Bulletin PFC

Phonologie du Français Contemporain
Usages, Variétés et Structure

sous la Direction de

Jacques Durand, Université Toulouse-Le Mirail
Bernard Laks, MODYCO Paris-X Nanterre
Chantal Lyche, Oslo et Tromsø

Bulletin n°6

Coordination : Anne Catherine Simon,
Geneviève Caelen-Haumont
et Claudine Pagliano

Prosodie du français contemporain

L'autre versant de PFC

2006

Edition Nicole Serna

Equipe de Recherche en Syntaxe et Sémantique
UMR 5610
CNRS & Université de Toulouse-Le Mirail

SOMMAIRE

Introduction

Anne Catherine SIMON

Parti de la constatation que de nombreux travaux sur la phonologie du français s'appuyaient sur des données étroites, le projet *Phonologie du français contemporain : usages, variétés, structure* s'est donné comme premier objectif de renouveler la base empirique des études phonologiques en constituant un corpus de référence afin d'offrir à terme une vision globale de la phonologie du français. Dans un premier temps, la direction du projet (Jacques Durand, Bernard Laks, Chantal Lyche) s'est attachée à définir un protocole commun à tous les points d'enquête de façon à recueillir des données fiables et comparables. Ce protocole a été conçu en sorte que l'on puisse établir l'inventaire phonémique de chaque locuteur et, très tôt, la flexibilité de l'outil Praat a permis d'élargir le domaine des recherches pour inclure l'étude du schwa et de la liaison par le biais d'un codage spécifique.

La chronologie a été la suivante. On s'est d'abord intéressé à la phonologie et on a défini un protocole pour rassembler les données et permettre une identification de l'inventaire phonémique des locuteurs ; afin de faciliter l'accès aux données nécessaires pour l'analyse des deux incontournables de la phonologie du français – le schwa et la liaison –, on a conçu des systèmes de codage sous Praat[1] ; finalement, en 2003, on s'est tourné vers le suprasegmental. Geneviève Caelen-Haumont a été la première coordinatrice de l'étude de la prosodie au sein du projet PFC (2003-2004).

En juillet de cette année-là, Jacques Durand, Bernard Laks et Chantal Lyche invitaient quelques spécialistes de la prosodie du français aux 5e Journées PFC à Toulouse-le-Mirail. Entre autres chercheurs, Véronique Aubergé et Philippe Martin traçaient une première esquisse des problématiques du domaine, et donnaient un aperçu des outils de transcription et d'annotation existants. Les membres de PFC ont aussi pris connaissance des projets sur la variation prosodique hors du domaine francophone[2]. En effet, peu de choses existaient pour le français, si l'on excepte le projet AMPER (Atlas Multimédia Prosodique de l'Espace Roman) et les travaux déjà anciens de Fernand Carton (par exemple, Carton et al. 1983). La décision a ensuite été prise de proposer un protocole de codage minimal pour la prosodie, qui serait appliqué pour certains points d'enquêtes par les équipes disposant de spécialistes en la matière. Le 3e numéro du *Bulletin PFC* offre un aperçu des premières suggestions pour une prise en compte de la prosodie (Eychenne & Mallet 2004).

En 2004, un premier *atelier prosodie* réunissait quelques chercheurs à l'Université de Caen[3] pour réfléchir à un protocole de codage minimal – qui devait être partageable par des chercheurs de diverses orientations théoriques, et applicable par des codeurs non spécialistes de la prosodie après un entraînement raisonnable. À cette occasion, une série de débats théoriques et méthodologiques passionnants ont été ouverts :

[1] Les sommaires des premiers *Bulletins PFC* (consultables sur http://www.projet-pfc.net/) retracent cette orientation de départ.

[2] Voir la communication d'Anne Catherine Simon à Toulouse en juillet 2003 : "Les projets internationaux dans le domaine de la variabilité prosodique".

[3] Atelier organisé les 7-8 mai 2004 à l'Université de Caen, par Jacques Durand, Bernard Laks et Chantal Lyche, avec le soutien de l'OFNEC. Étaient présents Cyril Auran, Geneviève Caelen-Haumont, Annelise Coquillon, Anne Lacheret, Chantal Lyche, Philippe Martin, Piet Mertens, François Poiré et Anne Catherine Simon.

- Quel domaine de codage adopter pour l'étude de la variation prosodique ? La notion de proéminence a-t-elle une réalité perceptuelle ou théorique et peut-elle servir de point de départ à la définition d'un tel domaine[4] ?
- Faut-il privilégier une transcription et un codage de la prosodie directement utiles à l'étude du schwa et de la liaison (impliquant un modèle phonologique sous-jacent), ou une approche descriptive large de la variation prosodique en français ?
- Le codage doit-il être manuel (auditif), partiellement ou totalement automatisable ? Pour quels logiciels et quel type d'automatisation opter ?

À l'issue de cet atelier, il a été décide que le travail en prosodie devait être différencié : (i) au service du segmental, (ii) en vue de décrire les caractéristiques diatopiques.

Où en est-on aujourd'hui ? Depuis 2005, j'ai repris la coordination du versant "prosodie" au sein de PFC. Plusieurs débats initiés à Caen en 2004 sont encore en cours actuellement, comme en témoignent certaines contributions à ce numéro. D'autre part notre maîtrise des outils de transcription et de codage a considérablement progressé (développement de Prosogramme et de Melism ; amélioration de l'interfaçage entre les logiciels WinPitch et Praat) – et on entrevoit dès lors comment concilier les apports d'un codage au service de l'analyse du domaine prosodique du schwa et de la liaison, et d'un codage qui permet d'établir sur quels aspects strictement suprasegmentaux plusieurs variétés de français diffèrent (typologies et inventaires de contours intonatifs ; aspects rythmiques ou de registre). Concrètement, un second *atelier prosodie* est organisé à Paris en février 2006, et il s'adresse cette fois aux doctorants et jeunes chercheurs du projet soucieux de maîtriser ces logiciels. Ceci augure d'un développement prochain des recherches prosodiques au sein de PFC, et le colloque international de Louvain-la-Neuve en juillet 2006 fera d'ailleurs une large place à la prosodie.

Présentation des contributions

Les neuf contributions qui composent ce *Bulletin* sont regroupées en trois sections.

Une première partie, ***Prosodie : outils d'annotation et codage***, permet de faire le point sur les outils actuellement à disposition des chercheurs. Dans un très riche et complet article collectif, Elisabeth Delais-Roussarie (coord.), Geneviève Caelen-Haumont, Daniel Hirst, Philippe Martin et Piet Mertens présentent deux outils de transcription et d'annotation prosodiques (Praat et WinPitch) et trois programmes d'étiquetage (semi-)automatiques pour l'information mélodique (et, dans une moindre mesure, accentuelle et rythmique) : Prosogramme, Momel-Intsint et Melism. Les deux articles suivants illustrent chacun un des objectifs fondamentaux de "PFC prosodie" : d'une part il s'agit de mettre en valeur le lien entre la prosodie et la variation du schwa et de la liaison ; d'autre part, il s'agit de décrire et de comparer les caractéristiques prosodiques de différentes variétés de français.

Dans leur article, Anne Lacheret et Chantal Lyche affinent et améliorent considérablement le protocole de codage permettant d'explorer le rôle des facteurs prosodiques dans l'analyse phonologique du schwa. Elles posent en outre des questions importantes sur le repérage et la sélection des passages pertinents à coder, étant entendu que le caractère coûteux d'un codage prosodique empêche qu'on l'applique de manière exhaustive aux données codées par ailleurs pour d'autres phénomènes segmentaux. Le système de codage qui résulte de leurs hypothèses théoriques sur le schwa en français et de pré-tests sur les données se décline dans un protocole standard et étendu.

[4] Cette question fait l'objet d'une discussion dans deux articles du présent *Bulletin* (Poiré ; Martin).

L'autre objectif de "PFC prosodie" est rencontré dans l'article de Brechtje Post, Elisabeth Delais-Roussarie et Anne Catherine Simon. Le système IVTS (pour *Intonational Variation Transcription System*) mis au point par les auteurs[5] est un protocole de codage multilinéaire qui vise à capturer les paramètres prosodiques (phonétiques et phonologiques) par lesquels des variétés de français peuvent se distinguer les unes des autres. Applicable sous Praat, ce système de codage permet d'annoter de manière autonome les aspects rythmiques, phonétiques locaux et globaux, et phonologiques. Le niveau d'annotation phonologique est fondé sur une approche métrique-autosegmentale de la prosodie, et le domaine de codage (ID) s'établit à partir d'un repérage auditif des syllabes proéminentes (avec les difficultés inhérentes à cette procédure ; voir les contributions de la section 2).

La deuxième section de ce *Bulletin* présente les suites d'une discussion mouvementée inaugurée lors de l'*atelier prosodie* de Caen en 2003. ***L'identification des proéminences*** est souvent au centre du travail du transcripteur, car elle permet dans bien des cas de décider du domaine même dans lequel s'effectuera une analyse phonétique ou phonologique. Or cette identification, qu'elle se fasse sur des bases perceptives ou instrumentales, est plus que problématique. La contribution de François Poiré retrace une expérience de codage des proéminences par sept codeurs sur un corpus de quelques minutes de parole spontanées. L'auteur y décrit la tâche à accomplir par les codeurs, et tente d'expliquer les grandes divergences inter-codeurs constatées à l'issue de ce périlleux exercice. À sa suite, Philippe Martin offre quelques pistes de réflexion pour comprendre pourquoi l'identification des proéminences en français spontané relève d'un art davantage que d'une procédure objectivable et répétable. Suivant le contexte et la qualification du codeur (locuteur non expert, phonéticien, phonologue), les catégories appliquées pour décider quelle syllabe peut prétendre à l'étiquette "proéminente" varient fortement, et ainsi varie le résultat du codage. Les deux auteurs se refusent cependant à conclure sur un constat d'échec, et fournissent des pistes pour améliorer l'identification perceptive des proéminences.

La troisième section du *Bulletin* fait état d'***analyses prosodiques sur quelques points d'enquête PFC***. En effet, même si les questions théoriques et méthodologiques ne sont pas toutes résolues, on dispose déjà de résultats issus d'analyses de données. La piste empruntée par Cécile Woehrling et Philippe Boula de Mareüil est originale : à partir de données de parole lue et spontanée provenant de six points d'enquête PFC (Normandie, Vendée, Suisse romande, Pays Basque, Languedoc et Provence), les auteurs ont testé comment des auditeurs naïfs d'Île-de-France identifiaient les accents régionaux, en fonction de l'origine géographique des locuteurs et de leur âge. Les résultats de l'expérience montrent que l'âge des locuteurs est un facteur qui a plus d'impact que le registre (lu vs. spontané), et que globalement trois accents émergent : français du Nord, français méridional et suisse romand.

L'étude d'Annelise Coquillon met en évidence un paramètre macro-prosodique, l'étendue du registre tonal, comme marqueur régional de la variété de français pratiquée par les locuteurs de son échantillon (Aix-Marseille). L'étude statistique de ce paramètre dans la parole de locuteurs appartenant à trois générations différentes montre qu'une étendue tonale importante semble être un trait régional mieux conservé par les générations plus âgées, ce qui corrobore une tendance de l'étude de Woehrling et Boula de Mareüil concernant l'impact de l'âge sur le "degré d'accent" (perçu ou produit) des locuteurs.

C'est le français parlé dans le canton de Vaud (Suisse romande) qui sert de point de départ à l'étude sur l'allongement vocalique menée par Helene N. Andreassen. À partir de l'analyse de ses données, l'auteur montre que l'allongement (en position finale absolue)

[5] IVTS s'inspire du projet IViE auquel Brechtje Post a collaboré, voir par exemple Grabe & Post (2002).

conserve une valeur distinctive en français de Vaud pour le marquage des formes féminines et plurielles. Elle discute ensuite, dans le cadre de la théorie de l'optimalité, la façon dont la durée contrastive est représentée dans la grammaire, sous la forme d'une spécification lexicale. Les deux propositions théoriques présentées mettent l'accent sur la représentation morique et le facteur morphologique.

La dernière contribution, due à Geneviève Caelen-Haumont, fait le lien entre plusieurs articles. De par sa méthodologie d'abord (codage de l'intonation en points cibles via le système Momel), présentée dans l'article de Delais-Roussarie *et al.* et utilisée également par Coquillon ; de par le corpus ensuite, qui comprend également des locutrices issues de 4 générations différentes (dans une même famille). Geneviève Caelen-Haumont s'intéresse aux mélismes, ces motifs mélodiques plus ou moins complexes ayant au moins une cible tonale réalisée dans l'aigu. Elle les décrit à l'aide de son outil Melism qui permet d'enrichir significativement le codage tonal. Sans être des marqueurs régionaux, les mélismes contribuent à l'expression sémantique et pragmatique de la subjectivité des locuteurs. L'étude permet de regrouper une grande variété des formes acoustiques de mélismes en familles tonales, qui à leur tour rendent possible la comparaison des mélismes réalisés par les différentes locutrices.

Références

Centre de dialectologie de l'Université de Grenoble. *Projet d'un Atlas Multimédia Prosodique de l'Espace Roman*. [Site web]. http://www.u-grenoble3.fr/dialecto/AMPER/amper.htm (consulté le 10 janvier 2005)

Carton, F., M. Rossi, D. Autesserre et P. Léon (1983). *Les Accents des Français*, Hachette, Paris.

Eychenne, J. et G. Mallet (eds.). (2004). *Du segmental au prosodique : protocoles, outils, extensions et travaux en cours*. *Bulletin PFC* 3. ERSS UMR 5610 CNRS & Université de Toulouse-Le Mirail.

Grabe, E. & B. Post. (2002). "Intonational Variation in English". In B. Bel & I. Marlin (eds), *Proceedings of the Speech Prosody 2002 Conference*, 11-13 Avril 2002, Aix-en-Provence: Laboratoire Parole et Langage, 343-346

Outils d'aide à l'annotation prosodique de corpus

Elisabeth DELAIS-ROUSSARIE[*]
Geneviève CAELEN-HAUMONT[♠]
Daniel HIRST[♠]
Philippe MARTIN[•]
Piet MERTENS[#]

1. Introduction

Depuis une dizaine d'années, on assiste à un renouveau des approches sur corpus en linguistique, et cela, quelle que soit la visée des recherches : grammaticale ou descriptive. En prosodie, plusieurs travaux sur l'intonation et sur l'étude des variations accentuelles et intonatives ont été menés à partir d'un travail sur corpus (cf., entre autres, Mertens (1987 et sq.), Simon (2004), Caelen-Haumont (2000 ; 2004), Portes (2004), Grabe et al. (2002), et les projets IViE et PFC). Ces différents travaux reposent généralement sur une annotation prosodique dont le but est double :

- encoder symboliquement les phénomènes prosodiques ;
- comparer différentes productions à partir d'une méthode d'encodage unifiée (cf. aussi sur ce point Post et. al. (ce volume)).

Annoter prosodiquement un corpus peut se faire en gros de deux façons distinctes : soit manuellement, soit automatiquement. Ceci étant, dans les deux cas, des outils sont nécessaires. Dans ce document, nous nous proposons donc de présenter deux types d'outils fréquemment utilisés pour effectuer des annotations prosodiques de corpus :

- des outils d'analyse prosodique permettant d'effectuer des annotations et de les sauvegarder : WINPITCH et PRAAT ;
- des programmes d'étiquetage (semi-)automatique d'information mélodique : le PROSOGRAPHE, MOMEL-INTSINT et MELISM.

