segmentale choisies pour mesurer l'alignement même.

En ce qui concerne les valeurs mélodiques, la valeur moyenne de f_0 de la cible L (fig. 4.9) était plus basse dans les questions (OM : 111 Hz ; AS : 158 Hz) que dans les affirmations (OM : 117 Hz ; AS : 178 Hz) chez les deux locuteurs. Chez OM, la valeur mélodique de H (fig. 4.10) était très similaire indépendamment de la modalité intonative (autour de 157 Hz), tandis que chez AS le pic accentuel était plus haut dans les questions (227 Hz) que dans les affirmations (207 Hz). L'ANOVA confirme nos observations : pour L, l'effet de modalité intonative est significatif chez les deux locuteurs (OM : $[F(1,105) = 49.8, p < .01]$; AS : $[F(1,135) = 25.4, p < .01]$) ; pour H, cet effet est significatif uniquement chez AS (OM : $[F(1,105) = 0.61, p > .01]$; AS : $[F(1,135) = 264.17, p < .01]$).

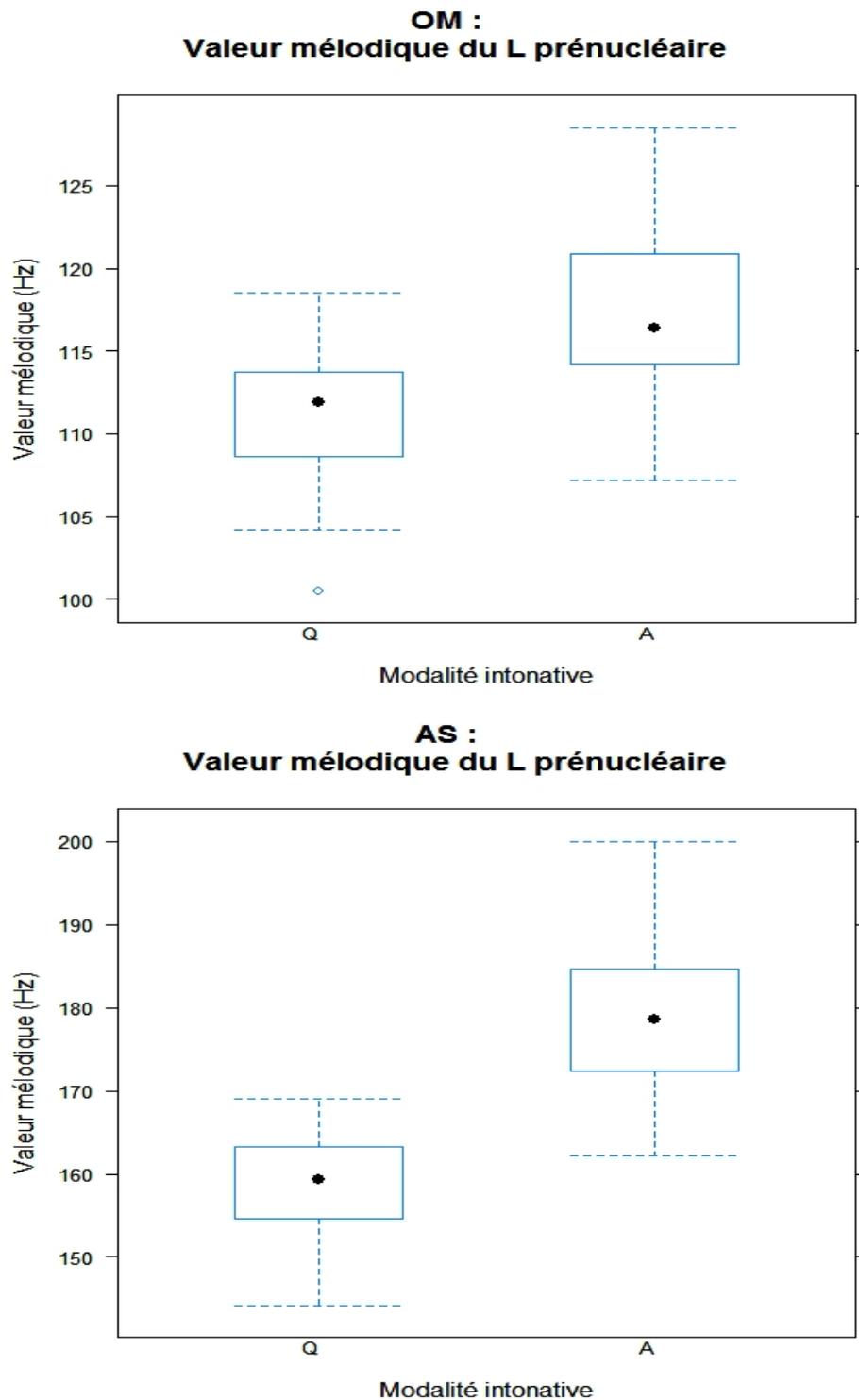


FIGURE 4.9 – Boxplots des valeurs mélodiques de la cible prénucléaire L pour OM (en haut) et AS (en bas), dans les questions (**Q**) et dans les affirmations (**A**).

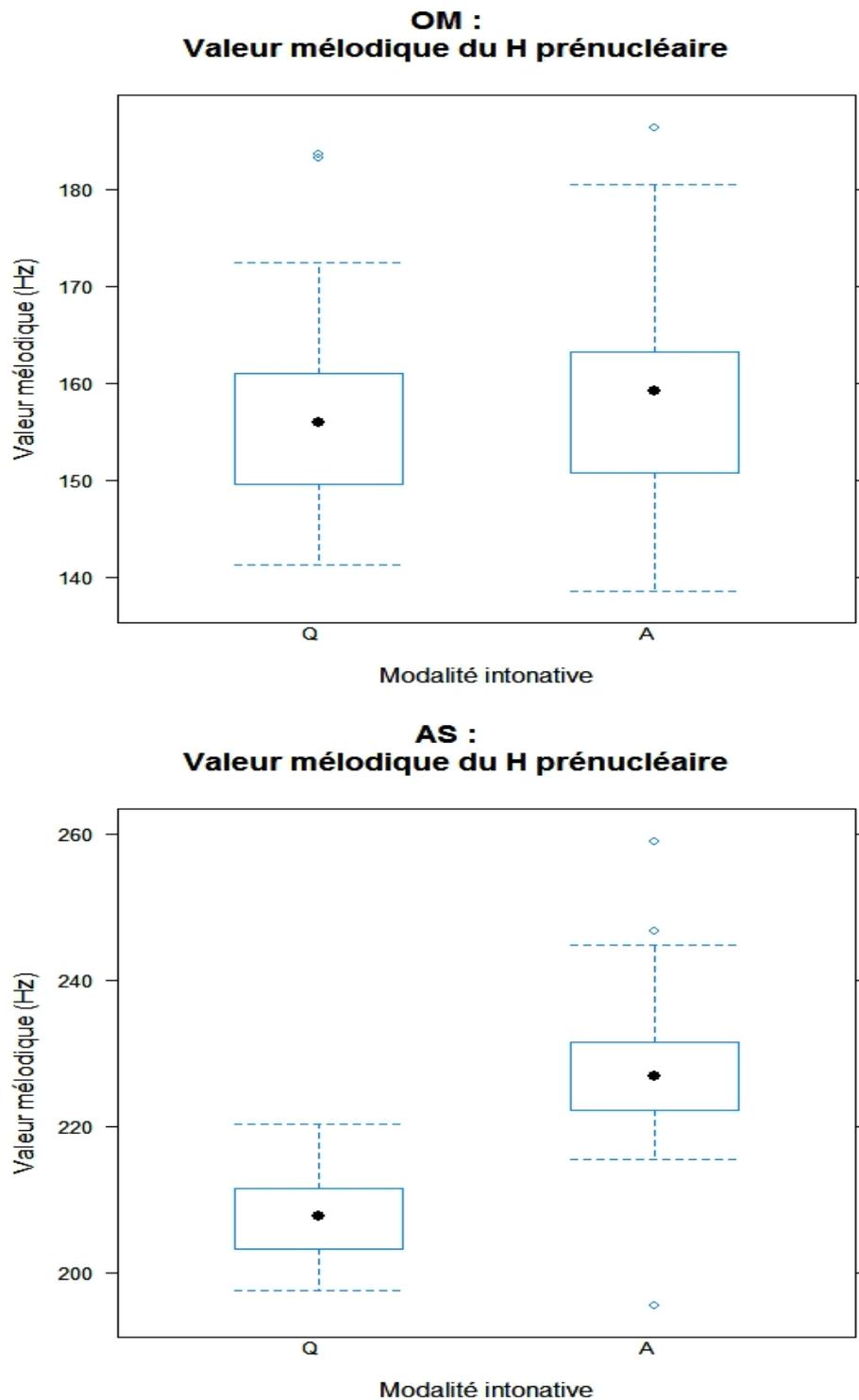


FIGURE 4.10 – Boxplots des valeurs mélodiques de la cible prénucléaire H pour OM (en haut) et AS (en bas), dans les questions (Q) et dans les affirmations (A).

Globalement, l'analyse acoustique de l'accent prénucléaire (LH*) montre que chez OM

la différence d'inclinaison de la pente interaccentuelle ne peut dépendre ni d'ajustements d'alignement ni d'ajustements mélodiques du pic prénucléaire, car sa réalisation ne varierait pas en fonction de la modalité intonative. Cependant, chez AS, le pic prénucléaire des questions est plus retardé et plus bas que celui des affirmations. Ce résultat pourrait suggérer un effet d'environnement tonal sur la pente et la forme de la descente de f_0 dans la région immédiatement suivante le pic prénucléaire. En particulier, l'alignement retardé du pic prénucléaire obligerait le locuteur AS à réaliser le mouvement descendant suivant dans un intervalle temporel plus court : la pente de la descente de f_0 serait donc plus raide pour un effet de " compression " (voir aussi citetCaspersetvanhauven93, Prieto95 pour des découvertes similaires). De plus, puisque dans les questions de AS le pic prénucléaire est plus bas, l'excursion mélodique entre ce pic et la fin de la descente de f_0 est plus petite, ce qui pourrait rendre la forme de la descente encore plus "aplatie". Finalement, l'effet de modalité sur l'alignement et la hauteur mélodique des cibles de l'accent prénucléaire L+H* soulève plusieurs questions sur l'interface phonétique/pragmatique, tout en suggérant l'existence de mécanismes de pré-planification du contour intonatif. Cependant, l'étude des effets possibles de modalité intonative sur l'implémentation de l'accent prénucléaire sera l'objet de nos recherches futures.

Alignement tonal et valeur mélodique de l'accent nucléaire

Nous avons aussi vérifié que la pente de la région interaccentuelle de f_0 peut être influencée par des variations temporelles ou mélodiques des cibles de l'accent nucléaire montant LH. Nous savons que les cibles tonales de l'accent nucléaire sont plus retardées dans les questions que dans les affirmations, et que l'effet de modalité peut aussi influencer leur valeur mélodique (voir chap. 2). Par conséquent, si l'alignement et la valeur mélodique du ton nucléaire L (c'est-à-dire de la cible tonale qui délimite l'extrémité droite de la région interaccentuelle) sont respectivement plus retardés et/ou plus bas dans les questions, on s'attend à ce que la pente interaccentuelle soit plus raide dans cette modalité. Pour l'alignement, nous avons trouvé que le L est ancré à l'attaque de la syllabe accentuée ($c\theta$) chez OM, bien que l'alignement du pic nucléaire relativement à l'attaque syllabique soit plus retardé dans les questions que dans les affirmations de 38 ms. La figure 4.11 illustre les

résultats d'alignement de l'accent montant pour ce locuteur. Nous avons effectué pour le locuteur OM une analyse de régression à effets mixtes, dans laquelle la modalité intonative était le facteur fixe et les mots étaient le facteur aléatoire. L'identité de la consonne d'attaque syllabique dans le mot ayant l'accent nucléaire n'a été pas prise en compte, car nous ne sommes pas intéressés ici aux effets de ce facteur. L'effet aléatoire des mots a été enlevé du modèle lorsque sa variance était proche de zéro. L'ANOVA du modèle a montré que l'effet de modalité n'est pas significatif sur L [$F(1,105) = 0.007, p < .01$], mais uniquement sur H [$F(1,105) = 44.86, p > .01$]². Chez AS, on voit que l'accent montant dans son ensemble était plus retardé dans les questions que dans les affirmations. En fait, l'effet de modalité est significatif sur l'alignement de L ([$F(1,125) = 28.67, p < 0.01$]) ainsi que sur celui de H ([$F(1,125) = 537.32, p < 0.01$]). Pour les figures concernant la réalisation des cibles nucléaires chez ce locuteur, nous renvoyons au chapitre 2.

2. Les résultats statistiques se réfèrent à l'alignement des cibles nucléaires avec l'attaque consonantique de la syllabe accentuée. En fait, chez OM l'alignement du ton nucléaire L avec cette frontière était stable, suggérant que la périphérie gauche de la syllabe est le point d' "ancrage" le plus probable pour ce locuteur. Pour le pic nucléaire H, les résultats de l'ANOVA sont très similaires indépendamment de la frontière ou de l'unité segmentale choisie pour mesurer l'alignement. Par souci de cohérence, les résultats d'alignement de H sont donc exposés relativement à l'attaque consonantique.

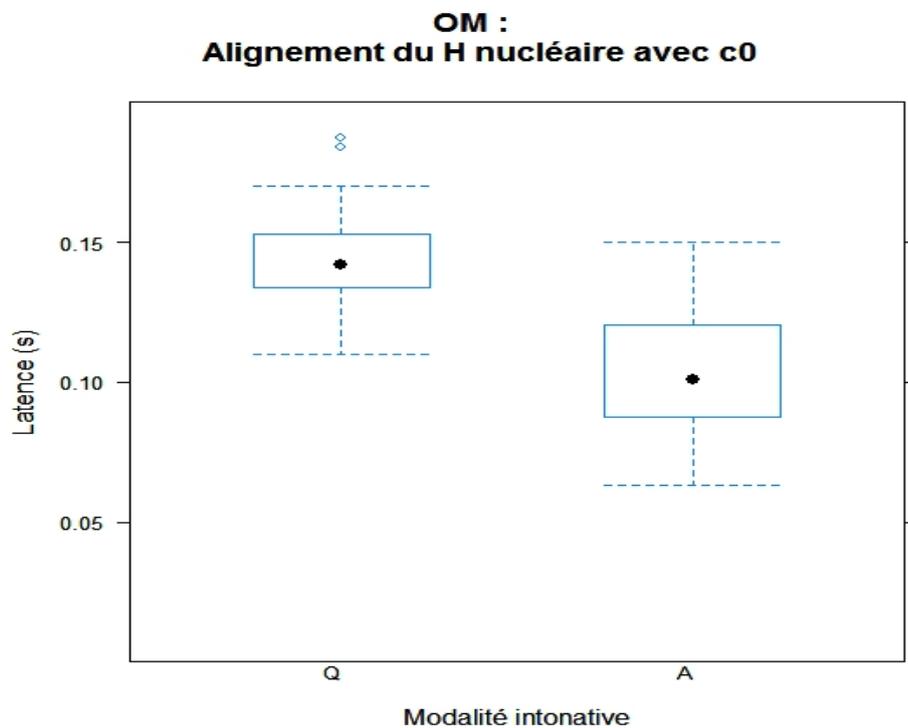
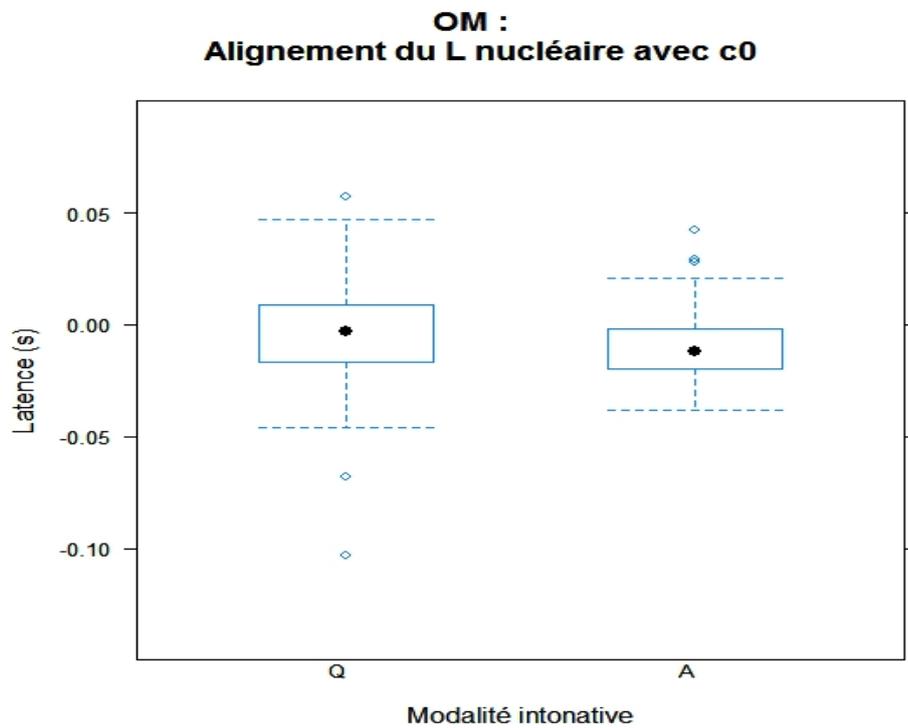


FIGURE 4.11 – Boxplots des latences des cibles nucléaires L et H relativement au début de la consonne (c_0) pour OM, dans les questions (**Q**) et dans les affirmations (**A**).

La valeur mélodique de L était elle aussi très stable quelle que soit la modalité interrogative chez OM (Q : 102 Hz ; A : 104 Hz), l'effet de modalité n'étant donc pas significatif [$F(1,103) = 3.25, p = .07$]. Cependant, pour ce même locuteur, le pic accentuel H est significativement plus élevé dans les questions (154 Hz) que dans les affirmations (129 Hz) [$F(1,105) = 262.33, p < .01$]. Les valeurs mélodiques des cibles nucléaires pour ce locuteur sont présentées dans la figure 4.12. Chez AS, l'effet de modalité est significatif pour L [$F(1,135) = 193, p < .01$] ainsi que pour H [$F(1,134) = 58.53, p < .01$], car chez ce locuteur, comme il est rappelé dans le chapitre 2, les deux cibles de l'accent nucléaire tendent à être respectivement plus basses et plus élevées dans les questions que dans les affirmations.

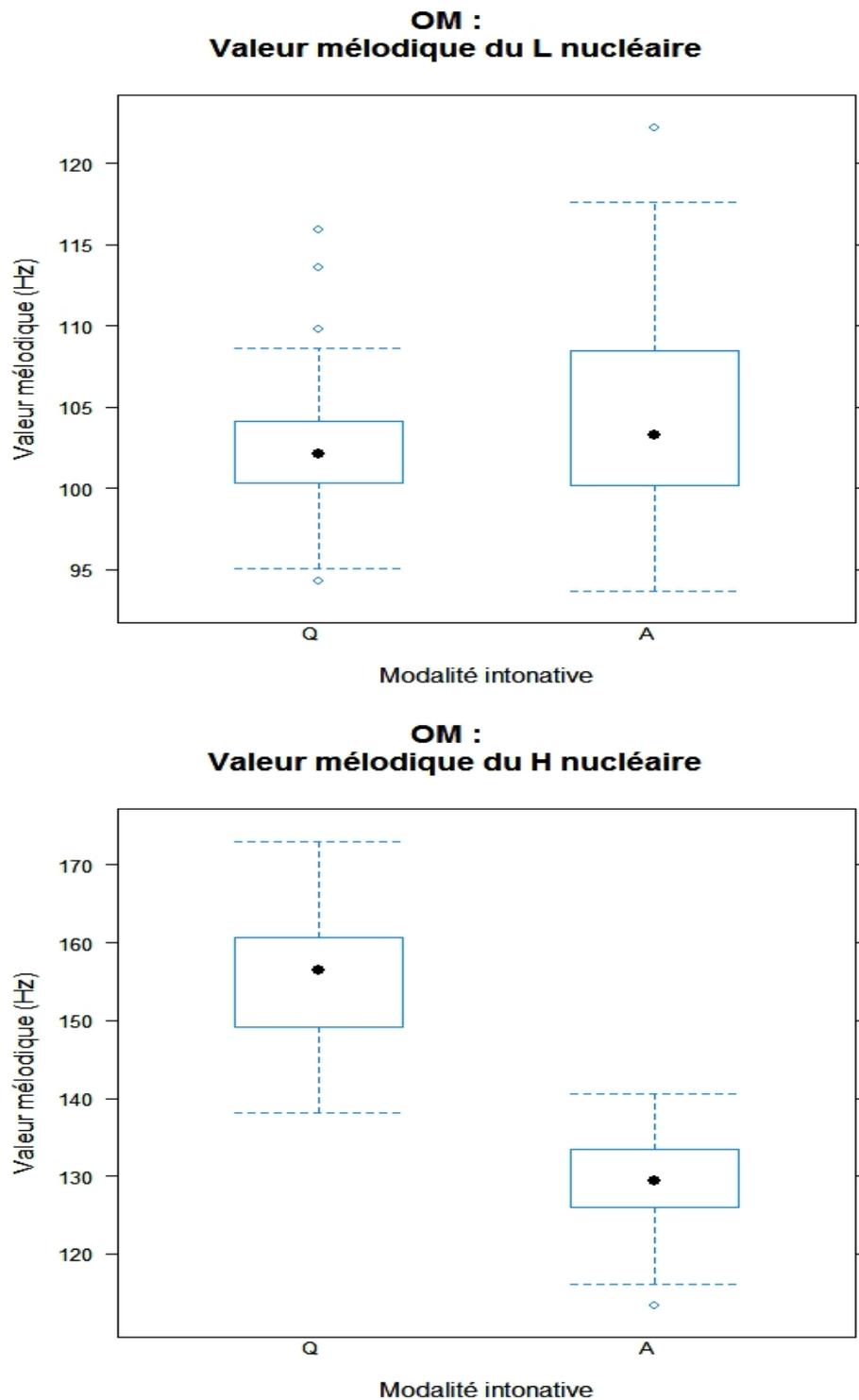


FIGURE 4.12 – Boxplots des valeurs mélodiques des cibles nucléaires L et H pour OM, dans les questions (Q) et dans les affirmations (A).

Pour résumer, chez OM les différences de forme et de pente de f_0 observées dans

la région interaccentuelle ne peuvent pas être provoquées par l’alignement ou la valeur mélodique de ton nucléaire L, car la réalisation de ce ton était très similaire dans les questions et dans les affirmations. En effet, pour ce locuteur le contraste de modalité dans l’accent mélodique se révèle uniquement dans l’alignement et la valeur mélodique du pic nucléaire. Au contraire, chez AS la différence de réalisation du ton nucléaire L pourrait avoir influencé la réalisation de la pente entre l’accent prénucléaire et l’accent nucléaire. Cependant, nous pensons que, bien que le L soit plus retardé dans les questions, cette subtile différence d’alignement (voir aussi le chapitre 2 pour des détails quantitatifs) ne peut pas expliquer immédiatement la différence de pente dans la région interaccentuelle chez ce locuteur.

Pente interaccentuelle

Dans les travaux inscrits dans le cadre de la Théorie AM, la pente de f_0 est souvent calculée comme le rapport entre l’excursion de f_0 et le temps employé pour effectuer la montée ou la descente de f_0 (Hz/ms), car on sait que l’interpolation entre des cibles de spécification tonale différente (L et H) est linéaire (voir, par exemple, Welby & Loevenbruck (2005)). Cependant, l’adoption *a priori* de ce type de mesure n’est pas justifiée pour évaluer l’inclinaison de la pente de f_0 comprise entre le pic prénucléaire et le L nucléaire des questions et des affirmations du napolitain. En fait, en première approximation, les affirmations semblent être caractérisées par la présence d’un minimum de f_0 à la fin de *La mamma*. De plus, dans les questions, le mouvement de descente de f_0 ne semble pas être uniforme, car il est beaucoup moins accéléré dans la région qui suit immédiatement l’accent prénucléaire que dans la portion de f_0 réalisée entre la fin de *La mamma* et le début de l’accent nucléaire.

Par conséquent, nous avons d’abord cherché à déterminer si un point d’inflexion maximal/minimal est présent par les trois méthodes de détection déjà employées dans le chapitre 2, la méthode manuelle, MOMEL et LSF. Cependant, l’inspection d’un certain nombre de cas nous a suggéré que les trois procédures sont insatisfaisantes pour expliquer la différence de forme de la région interaccentuelle. Ici, on va détailler les problèmes les plus fréquents rencontrés dans l’application des trois méthodes. L’utilisation de la mé-

thode manuelle était problématique surtout pour la localisation de ce point d'inflexion dans les questions. Dans cette modalité, la région suivant le H prénucléaire était caractérisée par une descente très douce des valeurs de f_0 , mais sans aucun pic maximal clairement identifiable. Par conséquent, il était difficile d'établir si un ton additionnel était présent, et, dans ce cas, son exacte localisation (fig. 4.13, en haut). Le choix de la localisation risquait d'être tout à fait arbitraire dans des cas comme celui illustré dans la fig. 4.3 (en haut) où, bien que l'interpolation entre le H prénucléaire et le L nucléaire ne soit pas strictement linéaire, il n'y a pas de point précis de changement de direction de la f_0 visible dans la trace acoustique. La méthode manuelle était plus précise dans les affirmations lorsqu'un minimum de f_0 était clairement visible dans la région suivant le H prénucléaire. Cependant, même dans cette modalité il était beaucoup plus difficile de localiser une cible basse en présence de perturbations microprosodiques (fig. 4.13, en bas).

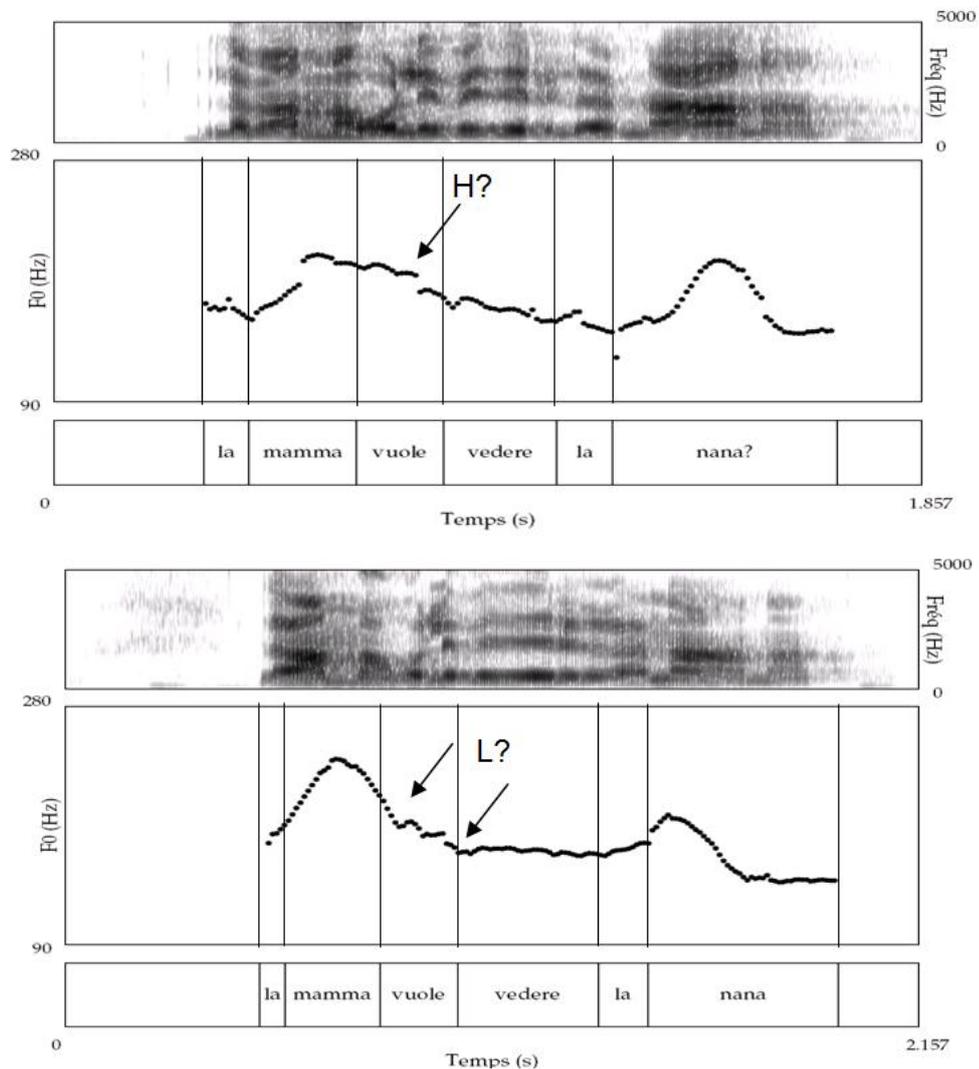


FIGURE 4.13 – Exemples de détection manuelle des cibles H et L dans la région suivant le H prénucléaire dans la phrase *La mamma vuole vedere la nana* (“La maman veut voir la naine”), prononcée comme une question (en haut) et comme une affirmation (en bas). Les flèches représentent de possibles localisations pour les cibles tonales. Locuteur AS.

L’utilisation de MOMEL n’a pas donné non plus de résultats satisfaisants car même dans les cas de détection apparemment les plus simples, il produisait des erreurs de localisation. En particulier, dans les affirmations, la cible localisée par cet algorithme est souvent beaucoup plus retardée par rapport à la position réelle du coude de f_0 , car elle est placée dans le plateau bas réalisé sur le syntagme verbal (fig. 4.14, en haut). De plus, dans certains cas, plusieurs points d’inflexion sont localisés par l’algorithme dans la région

interaccentuelle, de sorte que la stylisation de la courbe de f_0 pour la même phrase variait parmi les répétitions (fig. 4.14, en bas).

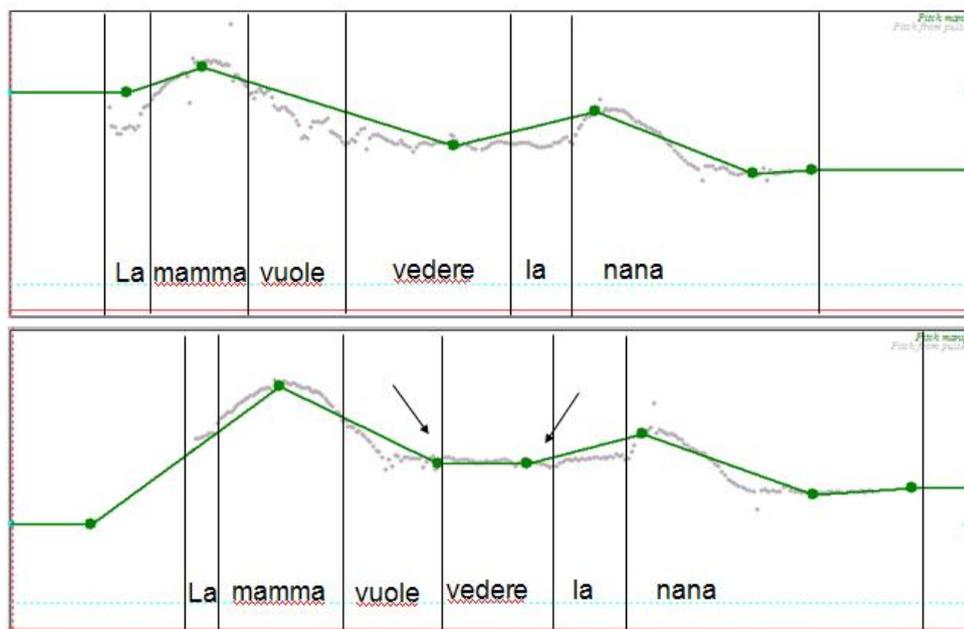


FIGURE 4.14 – Exemples de la détection, effectuée par MOMEL, de la cible L dans la région suivant le H prénucléaire dans la phrase *La mamma vuole vedere la nana* (“La maman veut voir la naine”), répétée deux fois comme affirmation par le locuteur AS. La ligne grise indique la courbe de f_0 . Dans chaque panneau, les points verts représentent les cibles détectées par l’algorithme pour chaque phrase, liées par des interpolations linéaires par le script MOMEL pour PRAAT.

Finalement, nous avons appliqué la méthode LSF pour vérifier l’existence d’une cible localisée en correspondance avec le point d’intersection de deux lignes droites ajustées dans la région interaccentuelle. Cette méthode semble être valide surtout pour la localisation de la cible basse dans les affirmations, tandis que beaucoup de variabilité dans la détection de la cible a été trouvée pour les questions. Dans la fig. 4.15, le point d’intersection des deux lignes ajustées à la région entre le pic prénucléaire et le début du dernier mot prosodique (*la nana* et *la rana*) est localisé tout de suite après le pic prénucléaire dans le premier exemple, mais près du début de l’accent nucléaire dans le deuxième exemple.

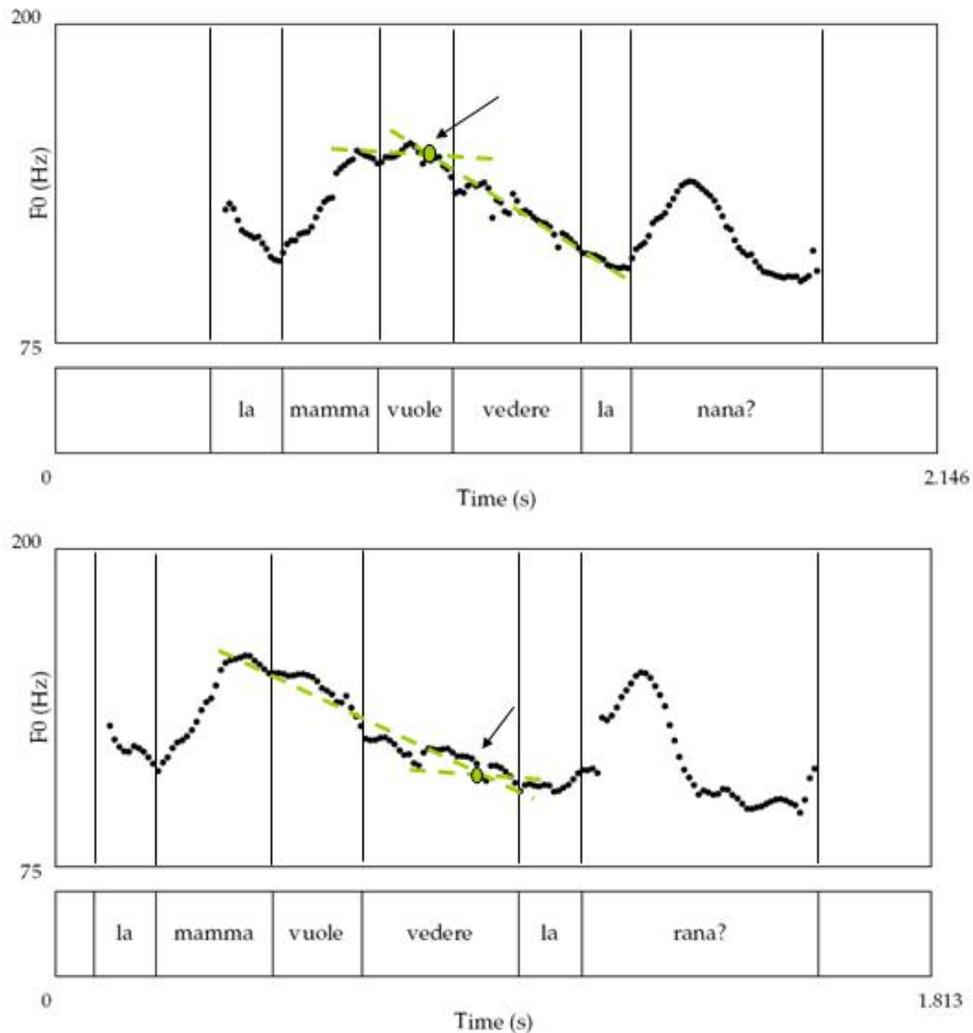


FIGURE 4.15 – Exemples de la détection, effectuée par le LSF, de la cible H dans la région suivant le H prénucléaire dans les phrases *La mamma vuole vedere la nana* (“La maman veut voir la naine”) et *La mamma vuole vedere la rana* (“La maman veut voir la grenouille”), prononcées comme questions par le locuteur OM. Dans chaque panneau, le point vert indique le point d’intersection des deux lignes droites ajustées à la région entre le pic prénucléaire et le début du dernier mot prosodique

On pourrait se demander si les erreurs de détection, qui sont plus importants surtout dans les questions, dépendent du fait que aucune cible n’est réalisée dans cette modalité. Dans les affirmations, au contraire, une cible L pourrait être réalisée entre le H prénucléaire et le L nucléaire. Cette hypothèse ne semble pas être justifiée, car elle impliquerait une différence structurelle entre les deux modalités. De plus, puisque la forme de la région qui suit l’accent prénucléaire est parabolique dans les questions, cela impliquerait

une interpolation non - linéaire entre l'accent prénucléaire et l'accent nucléaire seulement pour cette modalité intonative. Une alternative serait que les méthodes utilisées jusqu'ici ne puissent pas apprécier la différence de la forme du contour entre les deux modalités intonatives, car elles présupposent que les cibles tonales se manifestent comme des points d'inflexion minimaux et maximaux de la courbe de f_0 . Suivant ce raisonnement, on pourrait donc supposer que dans les questions la réalisation d'un ton entre le H prénucléaire et le L nucléaire n'est pas clairement visible comme un pic maximal de f_0 , de sorte qu'il est difficile de déterminer la position de la cible en se limitant aux méthodes de localisation utilisées jusqu'ici (chapitre 2).