Dans un premier temps, nous allons présenter les deux outils d'analyse, WINPITCH PRO et PRAAT. Après avoir rapidement présenté les fonctionnalités de ces deux logiciels, nous en proposerons une comparaison en insistant à la fois sur leur ergonomie, le type de fichier accepté en entrée, les sorties proposées, etc. Dans un second temps, nous présentons les trois systèmes d'annotation prosodique automatique fonctionnant sous PRAAT : le PROSOGRAPHE, MOMEL-INTSINT et MELISM. Après une présentation des fonctionnalités de ces outils, nous en proposerons une comparaison synthétique.

2. Outils d'analyse prosodique et d'aide à l'annotation

2.1. WINPITCH, un outil logiciel de transcription et d'alignement spécialisé pour l'analyse prosodique[1]

Comme son nom le suggère, WINPITCH est un logiciel d'analyse de la parole dédié dès le début de sa conception à l'étude des facteurs prosodiques (fréquence fondamentale, intensité, durée) et des formants. Il possède des fonctions originales spécialisées pour la

[*] CNRS, UMR 7110 / Laboratoire de Linguistique formelle, Paris 7
[♠] CNRS, UMR 6057 / Laboratoire Parole et Langage, Université de Provence
[•] UFR Linguistique, Université Paris 7 Denis Diderot
[#] Département de Linguistique, Université catholique de Louvain (KUL), Belgique
[1] Cette section a été rédigée par Philippe Martin. Le logiciel WINPITCHPRO peut être téléchargé sur le site http://www.winpitch.com

transcription, l'alignement et l'analyse des données prosodiques de grands corpus. WinPitch a été développé dans le cadre du projet européen C-ORAL-ROM (2005) en vue de la constitution de grands corpus de parole spontanée dans quatre langues romanes (italien, français, espagnol, portugais).

Comparé à d'autres logiciels de même type, WinPitch se distingue par son ergonomie poussée, et ses nombreuses fonctions spécialisées pour l'analyse prosodique :

1. Enregistrement du son de parole avec visualisation des courbes de signal de parole, d'intensité, de fréquence fondamentale et de spectrogramme en temps réel. Cette caractéristique permet un monitoring précis des conditions d'enregistrement : amplitude, écho, niveau de bruit ambiant, etc., caractéristiques très variables lors de la saisie de parole spontanée ;
2. Lecture et analyse de très grands fichiers (16 G Octets) ;
3. Transcription parole-texte en format Unicode (incluant donc l'API, les caractères chinois, des alphabets arabes, hébreux, tibétains, etc.) ;
4. Navigation aisée dans toute l'étendue du signal, avec sélection de zoom à 2 niveaux. La sélection à la souris d'un segment sonore du signal de parole entraîne automatiquement l'affichage de l'analyse acoustique correspondante (Fo, intensité, spectrogramme,…) ;
5. Play-back de segment sélectionné à vitesse programmable, permettant l'écoute ralentie pour une perception et une transcription plus efficaces. Trois moteurs de ralenti disponibles (Psola, autocorrélation, phase) ;
6. Fenêtre de transcription à clavier programmable avec sélection semi-automatique des segments à transcrire ;
7. Alignement à la volée : des textes existants peuvent être alignés à la volée en ralentissant le signal de parole en mode continu (*stream*), l'opérateur cliquant sur les éléments du texte au fur et à mesure de leur écoute. Une base de donnée est alors élaborée dynamiquement au fur et à mesure de la définition des segments temporels, y compris pour des textes avec plusieurs locuteurs annotés selon la convention CHILDES (cf. figure ci-après) ;

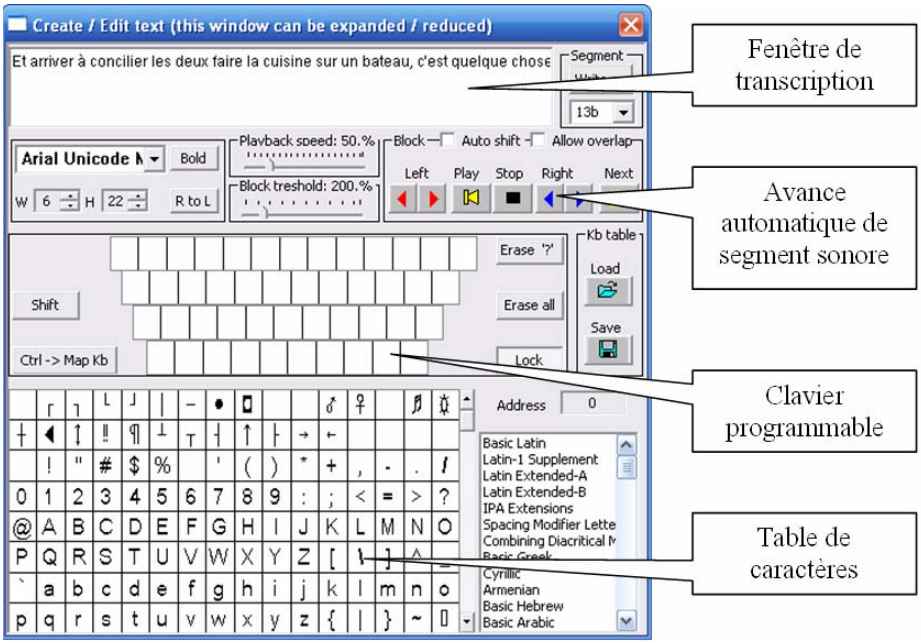
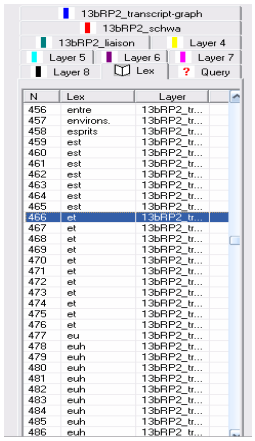


Figure 1. *Fenêtre de transcription et d'alignement sous WINPITCH PRO*

8. Moteur d'analyse des données intégré : la sélection d'un élément ou d'un groupe d'éléments du texte génère automatiquement l'écoute et l'analyse acoustique du segment sonore correspondant. Un lexique d'entrée est automatiquement généré à partir des unités du texte et permet la recherche immédiate des occurrences sonores correspondantes par un seul clic de souris. La recherche par caractéristiques morphologiques (préfixes, suffixes, etc.) est également possible ;
9. Sortie des données d'alignement en 2 formats aisément éditable : un format propriétaire wp2 et un format XML ;
10. 96 tires de transcription, chacune de caractéristiques programmables (police de caractères, couleur, allocation au texte, à l'étiquetage syntaxique, etc.) ;
11. Alignement précis et facile (à la souris) des chevauchements de tour de parole par examen spectrogramme à bande étroite pour la séparation des harmoniques de voix différentes ;
12. Quatre méthodes d'analyse de la fréquence fondamentale (peigne spectral, autocorrélation, AMDF, sélection harmonique) pour l'obtention de courbes correctes dans des cas difficiles d'enregistrements de parole spontanée très bruités ;
13. Morphing prosodique (et des formants) : modification par commande graphique des paramètres prosodiques de fréquence fondamentale, d'intensité et de durée de segments définis par l'utilisateur à l'intérieur du signal de parole ;
14. Analyse statistique multi-sélection (par tire, segments temporels, etc.) par basculement des données (en un seul clic) dans un tableur Excel® ;
15. Lecture de fichiers multimédia dans la plupart des formats y compris les formats image (wav, aiff, au, mp2, mp3, mp4, mpeg, mpg, wma, etc.). Conversion des formats et des fréquences d'échantillonnage intégrés. Transcription et alignement de fichiers multimédias ;
16. Annotation texte des courbes et spectrogrammes et sauvegarde en format image (4 formats de sortie disponibles) pour illustration d'articles, etc. ;
17. Compatible en lecture avec les fichiers Transcriber (trs) et en entrée/sortie avec les fichiers Praat (TextGrid) pour l'utilisation concurrente de ces programmes et le traitement de transcriptions existantes.

Voici, à titre indicatif, un exemple d'exploitation d'une transcription : recherche des contextes sonores et prosodiques de la conjonction "*et*" :

Le tableau d'entrée lexicale (généré automatiquement) contient toutes les occurrences des unités du texte ;
Cliquer sur une entrée "*et*" du tableau permet l'affichage immédiat du contexte et de son analyse acoustique ainsi que l'écoute du segment correspondant, à vitesse normale, accélérée ou ralentie ;
Morphing prosodique (modification de la courbe de Fo sur la syllabe accentuée "*Pay**an***" dans cet exemple).



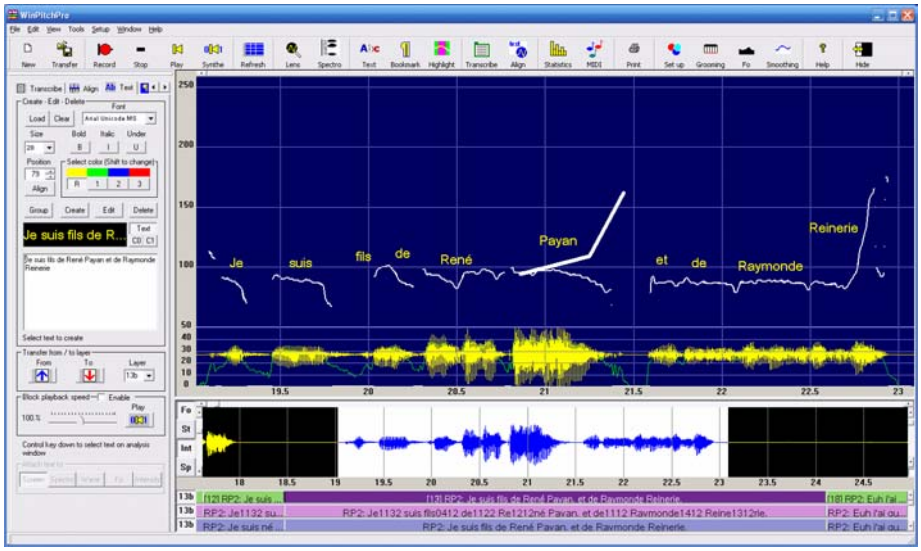


Figure 2. *Écran de WinPitch : fenêtres de commande (à gauche), de navigation (en bas à droite), d'analyse (en haut à droite)*

2.2. Praat

PRAAT est un logiciel d'analyse de la parole qui a été développé à l'Institut des Sciences Phonétiques de l'Université d'Amsterdam (Pays-Bas) par Paul Boersma et David Weenink et qui fonctionne sur plusieurs plates-formes[2]. Il peut être téléchargé à partir de l'adresse suivante : http://www.praat.org/. Ce programme offre la possibilité d'effectuer de multiples tâches :

1. Enregistrer des fichiers audio qui peuvent être ensuite analysés sous PRAAT. Ces fichiers peuvent être codés selon une multitude de formats audio[3] ;
2. Segmenter, transcrire et annoter des fichiers audio dont la taille peut aller jusqu'à 2 Gigabytes, c'est à dire 3 heures d'enregistrement stéréo de qualité CD ou 16 heures d'enregistrement mono à 22 kHz. Ces enregistrements peuvent avoir été effectués sous PRAAT ou peuvent provenir d'autres fichiers audio aux formats divers (cf. supra) ;
3. Effectuer des analyses phonétiques et acoustiques au niveau segmental. Le logiciel permet de calculer des paramètres prosodiques comme l'intensité, la fréquence fondamentale, le voisement, etc., et ceci selon plusieurs algorithmes, de mener des analyses spectrographiques et de donner des mesures précises telles que la durée du VOT des plosives, les valeurs des différents formants d'une voyelle, etc. ;

[2] Le programme PRAAT fonctionne sur les systèmes suivants : Windows (et plus précisément Win 95, 98, NT4, ME, 2000 et XP), Macintosh (systèmes 7.1 à 9.2), Linux, Sparc SOLARIS, HP Unix et Silicon Graphics IRIX (et plus précisément Indigo, Indy, O2, Onyx et autres).

[3] L'enregistrement des données audio peut se faire sous plusieurs formats qui sont synthétisés dans le tableau ci-dessous :

FORMATS POSSIBLES POUR LA SAUVEGARDE (ET LA LECTURE) DES DONNÉES AUDIO	
AIFF	NIST
WAV	Kay Sound file
AIFC	RAW 16 bits big Endian file
AU	RAW 16 bits Little Endian file

4. Étudier les paramètres prosodiques (F0, durée et intensité) et modifier par stylisation des courbes de fréquence fondamentale et d'intensité ;
5. Effectuer des manipulations et des modifications du signal de parole (utilisation de filtres ; analyse-synthèse, etc.) ;
6. Construire des outils d'apprentissage (Réseau de neurones et élaboration de grammaires dans le cadre de la Théorie de l'Optimalité (OT : Optimality Theory) ;
7. Écrire des scripts pour effectuer plus rapidement certaines tâches d'analyse, d'extraction d'information ou d'édition, etc.

Nous n'allons pas présenter ici en détail toutes ces fonctionnalités. Nous renvoyons donc le lecteur intéressé à Lieshout (2004) et à Delais et. al (2003). En revanche, nous souhaitons nous attarder sur deux points, (i) l'ergonomie du logiciel et son interface utilisateur, (ii) les fichiers d'annotation.

Le logiciel PRAAT propose une interface utilisateur assez déroutante au premier abord, dans la mesure où elle est différente de celle fréquemment rencontrée. Ainsi, au lancement du programme PRAAT, deux fenêtres s'ouvrent à l'écran (cf. figure 3).

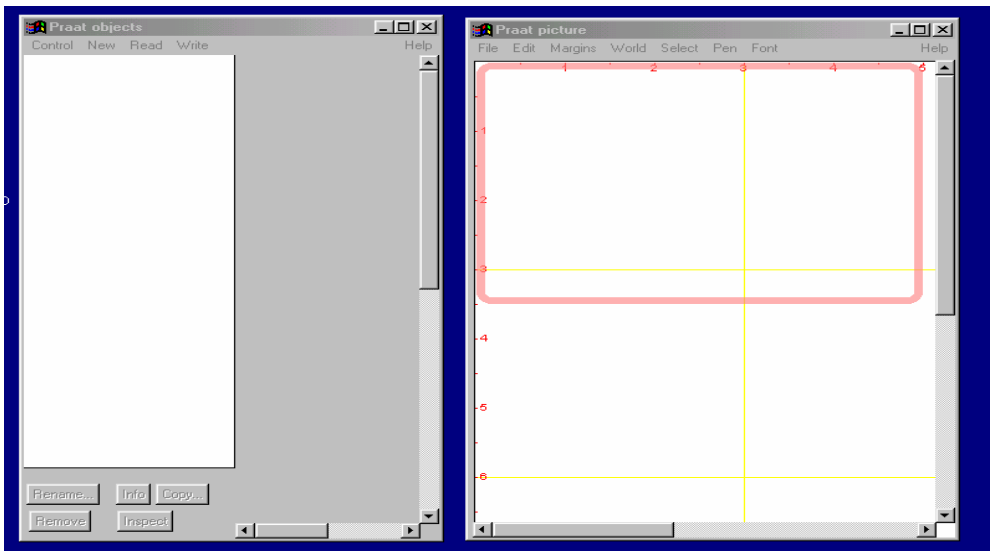


Figure 3. *Écran à l'ouverture de PRAAT*

La fenêtre de gauche est intitulée "**Praat objects**" et sert à "lister" les différents objets (fichiers sons, fichiers d'annotation, etc.) à partir desquels sont effectuées les analyses ou qui en sont le résultat. La fenêtre de droite, intitulée "**Praat picture**", est utilisée pour reproduire des figures (sonagramme, courbe de F0, etc.) qui pourront être exportées vers d'autres logiciels (traitement de texte, etc.).

Les menus déroulants [Control], [New], [Read], [Write] et [Help] sont accessible à partir de la fenêtre "**Praat objects**". Une description des fonctions accessibles par ces menus est donnée ci-après :

	Descriptif des fonctions accessibles à partir des menus déroulants
[Control]	Édition de scripts, Lancement de scripts, Configuration (taille des buffers, etc.),
[New]	Création de fichiers divers (enregistrement audio, création de sons, fichiers d'annotation), création de grammaires OT, création de sons par synthèse articulatoire, etc.
[Read]	Ouverture de fichiers existants (annotation, fichiers audio, fichiers d'annotations effectuées sous XWaves, etc.)
[Write]	Sauvegarde des fichiers sous des formats particuliers (fichiers binaires, fichiers texte, etc.)
[Help]	Accès au manuel et à plusieurs tutorials (introduction à Praat, introduction à l'édition de scripts, etc.), accès à des informations pratiques (FAQ, Éditeurs, etc.).

Il est important de comprendre que tout le fonctionnement de PRAAT est basé sur la notion d'objet : les fichiers audio, les fichiers de transcription et les manipulations du signal sont des objets. Dès qu'ils sont ouverts ou créés par l'utilisateur, ils apparaissent dans une fenêtre listant les objets ("**Praat objects**") et peuvent être sélectionnés afin que des opérations soient effectuées. Ainsi, par exemple, pour transcrire un fichier son, l'utilisateur doit tout d'abord appeler ce fichier. Ensuite, il doit sélectionner cet objet "son" dans le fenêtre "**Praat objects**" et, une fois la sélection effectuée, il doit cliquer sur le bouton "Annotate" et choisir "to Textgrid" afin de créer un fichier de transcription appelé TextGrid. Celui-ci figurera alors dans la liste des objets et pourra être appelé pour être visualisé.

Passons maintenant au second point mentionné précédemment, à savoir les fichiers d'annotation. Toutes les annotations et les segmentations effectuées sont PRAAT sont enregistrées dans un format propriétaire appelé TextGrid. Elles sont visualisables sous PRAAT parallèlement au signal et à d'autres courbes (F0, intensité, etc.) dans une fenêtre :

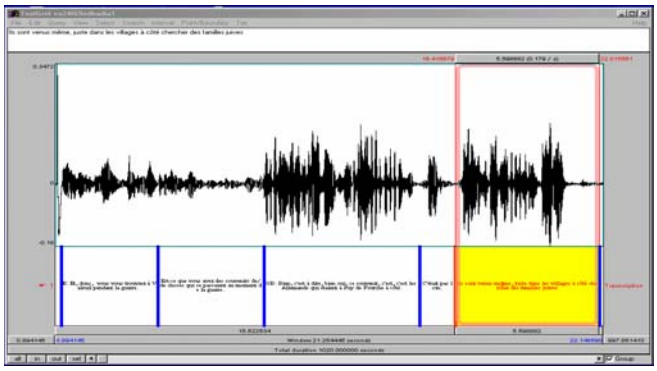


Figure 4. *Fenêtre de segmentation, de transcription et d'alignement*

Les informations continues dans les fichiers d'annotation sont encodées dans un format textuel. Sont indiqués pour chaque intervalle créé les temps de début et de fin, ainsi que l'étiquette associée à l'intervalle. Le fichier *TextGrid* correspondant à la transcription présentée ci-dessus est donné dans la figure 5 qui suit.

Sous PRAAT, il est possible de faire des requêtes sur le contenu des fichiers TextGrid. En revanche, ce format ne peut pas être utilisé avec d'autres outils d'annotation linguistique comme par exemple les étiqueteurs morphologiques ou les

analyseurs syntaxiques. Cela constitue sans aucun doute l'une des limites de PRAAT. Cependant, il suffit d'un petit script pour transformer un TextGrid en fichier texte lisible par une application de ce genre, ou inversement de transformer la sortie d'un étiqueteur en fichier TextGrid.

```
File type = "ooTextFile"
Object class = "TextGrid"
xmin = 0
xmax = 1020
tiers? <exists>
size = 1
item []:
    item [1]:
        class = "IntervalTier"
        name = "Transcription"
        xmin = 0
        xmax = 1020
        intervals: size = 51
        intervals [1]:
            xmin = 0
            xmax = 1.1406789654355898
            text = ""
        intervals [2]:
            xmin = 1.1406789654355898
            xmax = 4.8732672929524963
            text = "E: Et, donc,  vous vous trouviez à Valeuil pendant la guerre. "
        intervals [3]:
            xmin = 4.8732672929524963
            xmax = 8.9946669045857472
            text = " Est-ce que vous avez des souvenirs de/ de choses qui se passaient
au moment de la guerre."
        intervals [4]:
            xmin = 8.9946669045857472
            xmax = 15.034119251845809
            text = "OB: Bien, c'est à dire, bien oui, ce souvenir, c'est, c'est les
Allemands qui étaient à Puy de Fourche à côté."
```

Figure 5. *Transcription au format TextGrid (format PRAAT)*

2.3. Synthèse et comparaison[4]

Dans cette section, nous allons présenter sous forme de tableau les spécificités des deux logiciels. Cela permettra de mettre en avant leurs différences, ainsi que leurs avantages. Pour faire cette synthèse, nous nous intéressons plus particulièrement aux points suivants : aspects généraux, tâche d'enregistrement et de lecture des fichiers audio, tâche d'annotation, tâche d'étude des paramètres prosodiques et de morphing.

Fonction évaluée	WINPITCH	PRAAT
Généralités		
Plateforme acceptée	Windows	Windows, Unix, Mac. Logiciel multiplateforme
Ergonomie	Interface utilisateur très conviviale	Interface utilisateur déroutante au premier abord.
Possibilités et fonctionnalités	Logiciel permettant: - l'annotation de corpus ; - l'étude des paramètres prosodiques ; - l'étude des phénomènes segmentaux ;	Logiciel permettant : - l'annotation de corpus ; - l'étude des paramètres prosodiques ; - l'étude des phénomènes segmentaux ;

[4] La rédaction de la section 2.3 a été faite par Elisabeth Delais-Roussarie.

	- la formulation de requêtes contextuelles sur des données alignées.	- la construction de grammaire OT ; - l'écriture de scripts pour effectuer des tâches, etc.
Lecture et enregistrement des fichiers sonores		
Formats acceptés en entrée et en sortir	Sur ce point, les deux logiciels sont assez comparables : ils acceptent en entrée comme en sortie un nombre important de formats audio (wav, aiff, etc.). WinPitch accepte aussi des formats vidéo.	
Gestion des fichiers sons volumineux	Les deux logiciels permettent de travailler sur des fichiers audio volumineux.	
Tâche d'enregistrement	WinPitch permet de faire des enregistrements avec visualisation en temps réel des différentes courbes (signal, F0, etc.), d'où une possibilité de monitoring.	Des enregistrements peuvent être faits sous PRAAT mais l'interface est peu ergonomique. Pas de monitoring précis.
Tâche de lecture	Possibilité de lire tout ou partie des fichiers audio. Gestion aisée de cette tâche avec des touches sur lesquelles il faut cliquer Lecture à plusieurs vitesses, ce qui est intéressant en cas d'alignement ou de transcription	Lecture possible de la totalité ou d'une partie du fichier audio. L'interface proposée pour cette tâche n'est pas très ergonomique.
Segmentation et annotation de données audio		
Tâche de transcription et d'annotation alignée	WINPITCH permet : - d'aligner et d'annoter sur plusieurs tires ; - d'aligner à la volée si la transcription a déjà été faite sous un autre format (texte, fichier CLAN, etc.) ; - de faire des requêtes sur les fichiers annotés.	PRAAT permet : - d'aligner et annoter sur plusieurs tires ; - d'effectuer des requêtes sur les étiquettes assignées aux intervalles. En revanche, aucune possibilité d'alignement à la volée.
Fichiers acceptés en entrée	WinPitch accepte les fichiers de transcription de Transcriber (format .trs) et en lecture-écriture de Praat (format TextGrid)	PRAAT accepte en lecture les fichiers d'annotation faits sous XWaves.
Fichiers obtenus en sortie	Les annotations et transcriptions peuvent être sauvegardées sous un format propriétaire (wp2), sous format Praat (TextGrid) et sous XML	Les annotations peuvent uniquement être sauvegardées sous le format propriétaire de Praat, TextGrid.

Étude des paramètres prosodiques		
Paramètres pris en compte	Les deux logiciels permettent d'étudier : - les paramètres de fréquence fondamentale (plusieurs méthodes d'extraction de F0, dont AMDF, Autocorrélation, etc.) ; - l'intensité ; - la durée. Pour F0, plusieurs unités peuvent être choisies pour visualiser la courbe.	
Modifications des paramètres	Sous WinPitch, il est possible de modifier les paramètres prosodiques à l'aide de la souris (modification des courbes, abaissement, etc.). La courbe résultant des modifications peut être écoutée en analyse-synthèse.	Il est possible de modifier les paramètres prosodiques (détermination de points cibles, simplification des courbes, etc.) et d'écouter, grâce à l'analyse synthèse, les courbes modifiées. L'interface de PRAAT pour effectuer ces tâches est moins conviviale que celle de WINPITCH.

3. Annotation prosodique automatique

3.1. PROSOGRAPHE et prosogramme[5]

Le PROSOGRAPHE est un outil pour la transcription de la prosodie dans la parole, plus particulièrement pour la transcription des aspects mélodiques. La représentation obtenue, appelée prosogramme, affiche la courbe stylisée des variations mélodiques audibles dans la parole analysée. Cette courbe est alignée avec la transcription phonétique et éventuellement avec d'autres couches d'annotation, selon le choix de l'utilisateur.

La **particularité** du prosogramme, par rapport à d'autres approches de stylisation, réside dans le fait que la stylisation repose sur un **modèle de la perception tonale** chez l'auditeur humain. Les variations mélodiques inférieures au seuil de perception (seuil de glissando) apparaissent comme des traits plats, les variations au-dessus du seuil comme des traits inclinés ou éventuellement des courbes (en cas de séquence de mouvements de pente différente). Afin de faciliter l'interprétation des intervalles mélodiques, ces données sont affichées sur une échelle musicale en demi-tons, similaire à la portée musicale. Ainsi on obtient une transcription à la fois objective et quantifiée des phénomènes mélodiques perçus par l'auditeur moyen et susceptibles de jouer un rôle dans la communication parlée. Les données quantitatives sont récupérées dans un fichier de sortie, en vue de leur utilisation dans d'autres applications (comme la resynthèse ou les manipulations de contours).

L'utilisation du modèle perceptif, alliée à la segmentation phonétique, résulte en une transcription lisible, qui efface les variations inaudibles et les phénomènes de microprosodie co-intrinsèques. Grâce à la simulation de la perception tonale, on obtient une représentation mélodique générale, et non pas spécifique pour telle ou telle langue, et en même temps indépendante de toute théorie linguistique de l'intonation. Le but est en effet de fournir une représentation de l'image auditive, plutôt que de donner la représentation du

[5] La section 3.1 a été faite par Piet Mertens.

contour dans le cadre de tel ou tel modèle phonologique. Depuis la distribution de l'outil, la méthode a été appliquée au français, à l'anglais, à l'allemand, à l'italien, au néerlandais, au turc, parmi d'autres.

La stylisation suppose une **segmentation** du signal de parole en noyaux syllabiques ou en unités de taille similaire. Plusieurs types de segmentation sont utilisés.

1. Dans la **segmentation en noyaux vocaliques**, on délimite le noyau d'intensité élevée à l'intérieur de chaque voyelle de la transcription phonétique. Dans cette approche, le choix de la voyelle au détriment de la syllabe est motivé par le fait que les règles de syllabation peuvent varier d'une langue à l'autre (à cause des consonnes syllabiques par exemple), et qu'il est par conséquent impossible d'arriver à une syllabation universelle. Cependant, ce type de segmentation présente l'inconvénient majeur du temps requis pour l'alignement phonétique manuel (environ 90 minutes pour une minute de parole).
2. Pour remédier à cet inconvénient, on a prévu la possibilité d'utiliser la **segmentation existante**, fournie par une **application externe** (alignement par reconnaissance automatique, segmentation automatique en syllabes), qui doit alors être reprise dans le fichier TextGrid fourni en entrée.
3. La **segmentation automatique** présente des avantages évidents. Un type particulier fournit une segmentation **en sommets de sonie**, basée sur l'évolution du paramètre de sonie (loudness). Elle est automatique et donne des résultats assez proches d'une segmentation en noyaux syllabiques. Vu qu'elle élimine tout le travail de segmentation manuelle, les résultats sont tout à fait acceptables. Cette méthode de segmentation sera distribuée sous peu, et elle est déjà disponible dans le cadre de projets de recherche communs.

Le prosographe permet plusieurs **variantes de visualisation**, selon la taille du graphique et le nombre de données affichées. La version compacte est conçue pour la transcription de corpus entiers. Comme elle indique également la calibration de l'axe du temps et de l'axe de la hauteur, la version standard ("wide") convient mieux aux analyses d'extraits courts. Selon la quantité de données affichées, on distingue deux variantes.

1. La version "simple" ne montre que la stylisation, le voisement et les frontières de portions stylisées.
2. La version "riche" affiche également les paramètres de fréquence fondamentale et d'intensité, ce qui permet de vérifier le traitement.

L'outil détecte les changements abrupts de fréquence fondamentale, qui sont le plus souvent le résultat d'une erreur de détection (saut d'octave, creaky voice). Ces discontinuités sont signalées par un signe dans le prosogramme (un x dans un petit cercle). Comme ces sauts entraîneraient une erreur de stylisation, l'intervalle temporel à styliser est raccourci à l'instant de la première discontinuité.

Il est possible d'ajuster le seuil de glissando utilisé dans la stylisation. Pour une discussion sur le choix du seuil, nous renvoyons à l'article de Mertens (2004) "Un outil pour la transcription de la prosodie dans les corpus oraux". *Traitement Automatique des langues* 45 (2), 109-130. Le PROSOGRAPHE fournit comme **sortie** les données suivantes.

1. Un ou plusieurs fichiers graphiques (au format EPS, encapsulated Postscript) qui contiennent la stylisation obtenue, alignée avec la transcription phonétique, et éventuellement d'autres couches d'annotation, selon le choix de l'utilisateur. Les portions stylisées (noyaux syllabiques) et les régions

voisées sont également indiquées. Les formats d'affichage (cf. supra) varient selon le choix de l'utilisateur.

2. Un fichier (au format PitchTier de Praat) qui contient la stylisation (valeurs de fréquence en Hz) qui peut être utilisée (sous Praat, par exemple) en resynthèse ou pour des manipulations de courbe mélodique.