Par conséquent, nous avons décidé de mieux analyser les propriétés dynamiques de la région interaccentuelle par sa modélisation statistique. En effet, nous avons supposé que si la différence de forme entre les questions et les affirmations est systématique, la région interaccentuelle devrait être modélisée différemment selon la modalité intonative, les questions nécessitant une technique de régression beaucoup plus sophistiquée que celle utilisée par la méthode LSF et MOMEL (voir chapitre 2). Pour cela, nous avons d'abord extrait les valeurs de f_0 de 40 points équidistants à partir du H prénucléaire jusqu'au L nucléaire. L'extraction des points a été faite à l'aide d'un script PRAAT écrit par Bert Remjisen et modifié par l'auteur. Ensuite, nous avons décidé de nous concentrer sur les premiers 20 points, c'est-à-dire sur la région entre le H prénucléaire et le milieu de la région interaccentuelle. Ce choix est dû au fait que (1) la différence de forme de f_0 était plus évidente dans la région entre la fin de *La mamma* et la fin de la parole suivante, *vuole*; (2) l'inspection d'une région de f_0 de taille plus petite nous a permis de minimiser les effets possibles de la microprosodie sur la modélisation de la courbe de f_0 . La normalisation des mouvements de f_0 dans le temps nous a permis d'examiner de plus près la forme du contour, indépendamment des variations possibles de vitesse d'élocution liées à la modalité intonative (van Heuven & van Zanten, 2005) ou au locuteur (tels que le sexe)³.

Les valeurs de f_0 (axe des ordonnées) normalisées dans le temps à 20 points de mesure (axe des abscisses) sont montrées dans la fig. 4.16 respectivement pour le locuteur OM (en

3. La distance moyenne entre un point et les successifs est cependant stable, car elle est de 17 ms chez OM (dans les questions et dans les affirmations) et de 19 ms (dans les questions) et 18 ms (dans les affirmations) chez AS.

haut) et AS (en bas). Pour chaque locuteur, les résultats sont présentés séparément pour les affirmations (gauche) et les questions (droite). Notre exploration a commencé par le lissage de la courbe de f_0 entre le pic prénucléaire H et le milieu de la région préaccentuelle à l'aide d'un modèle de régression polynomial localement pondéré (*local weighted polynomial regression*, ou tout simplement *lowess*, Baayen (à paraître)). Le principe sous-jacent au *lowess* est que la surface de régression est bien estimée par un polynôme de bas degré ajusté localement aux données ; par conséquent, comme d'autres fonctions de lissage, elle permet aux données mêmes de déterminer la forme de la courbe ajustée, sans hypothèse *a priori* sur la forme de la courbe.

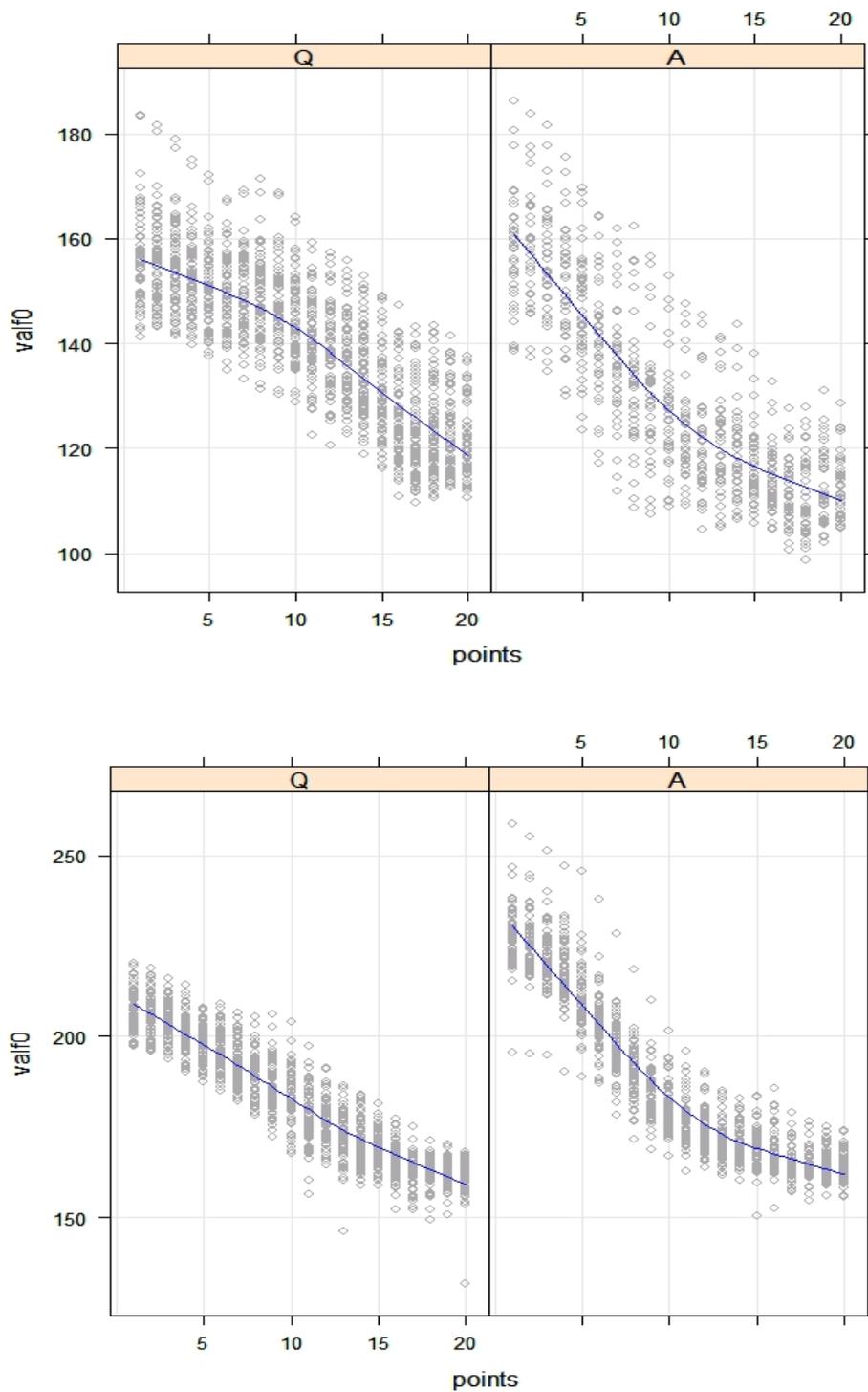


FIGURE 4.16 – Valeurs de f_0 à 20 points de mesure pour OM (en haut) et AS (en bas). Dans chaque graphique, les données sont illustrées séparément pour les questions (Q, à gauche) et les affirmations (A, à droite). Le lissage *lowess* est indiqué par la ligne continue.

Chez les deux locuteurs la courbe de f_0 a été réalisée de façon différente dans les affirmations et dans les questions. Dans les affirmations, la f_0 descend rapidement du pic accentuel prénucléaire jusqu'à la région autour des points 10-15, où l'inclinaison de la pente change brusquement et devient plus douce. Dans les questions, la forme de la région postaccentuelle de f_0 est réalisée de façon légèrement différente entre les deux locuteurs. Chez OM, cette région offre à la forme d'une parabole convexe vers le bas, dont le point d'inflexion est localisé autour des points 11-12. La parabole est plus plate dans les questions de AS, où les variations d'alignement et de valeur mélodique peuvent avoir influencé la réalisation de la pente interaccentuelle. Ici, en fait, la f_0 semble décliner de façon presque linéaire jusqu'au points 9-11, où un petit changement d'inclinaison de la pente est visible. Il faut noter cependant que la pente de la région de f_0 avant le point d'inflexion continue à être plus douce dans les questions que dans les affirmations même chez ce locuteur. En particulier, la valeur mélodique au point 11 (c'est à dire, dans la région où le point d'inflexion de la parabole semble être localisé) est plus basse que celle au point 1 (correspondant au H prénucléaire), tant pour le locuteur OM (138 Hz) que pour AS (187 Hz).

Notre hypothèse est qu'un ton est présent dans les deux modalités et dans la même région temporelle, mais qu'il est spécifié différemment dans les questions et dans les affirmations. Plus précisément, on suppose que la forme convexe des questions est due à la présence d'un ton haut H, responsable de maintenir élevées les valeurs de f_0 jusqu'au point 11. Après ce point, la pente de f_0 décline de façon linéaire jusqu'à ce que la cible L^* de l'accent nucléaire L^*+H soit atteinte. Au contraire, dans les affirmations la pente très raide dans la région postaccentuelle peut être due à l'insertion d'un ton L, après lequel un plateau bas de f_0 est réalisé jusqu'au début de l'accent nucléaire. Si cette hypothèse est vraie, nous nous attendons à ce que la région interaccentuelle dans les questions soit mieux modélisée par une régression non-linéaire, telle que la régression logistique. Pour les affirmations, la régression linéaire par morceaux (*linear piecewise regression*, Baayen, à paraître) devrait fournir des résultats plus satisfaisants. Pour les détails concernant la modélisation logistique, nous renvoyons au chap. 2. Dans la régression linéaire par morceaux, la courbe est estimée par l'utilisation de deux (ou plus) régressions linéaires simples, car on

suppose que la relation entre deux variables (tels que la valeur mélodique et le temps) est linéaire jusqu'à un certain point ; après ce point, la pente du modèle change brusquement. Chaque pente de régression est une *spline* et le point de changement de pente est défini comme "point de rupture" ou un "point d'inflexion" (*breakpoint*). Il faut noter que cette procédure de modélisation est la même que la technique de modélisation sous-jacente à l'algorithme LSF, où le coude de f_0 (correspondant au point d'inflexion de la régression par morceaux) est trouvé à l'intersection de deux lignes droites (correspondantes à deux *splines*). Cependant, la régression par morceaux est un modèle *statistique*, qui suppose que la valeur moyenne du *breakpoint* est déjà connue (son but principal étant d'obtenir les coefficients de régression des *splines*). Au contraire, le LSF n'est pas une procédure statistique, dans la mesure où elle permet de détecter la position du point d'inflexion séparément pour chaque observation.

Pour tester notre hypothèse, nous avons d'abord ajusté la courbe logistique à la région entre le pic prénucléaire et le milieu de la région interaccentuelle, séparément pour les questions et les affirmations. En accord avec nos prédictions, la régression logistique a montré un bon degré d'ajustement pour les questions. La modélisation de la courbe de f_0 a été effectuée à l'aide de l'environnement de travail R. Il faut noter que le développement des modèles non-linéaires dans R est relativement nouveau, de sorte qu'il reste encore très problématique d'évaluer de façon quantitative leur ajustement (par exemple, il n'y a pas de critère correspondant à celui du R^2 , utilisé souvent pour évaluer la qualité d'ajustement des régressions linéaires). Toutefois, la valeur très basse de l'erreur standard des résidus (OM : 8 ; pour AS : 5.6) ainsi que l'inspection des plots *quantile-quantile* (un plot diagnostic qui permet de déterminer si la distribution des résidus est normale, voir la fig. A.1 et fig. A.2 dans l'Appendix) nous indiquent que la région interaccentuelle des questions peut être bien modélisée à l'aide d'une courbe logistique. Il faut noter aussi que le point d'inflexion de cette courbe est assez proche pour les deux locuteurs, étant respectivement placé au point 13 pour OM et 9 pour AS. En particulier, chez les deux locuteurs ces points sont alignés avec la fin (pour OM) ou le début (pour AS) de la syllabe *vu-* de *vuole*, qui suit immédiatement le mot accentué *La mamma*. La valeur mélodique du point d'inflexion était de 137 Hz chez OM et de 187 Hz chez AS.

Lorsque le modèle est ajusté aux affirmations, l'algorithme n'arrive à converger ni chez OM ni chez AS, confirmant aussi que, contrairement aux résultats des questions, la région interaccentuelle des affirmations n'est pas caractérisée par une forme sigmoïde. Une régression linéaire par morceaux a donc été appliquée aux affirmations. Afin de trouver le point d'inflexion du modèle le plus probable, nous avons adopté une méthode déjà utilisée par Baayen (à paraître) : d'abord, une série de modèles (un pour chaque point possible d'inflexion de la courbe) a été ajustée pour cette modalité intonative ; ensuite, nous avons sélectionné le modèle pour lequel la déviance est la plus petite. Pour chaque locuteur, le meilleur modèle est celui dont le point d'inflexion était localisé au point 11 de la courbe normalisée de f_0 . La variance expliquée par la régression par morceau était très élevée, indiquant que le modèle choisi était bien ajusté aux données ($R^2 = 0.73$ pour OM et $R^2 = 0.93$ pour AS). En particulier, le point d'inflexion était localisé dans la première syllabe de *vuole*. Sa valeur mélodique est plus basse que celle trouvée pour les questions, car elle était de 122 Hz pour OM et de 174 Hz pour AS. De plus, chez les deux locuteurs la pente de f_0 est plus raide avant le point d'inflexion (valeur moyenne pour OM = -3.8 ; pour AS = -5.6) que après ce point (valeur moyenne égale à -1.3 chez les deux locuteurs). L'ajustement de la régression logistique (aux questions) et de la régression par morceaux (aux affirmations) est exposé dans les figures 4.17 et 4.18 respectivement pour OM et AS.

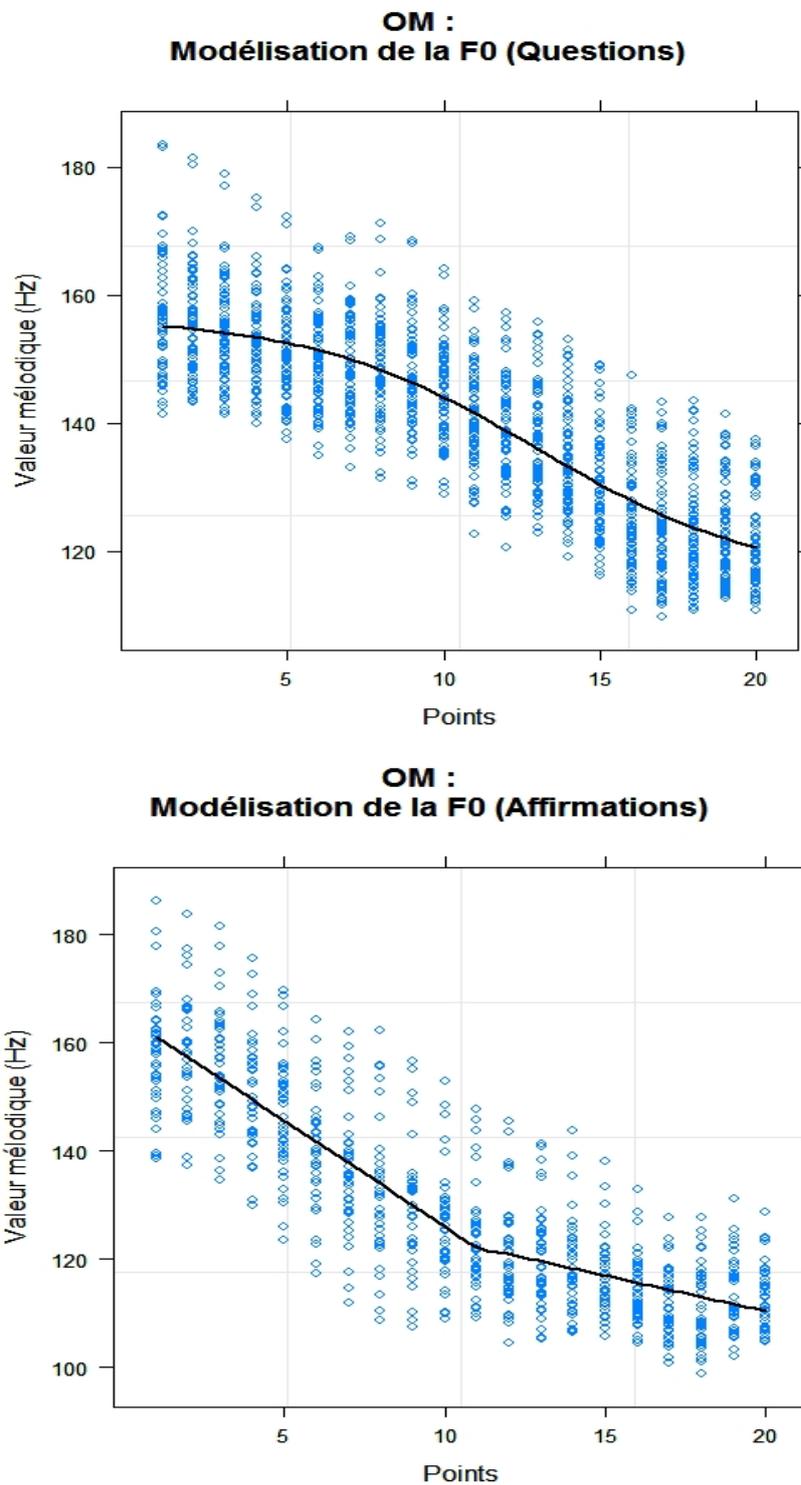


FIGURE 4.17 – Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression logistique, pour les questions (en haut) ; et par la régression linéaire par morceaux, pour les affirmations (en bas). Locuteur OM.

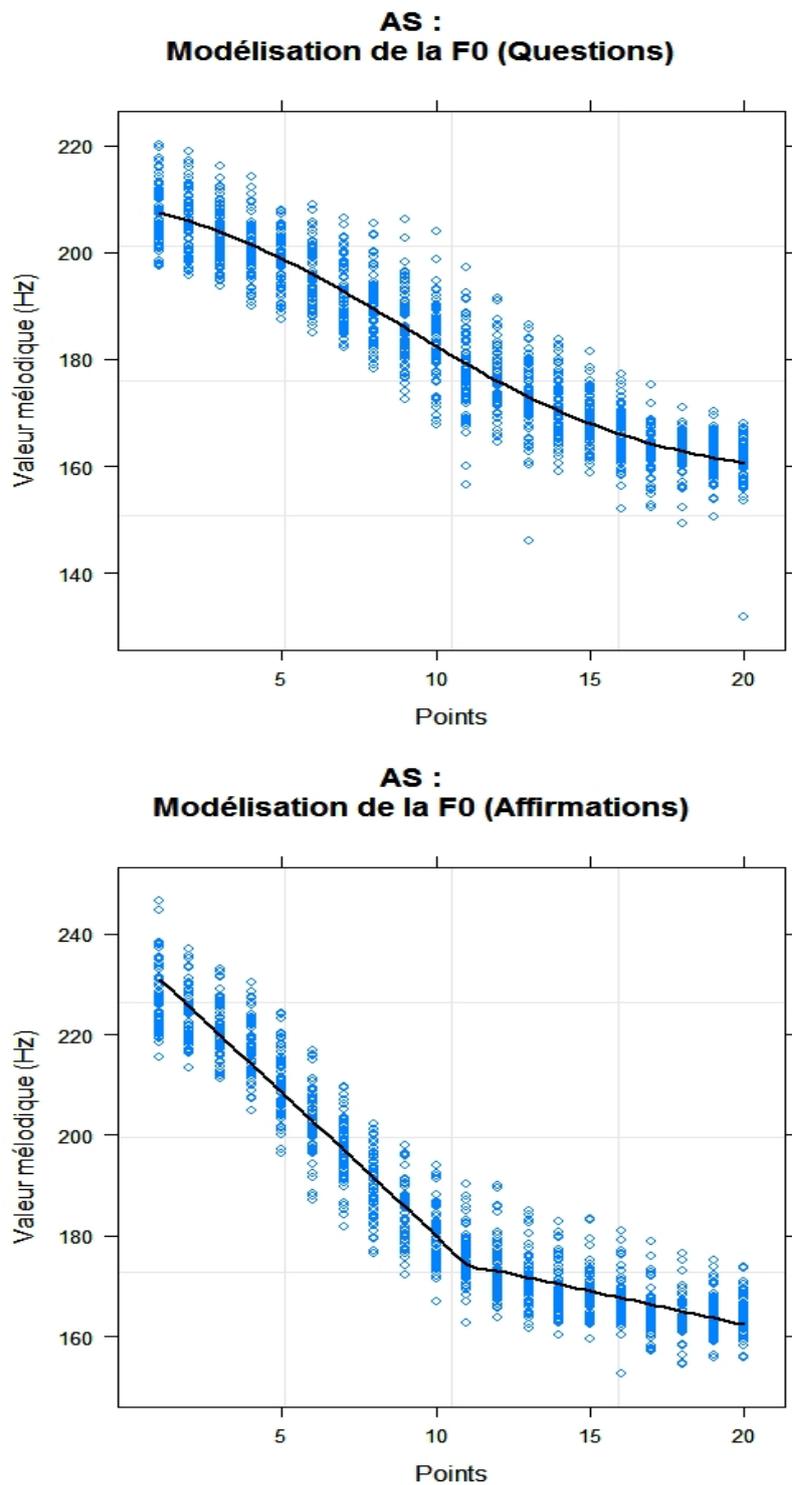


FIGURE 4.18 – Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression logistique, pour les questions (en haut) ; et par la régression linéaire par morceaux, pour les affirmations (en bas). Locuteur AS.

De plus, afin de comparer directement les deux modalités intonatives, nous avons décidé d'ajuster une autre régression par morceaux à la région entre les points 1-21 de chaque locuteur, dans laquelle le contraste entre les questions et les affirmations est la variable explicative. Le point d'inflexion a été placé au point 12 pour OM et 10 pour AS, c'est-à-dire la distance moyenne entre le point d'inflexion pour les questions et les affirmations. Pour les deux locuteurs, la valeur du point d'inflexion est plus élevée dans les questions (OM : 140 Hz ; AS : 182 Hz) que dans les affirmations (OM : 120 Hz ; AS : 178 Hz), et cette différence est significative pour les deux locuteurs (OM : $t = -26.16$, $p < .01$; AS : $t = -7.07$, $p < .01$). En outre, la pente est plus raide dans les affirmations que dans les questions *avant* le point d'inflexion (le coefficient de descente étant plus élevé dans les affirmations de 2.07 chez OM et de 2.9 chez AS), mais la situation s'inverse *après* ce point (le coefficient de descente étant plus élevé dans les questions de 1.57 chez OM et de 0.92 chez AS). En effet, l'analyse de régression nous a confirmé un effet significatif de la modalité intonative sur la pente avant (OM : $t = -16.38$, $p < .01$; AS : $t = -29.08$, $p < .01$) et après (OM : $t = 8.59$, $p < .01$; AS : $t = 7.6$, $p < .01$) le point d'inflexion. De plus, pour AS, le test du rapport de vraisemblance a montré que les questions étaient mieux modélisées par la régression linéaire par morceaux que par une régression linéaire simple (c'est-à-dire, par deux lignes droites plutôt que par une seule), ce qui soutient donc l'existence d'un point d'inflexion réel même pour ce locuteur [$F(2, 1418) = 43.7$, $p < .01$].

4.2.3 Discussion

En italien napolitain, les questions et les affirmations ne diffèrent pas uniquement dans le détail d'alignement de l'accent nucléaire. En fait, nous avons observé que lorsque deux accents montants sont réalisés, il y a aussi une différence dans la forme et la pente de la région interaccentuelle qui lie la cible H de l'accent prénucléaire et la cible L de l'accent nucléaire. L'existence d'une telle différence nous a permis d'exclure l'hypothèse que la région entre l'accent prénucléaire et l'accent nucléaire est caractérisée par une simple interpolation. Selon la Théorie AM, deux cibles H et L doivent être liées par une ligne droite. Cela n'est pas le cas dans les questions et les affirmations du napolitain : dans les questions, la région interaccentuelle a une forme parabolique ; dans les affirmations, un

coude bas de f_0 est visible après l'accent prénucléaire. De plus, nous avons exclu que la différence entre les deux modalités dépende de l'application d'autres règles d'interpolation, telles que la *tone linking rule*. En effet, suivant cette règle, la région interaccentuelle des questions aurait dû être caractérisée par une interpolation linéaire.

De plus, la différence de pente ne peut pas être due à une différence dans la spécification phonologique de l'accent prénucléaire. Chez les deux locuteurs, l'accent prénucléaire est caractérisé par une montée de f_0 , dont le début est localisé à l'attaque consonantique de la syllabe accentuée. De plus, le pic prénucléaire est aligné avec la fin de la syllabe. Ce patron est similaire à celui trouvé pour l'accent prénucléaire dans d'autres langues (Caspers & van Heuven, 1993; Prieto *et al.*, 1995; Arvaniti *et al.*, 1998; Ladd *et al.*, 1999; Ladd & Schepman, 2003, *inter alia*) et suggère que dans les deux modalités intonatives l'accent prénucléaire est montant. En particulier, l'analyse de cet accent comme (LH*) est en accord avec l'étude de Ladd & Schepman (2003), qui a montré que dans les accents prénucléaires de l'anglais, le minimum de f_0 , qui est aligné systématiquement avec l'attaque de la syllabe accentuée, constitue une cible phonologique (L). De plus, nos résultats sont en accord avec l'étude préliminaire de Gili Fivela & D'Imperio (2003) qui a montré qu'en italien il n'y a pas de distinction phonologique claire entre les deux accents prénucléaires H* et L+H*. Comme on l'a déjà souligné, la transcription des deux cibles entre parenthèse nous permet aussi de distinguer cet accent de l'accent nucléaire L+H*, dont le pic semble être caractérisé par une association secondaire avec la première mora de la syllabe accentuée (Prieto *et al.*, 2005).

En outre, nous avons exclu que la différence dans la région interaccentuelle soit fonction de la vitesse d'élocution ou d'ajustements mélodiques ou temporels des tons qui circonscrivent cette région (c'est-à-dire, le H prénucléaire et le L nucléaire). On se souviendra qu'une prédiction possible est qu'un ton L soit inséré dans les questions et dans les affirmations, mais que le mouvement tonal dans les questions ne peut pas atteindre sa cible à cause des contraintes de pression temporelle. Cette hypothèse n'est pas confirmée par nos données. Chez les deux locuteurs, la vitesse d'élocution ne semble pas jouer un rôle clair dans la distinction entre les questions et les affirmations. En ce qui concerne l'implémentation phonétique des tons de l'accent prénucléaire et de l'accent nucléaire, au-

cune différence entre les deux modalités n'a été trouvée pour le locuteur OM. Cependant, la région interaccentuelle des questions présente une courbature importante de la f_0 dans la région suivant l'accent prénucléaire. Cela indique que chez ce locuteur la différence de forme et de pente de la f_0 est authentique. La forme parabolique est moins convexe dans les questions d'AS. Cela peut être dû à un effet d'*undershoot* tonal : puisque chez AS les pics prénucléaires sont plus bas et plus retardés dans les questions que dans les affirmations, la transition entre le H prénucléaire et le ton réalisé dans la région postaccentuelle est aussi plus courte et plus aplatie. Il faut noter cependant que même chez AS nous avons trouvé une différence significative dans l'inclinaison de la pente avant et après le point d'inflexion de la courbe de f_0 .

Ces résultats nous ont conduits à considérer l'existence d'un ton additionnel dans la région suivant l'accent prénucléaire, spécifié différemment dans les questions (H) et dans les affirmations (L). Nous avons aussi supposé que, bien que dans les questions aucun pic de f_0 ne soit visible dans la trace acoustique, la présence d'un ton H aurait dû empêcher une chute brusque de la f_0 dans la région après le pic prénucléaire. L'utilisation d'une technique de régression non-linéaire a donc été nécessaire afin de mieux explorer les caractéristiques dynamiques de la courbe de f_0 . Cette fonction est plus sophistiquée que les fonctions linéaire et quadratique, qui sont sous-jacentes aux méthodes de détection des cibles tonales utilisées jusqu'ici (respectivement, le LSF et MOMEL), et elle permet donc une modélisation plus précise de la courbe observée. Les résultats de la régression logistique montrent la présence un ton H, qui est à l'origine de l'inclinaison douce de la pente dans le pic prénucléaire. De plus, la différence de forme du contour est confirmée par le fait que la régression logistique n'a pas pu être appliquée avec succès pour les affirmations. En fait, pour cette modalité (où un coude de f_0 était clairement visible dans la région postaccentuelle), la modélisation statistique de la courbe a été effectuée par un modèle de régression plus simple, la régression linéaire par morceaux. Deux autres preuves semblent soutenir notre hypothèse. En premier lieu, la valeur mélodique du point d'inflexion de la courbe de f_0 est plus élevée dans les questions que dans les affirmations. Cela suggère que la cible tonale des questions est haute (H), tandis que la cible tonale des affirmations est basse (L). De plus, cette différence de spécification tonale peut rendre compte

du changement d'inclinaison de la pente après le point d'inflexion. Dans les questions, la pente devient plus raide parce que la f_0 doit descendre de la cible H située après le pic prénucléaire jusqu'au L* de l'accent nucléaire L*+H. Dans les affirmations, la pente est beaucoup plus douce, car elle descend du L postaccentuel jusqu'au L de l'accent nucléaire L+H*.

Une fois que l'existence de ce ton a été établie, on peut alors se poser la question de sa représentation phonologique. Deux hypothèses sont possibles : le ton peut faire partie de l'accent mélodique prénucléaire ou bien être un ton de frontière. La première hypothèse doit être rejetée. Comme on l'a vu plus tôt, plusieurs preuves internes et externes nous ont amené à transcrire l'accent prénucléaire comme (LH*). Si le ton qui suit le pic prénucléaire est le ton *trailing* de cet accent, nous serons donc amenés à réanalyser l'accent prénucléaire comme une configuration tritonale, spécifiée différemment dans les questions (LHH) et dans les affirmations (LHL). Plusieurs arguments nous empêchent de favoriser cette hypothèse, tel que le problème de la multiplication de configurations qui ne sont pas possibles (Grice, 1995).

L'hypothèse favorisée ici est que ce ton marque la fin d'un constituant prosodique de rang inférieur au syntagme intermédiaire. Nous rappelons que chacune des phrases utilisées dans ce corpus a été prononcée comme un syntagme intonatif contenant un seul syntagme intermédiaire. L'accent nucléaire, en position finale de phrase, était donc suivi par un accent de phrase L- et un ton de frontière L%. Aucune rupture perceptive n'était réalisée après *La mamma*, dont la force est comparable à celle trouvée à la fin du syntagme intermédiaire. Pourtant, dans ces phrases un ton est réalisé juste après la fin du premier mot prosodique. Or, notre proposition provisoire est que ce constituant est le syntagme accentuel (*Accentual Phrase*, AP) et son domaine est celui de la proéminence accentuelle (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988). L'existence de ce constituant dans des langues à accent dynamique a déjà été suggérée par Beckman & Pierrehumbert (1986), et très récemment des preuves empiriques ont été apportées pour une des ces langues, l'arabe du Caire (Hellmuth, 2004, 2007). L'existence d'un constituant inférieur au syntagme intermédiaire en napolitain sera l'objet principal d'étude pour l'Expérience II décrite dans la section 4.3 de ce chapitre. De plus, la variabilité de forme et

de pente de f_0 reflète une régularité déjà trouvée par D’Imperio (2002a); Prieto *et al.* (2003) dans les constituants longs focalisés en napolitain. En fait, dans ces constituants, la f_0 ne descend pas après l’accent prénucléaire dans les questions, tandis qu’elle descend rapidement dans les affirmations.

On pourrait se demander aussi pourquoi le point d’inflexion de la courbe, qui devrait correspondre à la cible phonologique, est située au delà de la fin du premier mot prosodique (dans la première syllabe de *vuole*. Si les frontières du AP coïncident avec les frontières du mot prosodique, nous nous attendrions à ce que le ton de *AP* soit réalisé à la fin de *La mamma*, et non après. Cependant, des cas de violation de contraintes d’alignement ont déjà été répertoriés dans la littérature. Par exemple, dans sa description du hollandais de Roermond, Gussenhoven (2000) a montré que le ton de frontière H% s’associe optionnellement à la syllabe tonique, mais uniquement lorsque sa deuxième mora n’est pas occupée par un ton précédent. De contraintes de pression temporelle pourraient aussi expliquer l’alignement du ton de AP en italien : le ton de *AP* tend à être réalisé à la fin de l’*AP*, mais uniquement lorsque les conditions d’espace temporel sont favorables. Dans l’expérience décrite ci-dessous, le pic prénucléaire était réalisé dans la consonne geminée de *La mamma*. Par conséquent, il est possible que dans ces phrases il n’y avait pas d’espace suffisant pour compléter le mouvement descendant du pic H* jusqu’au ton de *AP*. Si cela est vrai, on s’attendrait à ce que ce ton soit réalisé à la fin de l’*AP* lorsque la distance entre la syllabe accentuée et la fin de l’AP est suffisamment large. L’expérience II nous éclairera sur ce point.

Enfin, deux derniers points méritent d’être soulignés. Dans notre analyse phonologique, la variabilité de forme et de pente de f_0 a été interprétée comme le résultat d’une différence dans la spécification tonale de tons H_{AP} et L_{AP} . Puisque cette thèse est inscrite dans le cadre de la Théorie AM, nous n’avons pas pris en considération le fait qu’une telle différence puisse être le résultat d’une interpolation non-linéaire, puisque de telles interpolation ne sont pas permises par la Théorie⁴. De façon similaire, nous n’avons pas considéré

4. D’un point de vue mathématique, l’interpolation *sagging* entre deux pics successifs n’est pas non plus une interpolation non-linéaire. Elle peut être définie plutôt comme une interpolation linéaire non monotonique car l’insertion d’une cible phonétique basse entre les deux pics provoque une interpolation linéaire entre chacun des deux pics

l'hypothèse selon laquelle cette variabilité dépende de l'existence de tons dynamiques, car un présupposé de la Théorie AM est que les unités de base de l'intonation sont des tons statiques définis uniquement par leur alignement et leur valeur mélodique. Cependant, l'importance phonologique des propriétés dynamiques du contour a été récemment souligné par plusieurs auteurs dans des cadres de travail différents (Kohler, 1991; Xu & Wang, 2001; Niebuhr, 2003, 2007; D'Imperio, 2000), de sorte que nous ne pouvons pas exclure d'autres interprétations de nos données. Cela va au delà des buts de cette thèse, mais il sera l'objet de nos études dans le futur.

4.3 Expérience II

Selon la Théorie AM le contour intonatif est composé par des syntagmes phonologiques organisés de façon hiérarchique, tels que le syntagme intonatif (*Intonational Phrase, IP*), le syntagme intermédiaire (*intermediate phrase, ip*) ou le syntagme accentuel (*Accentual Phrase, AP*; voir aussi le par. 1.3.2 dans l'introduction de cette thèse). Ces constituants, marqués par des tons périphériques (tels que les tons de frontière, les accents de syntagme et les tons de syntagme accentuel), peuvent être signalés également par d'autres indices de nature acoustique et perceptive. Par exemple, un corrélat important de constituance prosodique est le degré d'allongement des segments ou des unités prosodiques proches des frontières des constituants (*preboundary lengthening* ou, plus généralement, *final lengthening*). L'allongement final tend à être plus important à la fin du syntagme intonatif qu'à la fin des constituants de niveaux plus bas (Beckman & Edwards, 1990, *inter alia*). Le domaine de l'allongement reste cependant problématique. Plusieurs études dans la littérature (concernant surtout l'anglais) ont observé que l'allongement final se produit sur la syllabe ou la rime à la fin du constituant (Klatt, 1975; Wightman *et al.*, 1992, *inter alia*). Cependant, des travaux dans d'autres langues ont montré que l'allongement final peut s'étendre jusqu'à la syllabe tonique pénultième ou antepénultième du mot en position finale de constituant. De plus, la distribution de l'allongement n'est pas uniforme, car le degré d'allongement dans les syllabes qui se trouvent entre la syllabe tonique et la syllabe finale est beaucoup moins important (Turk & Hufnagel, 2007). D'un point de vue

perceptif, les niveaux de constituance prosodique peuvent être signalés par de degrés différents de rupture perceptive, selon la position du constituant à l'intérieur de la hiérarchie prosodique. Donc, la fin du syntagme intonatif est marquée par une rupture perceptive plus importante que celle qui marque la frontière droite du syntagme intermédiaire ou du syntagme accentuel. Dans la transcription ToBI (*TOne and Break Index*, Beckman & Elam, 1994) de l'anglais, la discontinuité perceptive à la fin du syntagme intonatif est marquée par un *index* (B4) supérieur à celui du syntagme intermédiaire (B3).

En italien, le syntagme intonatif a été attesté dans plusieurs variétés régionales (Grice *et al.*, 2005). Des preuves de son existence ont été aussi fournies dans le cadre de la Phonologie Prosodique (Nespor & Vogel, 1986; Frascarelli, 1997). L'existence du syntagme intermédiaire est plus problématique, et à notre connaissance des preuves empiriques ont été fournies uniquement pour les variétés parlées à Florence, Bari et Naples (voir (Grice *et al.*, 2005) et références y citées). En particulier, D'Imperio (2002a) a montré que dans les questions du napolitain les constituants longs focalisés sont caractérisés par un mouvement descendant de f_0 , dont le début est ancré à la dernière syllabe tonique du constituant. Cet accent, similaire au *sentence accent* du suédois (Bruce, 1977), a été analysé comme un accent de phrase HL- qui délimite la frontière droite du syntagme intermédiaire.

Bien que la structure en constituants prosodiques et la structure en constituants syntaxiques ne soient pas isomorphes, les informations transmises par le module syntaxique peuvent influencer le découpage prosodique (*prosodic phrasing*). Par exemple, en italien, comme dans d'autres langues, les propositions relatives provoquent souvent l'insertion d'une frontière de syntagme intonatif à la périphérie du constituant (Nespor & Vogel, 1986).