La **plage de hauteur** affichée (en demi-tons) s'ajuste automatiquement suivant les valeurs de fréquence fondamentale dans le signal de parole. Cependant l'utilisateur garde la possibilité de fixer manuellement ces valeurs.

L'utilisateur peut sélectionner dans le fichier d'**annotation** en entrée (TextGrid) les couches ("tiers") qui seront affichées avec la stylisation : par exemple, la transcription orthographique, la notation en tons, les syllabes, l'accentuation, etc.

L'outil d'annotation a été réalisé comme un **script** pour le logiciel Praat, qui peut être considéré comme un des logiciels d'analyse phonétique les plus réussis. Actuellement ce logiciel est utilisé par de nombreux phonéticiens et linguistes. Il est disponible sur les systèmes de gestion majeurs (Windows, Mac, Linux). Il accepte la plupart des formats de fichier son utilisés en phonétique expérimentale et en traitement du signal. Les résultats d'analyse (paramètres acoustiques, décision de voisement, segmentation, stylisation, etc.) peuvent être sauvegardés dans des fichiers et récupérés dans d'autres étapes de traitement (resynthèse, manipulations des variations mélodiques, de la durée, segmentation corrigée, etc.). Grâce aux scripts (programmes qui combinent des traitements prévus dans le logiciel), il est possible de réaliser des traitements complexes ou de les appliquer à un nombre élevé de données (fichiers).

Des **outils annexes** permettent de faciliter l'analyse ou de l'appliquer à des données provenant de logiciels différents. Rappelons que l'outil de base permet déjà d'analyser plusieurs fichiers sons d'affilée, sans intervention de la part de l'utilisateur.

1. Calcul des paramètres pour un ensemble de fichiers (utilisation de wildcard).
2. Conversion de fichiers d'alignement phonétique ("labeling") en fichier TextGrid.

Parmi les extensions envisagées du prosographe, il y a la détection automatique des pauses et et le calcul de la proéminence syllabique.

À titre indicatif, nous fournissons ci-après les deux prosogrammes faits pour un même fichier audio contenant un extrait de la lecture du texte présenté aux locuteurs dans le cadre du projet PFC. Dans le premier cas, le prosogramme a été calculé à partir d'une segmentation manuelle au niveau phonémique, alors que dans le second cas, la segmentation des noyaux syllabiques a été faite automatiquement.

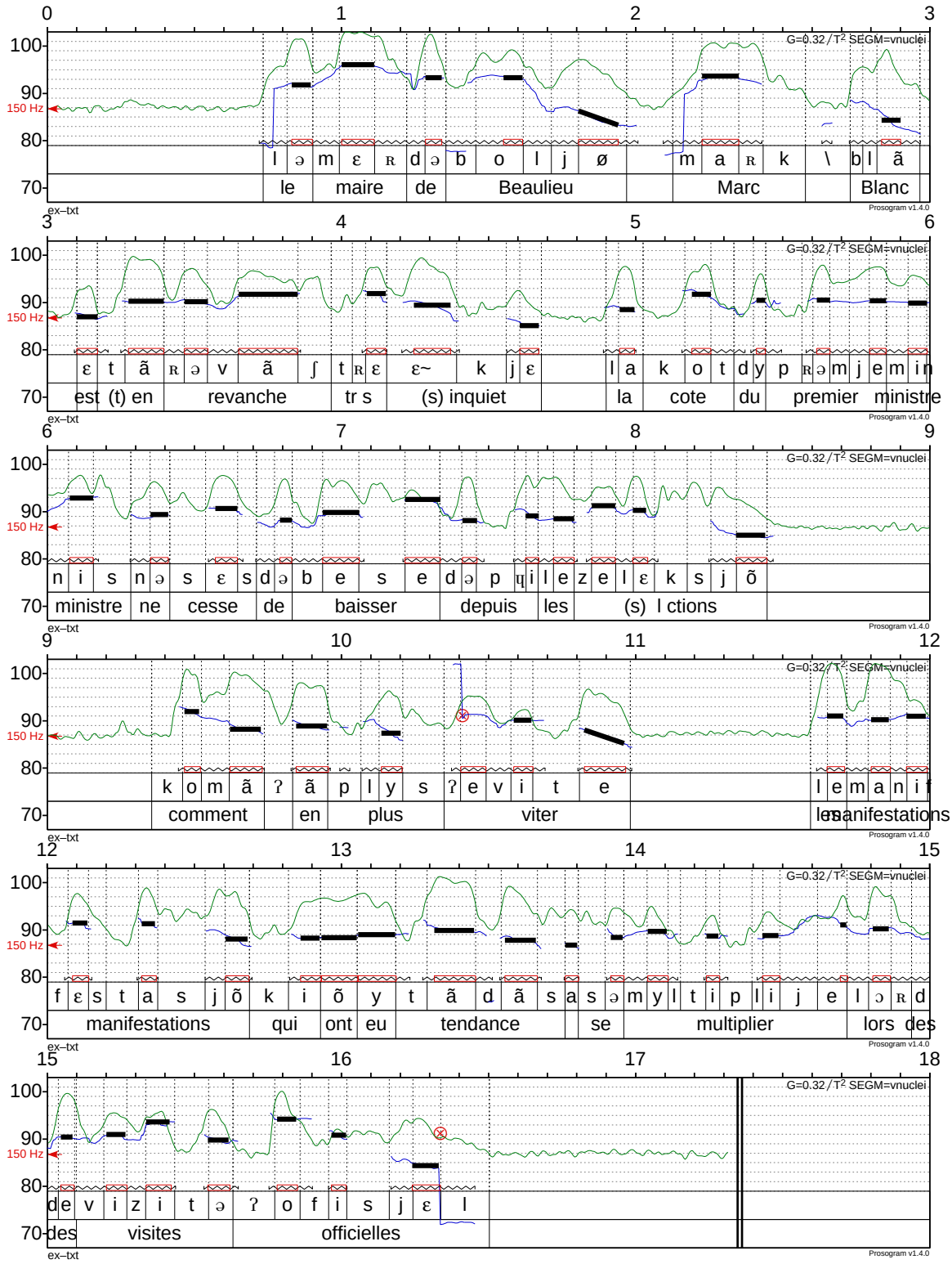


Figure 6. *Prosogramme fait à partir d'une segmentation manuelle*

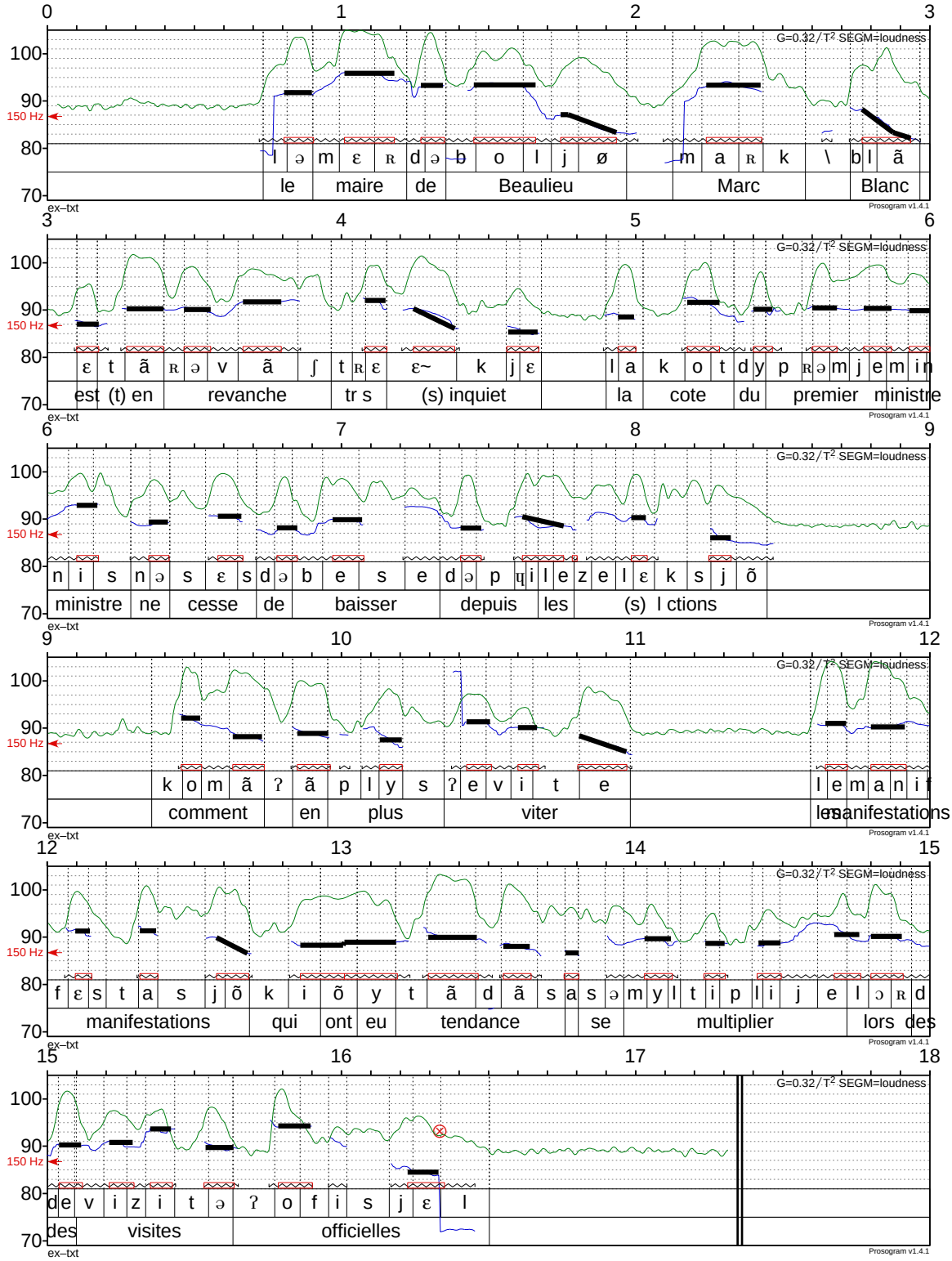


Figure 7. *Prosogramme fait à partir d'une segmentation automatique*

3.2. MOMEL-INTSINT[6]

Nous présentons ici une approche qui permet d'encoder automatiquement et de façon réversible les informations prosodiques, et plus particulièrement mélodiques, contenues dans le signal de parole. Cette approche se base sur une distinction entre quatre niveaux d'analyse des phénomènes mélodiques :

1. le niveau physique (ou acoustique) qui contient des informations très riches et continues (valeurs continues des paramètres prosodiques).
2. le niveau phonétique qui est conçu comme une copie du signal obtenue automatiquement à partir de la détermination de points cibles par MOMEL.
3. le niveau phonologique de surface où les points cibles déterminés par MOMEL sont encodés de façon discrète avec INTSINT.
4. le niveau phonologique profond qui assure le lien entre la représentation phonologique d'un énoncé et les autres niveaux de description linguistique (syntaxe, sémantique, etc.).

Les niveaux phonétique et phonologique de surface peuvent être dérivés automatiquement grâce à des programmes utilisables sous Praat et téléchargeables à l'adresse suivante : http://aune.lpl.univ-aix.fr:16080/~auran/francais/index.html. Dans ce document, nous allons brièvement expliquer comment fonctionnent ces deux outils, à savoir MOMEL et INTSINT.

3.2.1. MOMEL

Différents modèles ont été proposés pour modéliser et styliser les courbes de fréquence fondamentale. L'algorithme MOMEL (cf. Hirst et Espesser 199), Hirst et al. 1997) décompose la courbe de fréquence fondamentale brute en deux types d'éléments :

- les éléments micro-prosodiques qui correspondent aux variations mélodiques à court terme liées à la nature des segments ;
- les éléments macro-prosodiques qui rendent compte des variations mélodiques à plus long terme. Les courbes macro-prosodiques sont modélisées en utilisant une fonction spline quadratique.

L'algorithme MOMEL fournit en sortie une séquence de points cibles qui correspondent à des cibles linguistiquement pertinentes. Un exemple est donné ci-dessous.

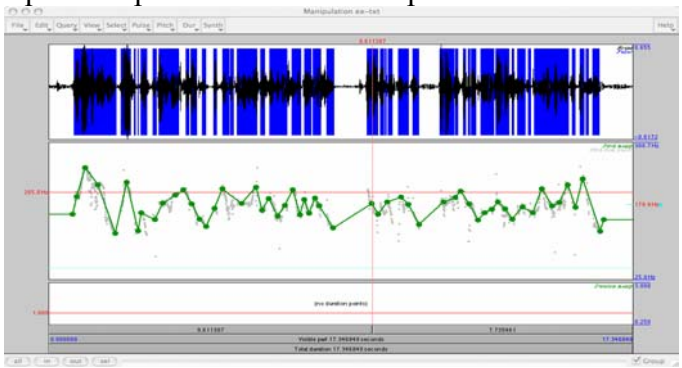


Figure 8. *Manipulation de F0 par l'algorithme MOMEL*

[6] Section rédigée par Elisabeth Delais-Roussarie, en collaboration avec Daniel Hirst (en ce qui concerne Intsint).

3.2.2. INTSINT et la représentation phonologique de surface

Les points-cibles modélisés par l'algorithme MOMEL peuvent être interprétés de différentes façons. Ils peuvent servir à générer la courbe de fréquence fondamentale à partir d'une interpolation par fonction spline quadratique (sur ce point, voir Mixdorff & Fujisaki 2000). Dans la figure ci-dessous, la courbe proposée a été générée à partir des points cibles déterminés par MOMEL (figure précédente) par une interpolation.

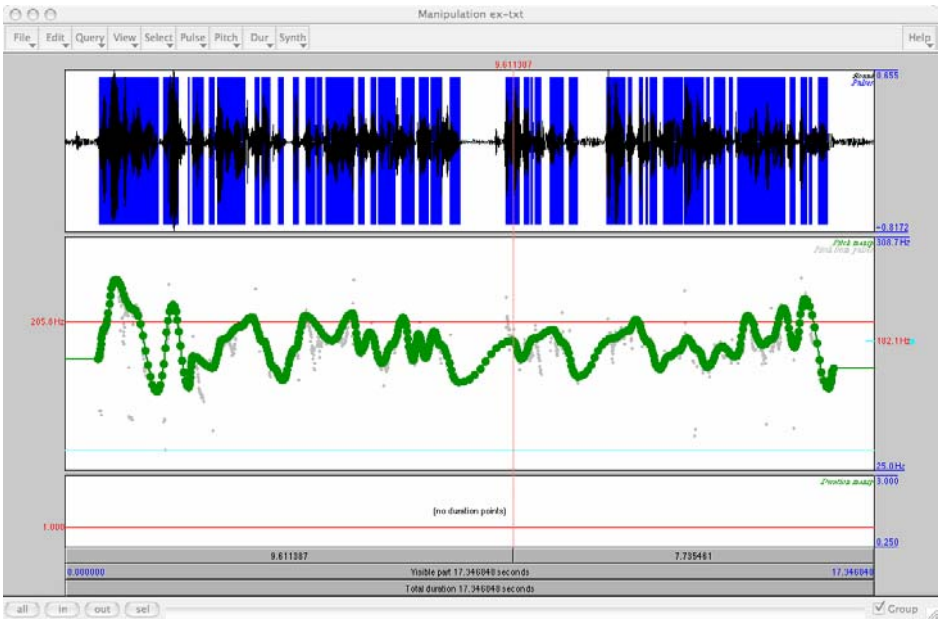


Figure 9. *Courbe générée par MOMEL à partir des points cibles*

Il est également possible d'utiliser ces points pour obtenir une représentation discrète de la courbe de fréquence fondamentale. Cela peut se faire en attribuant comme étiquette à ces points des symboles phonologiques extraits de l'alphabet INTSINT (cf. Hirst et Di Cristo 1998). Le système de codage INTSINT repose sur une distinction entre :

- des valeurs définies globalement relativement au registre de chaque locuteur, à savoir : Top (T), Mid (M) et Bottom (B) ;
- des valeurs définies localement relativement à la valeur associée au point cible précédent : Higher (H), Same (S) et Lower (L) ;
- des valeurs permettant de rendre compte de changements mélodiques de faible ampleur. Elles sont définies relativement aux valeurs et points cibles précédents : Upstepped (U) et Downstepped (D).

La représentation phonologique de surface encodée à l'aide des symboles Intsint peut être générée automatiquement à partir de l'extraction des points cibles effectuée par MOMEL. Si les calculs sont faits sous PRAAT à l'aide des scripts téléchargeables, la représentation phonologique INTSINT prend la forme d'une tire dans un fichier d'annotation (TextGrid).

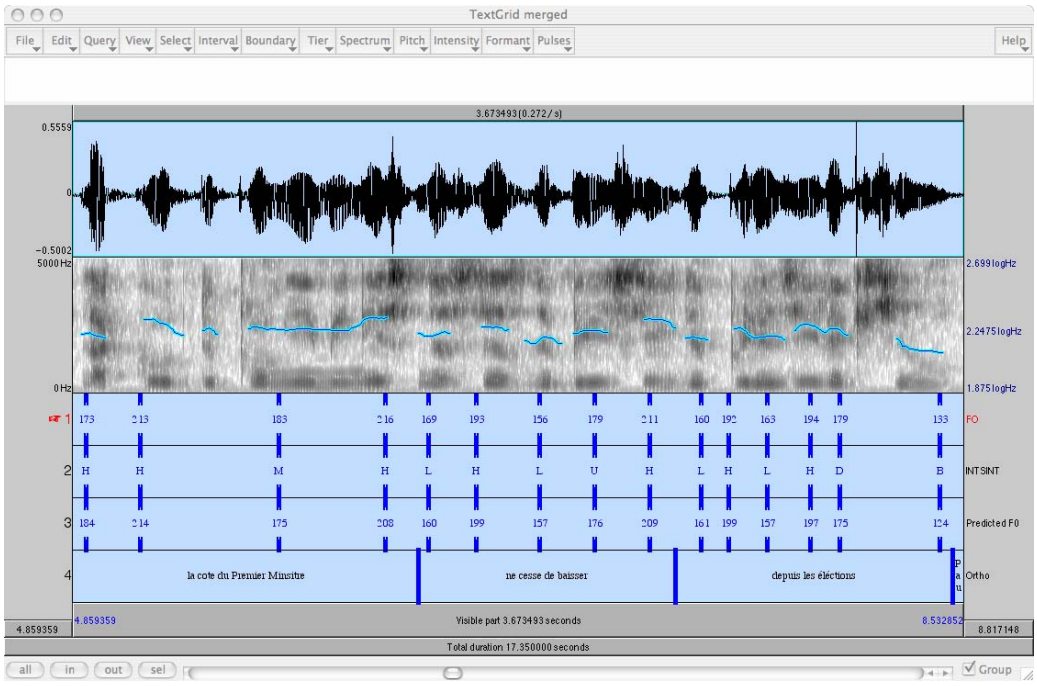


Figure 10. *Codage symbolique INTSINT dans un TextGrid*

3.3. MELISM[7]

MELISM est une procédure de codage discret de la mélodie qui a été récemment développée et est intégrée sous PRAAT sous la forme d'un script (Caelen-Haumont et Auran 2004). Ce codage permet d'analyser avec précision les excursions et les proéminences mélodiques se réalisant sur des séquences sonores courtes (mot ou suite de mots). Pour fonctionner, MELISM suppose :

- une segmentation préalable du fichier audio en unités linguistiques pertinentes (mots, mots prosodiques, etc.) sur lesquelles l'étude des excursions mélodiques se fera,
- une stylisation de la courbe de fréquence fondamentale par détermination de points cibles effectuée par l'algorithme MOMEL (cf. § 3.2.1).

MELISM partage l'amplitude tonale du locuteur en 9 niveaux (niveaux absolus exprimés en demi-tons) en partant de la courbe de fréquence fondamentale stylisée avec MOMEL. Ces niveaux sont ensuite utilisés pour encoder de manière discrète les excursions et proéminences mélodiques. Estimant que ces proéminences mélodiques n'ont pas été suffisamment explorées avec des outils adéquats, nous avons reconsidéré la division du registre du locuteur selon 4 niveaux, usuelle depuis Delattre (1966). Pour ce faire, nous avons en fait divisé chaque niveau en 3, le "cœur" représentant 50% du niveau originel (modèle de Delattre), chacune des marges inférieure et supérieure, 25%. En additionnant 2 à 2 ces 25%, nous avons donc une succession de 7 plages de 50% (échelle logarithmique), et avec les marges extrêmes de 25% (niveau le plus aigu et niveau le plus grave du registre du locuteur), nous arrivons ainsi à 9 niveaux (tableau 1) : aigu (a), supérieur (s), haut (h), élevé (e), moyen (m), centré (c), bas (b), inférieur (i), grave (g). Les cases en jaune correspondent aux séquences tonales qui identifient un mélisme (Caelen-Haumont, 2004).

[7] Cette section a été rédigée par Geneviève Caelen, avec la collaboration d'Elisabeth Delais-Roussarie.

Tableau 1. *Matrice des séquences tonales pour la description des configurations mélodiques des mots, en particulier les proéminences mélodiques subjectives (ou mélismes) sur fond jaune, ayant pour corrélat mélodique, les niveaux les plus aigus*

ton	Mélismes aigu a	supra s	haut h	élevé e	moyen m	centré c	bas b	infra c	grave g
a	aa	as	ah	ae	am	ac	ab	ai	ag
s	sa	ss	sh	se	sm	sc	sb	si	sg
h	ha	hs	hh	he	hm	hc	hb	hi	hg
e	ea	es	eh	ee	em	ec	eb	ei	eg
m	ma	ms	mh	me	mm	mc	mb	mi	mg
c	ca	cs	ch	ce	cm	cc	cb	ci	cg
b	ba	bs	bh	be	bm	bc	bb	bi	bg
i	ia	is	ih	ie	im	ic	ib	ii	ig
g	ga	gs	gh	ge	gm	gc	gb	gi	gg

La figure suivante montre un textgrid de PRAAT avec diverses tires propres à la procédure MELISM, avec respectivement de haut en bas, la fenêtre “manipulation” de PRAAT, comportant le signal de parole, la courbe mélodique stylisée, puis le codage de la procédure MELISM, présentant de bas en haut :

- les valeurs de F0 ;
- leur conversion en demi-tons ;
- le codage en cibles tonales selon la procédure MELISM (premier passage avant segmentation) ;
- le recodage au sein de la séquence segmentée, quelle que soit sa nature : mot, groupe, énoncé ... (deuxième passage), et détermination de la cible la plus haute au sein de l’unité segmentée ;
- l'annotation en séquences mono- ou bitonales (les “syllabes tonales”). Cette tire peut constituer une tire d’annotation mélodique à l’usage des utilisateurs de PFC ;
- la segmentation en mots.

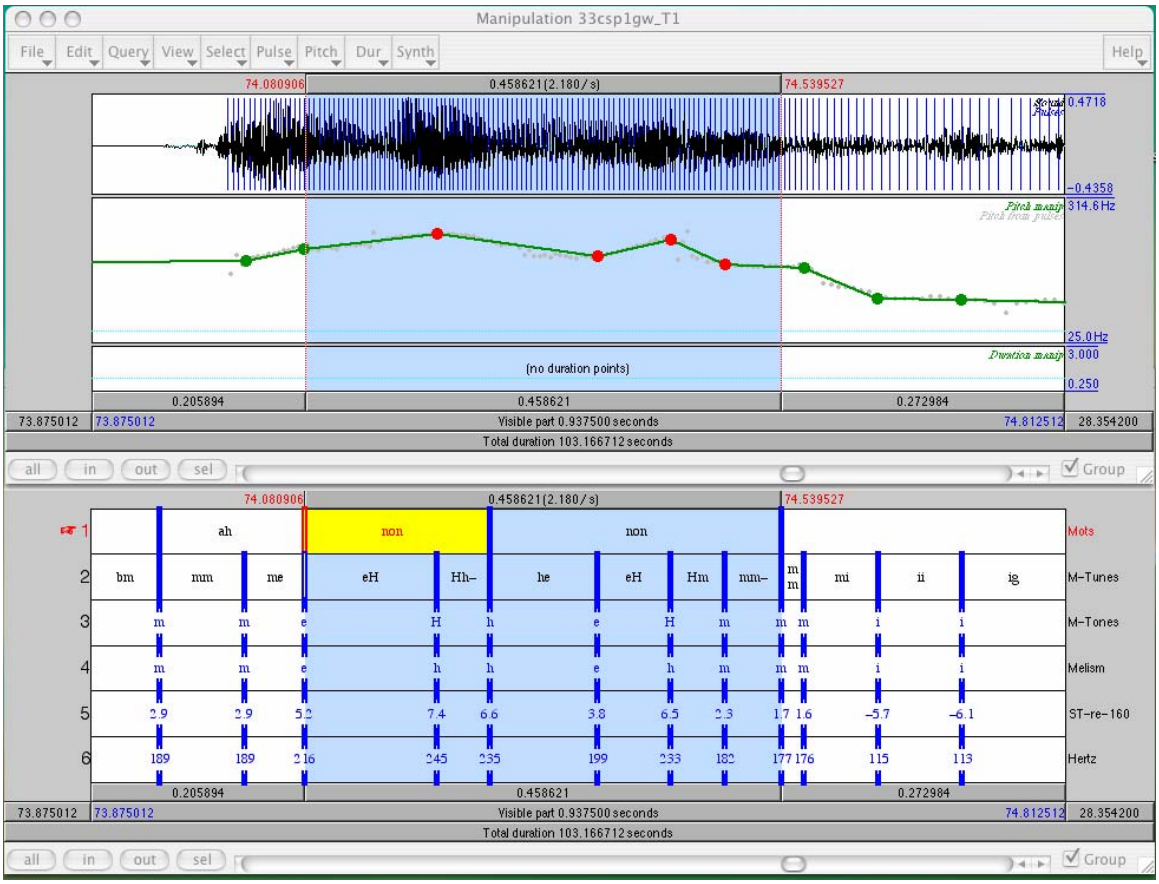


Figure 11. *Codage MELISM et TextGrid*

3.4. Synthèse[8]

Les trois systèmes automatiques d'annotation prosodiques que nous venons de présenter partagent certaines caractéristiques :

- ils fonctionnent tous trois sous PRAAT sous forme de script ;
- ils permettent de modéliser la courbe de fréquence fondamentale de façon à générer une analyse phonétique au sens de Hirst et al, 1998b, c'est-à-dire une représentation simplifiée de la courbe réelle ;

En revanche, ils se distinguent les une des autres sur certains points :

- MOMEL-INTSINT et MELISM proposent, contrairement au PROSOGRAPHE, une représentation phonologique de surface sous la forme d'un codage symbolique discret. Bien que les prosogrammes puissent servir de base à la génération d'une représentation phonologique de surface du type de celle défendue par Mertens (1987 et sq.), ils n'ont pas la forme d'un encodage phonologique discret ;
- le prosographe propose une modélisation sur base perceptive, contrairement à MOMEL-INTSINT et à MELISM qui construisent leur modélisation sur des bases acoustiques. Dans certains cas, les résultats obtenus diffèrent comme l'a montré Olivier (2005) ;

[8] Cette partie a été rédigée par Elisabeth Delais-Roussarie.

- le PROSOGRAPHE et MELISM supposent une segmentation préalable du continuum sonore en unités linguistiques pertinentes (mots prosodiques, mots, etc.), contrairement à MOMEL-INTSINT.

4. Conclusion

Dans cette contribution, nous avons présenté les caractéristiques de deux outils logiciels pouvant aider à l'annotation orthographique et prosodique de corpus oraux, PRAAT et WINPITCH ; puis nous avons expliqué comment fonctionnent trois algorithmes de stylisation et d'encodage de la courbe mélodique, le prosographe, MOMEL-INTSINT et MELISM.

Parmi les deux outils logiciels, WINPITCH offre certains avantages par rapport à Praat : (i) une ergonomie plus grande ; (ii) des fonctionnalités intéressantes pour effectuer un alignement entre transcription et signal de parole.

Ceci étant, PRAAT permet, au moyen de scripts, d'effectuer de nombreuses tâches comme des stylisations de F0, des conversions de fichiers, etc. Aussi, nous pensons que l'utilisation des deux outils en parallèle est très recommandée.

Références

Alessandro d', C. et P. Mertens (1995). "Automatic pitch contour stylization using a model of tonal perception". *Computer Speech and Language* 9(3), 257-288.

Caelen-Haumont, G. et B. Bel (2000). "Le caractère spontané dans la parole et le chant improvisés : de la structure intonative au mélisme". *Revue Parole* 15-16, 251-302.

Caelen-Haumont, G. (2004). "Valeurs pragmatiques de la proéminence prosodique lexicale: de l'outil vers l'analyse". Actes des *Journées d'Étude sur la Parole (JEP)*, 19-22 avril 2004, Fès, Maroc, 105-108.

Caelen-Haumont, G. et C. Auran (2004). "The phonology of melodic prominence: the structure of melisms". Actes de *Speech Prosody 2004*, 23-26 mars 2004, Nara, Japon, 143-146.

Caelen-Haumont, G. et C. Auran (2004). "INTSMEL : un outil pour l'analyse des contours proéminents de F0". *Bulletin PFC 3*, 115-125. 2004. Consultable à : http://www.projet-pfc.net/?Accueil%20PFC:Bibliothèque%20PFC:Bulletins%20PFC

Corpus C-ORAL-ROM (2005). Les informations sur le projet sont disponibles sur le site : http://lablita.dit.unifi.it/coralrom

Delais-Roussarie, E., A. Meqqori et J.-M. Tarrier (2003). "Annoter et segmenter des données de parole sous PRAAT". In E. Delais-Roussarie et J. Durand (eds.), *Corpus et Variation en Phonologie*, Toulouse : Presses Universitaires du Mirail, 159-185.

Grabe, E. et B. Post (2002). "Intonational variation in the British Isles". In B. Bel et I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 conference*, 11-13 avril 2002. Aix-en-Provence : Laboratoire Parole et Langage, 343-346.

Hirst, D.J. (2001). "Automatic analysis of prosody for multilingual speech corpora". In E. Keller, G. Bailly, J. Terken et M. Huckvale (eds.), *Improvements in Speech Synthesis*, UK : John Wiley, 320-327.

Hirst, D.J. et A. Di Cristo (1998a). "A survey of intonation systems". In Hirst, D.J., Di Cristo, A. (Eds), *Intonation Systems: a Survey of Twenty Languages*. Cambridge : Cambridge University Press, 1-44

Hirst, D.J., A. Di Cristo et R. Espesser (1998b). “Levels of representation and levels of analysis for the description of intonation systems”. In M. Horne (ed.), *Prosody: Theory and Experiment*, Dordrecht : Kluwer Academic Publishers.