De plus, D'Imperio *et al.* (2007b) ont trouvé qu'en italien, dans des affirmations de type S-V-O (Sujet-Verb-Objet), une frontière de constituant tend à être insérée à la fin du Sujet lorsqu'il est composé d'une tête nominale ramifiée au niveau syntaxique (*La boliviana di Badalona*) ou prosodique (*Esmeraldina Secondigliano*). Ce constituant, correspondant probablement au syntagme intermédiaire (Frota *et al.*, 2007; D'Imperio *et al.*, 2007b), est signalé par plusieurs indices. D'Imperio *et al.* (2007b) précisent que dans leur corpus pour l'italien (constitué par des phrases affirmatives à focalisation large), la fin de ce constituant

est signalée par un ton de périphérie haut, réalisé à la frontière droite du constituant sous la forme d'une montée de continuation (surtout après un accent montant) ou d'un plateau de f_0 (surtout après un accent descendant). Frota *et al.* (2007) relèvent aussi la présence d'une rupture prosodique à la fin du Sujet (S).

A part le syntagme intonatif et intermédiaire, aucun autre niveau de constituance prosodique n'a été attesté en italien jusqu'ici. Cependant, dans le chapitre précédent, nous avons suggéré que la variabilité de forme et de pente de f_0 trouvée dans les questions et dans les affirmations du napolitain dépende de l'insertion d'un ton périphérique entre l'accent prénucléaire (LH*) et l'accent nucléaire montant des questions (L*+H) et des affirmations (L+H*). En particulier, notre hypothèse est que ce ton marque la frontière droite d'un constituant de niveau bas, constituant que nous avons provisoirement défini comme le syntagme accentuel.

L'idée de l'existence de l'*AP* en italien est motivée par plusieurs observations. En premier lieu, l'*AP* a été attesté dans de nombreuses langues à accent non-dynamique (*non-stress accent languages*, c'est-à-dire, des langues dans lesquelles la proéminence métrique est marquée principalement par la f_0 , Beckman (1986)), telles que le japonais (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988) ou le coréen (Jun, 1993). Parmi ces langues, on retrouve aussi une langue romane, le français (Jun & Fougeron, 2002). De plus, comme il a été admis par (Beckman & Pierrehumbert, 1986), l'existence de ce constituant est théoriquement possible pour les langues à accent dynamique (*stress accent languages*, où la proéminence est signalée par la combinaison de l'intensité, de la f_0 et de la durée, Beckman, 1986). En effet, récemment, des preuves empiriques de l'existence de l'*AP* ont été trouvées pour une langue à accent dynamique, l'arabe du Caire (Hellmuth, 2004, 2007).

De plus, dans une approche désormais très ancienne de la Phonologie Prosodique (Nespor & Vogel, 1986), il a été affirmé qu'un niveau de constituance typique de l'italien est le syntagme phonologique (*phonological phrase*, ϕ), qui inclut une tête lexicale et tous les compléments qui se trouvent dans le côté non-récursif (le côté gauche en italien). Par exemple, selon Nespor & Vogel (1986) un syntagme tel qu'*il benevolo manovale* (litt. "le bienveillant ouvrier") constitue un seul ϕ , car l'adjectif *benevolo* est à gauche de la tête

manovale. Au contraire, *l'agronomo di Benevolo* (litt. “l’agronome de (M.) Benevolo”), est constitué par deux ϕ , *l'agronomo* et *di Benevolo* car le syntagme prépositionnel est à la droite de sa tête. Il faut noter que, à cause de la définition strictement syntaxique du ϕ , il n’y a pas de correspondance univoque entre ce niveau et ceux définis par la Théorie AM. Par conséquent, le domaine de ϕ semble correspondre parfois au domaine du syntagme intermédiaire, mais aussi, parfois, à celui du syntagme accentuel.

Dans cette section, nous allons aborder la question de l’existence d’un troisième niveau de constituance, de rang inférieur au syntagme intermédiaire et au syntagme intonatif : le syntagme accentuel. S’il existe, nous nous attendons à ce que la fin de ce constituant soit marquée par un ton de périphérie différent de l’accent de phrase et du ton de frontière. Dans l’Expérience I, nous avons trouvé qu’un ton de périphérie est inséré entre l’accent prénucléaire et l’accent nucléaire, c’est-à-dire dans la région de f_0 entre le syntagme nominal *La mamma*, et le syntagme verbal *vuole vedere la X*. Nous n’avons pas vérifié si l’insertion de ce ton est influencée par des contraintes syntaxiques (car la fin de *La mamma* coïncide avec la fin de ϕ) ou par des contraintes purement prosodiques (car la fin de *La mamma* pourrait aussi coïncider avec la fin d’un *AP*). Si l’insertion du ton de frontière dépend de contraintes syntaxiques, nous nous attendons à ce qu’il n’y ait pas de ton de frontière entre *il benevolo* et *il manovale*, mais uniquement entre *l'agronomo* et *di Benevolo*. Au contraire, si le constituant est défini uniquement sur la base des contraintes prosodiques, le ton de frontière d’*AP* sera réalisé à la fin du premier syntagme dans les deux cas. De plus, dans l’Expérience I, nous avons montré que la spécification phonologique de ce ton dépend de la modalité intonative, car il est bas dans les affirmations et haut dans les questions. Par conséquent, nous émettons l’hypothèse que le ton d’*AP* sera spécifié comme L_{AP} dans les affirmations et comme H_{AP} dans les questions.

Ce troisième constituant sera identifié non seulement par le ton de frontière, mais aussi par d’autres indices de *phrasing*. En particulier, le degré d’allongement final et de rupture perceptive seront moins importants à la fin du syntagme accentuel qu’à la fin du syntagme intermédiaire et du syntagme intonatif. En ce qui concerne l’allongement final, il faut noter qu’en italien aucune étude systématique n’a été faite pour tester dans quelle mesure les niveaux de constituance prosodique influencent l’application et la dis-

tribution de l’allongement final sur les segments proches des frontières prosodiques⁵. Par conséquent, nous ne savons pas si le mécanisme d’allongement final en italien s’applique à la syllabe final du constituant (comme en anglais), ou s’il s’étend à d’autres portions du constituant même. Cela aussi sera vérifié dans notre expérience.

Au niveau méthodologique, nous avons montré dans l’Expérience I que l’utilisation de méthodes sophistiquées de modélisation de la courbe de f_0 peut nous aider dans la détection acoustique des cibles tonales. En particulier, nous avons suggéré que, dans les questions, le ton haut coïncide avec le point d’inflexion du mouvement descendant convexe qui suit immédiatement l’accent prénucléaire. De plus, plusieurs facteurs phonétiques peuvent rendre ce ton moins " visible " dans le signal acoustique. Par exemple, l’alignement retardé du pic prénucléaire peut avoir provoqué un " aplatissement " du mouvement descendant interaccentuel chez un des deux locuteurs. De façon similaire, dans l’Expérience II, nous nous attendons à ce que les tons de frontière d’*AP* ne soient pas toujours visibles comme des points d’inflexion de la courbe de f_0 . Cependant, nous supposons que, même lorsque la cible ne sera pas bien visible, sa présence sera toujours signalée par une variabilité systématique de la pente de la région entre l’accent prénucléaire et nucléaire.

4.3.1 Méthodes

Corpus et enregistrement

Le corpus utilisé était composé de trois phrases dans lesquelles la position du mot-cible proparoxyton *benevolo* a été modifiée à l’intérieur du Sujet ramifié (Table 4.1). Le Sujet était en fait constitué d’un syntagme nominal complexe. Dans les deux premières phrases, *Benevolo*, considéré comme un nom de famille, est en position finale du syntagme nominal complexe et il constitue donc la tête du syntagme prépositionnel attaché à *l’agronomo* (“l’agronome”). Dans la troisième phrase, il est en position non-finale, car, utilisé dans son statut d’adjectif (“bienveillant”), il est attaché à gauche de la tête nominale *manovale* (“ouvrier”).

5. A notre connaissance, les seules travaux existants se sont intéressés aux effets des frontières prosodiques sur la durée de la voyelle finale dans les mots oxytons, car ce contexte déclenche souvent l’application d’une règle de sandhi typique de l’italien, le *Raddoppiamento Sintattico* (Nespor & Vogel, 1986; D’Imperio & Fivela, 2003; Petrone, 2005).

Dans la condition **f-IP**, le premier syntagme nominal est suivi par une phrase relative, et nous nous attendons donc à ce que la fin du mot cible coïncide avec une frontière de syntagme intonatif. Dans la condition **f-ip**, une frontière de syntagme intermédiaire est attendue à la droite du syntagme nominal complexe (et donc, du mot-cible), car nous savons qu'en italien une frontière de syntagme intermédiaire tend à être insérée à la fin du Sujet ramifié (voir aussi le par. 1.3.2). Pour la même raison, une frontière de syntagme intermédiaire sera aussi réalisée à la fin du syntagme nominal complexe dans la condition **f-AP**. Par conséquent, la fin de *benevolo* dans cette dernière condition coïncidera avec la fin du syntagme accentuel.

Dans les trois conditions (illustrées dans la table 4.1, le Sujet était susceptible d'avoir deux accents, un prénucléaire sur la tête du syntagme nominal (*agronomo* ou *manovale*) et un nucléaire sur le mot-cible. Les deux syllabes accentuées étaient toujours séparées par quatre syllabes atones pour éviter des effets possibles de "clash" accentuel (Silverman & Pierrehumbert, 1990). L'accent nucléaire sera suivi par un accent de phrase et un ton de frontière (dans la condition **f-IP**) ou uniquement par un accent de phrase (dans les conditions **f-ip** et **f-AP**) correspondant à la fin du syntagme nominal complexe. Si un troisième niveau de constituance existe en italien, un ton périphérique sera réalisé à l'intérieur du Sujet ramifié, entre l'accent prénucléaire et l'accent nucléaire, ce ton étant aligné avec la fin du premier mot prosodique. Il faut noter que la comparaison entre les conditions **f-IP** et **f-ip** d'un côté et de la condition **f-AP** de l'autre côté est cruciale pour évaluer si l'insertion de ce ton dépend des contraintes syntaxiques ou prosodiques. En fait, si le ton marque la fin de ϕ , il sera réalisé uniquement après *l'agronomo*, dans **f-IP** et **f-ip**. Au contraire, s'il marque l'*AP*, il sera réalisé dans les trois conditions, sa présence étant indépendant du type de ramification de la tête nominale. En particulier, un ton d'*AP* sera aussi réalisé à la fin de *benevolo* dans la condition **f-AP**. De plus, si ce constituant est inférieur au syntagme intermédiaire, nous nous attendons à ce que le degré d'allongement final soit inférieur dans la condition **f-AP** par rapport à la condition **f-ip**.

f-IP	[[L'agronomo] _N [di Benevolo] _{SP}] _{SN/IP} [mangiando un panino] _{P'/IP} [si strozzò] _{SV/IP} litt. "L'ouvrier de Benevolo, en mangeant un panini, s'étrangla"
f-IP	[[L'agronomo] _N [di Benevolo] _{SP}] _{SN/ip} [mangiava un panino] _{SV/IP} litt. "L'ouvrier de Benevolo mangeait un panini"
f-AP	[[Il Benevolo] _{Adj/AP} [manovale] _N] _{SN} [mangiava un panino] _{SV/IP} litt. "Le bienveillant ouvrier mangeait un panini"

TABLE 4.1 – Phrases du corpus. Les crochets indiquent les frontières des constituants syntaxiques et prosodiques. *N* = Nom; *V* = Verbe; *Adj* = Adjectif; *SN* = syntagme nominal; *SV* = syntagme verbal; *P* = Phrase; *IP* = *Intonational Phrase* (Syntagme intonatif); *ip* = *Intermediate Phrase* (Syntagme Intermédiaire); *AP* = *Accentual Phrase* (Syntagme Accentuel).

Le corpus a été enregistré avec cinq locuteurs napolitains. Trois locuteurs (FC, LL et DD) étaient de sexe masculin, et deux (AS et MP) de sexe féminin. DD et AS avaient déjà participé à l'expérience décrite dans le chapitre 2. Les phrases devaient être lues comme affirmations à focalisation large ou questions oui/non (**A/Q**). En napolitain, les affirmations à focalisation large sont caractérisées par un accent nucléaire H+L* (D'Imperio, 2002a). D'Imperio (2002a) relève que le pic de cet accent est acoustiquement peu saillant, de sorte qu'il est souvent très difficile à localiser dans le spectre acoustique. Au contraire, la fin de la descente est bien visible à l'intérieur de la syllabe accentuée.

L'utilisation d'affirmations à focalisation large au lieu de celles à focalisation étroite n'est pas problématique, car nous avons supposé que si un ton d'*AP* est spécifié différemment dans les affirmations et dans les questions, sa présence sera indépendante du type de focalisation des constituants (large vs. étroite).

L'enregistrement a été réalisé au domicile des locuteurs, à l'aide d'un enregistreur MD et d'un microphone sur pied placé à une distance constante du sujet. Chaque phrase, présentée en ordre aléatoire sur papier A4 et parmi neuf phrases de distraction, a été répétée sept fois. Au total, 210 énoncés ont été enregistrés (3 phrases X 2 modalités intonatives X 7 répétitions X 5 locuteurs).

Analyse auditive et acoustique

Nous avons exploré l'existence d'indices différents de constituance prosodique de nature auditive et acoustique. Au niveau auditif, un transcripteur (l'auteur) a annoté les niveaux de constituance en termes de *phrasing* perçu. Cela a été fait de façon similaire à la transcription des *Break Indices* dans l'annotation ToBI (Beckman & Elam, 1994)⁶. En particulier, nous avons examiné la présence d'une frontière de *phrasing* à la fin et à l'intérieur du Sujet pour l'ensemble du corpus, car nous avons considéré l'hypothèse que si trois niveaux de constituance existent en italien napolitain, leur frontière perceptive sera localisée à la fin du syntagme nominal complexe pour le syntagme intonatif et intermédiaire ; entre l'adjectif et sa tête (condition **f-AP**) ou entre le nom et le syntagme prépositionnel (**f-IP** et **f-ip**) pour le syntagme accentuel. Une autre possibilité est que si le troisième niveau de constituance est ϕ et non l'*AP*, le degré de rupture perceptive entre le premier et le deuxième mot prosodique sera plus élevé dans les conditions **f-IP** et **f-ip** que dans **f-AP**. Il faut noter cependant que dans cette étape de l'analyse, nous n'avons fait aucune supposition quant à la nature de cette frontière, c'est-à-dire si elle signale la présence d'un syntagme intonatif, intermédiaire, phonologique ou accentuel. L'analyse acoustique a été cruciale pour l'étiquetage des niveaux de *phrasing*. L'analyse acoustique a été effectuée sur le plan tonal ainsi que sur le plan segmental de la façon suivante.

D'abord, le patron tonal de chaque phrase prononcée par les cinq locuteurs a été examiné. Nous avons donné une caractérisation phonologique (1) de la spécification tonale du ton à la périphérie du Sujet ramifié et (2) des accents prénucléaires et nucléaires réalisés à l'intérieur du Sujet même. Cette analyse a deux buts. En premier lieu, nous voulons vérifier l'existence d'une frontière de syntagme intonatif ou intermédiaire à la fin du Sujet. En particulier, nous nous attendons à ce que la fin du syntagme intermédiaire soit marquée par un ton de frontière haut, réalisé comme un contour de continuation ou encore comme un plateau haut de f_0 (D'Imperio *et al.*, 2007b). De plus, nous voulons évaluer l'impact du choix phonologique des accents mélodiques sur la réalisation de la région interaccentuelle

6. Nous sommes conscients du fait que l'analyse auditive d'un seul transcripteur n'est pas souhaitable pour parvenir à une transcription auditive fiable. En fait, une bonne transcription auditive demanderait au moins deux analystes dont évaluer les cas consensuels de *phrasing* perçu (Pitrelli *et al.*, 1994). Néanmoins, nous avons décidé d'inclure cette analyse, tout à fait préliminaire, en accord avec d'autres travaux sur la constituance prosodique en italien (D'Imperio & Fivela, 2003; Petrone, 2005)

de f_0 , c'est-à-dire dans la région où le ton de frontière d'*AP* est attendu.

La région interaccentuelle à l'intérieur du Sujet a été l'objet d'une analyse acoustique très détaillée. En particulier, nous avons d'abord extrait les vingt premiers points situés dans la région de f_0 entre le pic maximal de f_0 du premier mot et le début de la syllabe accentuée du deuxième mot. Puis nous avons fait une modélisation statistique de cette région. La modélisation a été effectuée pour les trois conditions **f-IP**, **f-ip** et **f-AP** car nous avons émis l'hypothèse que si un ton d'*AP* était présent dans la région interaccentuelle, il doit se réaliser indépendamment du type de ramification du Sujet. En particulier, nous nous attendons à ce que la valeur de f_0 du point d'inflexion de la courbe modélisée soit plus basse dans les affirmations (où ce point coïncide avec un ton L_{AP}) que dans les questions (où il coïncide avec H_{AP}). Par conséquent, nous supposons aussi que la pente allant de la courbe de l'accent prénucléaire jusqu'au point d'inflexion sera systématiquement plus raide dans les affirmations que dans les questions, car le mouvement de f_0 doit atteindre une valeur plus basse pour cette modalité intonative.

Sur le plan segmental, nous avons vérifié si la présence d'une frontière de constituant prosodique est signalée par l'allongement final. Compte tenu de l'absence d'études antérieures en italien, nous avons décidé de mesurer la durée des chaque syllabes à l'intérieur du mot clé *benevolo*. En particulier, nous avons mesuré : (1) la durée de la syllabe finale *-lo* ; (2) la durée de la syllabe tonique *-ne-* et du pied accentuel *-nevo-*. Bien que la durée de la syllabe atone prétonique puisse aussi présenter des effets d'allongement (Turk & Hufnagel, 2007), nous ne pouvons pas tester leur présence dans notre corpus. En fait, la syllabe initiale *be-* était en position intervocalique, un contexte phonotactique qui déclenche dans plusieurs variétés de l'italien (y compris le napolitain) le redoublement de la durée de la consonne (Albano Leoni, 1998). Indépendamment de la distribution de l'allongement, nous nous attendons à ce que l'allongement soit progressivement plus important dans la condition **f-AP**, **f-ip** et **f-IP**.

4.3.2 Résultats

Transcription auditive

La table 4.2 montre la proportion de ruptures prosodiques perçues à l'intérieur et à la fin du Sujet (S). Les résultats sont illustrés pour les trois conditions **f-IP**, **f-ip** et **f-AP**, dans les questions (en haut) et dans les affirmations (en bas). Les données sont regroupées pour les cinq locuteurs. Comme on peut le remarquer, une rupture prosodique a été perçue entre le syntagme nominal en position de S et le syntagme verbal suivant dans les trois conditions **f-IP**, **f-ip** et **f-AP**. Le pourcentage de ruptures prosodiques à la fin de S est très élevé, surtout dans **f-IP**. En fait, dans cette condition, la présence de la proposition relative a provoqué systématiquement l'insertion d'une frontière de constituant après S. Au contraire, aucune rupture n'a été perçue à l'intérieur de S, dont la force pouvait être similaire à celle trouvée à la fin du même constituant. Ce résultat est similaire dans les trois conditions (**f-IP**, **f-ip** et **f-AP**), suggérant donc que *lephrasing* est indépendant de la structure syntaxique du Sujet. Il faut aussi remarquer que les résultats de la transcription auditive sont similaires pour les questions et les affirmations, ce qui suggère qu'il n'y a pas de différence de *phrasing* selon la modalité intonative.

Questions		
Cond.	Rupture perceptive	
	Dans S	A la fin de S
f-IP	0	100
f-ip	0	97
f-AP	0	100

Affirmations		
Cond.	Rupture perceptive	
	Dans S	A la fin de S
f-IP	0	100
f-ip	0	97
f-AP	0	80

TABLE 4.2 – Proportion de ruptures prosodiques dans les questions (en haut) et dans les affirmations (en bas), séparément pour les trois conditions **f-IP**, **f-ip** et **f-AP**.

Choix phonologiques

Dans cette section nous allons décrire le patron intonatif utilisé dans la lecture des phrases par les cinq locuteurs. En particulier, nous allons détailler la typologie des tons périphériques trouvés à la fin du syntagme nominal complexe en position de Sujet (S), ainsi que la typologie d’accents mélodiques trouvés à l’intérieur de S. Vu le nombre élevé de ruptures prosodiques sur l’ensemble des phrases (par. 4.3.2), nous avons décidé d’analyser toutes les productions qui avaient été l’objet de la transcription auditive. Nous nous attendons à ce que ces phrases contiennent une frontière de syntagme intonatif ou intermédiaire à la fin de S, mais pas entre les deux mots prosodiques de S. En fait, aucune rupture perceptive n’a été perçue à l’intérieur de S, dont la force pourrait être comparable à celle du syntagme intermédiaire. L’analyse des indices de *phrasing* a été complétée par l’inspection de la pause (cette section), de la pente interaccentuelle (section 4.3.2.3) et de l’allongement final (section 4.3.2.4).

La typologie des tons périphériques et des accents mélodiques est illustrée dans la table 4.3. Comme on peut le remarquer, le ton périphérique L est le ton le plus fréquemment réalisé à la fin de S (surtout dans les questions). Cette analyse est très différente de celle effectuée par D’Imperio *et al.* (2007b) pour l’italien. Dans ce travail, les auteurs avaient remarqué que dans les affirmations, un ton périphérique H était inséré à la fin du constituant ramifié en position de Sujet, lorsqu’il était prononcé comme un seul constituant prosodique. En particulier, ils avaient trouvé que dans les affirmations (la seule modalité

intonative examinée dans leur étude), la dernière syllabe tonique du mot cible est associée à un accent mélodique montant suivi par un ton de frontière haut. Au niveau phonétique, le ton de frontière est signalé par le déplacement du pic accentuel dans la syllabe finale du mot (contour de continuation) ou encore par un plateau haut de f_0 , visible sur la syllabe finale du mot (*sustained pitch*). Il faut noter que les mots utilisés dans le travail de D’Imperio *et al.* (2007b) étaient uniquement paroxytons (ex. *BadaLona*), c’est-à-dire qu’une seule syllabe atone était présente entre la syllabe tonique et la fin du constituant.

Questions							
Cond.	Acc. prén.	Acc. nucl.			Tons per.		Pause
	L+H*	L+H*	!H*	H+L*	H	L	
f-IP	100	11	3	86	0	100	0
f-ip	100	63	11	26	3	97	0
f-AP	100	60	3	37	9	91	0

Affirmations							
Cond.	Acc. prén.	Acc. nucl.			Tons per.		Pause
	L+H*	L+H*	!H*	H+L*	H	L	
f-IP	100	26	0	74	17	83	17
f-ip	100	79	0	21	24	76	9
f-AP	100	77	0	23	26	74	9

TABLE 4.3 – Distribution de différents types d’accents prénucléaires, nucléaires, de tons périphériques à la fin de S et de pauses. Les résultats sont montrés en termes de pourcentage relativement à l’ensemble des phrases du corpus.

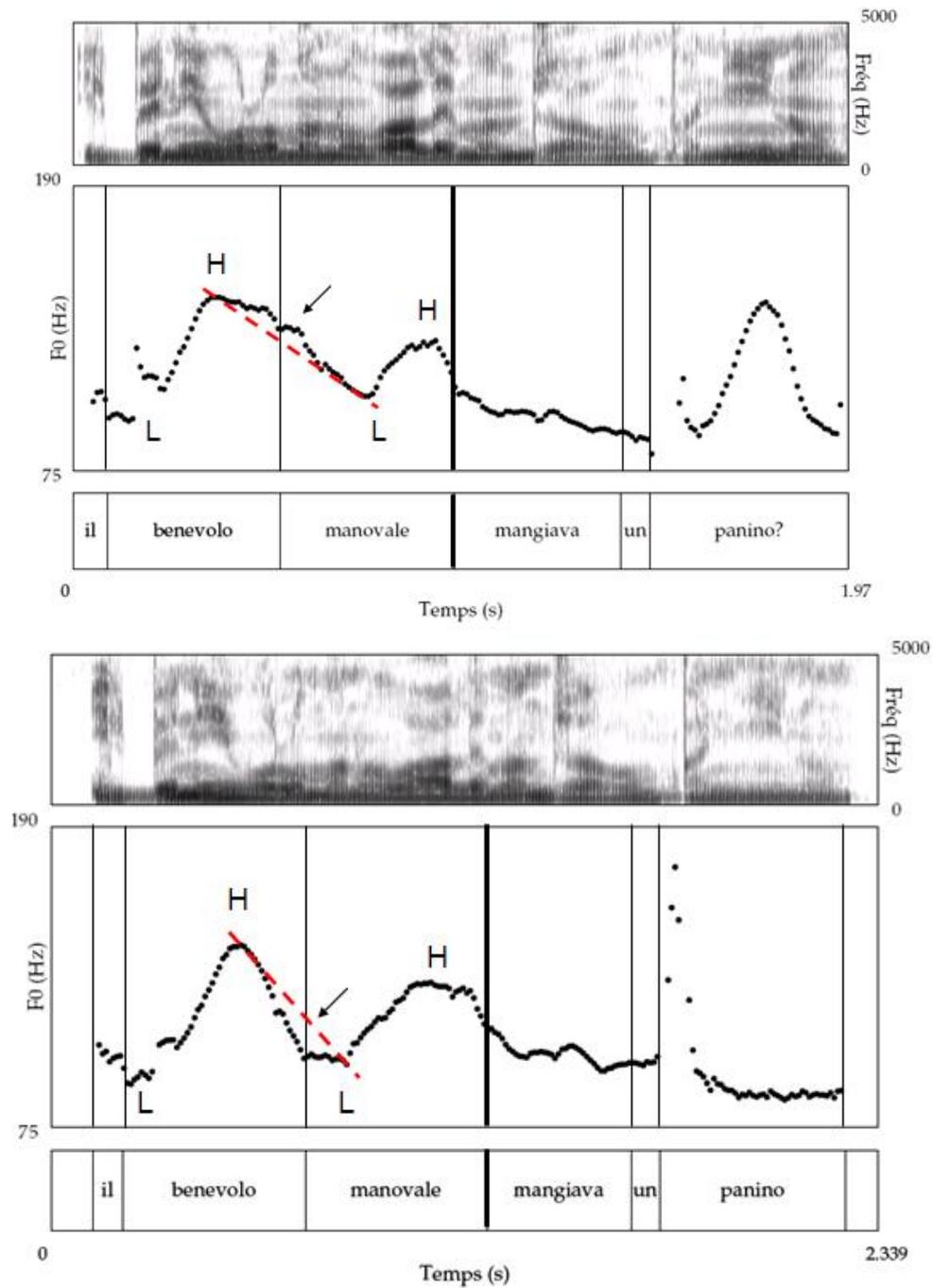


FIGURE 4.19 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *Il benevolo manovale mangiava un panino* prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, et la ligne épaisse la fin du constituant en position de Sujet. La ligne rouge tiretée schématise l'interpolation linéaire prédite par Pierrehumbert (1980) entre les deux accents. Locuteur FC.

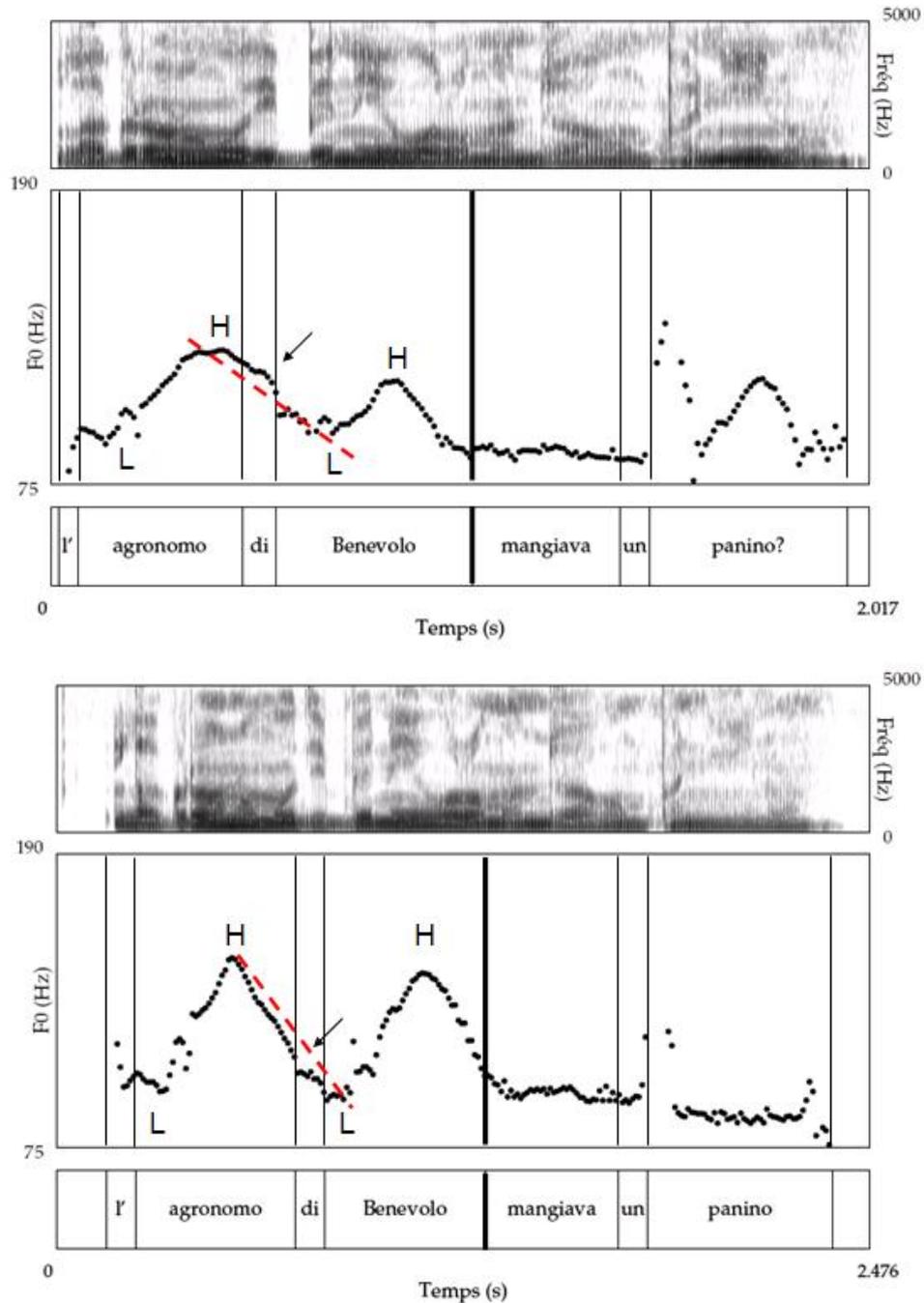


FIGURE 4.20 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *L'agronomo di Benevolo mangiava un panino* prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, la ligne épaisse la fin du constituant en position de Sujet. La ligne rouge tiretée schématise l'interpolation linéaire prédite par Pierrehumbert (1980) entre les deux accents. Locuteur FC.

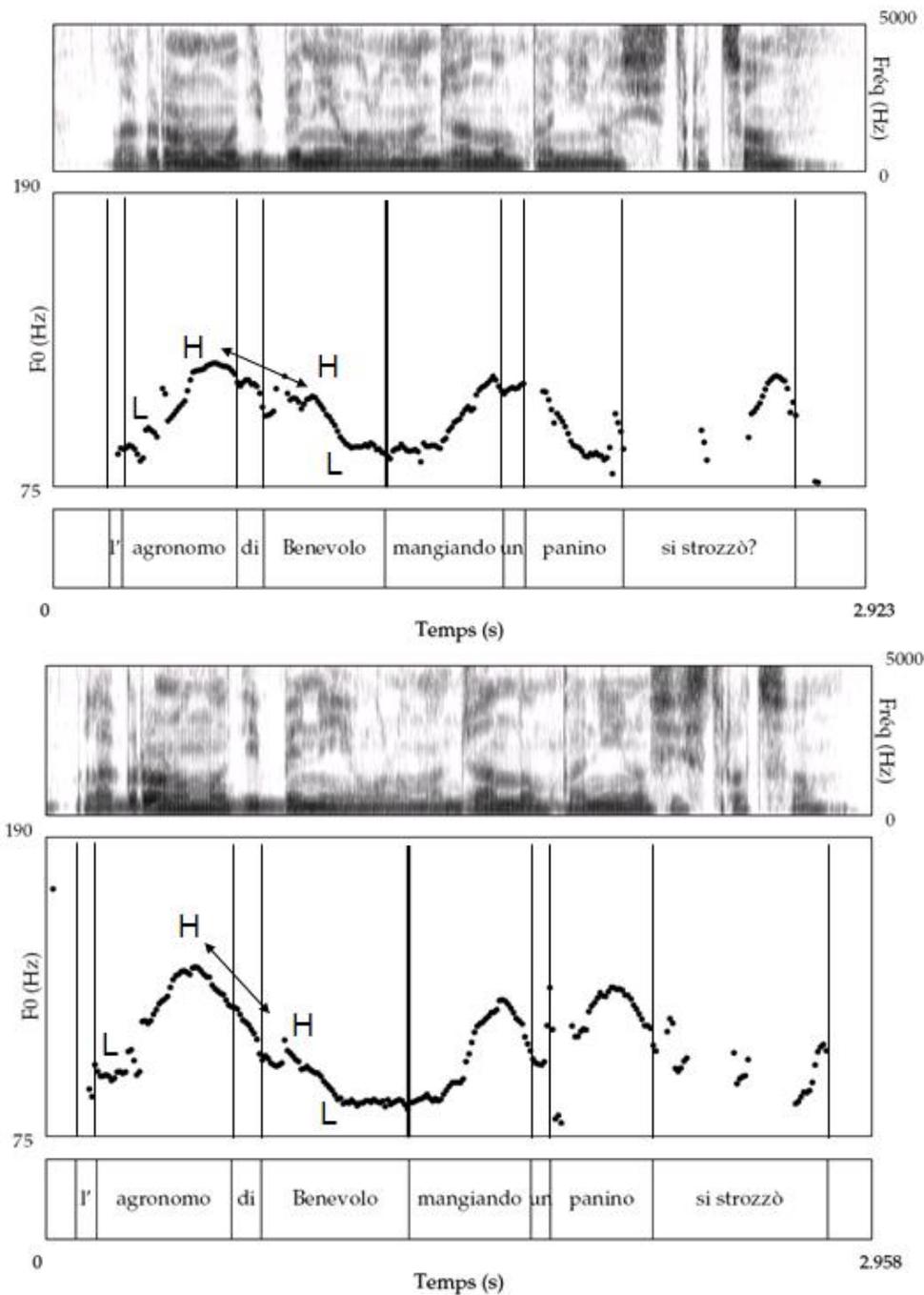


FIGURE 4.21 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *L'agronomo di Benevolo, mangiando un panino, si strozzò* prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, la ligne épaisse la fin du constituant en position de Sujet. La flèche indique la pente interaccentuelle. Locuteur FC.

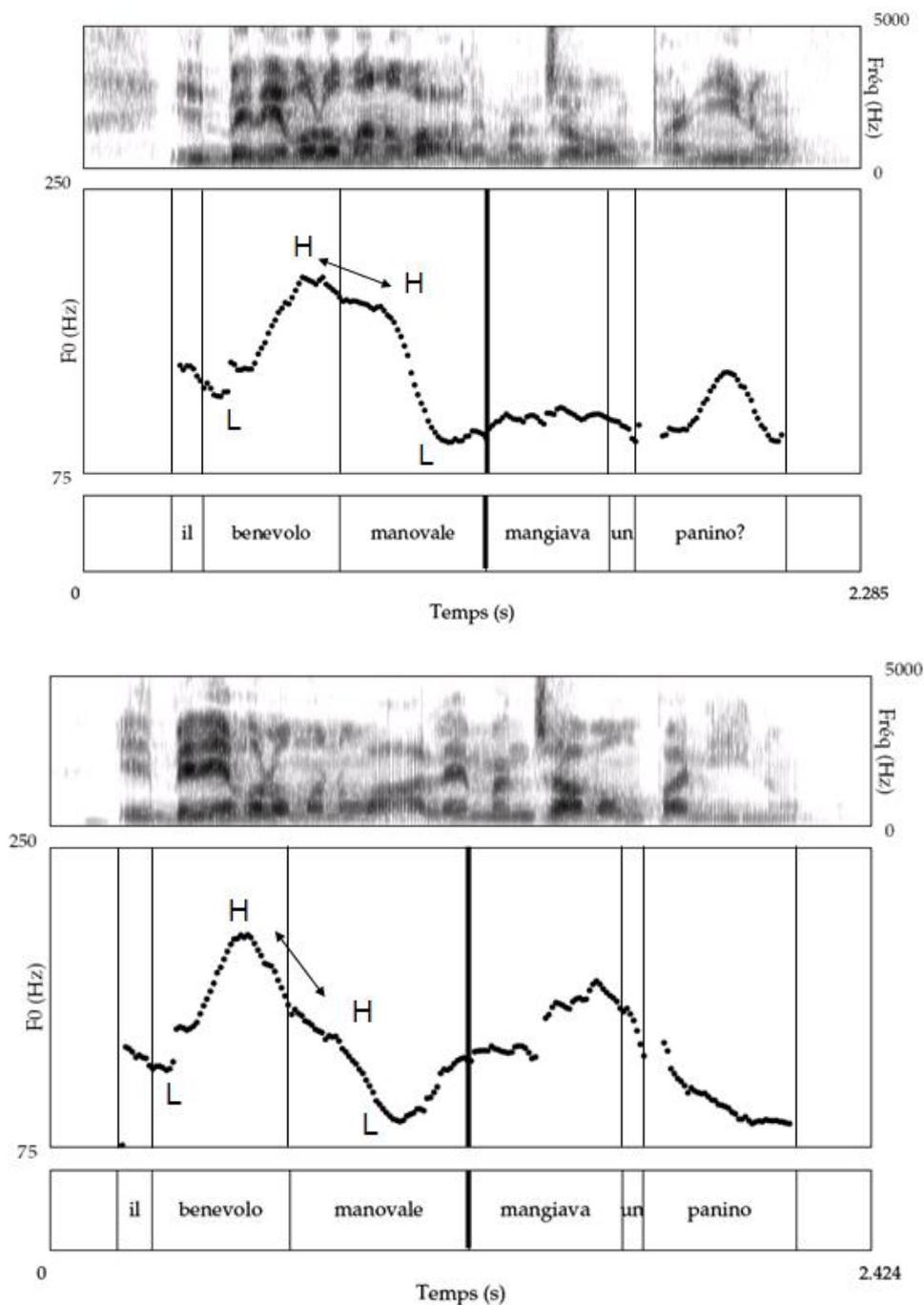


FIGURE 4.22 – Spectrogramme, courbe de f_0 et cibles tonales pour la phrase *L'agronomo di Benevolo mangiava un panino* prononcée comme une question (en haut) et une affirmation (en bas). Les lignes continues indiquent les frontières des mots, la ligne épaisse la fin du constituant en position de Sujet. La flèche indique la pente interaccentuelle. Locuteur DD.

Nos données suggèrent cependant que dans plusieurs cas, le pic final de f_0 dans la

syllabe post-tonique signale uniquement la cible de l'accent mélodique, suivi par une descente de f_0 jusqu'à la cible périphérique L, à la fin du constituant. Dans les cas des mots paroxytons, la descente finale de f_0 jusqu'à la cible pourrait être coupée, de sorte que la cible L est *undershot* ou réalisée dans la région immédiatement suivant la fin du constituant. Prenons les exemples illustrés dans la fig. 4.19. Cette figure illustre la prononciation de la phrase *Il benevolo manovale mangiava un panino* comme une question (en haut) et une affirmation (en bas), effectuée par le locuteur FC. Dans cette phrase, le mot à la fin du constituant est paroxyton, et la syllabe tonique *-va-* est caractérisée par une montée de f_0 , dont le début est à l'attaque consonantique, et la fin dans la syllabe postaccentuelle. A première vue, cette configuration est très similaire au contour de continuation illustré par D'Imperio *et al.* (2007b).

La phrase *L'agronomo di Benevolo mangiava un panino* (fig. 4.20) nous offre une interprétation différente. Cette phrase a été prononcée comme une question et une affirmation par le même locuteur. Le *phrasing* est le même que dans la phrase précédente, et la seule différence est que le mot à la fin du sujet est, cette fois-ci, proparoxyton. La figure nous montre cependant que la montée de f_0 sur la syllabe tonique finale est suivie par une descente de f_0 , qui se termine par une cible basse à la fin du constituant. Ce patron est indépendant de la modalité intonative, et la table 4.3 nous indique qu'on le retrouve même dans la condition **f-IP**.

Comment peut-on interpréter le déplacement temporel du pic accentuel? Dans les conditions **f-AP** et **f-ip**, la syllabe tonique du deuxième mot est caractérisée par une montée de f_0 au début de la syllabe, qui se termine dans la syllabe post-tonique. Cet alignement retardé, typique de trois de nos locuteurs (FC, MD, LL), est tout à fait similaire à celui de l'accent prénucléaire (LH*), réalisé sur la syllabe accentuée du premier syntagme nominal à l'intérieur de S. Par conséquent, nous supposons que, pour ces trois locuteurs, les conditions **f-AP** et **f-ip** sont généralement caractérisées par deux accents (LH*), dont le deuxième est suivi par un ton de frontière (voir table 4.3).

Les figures 4.19 et 4.20 nous montrent aussi que la pente interaccentuelle est très différente selon la modalité intonative. Dans les questions, la f_0 descend doucement du pic H dans la première syllabe tonique jusqu'à la fin du premier mot prosodique. Après

cette région, la pente de f_0 devient plus raide jusqu'à la cible basse du deuxième accent L+H*. Dans les affirmations, la descente de f_0 après le premier pic est beaucoup plus rapide, de sorte qu'un minimum de f_0 est visible à la fin du mot. Ce patron est similaire à celui trouvé dans l'Expérience I, suggérant la présence d'un ton de frontière différemment spécifié selon la modalité. Il faut noter aussi que cet effet de modalité a été observé dans les deux conditions. Cela nous amène à exclure que le ton de frontière signale la fin du ϕ , car il est indépendant de la position syntaxique du complément relativement à sa tête.

En ce qui concerne la condition **f-IP**, le premier accent (LH*) est suivi chez les trois locuteurs par un accent descendant, dont le pic, réalisé avant le début de la syllabe accentuée, était souvent difficilement localisable dans le signal acoustique (fig. 4.21). Le pic était suivi par une descente rapide de f_0 jusqu'à la syllabe accentuée, où un coude bas de f_0 était réalisé. Cette configuration est typique des affirmations et des questions de tous les locuteurs. Une hypothèse possible, est que la descente de f_0 entre les deux pics est due au fait que la f_0 doit attendre la valeur de la cible de l'accent nucléaire H+L*, cible *downstepped*. Cette explication repose sur une analogie faite avec les résultats de D'Imperio (2002a), selon lesquelles l'accent nucléaire H+L* des affirmations à focalisation large est caractérisé par un pic peu proéminent et intrinsèquement abaissé. Par conséquent, cet accent a été étiqueté comme H+L*.

Il faut noter aussi que, bien qu'aucune cible intermédiaire n'ait été trouvée entre (LH*) et H+L*, une différence subtile de la pente interaccentuelle est encore visible : dans les questions, la descente de la pente est plus douce que dans les affirmations. Cela pourrait suggérer que les deux cibles hautes délimitant la région interaccentuelle pourraient avoir provoqué un *undershoot* du ton de frontière. Nous rappelons que l'alignement du pic du premier accent est toujours retardé, tandis que le ton *trailing* H du deuxième accent est réalisé avant la syllabe tonique. A cause de cela, il est possible que la f_0 n'arrive pas à atteindre sa cible après le premier pic ni dans les affirmations (où le L_{AP} provoquerait un coude de f_0 autour de fin du premier mot), ni dans les questions (où le mouvement de f_0 n'est pas caractérisé par une forme convexe, dont la présence signale l'insertion du H_{AP}). Si cela est vrai, nous nous attendons à ce que la différence de pente entre les deux accents soit systématique.

Pour le quatrième locuteur (DD), le deuxième accent est toujours descendant, indépendamment de la modalité intonative et de la condition de *phrasing*. La réalisation acoustique de cet accent est très similaire à celui trouvé dans la condition **f-IP** pour FC, MD et LL, et a donc été étiqueté comme H+L* (voir fig.4.22). De plus, chez DD, le degré de rupture perceptive dans les conditions **f-AP** et **f-ip** était très élevé, et il ressemblait à celui trouvé dans la condition **f-IP**. Par ailleurs, il est intéressant de noter que chez DD la fin de S était parfois marquée par une pause silencieuse, et cela a été trouvé dans les trois conditions de *phrasing*. On pourrait donc supposer que dans les trois conditions le locuteur a réalisé une frontière de syntagme intonatif. En fait, la fig. 4.22 nous montre aussi que la différence de pente interaccentuelle dans les trois conditions reflète les observations faites pour les trois locuteurs FC, MD et LL pour la condition **f-IP**. Cela suggère que, malgré la différence dans le choix phonologique des accents mélodiques, la spécification tonale du ton de syntagme accentuel est toujours présente.

Le cinquième locuteur (AS) présente un patron différent selon la modalité intonative et le type de *phrasing*. Les affirmations dans les conditions **f-AP** et **f-ip** sont caractérisées par deux accents montants, dont le début est réalisé à l'attaque de la syllabe accentuée. Les questions présentent une montée de f_0 dans la première syllabe tonique suivie par une descente plus douce jusqu'à la deuxième syllabe tonique. Le début de la descente était signalé par un petit haussement de la f_0 en correspondance avec la syllabe prétonique, suggérant la présence d'un accent H+L*. Les données de AS relatives aux différences de pente interaccentuelle ne seront pas analysées ici car nous ne pouvons pas évaluer si et dans quelle mesure ces éventuelles différences dépendent du choix de l'accent mélodique ou si elles sont "authentiques".

Pour résumer, la fin du Sujet semble être caractérisée, dans nos données, par l'insertion d'un ton bas (et, moins souvent, d'un ton haut) dans les questions et dans les affirmations, indépendamment du type de *phrasing*. Dans les trois conditions de *phrasing*, l'accent mélodique sur la première syllabe tonique est un accent prénucléaire (LH*). Beaucoup plus de variabilité a été trouvée pour le choix phonologique du deuxième accent. Chez trois locuteurs, les conditions **f-AP** et **f-ip** sont caractérisées par l'insertion d'un accent montant sur la syllabe tonique finale, ce qui suggère la présence d'un (LH*).

Ce patron accentuel est similaire à celui observé par D’Imperio *et al.* (2007b), suggérant l’insertion d’une frontière de constituant phonologique à la fin du Sujet ramifié de rang inférieur au syntagme intonatif. Cependant, lorsqu’une frontière de syntagme intonatif est insérée (provoquée par l’enchâssement de la proposition relative) dans la condition **f-IP** le deuxième accent est plutôt descendant. Indépendamment du choix accentuel, la région interaccentuelle présente des différences de la forme et de la pente de f_0 selon la modalité intonative, suggérant la présence d’un ton de syntagme accentuel dans les questions (H_{AP}) et dans les affirmations (L_{AP}).

Pente interaccentuelle

Compte-tenu de la variabilité d’accentuation décrite dans le paragraphe précédent, l’analyse des valeurs de pente a été effectuée séparément pour chaque sujet. Pour les quatre locuteurs FC, LL, MD et DD une analyse globale a été effectuée pour les conditions **f-AP** et **f-ip**, c’est-à-dire dans les conditions où on s’attend à ce que la fin de S soit signalée par une frontière de syntagme intermédiaire. Chez les trois premiers locuteurs, nous avons modélisé la région de f_0 qui se trouve entre le pic du premier accent et la cible L du deuxième accent, lorsque le deuxième accent était montant (le type d’accent le plus fréquent pour ces locuteurs). La modélisation a été effectuée par des régressions linéaires par morceaux, car nous avons observé dans le signal acoustique la présence d’un ton additionnel de syntagme accentuelle entre les deux accents.

Chez DD, le deuxième accent était descendant ($H+L^*$) dans les deux conditions de *phrasing*. La pente interaccentuelle entre les deux accents ne présentait pas de cible additionnelle. En particulier, dans les questions il était impossible de localiser de façon objective la position du ton H_{AP} entre les deux pics accentuels, car la f_0 présente une configuration à plateau descendant (fig. 4.22, en haut). Dans les affirmations, le ton L_{AP} n’était pas réalisé, bien que la pente interaccentuelle semble être plus raide que dans les questions (fig. 4.22, en bas). Nous avons supposé que, à cause du voisinage de deux pics accentuels, le ton de syntagme accentuel chez ce locuteur est *undershot* dans les deux modalités intonatives. Si cela est vrai, la différence de spécification tonale H_{AP}/L_{AP} se reflétera uniquement dans l’inclinaison de la pente. Dans les questions, le ton H_{AP} pro-

voquera une inclinaison très douce de la pente entre le premier et le deuxième pic ; dans les affirmations, un ton L_{AP} sous-jacent sera responsable de l'inclinaison plus raide de la pente interaccentuelle. Vu l'absence d'une cible H_{AP} ou L_{AP} dans la trace acoustique, la modélisation de cette région a été faite à l'aide d'une régression linéaire simple⁷.

Il faut noter que dans cette expérience, nous n'avons pas utilisé de techniques de modélisation non-linéaires, telle que la régression logistique, qui avait déjà été employée dans l'Expérience I pour les questions. En fait, l'existence d'effets possibles d'environnement tonal sur la réalisation de la cible H_{AP} pourrait rendre la modélisation du mouvement descendant convexe très difficile. Par conséquent, nous avons modélisé la région interaccentuelle à l'aide de techniques de modélisation linéaire, afin de vérifier s'il existe un troisième niveau de *phrasing* qui puisse rendre compte de la différence systématique de pente entre deux accents observée à l'intérieur du même syntagme intermédiaire.

La procédure pour la modélisation de la pente interaccentuelle est très similaire à celle adoptée dans l'Expérience I. Nous avons d'abord extrait les valeurs de vingt premiers points entre le pic prénucléaire et la cible du deuxième accent. Les questions et les affirmations ont été ensuite comparées par une régression à effets mixtes ajustée à cette région, où la modalité intonative était l'effet de traitement. Puisque l'identité segmentale de S était légèrement différente entre **f-AP** et **f-ip**, le facteur " mot " a été considéré comme l'effet aléatoire. L'intercept aléatoire a été enlevé du modèle lorsqu'il était proche de zéro.

Nous allons d'abord détailler les résultats obtenus pour la régression linéaire par morceaux, utilisée pour modéliser la pente chez FC, LL et MD. De façon similaire à l'Expérience I, l'exploration des nos données a commencé par l'utilisation de la technique de lissage *lowess* ajustée à la région interaccentuelle. Nous avons ensuite effectué plusieurs régressions par morceaux pour chaque locuteur et selon chaque modalité intonative, c'est-à-dire une régression pour chaque point d'inflexion possible. Pour le locuteur FC, les meilleurs modèles - c'est-à-dire dont la déviance était la plus petite - étaient ceux dont le point d'inflexion était localisé au point 10 dans les questions, et au point 14 pour les

7. Le locuteur AS a été exclus des analyses de pente car la modélisation de la région interaccentuelle aurait pu être biaisée par le choix différent du deuxième accent mélodique, qui était montant dans les affirmations et descendant dans les questions (par. 4.3.2).

affirmations. Pour le locuteur LL, le point d'inflexion était à 10 et 11 respectivement dans les questions et dans les affirmations. Chez MD, le point d'inflexion des affirmations était à 12, mais il n'y avait pas de point d'inflexion pour les questions, les variations de déviance entre une régression et l'autre étant très similaires pour cette modalité intonative. En d'autres termes, la pente des questions semblait purement linéaire. Afin de comparer les deux modalités intonatives pour chaque locuteur, nous avons appliqué aux données une autre régression linéaire par morceaux, dans laquelle le point d'inflexion a été placé à une distance moyenne entre le point d'inflexion des questions et des affirmations pour FC et LL. Pour MD, le point d'inflexion de la pente a été fixé à 12, le *breakpoint* des affirmations.

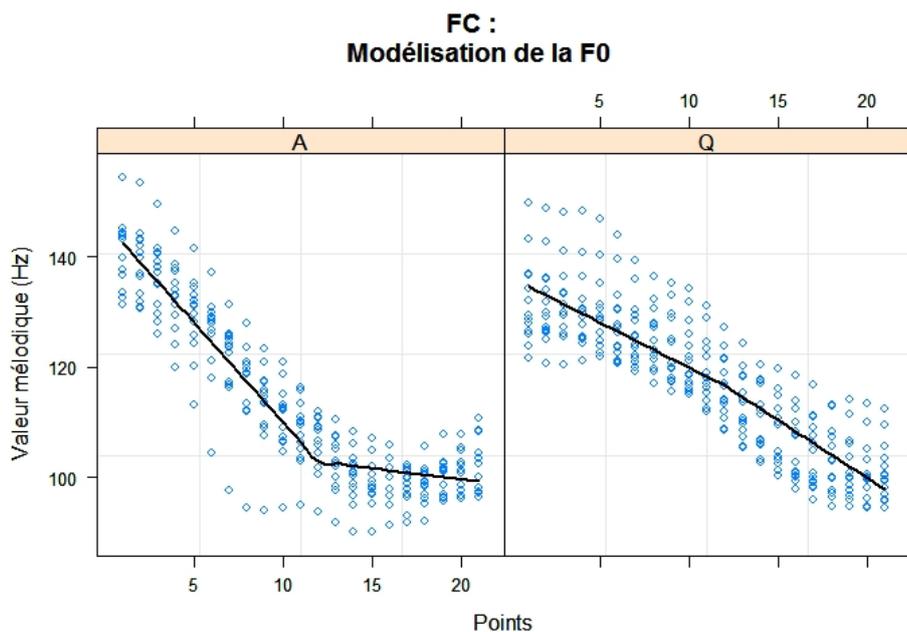


FIGURE 4.23 – Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression linéaire par morceaux dans les affirmations (**A**, à gauche) et dans les questions (**Q**, à droite). Locuteur FC.

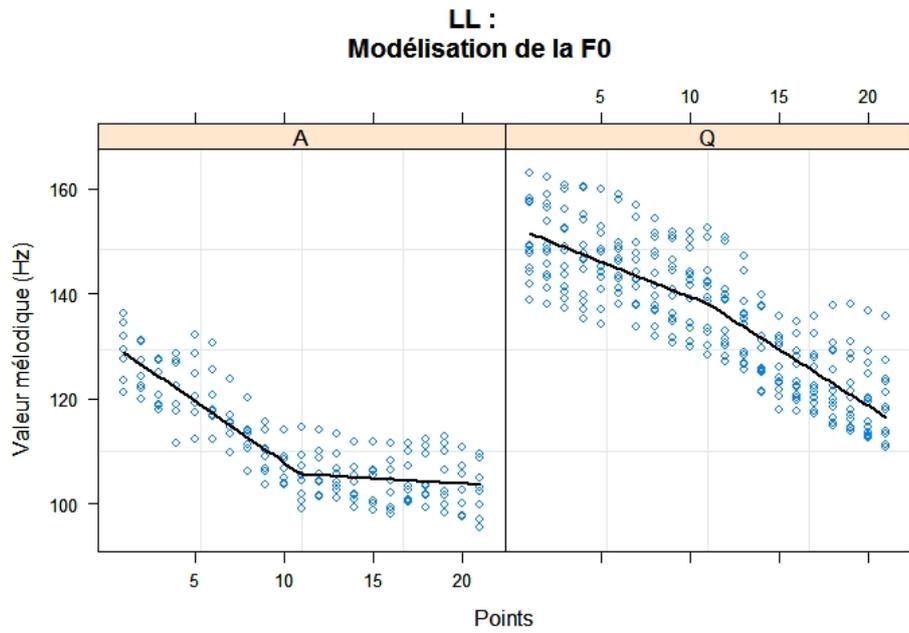


FIGURE 4.24 – Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression linéaire par morceaux dans les affirmations (**A**, à gauche) et dans les questions (**Q**, à droite). Locuteur LL.

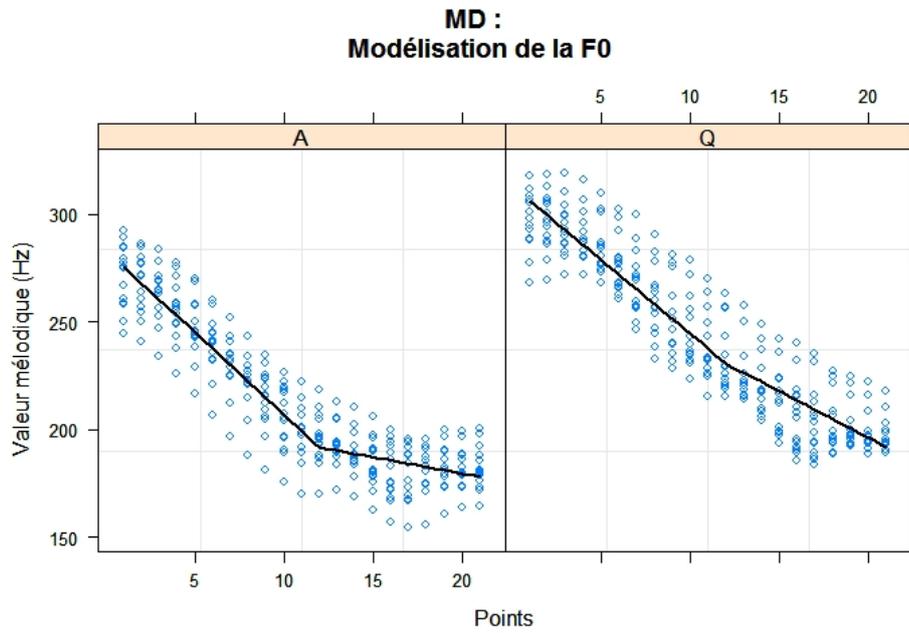


FIGURE 4.25 – Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression linéaire par morceaux dans les affirmations (**A**, à gauche) et dans les questions (**Q**, à droite). Locuteur MD.

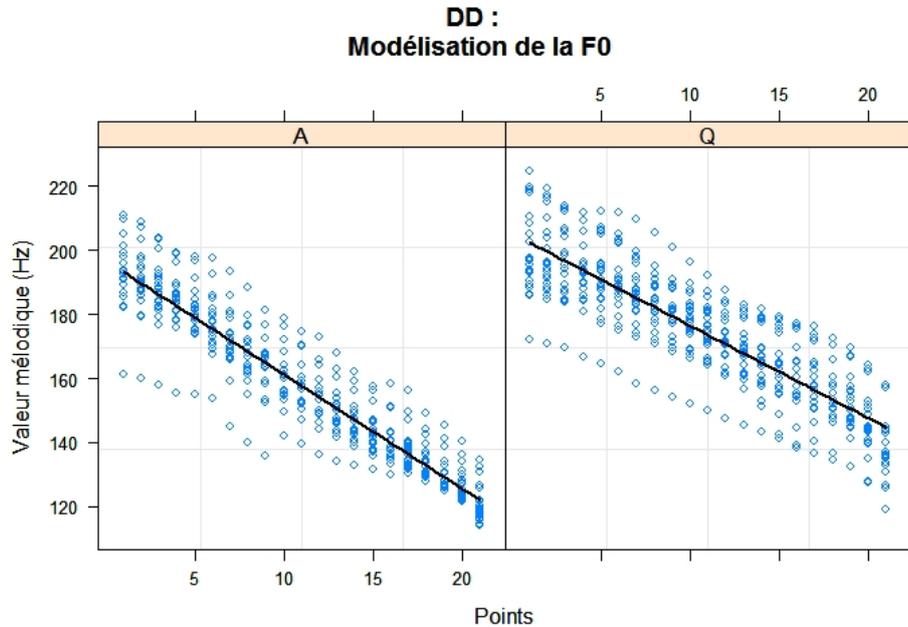


FIGURE 4.26 – Valeurs de f_0 observées (points) et attendues (ligne continue) par la régression linéaire dans les affirmations (**A**, à gauche) et dans les questions (**Q**, à droite). Locuteur DD.

Nous avons trouvé que la valeur mélodique du point d'inflexion était plus basse dans les affirmations que dans les questions pour les trois locuteurs FC (**A** : 100 Hz ; **Q** : 113 Hz $t = -13.12$, $pMCMC < .01$), MD (**A** : 192 Hz ; **Q** : 230 Hz $t = -20.07$, $pMCMC < .01$) et LL (**A** : 101 Hz ; **Q** : 137 Hz $t = -27.58$, $p < .01$). Chez FC et LL, la pente était plus raide dans les affirmations que dans les questions avant le point d'inflexion (le coefficient de pente étant plus élevé dans les affirmations de 1.7, et 1.01 respectivement pour les deux locuteurs) mais plus douce après ce point (le coefficient de pente étant plus élevé dans les questions de 1.96 et 2.01). L'effet de modalité est en fait significatif sur la pente avant le point d'inflexion (FC : $t = -11.77$, $pMCMC < .01$; LL : $t = -4.83$, $pMCMC < .01$) et après celui-ci (FC : $t = 7.76$, $pMCMC < .01$; LL : $t = 9.55$, $p < .01$). Pour le locuteur MD, la différence de pente avant le point d'inflexion était au-dessous du niveau de significativité avant le point d'inflexion ($pMCMC = .016$) mais elle était significative

après ce point ($t = 6.9$, $pMCMC < .01$). Les résultats obtenus pour MD nous rappellent donc ceux obtenus pour AS dans l'Expérience I, où la forme parabolique de la région postaccentuelle des questions était aplatie au point de donner une pente presque linéaire. Il faut noter que pour les trois locuteurs, le point d'inflexion du modèle de régression était localisé dans la région après la fin du premier mot prosodique. En particulier, dans la condition **f-AP**, le point d'inflexion était respectivement à 20 et 51 ms de la fin de *Benevolo* dans les affirmations et questions de FC, et à 27 et 58 ms de cette frontière chez LL. Pour MD, le point d'inflexion dans les affirmations était retardé de 102 ms, et il est était donc aligné avec la fin de la première syllabe du syntagme verbal *mangiava*. Finalement, nous avons ajustée le modèle de régression simple à la région interaccentuelle du locuteur DD. Puisqu'il était impossible de localiser le début de l'accent descendant à la main, une ligne droite a été ajustée à la région entre le premier pic accentuel et le début de la syllabe tonique finale, syllabe où la cible L* était réalisée. Comme on s'y attendait, le coefficient de la pente était plus élevé dans les questions (-2.8) que dans les affirmations (-3.5), cette différence étant significative ($t = 6.58$, $pMCMC < .01$). Les fig. 4.23, 4.24, 4.25 et 4.26 illustrent l'ajustement du modèle de régression chez les quatre locuteurs.

Durée

La fig. 4.27 montre les valeurs moyennes de durée (ordonnées) pour chaque syllabe du mot-cible *Benevolo* (abscisses). Les résultats sont illustrés séparément pour les affirmations (panneau à gauche) et pour les questions (à droite) et, dans chaque panneau, en fonction des trois niveaux de constituance prosodique.

Contrairement à nos attentes, ni la durée de la syllabe finale ni celle de la syllabe antépénultième sont fonction de la constituance prosodique, tandis que les effets du niveau de constituance se manifestent surtout sur la durée de la syllabe tonique : la syllabe *-ne-* tend à être progressivement plus longue lorsque que le mot-clé se trouve en position finale de syntagme accentuel (**f-AP**), de syntagme intermédiaire (**f-ip**) et de syntagme intonatif (**f-IP**). Ces effets semblent dépendre également de la modalité intonative, car la syllabe accentuée est plus longue dans les affirmations que dans les questions. En particulier, la durée de la syllabe accentuée est plus allongée en position finale de *ip* que d'*AP* de 27 ms

dans les affirmations et de 16 ms dans les questions (valeurs moyennes globales).

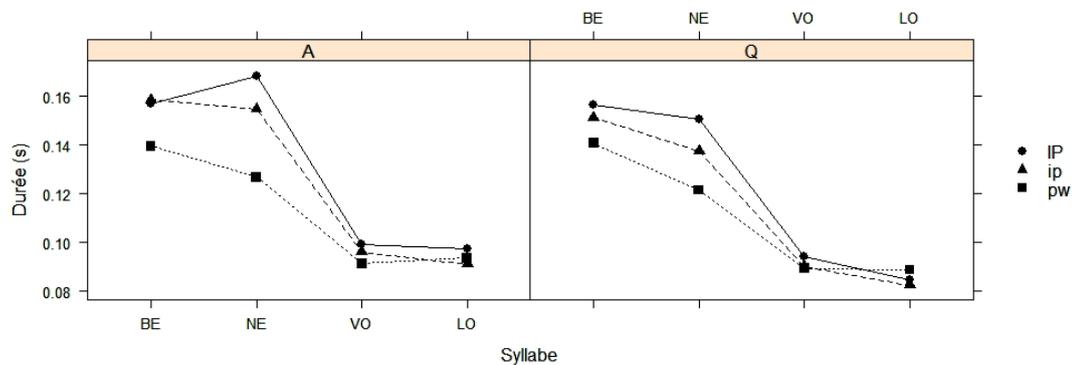


FIGURE 4.27 – Durée moyenne des syllabes de *Benevolo*. Les résultats, regroupés pour les cinq locuteurs, sont présentés en fonction de la modalité intonative et du niveau de constituance prosodique.

Durée de la syllabe tonique

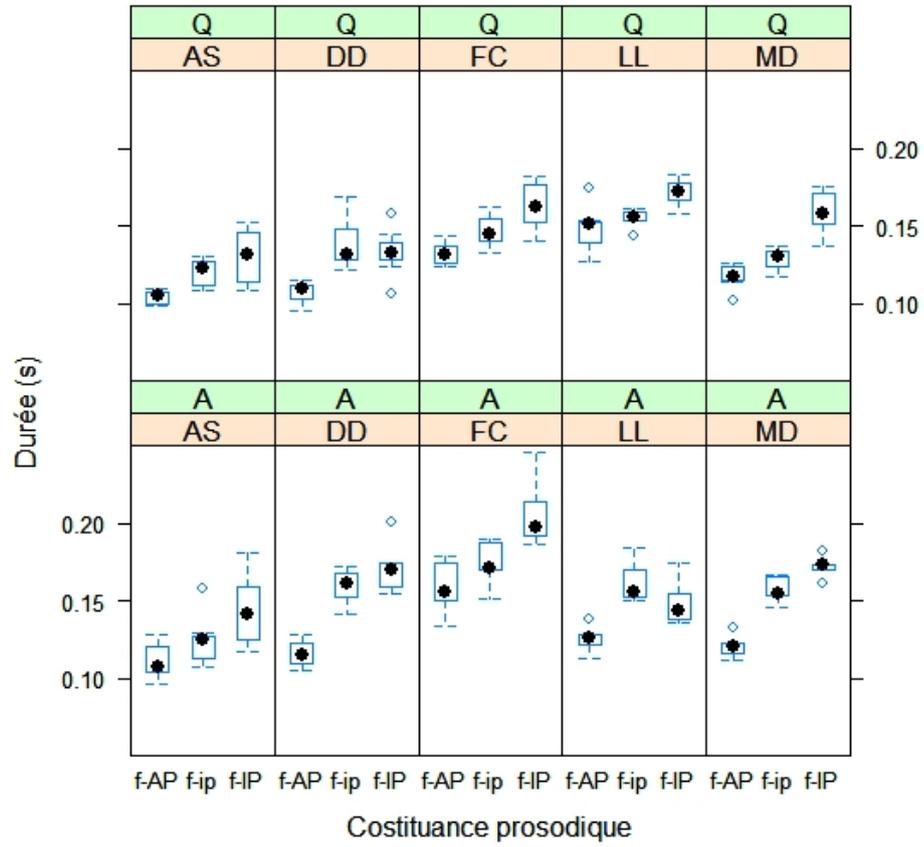


FIGURE 4.28 – Durée moyenne de la syllabe tonique. Les résultats, séparés par locuteur, sont fonction de la modalité intonative et du niveau de constituance prosodique.

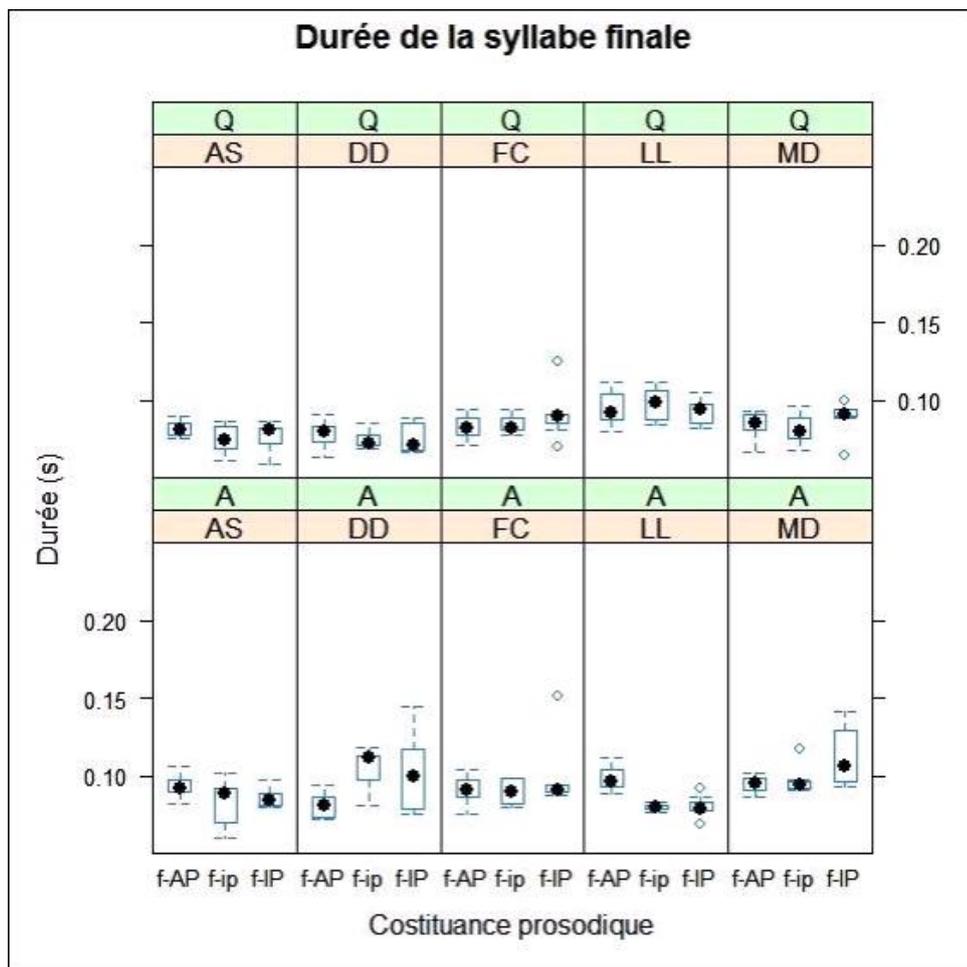


FIGURE 4.29 – Durée moyenne de la syllabe finale. Les résultats, séparés par locuteur, sont fonction de la modalité intonative et du niveau de constituance prosodique.

La Fig. 4.28 illustre les résultats de durée de la syllabe tonique *-ne-* (ordonnées) par niveaux de constituance (abscisses), pour les questions (en haut) et pour les affirmations (en bas), pour chaque locuteur. Comme on peut le remarquer, il y a une tendance à allonger progressivement la durée de la syllabe de la condition **f-AP** jusqu'à **f-IP** dans les deux modalités intonatives. Cela est vrai surtout chez AS, FC, MD et - uniquement dans les affirmations - chez LL, où la durée de la syllabe est plus petite pour **f-ip** que pour **f-IP**. Cela pourrait suggérer l'existence de trois constituants prosodiques différents. Chez DD et dans les questions de LL, au contraire, la différence entre **f-ip** et **f-IP** semble beaucoup plus faible.

Généralement, la durée de la syllabe finale n'est pas influencée par le voisinage des frontières de constituants, car elle est autour de 90 ms (dans les affirmations) et de 80 ms

(dans les questions) indépendamment du *phrasing*. Ces résultats sont en désaccord avec ceux de littérature sur l’allongement final, qui montrent que l’effet de constituance prosodique se reflète dans la durée de la syllabe finale et qu’il peut optionnellement s’étendre aux syllabes précédentes. Au contraire, le graphique 4.29 nous montre que la durée de la syllabe finale est stable quel que soit le niveau de constituance, surtout dans les questions. Les affirmations sont caractérisées par une variabilité plus importante chez tous les locuteurs. Par exemple, chez MD et DD, la fin du syntagme intonatif est signalée par un allongement important de la syllabe finale relativement à la condition **f-ip**.

Nous avons effectué une série d’analyses de régression linéaire à effets mixtes sur le jeu de données regroupées pour les cinq locuteurs, avec la position à l’intérieur du constituant prosodique (**f-IP/f-ip/f-AP**) et la modalité intonative (**A/Q**) comme facteurs de traitement, et les locuteurs comme facteur aléatoire⁸. En ligne avec nos observations, les résultats statistiques montrent que l’effet de constituance prosodique est significatif sur la durée de la syllabe accentuée (**f-ip** vs **f-AP** : $t = -7.05$, $pMCMC < .01$; **f-ip** vs **f-IP** : $t = 3.26$, $pMCMC < .01$). De plus, cette syllabe est plus longue dans les affirmations que dans les questions, en position finale de syntagme intermédiaire (**A** vs **Q** : $t = -4.45$, $pMCMC < .01$) et de IP ($t = -4.40$, $pMCMC < .01$). Au contraire, il n’y a pas de différence significative de modalité intonative dans **f-AP** ($t = -1.47$, $p > .01$). Les interactions sont également non significatives. L’analyse de régression a été effectuée aussi sur la durée de la syllabe finale et pré-finale⁹. Pour les deux mesures, ni l’effet de constituance ni l’effet de modalité ni les interactions n’étaient significatifs ($pMCMC > .01$).

La fig. 4.27 montre aussi que la durée de la syllabe initiale *Be-* est beaucoup plus petite en position finale d’*AP* que de *ip* et *IP*. Cela ne peut pas être dû à l’effet de la constituance prosodique. En fait, nous savons qu’en italien la consonne /b/ tend à être géminée en position intervocalique (Albano Leoni, 1998). C’est le cas dans les conditions **f-IP** et **f-ip** (*di Benevolo*), mais non dans **f-AP** (*il Benevolo*). Si dans notre expérience

8. Pour la comparaison par paires des trois niveaux de constituance prosodique, le niveau de référence choisi était **f-ip** (voir le par. 2.2.5 dans le chapitre 2 pour les détails sur la comparaison par paires dans les analyses de régression).

9. L’existence de différences dans la durée des syllabes parmi les sujets aurait dû aussi être explorée. Cependant, vu le nombre petit d’observations, nous avons préféré considérer les sujets comme l’effet aléatoire des régressions mixtes. L’exploration plus détaillée de nos découvertes sur l’allongement final sera l’objet de nos études dans le futur.

l’allongement de la syllabe est dû uniquement à des contraintes phonotactiques, il n’y aura pas de différences de la durée de la consonne pour les conditions **f-IP** et **f-ip**, mais uniquement entre ces deux conditions et **f-AP**. De plus, aucune différence ne sera attendue pour la durée de la voyelle suivante. Cela est confirmé par les analyses statistiques : la durée de /b/ est plus longue pour **f-ip** que pour **f-AP** ($t = -8.35, p < .01$), mais aucune différence n’a été trouvée ni entre **f-ip** et **f-IP**, ni lorsque la durée de la voyelle a été considérée en fonction des trois niveaux de constituance ($pMCMC > .01$).