Hirst, D.J. et R. Espesser (1993). “Automatic Modelling of Fundamental Frequency using a quadratic spline function”. *Travaux de l’Institut de Phonétique d’Aix-en-Provence* 15, 75-85.

Hirst, D.J., N. Ide et J Véronis (1994). “Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTEXT project”, *Proceedings of the ESCA/IEEE Workshop on Speech Synthesis*, septembre 1994, New York.

Lieshout van, P. (2004). *PRAAT Short Tutorial: An introduction.* Peut être obtenu sur demande à partir de http://ots.utoronto.ca/users/vanlieshout/

Mertens, P. (1987). *L'intonation du français. De la description linguistique à la reconnaissance automatique*, Thèse de Doctorat, Université de Louvain.

Mertens, P. (2004). “Un outil pour la transcription de la prosodie dans les corpus oraux”. *Traitement Automatique des langues* 45 (2), 109-130.

Mertens, P. (2004). “The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model”. In B. Bel et I. Marlien (eds.) *Proceedings of Speech Prosody 2004*, Nara , Japon, 23-26 mars 2004.

Mertens, P. et Ch. d’Alessandro (1995). “Pitch contour stylization using a tonal perception model”. *Proc. Int. Congr. Phonetic Sciences 13*, Stockholm 1995, 228-231.

Mixdorff, H., & H. Fujisaki. (2000). “Symbolic versus quantitative descriptions of f0 contours in German: Quantitative modelling can provide both”. *Proceedings Prosody 2000: Speech recognition and Synthesis,* octobre 2000, Kraków.

Oliver, D. (2005). “Deriving pitch accent classes using automatic F0 stylisation and unsupervised clustering techniques”. In *Proceedings of Second Baltic Conference on Human Language Technologies*, 4-6 avril 2005, Tallinn, Estonia, 161-166

Portes, C. (2004). *Prosodie et économie du discours : Spécificité phonétique, écologie discursive et portée pragmatique de l’intonation d’implication*, Thèse de Doctorat, Université de Provence.

Simon, A.C. (2004). *La structuration prosodique du discours en français*. Berne : Peter Lang.

Le rôle des facteurs prosodiques dans l'analyse du schwa et de la liaison

Anne LACHERET[1]
Chantal LYCHE[2]

0. Introduction

Toute exploitation de base de données présuppose la mise à disposition d'une documentation minutieuse, enrichie d'une annotation précise où chaque choix effectué se voit explicité et justifié. La base de données PFC (Phonologie du Français Contemporain), conçue pour devenir à terme une base de référence pour la phonologie du français, ne saurait faillir à ces exigences. Nous ne reviendrons pas ici sur les choix méthodologiques profonds (Durand et Lyche 2003) qui sous-tendent son organisation, mais rappellerons brièvement que les objectifs prioritaires de PFC se déclinent en trois points : l'inventaire phonémique des locuteurs, les phénomènes de schwa et de liaison, les incontournables de la phonologie du français. L'inventaire phonémique sera établi à partir d'une liste de mots, puis affiné grâce à la lecture du texte et à travers les différents enregistrements de parole spontanée. Comme préambule à une analyse du schwa et de la liaison, nous avons opté pour une démarche de codage effectué dans PRAAT sur des tires spécifiques et à partir de la transcription graphique (Durand et Lyche 2003). Une fois les systèmes de codage schwa et liaison mis en place, il nous est paru opportun de tirer profit de la flexibilité de PRAAT pour envisager un codage de la prosodie, domaine interdépendant du segmental et soumis par définition à une variation extrême. Malgré les obstacles évidents[3], la notion même de codage prosodique s'est imposée dans la mesure où seul un codage rigoureux et systématique permettra de mettre en exergue l'ensemble des facteurs qui interagissent dans la construction de ce que l'on catégorise à l'aide du label 'prosodie', tout comme le poids de ces facteurs les uns par rapport aux autres. Étant donné la complexité du champ prosodique, nous ne saurions prétendre à une quelconque exhaustivité et nous avons défini deux objectifs pour la prosodie dans PFC : (a) mettre en valeur le lien entre deux phénomènes hautement variationnels (le schwa et la liaison) et la prosodie, et (b) dégager les caractéristiques prosodiques générales de chaque variété de français étudiée dans un but purement descriptif et comparatif. Nous proposons de restreindre notre propos au point (a) en adoptant une approche par codage, dans le même esprit que ce qui a été mis en place pour les phénomènes segmentaux schwa et liaison. Nous justifierons tout d'abord notre ambition de relier le segmental au prosodique d'un point de vue purement phonologique avant d'expliciter notre méthodologie et son cheminement, pour proposer dans une dernière partie un système de codage.

[1] Institut Universitaire de France, Paris, CRISCO, Université de Caen.
[2] Université d'Oslo/Tromsø.

[3] Ces obstacles sont liés à plusieurs types de difficultés : il s'agit tout d'abord de la complexité intrinsèque de la matière à transcrire (Lacheret-Dujour et Beaugendre 1999), mais également de problèmes inhérents au choix même du logiciel PRAAT. Si ce dernier remplit parfaitement sa fonction de logiciel de traitement et d'analyse phonétique, en revanche il n'a pas été conçu pour servir de logiciel de codage, fonction que nous lui faisons portant remplir par l'utilisation de tires parallèles où figurent sur un même plan données et métadonnées (Baude, Jacobson, Tchobanov et Walter 2005).

1. Nature du lien entre la prosodie et le schwa/la liaison : le regard du phonologue

Le schwa et la liaison constituent sans nul doute les deux phénomènes les plus débattus de la phonologie du français, notoriété qui s'explique par l'impact que les solutions envisagées ont eu dans les grands débats théoriques. Schwa et liaison partagent en effet certaines caractéristiques : ce sont deux phénomènes variables, les segments impliqués, qui possèdent normalement un corrélat graphique, sont soit présents soit totalement absents, et les analyses proposées convergent vers des problématiques similaires (insertion, élision, segment flottant ?). Nous avons choisi de limiter ici notre réflexion au schwa, estimant qu'il serait loisible de présenter des arguments proches pour le traitement de la liaison[4]. Nous nous pencherons plus particulièrement sur la question de la variation et tenterons de mettre en exergue la relation entre un ensemble de facteurs prosodiques et la présence/absence de ce segment. Ces facteurs prosodiques ne sont bien évidemment pas déterminants à eux seuls, ils fonctionnent en complémentarité à tout un ensemble de paramètres largement décrits dans la littérature (par ex. Malécot 1976, Hansen 1994, Walker 1996, Racine et Grosjean 2002).

Une approche systématique de la variabilité du schwa présuppose l'élaboration de données statistiques de grande ampleur que seul un codage méthodique des transcriptions permettra de fournir. La méthodologie adoptée dans PFC, explicitée dans Durand et Lyche (2003), préconise un point de vue théorique aussi neutre que faire se peut. Ainsi, par exemple, le codage envisagé ne présuppose aucune représentation spécifique de la voyelle, pas plus qu'il ne fait de prédictions sur sa nature. Il se limite à prendre en compte quelques uns des facteurs qui conditionnent la présence ou l'absence de la voyelle. Le codage de la prosodie se propose tout naturellement d'enrichir cet aspect à l'aide de données nouvelles, mais son intérêt ne se limite pas à cet angle du problème, il s'étend également à celui de la représentation sous-jacente du schwa, sujet qui n'a toujours pas reçu dans la littérature de réponse entièrement satisfaisante. L'objectif de cette section est de proposer un ensemble d'hypothèses que le codage prosodique nous permettra de tester. Ces hypothèses s'appuient sur des considérations qui relèvent de la phonologie segmentale (1.1), mais aussi du rythme (1.2 à 1.4) et vers lesquelles nous nous tournons maintenant.

1.1. La représentation sous-jacente de schwa

Nous ne prétendons pas entreprendre ici un tour d'horizon complet de la question, mais nous limiterons à rappeler quelques-uns des débats en cours. Citons tout d'abord le conflit maintenant classique entre l'élision et l'insertion, deux positions qui s'opposent encore dans des travaux récents (par ex. Fougeron et Steriade 1997 *vs*. Cotté 2001). Les "insertionnistes", qui se considèrent moins abstraits que ceux qui se fondent sur un processus d'élision (Klausenburger 1994), se réclament de Martinet (1969, 1972). Martinet propose que le schwa s'oppose aux autres phonèmes, son occurrence résultant d'une insertion dont la finalité serait d'alléger certains groupes consonantiques. Le schwa se voit ainsi attribuer un rôle de "lubrifiant phonétique"[5]. Dans le cadre de la phonologie générative classique, cette position fut vivement critiquée au regard des difficultés de son implémentation (*p(e)louse* mais pas **p(e)*lace pour *place*)[6], et aussi bien Schane (1968) que Dell (1973[85]) s'opposent à une solution par insertion pour prôner une solution par

[4] Soulignons cependant que les données sont beaucoup plus maigres pour la liaison que pour le schwa.

[5] Pourtant, plutôt que véritablement inséré, ce "lubrifiant phonétique" pourrait être conditionné dans ses apparitions par des contraintes rythmiques. C'est la position d'Angoujard (1997) qui, tout comme Martinet, réfute l'existence de schwas lexicaux.

[6] Voir à ce sujet l'approche de Bibeau (1975) qui montre toutes les limites d'une théorie d'insertion.

troncation. Dell (1973[85]) postule une voyelle sous-jacente, semblable dans sa représentation aux autres voyelles, mais dont la particularité réside dans son alternance systématique avec zéro ou avec la voyelle /ɛ/ (*étiqu**e**tte vs étiqu**e**ter*). Un ensemble de règles ordonnées lui permet de rendre compte de la chute de la voyelle, règles qui présentent un caractère facultatif dans les cas de la variation observée.

Dans le cadre plurilinéaire qui fait suite, les représentations s'enrichissent et le schwa devient une voyelle non ancrée dans le squelette (Tranel 1984, 1987, 1998, 2000, Encrevé 1988), ou encore la voyelle qui, par défaut, remplit un noyau vide (Anderson 1982), la présence à la surface du schwa étant alors déterminée par des règles de syllabation. Cette mise à l'index de la voyelle par le biais du statut spécial qu'elle se voit attribuer a pour but de légitimer le caractère intrinsèquement faible du schwa. Si le schwa est une voyelle différente des autres voyelles dans sa forme de base (une voyelle non ancrée ou un noyau vide), il devrait s'ensuivre que son absence à la surface ne laisse aucune trace prosodique, ce qui ne serait pas le cas si une voyelle pleine venait à être éliminée[7]. Ces traces sont à rechercher soit dans la syllabe qui précède le schwa, soit dans la syllabe qui le suit. Si dans la syllabe qui précède schwa, la voyelle est systématiquement allongée, on pourrait invoquer une forme d'allongement compensatoire, comme on l'observe dans plusieurs variétés de français qui maintiennent l'allongement (Montreuil 2003, Girard et Lyche 2004, Andreassen ce volume). L'absence de marque, par contre, ne nous renseigne aucunement sur la nature du schwa. De même, si la voyelle qui suit un schwa absent porte régulièrement une proéminence, on pourrait conclure à la présence d'une voyelle sous-jacente laissant une marque prosodique. Dans le cas contraire, l'ambiguïté persiste. Le codage de la prosodie devrait faciliter la réponse à la question suivante : **la voyelle qui précède ou qui suit un schwa absent se distingue-t-elle de quelque manière que ce soit d'une voyelle qui suit ou précède une voyelle pleine[8] ?**

1.2. Le codage prosodique, un complément au codage schwa

Toute influence prosodique que l'on pourra déceler dans le comportement du schwa vient s'ajouter à d'autres paramètres linguistiques dont certains sont déjà pris en compte par le codage schwa. Il semble donc naturel d'envisager le codage prosodique dans le même esprit que celui du schwa déjà mis en place dans le projet PFC. Dans la mesure où le travail classique de Dell (1973[85]) a enrichi notre réflexion pour ce dernier codage, il n'est pas inutile d'en rappeler les grands axes. Pour Dell, le schwa est une voyelle sous-jacente /ə/ dont le corrélat graphique est la lettre *e*. Nous adoptons ainsi la position de Dell (Durand et Lyche 2003) et préconisons le codage de toute voyelle graphique *e* sauf si elle se trouve en position interne de mot après un groupe <obstruante-liquide>. On sait en effet que cet environnement a entraîné la lexicalisation des schwas et le coder n'apporterait aucune information. Par contre, tout graphème *e* qui contredit cette prédiction sera évidemment codé, comme ce serait le cas si le schwa de *vendredi,* par exemple, n'était pas prononcé. Mais le codage ne se limite pas aux graphèmes *e*, il concerne également toute consonne finale prononcée, étant entendu qu'un tel contexte constitue un site d'apparition potentielle

[7] Verluyten (1982, 1985) réfute une utilisation diacritique du /ə/ et postule une seule voyelle sous-jacente /œ/. S'inspirant du modèle rythmique de Schane (1979), il analyse l'alternance schwa-zéro en termes de structure rythmique de la phrase composée d'une suite de syllabes faibles et fortes : *souvenir* correspondrait à un rythme syllabique fort-faible-fort.

[8] Une question subsidiaire vient se greffer à cette interrogation : quel est le rôle du groupe consonantique qui résulte de la chute/absence de la voyelle, sa nature est-elle déterminante pour la force de l'accent ? Ces questions illustrent la limite du codage et la nécessité de le compléter à l'aide d'études de laboratoire (cf. *infra*, conclusion).

de schwa, que la voyelle soit présente dans la graphie ou non. Il peut s'agir dans ce cas, soit d'un schwa prépausal comme dans *bonjour-e*, soit de la présence d'un schwa non suivi d'une pause marquée (*donc-e ça a fonctionné comme ça*) et sans montée mélodique.

Pour Dell, un schwa sous-jacent non affecté par la batterie de règles de troncation apparaît à la surface sous la forme d'une voyelle centrale ouverte. Certaines de ces règles ont un caractère obligatoire alors que d'autres sont facultatives, lui permettant par ce biais de rendre compte de la variation observée. Un schwa interne après une seule consonne prononcée chute systématiquement, tout comme un schwa final, alors qu'un schwa de monosyllabe et un schwa initial de mot peuvent tomber ou se maintenir dans les mêmes conditions. En finale de mot après deux consonnes prononcées, un schwa est également susceptible d'apparaître. C'est ce canevas qui a été suivi pour l'élaboration du codage.

Le codage schwa est un codage numérique à 4 chiffres où le premier chiffre rend compte de la présence ou de l'absence de la voyelle ; le deuxième chiffre indique la position de la syllabe dans le mot (initiale, médiane, finale) ou s'il s'agit d'un schwa de monosyllabe ; le troisième chiffre spécifie le contexte gauche (voyelle, consonne, groupe consonantique simplifié, pause) de la syllabe contenant le schwa en question ; le quatrième chiffre s'intéresse à la syllabe qui suit le schwa (syllabe à initiale consonantique ou vocalique, pause faible ou forte). Ce codage couplé au classeur schwa, développé dans le cadre du projet par Meqqori (Meqqori et Durand 2003), fournit d'importantes données statistiques qui ont déjà permis une comparaison de différentes variétés de français. Durand et Eychenne (2004) par exemple étudient les sites les premiers touchés lors du passage d'une variété conservatrice (Douzens), où un maximum de schwas sont maintenus, à une variété innovante (Biarritz), où les schwas sont beaucoup moins stables. Nous survolerons maintenant le comportement de schwa dans l'ensemble des positions qu'il est susceptible d'occuper dans le mot lexical, en tentant de dégager les liens entre ce comportement et les facteurs prosodiques.