4.3.3 Discussion

Dans cette expérience, nous avons testé l’existence d’un troisième niveau de constituance prosodique, que nous avons défini comme le syntagme accentuel. Notre hypothèse était que si ce constituant existe, il doit être signalé par des indices de *phrasing* perceptifs et acoustiques. En particulier, nous avons supposé que le syntagme accentuel doit être signalé par un ton de périphérie différent de l’accent de phrase et par un degré d’allongement et de rupture perceptive inférieurs à ceux trouvés pour le syntagme intermédiaire.

Nos données semblent confirmer cette hypothèse. En premier lieu, de façon similaire à l’Expérience I, nous avons trouvé que, à l’intérieur du même syntagme intermédiaire, la région de f_0 entre l’accent prénucléaire et l’accent suivant est caractérisée par une différence systématique de forme et de pente selon la modalité intonative. L’Expérience II nous suggère aussi que la façon dont le ton se manifeste dans le signal acoustique dépend partiellement du type d’accent mélodique. Dans les questions, lorsque l’accent prénucléaire (LH*) est suivi par un autre accent montant, la f_0 descend graduellement du pic prénucléaire au L suivant. A la fin du premier mot prosodique (*Agronomo* ou *benevolo*) la pente de f_0 change brusquement et devient plus raide pour atteindre la cible basse du deuxième accent mélodique. Chez les locuteurs FC et LL, la présence d’une cible après le premier accent est responsable du changement brusque des valeurs de la pente interaccentuelle dans les questions et dans les affirmations. Chez MD, la forme de la f_0 est plus aplatie dans les questions, et la différence de coefficient de la pente avant le point d’inflexion de la courbe est juste au dessous du niveau de significativité. La différence des valeurs de la pente après ce point est significative, dans la mesure où la présence d’une

cible basse dans les affirmations provoque une configuration à plateau basse entre la cible même et le début du deuxième accent. On pourrait supposer que dans les affirmations de MD un ton L_{AP} est associé à la fin du premier mot prosodique, tandis qu'aucun ton n'est présent dans les questions. Cette hypothèse doit être rejetée, car elle implique une différence structurelle entre les deux modalités. Une autre hypothèse est que même lorsque les tons d' AP ne sont pas réalisés comme des points d'inflexion bien visibles dans la courbe de f_0 , leur présence "sous-jacente" provoque des variations plus subtiles sur les propriétés dynamiques du contour. Selon cette hypothèse, le ton H_{AP} sous-jacent chez le locuteur MD ne permet pas un changement brusque de la pente, comme c'est le cas dans les affirmations après le point d'inflexion. Chez DD, un accent montant $L+H^*$ et un accent descendant H^*+L délimitent la région prénucléaire dans les questions et dans les affirmations. Contrairement aux prédictions de la Théorie AM, les deux pics ne sont pas liés par une interpolation *sagging* (due à l'insertion d'une cible phonétique basse entre les deux accents), mais par une ligne droite du premier au deuxième pic. Cependant, la différence entre les deux accents est encore signalée par une simple différence d'inclinaison de pente, qui est plus douce dans les questions et plus raide dans les affirmations. Il faut noter encore que, pour les quatre locuteurs examinés, les valeurs d'intercepte sont toujours plus élevées dans les questions que dans les affirmations, ce qui implique, à nouveau, l'existence d'un ton H_{AP}/L_{AP} dans les deux modalités. En particulier, ce ton semble aligné dans la région après la fin du premier mot accentué, suggérant qu'il n'y a pas d'association stricte à la périphérie droite du constituant prosodique.

Une autre observation mérite d'être faite. Jusqu'à alors, nous avons supposé que le troisième constituant évoqué est le syntagme accentuel. On pourrait objecter que ce constituant est plutôt le mot prosodique. Par exemple, une preuve de cette interprétation alternative a été trouvée dans une autre langue romane, le portugais, où la fin du premier mot prosodique est optionnellement signalée par un ton de périphérie haut (D'Imperio *et al.* (2007b) et références y citées). En effet, dans les deux expériences présentées dans ce chapitre, nous nous sommes intéressés uniquement à la région de f_0 entre le premier et le deuxième mot prosodique. Bien que l'existence d'un troisième niveau de constituance soit donc claire en italien, il s'avère encore difficile de définir son domaine, et notre proposition

d'*AP* reste encore provisoire.

En ce qui concerne les données relatives à la durée segmentale, elles renforcent l'idée que ce troisième constituant, marqué au niveau tonal, ne peut pas correspondre au syntagme intermédiaire. En effet, nous avons trouvé que le degré d'allongement final est plus important en présence d'une frontière de syntagme intermédiaire que d'une frontière d'*AP*. Un résultat très important est que le voisinage des frontières prosodiques n'influence pas la durée de la syllabe finale du constituant, mais uniquement la durée de la syllabe tonique. Ceci est en désaccord avec la littérature concernant l'allongement final, qui a mis en évidence que l'allongement final dû à la présence d'une frontière prosodique se reflète dans la durée de la syllabe (ou de la rime) finale et, optionnellement, dans la durée des segments ou des syllabes précédentes. Par exemple, selon Turk & Hufnagel (2007), l'allongement acoustique peut s'étendre de la syllabe finale jusqu'à la syllabe tonique du mot, tandis que les syllabes intermédiaires sont sujettes à des degrés d'allongement inférieurs à ceux trouvés dans la syllabe finale et tonique. La présence d'un accent mélodique ne semble pas jouer de rôle sur l'allongement de la syllabe tonique (Turk & Hufnagel, 2007). Cependant, à notre connaissance, aucune étude acoustique n'a montré jusqu'à maintenant que la syllabe tonique est le seul "attracteur" de l'allongement final.

Une possible interprétation de nos résultats nous est offerte par de récents développements de la notion de π - *gesture* dans le cadre de la Phonologie Articulatoire (Byrd & Saltzman, 1998, 2003). Selon cette approche, l'allongement final est dû à un geste articulatoire, le π - *gesture*, qui contrôle la vitesse d'articulation des segments en présence des frontières prosodiques. Lorsqu'une frontière prosodique est présente, le π - *gesture* ralentit l'articulation des segments avec lesquels il se superpose, ce qui est à l'origine de l'allongement prosodique. Très récemment, Riggs & Byrd (2007) ont proposé deux interprétations du π - *gesture*, l'*extension account* et le *pitch account*. Selon la première interprétation, la durée du π - *gesture* est fixe, indépendamment de la durée des segments avec lesquels il se superpose. Par conséquent, lorsque la durée d'articulation des segments de la syllabe finale est inférieur à la durée d'activation du π - *gesture* (par exemple, pour la présence d'une voyelle relâchée ou *lax*; pour l'absence d'une coda consonantique), l'allongement peut s'étendre à gauche de la syllabe même. Selon le *pitch account*, la présence d'un accent

mélodique autour de la frontière prosodique “attire” le π - *gesture* : la syllabe accentuée va être la seule syllabe allongée, tandis que ni la syllabe finale ni les syllabes intermédiaires ne subiront d’allongement. Tandis que les données articulatoires pour l’anglais semblent soutenir l’*extension account*, nos données pour l’italien semblent plutôt fournir une preuve acoustique pour le *pitch account*.

Pour résumer, en italien napolitain le contraste entre questions et affirmations est signalé acoustiquement par une différence d’alignement de l’accent nucléaire (L^*+H/L^*+H) ainsi que par une différence dans la spécification tonale d’un ton de périphérie, que nous avons provisoirement attribué à la présence du syntagme accentuel. En perception, plusieurs études (D’Imperio, 2000 ; voir aussi chap. 3) ont montré que la différence d’alignement est un indice robuste pour l’identification intonative. Une hypothèse possible est que même la différence de spécification tonale peut être exploitée pour la perception des questions et des affirmations. Cette hypothèse sera testée dans le chapitre suivant.

Effets du ton de Syntagme Accentuel sur la perception du
contraste intonatif

5.1 Problématique

Un présupposé fondamental de la Théorie AM est que le contour intonatif est composé d'une séquence d'accents mélodiques et tons périphériques générés par une grammaire à états finis (Pierrehumbert, 1980). Le contour se construit de gauche à droite en passant d'un état au suivant : on réalise d'abord un accent mélodique ou plus, ensuite un accent de syntagme et enfin un ton de frontière. La grammaire de Pierrehumbert (1980) ne donne aucune information sur la probabilité des transitions d'un état à l'autre, car elle est purement "non-déterministe". En d'autres termes, selon Pierrehumbert (1980) toutes les combinaisons entre les états du système sont possibles. Par exemple, un accent $L+H^*$ ou un accent L^*+H peuvent se combiner librement avec un ton périphérique haut ou bas. La libre combinaison de ces éléments dépend du fait qu'ils contribuent de façon indépendante au sens intonatif du contour, contour qui est en fait le résultat de l'addition de la signification de chaque morphème tonal (Pierrehumbert & Hirschberg, 1990). Le domaine d'interprétation de chaque morphème dépend de son domaine phonologique. Les

accents mélodiques véhiculent une information sur le statut des mots auxquels ils sont associés, tandis que les accents de syntagmes et les tons de frontière indiquent si le contenu propositionnel de la phrase doit être évalué relativement aux syntagmes intermédiaires et intonatifs précédents ou suivants. Par exemple, selon Pierrehumbert & Hirschberg (1990), en anglais américain la modalité “déclarative neutre” (*neutral declarative*) est le résultat de la combinaison de l’accent H* (qui véhicule le sens de “nouveau” dans les discours) avec les tons périphériques L- et L% (qui véhiculent l’idée que la phrase ne doit pas être interprétée en relation avec d’autres constituants prosodiques), tandis que la modalité “question” peut être véhiculée par la combinaison de H* (pour les *high-rise questions*) ou L* (pour les *yes/no questions*) avec H- et H%. Une implication importante de la compositionnalité est que les tons sont choisis *localement*, car la présence d’un type d’accent ou de ton périphérique ne peut pas être utilisée pour prédire les types de tons successifs. Par conséquent, lorsqu’un mot contient un accent H*, nous ne pouvons pas prédire de cette information si la phrase dans laquelle l’accent est inséré est une question ou une affirmation. L’ambiguïté dans la modalité intonative est résolue uniquement lorsque l’information relative à la spécification tonale des tons de périphérie (et, en particulier, du ton de frontière) est disponible.

L’interprétation stricte de la compositionnalité du contour intonatif est problématique au regard des données du napolitain. En napolitain, le contraste entre questions et affirmations est véhiculé principalement par le contraste d’alignement de l’accent nucléaire, car l’accent de syntagme et le ton de frontière sont les mêmes dans les deux modalités. Donc, dans une phrase contenant un seul accent mélodique (l’accent nucléaire), le contour intonatif sera construit par la séquence L+H* L- L% dans les affirmations et L*+H L- L% dans les questions. Dans le chapitre 3, nous avons montré que l’alignement de cet accent est un indice robuste pour la perception du contraste, ce qui est en accord avec les études de D’Imperio (2000). Cependant, les découvertes dans le chapitre 4 révèlent que les contours mélodiques des questions et des affirmations ne diffèrent pas seulement dans l’accent nucléaire, mais aussi dans la spécification d’un ton de périphérie qui marque la fin du syntagme accentuel. En particulier, dans les phrases dans lesquelles l’accent nucléaire est précédé par un accent prénucléaire, la région interaccentuelle est composée par un ton

H_{AP} dans les questions et L_{AP} dans les affirmations. Cette découverte semble contraster avec l'idée que les tons dans un contour sont choisis de façon indépendante, car dans notre corpus aucun cas n'a été trouvé dans lequel le ton L_{AP} suivait L^*+H dans les questions et/ou le ton H_{AP} suivait $L+H^*$ dans les affirmations.

Une question est laissée ouverte par nos découvertes : pourquoi les tons de AP et les accents nucléaires subissent-ils des contraintes de cooccurrence ? Une hypothèse possible est que la différence de spécification phonologique entre L_{AP} et H_{AP} peut être utilisée systématiquement pour renforcer le contraste entre questions et affirmations. Cette hypothèse repose sur deux considérations. En premier lieu, en italien le contraste de modalité intonative n'est pas signalé par des différences morphosyntaxiques, comme il peut l'être dans d'autres langues telles que le français ou l'anglais. En fait, le contraste entre questions et affirmations est confié uniquement à l'intonation. En italien napolitain, le contraste intonatif est signalé par une différence très subtile d'alignement entre les deux accents : les deux cibles de l'accent nucléaire LH tendent à être réalisées dans la syllabe accentuée, et le pic des questions peut être plus retardé de celui des affirmations de 45 ms (D'Imperio, 2000). Par conséquent, on pourra supposer que l'information d'alignement véhiculée par l'accent nucléaire est renforcée par un ton de syntagme accentuel contrastif entre les deux modalités.

De plus, dans une perspective "universaliste", l'utilisation d'un ton bas L_{AP}/H_{AP} semble être cohérent avec le contraste entre le pic anticipé/retardé dans la détermination du sens intonatif. En particulier, selon (Gussenhoven, 2002, 2004), l'association forme-fonction dépend de codes biologiques qui influencent la production et la perception de la fréquence fondamentale, tels que le Code de Fréquence. Le Code de Fréquence (Ohala, 1983) est lié à l'idée que dans l'univers animal la corrélation entre la taille du larynx et la fréquence de vibration des cordes vocales est exploitée pour exprimer des relations de pouvoir entre les individus : une voix dont le *pitch* est bas indique un animal de taille grosse et donc dominant ; une voix dont le *pitch* est haut signale que l'animal est de taille petite et donc faible, soumis. Ce contraste de *pitch* peut être employé pour véhiculer des informations grammaticales. Par exemple, l'utilisation d'un ton bas/haut peut signaler que le locuteur est plus certain/incertain, et donc qu'il affirme ou qu'il pose une ques-

tion. Puisque il n'y a pas de correspondance univoque (*one-to-one mapping*) entre forme et fonction, l'interprétation informationnelle du Code de Fréquence peut être réalisée de différentes manières. Par exemple, les locuteurs mettent plus de temps à produire d'un ton H ou d'un pic retardé ; par conséquent, tant l'alignement retardé que la valeur mélodique haut d'une cible tonale peuvent être exploités pour produire une question. Selon cette perspective, en italien napolitain le contraste de modalité intonative sera signalé par les deux indices : le pic retardé de l'accent nucléaire L^*+H et le ton H_{AP} contribuent à signaler que le locuteur prononce une question, tandis que le pic anticipé de l'accent nucléaire $L+H^*$ et le ton L_{AP} signalent que le locuteur prononce une affirmation.

Si la différence de spécification phonologique est utilisée pour renforcer le contraste intonatif, nous supposons aussi qu'elle sera utilisée systématiquement dans la perception des questions et des affirmations. Dans le chap. 5, cette hypothèse est testée par deux expériences de production. La première (Expérience I) vise à déterminer si les auditeurs napolitains sont capables de percevoir le contraste intonatif même lorsque l'information nucléaire n'est pas disponible. Pour cela, nous avons utilisé des stimuli naturels. Par conséquent, ni la valeur mélodique du ton de frontière ni d'autres possibles corrélats de modalité (tels que la vitesse d'élocution) n'ont été explicitement contrôlés. En accord avec la procédure adoptée dans le chap. 3, les participants étaient soumis à une tâche d'identification. La deuxième expérience (Expérience II) est une étude plus contrôlée sur les propriétés sémantiques du ton de *AP* et sur leur contribution dans la perception du contraste intonatif. Les stimuli utilisés dans cette expérience ont été créés par la manipulation de hauteur du ton de *AP*. Les participants étaient soumis à un *semantic differential task* (voir Dombrowski, 2003; Ambrazaitis, 2005 pour les détails sur cette procédure).

5.2 Expérience I

Le but de cette expérience est de vérifier si la perception des questions et des affirmations peut être effectuée même lorsque l'indice le plus robuste pour l'identification du contraste (l'alignement tonal de l'accent nucléaire) n'est pas disponible. En particulier, la

région de f_0 entre deux accents montants (un prénucléaire et un nucléaire) est caractérisée par l'insertion d'un ton L_{AP} dans les affirmations et d'un ton H_{AP} dans les questions. Si la différence de spécification du ton de syntagme accentuel véhicule une information sur le contraste intonatif, nous nous attendons à ce que les auditeurs identifient plus de questions et plus d'affirmations lorsque, respectivement, un H_{AP} ou un L_{AP} est présent dans le signal. Au contraire, si l'information de modalité est véhiculée uniquement par l'accent nucléaire, l'identification intonative pour des stimuli contenant uniquement le ton de AP sera autour du niveau de hasard.

De plus, nous savons que d'autres indices de nature segmentale et prosodique peuvent influencer la réalisation des questions et des affirmations. Par exemple, D'Imperio (2000) a trouvé qu'en napolitain l'identification intonative de stimuli manipulés relativement à l'alignement de l'accent nucléaire dépend partiellement de la modalité originale des phrases utilisées pour la création des stimuli. D'Imperio (2000) suppose aussi que ces indices doivent être de type segmental (tels que le tilt et la balance spectrale ou l'intensité). Des effets similaires ont été trouvés aussi dans l'expérience de perception décrite dans le chap. 3. La vitesse d'élocution semble aussi un corrélé prosodique secondaire pour signaler le contraste phonologique, car, globalement, les questions tendent à être prononcées plus rapidement que les affirmations (D'Imperio, 2000 ; voir aussi les résultats obtenus pour le napolitain dans le chap. 2 et dans le chap. 4 de cette thèse). Par conséquent, si ces indices secondaires influencent l'identification lexicale, les auditeurs napolitains pourront identifier le contraste entre questions et affirmations même lorsque l'information tonale (de l'accent nucléaire et du ton de AP) ne sera pas disponible.

5.2.1 Préparation des stimuli

Trois phrases, extraites du corpus enregistré pour l'Expérience I décrite dans le chap. 4, ont été la base pour cette expérience. Les phrases étaient composées : (1) d'un syntagme nominal au début de l'énoncé (*La mamma*), qui portait un accent prénucléaire sur *mam-* dans les deux modalités intonatives ; (2) d'un syntagme verbal (*vuole vedere*), dans lequel le ton d' AP était réalisé dans la syllabe *vu-* de *vuole* ; et (3) d'un syntagme nominal à la fin de l'énoncé, qui portait un accent nucléaire. Le dernier syntagme nominal était

constitué par un nom propre paroxyton (*Lina*, *Rina* ou *Bina*).

Les phrases ont été lues par le locuteur napolitain OM comme des questions et des affirmations. Par conséquent, les deux modalités intonatives étaient différentes pour la spécification phonologique du ton de *AP* (H_{AP} vs L_{AP}) et de l'accent nucléaire ($L+H^*$ vs. L^*+H). L'accent prénucléaire était montant (LH^*) indépendamment de la modalité intonative.

Chaque production a été coupée à deux positions différentes dans la phrase : à la fin du mot *mamma* et à la fin de *vuole*. Le premier groupe de stimuli contenait l'accent prénucléaire, suivi par une petite portion du mouvement descendant de f_0 du pic prénucléaire au ton de *AP* (condition **Pren**). Le deuxième contenait l'accent prénucléaire suivi par le ton de *AP* (condition **AP**). Les phrases entières ont été incluses dans le corpus comme groupe de contrôle, car dans ces stimuli l'accent nucléaire est lui aussi présent (condition **Nucl**). La Fig. 5.1 schématise les trois conditions expérimentales. Nous avons préféré couper des stimuli à la fin des mots plutôt que dans la région suivant immédiatement l'accent prénucléaire ou le ton de *AP* pour rendre les stimuli plus naturels. Au total, le corpus était constitué de 18 stimuli.

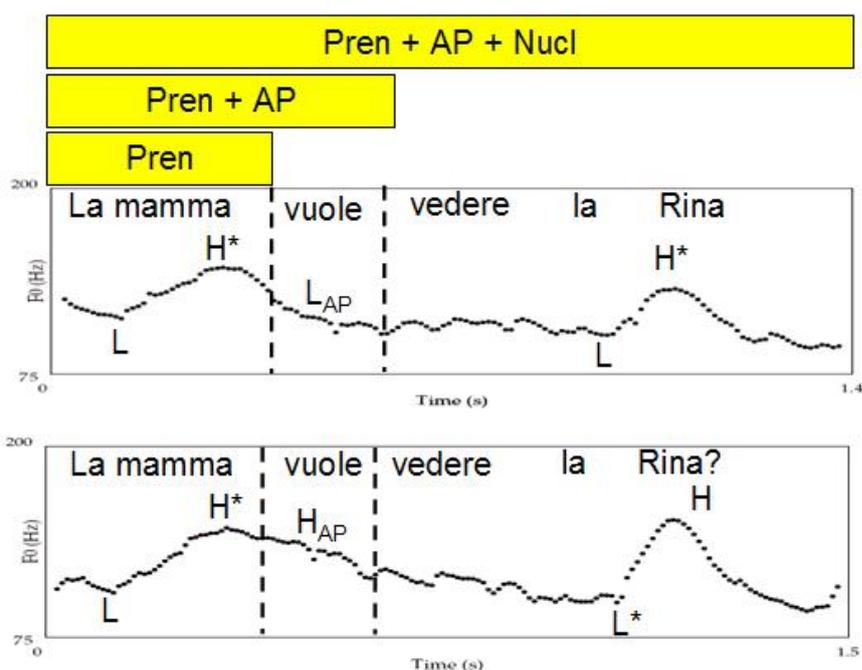


FIGURE 5.1 – Représentation schématisée des trois conditions expérimentales **Pren**, **AP** et **Nucl**.

5.2.2 Déroulement de l'expérience

L'ensemble des stimuli a été reproduit sur un ordinateur portable à l'aide de PERCEVAL (André *et al.*, 2007). Au début de l'expérience, nous avons prévenu les sujets qu'ils allaient écouter des phrases ou fragments de phrases, comme si le locuteur interrompait soudainement la formulation de sa phrase. Les auditeurs devaient donc identifier si les stimuli étaient des questions ou des affirmations. Tous les stimuli ont été présentés de façon binaurale par des casques professionnels (Sennheizer HD 497) dans une chambre silencieuse chez le locuteur. Dans le même temps, les deux étiquettes de modalité étaient visualisés sur l'écran de l'ordinateur dans le même ordre (1. *Domanda*, 2. *Affermazione*) (“Question”/“Affirmation”).

Les fragments de phrase ont été présentés séparément 5 fois dans 3 blocs randomisés. Le premier bloc contenait les stimuli avec l'accent prénucléaire, le deuxième ceux contenant aussi le ton de *AP* et le troisième les phrases entières. Les blocs étaient séparés par une pause de 4 secondes, et une pause de 60 secondes était insérée entre un bloc et le suivant. Neuf napolitains devaient identifier une des deux modalités intonatives de la paire minimale en choisissant sur le clavier la flèche droite pour la question, et la flèche gauche pour l'affirmation. L'expérience durait moins de 10 minutes.

La tâche a été précédée par une phase d'entraînement dans laquelle les auditeurs devaient identifier dix essais sélectionnés parmi : (1) l'ensemble des fragments des conditions **Pren** et **AP** utilisés dans cette expérience et (2) l'ensemble de phrases longues d'un corpus utilisé pour une autre expérience (Petrone & Ladd, 2007). Les essais étaient prononcés par deux locuteurs différents comme questions ou affirmations. Une telle procédure expérimentale a été adoptée pour éviter des effets éventuels d'apprentissage dans la tâche d'identification.

5.2.3 Analyse statistique

Nous savons que le modèle logistique est généralement utilisé pour modéliser des données dont la variable dépendante est binaire et la variable indépendante est (ou est considérée comme) numérique. Par exemple, dans le chap. 5, le score d'identification était

binaire, tandis que le continuum de 14 pas d’alignement a été pris en continu, ce qui nous a permis de modéliser la relation entre les deux variables par l’utilisation d’un modèle logit. Cette procédure ne peut pas être adoptée ici, car le nombre de niveaux de variables indépendantes est trop bas pour pouvoir prendre au moins un des deux en continu (deux niveaux pour le facteur “modalité intonative” du stimulus de base : **Q/A** ; trois niveaux pour le facteur “composition tonale” : **Pren/AP/Nucl**).

Par conséquent, l’analyse statistique a inclus une série d’analyses linéaires à effets mixtes à un facteur fixe (modalité intonative) ou à deux facteurs fixes (modalité intonative et composition tonale). Les auditeurs étaient le facteur aléatoire¹. L’utilisation des modèles de régression linéaire pour modéliser des données catégoriques est tout à fait justifiée d’un point de vue statistique. Elle est même préférable relativement à une analyse ANOVA omnibus, normalement utilisée pour des données dont la variable indépendante est catégorielle. En fait, l’ANOVA omnibus nous révèle l’effet global de la variable indépendante ; cependant, lorsque les variables catégorielles sont composées de plus de deux niveaux, elle ne restitue pas d’information sur les contrastes parmi les niveaux. Au contraire, l’analyse de régression nous offre des informations additionnelles sur la comparaison par paires des niveaux.

5.2.4 Résultats

Si l’identification des questions et des affirmations dépend uniquement de la différence de catégorie tonale de l’accent nucléaire, l’effet de modalité sera significatif uniquement dans les stimuli de contrôle (condition **Nucl**). En d’autres termes, dans les (conditions **Pren** et **AP**), le score de réponse doit être autour du niveau de hasard, car dans ces stimuli l’information relative à l’accent nucléaire n’est pas disponible. L’histogramme dans la fig. 5.2 nous montre cependant une situation différente. Dans cette figure, le score moyen des réponses “questions” est fonction de la composition tonale des stimuli. Les résultats sont

1. Comme nous l’avons dit précédemment, les stimuli étaient construits sur trois phrases interrogatives et trois phrases affirmatives, ayant la même structure segmentale dans les conditions **Pren** et **AP** (les conditions expérimentales importantes pour notre hypothèse), mais non dans la **Nucl** (par. 5.2.1). Dans des analyses statistiques préliminaires, le facteur “mot” a été pris en compte comme un facteur aléatoire additionnel dans l’analyse, mais il a été exclu car la variance expliquée par ce facteur était proche de zéro.

illustrés séparément pour les stimuli dont la base intonative était une question ou une affirmation.

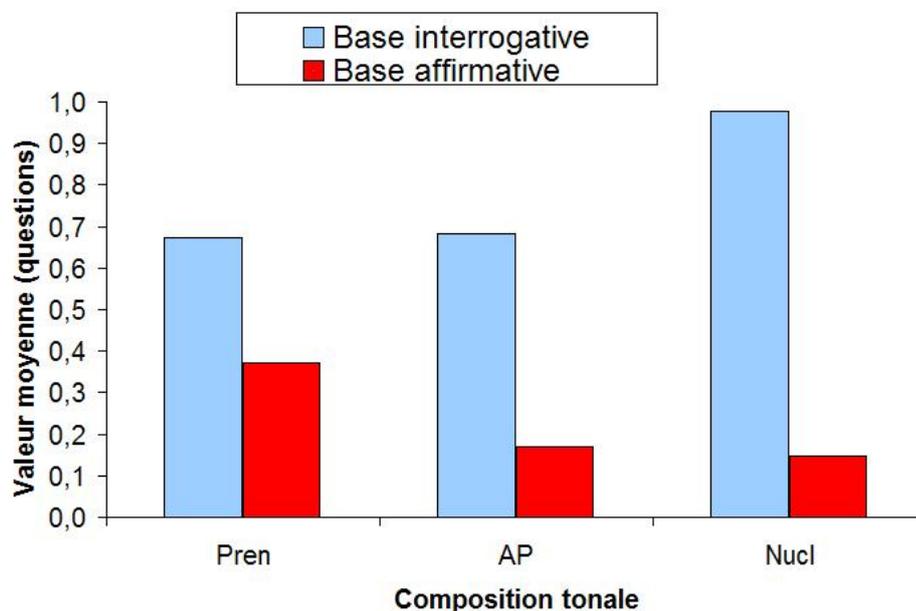


FIGURE 5.2 – Histogramme de réponses “questions” pour les stimuli dont la base intonative était une question (bleue) ou une affirmation (rouge), pour les conditions **Pren**, **AP** et **Nucl**. Les résultats sont regroupés pour tous les locuteurs

Lorsque les auditeurs ont écouté les stimuli dans **Pren**, ils étaient déjà capables de distinguer les questions des affirmations. Pour les stimuli dont la base intonative était une question, le score de réponse est de 0.67, tandis qu’il est de 0.37 pour ceux créés à partir d’une base affirmative. L’analyse de régression à un facteur fixe effectuée sur ce sous-ensemble des stimuli nous confirme que l’effet de la base intonative est significatif [$t = -5.7$, $pMCMC < .01$], suggérant l’existence d’indices de modalité dans la région prénucléaire.

Dans la condition **AP**, les réponses aux stimuli sont aussi loin du niveau de hasard, et l’effet de la base pour ce sous-ensemble est aussi significatif [$t = -10.49$, $pMCMC < .01$]. Cependant, l’importance de l’effet est différente selon la base intonative. Dans les stimuli créés à partir d’une affirmation, la présence du ton L_{AP} semble cruciale pour l’identification intonative. En particulier, le score des réponses “questions” diminue à 0.2 relativement aux résultats obtenus dans la condition **Pren**. Pour les stimuli de la même condition, mais créés à partir d’une question, il n’y a pas d’augmentation dans le score

de réponses “questions” (0.68) relativement à la condition **Pren**. L’analyse de régression nous révèle en fait que la différence entre les deux conditions est significative pour les affirmations [$t = 3.08$, $pMCMC > .01$], mais pas pour les questions [$t = 0.16$, $pMCMC > .01$]. Cela pourrait suggérer que le ton L_{AP} est un indice plus robuste de modalité intonative que le ton H_{AP} . En fait, le graphique nous montre que pour les questions le score de réponse augmente uniquement dans la condition **Nucl** (0.97, $t = -0.96$, $pMCMC < .01$), ce qui pourrait indiquer que pour les auditeurs l’information d’alignement de l’accent nucléaire est cruciale pour l’identification des questions. Il faut noter aussi que la présence de l’accent nucléaire détermine une chute ultérieure du score de réponse à 0.01 dans les stimuli créés à partir d’une affirmation relativement à la condition **AP** [$t = 8.05$, $pMCMC < .01$].

5.2.5 Discussion

La littérature précédente sur le napolitain (D’Imperio & House 1997; D’Imperio 2000, 2002a; D’Imperio *et al.* 2007, *inter alia*) s’était occupée uniquement de l’impact de l’accent nucléaire sur la perception du contraste entre les deux modalités. En particulier, D’Imperio (2000) a montré que l’alignement des cibles LH est un indice très robuste pour l’identification des questions vs. affirmations. D’Imperio (2000) avait aussi exploré l’existence d’autres indices de modalité liés à f_0 et toujours localisés au niveau de l’accent nucléaire, tels que la hauteur mélodique et la forme du contour. L’idée que les informations véhiculées par l’accent mélodique, et en particulier l’alignement tonal, soient cruciales pour la détermination du contraste phonologique a été aussi à la base de nos expériences de production et perception décrites dans les chapitres 2 et 3. Cependant, la découverte d’un ton de frontière de AP 4, nous a amenés à considérer l’hypothèse que les auditeurs puissent exploiter la différence de spécification tonale de l’ AP pour l’identification intonative.

Les résultats de cette expérience, bien que préliminaires, sont très intéressants, car ils vont dans la direction de notre hypothèse. En particulier, ils montrent que les auditeurs napolitains sont capables de reconnaître les questions et les affirmations même lorsque l’accent nucléaire n’est pas disponible.

Un premier résultat est que la région prénucléaire est déjà très informative pour l’iden-

tification du contraste intonatif. Dans la condition **Pren**, les fragments de phrases contenaient uniquement l'accent prénucléaire. Pourtant, l'effet de modalité est significatif indépendamment de la base intonative de l'énoncé. Il faut noter que cela ne peut pas dépendre d'une différence phonologique de l'accent prénucléaire, car il était (LH*) tant dans les questions que dans les affirmations. Puisque nous avons utilisé des stimuli naturels, plusieurs indices peuvent avoir contribué à la perception du contraste, tels que la vitesse d'élocution, les caractéristiques spectrales ou même les différences dans la réalisation de l'accent prénucléaire (LH*). Nous rappelons que le locuteur OM (dont les productions ont été la base pour la création des stimuli utilisés) a prononcé le premier syntagme nominal (*La mamma*) plus rapidement dans les questions que dans les affirmations. De plus, nous avons trouvé que chez le même locuteur, le début de l'accent montant est plus bas dans les questions, ce qui reflète une différence déjà trouvée dans le L de l'accent montant nucléaire.

Un résultat très important est celui relatif à la condition **AP**. Dans cette condition, les stimuli étaient constitués par le fragment *La mamma vuole*, qui contenait l'accent prénucléaire ainsi que le ton de AP, spécifié comme L_{AP} dans les affirmations et H_{AP} dans les questions. Si la perception du contraste entre questions et affirmations pour ces stimuli dépend uniquement de facteurs phonétiques tels que la vitesse d'élocution ou de différences dans la réalisation de l'accent prénucléaire, le score de réponses devrait être similaire à celui obtenu dans la condition **Pren**. Au contraire, dans le stimuli créés à partir d'une affirmation, l'identification des affirmations est beaucoup plus élevée dans la condition **Pren**, suggérant que la présence du L_{AP} est cruciale pour l'identification intonative (0.82). La présence du H_{AP} n'a pas d'impact important, et le score d'identification reste très similaire à celui obtenu dans la condition **Pren** (0.68). L'amélioration du score d'identification des affirmations comparé aux résultats obtenus pour les questions indique, à nouveau, que le score de réponse ne peut pas dépendre d'une différence de vitesse d'élocution, mais d'une différence au niveau tonal. En fait, si le score est fonction de la vitesse d'élocution, on ne comprendra pas pourquoi cet effet doit amener à une amélioration du score d'identification dans une modalité intonative et non dans l'autre. Une autre hypothèse, qui nous semble plus probable, est que le L_{AP} soit un indice moins ambigu pour

les affirmations que le H_{AP} pour les questions². Nous savons que généralement les tons bas ont tendance à véhiculer le sens de “finalité” et donc d’ “affirmation”, tandis que les tons hauts celui de “continuation” ou encore d’ “interrogation”. Ces associations de forme et fonction semblent dépendre de codes biologiques qui influencent la production de la f_0 , tel que le Code de Production ou le Code de Fréquence (Gussenhoven, 2002, 2004). Par conséquent, il est possible que la présence du ton H_{AP} ait été interprétée comme un indice de continuation, et c’est uniquement grâce à l’ajout de l’accent nucléaire L^*+H que l’interprétation du contour a été déplacée vers les questions. Des phénomènes similaires de “déplacement” dans l’interprétation des contours montants sont déjà connus en littérature. Par exemple, Caspers (1998) a observé qu’en hollandais lorsque les accents L^*+H et H^* sont suivis par un plateau haut de f_0 (en termes autosegmentaux, par un accent de syntagme H-) le contour est interprété comme un contour de continuation, tandis que la présence additionnelle d’un ton $H\%$ provoque une augmentation de la perception des questions.

Si cette hypothèse est vraie, cela indiquerait que le paradigme de la perception catégorielle (et en particulier, le test d’identification) n’est pas capable de rendre compte de l’espace sémantique couvert par le ton de AP , et donc de rendre compte de sa contribution dans la perception des catégories pragmatiques de “question” et “affirmation”. Un paradigme plus prometteur pour répondre à ce genre de questions est, à notre avis, le *semantic differential task*. Ce paradigme a donc été utilisé dans l’Expérience II.

5.3 Expérience II

Dans cette expérience, la contribution du ton de AP dans le sens intonatif a été vérifiée de façon plus contrôlée que dans l’expérience précédente. Puisque l’identification des questions et des affirmations semble dépendre de plusieurs indices segmentales et supraseg-

2. Cette hypothèse est cependant en contradiction avec l’hypothèse qu’on a formulé dans le chapitre 4, selon laquelle, en italien, la fin des syntagmes intermédiaires en position non-finale de syntagme intonatif est signalée principalement par un ton de frontière bas. Puisque les stimuli de perception étaient des fragments de phrase, on aurait dû s’attendre à ce que le ton L véhicule l’idée de continuation et donc qu’il était plus ambigu que le H pour les auditeurs napolitains. Une possible explication est que le mouvement descendant à la fin des syntagmes intermédiaires non-finaux est le résultat d’une simple interpolation. Cette hypothèse ne sera pas testée dans ce travail.

mentales, dans l'Expérience I nous n'avons pas pu mieux évaluer l'impact informationnel du ton de *AP*. De plus, nous avons suggéré que la différence de résultats dans le score d'identification entre les stimuli dont la base intonative était une question/affirmation peut avoir été le résultat d'un biais dû à l'adoption d'un test d'identification.

Ce dernier point nous ramène à des considérations d'ordre méthodologique. Nous rappelons que cette thèse a des buts théoriques et méthodologiques. Nous ne sommes pas seulement intéressés à comment la variabilité (phonétique ou phonologique) peut influencer la détermination des contrastes intonatives, mais aussi à explorer comment de différentes techniques ou procédures pour l'extraction de l'information tonale peuvent soutenir notre recherche. En perception, nous avons utilisé jusqu'à maintenant le paradigme de perception catégorielle, car nous avons exploré l'influence du plan segmental sur un contraste, celui d'alignement de l'accent nucléaire, dont la nature catégorielle avait été déjà établie par la littérature (chap. 3). Cependant, nous savons également que le paradigme de perception catégorielle a été appliqué à l'intonation avec des résultats très différents dans la littérature (nous renvoyons à Gussenhoven, 2004 pour un résumé de ces résultats). Par exemple, Kohler (1987) a montré que le test d'identification amène à des résultats assez clairs sur la perception du contraste entre le pic avancé/moyen/retardé en allemand. De l'autre côté, dans des études successives, Niebuhr & Kohler (2004) et Niebuhr (2007) ont montré que l'utilisation de ce paradigme est souhaitable uniquement pour certains contrastes : en particulier, Niebuhr (2007) a trouvé qu'en allemand le score d'identification offre à une forme sigmoïde lorsque le pic de f_0 est aligné avec l'onset consonantique/le début de la voyelle accentué tandis que la fonction d'identification est moins abrupte lorsque la différence entre pic avancé/retardé se réalise à l'intérieur de la voyelle.

Récemment, d'autres études ont utilisé le *semantic differential task* pour mieux étudier la relation entre forme et fonction de l'intonation. Le *semantic differential task* repose sur l'idée que cette association doit être étudiée à l'aide d'une tâche de *rating* multidimensionnelle, dans laquelle les sujets doivent donner des jugements sur plusieurs échelles sémantiques. En particulier, dans les études sur l'intonation, cette tâche a été utilisée par Uldall (1960) et, plus récemment par Grabe *et al.* (1998); Dombrowski (2003); Kohler

(2005); Radtchke & Harrington (à paraître). Parmi ces études, il faut rappeler celle de Radtchke & Harrington (à paraître), car ils ont utilisé le *semantic differential task* pour évaluer l'existence d'une distinction catégorielle de hauteur entre trois accents mélodiques en allemands, H*, H*+!H et H+L*. Pour cela, ils ont créé des stimuli résynthésés par l'augmentation progressive de la valeur mélodique du ton L de l'accent H+L*. Le continuum de hauteur ainsi créé s'étendait sur les trois catégories, de l'accent H+L* aux accents H*+!H et H*. Ensuite, ils ont demandé aux auditeurs de donner leurs jugements sur une série d'échelles sémantiques. Les échelles avaient été choisies sur la base d'hypothèses sur la signification linguistique et paralinguistique véhiculée par les trois accents. Les résultats de leur l'expérience suggèrent que le continuum était partitionné par les auditeurs en deux catégories (H* et H+L*). En particulier, une série de t-test effectuée sur les trois groupes de stimuli correspondant aux trois accents a confirmé que les jugements des auditeurs pour l'accent H+!H* ne différaient pas de ceux pour H+L*. Cela suggère donc qu'en allemand il n'a pas de distinction catégorielle entre H+!H* et H+L*.

De façon similaire à Radtchke & Harrington (à paraître), dans l'Expérience II nous avons exploré l'existence d'une distinction catégorielle de hauteur entre les deux tons H_{AP} et L_{AP} à l'aide du *semantic differential task*. Les échelles ont été construites sur la base d'hypothèses *a priori* sur le sens linguistique et paralinguistique du ton périphérique. Ce choix est dû au fait que nous ne savons pas encore si le sens véhiculé par ce ton est de nature informationnelle ou affective, c'est-à-dire si son utilisation est liée au contraste entre questions et affirmations ou à la transmission d'informations sur l'état émotionnel du locuteur (politesse, amitié, etc). En fait, l'existence de morphèmes affectifs est déjà attestée en littérature. En particulier, Grabe *et al.* (1998) ont trouvé qu'en hollandais les tons %H (*high prehaed*) et %L (*low prehaed*) véhiculent l'idée d'un personne sociable et gentille, mais aucune information relative au contenu propositionnel du message (mais voir Gussenhoven, 2004 pour une interprétation différente). Si le ton de AP est utilisé pour renforcer le contraste entre questions et affirmations, nous supposons que sa présence va influencer les jugements des locuteurs sur les échelles linguistiques. Cependant, s'il a une valeur purement affective (comme dans le cas de l'hollandais), il ira influencer exclusivement les réponses sur les échelles paralinguistiques.

5.3.1 Préparation des stimuli

Le corpus utilisé dans cette expérience était constitué par des stimuli resynthétisés à partir de la phrase *La mamma vuole vedere la Dina*, prononcée comme une affirmation à focalisation étroite sur *Dina* par le locuteur OM. Cette phrase contenait un accent prénucléaire (LH*), dont les cibles tonales étaient alignées respectivement avec le début et la fin de la syllabe *mam-* de *mamma*; un ton de frontière de *AP* sur la syllabe *vu-* de *vole* et un accent nucléaire L*+H, dont les cibles tonales étaient alignées respectivement avec le début de l'attaque consonantique et avec la voyelle de la syllabe *Di-* de *Dina*.

La phrase a été d'abord coupée en correspondance avec l'attaque de la syllabe nucléaire pour éviter que l'information d'alignement véhiculée par l'accent nucléaire puisse aider les auditeurs dans la perception des questions et des affirmations.

La courbe de f_0 du fragment ainsi obtenu a été stylisée par une série d'interpolations linéaires qui liaient 5 points : un point au début de l'énoncé, deux points respectivement au début et à la fin de la montée de f_0 de l'accent prénucléaire, un point en correspondance avec le ton de frontière L_{AP} , et un point à la fin du fragment même (correspondant à la position du L nucléaire). Les valeurs mélodiques du début de l'énoncé, de l'accent prénucléaire et du L nucléaire étaient intermédiaires entre celles des affirmations et des questions. Ce choix a été fait pour éviter des effets éventuels des valeurs mélodiques de l'accent prénucléaire ou du ton nucléaire L. En particulier, les valeurs de f_0 au début de la phrase et au début de la montée prénucléaire ont été fixées à 114 Hz, la valeur du pic prénucléaire à 157 Hz, celle de L_{AP} à 120 Hz et celle de la cible nucléaire L à 103 Hz. Les valeurs d'alignement de l'accent prénucléaire et du ton L nucléaire étaient celles trouvées dans le stimulus naturel, car dans le chapitre 4 nous avons trouvé que chez ce locuteur la localisation temporelle de ces cibles ne différait pas selon la modalité intonative. Ensuite, nous avons créé un continuum dans le domaine mélodique par l'augmentation progressive de la valeur de f_0 du ton L_{AP} jusqu'à une valeur similaire à celle du pic prénucléaire en 10 pas de 0.5 demi-tons³. En suivant le critère de discriminabilité de 3 demi-tons fixé par

3. Une autre possibilité était d'augmenter la valeur mélodique de L_{AP} jusqu'à la valeur mélodique correspondante au point d'inflexion de la courbe logistique ajustée aux questions, qui était de 138 Hz chez OM (voir chap. 4). Cela a été rejeté pour deux raisons. En premier lieu, nous ne sommes pas sûrs que le point d'inflexion de la courbe corresponde à la cible H_{AP} , car il pourrait être signalé uniquement

't Hart (1981), on s'attend à voir la frontière entre L_{AP} et H_{AP} autour des stimuli H5 et H6. Nous avons donc émis l'hypothèse que les cinq premiers stimuli (H1 - H5) correspondent au L_{AP} et les cinq derniers (H6 - H10) à H_{AP} . Les stimuli, resynthétisés à l'aide de PSOLA (Carpentier & Moulines, 1990), ont été coupés dans deux positions différentes : après *La mamma* et après *La mamma vuole*. Les fragments plus petits contenaient uniquement l'accent prénucléaire et la première portion de la chute de f_0 du pic prénucléaire vers le ton de frontière de *AP*. Le ton de *AP* était présent dans le deuxième groupe de stimuli. Les figures 5.3 et 5.5 schématisent respectivement la manipulation de hauteur du ton L_{AP} dans la courbe stylisée et la composition tonale des fragments de phrase.

Finallement, deux phrases de contrôle ont été ajoutées, dans lesquelles les valeurs aux extrémités du continuum ont été combinées respectivement avec la configuration nucléaire en montée-descente des affirmations et des questions. L'alignement et la hauteur mélodique du pic de l'accent nucléaire et du ton de frontière L- correspondaient aux valeurs moyennes trouvées pour ce locuteur (fig. 5.5).

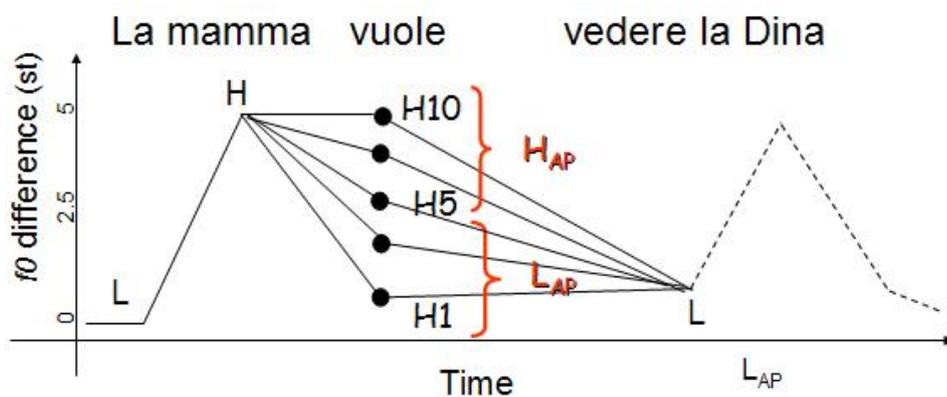


FIGURE 5.3 – Représentation schématisée du continuum dans le domaine mélodique créé par l'augmentation progressive de la valeur mélodique du ton L_{AP} . L'association des pas du continuum avec les tons de *AP* est indiquée à droite

par l'absence d'une chute soudaine de f_0 après le pic prénucléaire, qui offre au contour de f_0 une forme parabolique. En fait, l'interpolation linéaire entre ce point et le pic prénucléaire aurait créé un contour très différent de celui convexe trouvé dans cette modalité.

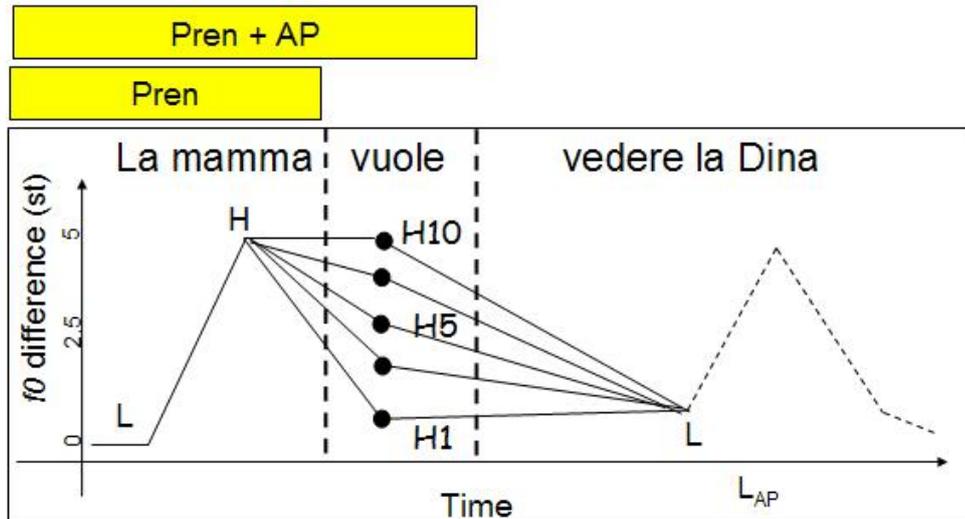


FIGURE 5.4 – Représentation schématisée des trois conditions expérimentales **Pren** et **AP**.

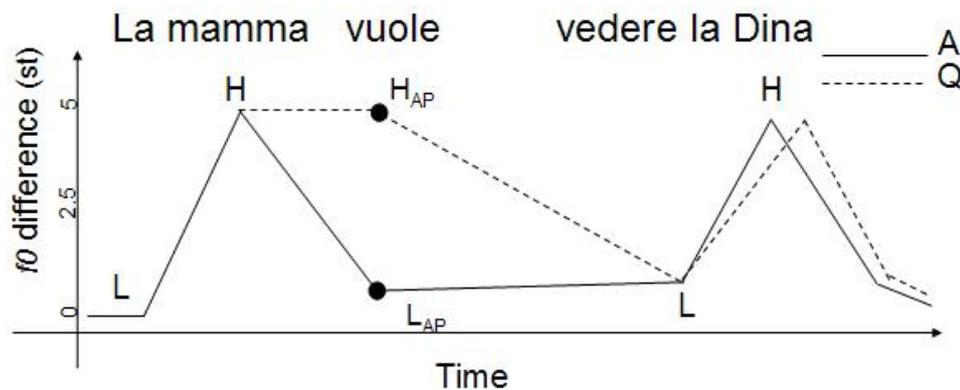


FIGURE 5.5 – Représentation schématisée des deux phrases de contrôle. La ligne continue indique le stimulus représentatif des affirmations, la ligne tiretée celui représentatif des questions.

5.3.2 Échelles sémantiques

Les cinq échelles sémantiques que nous avons choisies se basent sur des hypothèses *a priori* sur les propriétés sémantiques (linguistiques et paralinguistiques) des tons L_{AP} et H_{AP} et en particulier sur la contribution des deux tons dans le contraste entre questions et affirmations.

La première échelle (“engagement”) repose sur une hypothèse sur le sens informationnel véhiculé par les tons L_{AP} et H_{AP} dans l’interaction discursive. En particulier, en suivant

Marandin et al. (2006) Delais-Roussairie (2005), nous supposons que l'utilisation des ces deux tons signalent comment le locuteur pense que son message va être "reçu" par l'auditeur. Selon cette hypothèse, la présence d'un ton bas signale que le locuteur pense que le contenu propositionnel de son message va être accepté par l'auditeur indépendamment du fait que ses croyances soient compatibles avec celles du locuteur. Par conséquent, le locuteur pense que l'auditeur continuera la conversation en fonction. En utilisant un contour montant, le locuteur signale de penser que ces croyances sont incompatibles avec celles de son interlocuteur, mais qu'il s'engage à réviser ses propos. Les étiquettes choisies pour indiquer ce contraste sont *contestabile/incontestabile* ("contestable/incontestable").

La deuxième échelle ("puissance") repose sur l'idée que les tons bas et hauts véhiculent respectivement "certitude" et "incertitude". Selon Gussenhoven (2002), le sens de certitude est lié à l'interprétation informationnelle du Code de Fréquence, car il conduit à la distinction entre affirmations et questions. Les étiquettes utilisées pour cette échelle sont *insicuro/sicuro* ("incertain/certain").

Les troisième ("activité"), quatrième ("évaluation") et cinquième ("soumission") échelles sont aussi liées à l'interprétation affective du Code de Fréquence, selon lequel les tons bas véhiculent l'idée d'une personne qui n'est pas intéressée par le discours, hostile et dominant, tandis que les tons hauts véhiculent l'idée d'une personne engagée sur le plan émotionnel, sociable et soumise. Les étiquettes pour les trois échelles sont respectivement *distaccato/coinvolto* ("détaché/engagé"), *amichevole/ostile* ("sociable/hostile") et *sottomesso/autoritario* ("soumis/autoritaire").

Vue la distinction de sens associée à l'interprétation informationnelle et attitudinale du Code de Fréquence, nous nous attendons à ce que le pôle gauche de chaque échelle soit associé par les auditeurs au ton H_{AP} tandis que le pôle droit soit plus approprié pour L_{AP} .

5.3.3 Déroulement de l'expérience

L'expérience a eu lieu dans une salle silencieuse chez un des neuf participants. Tous les stimuli ont été reproduits sur un ordinateur portable grâce à PERCEVAL. Les auditeurs écoutaient d'abord les stimuli coupés (conditions **Pren** et **AP**) dont l'ordre de

présentation était aléatoire. Les phrases de contrôle étaient présentées à la fin de la session.

Chaque stimulus écouté était associé avec chaque échelle sémantique. Les échelles étaient visualisées les unes après les autres sur l'écran de l'ordinateur. La réponse était donnée en appuyant sur un des sept boutons correspondant aux étiquettes des échelles visualisées. Les stimuli étaient évalués entre -3 et 3, c'est-à-dire entre les extrêmes aux deux pôles. Les sujets devaient interpréter les valeurs des échelles de la façon suivante : "0" comme "neutre", "+/-1" comme "un peu", "+/-2" comme "assez" et "+/-3" comme "très". Chaque stimulus était présenté deux fois de suite, pour renforcer l'impression auditive des stimuli et rendre la tâche plus simple. Les auditeurs pouvaient répondre uniquement après la deuxième répétition.

Au début de l'expérience, nous avons prévenu les sujets qu'ils allaient écouter des phrases ou fragments de phrases, comme s'ils entraient soudainement dans une salle où une personne parlait avec une autre. Dans ce "scénario", Le locuteur va s'interrompre dès qu'il entend que la porte s'ouvre. Les sujets devaient donc donner des jugements sur l'impression qu'ils avaient du locuteur. Les instructions étaient aussi visualisées sur l'écran de l'ordinateur au début de l'expérience.

L'ensemble des stimuli était précédé par une phase d'entraînement des 10 fragments du corpus (**Pren** et **AP**), chacun associé avec une des 5 échelles sémantiques. À la fin de la phase d'entraînement, les auditeurs avaient la possibilité de poser des questions sur les étiquettes utilisées. A la fin de l'expérience, nous avons demandé aux locuteurs de donner leur opinion sur la difficulté de la tâche et sur le caractère naturel des stimuli. Généralement, les auditeurs disaient avoir bien compris le sens des étiquettes utilisées, mais certains parmi eux ont contesté leur utilisation pour cette tâche⁴. Les stimuli étaient jugés naturels. L'expérience durait environ 20 minutes.

5.3.4 Analyse statistique

Les analyses statistiques des échelles sémantiques se basent sur des ANOVA omnibus, qui testent l'effet de chaque échelle sur l'ensemble des réponses (Kohler, 2005, *inter alia*).

4. Par exemple, un locuteur a admis avoir bien compris que les fragments appartenaient à des questions et à des affirmations, mais il n'a pas compris pourquoi nous avons utilisé certaines étiquettes (tel que "certain/incertain") pour définir le contraste.

Si l'ANOVA omnibus est significatif, une série de t-tests est effectuée pour comparer par paires l'ensemble des réponses obtenues pour chaque pas de l'échelle. Ambrazaitis (2005) suggère aussi d'effectuer des t-tests sur deux stimuli A et B uniquement lorsque leur différence de score est de plus de 1 pas sur une échelle de 7 points (par exemple, si le stimulus A est évalué à 2.5 et B à 3.5).

Cependant, nous ne sommes pas d'accord sur le type de tests paramétriques utilisés. En premier lieu, il faudrait prendre avec précaution les résultats de l'ANOVA omnibus comme critère de sélection pour évaluer la pertinence statistique des contrastes par paires. Le test ANOVA, comme nous le savons, est généralement utilisé pour explorer la relation entre une variable dépendante continue et une variable indépendante catégorielle. Dans les études de perception de l'intonation, la variable indépendante est souvent donnée dans un continuum de pas créés par la manipulation d'une ou plus conditions expérimentales (tels que l'alignement tonal ou la valeur mélodique de f_0), et est ensuite traitée sur le plan statistique comme une variable numérique. L'ANOVA omnibus nous restitue l'effet global de la variable indépendante, sans nous donner d'informations sur le type de relation existant entre (la variable dépendante et la variable indépendante (y-t-il une corrélation linéaire? Positive ou négative?)).

De plus, il est très difficile d'interpréter les résultats du t-test appliqué à des paires d'un continuum. Prenons l'exemple d'un continuum de 7 points sur une échelle bipolaire. Dans cette échelle, les réponses pour le stimulus A sont évaluées à 4 et celles pour les stimuli B sont évaluées à 5, et le t-test nous révèle que cette différence est significative. Comment interpréter au niveau théorique cette différence? Signifie-t-elle que les stimuli entre les pas 1 à 4 véhiculent une signification et ceux entre les pas 5 à 7 une autre signification? Et comment le continuum va-t-il être partitionné, lorsque la comparaison d'une autre paire de stimuli (ex. stimulus C/stimulus D) est elle aussi significative?

Nous pensons que l'analyse de régression est plus appropriée que l'ANOVA omnibus lorsque la variable indépendante est numérique. Elle prend en compte la relation d'ordre entre les variables dépendante et indépendante et donc nous restitue une information sur la tendance globale des réponses en fonction du continuum de pas créé par la manipulation de la condition expérimentale. Par exemple, pour l'échelle bipolaire "certain/incertain",

nous supposons qu'il y a une corrélation négative entre l'augmentation de hauteur du ton LA_P et le degré de "certitude". Si cela est vrai, on s'attendra à ce que le score de réponses puisse être modélisé par une droite décroissante du pas H1 au pas H10 de notre continuum. En particulier, nous avons effectué une série de régressions linéaires à effets mixtes, effectuées séparément pour les fragments courts et de taille moyenne pour chacune des cinq échelles. Les sujets constituaient l'effet aléatoire. Le niveau de significativité était de $pMCMC < .01$. De plus, la comparaison par paire est effectuée sur la base d'hypothèses sur le partitionnement du continuum. Puisque nous avons émis l'hypothèse que les cinq premiers pas (H1 - H5) correspondent à la catégorie LA_P et les cinq derniers (H6 - H10) à HA_P , le t-test sera effectué uniquement sur les deux groupes de stimuli.

5.3.5 Résultats

La fig. 5.6 montre l'évaluation du continuum pour les stimuli dans la condition **Pren**, c'est-à-dire ceux qui contiennent uniquement l'accent prénucléaire (LH*) ainsi que la portion initiale de la chute de f_0 vers le ton de frontière de AP. Le score moyen des réponses (ordonnées) est fonction des 10 pas de la manipulation de hauteur du ton de frontière. Les données sont illustrées séparément pour chaque échelle sémantique. Les réponses ne varient pas en fonction de la manipulation de hauteur : elles sont généralement autour du niveau zéro, c'est-à-dire que le locuteur semblait "neutre". L'analyse de régression révèle en fait que l'effet de hauteur sur le jugement des locuteurs n'est significatif pour aucune des échelles ($pMCMC < .01$). Cela était attendu, car dans cette expérience les facteurs phonétiques tels que la vitesse d'élocution ou les différences d'implémentation de l'accent prénucléaire ont été neutralisés.

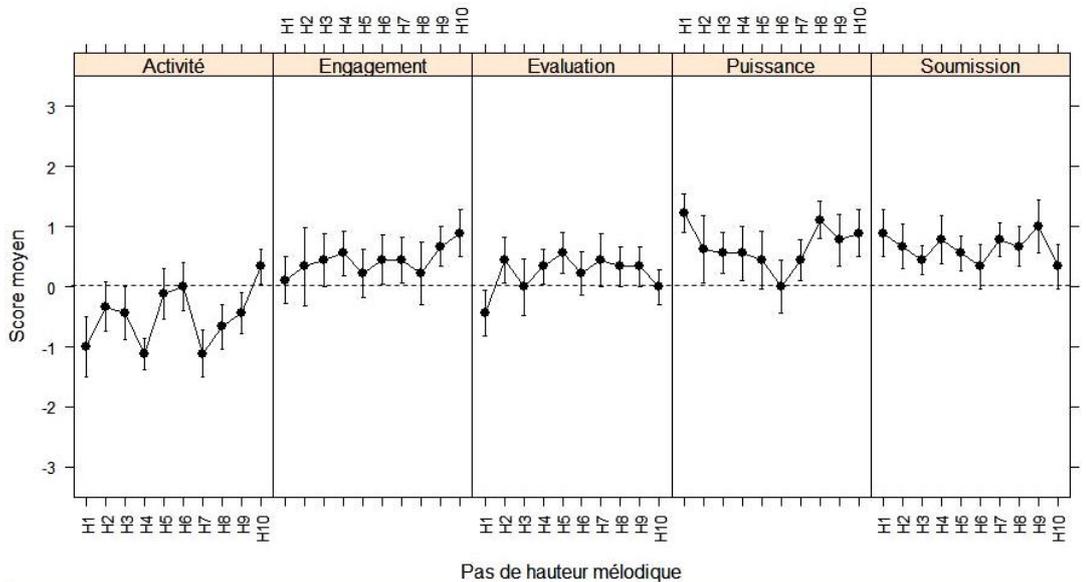


FIGURE 5.6 – Score moyen des jugements des auditeurs pour les cinq échelles sémantiques dans la condition **Pren**. La ligne tiretée signale le pas “zéro” de chaque échelle, correspondant à l’interprétation des stimuli comme “neutres”. Les résultats sont regroupés pour les neuf auditeurs.

La fig. 5.7 nous présente une situation différente. Cette figure illustre les résultats des jugements des auditeurs pour les stimuli contenant l’accent prénucléaire et le ton de frontière de *AP*. Comme on peut le remarquer, les réponses ne varient pas de façon drastique en fonction de la manipulation de hauteur tonale. Cela pourrait être dû à des facteurs différents. Par exemple, pour les échelles linguistiques (“engagement” et “puissance”), il pourrait indiquer que la différence de spécification tonale de l’*AP* est un corrélé secondaire pour le contraste entre questions et affirmations. Cependant, puisque la tendance à la centralisation des réponses a été trouvée aussi pour les échelles paralinguistiques (“activité”, “évaluation” et “soumission”), une hypothèse plus probable est que le score des réponses a été influencé par la nature même des stimuli, qui, étant coupés, auraient pu provoquer une incertitude plus importante dans les jugements des locuteurs. Cependant, nous avons trouvé que le score moyen des réponses diminue progressivement du pas H1 à H10 dans deux échelles sémantiques, l’échelle d’“activité” [$t = -3.1, pMCMC < .01$] et celle de “puissance” [$t = -2.64, pMCMC < .01$]. Dans l’échelle d’“activité” la valeur moyenne à l’extrémité gauche du continuum (H1) est près de zéro (-0.4). Ensuite, le score

moyen diminue progressivement, de sorte qu'à l'extrémité droite du continuum (H9-H10) la valeur moyenne des réponses est de -1.6. Cela signifie que lorsque la valeur mélodique du ton de frontière augmente, le locuteur est jugé de "neutre" à "un peu plus engagé" dans le discours. Dans l'échelle de "puissance", les valeurs moyennes des réponses diminuent de 1.3 (H1) à 0.2 (H9) et -0.3 (H10). Cela signifie que l'augmentation de hauteur du ton de frontière provoque un déplacement du sens d'"un peu certain" à "neutre". Le t-test, effectué séparément sur les deux échelles, nous révèle une différence significative entre le groupe de stimuli H1 - H5/H5 - H10 sur l'échelle d'"activité" [$t = 3.02, p < .01$], mais pas sur celle de "puissance" ($p > .01$) L'effet de la manipulation de hauteur sur les autres échelles n'était pas significatif, ce qui indique qu'elles ne sont pas capables de rendre compte de l'espace sémantique couvert par le ton de AP.

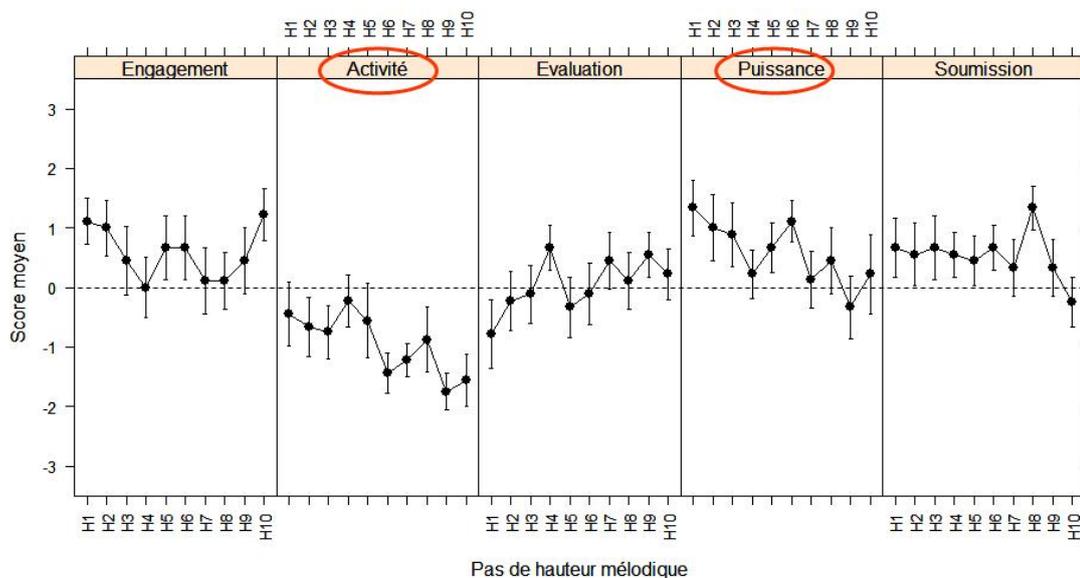


FIGURE 5.7 – Score moyen des jugements des auditeurs pour les cinq échelles sémantiques dans la condition AP. La ligne tiretée signale le pas "zéro" de chaque échelle, correspondant à l'interprétation des stimuli comme "neutres". Les résultats sont regroupés pour les neuf auditeurs.

Nous allons maintenant détailler les résultats obtenus pour les phrases de contrôle (**Nucl**, fig. 5.8). Pour chaque échelle, les valeurs des réponses (ordonnées) sont fonction des deux stimuli choisis comme représentatifs du contraste entre affirmations (à gauche) et questions (à droite). Les deux stimuli diffèrent pour la spécification phonologique du ton

de frontière (L_{AP}/H_{AP}) et de l'accent nucléaire ($L+H^*/L^*+H$). Comme on le remarque, la différence entre les deux stimuli est plus importante que celle trouvée dans les stimuli contenant uniquement l'accent prénucléaire et le ton de frontière. En particulier, dans l'échelle d' "engagement", le score moyen de réponses est de 1.3 pour l'affirmation ("assez incontestable") à -1 dans la question ("un peu contestable"). Donc, la différence est de deux pas dans l'échelle. Dans l'échelle d'"activité", le score était de 0.5 pour l'affirmation et de -1.7 pour la question. La différence dans l'échelle est donc d'un pas : le locuteur est jugé "neutre" dans les affirmations et "un peu engagé" dans les questions. Le modèle mixte confirme une différence significative entre le deux stimuli sur l'échelle d'"activité" [$t = -2.99, pMCMC < .01$] et sur celle d'"engagement" [$t = -2.33, pMCMC < .01$]. Les résultats pour les phrases de contrôle doivent être considérés avec précaution (car le nombre d'observations est très bas : chaque échelle contient 18 observations). Cependant, il est intéressant de noter qu'il y a une superposition partielle entre les propriétés sémantiques du ton de frontière et de l'accent nucléaire, car l'échelle paralinguistique d'"activité" est partagée parmi le groupe de stimuli contenant le ton de frontière et celui contenant aussi l'accent nucléaire. Les deux échelles informationnelles "puissance" et "engagement" contribuent à définir respectivement les propriétés sémantiques du ton de frontière et de l'accent nucléaire.

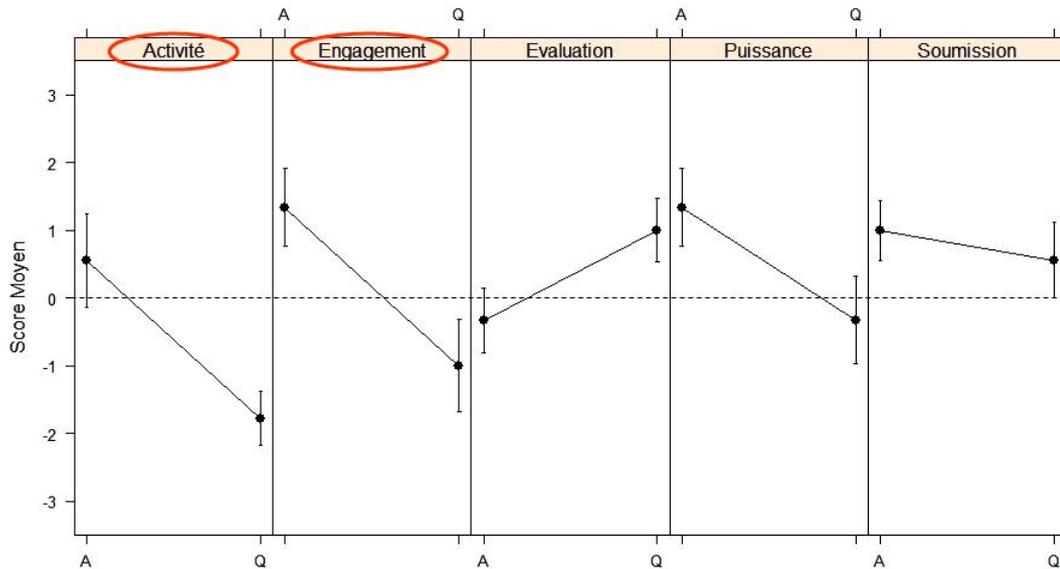


FIGURE 5.8 – Score moyen des jugements des auditeurs pour les cinq échelles sémantiques dans la condition **Nucl**. La ligne tiretée signale le pas “zéro” de chaque échelle, correspondante à l’interprétation des stimuli comme “neutres”. Les résultats sont regroupés pour les neuf auditeurs.

5.3.6 Discussion

Dans cette expérience, nous avons vérifié si les propriétés sémantiques des tons L_{AP} et H_{AP} sont exploitées pour la distinction du contraste entre questions et affirmations en napolitain. Les résultats du *semantic differential task* nous révèlent que le ton de AP véhicule un sens informationnel et affectif. Cependant, nous avons trouvé une différence dans les jugements des locuteurs selon la spécification tonale de ce ton périphérique. En premier lieu, lorsque les stimuli contient un ton L_{AP} , le locuteur est jugé plus certain. Cet effet est graduel mais significatif car, dans nos stimuli, le degré de certitude du locuteur diminue progressivement avec l’augmentation de la hauteur du ton de AP . Ce résultat est très intéressant, car nous savons que l’idée de certitude est lié à la notion d’“assertion” et donc à la modalité affirmative. Cela est aussi congruent avec les résultats obtenus pour les stimuli de contrôle, où la présence additionnelle de l’accent nucléaire $L+H^*$ modifiait une autre échelle linguistique, liée au contraste entre affirmations et questions : celle d’“engagement”. En particulier, le locuteur qui prononce une affirmation est jugé plus

incontestable, moins prêt à réviser son message. L'idée d'incontestabilité est de fait liée aux contours descendants selon Marandin et al. (2006). Nous rappelons aussi que Gussenhoven (2002) a affirmé que le contraste entre tons hauts/bas est souvent substitué avec celui entre pic anticipé/retardé pour signaler la différence de sens entre questions et affirmations. Par conséquent, le fait que l'idée d'incontestabilité est véhiculée par de stimuli contenant aussi l'accent $L+H^*$ est tout à fait compatible avec la grammaticalisation du Code de Fréquence.

Au contraire, le ton de H_{AP} semble donner des informations uniquement sur l'état émotionnel du locuteur, car sa présence indique que le locuteur qui semble plus engagé dans le discours. Cette attitude semble être véhiculée aussi par l'accent L^*+H . En fait, le score de réponse dans l'échelle d'"activité" est plus élevé pour le stimulus représentatif des affirmations que pour les stimuli contenant uniquement le ton de AP . Cela suggère que l'idée de engagement émotif véhiculée par le ton H_{AP} est renforcée par l'accent nucléaire. Cependant, l'accent nucléaire semble être l'indice le plus fort pour la modalité interrogative. En fait, le stimulus de contrôle représentatif des questions a été jugé "plus contestable", tandis que les réponses sur les échelles linguistiques pour les stimuli contenant uniquement le H_{AP} ne variaient pas de façon significative.

Pour résumer, le ton de L_{AP} et l'accent nucléaire $L+H^*$ véhiculent une information sur le message, lié donc à la modalité intonative des affirmations. Au contraire, le ton de H_{AP} ne semble pas capable d'influencer les réponses des auditeurs sur les échelles linguistiques employées ici, car il véhicule uniquement une information sur l'état émotionnel du locuteur. Ces résultats sont tout à fait compatibles avec ceux obtenus dans l'Expérience I, car les deux tests montrent que le ton de L_{AP} est plus informatif que celui de H_{AP} pour véhiculer la modalité intonative. La différence entre les deux tons de AP n'est pas claire, car elle semble suggérer l'existence d'un morphème affectif en napolitain, mais uniquement dans les questions. Bien que les morphèmes affectifs aient été déjà attestés en hollandais, cette explication nous semble peu probable, car elle implique une différence fonctionnelle pour deux tons dans la même position structurelle (le sens informationnel pour le L_{AP} et attitudinal pour H_{AP}). Deux autres explications sont plausibles. Nous pouvons supposer qu'en napolitain, la présence d'un ton bas est un indice plus robuste pour les affirmations que le ton haut pour les questions. Compte tenu du fait que l'indice le plus robuste pour

signaler les questions est le déplacement du pic de l’accent nucléaire, on pourrait supposer que dans cette variété le ton haut est plus ambigu pour signaler cette modalité intonative. Cela est aussi en accord avec des études effectuées dans plusieurs langues, qui ont montré que le ton bas est plus important et plus stable pour véhiculer le sens d’ “affirmation” que le ton haut pour véhiculer le sens de “question” (Gussenhoven, 2002).

Une autre explication est que la façon dont nous avons construit nos stimuli n’est pas appropriée pour apprécier l’influence de la région prénucléaire dans le contraste intonatif. En premier lieu, le stimulus de base de nos manipulations de *scaling* était une affirmation. Nous savons que d’autres propriétés différentes que la f_0 peuvent être utilisées pour la perception des patrons intonatifs, telles que les changements spectraux ou l’intensité (voir par exemple Pierrehumbert & Steele (1989) pour l’anglais et D’Imperio (2000) pour le napolitain). Par conséquent, il est possible que l’utilisation d’une base affirmative dans notre expérience ait biaisé les résultats pour les stimuli dont la valeur de f_0 du ton de AP était élevé (ce patron mélodique étant, au contraire, caractéristique des questions). De plus, en accord avec la Théorie AM, nous avons interprété la différence dans la région interaccentuelle des questions et des affirmations en termes de tons de périphérie statiques (L_{AP} / H_{AP}). Cependant, dans le chap. 4 nous avons montré que les deux modalités étaient caractérisées par une différence dans les propriétés dynamiques de la courbe de f_0 . En particulier, la région suivant le pic prénucléaire dans les questions offre à une forme parabolique. Cette différence de forme n’a pas été “reproduite” dans la construction des stimuli pour l’Expérience II de perception, car les stimuli ont été construits par l’augmentation de la valeur mélodique d’un point dans la courbe (correspondant au ton L_{AP}) et par la liaison de ce point aux points voisins (correspondant au pic prénucléaire et au ton nucléaire L) par des lignes droites. Par conséquent, nous pensons que nos stimuli n’ont pas pu rendre compte des propriétés acoustiques de la région prénucléaire des questions. La caractérisation perceptive de ces caractéristiques dynamiques sera l’objet de nos études dans le futur.