1.3. La position finale de mot

Dans la plupart des variétés de français, un schwa final n'est pas réalisé lorsqu'il est précédé d'une seule consonne prononcée, alors qu'un environnement CCə#C favorise dans certaines circonstances l'apparition de la voyelle. Par ailleurs Léon (1966) rapporte une série de tests qui montrent clairement l'importance du facteur rythmique sur la présence d'un schwa final : le schwa final de *porte* est normalement réalisé dans *porte-plume* mais pas dans *porte-parapluie*[9], mot syllabiquement plus lourd. Le codage schwa tel qu'il a été défini n'apporte aucune information sur le rôle que la longueur des constituants peut jouer sur la présence/absence de schwa et pourtant cette relation étroite a également été relevée dans des variétés où le schwa fait preuve de stabilité. Le français méridional maintient tout schwa final à condition que ce dernier apparaisse en contexte prépausal ou qu'il précède un mot à initiale consonantique (Durand, Slater et Wise 1987).

(1) a. *ma mère* /mamɛʁə/
b. *la mère de Paul* /lamɛʁədəpɔl/
c. *sa mère était inquiète* /samɛʁete…/

[9] Léon (1966 : 119) observe très finement : "Le lien qui semble exister entre la chute du E muet et la longueur du mot est un phénomène à rapprocher du raccourcissement des voyelles, constaté à mesure de l'allongement d'un même mot." Les données dont nous disposons devraient nous permettre d'opérer des mesures systématiques de longueur de schwa dans les différents contextes.

En (1c), l'absence du schwa s'explique par l'initiale vocalique de la syllabe qui le suit. Or, Watbled (1991) relève que dans le français de Marseille, un schwa final peut faire surface dans l'environnement de (1c) lorsque le mot suivant est monosyllabique.

(2) a. *robe ample* /rɔbəɑ̃plə/
b. *la reine Amélie* /ʁɛnameli/

La présence du schwa en (2a) garantit un rythme Fort-Faible-Fort (Verluyten 1985) en admettant qu'une syllabe finale de mot soit toujours forte à moins qu'elle ne contienne un schwa. Il semblerait donc que la longueur des constituants représente un paramètre non négligeable pour le maintien ou non du schwa, comme Delattre (1966) l'avait déjà mentionné pour la liaison. **Nous en concluons que tout codage prosodique doit nous permettre de mettre en relation la présence/absence de schwa avec la position effective de l'accent de groupe.**

Un schwa final ne correspond pas systématiquement à une voyelle graphique et certaines variétés sont friandes d'insertion : un tel schwa peut indiquer la fin d'un énoncé (schwa prépausal), auquel cas il jouit d'un rythme particulier (montée mélodique marquée sur la syllabe qui le précède et chute abrupte sur le schwa (Fónagy 1989, Hansen 1997), ou bien il est simplement inséré pour donner plus de poids à un mot grammatical, comme dans *donc-e* ou *avec-e*[10]. Il semble utile de s'interroger sur les caractéristiques de ce schwa final ainsi que sur les facteurs qui le distinguent du *euh* dit d'hésitation. La longueur étant tout naturellement le premier critère que l'on invoque pour assurer cette distinction (Candea 2000), **nous demanderons au codage prosodique de nous permettre non seulement de distinguer les contextes VCə#CV (*avec-e ta sœur*) de VCə#pause/hésitation (*bonjour-e*…) mais de signaler également les voyelles allongées**.

La longueur vocalique n'a pas de caractère contrastif dans la plupart des variétés du français, mais son impact théorique est non négligeable. Fougeron et Steriade (1997) rapportent les résultats d'une expérience qui met en valeur les éléments permettant aux auditeurs de distinguer deux séquences segmentalement identiques comme *pas de rôle* /padrol/ et *pas drôle* /padrol/. Elles relèvent que les auditeurs discriminent difficilement les deux séquences, mais que ceux qui sont les plus performants dans cette tâche de discrimination sont sensibles aux facteurs acoustiques différenciant les deux suites : le contact linguopalatal dans [d] (plus important pour la suite *d'r*), la durée acoustique du [d] (plus long dans *d'r*) et enfin la longueur de la voyelle précédente, le [a] étant plus long devant *drôle*. Ce travail prend tout naturellement sa place au centre du débat sur la nature de la représentation de schwa. En effet, si ces résultats pouvaient être confirmés sur des données spontanées, ils soutiendraient une approche en faveur d'un segment sous-jacent éliminé plutôt qu'une insertion. Les distinctions de durée effectuées par le codage prosodique sont probablement trop grossières pour étudier ce dernier facteur à grande échelle sur de la parole spontanée. Il n'en reste pas moins vrai que la segmentation syllabique nécessaire au codage facilite largement l'analyse de ces mesures de durée et constitue un préalable précieux pour envisager une étude de laboratoire écologique.

[10] Plénat (1993) a montré que le mot minimal en français est un pied qui prend lui-même une forme minimale (bimorique) et une forme maximale (dissyllabique ou trisyllabique).

1.4. La position interne

Le schwa est maintenu dans cette position après une seule consonne prononcée, uniquement dans les variétés du midi ou lorsqu'il est assujetti à un accent focal (*douc**e**ment !*). Puisqu'un schwa interne se caractérise par sa faiblesse et qu'il résiste mal dans les variétés intermédiaires entre le français du midi et celui du nord (Durand et Eychenne 2004), il importe d'étudier sa valeur accentuelle. On suppose évidemment que dans les variétés du midi, un tel schwa ne fera pas l'objet de proéminence, mais encore s'agit-il de le démontrer. Dans les régions où il est absent, on se concentrera sur les traces qu'il est susceptible de laisser et en particulier sur la longueur de la voyelle précédente qui serait soumise à une forme d'allongement compensatoire (Montreuil 2003) : le /a/ de *c**a**peline* serait plus long que celui de *c**a**pitale*. Dans un contexte CCəC, un schwa se maintient et cela sans exception, si le groupe consonantique constitue un groupe d'attaque de syllabe, i.e. un groupe <obstruante–liquide> (OL). Ce caractère régulier de la présence de la voyelle a incité de nombreux chercheurs à ne plus voir là un schwa, mais une voyelle pleine. La situation est cependant plus opaque lorsque l'on envisage des groupes consonantiques autres que OL. Le groupe consonantique LO, par exemple, ne constitue jamais un groupe d'attaque, il est scindé en coda+attaque (*dépar-tement*). Le schwa se maintient généralement dans de tels contextes, mais il perd son caractère obligatoire à mesure qu'il s'éloigne de la fin du mot, site potentiel d'accent (Niedle 1979, Bouchard 1981) : *département* [depaʁtəmɑ̃], mais *départemental* [depaʁtmɑ̃tal][11]. Le codage prosodique doit être en mesure de nous fournir des données fiables, de confirmer ou d'infirmer ces observations, de nous dire dans quelle mesure la longueur de l'item lexical et/ou la distance du schwa de la fin du mot influent sur la présence/absence du segment. Afin d'être à même de calculer cette distance, nous demanderons au codage prosodique d'indiquer la position du schwa dans le mot, mais aussi de **nous renseigner sur le nombre de syllabes effectivement prononcées pour chaque item étudié.**

1.5. La position initiale et schwa de monosyllabe

Ces deux contextes se situent au cœur de la variation du schwa dans la majeure partie des variétés de français, se propageant également dans les variétés du midi (Durand et Eychenne 2004). Dell (1973[85]) invoque des règles facultatives pour rendre compte de cette variabilité dont le fonctionnement n'a pas encore été totalement élucidé, même si de nombreux chercheurs ont mis au jour un assortiment de paramètres, mais sans jamais pouvoir offrir une vision complète de leur interaction. Racine et Grosjean (2002) envisagent par exemple sept facteurs largement décrits dans la littérature, pouvant influencer la prononciation du schwa initial : 1. fréquence lexicale ; 2. fréquence estimée de la prononciation ; 3. environnement consonantique ; 4. force articulatoire ; 5. effacement précédent[12] ; 6. vitesse d'articulation ; 7. importance discursive[13]. Ils reconnaissent qu'aux paramètres explorés dans leur travail viennent s'ajouter d'autres éléments, comme par exemple la classe sociale, le niveau de langue, la prosodie et regrettent que de par leurs conditions

[11] "the further the schwa is from the stressed syllable, the more it tends to be dropped." (Bouchard 1981 : 22). Bouchard rend compte de cette distinction en attribuant un accent secondaire au schwa de *gouvernement* et un accent tertiaire à celui de *gouvernemental.*

[12] "Plus le pourcentage d'effacement du E caduc dans les mots qui précèdent le mot-cible est élevé, plus le taux d'effacement du E caduc dans le mot-cible devrait l'être aussi." (Racine et Grosjean 2002 : 317).

[13] L'expérience montre que seuls les facteurs 4 et 7 n'influent pas sur le comportement de la voyelle. Le facteur 7 s'inspire, entre autres, de Grammont (1914) selon lequel la prononciation du schwa participe à la mise en relief d'un mot.

d'expérience[14], ils ne soient pas en mesure de présenter une analyse complète de tous les facteurs en jeu. Ils soulignent de surcroît la nécessité d'étudier de plus près la longueur de la syllabe lorsque le schwa est prononcé. Sur la base de quelques mesures, ils établissent d'ailleurs l'existence d'une corrélation étroite entre la longueur de la voyelle et sa propension à chuter dans le même contexte : en d'autres termes, plus la fréquence de chute est élevée, plus la voyelle réalisée sera brève. Une telle observation, si elle est confirmée sur un large corpus, conforterait le côté exceptionnel de la voyelle, phonologiquement distincte de toutes les autres voyelles mais également phonétiquement faible lorsqu'elle se réalise. Rappelons que les positions classiques se préoccupent assez peu de la réalisation phonétique du schwa, soulignant simplement qu'il ne s'agit pas d'un schwa phonétique mais plutôt d'une voyelle dont le timbre ne se distingue pas de celui de *seul* ou qui, dans certains contextes, peut être assimilé à celui de *peu.*

Walter (1977) remarque qu'un ensemble de mots maintiennent leur schwa en syllabe initiale et que la prononciation avec schwa tend à se lexicaliser. Hansen (1994) étudie comment la fréquence lexicale, la nature des consonnes entourant le schwa et la morphologie conditionnent cette lexicalisation. Mais la cause même de cette lexicalisation n'a pas été envisagée et la question à laquelle toute étude approfondie se doit de répondre peut être formulée comme suit : pourquoi cette propension à maintenir le schwa se manifeste-t-elle exclusivement en position initiale? En quoi la position initiale se distingue-t-elle des autres positions du mot ? La première réponse à cette interrogation nous est fournie par le développement en français d'un accent initial (Lucci 1983, Lyche et Girard 1995, Di Cristo 1999), très répandu chez les habitués de la parole publique, et qui redonne au mot lexical toutes ses lettres de noblesse. Le lien entre l'accent et la tendance à singulariser l'initiale n'avait pas échappé à Delattre (1966 : 29) même s'il le restreint à l'initiale de phrase : "Très légère dans le parler parisien cultivé, plus marquée dans le parler parisien des faubourgs, cette attraction de l'initiale se révèle, en français, dans l'accent d'insistance." Il poursuit en mettant en valeur ce facteur dans la chute de schwa : "[l]a grande loi qui va présider au jeu des *e* instables de monosyllabes à l'initiale […] : *le facteur psychologique (attraction de la position initiale de phrase) joue contre le facteur mécanique (force d'articulation consonantique combinée avec aperture consonantique) et l'emporte généralement, mais d'autant moins nettement que ce dernier facteur lui oppose plus de résistance."*

Ces remarques de Delattre couplées aux travaux ci-dessus mentionnés nous invitent à nous interroger sur le rôle effectif de la position initiale pour le maintien du schwa. Peut-on envisager de retrouver le même type de contraintes au niveau lexical qu'au niveau post-lexical ? **Nous demanderons donc au codage prosodique de nous renseigner sur la position de la voyelle dans le mot dans sa réalisation concrète.** En considérant uniquement l'output, nous ne dupliquons pas le codage schwa qui renseigne sur la position de la voyelle dans la forme de base du mot. Ainsi une voyelle cataloguée médiane pour le codage schwa est-elle susceptible d'être initiale dans le codage prosodique, la voyelle médiane de *devenir* par exemple, étant comptabilisée comme initiale si le premier schwa n'est pas prononcé.

Les défis théoriques lancés par la position initiale ont été relevés en particulier par Scheer (2000) qui, dans le cadre de la phonologie du gouvernement, propose une solution élégante à la présence obligatoire du schwa et de la liquide dans *grenouille* alors que les deux segments peuvent être omis dans *autrement* [otmə̃]. Son analyse repose sur l'existence d'une syllabe vide CV (Lowenstamm 1999) à l'initiale d'un mot, syllabe

[14] Il s'agissait pour chaque sujet de lire plusieurs fois une histoire afin de la mémoriser et ensuite de la raconter le plus fidèlement possible sans avoir accès au texte.

soumise aux principes de gouvernement. Pour simplifier, nous dirons que la voyelle vide initiale doit être légitimée par le Gouvernement Propre, ce qui est possible uniquement si le schwa est réalisé, une voyelle initiale pleine n'étant pas soumise aux mêmes contraintes[15]. Si Charette (1991)[16], dans le même cadre théorique, rend également compte sans difficulté de la tendance à la lexicalisation du schwa à l'initiale devant un groupe OL (*secret*), son analyse, pas plus qu'aucune autre, ne propose de solution satisfaisante pour exprimer la distinction entre *le secret*, où le schwa est plutôt stable, et *le secrétaire,* où le schwa chute plus fréquemment (*le premier s(e)crétaire du parti socialiste*). Le seul élément qui sépare ces deux suites, en plus de la fréquence lexicale, est à chercher dans la longueur syllabique (LG). On retrouve un comportement identique avec les monosyllabes, où un schwa tombe plus facilement dans un contexte Cə#CV que dans un contexte Cə#CCV lorsque le groupe consonantique n'est pas de type <obstruante+liquide>. Dans ce deuxième cas, force nous est de distinguer entre Cə#CCV(C)# et Cə#CCVC(C)V : le schwa dans *pas de pneu* se maintient normalement alors qu'il est beaucoup plus fragile dans *pas d(e) pneumologue*[17].

Rappelons pour conclure la liste des exigences que nous formulons par rapport au codage prosodique en vue d'un approfondissement de l'analyse du schwa. Le codage prosodique renseignera le codage schwa comme suit et nous saurons précisément :

1. Le nombre de syllabes du mot effectivement réalisées, ceci afin de pouvoir calculer le nombre de syllabes séparant le schwa de la proéminence la plus proche.
2. La position du schwa dans le mot tel qu'il est réalisé.
3. Si un schwa prononcé est perçu ou non comme proéminent.
4. Si un schwa prononcé est perçu comme long ou non.
5. Si le schwa est suivi d'une pause.
6. La position du schwa dans l'énoncé.