CHAPITRE 6

Conclusion

Cette thèse s'est donné pour objectif d'évaluer le rôle de la variabilité phonétique dans la représentation des contours intonatifs et de leur sens. Le thème de la variabilité a été abordé en production et en perception. En particulier, nous avons exploré comment la réalisation acoustique des cibles tonales est influencée par plusieurs facteurs de variabilité, et aussi comment cette variabilité est traitée par l'auditeur dans l'identification des contrastes phonologiques. L'étude de la variabilité nécessite également une méthodologie appropriée. Le deuxième but de cette thèse était donc de vérifier comment différentes méthodes d'extraction de l'information tonale peuvent nous aider dans l'interprétation des données.

Nous allons commencer ce chapitre par le résumé des découvertes les plus importantes de notre thèse (par. 6.1). Puis, nous allons faire le point sur les apports méthodologiques que ce travail peut offrir (par. 6.2). Nous tenterons aussi d'analyser ce que nos résultats peuvent apporter aux problématiques plus générales liées à la représentation phonétique et phonologique de l'intonation (par. 6.3). Enfin, de possibles conséquences sur le sens intonatif seront aussi esquissées (par. 6.4).

6.1 Résumé des découvertes

Notre travail a démarré par l'exploration de l'interaction entre variabilité segmentale et alignement tonal dans la réalisation (chap. 2) et la perception (chap. 3) des accents nucléaires $L^*+H/L+H^*$ et de l'accent de phrase $L-$ en napolitain. Ce choix est motivé par le fait que la différence systématique d'alignement avancé/retardé entre ces deux accents et l'accent de phrase suivant est un indice robuste pour signaler le contraste entre affirmations à focalisation étroite et questions oui/non (D'Imperio, 2000).

En ligne avec cette étude, nous avons trouvé dans le chapitre 2 que l'alignement acoustique des trois cibles $L1$, H et $L2$ varie en fonction de la modalité intonative. La manipulation de la classe consonantique de l'attaque syllabique influence surtout la réalisation de la cible nucléaire L , qui tend à être alignée autour de l'attaque syllabique. Cependant, les effets d'une telle manipulation sont difficiles à interpréter, car ils dépendent aussi de l'interaction avec la modalité intonative et du locuteur. En particulier, la différence d'alignement du L nucléaire entre les questions et les affirmations est significative lorsque l'attaque consonantique est une consonne nasale ou liquide, tandis qu'aucun effet a été trouvé en présence de l'obstruante. La valeur mélodique de la cible nucléaire L est plus élevée dans les questions que dans les affirmations chez les locuteurs AS et DD. De plus, le L est plus élevé avec les nasales que avec les liquides et les obstruantes, mais la grandeur du contraste entre les trois classes varie selon le locuteur. La manipulation de modalité intonative influence aussi la valeur du pic accentuel, mais les résultats sont incohérents parmi les locuteurs. Chez AS, le pic accentuel est plus élevé dans les questions, tandis que chez DD aucune différence a été trouvée parmi les deux modalités. L'absence de différences chez DD pourrait être due à un rehaussement mélodique du pic de l'accent nucléaire $L+H^*$ à la suite du fait que chez ce locuteur les affirmations ont été lues avec une double focalisation. De plus, chez AS, la valeur mélodique de H dépend aussi de la classe consonantique de l'attaque, ce qui suggère que les effets microprosodiques peuvent s'appliquer aussi sur des distances temporelles assez longues (la cible H étant réalisée à l'intérieur de la voyelle accentuée). Les différences de pente pour la montée LH et la descente $HL-$ semblent être justifiées par des ajustements de la valeur mélodique et de

l'alignement des trois cibles.

A part l'identité segmentale, la structure de la syllabe accentuée influence également l'alignement tonal. L'alignement du pic accentuel relativement à la fin de la voyelle est plus retardé en présence d'une syllabe fermée que d'une syllabe ouverte, dans les deux accents nucléaires $L+H^*$ et L^*+H (D'Imperio, 2000). Dans le chapitre 3, nous avons rendu compte des résultats d'un test d'identification, dans lequel un corpus de stimuli resynthétisés a été manipulé quant à l'alignement de la configuration en montée descente des cibles L1, H et L2 ainsi qu'à la durée de la structure syllabique. Nous avons trouvé que l'alignement avancé/retardé de l'accent nucléaire et de l'accent de phrase L- est exploité par les auditeurs napolitains comme un indice pour l'identification des affirmations/questions, et que le déplacement de l'interprétation du contour vers la modalité "question" est plus important en présence d'une syllabe ouverte qu'en présence d'une syllabe fermée. Cela est cohérent avec le fait que la cible acoustique H tend à être réalisée déjà dans la portion finale de la voyelle dans les questions ayant une syllabe accentuée ouverte. Au contraire, dans les questions ayant une syllabe fermée le pic accentuel est réalisé à la fin de la voyelle ou dans la coda consonantique suivante. L'effet de structure syllabique peut être interprété en termes de différences de durée segmentale. En fait, lorsque la durée de la voyelle accentuée et de la coda consonantique suivante étaient ambiguës entre celle typique des syllabes fermées/ouvertes, l'effet de structure syllabique sur l'identification intonative n'était plus significatif. De plus, les auditeurs napolitains semblent exploiter la variabilité systématique d'alignement pour l'identification des mots en contraste de structure syllabique et durée segmentale (tels que *nonno/nono*). Cet effet est significatif uniquement lorsque l'indice le plus robuste de gémination consonantique (la durée) est ambiguë. Dans ce cas, le déplacement du pic accentuel vers la fin de la syllabe consonantique provoque aussi un déplacement de l'interprétation des stimuli vers *nonno*. Cela semble en fait cohérent avec l'alignement retardé typique des mots ayant une structure syllabique fermée. Les résultats obtenus pour l'identification lexicale sont aussi en ligne avec nos précédents travaux (D'Imperio *et al.*, 2007).

Les affirmations et les questions du napolitain sont aussi caractérisées par une différence dans la région de f_0 entre l'accent prénucléaire et l'accent nucléaire. Cette différence

a été définie en termes de variations de forme et de pente interaccentuelle (voir aussi le par. 1.2 pour la définition de “forme du contour” adoptée dans ce travail). Dans les affirmations, la f_0 descend rapidement de l’accent prénucléaire jusqu’à la région autour de la fin du premier mot prosodique, où un coude de f_0 est clairement visible dans le tracé acoustique. Dans les questions, les valeurs de f_0 après le pic prénucléaire restent élevées, et la f_0 a une forme convexe. Les deux chapitres expérimentaux suivants ont été donc consacrés à l’étude acoustique (chap. 4) et perceptif (chap. 5) de cette différence de forme et de pente de f_0 dans la région prénucléaire.

La première expérience de production décrite dans le chap. 4 nous a révélé que cette différence est “authentique”, car elle n’est pas provoquée par des variations possibles de vitesse d’élocution ou par des ajustements mélodiques et d’alignement des cibles tonales voisines. Les résultats de la modélisation (linéaire et non-linéaire) de la courbe interaccentuelle nous ont aussi montré que dans les deux modalités le point d’inflexion de la courbe modélisée est situé autour de la fin du premier mot prosodique dans les deux modalités intonatives. Nous avons émis l’hypothèse que la variabilité de forme et de pente est provoquée par l’insertion d’un ton, différemment spécifié dans les questions et dans les affirmations. Ce ton marquerait la fin d’un constituant prosodique de niveau bas, que nous avons défini provisoirement comme le syntagme accentuel. En particulier, nous avons suggéré que les questions sont caractérisées par l’insertion d’un ton haut H_{AP} et les affirmations par l’insertion d’un ton bas L_{AP} . La deuxième expérience de production décrite dans le même chapitre semble soutenir cette hypothèse, car l’insertion de ce ton signifierait la fin d’un constituant prosodique de rang inférieur au syntagme intonatif et au syntagme intermédiaire (c’est-à-dire, les seuls niveaux de *phrasing* attestés jusqu’à maintenant en italien, Grice *et al.* (2005)). De plus, la présence d’une frontière de syntagme accentuel serait signalée par un degré d’allongement de la durée segmentale inférieure à celui trouvé à la fin du syntagme intermédiaire. L’allongement final se produit uniquement sur la syllabe tonique en position finale du constituant prosodique, mais non sur la syllabe atone finale. Une possible explication, inspirée de travaux récents dans le cadre de la Phonologie Articulatoire (Riggs & Byrd, 2007), est que l’accent mélodique “attire” l’allongement final sur la dernière syllabe tonique accentuée. Cette hypothèse sera testée dans nos recherches

futures.

Nous avons supposé que la variabilité systématique de la région interaccentuelle peut être utilisée pour renforcer le contraste entre questions et affirmations, dont l'indice principal est l'alignement de l'accent nucléaire. Si cela est vrai, nous nous attendions à ce que la différence de spécification tonale entre L_{AP} et H_{AP} puisse être exploitée par les auditeurs napolitains pour l'identification du contraste de sens intonatif. Cette hypothèse a été testée dans le chap. 5 par deux expériences de perception. Dans un test d'identification, nous avons utilisé un corpus d'énoncés naturels prononcés comme questions/affirmations et déjà employés pour d'autres analyses acoustiques (chapitres 2 et 4). Le coupage de chaque énoncé à la fin du premier et du deuxième mot prosodique nous a permis d'"isoler" les effets de la région prénucléaire sur l'identification intonative. Nous avons trouvé que l'identification des affirmations et des questions est au dessous du niveau de hasard déjà lorsque le premier mot prosodique de l'énoncé (contenant l'accent prénucléaire (LH*)) est le seul disponible aux auditeurs. Cela pourrait être dû à des différences phonétiques entre les affirmations et les questions (telle que la vitesse d'élocution), qui ont été exploitées comme indices secondaires de modalité intonative. De plus, lorsque l'information du ton L_{AP} est disponible, l'identification des affirmations augmente de façon significative. Le ton H_{AP} est un indice beaucoup plus ambigu pour les questions, de sorte qu'aucune différence significative n'a été trouvée dans l'identification des stimuli contenant uniquement l'accent prénucléaire et ceux contenant aussi le ton H_{AP} . Les propriétés sémantiques des tons L_{AP} et H_{AP} ont été mieux explorées par le *semantic differential task*. Dans ce test, nous avons utilisé des stimuli resynthésés, dans lesquels la hauteur mélodique du ton de syntagme accentuel a été modifiée. Les résultats indiquent que le ton L_{AP} véhicule l'idée de "certitude", qui est associée à la modalité affirmative selon l'interprétation informationnelle du Code de Fréquence (Gussenhoven, 2002). Cependant, le ton H_{AP} semble véhiculer uniquement la signification paralinguistique d'"engagement émotionnel". Globalement, le test d'identification et le *semantic differential task* semblent suggérer que c'est uniquement le ton L_{AP} qui contribue au sens intonatif. Cependant, une possible hypothèse qui pourrait expliquer cette "discrepance" entre les tons L_{AP}/H_{AP} est que la construction des stimuli pour le *semantic differential task* n'a pas tenu compte de propriétés dynamiques

trouvés dans la région interaccentuelle des questions, qui auraient pu être une information importante pour la perception du ton H_{AP} . Cette hypothèse sera également testée dans nos recherches futures.

6.2 Apports méthodologiques

Au cours de cette thèse, nous avons utilisé et comparé plusieurs méthodologies pour mieux évaluer comment la variabilité phonétique influence la production et la perception de l'intonation.

L'existence d'effets segmentaux sur la f_0 nous a ramené d'abord face au problème de détermination de la localisation exacte des cibles tonales dans le tracé acoustique (chapt. 2). Dans les études sur l'alignement tonal, les cibles tonales sont souvent localisées manuellement comme des points minimales (pour le L) ou maximales (pour le H) de la courbe de f_0 (Arvaniti *et al.*, 1998; Ladd *et al.*, 1999; Prieto *et al.*, 1995, *inter alia*). En présence de perturbations segmentales, les annotateurs suivent des critères subjectives pour la localisation manuelle des cibles. Par exemple, on peut localiser la cible L quelques millisecondes avant ou après la perturbations, ou même négliger les effets microprosodiques. D'autres travaux ont utilisés de procédures automatiques pour la modélisation de la courbe (Hirst *et al.*, 2000; D'Imperio, 2000; Welby, 2003; Xu & Wang, 2001; Prom-on & Xu, en préparation). Cependant, il est encore difficile de déterminer quelle méthode de localisation des cibles est la "meilleure". A notre connaissance, la seule étude qui s'est interrogée à ce sujet est celle de del Giudice *et al.* (2007). del Giudice *et al.* (2007) ont montré que la détection manuelle est une méthode valide pour la localisation de la cible tonale L (notamment beaucoup plus difficile à détecter que le H) car, lorsque les annotateurs suivent un protocole commune, l'étiquetage manuel est très cohérent.

Dans notre thèse, nous avons comparé la méthode manuelle, le LSF (D'Imperio, 2000) et MOMEL (Hirst *et al.*, 2000) pour la localisation du L des accents nucléaires L+H* et L*+H du napolitain (chap. 2). Nos résultats suggèrent que la "cohérence" est un critère nécessaire mais pas suffisant pour évaluer la qualité de la procédure adoptée, et que le choix méthodologique a des conséquences sur l'interprétation phonologique du contour.

La localisation manuelle a révélé une forte cohérence mais elle a effacé une information phonologique : celle de la différence d’alignement de la cible L entre les deux accents. L’importance de cette différence pour l’identification du contraste entre questions et affirmations a été soulignée par D’Imperio (2000) (voir aussi dans l’introduction, le par. 1.3.1). La détection de MOMEL est plus variable ; l’algorithme a cependant capturé la différence d’alignement entre les deux accents. Les résultats de la détection du LSF sont cohérents, et la différence phonologique d’alignement du L est capturée.

Il faut noter que, bien que les méthodes algorithmiques nous semblent plus fiables pour apprécier le contraste d’alignement du napolitain, nous ne pensons pas que le LSF est la méthode la meilleure “tout court”. En premier lieu, l’exacte localisation des cibles effectuée par le LSF dépend de la taille de la fenêtre d’analyse utilisée (voir aussi del Giudice *et al.* (2007)). Par conséquent, il est crucial de spécifier de façon explicite la région dans laquelle l’algorithme est ajusté. De plus, les perturbations microprosodiques peuvent être une source d’erreurs même pour les procédures automatiques (del Giudice *et al.*, 2007). Dans notre étude, nous avons corrigé manuellement les cas non-ambigus d’erreurs effectuées par le LSF, tandis que lorsque la détection manuelle et celle du LSF étaient également plausibles, la détection du LSF a été préférée.

De plus, bien que les points d’inflexion soient la manifestation la plus facilement mesurable des cibles tonales, l’équation entre cibles et points d’inflexion n’est pas appropriée pour décrire la manifestation acoustique du ton de syntagme accentuel du napolitain, et cela même en absence de perturbations segmentales. En particulier, dans le chapitre 4, nous avons remarqué que le ton H_{AP} n’est identifiable avec aucun pic maximal local, la région de f_0 après l’accent prénucléaire offrant plutôt à une forme convexe de la courbe. De plus, l’environnement tonal influence autant la réalisation des deux tons H_{AP} et L_{AP} , que le contraste de spécification tonale peut se réduire à une variabilité de l’inclinaison de la pente interaccentuelle de f_0 . Ni la méthode manuelle, ni le LSF ni MOMEL n’ont réussi à capturer cette différence subtile entre questions et affirmations. Une possible explication est que les trois méthodes impliquent que la cible tonale coïncide avec un point de changement soudain de direction de la courbe de f_0 , ce qui n’est pas approprié surtout pour décrire la forme et la pente de f_0 dans questions (voir aussi le chapitre 4, par. 4.2

pour plus de détails sur cette hypothèse). Nous avons donc décidé d'examiner de plus près les propriétés dynamiques de la région interaccentuelle par l'utilisation combinée de techniques de modélisation statistique linéaires et non-linéaires. En particulier, nous avons pu quantifier la différence entre les deux modalités intonatives par l'utilisation de la régression logistique, de la régression linéaire par morceaux et de la régression linéaire simple.

Globalement, notre étude suggère que l'utilisation combinée de plusieurs méthodes (le LSF et la méthode manuelle ; la régression logistique et celle linéaire) peut nous aider dans la détermination l'exacte localisation des cibles. Le développement de procédures algorithmiques sophistiquées d'un point de vue mathématique peut être également un instrument méthodologique d'importance cruciale pour la détection des cibles, surtout lorsqu'elles ne sont pas réalisées comme des points d'inflexion locales. De plus, plusieurs travaux nous suggèrent l'existence d'une interaction entre les perturbations segmentales et l'intonation, en production et en perception. En particulier, Kohler (1987), Niebuhr (à paraître) et Silverman (1987) ont montré que que la micro- et macroprosodie interagissent dans l'identification intonative et segmentale, et que l'influence des effets microprosodiques dépend de la langue considérée. De plus, Silverman (1987) a montré que l'intégration de l'information microprosodique peut améliorer la synthèse de f_0 en anglais (Silverman, 1987). De façon similaire, il serait intéressant de vérifier si l'intégration de l'information microprosodique puisse aussi améliorer la modélisation de la f_0 effectuée par les procédures algorithmes déjà existantes (telles que MOMEL ou le LSF).

Nous voulons aussi mentionner quelques contributions que notre travail pourrait apporter au traitement statistique des données. Comme on l'a remarqué au cours de cette thèse (et surtout dans les chapitre 2), la réalisation des cibles tonales est fortement influencée par la variabilité segmentale, au point que la significativité d'autres facteurs expérimentaux peut dépendre de l'interaction avec l'identité segmentale de l'attaque consonantique. Si, dans le chapitre 2, on avait ignoré d'évaluer l'exacte contribution de la classe consonantique sur l'alignement du L, les résultats des ANOVA ou des analyses de régression pour la manipulation de la modalité intonative auraient pu être biaisés. En fait, nous avons trouvé que la grandeur des effets de modalité intonative sur le L nucléaire

dépende de la classe consonantique de l'attaque consonantique. Par conséquent, bien que, dans les expériences de production, l'utilisation de matériel contenant plusieurs types de segments puisse conduire à une meilleure compréhension de l'interaction entre plan segmentale et suprasegmentale, nous pensons qu'il faut séparer les classes consonantiques dans l'analyse statistique (au lieu de regrouper les catégories segmentales), car l'effet de mode d'articulation influence la localisation des cibles tonales ainsi que leur détection.

De plus, nous pensons que l'utilisation des analyses de régression puisse nous restituer des informations plus précises que les ANOVA omnibus sur les effets d'un facteur catégoriel, lorsqu'il est composé de plus de deux niveaux. Comme on se souviendra, dans le chapitre 2, le facteur "classe consonantique" était composé de trois niveaux (nasales/liquides/obstruantes). En comparant les modèles de régression et les ANOVA omnibus des modèles mêmes, nous avons trouvé que souvent des contradictions apparaissent entre significativité des coefficients et analyses de variance. Cela est dû au fait que les ANOVA omnibus nous restituent une information sur les effets globales d'un facteur, tandis que les analyses de régression nous restituent des informations sur les contrastes entre les niveaux d'un même facteur. Par conséquent, c'était uniquement par les modèles de régression que les différences entre les trois classes consonantiques ont pu mieux ressortir. L'analyse de régression a été utilisée aussi dans les chapitres 3, 4 et 5 pour modéliser l'effet des facteurs continus ou pour expliquer les contrastes entre trois ou plus niveaux d'un facteur catégoriel.

Finalement, un autre problème que nous avons rencontré dans notre travail est comment traiter la variabilité inter-locuteurs. Deux méthodologies statistiques ont été proposées, dont le choix dépend essentiellement du nombre des participants à l'expérience et des objectifs de l'expérimentateur (voir aussi le par. 2.2.5 dans le chapitre 2). Lorsque le nombre des sujets était petit (comme dans le chap. 2, où nous avons analysé les productions de deux seuls locuteurs), les analyses statistiques ont été effectuées séparément par sujet. Cela nous a permis aussi de quantifier les variations inter-locuteurs. Lorsque le nombre de sujets était plus élevé (comme dans les tests de perception dans les chapitres 3 et 5, ou pour les analyses de durée dans le chapitre 4), les sujets ont été considérés comme échantillonnés de façon aléatoire d'une population plus large. Par conséquent, nous avons

préféré utiliser des modèles à effets mixtes, où le facteur “sujets” a été introduit comme facteur aléatoire. Cela nous a permis d’évaluer les effets de facteurs fixes, tout en contrôlant la variance apportée par les sujets dans le modèle. Nous pensons cependant que le traitement statistique de la variabilité des sujets est une tâche très difficile, et que même ces méthodologies présentent des problèmes. En premier lieu, au niveau statistique, il est encore difficile d’établir ce qui signifie “un nombre petit/élevé” des sujets. Quand peut-on considérer un facteur fixe ? Quand le nombre des sujets est-il si élevé que l’on peut le considérer comme aléatoire ? De plus, même lorsque le nombre de sujets est assez élevé (disons, plus de cinq participants), il peut être encore très utile d’évaluer exactement les effets de la variabilité inter-locuteurs. En fait, cela peut nous restituer des informations importants sur les stratégies adoptées par chaque locuteur pour la production ou perception de l’intonation. Cependant, dans ce cas, considérer le facteur “sujet” en fixe peut rendre l’analyse statistique plus délicate, car l’introduction d’un facteur fixe à nombreux niveaux influence aussi le nombre de degré de liberté du modèle. De plus, l’interprétation globale des effets expérimentaux devient, elle aussi, plus difficile.

6.3 Variabilité, représentation phonétique et phonologique

Dans les études sur la parole, le statut de la variabilité phonétique est très problématique. Sur le plan segmental, elle a été traditionnellement regardée comme le résultat de la simple actualisation physique des catégories phonologiques, de sorte que les chercheurs ont souvent fait coïncider la recherche des éléments “pertinents” au sens linguistique avec la recherche des “invariants”(Lieberman *et al.*, 1967; Stevens & Blumstein, 1978, *inter alia*). Selon une autre perspective, puisque la variabilité est inhérente à la parole, un état “au dehors” de la variabilité est une pure abstraction. Au contraire, le “*lack-of-invariance problem*” ne constituerait pas un problème car “*[i]t is precisely the variability in the signal which permits listeners to understand speech in a variety of contexts, and spoken by a variety of speakers. Instead of searching invariance in the signal, we think it makes more sense to try to understand how it is that listeners deal with the variability which*

is there”(Elman & McClelland, 1986, pg. 360). Récemment, l’importance de la variabilité phonétique a en fait été souligné par un nombre de plus en plus croissant de travaux, qui montrent que les variations systématiques du détail acoustique peuvent être exploitées pour véhiculer des informations sur la structure grammaticale de la langue, même lorsqu’elles sont distribués globalement dans le signal (voir Hawkins & Smith, 2001; Hawkins, 2003 *inter alia*).

Le problème de définir le statut de la variabilité concerne aussi un autre plan beaucoup moins étudié, celui de l’intonation. Nous avons vu dans l’introduction de cette thèse que la littérature sur l’alignement tonal nous offre des résultats assez controversés à ce sujet. D’un côté, il a été souvent affirmé que la réalisation des cibles tonales obéit à des principes de synchronie et de stabilité (Arvaniti *et al.*, 1998, Xu, 1998, Ladd *et al.* (2000), Ishihara (2003), *inter alia*). Par ailleurs, l’“invariance” d’alignement et de valeur mélodique des cibles H et L a été donc utilisée, dans des études, comme preuve empirique du fait que les unités d’analyse de l’intonation sont les tons statiques (Ladd *et al.*, 1999). De l’autre côté, plusieurs études ont montré que les cibles tonales (et surtout la réalisation de la cible H) peuvent être soumises à plusieurs facteurs de variabilité, tels que l’identité segmentale, la structure syllabique ou la vitesse d’élocution. Cette variabilité acoustique a été expliquée de façon différente. Par exemple, la variabilité systématique due à des contraintes structurels a été intégrée dans des modèles soutenant l’ancrage segmental (tel que la *revised structural explanation* de Ladd *et al.*, 2000 ; voir aussi l’introduction de cette thèse). Dans d’autres cas, la variabilité acoustique a été expliquée par l’admission que d’une fenêtre temporelle assez large dans laquelle la cible peut être ancrée (voir la notion de *anchorage* chez Welby & Loevenbruck, 2005). Selon une autre interprétation, l’alignement est un phénomène simplement quantitatif, sans aucune conséquence phonologique (Atterer & Ladd, 2004; Ladd, 2008).

En italien napolitain, cependant, la variabilité d’alignement a une valeur phonologique, dans la mesure où le déplacement de l’accent nucléaire LH provoque un changement de la perception des affirmations à focalisation étroite aux questions oui/non. Cela signifie que les auditeurs napolitains “connaissent” comment le continuum d’alignement acoustique est partitionnée dans leur variété, et exploitent cette information dans le *mapping*

entre cibles acoustiques et cibles perceptives. Cette position avait été déjà été soutenue par D’Imperio (2000), qui avait montré que le contraste d’alignement est catégoriel, et qu’il est interprétable au niveau phonologique comme dû à une différence d’association (L+H* vs. L*+H). Dans notre thèse, nous suggérons aussi que les auditeurs sont capables de faire correspondre un input variable avec une catégorie phonologique spécifique tout en tenant compte des effets contextuels sur l’input même. Prenons les résultats des expériences décrites dans le chapitre 3. En premier lieu, nous avons trouvé que les auditeurs napolitains ne sont pas seulement capables de “partitionner” le continuum d’alignement pour l’identification des affirmations et des questions, mais ils peuvent aussi interpréter ces détails d’alignement en relation avec la structure segmentale (ie., le contraste entre syllabes fermées et ouvertes). Donc, l’information d’alignement et celle de structure syllabique sont élaborées simultanément pour la perception des catégories intonatives. De plus, l’information d’alignement est employée pour l’identification des paires lexicales minimales en contraste de structure syllabique et de longueur consonantique. Dans les expériences décrites dans le chapitre 3, de tels effets sont très “graduels” et ils sont visibles uniquement lorsque la durée segmentale est ambiguë. Si nous pensons aux catégories linguistiques comme “centres de densité” (Post *et al.*, 2007) , nous pouvons imaginer que, pour les contrastes de gémation consonantique, l’alignement tonale ne joue aucun rôle linguistique au centre de la catégorie (car la durée est l’indice principale de la gémation). Cependant, lorsqu’on s’éloigne du centre (par exemple, lorsque la durée devient de plus en plus ambiguë), le détail d’alignement commence à être exploité par les auditeurs pour l’identification lexicale. En d’autres termes, à la périphérie des catégories linguistiques, les détails contextuels peuvent influencer la position de la frontière catégorielle. De plus, il faut noter que l’existence d’une interaction entre la perception de l’alignement et le plan segmental est contraire aux modèles, tels que celui de Levelt (1989, 1993), dans lesquels l’élaboration de la parole est le résultat de la contribution indépendante de différentes sources d’information, et la prosodie est générée uniquement après le complètement de l’accès lexical et des procédures de construction syntactique. De plus, nos résultats sont en contraste avec les modèles de perception dans lesquels la prosodie n’est pas prise en compte pour l’identification lexicale (Gaskell *et al.*, 1995; Norris *et al.*, 2000; Stevens,

2002).

Une autre découverte importante est que le contraste de sens intonatif est renforcé par plusieurs indices tonales, distribués dans plusieurs portions de l'énoncé. En particulier, en napolitain, le contraste entre questions et affirmations est signalé pas seulement par l'accent nucléaire, mais aussi par une variabilité systématique de la forme et de la pente de f_0 dans la région prénucléaire. Les résultats préliminaires des tests de perception (chap. 5) suggèrent que cette variabilité est exploitée même lorsque l'information nucléaire n'est pas disponible. Cela évoque les résultats d'études récentes sur le plan segmental, qui ont montré l'existence d'indices acoustiques distribués sur des portions longues du signal utilisés pour la perception des contrastes phonologiques. Par exemple, on se souviendra du travail de Warren & Marslen-Wilson (1987) sur le rôle de la coarticulation anticipatoire sur l'identification des voyelles, ou encore du travail de Nguyen & Hawkins (1999) sur les effets du voisement d'une coda syllabique dans l'identification lexicale (voir Hawkins & Smith (2001) pour une modélisation de ces effets). De façon similaire, en napolitain, la relation phonétique à distance entre la région nucléaire et celle prénucléaire permettrait aux locuteurs de renforcer le contraste de sens entre questions et affirmations, pour lequel aucun indice morpho-syntaxique n'est utilisé et dont l'indice le plus important est la subtile variation d'alignement tonal à l'intérieur de la syllabe nucléaire. Il faut noter que nous avons émis l'hypothèse que la variabilité de pente de la chute prénucléaire est due au contraste phonologique entre deux tons, L_{AP} et H_{AP} . Cependant, l'expérience I décrite dans le chapitre 5 nous donne une information supplémentaire, car nous avons trouvé que les auditeurs napolitains sont capables de distinguer une question d'une affirmation très tôt, c'est-à-dire même lorsque la portion contenant uniquement l'accent prénucléaire montant est disponible. Il est encore difficile d'établir quel sont les indices exploités par les auditeurs dans ce cas (les différences d'implémentation phonétique de l'accent prénucléaire, la vitesse d'élocution, les changements spectraux ou d'intensité, etc.). Ce résultat suggère donc l'existence de détails phonétiques qui peuvent être utilisés comme corrélats perceptifs du contraste intonatif, qui aident à renforcer les contrastes phonologiques entre H_{AP} et L_{AP} (dans la région prénucléaire) et entre L^*+H et $L+H^*$ (dans la région nucléaire) et qui sont distribués, à nouveau, sur des sections très longues du signal.

Une autre observation mérite d'être faite. Dans notre thèse, nous avons trouvé une variabilité inter-locuteurs assez important. Par exemple, dans le chapitre 2, nous avons trouvé que la réalisation de l'accent nucléaire interagit avec l'identité de la consonne en attaque consonantique et la modalité intonative de façon partiellement différente selon le locuteur ; dans le chapitre 4, nous avons trouvé un certain degré de variabilité inter-locuteurs dans l'implementation de l'accent prénucléaire montant (LH*) et de la chute prénucléaire. Comment cette variabilité peut-elle être expliquée ? A ce stade de la réflexion, nous émettons uniquement des conjectures. Si nous adoptons la notion de "prototype" même pour les catégories intonatives, nous pouvons penser que plusieurs indices, même partiellement différents selon le locuteur, peuvent être utilisés pour signaler des contrastes intonatifs. Donc, bien qu'en napolitain l'alignement de la montée soit très important pour la distinction entre les catégories L*+H et L+H*, les auditeurs peuvent optionnellement choisir d'utiliser d'autres indices secondaires, dans le domaine de la f_0 ou encore dans d'autres domaines prosodiques. Par exemple, le locuteur peut décider de retarder uniquement le pic nucléaire, et au même temps de hausser sa valeur mélodique de sorte que le locuteur puisse percevoir sa phrase comme une "question". Dans le même temps, le locuteur devrait avoir une capacité de computation adaptative aux stratégies différentes des locuteurs, qui lui permettrait d'employer simultanément plusieurs indices pour l'interprétation linguistique. En d'autres termes, la variabilité inter-locuteur pourrait être due au fait que les locuteurs utilisent des stratégies différentes pour obtenir le même but perceptif. Cette hypothèse sera mieux explorée dans nos recherches futures.

Pour résumer, nous pensons que la représentation phonétique des cibles tonales est le résultat de l'interaction de plusieurs facteurs de variabilité phonétique et phonologique et que, lorsque cette variabilité est systématique, elle peut aider les auditeurs dans la perception des contrastes phonologiques. De plus, l'existence d'indices tonales distribués globalement dans l'énoncé suggère que la région nucléaire et prénucléaire coopèrent dans la détermination du contraste de sens intonatif.

6.4 Conséquences sur le sens intonatif

Nous avons commencé cette thèse par l'exploration du concept de la compositionnalité intonative. Dans la Théorie Métrique-Autosegmentale (Pierrehumbert, 1980), la compositionnalité des contours intonatifs à deux conséquences : sur la représentation formelle de l'intonation et sur sa représentation fonctionnelle. Les deux représentations sont liées de façon très étroite. D'un côté, la machine à états finis du modèle de Pierrehumbert nous informe sur comment les catégories tonales se combinent les unes avec les autres. Le contour intonatif est interprété comme une séquence de tons (correspondant aux états de la machine) dont la combinaison est libre. En fait, la machine nous ne donne pas d'information sur la probabilité avec laquelle les tons co-occurrent : virtuellement, toutes les combinaisons sont possibles. Cette liberté de choix est la conséquence du fait que chaque ton contribue au sens du contour de façon indépendante, le domaine d'interprétation des tons correspondant à son domaine phonologique (Pierrehumbert & Hirschberg, 1990). En particulier, selon les deux auteurs, les accents mélodiques véhiculent des informations sur les mots auxquels ils sont associés, et les tons périphériques (accent de phrase, ton de frontière) nous indiquent comment interpréter un constituant prosodique en relation aux autres du même rang hiérarchique. Aucun intérêt particulier n'est accordé, cependant, à la contribution spécifique de la région prénucléaire au sens intonatif.

En italien napolitain, la combinaison des tons dans un contour intonatif semble être conditionnée par des contraintes de nature fonctionnelle. En production, la réalisation du contraste intonative entre questions et affirmations est véhiculée par deux types d'informations tonales : une liée à l'accent nucléaire (le contraste entre $L+H^*$ et L^*+H), et l'autre liée au ton de AP qui le précède (le contraste entre H_{AP} et L_{AP}). Les accents mélodiques et les tons de syntagme accentuels sont combinés de façon très systématique dans le contour : l'accent mélodique $L+H^*$ est toujours précédé par le ton L_{AP} dans les affirmations à focalisation étroite, et l'accent mélodique L^*+H est précédé par le ton H_{AP} dans les questions oui/non. De plus, les résultats des expériences de perception suggèrent que le contraste de spécification tonale entre les tons H_{AP} et L_{AP} peut être utilisé aussi pour l'identification du contraste intonatif.

Cela suggère trois possibles conséquences pour la représentation du sens intonatif. En premier lieu, la partie “prénucléaire” du contour intonatif semble jouer un rôle important pour la transmission du sens. En fait, en italien napolitain, les propriétés sémantiques du ton de syntagme accentuel sont cohérentes avec celle de l’accent nucléaire dans la réalisation du contraste entre questions et affirmations. Nous savons que le “nucleus” (c’est-à-dire, la composition de l’accent nucléaire, et des tons périphériques suivants) est généralement considéré comme le cœur sémantique des énoncés. Cependant, si la région prénucléaire est elle aussi importante pour le sens intonatif, il faudrait vérifier quel est son poids informationnel, et comment il interagit avec celui du nucleus. De plus, nos données suggèrent que le choix des tons dans un contour implique un certain degré de pré-planification du contour intonatif. En fait, si la construction du contour intonatif procédait strictement de gauche à droite, nous aurions dû attendre que la région prénucléaire fusse la même indépendamment de la modalité intonative. Finalement, les restrictions de cooccurrence des tons suggèrent qu’en napolitain la construction du contour intonatif n’est pas expliquée par une interprétation stricte de l’approche compositionnelle.

Nous nous sommes conscients du fait que dans cette étude, nous nous sommes concentrés sur un seul contraste de sens, celui entre les questions oui/non et les affirmations à focalisation étroite. Par conséquent, nous ne pouvons pas prédire par ce travail seul si et comment les tons se combinent pour la construction d’autres contours intonatifs en italien napolitain. Par exemple, nous ne pouvons pas exclure que le ton L_{AP} et l’accent mélodique L^*+H se combinent pour la construction d’un type de contour interrogatif présentant une valeur pragmatique différente que la question oui/non. Nous pensons donc qu’une étude plus systématique doit être effectuée pour mieux évaluer l’existence de contraintes de co-occurrences des tons pour l’italien. De plus, nous pensons que l’exploration de ces problématiques présuppose l’abandon de l’équation forme-modalité intonative, et que la construction d’une théorie du sens intonatif de l’italien doive plutôt prendre en compte l’exacte contribution des catégories tonales dans le discours.

Bibliographie

- Albano Leoni, F. 1998. *Manuale di Fonetica*. Roma : Il Carroccio.
- Ambrazaitis, G. 2005. Between fall and fall-rise. Substance-function relations in German phrasefinal intonation contours. *Phonetica*, **62**, 196–214.
- Anderson, S. 1978. Tone features. *Pages 72–106 of* : Fromkin, V. (ed), *Tone : a linguistic survey*. New York, USA : Academic Press.
- André, C., Ghio, A., Cavé, C., & Teston, B. 2007. *PERCEVAL : PERCEPTION EVALUATION Auditive and Visuelle (Version 3.0.4) [Computer program]*. <http://www.lpl.univ-aix.fr/~lpl/dev/perceval>.
- Arvaniti, A., & Ladd, D. R. 1995. Tonal alignment and the representation of accentual targets. *Page 220–223 of* : *Proceedings of the 13th international congress of phonetic sciences*, vol. 4.
- Arvaniti, A., Ladd, D. R., & Mennen, I. 1998. Stability of tonal alignment : the case of Greek prenuclear accents. *Journal of Phonetics*, **26**(1), 3–25.
- Arvaniti, A., Ladd, D. R., & Mennen, I. 2000. What is a starred tone ? Evidence from Greek. *Pages 132–151 of* : Broe, M., & Pierrehumbert, J. B. (eds), *Papers in Laboratory Phonology V : Acquisition and the Lexicon*. Cambridge : Cambridge University Press.

- Atterer, A., & Ladd, D. R. 2004. On the phonetics and phonology of “segmental anchoring” of F0 : evidence from German. *Journal of Phonetics*, **32**, 177–197.
- Baayen, R. H. à paraître. *Analyzing Linguistic Data : A practical introduction to Statistics*. Cambridge : Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. à paraître. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **special issue on Emerging Data Analysis Techniques**.
- Beckman, M. E. 1986. *Stress and non-stress accent*. Dordrecht : Foris.
- Beckman, M. E., & Edwards, J. 1990. Lengthenings, shortenings, and the nature of prosodic constituency. *Page 152–178 of* : Kingston, J., & Beckman, M. E. (eds), *Papers in Laboratory Phonology I : Between the Grammar and the Physics of Speech*. Cambridge : Cambridge University Press.
- Beckman, M. E., & Elam, G. Ayers. 1994. *Guidelines for ToBI labelling*. <http://www.tobihome.org>.
- Beckman, M. E., & Pierrehumbert, J. B. 1986. Intonational Structure in Japanese and English. *Phonology Yearbook*, **3**, 255–310.
- Bertinetto, P.M. 1980. The perception of stress by Italian speakers. *Journal of Phonetics*, **8**, 385–395.
- Blumstein, S. E. 1986. On Acoustic invariance in speech. *Pages 178–193 of* : Perkell, J., & Klatt, D. (eds), *Invariance and Variability in Speech Processes*. Hillsdale : LEA.
- Boersma, P., & Weenink, D. 2007. *Praat : doing phonetics by computer (Version 4.5.18) [Computer program]*. <http://praat.org/>.
- Bolinger, D. 1951. Intonation : levels versus configurations. *Word*, **7**, 199–210.
- Bolinger, D. 1958. A theory of pitch accent in English. *Word*, **14**, 109–149.

- Bolinger, D. 1986. *Intonation and its Parts*. Palo Alto, California : Stanford University Press.
- Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Ph.D. thesis, Lund University, Lund, Sweden.
- Bruce, G. 1983. Accentuation and timing in Swedish. *Folia Linguistica*, **8**, 221–238.
- Byrd, D., & Saltzman, E. 1998. Intra-gestural dynamics of multiple phrasal boundaries. *Journal of Phonetics*, **26**, 173–199.
- Byrd, D., & Saltzman, E. 2003. The elastic phase : Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, **31**(2), 149–180.
- Campione, E., Hirst, D., & Véronis, J. 1997. *Automatic Stylistic and Symbolic Coding of F0 : Implementations of the INTSINT Model*. <http://sites.univ-provence.fr>.
- Carpentier, F., & Moulines, E. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, **9**(4), 453–457.
- Caspers, J. 1998. Who's the next ? The melodic marking of question vs. continuation in Dutch. *Language and Speech*, **41**, 375–398.
- Caspers, J., & van Heuven, V. J. 1993. Effects of time pressure on the phonetic realisation of the Dutch accent-lending pitch rise and fall. *Phonetica*, **50**, 161–171.
- Chomsky, N., & Halle, M. 1968. *The Sound Pattern of English*. New York : Harper and Row.
- Clark, H. H. 1973. The language-as-fixed-effect fallacy : A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **special issue on Emerging Data Analysis Techniques**(12), 335–359.
- Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge : Cambridge University Press.

- Dainora, A. 2002. Does intonational meaning come from tones or tunes ? Evidence against a compositional approach. *Pages 235–238 of* : B.Bel, & Marlien, I. (eds), *Proceedings of Speech Prosody 2002*.
- Dainora, A. 2006. Modeling intonation in English : a probabilistic approach to phonological competence. *Pages 107–132 of* : Goldstein, L. M., Whalen, D., & Best, C. T. (eds), *Papers in Laboratory Phonology VIII : Varieties of Phonological Competence*. Berlin : Mouton de Gruyter.
- del Giudice, A., Shosted, R.K., Davidson, C., Salihie, M., & Arvaniti, A. 2007. Comparing methods for locating pitch “elbows”. *Pages 1117–1120 of* : *Proceedings of the 16th International Congress of Phonetic Sciences*, vol. 2.
- Delattre, P. 1966. Les dix intonations de base du français. *The French Review*, **40**(1), 1–14.
- Di Cristo, A. 1985. *De la microprosodie à l'intonosyntaxe*. Aix-en-Provence, France : American Institute of Physics.
- Di Cristo, A. 2004. La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonctions. *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, **23**, 67–211.
- Di Cristo, A., & Hirst, D. J. 1993. Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix*, **15**, 9–24.
- D'Imperio, M. 1997. Narrow focus and focal accent in the Neapolitan variety of Italian. *Page 87–90 of* : Botinis, A., Kouroupetroglou, G., & Carayiannis, G. (eds), *Intonation : Theory, Models and Applications. Proceedings of the ESCA Workshop*.
- D'Imperio, M. 1999. Tonal structure and pitch targets in Italian focus constituents. *Page 1757–1760 of* : *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 3.

- D'Imperio, M. 2000. *The Role of Perception in Defining Tonal Targets and their Alignment*. Ph.D. thesis, The Ohio State University, Ohio State, USA.
- D'Imperio, M. 2001. Focus and tonal structure in Neapolitan Italian. *Speech Communication*, **33**(4), 339–356.
- D'Imperio, M. 2002a. Italian intonation : An overview and some questions. *Probus*, **14**, 37–69.
- D'Imperio, M. 2002b. Tonal structure and pitch targets in Italian focus constituents. *Catalan Journal of Linguistics*, **2**, 53–63.
- D'Imperio, M. 2005. *Aspects Prosodiques de l'italien de Naples : Proéminence, Catégories Accentuelles et Alignement Tonal*. Mémoire d'habilitation à diriger des recherches.
- D'Imperio, M., & Fivela, B. Gili. 2003. How many levels of phrasing? Evidence from two varieties of Italian. *Pages 38–57 of : Local, J., Ogden, R., & Temple, R. (eds), Papers in Laboratory Phonology VI : Phonetic Interpretation*. Cambridge : Cambridge University Press.
- D'Imperio, M., & House, D. 1997. Perception of questions and statements in Neapolitan Italian. *Page 251–254 of : Kokkinakis, G., Fakotakis, N., & Dermatas, E. (eds), Proceedings of Eurospeech '97*, vol. 1.
- D'Imperio, M., & Rosenthal, S. 1999. Phonetics and phonology of main stress in Italian. *Phonology*, **16**, 1–28.
- D'Imperio, M., Espesser, R., Loevenbruck, H., Menezes, C., Nguyen, N., & Welby, P. à paraitre. Are tones aligned with articulatory events? Evidence from Italian and French. *In : Cole, J., & Hualde, J. (eds), Papers in Laboratory Phonology iX : Change in Phonology*. The Hague : Mouton de Gruyter.
- D'Imperio, M., Petrone, C., & Nguyen, N. 2007. Effects of tonal alignment on lexical identification in Italian. *Pages 79–106 of : Gussenhoven, C., & Riad, T. (eds), Tones and Tunes*. The Hague : Mouton de Gruyter.

- D'Imperio, M., Elordieta, G., S.Frota, Prieto, P., & Vigário, M. 2007b. Intonational phrasing in Romance : The role of syntactic and prosodic structure. *Pages 59–97 of : S. Frota, M. Vigario, & Freitas, M.-J. (eds), Prosodies. Phonetics and Phonology Series.* Berlin : Mouton de Gruyter.
- Dohen, M. 2006. *Deixis prosodique multisensorielle : production et perception audiovisuelle de la focalisation contrastive en français.* Ph.D. thesis, Institut de la Communication Parlée, Grenoble, France.
- Dombrowski, E. 2003. Semantic features of accent contours : Effects of F0 peak position and F0 time shape. *Pages 1217–1220 of : M. J. Solé, D. Recasens, J. Romero (ed), Proceedings of the 15th international congress of phonetic sciences, vol. 1.* Barcelona, Spain : Causal Productions.
- Elman, J., & McClelland, J. 1986. Exploiting lawful variability in the speech wave. *Pages 360–381 of : Perkell, J., & Klatt, D. (eds), Invariance and Variability in Speech Processes.* Hillsdale : LEA.
- Esposito, A., & di Benedetto, M.G. 1999. Acoustic and perceptual analysis of Italian stop consonants. *The Journal of the Acoustical Society of America*, **106**(4), 2051–2062.
- Farnetani, E., & Kori, S. 1991. Rhythmic structure in Italian noun phrases : a study on vowel durations. *Phonetica*, **47**, 50–65.
- Frascarelli, M. 1997. The Phonology of Focus and Topic in Italian. *The Linguistic Review*, **14**, 221–248.
- Frota, S., D'Imperio, M., Elordieta, G., Prieto, P., & Vigário, M. 2007. The phonetics and phonology of intonational phrasing in Romance. *Pages 131–154 of : P. Prieto, J. Mascaró, & Solé, M.-J. (eds), Segmental and Prosodic issues in Romance Phonology.* Amsterdam/Philadelphia : John Benjamins Publishing.
- Féry, C. 1993. *German Intonational Patterns.* Tübingen, Germany : Niemeyer.
- Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. 1995. A connectionist model of phonological representation in speech perception. *Cognitive Science*, **19**, 407–439.

- Gili Fivela, B., & D'Imperio, M. 2003. *Tonal alignment of prenuclear accent in Italian*. poster presented at the 2nd Conference on *Tone and intonation in Europe*, Santorini, Greece.
- Gili Fivela, B., & D'Imperio, M. 2007. *High peak vs. high plateau in the identification of contrastive accents in Italian*. poster presented at the 3rd Conference on *Tone and intonation in Europe*, Lisbon, Portugal.
- Gili Fivela, B., & Savino, M. 2003. Segments, syllables and tonal alignment : A study on two varieties of Italian. *Page 2933–2936 of* : M. J. Solé, D. Recasens, J. Romero (ed), *Proceedings of the 15th international congress of phonetic sciences*, vol. 1. Barcelona, Spain : Causal Productions.
- Goldsmith, J. 1976. *Autosegmental Phonology*. Ph.D. thesis, Massachusetts Institut Of Technology, Cambridge, Massachusetts.
- Grabe, E., Gussenhoven, C., Haan, J., Marsi, E., & Post, B. 1998. Preaccentual pitch and speaker attitude in Dutch. *Language and Speech*, **41**, 63–85.
- 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., D'Imperio, M., Savino, M., & Avesani, C. 2005. Towards a strategy for labelling varieties of Italian. *Pages 55–83 of* : Jun, S.-A. (ed), *Prosodic Typology and Transcription : A Unified Approach*. Oxford : Oxford University Press.
- Grice, M., Baumann, S., & Benzmueller, R. 2007. German intonation in Autosegmental-Metrical Phonology. *Pages 55–83 of* : Jun, S.-A. (ed), *Prosodic Models and Transcription : Towards Prosodic Typology*. Oxford : Oxford University Press.
- Gussenhoven, C. 1984. *On the Grammar and Semantics of Sentence Accents*. Foris.

- Gussenhoven, C. 1988. Autosegmental Studies on Pitch Accent. *Pages 95–121 of* : van der Hulst, H., & Smith, N. (eds), *Adequacy in intonation analysis : the case of Dutch*. Dordrecht : Foris.
- Gussenhoven, C. 2000. The boundary tones are coming : On the nonperipheral realisation of boundary tones. *Pages 132–151 of* : Broe, M., & Pierrehumbert, J. B. (eds), *Papers in Laboratory Phonology V : Acquisition and the Lexicon*. Cambridge : Cambridge University Press.
- Gussenhoven, C. 2002. Intonation and interpretation : Phonetics and phonology. *Pages 45–57 of* : Bel, B., & Marilier, I. (eds), *Proceedings of Speech Prosody*, vol. 1.
- Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge : Cambridge University Press.
- Haan, J. 2001. *Speaking of questions : an explanation of Dutch question intonation*. Ph.D. thesis, LOT, Utrecht, Netherlands.
- Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. The Hague : Mouton.
- Hawkins, S. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, **31**, 375–405.
- Hawkins, S., & Smith, R. 2001. Polysp : A polysystemic, phonetically rich approach to speech understanding. *Italian Journal of Linguistics*, **13**, 99–188.
- Hellmuth, S. 2004. *Prosodic weight and phonological phrasing in Cairene Arabic*. Ph.D. School of Oriental & African Studies, University of London. <http://www.sfb632.uni-potsdam.de/homes/samhellmuth>.
- Hellmuth, S. 2007. The relationship between prosodic structure and pitch accent distribution : Evidence from Egyptian Arabic. *The Linguistic Review*, **24**, 291–316.

- Hirata, Y., & Forbes, C. 2007. Vowel-onset interval as a timing unit for singleton/geminate stop distinction in Japanese. *Pages 349–354 of : Proceedings of the 16th International Congress of Phonetic Sciences*, vol. 2.
- Hirst, D. J. 1983. Structures and categories in prosodic representations. *Pages 93–109 of : Cutler, A., & Ladd, D. R. (eds), Prosody : Models and Measurements*. Berlin : Springer.
- Hirst, D. J., & Espesser, R. 1993. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, **15**, 71–85.
- Hirst, D. J., Di Cristo, A., & Espesser, R. 2000. Levels of representation and levels of analysis for the description of intonation systems. *Pages 51–87 of : Horne, M. (ed), Intonation : Theory and Experiment*. Dordrecht : Kluwer Academic Press.
- Ishihara, T. 2003. A phonological effect on tonal alignment in Tokyo Japanese. *Page 615–618 of : M. J. Solé, D. Recasens, J. Romero (ed), Proceedings of the 15th international congress of phonetic sciences*, vol. 1. Barcelona, Spain : Causal Productions.
- Jun, S.-A. 1993. *The Phonetics and Phonology of Korean Prosody*. Ph.D. thesis, The Ohio State University, Ohio State, USA.
- Jun, S.-A., & Fougeron, C. 2000. A phonological model of French intonation. *Pages 209–242 of : Botinis, A. (ed), Intonation : analysis, modelling and technology*. Boston : Kluwer Academic Press.
- Jun, S.-A., & Fougeron, C. 2002. The Realizations of the accentual phrase in French intonation. *Probus*, **14**, 147–172.
- Karpinski, A. 2007. *Course Notes on Graduate Statistics and Data Analysis*. <http://www.temple.edu/psychology/FacultyWebs/Karpinski/>.
- Klatt, D. H. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129–140.

- Knight, R.-A. 2003. *Peaks and Plateaux : The Production and Perception of Intonational High Targets in English*. Ph.D. thesis, University of Cambridge, Cambridge.
- Knight, R.-A., & Nolan, F. 2006. The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association*, **36**, 21–28.
- Kohler, K. 1987. Categorical pitch perception. *Page 331–333 of : Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 5.
- Kohler, K. 2005. Timing and communicative functions of pitch contours. *Phonetica*, **62**, 88–105.
- Kohler, K. J. 1990. Macro and micro F0 in the synthesis of intonation. *Page 115–138 of : Kingston, J., & Beckman, M. E. (eds), Papers in Laboratory Phonology I : Between the Grammar and the Physics of Speech*. Cambridge : Cambridge University Press.
- Kohler, K. J. 1991. A model of German intonation. *AIPUK*, **25**, 295–360.
- Kvale, K. 1993. *Segmentation and labelling of speech*. Ph.D. thesis, The department of telecommunications of the Norwegian institute of technology, Universitetet I Trondheim, Norway.
- Ladd, D. R. 1983. Phonological features of intonational peaks. *Language*, **59**, 721–759.
- Ladd, D. R. 2008. *Intonational Phonology*. Ii edn. Cambridge : Cambridge University Press.
- Ladd, D. R., & Schepman, A. 2003. ‘Sagging transitions’ between high pitch accents in English : Experimental evidence. *Journal of Phonetics*, **31**, 81–112.
- Ladd, D. R., Faulkner, D., Faulkner, A., & Schepman, A. 1999. Constant ‘segmental’ anchoring of F0 movements under changes in speech rate. *The Journal of the Acoustical Society of America*, **106**(3), 1543–1554.
- Ladd, D. R., Mennen, I., & Schepman, A. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *The Journal of the Acoustical Society of America*, **107**(5), 2685–2696.

- Lahiri, A., & Marslen-Wilson, W. 1991. The mental representation of lexical form : a phonological approach to the recognition lexicon. *Cognition*, **38**, 245–294.
- Leben, W. 1973. *Suprasegmental Phonology*. Ph.D. thesis, Massachusetts Institut Of Technology, Cambridge, Massachussets.
- Levelt, W. J. M. 1989. The speaker as information processor. *Pages 1–28 of* : Levelt, W. J. M. (ed), *Speaking : from intention to articulation*. Cambridge : The MIT Press.
- Levelt, W. J. M. 1993. Accessing words in speech production : Stages, processes and representations. *Pages 241–251 of* : Reuland, E., & Abraham, W. (eds), *Knowledge and language I*. Dordrecht : Kluwer Academic Publishers.
- Liberman, A. M., Cooper, E. S., Shankweiler, D. P., & Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review*, **74**, 431–461.
- Liberman, M. 1975. *The Intonational System of English*. Ph.D. thesis, Massachusetts Institut Of Technology, Cambridge, Massachussets.
- Liberman, M., & Prince, A. 1977. On stress and linguistic rhythm. *Linguistic Inquiry*, **8**, 249–336.
- Liberman, M., & Sag, I. 1974. Prosodic form and discourse function. *CLS*, **10**, 487–497.
- Lindblom, B. E. F. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, **35**(11), 1773–1781.
- Marotta, G. 1985. *Modelli e Misure Ritmiche*. Bologna : Zanichelli.
- Mattei, M., & di Benedetto, M.G. 2000. *Acoustic analysis of singleton and geminate nasals in Italian*. <http://www.essex.ac.uk/web-sls/papers/00-02/00-02.html>.
- Nespor, M., & Vogel, I. 1986. *Prosodic phonology*. Dordrecht : Foris.
- Nguyen, N. 2004. Perception de la parole. *In* : Duran, J., Nguyen, N., Rey, V., & Wauquier-Gravelines, S. (eds), *Phonologie et Phonétique : Approches Contemporaines*. Hermès.

- Nguyen, N., & Hawkins, S. 1999. Implications for word recognition of phonetic dependencies between syllable onsets and codas. *In : Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 3.
- Niebuhr, O. à paraître. The coding of intonational meanings beyond f_0 . *The Journal of the Acoustical Society of America*, **142**, 1252–1263.
- Niebuhr, O. 2003. Perceptual study of timing variables in F0 peaks. *Page 615–618 of : M. J. Solé, D. Recasens, J. Romero (ed), Proceedings of the 15th international congress of phonetic sciences*, vol. 1. Barcelona, Spain : Causal Productions.
- Niebuhr, O. 2007. The signalling of German rising-falling intonation categories-the interplay of synchronization, shape, and height. *Phonetica*, **64**, 174–193.
- Niebuhr, O., & Kohler, K. 2004. Perception and cognitive processing of tonal alignment in German. *Pages 155–158 of : Proceedings of the TAL Conference*.
- Norris, D., McQueen, J., & Cutler, A. 2000. Merging information in speech recognition : Feedback is never necessary. *Behavioral and Brain Sciences*, **23**, 299–325.
- Ohala, J. 1983. Cross-language use of pitch : An ethological view. *Phonetica*, **40**, 1–18.
- Peterson, G. E., & Lehiste, I. 1960. Duration of syllable nuclei in English. *Journal of Acoustical Society of America*, **32**, 693–703.
- Petrone, C. 2005. Effects of Raddoppiamento Sintattico on tonal alignment in Italian. *Pages 1425–1428 of : Proceedings of Interspeech 2005, Lisbon, Portugal*, vol. 2.
- Petrone, C., & Ladd, D. R. 2007. Sentence domain effects on tonal alignment in Italian ? *Pages 1253–1256 of : Proceedings of the 16th International Congress of Phonetic Sciences*, vol. 2.
- Pierrehumbert, J. B. 2000. The phonetic grounding of phonology. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, **5**, 7–23.
- Pierrehumbert, J. B., & Beckman, M. E. 1988. *Japanese Tone Structure*. Cambridge, Massachussets : The MIT Press.

- Pierrehumbert, J. B., & Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. *Pages 271–311 of* : Cohen, P.R., Morgan, J., & Pollack, M.E. (eds), *Intentions in Communication*. Cambridge : MIT Press.
- Pierrehumbert, J. B., & Steele, S. 1989. Categories of tonal alignment in English. *Phonetica*, **46**(3), 181–196.
- Pierrehumbert, J. B., Beckman, M. E., & Ladd, D. R. 2000. Conceptual foundations of phonology as a laboratory science. *Pages 273–303 of* : Burton-Roberts, N., Carr, P., & Docherty, J. (eds), *Prosody : theory and experiments*. Oxford : Oxford University Press.
- Pierrehumbert, J.B. 1980. *The Phonology and Phonetics of English intonation*. Ph.D. thesis, Massachusetts Institut Of Technology, Cambridge, Massachusetts.
- Pike, K. 1945. *The Intonation of American English*. Ann Arbor : University of Michigan Press.
- Pike, K. 1948. *Tone Languages*. Ann Arbor, USA : University of Michigan Press.
- Pinheiro, J., & Bates, D. 2000. *Mixed-effects models in S and S-Plus*. Statistics and Computing Series. New York, NY : Springer-Verlag.
- Pitrelli, J. F., Beckman, M. E., & Hirschberg, J. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Pages 123–126 of* : *Proceedings of the ICSLP*, vol. 1.
- Post, B., D’Imperio, M., & Gussenhoven, C. 2007. Fine phonetic detail and intonational meaning. *Pages 191–196 of* : *Proceedings of the 16th International Congress of Phonetic Sciences*.
- Prieto, P. 2002. Coarticulation and stability effects in tonal clash contexts in Catalan. *Pages 45–57 of* : Bel, B., & Marilier, I. (eds), *Proceedings of Speech Prosody*, vol. 1.
- Prieto, P., & Torreira, F. à paraître. The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics*.

- Prieto, P., van Santen, J. P. H., & Hirschberg, J. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics*, **23**, 429–451.
- Prieto, P., D’Imperio, M., & Fivela, B. Gili. 2003. Tonal structure and pitch targets in Italian focus constituents. *Catalan Journal of Linguistics*, **2**, 55–65.
- Prieto, P., D’Imperio, M., & Fivela, B. Gili. 2005. Pitch accent alignment in Romance : primary and secondary associations with metrical structure. *Language and Speech (special issue on “Variation in Intonation”)*, **48**(4), 359–396.
- Prince, A., & Smolensky, P. 1993. *Optimality Theory : Constraint Interaction in Generative Grammar*. Ms. Rutgers University and Johns Hopkins University : Zanichelli.
- Prom-on, S., & Xu, Y. en préparation. *Modeling tone and intonation in Mandarin and English as a process of target approximation*. <http://www.phon.ucl.ac.uk/home/yi/publications.html>.
- R Development Core Team. 2008. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Radtchke, T., & Harrington, J. à paraître. The variability of early accent peaks in Standard German. *In : Papers in Laboratory Phonology X*.
- Repp, B. 1984. Categorical perception : issues, methods, findings. *Pages 231–261 of : Lass, N. (ed), Speech and Language : Advances in Basic Research and Practice*. Orlando : Academic Press.
- Rietveld, T., & Gussenhoven, C. 1995. Aligning pitch targets in speech synthesis : effects of syllable structure. *Journal of Phonetics*, **23**, 375–385.
- Riggs, D., & Byrd, D. 2007. *The scope of phrasal lengthening in articulation : Prosody and prominence*. talk presented at the 11th Conference on Laboratory Phonology, Wellington, New Zealand.
- Rossi, M. 1999. *L’intonation : le système du français*. Paris, France : Ophrys.

- Sag, I., & Liberman, M. 1975. The intonational disambiguation of indirect speech act. *CLS*, **11**, 487–497.
- Schepman, A., Lickley, R., & Ladd, D. R. 2006. Effects of vowel length and “right context” on the alignment of Dutch nuclear accents. *Journal of Phonetics*, **34**, 1–28.
- Scherer, K., Ladd, R. D., & Silverman, K. 1984. Vocal cues to speaker affect : Testing two models. *Journal of the Acoustical Society of America*, **76**, 1346–1356.
- Selkirk, E. 1978. On prosodic structure and its relation to syntactic structure. *Pages 111–140 of* : Fretheim, T. (ed), *Nordic Prosody*. Trondheim : TAPIR.
- Selkirk, E. 1984. *Phonology and Syntax : The Relation between Sound and Structure*. Cambridge : The MIT Press.
- Selkirk, E. 2000. The interaction of constraints on prosodic phrasing. *Pages 231–261 of* : Horne, M. (ed), *Prosody : theory and experiments*. Dordrecht : Kluwer Academic Press.
- Silverman, K. 1987. *The structure and processing of fundamental frequency contours*. Ph.D. thesis, University of Cambridge, Cambridge, Massachusetts.
- Silverman, K., & Pierrehumbert, J. B. 1990. The timing of prenuclear high accents in English. *Pages 72–106 of* : Kingston, J., & Beckman, M. E. (eds), *Papers in Laboratory Phonology I : Between the Grammar and the Physics of Speech*. Cambridge : Cambridge University Press.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Prince, C., Pierrehumbert, J. B., & Hirschberg, J. 1992. ToBI : a standard for labeling English prosody. *Pages 867–870 of* : *Proceedings of the 2nd International Conference on Spoken Language Processing*, vol. 2. Banff, USA : Causal Productions.
- Steedman, M. 2003. Information-Structural Semantics for English Intonation. *In* : *LSA Summer Institute Workshop on Topic and Focus, Santa Barbara*. Kluwer.
- Steele, S. A. 1986. Nuclear accent F0 peak location : effects of rate, vowel, and number of following syllables. *Journal of Acoustic Society, Supplement*, **80**, s51.

- Stevens, K. N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, **111**, 1872–1891.
- Stevens, K. N., & Blumstein, S. E. 1978. Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, **64**, 1358–1368.
- 't Hart, J. 1981. Differential sensitivity to pitch distance. *The Journal of the Acoustical Society of America*, **69**, 811–821.
- 't Hart, J., Collier, R., & Cohen, A. 1990. *A Perceptual Study of Intonation : An Experimental-Phonetic Approach*. Cambridge : Cambridge University Press.
- Trager, L., & Smith, H. L. 1951. *An outline of English structure*. Norman, Oklahoma : Battenburg Press.
- Turk, A., & Hufnagel, S. Shattuck. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, **35**, 445–472.
- Uldall, E. T. 1960. Attitudinal meanings conveyed by intonation contours. *Language and Speech*, **3**, 223–234.
- van Heuven, V., & van Zanten, E. 2005. Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, **47**, 87–99.
- van Santen, J. P. H. 2002. Quantitative modeling of pitch accent alignment. *Pages 107–112 of* : Bel, B., & Marilier, I. (eds), *Proceedings of Speech Prosody*, vol. 1.
- van Santen, J. P. H., & Hirschberg, J. 1994. Segmental effects on timing and height of pitch contours. *Page 719–722 of* : *Proceedings of the International Conference on Spoken Language Processing*, vol. 2.
- van Santen, J. P. H., & Mobius, Bernd. 1999. A quantitative model of f₀ generation and alignment. *Page 269–288 of* : Botinis, A. (ed), *Intonation Analysis, Modelling and Technology*. Netherlands : Kluwer Academic Press.
- Varnell Bosch, M. 2006. *The phonological role of tonal scaling in Majorcan Catalan interrogatives*. MA dissertation.

- Venables, W. N., & Ripley, B.D. 1999. *Modern Applied Statistics with S-PLUS*. Statistics and Computing Series. New York, NY : Springer-Verlag.
- Warren, P., & Marslen-Wilson, W. 1987. Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics*, **41**, 262–275.
- Welby, P. 2003. *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. Ph.D. thesis, The Ohio State University, Ohio State, USA.
- Welby, P., & Loevenbruck, H. 2005. Anchored down in anchorage : Syllable structure and segmental anchoring in French. *Journal of Linguistics*, **18**(1).
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, **3**, 1707–1717.
- Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, **55**(3), 179–203.
- Xu, Y. 2002. Articulatory Constraints and Tonal Alignment. *Pages 91–100 of* : Bel, B., & Marilier, I. (eds), *Proceedings of Speech Prosody*, vol. 1.
- Xu, Y., & Wang, Q. E. 2001. Pitch targets and their realization : Evidence from Mandarin Chinese. *Speech Communication*, **33**, 319–337.

	DD			AS		
	L1	H	L2	L1	H	L2
c0	- 377	- 358	- 298	- 301	- 363	- 340
v0	- 373	- 360	- 298	- 304	- 371	- 334
v1	- 289	- 314	- 251	- 278	- 326	- 351

TABLE A.1 – Valeurs de AIC des modèles d’alignement temporelle pour L1, H et L2. L’alignement est calculé en fonction de la distance absolue aux unités segmentales (**c0** = début de la consonne ; **v0/v1** = début/fin de la voyelle)

	DD			AS		
	L1	H	L2	L1	H	L2
syll %	477	358	298	301	363	340
voy %	373	360	298	304	371	334

TABLE A.2 – Valeurs de AIC des modèles d’alignement temporelle pour L1, H et L2. L’alignement est calculé en fonction de la distance relative aux unités segmentales (**syll %** = pourcentage d’alignement dans la syllabe ; **voy %** = pourcentage d’alignement dans la voyelle)

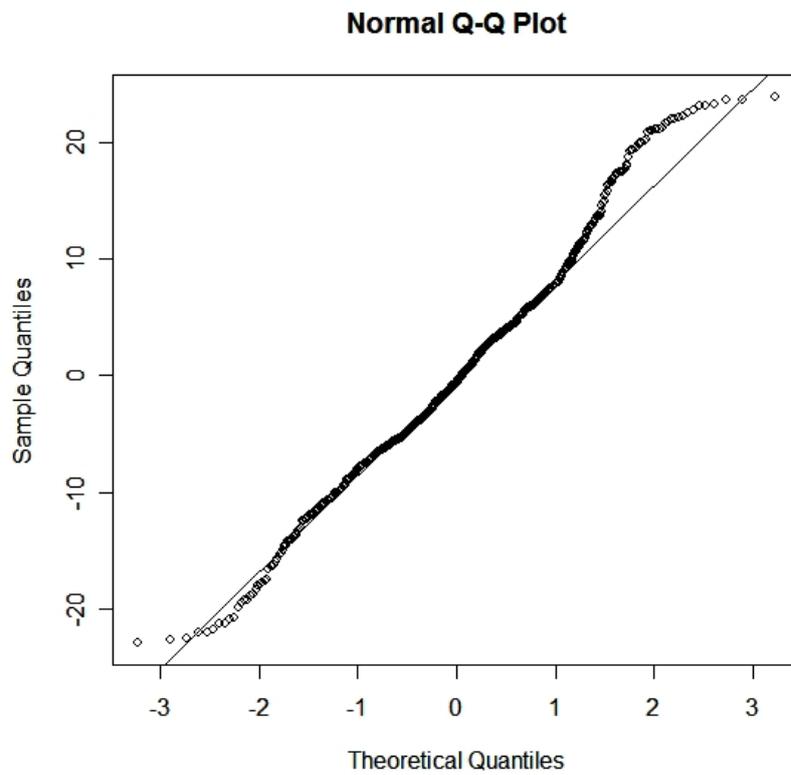
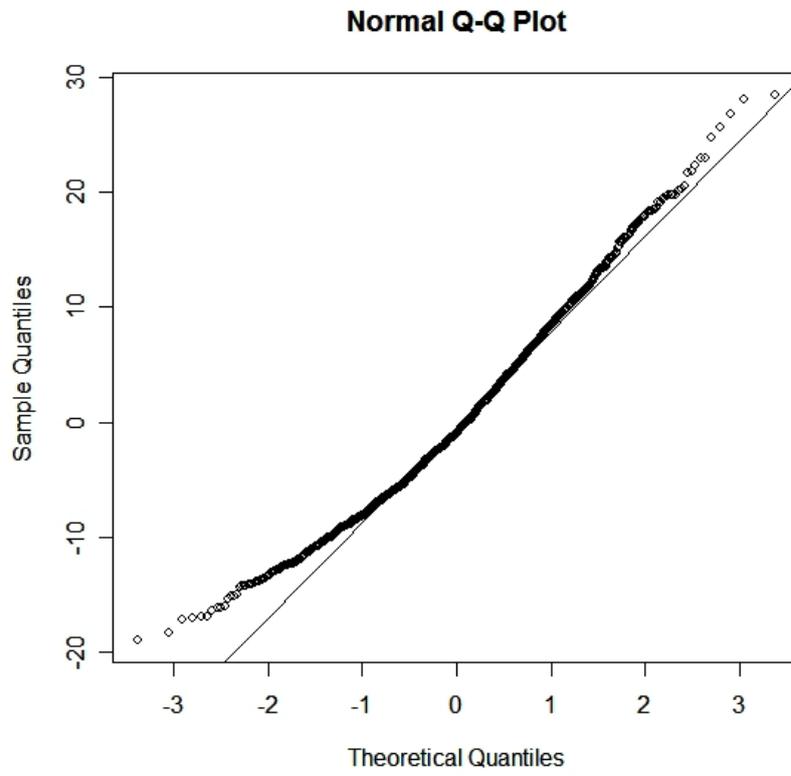


FIGURE A.1 – Quantile-Quantile plot des résidus de la régression logistique effectuée sur les questions (en haut) et de la régression linéaire par morceaux effectuée sur les affirmations (en bas). Locuteur OM. 289

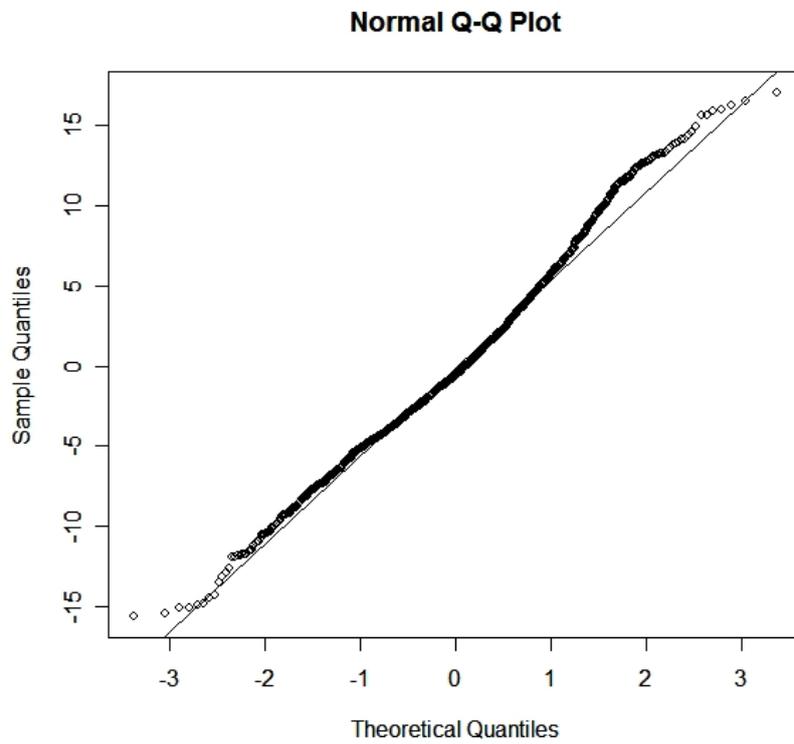
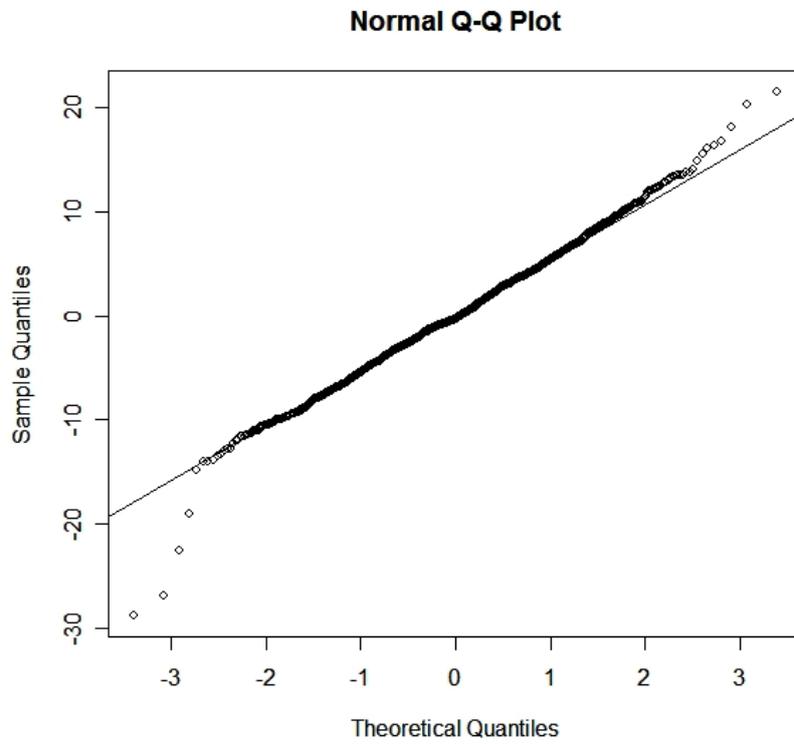


FIGURE A.2 – Quantile-Quantile plot des résidus de la régression logistique effectuée sur les questions (en haut) et de la régression linéaire par morceaux effectuée sur les affirmations (en bas). Locuteur AS.

	DD			AS		
	L1	H	L2	L1	H	L2
cons + voy	- 351	- 355	- 281	- 298	- 387	- 368
rime	- 345	- 353	- 282	- 294	- 369	- 360

TABLE A.3 – Valeurs de AIC des modèles d’alignement temporelle pour L1, H et L2. L’alignement est calculé en fonction de la durée des segments (**cons + voy** = durée de la consonne et de la voyelle ; **rime** = durée de la rime (/voyelle))