2. Questions de méthode

De la même façon que la constitution d'un corpus obéit toujours à des objectifs scientifiques clairement circonscrits[18], il est nécessaire de cibler finement les contextes d'analyse, *a fortiori* pour une étude prosodique. Effectivement, étant donné le coût de traitement que ce genre de travail exige (granularité du codage déjà relativement fine), des sélections drastiques de contexte s'imposent. Sur quelles bases théoriques et empiriques effectuer ces sélections ? Dans quelle mesure l'éclairage de premières données peut nous orienter pour stabiliser nos hypothèses et notre codage ? Que nous disent ces données préalables sur le potentiel contextuel des corpus (données bavardes ou silencieuses)[19] ? Enfin, de quelle façon réorientent-elles la démarche que nous nous étions fixée au départ pour l'analyse ? Voilà l'ensemble des questions qui sont abordées dans les paragraphes qui suivent.

[15] Voir Scheer (2000 : 122-124) pour une analyse détaillée.
[16] Voir Lyche et Durand (1996) pour une critique de l'analyse proposée.
[17] Dans le même ordre d'idées, on remarquera que certains locuteurs du français du midi insèrent un schwa dans *un pneu* ([pønø]) afin de briser le groupe consonantique, mais ne le font jamais dans *pneumatique* (J. Eychenne, p.c.).
[18] Le corpus PFC en est une illustration claire en particulier concernant la liste des mots et le texte à lire.
[19] Si, au bout du compte, nous avons limité notre étude au schwa et laissé la liaison de côté pour l'instant, c'est notamment dû à la faible quantité des contextes recueillis dans nos premières données.

2.1. La sélection des passages à coder

Une fois nos hypothèses posées, la question de la sélection des passages à coder mérite que l'on s'y attarde quelque peu. Pour ce qui est du schwa et de la liaison, les instructions données aux codeurs revêtent un caractère élémentaire : il s'agit de coder le texte dans son entier, 3 minutes des deux conversations pour le schwa et 5 minutes des deux conversations pour la liaison. Le codage, rappelons-le, est alphanumérique et s'effectue à la suite des lexèmes/graphèmes concernés sur une tire spécifique mais qui duplique la tire de transcription orthographique. Les consignes de sélection sont réduites à un minimum, puisqu'il est uniquement conseillé aux codeurs de choisir des passages où apparaissent un nombre relativement important des deux phénomènes. Ces consignes recommandent cependant de coder pour le schwa et la liaison les mêmes passages (et d'ajouter deux minutes pour la liaison). Les passages codés sont tous contigüs, les codeurs choisissant une séquence globale. Une telle approche pour le codage de la prosodie n'est pas envisageable, ne serait-ce qu'à cause du coût du codage qui exige une finalité bien précise. En d'autres termes, une sélection minutieuse et rigoureuse des passages à coder s'impose pour faire parler les données.

Précisons d'emblée que le codage de la prosodie se distingue significativement des codages précédents dans la mesure où les unités à coder ne sont plus des lexèmes/graphèmes mais des syllabes, et que la première tâche du codeur consistera a segmenter le signal en syllabes (voir 3.1). Plusieurs stratégies de sélection des passages restent néanmoins envisageables. Tout d'abord, on pourrait souhaiter maintenir le parallélisme avec le codage schwa et le codage liaison en proposant le codage aléatoire d'une portion du texte et d'une minute de chaque entretien par exemple, lors d'échanges reflétant au mieux le registre de conversation étudiée. Cette procédure peut fournir des résultats fructueux à condition que les occurrences recherchées soient suffisamment fréquentes. Or, les premiers résultats de PFC nous indiquent clairement que les instances de liaison sont fort rares dans le parler spontané et, même si un nombre considérable de segments sont associés à un codage schwa, rien ne nous garantit que n'importe quel passage codé pour le schwa présente des caractéristiques exploitables pour la prosodie. Il nous est apparu qu'une démarche non contrôlée ne donnerait pas de résultats pertinents, étant donné notre objectif particulier qui vise à mettre en relation la présence/absence du schwa (et de la liaison) avec la prosodie dans des contextes bien circonscrits. Une deuxième solution consisterait alors à sélectionner attentivement les passages à coder en ayant constamment à l'esprit les types de requêtes que nous souhaitons effectuer par la suite[20], de manière à tester des hypothèses phonologiques clairement formulées au préalable (cf. *infra* 2.4). Nous opterons bien entendu pour cette deuxième approche sans pour autant verser dans un optimisme débordant. La parole spontanée reste par définition spontanée, sans garantie aucune de fournir tous les types de contextes que nous nous proposons d'envisager, et nous serons contraints de mettre en place ultérieurement des procédures d'expériences complémentaires en laboratoire.

Environ cinq locuteurs sont retenus pour chaque point d'enquête soumis au codage prosodique et la sélection s'établit sur la base de la qualité du signal, sur la présence de variations mélodiques bien nettes, et sur l'intérêt des contextes. Les passages du texte à coder sont préétablis et favorisent l'étude de liaisons variables, les sites d'insertion attendue de schwa et les quelques instances de monosyllabes où l'absence de schwa est envisageable.

[20] Il s'agit par exemple d'optimiser le temps de sélection par l'utilisation des informations fournies par le classeur schwa.

(3) Texte PFC : passages à coder

Le Premier Ministre ira-t-il à Beaulieu
Le village de Beaulieu est en grand émoi
Le premier Ministre a en effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année
ses chemises en soie
4ème aux jeux olympiques de Berlin
son usine de pâtes italiennes
Qu'est-ce qui a donc valu à Beaulieu
car le Premier Ministre, lassé des circuits habituels qui tournaient toujours autour des mêmes villes
la campagne profonde
Le maire de Beaulieu, Marc Blanc, est en revanche très inquiet
Comment, en plus, éviter les manifestations qui ont eu tendance à se multiplier lors des visites officielles
les opposants de tous les bords manifestent leur colère
le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout est fait pour le protéger
un gros détachement de police
Un jeune membre de l'opposition aurait déclaré
S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction
l'Express, Ouest-Liberté et le Nouvel Observateur
Le sympathique maire de Beaulieu
Il a le sentiment de se trouver dans une impasse stupide
Il s'est, en désespoir de cause
pour vérifier si son village était vraiment une étape nécessaire dans la tournée prévue
Beaulieu préfère être inconnue et tranquille plutôt que de se trouver au centre d'une bataille politique

Nous préconisons le codage d'une minute des deux conversations (dirigée et libre) pour chacun des témoins retenus. Les passages sélectionnés sont discontinus et sont choisis de façon à mettre en valeur les patrons prosodiques qui interpellent le codeur, des patrons que l'on qualifiera de "marqués" par rapport à l'intuition du codeur de ce que peut être une prononciation non marquée ou standard[21]. Il s'agira également de privilégier ces contextes dans d'autres modalités de production : par exemple un contexte défini dans la conversation guidée sera également testé dans la conversation libre.

2.2. Notre premier codage et ses désillusions

La première version du codage proposée en 2003 (Lacheret, Lyche & Morel 2003b, 2004) s'articulait autour de 5 champs prosodiques qui nous renseignaient sur les proéminences perçues. Caractérisant à la fois les corrélats psycho-acoustiques d'une proéminence et ses caractéristiques distributionnelles, il donnait également des informations restreintes sur les pauses. Pour mémoire, nos objectifs étaient les suivants : (1) indiquer la configuration acoustique de la proéminence (type de pente et amplitude), (2) informer sur le registre intonatif et la longueur d'une syllabe perçue comme proéminente, (3) renseigner sur le type

[21] Exemples de prononciation marquée : la liaison réalisée dans *pâtes italiennes* ou le schwa prononcé dans *stage.*

de mot porteur d'une proéminence (outil *vs*. lexical), (4) fournir des indications sur le contexte droit d'une proéminence terminale (pause silencieuse *vs.* sonore, absence de pause), (5) compter les syllabes atones entre deux proéminences.

Très vite, ce codage est apparu problématique, tant du point de vue des objectifs attendus que des questions auxquelles il était censé répondre, et de l'exploitation systémique que l'on pouvait en faire. Si l'on part du principe, en effet, que toute transcription, *a fortiori* prosodique, repose sur un compromis à trouver entre la finesse des informations fournies et la nécessité de restreindre au maximum les informations (ni trop, ni trop peu d'informations), notre codage se révélait à la fois trop complexe et trop parcellaire. Trop complexe, étant donné le profil du codeur et de l'analyste d'une part (*a priori* novice en phonétique[22]), compte tenu de certains pré-requis linguistiques considérés comme allant de soi, ce qui n'est évidemment pas le cas (voir la distinction encore très problématique entre mots outils et mots lexicaux[23]). Trop parcellaire dans la mesure où ce codage repose essentiellement sur l'identification d'îlots de saillance prosodique sans en fixer précisément les domaines (seule la notion intuitive d'entrée lexicale mentale est retenue) et amalgame parfois dans un seul et même champ des informations qui peuvent être issues de contraintes fonctionnelles distinctes (voir champ 2 : registre intonatif et longueur). En outre, au fur et à mesure que nous regardions nos données pour nous assurer que nos hypothèses étaient effectivement falsifiables, il devenait très vite évident que celles-ci devaient être étayées et précisées. Ceci nous amène à une remarque d'ordre méthodologique générale : si l'opposition entre méthode hypothético-déductive et stratégie inductive est souvent posée de fait comme allant de soi, il s'agit dans la pratique d'un artefact. En effet, de même que l'éclairage des hypothèses par les données suppose un va-et-vient entre ces hypothèses et ces données, une méthode uniquement inductive est illusoire si on ne veut pas vider l'eau de la mer à la petite cuiller : on ne cherche pas n'importe quoi, n'importe où, des hypothèses posées préalablement orientent drastiquement la recherche. Soit le cheminement suivant (figure 1) :

- Dans une phase initiale, un premier jeu d'hypothèses va permettre de mettre en place un protocole de codage préalablement à la sélection des contextes à coder et à leur codage effectif. Une première exploration des données va nous amener à préciser nos hypothèses et à modifier notre protocole de codage le cas échéant[24] (cf. *infra* 2.3).
- Ce n'est que lorsque cette phase initiale est vraiment stabilisée que des requêtes automatiques peuvent être lancées afin de fournir des résultats statistiques qualitativement et quantitativement représentatifs pour l'analyse des données.

[22] La prise en compte des corrélats acoustiques de la proéminence suppose une expertise phonétique certaine et des outils logiciels dédiés.

[23] L'opposition ±clitique associée à la distinction mot outil *vs*. mot lexical est finalement peu apte à rendre compte des deux fonctionnalités accentuelles (emphase énonciative *vs*. démarcation de frontière).

[24] Ainsi, certains champs devaient être enrichis (ex. précision des contextes syllabiques des proéminences perçues, type et nature des pauses, tentative de distinction *a priori* entre *euh* d'hésitations et schwas terminaux), d'autres devaient être créés (contextes intra- *vs*. inter- mots prosodiques).

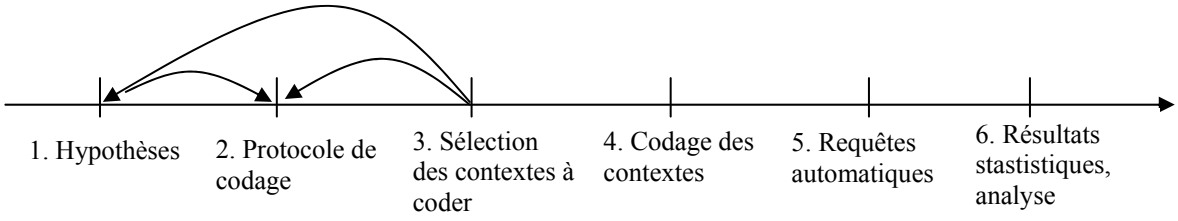


Figure 1. *La démarche sur corpus : des hypothèses aux données, des données aux hypothèses*

Restait à savoir enfin si le même type de codage pouvait être utilisé pour les différents modes de production analysés (lecture *vs*. conversation libre et dirigée). Bref, le problème de la granularité du codage qui se pose classiquement en linguistique de corpus, quel que soit le domaine d'investigation, restait ouvert et une année de discussions avec les experts[25], de réflexions et de présentation de l'évolution de nos travaux[26] a été nécessaire pour stabiliser notre protocole de codage.

2.3. L'éclairage de Brécey

Nos premiers efforts de codage de la prosodie ont porté sur un point d'enquête en Normandie, le village de Brécey proche d'Avranches. Brécey, petit bourg rural à vocation agricole, compte environ 2100 habitants (Girard et Lyche 2004) et 10 locuteurs ont été enregistrés selon le protocole PFC. Parmi ces 10 locuteurs, nous en avons choisi 5 pour le codage de la prosodie : trois hommes (âgés respectivement de 19, 53, 81 ans) et deux femmes (39 et 55 ans). Nous avons testé notre première version de codage sur des passages présélectionnés du texte, une minute de conversation dirigée et une minute de conversation libre. Les passages des conversations étaient extraits des séquences déjà codées pour le schwa et la liaison (cf. 2.1).

Les données sont modestes en position finale de mot où peu de schwas apparaissent dans le corpus. Dans tous les cas, il s'agit de marqueurs de planification du discours, des schwas prononcés essentiellement dans des connecteurs et anticipatoires d'une hésitation, que l'on pourrait qualifier de schwas "phatiques" : *et donc ça a fonctionné pendant je pense* [edõkə saafõksjɔne pɑ̃dɑ̃ ʒpɑ̃s]. Pour ce qui est des monosyllabes, relativement fréquents dans nos données, il semblerait que la consonne orpheline après la chute de schwa se greffe sur l'attaque de la consonne d'attaque de la syllabe suivante, créant de la sorte un groupe consonantique qui ne respecte pas obligatoirement la hiérarchie de sonorité. Le noyau de cette syllabe est proéminent, ce qui indiquerait que la création d'un groupe consonantique d'attaque "attire" l'accent. La présence de cet accent initial est également à la source de clashs accentuels : *au lycée de Mortain* [olise dmɔʁtɛ̃], où les deux dernières syllabes sont proéminentes.

(4) Proéminence perçue sur syllabe initiale

D'aller se colleter	[dale **skɔl**te]
Au lycée de Mortain	[olise **dmɔʁ**tɛ̃]
Entre te couper les cheveux	[ɑ̃tʁə **tku**pe leʃvɲ]

[25] Nous souhaitons remercier ici C. Auran, G. Caelen-Haumont, B. Laks, Ph. Martin, P. Mertens, M. Morel, F. Poiré, A.C. Simon, pour leurs conseils et critiques avisés.
[26] Voir Lacheret, Lyche et Morel (2003a,b), (2004a,b,c) (2005), Lacheret et Lyche (2005).

Dans certains cas, par contre, le monosyllabe non seulement maintient le schwa mais est accentué sous l'effet d'une instruction pragmatique : *je me suis marié l<u>e</u> 13 décembre*[27].

La position initiale de mot s'avère bien plus riche quantitativement que la position finale, même si l'on observe peu de schwas initiaux en dehors du préfixe *re*.

(5) Schwa absent en position initiale (hors *re-*)

Un petit peu	[œ̃ptipø]
Les cheveux	[leʃvø]
Je pense que c'était en seconde	[seteɑ̃sgõːd]
Ça devient quand même	[sa dvjɛ̃]
Ça a demandé du temps	[saa dmɑ̃de dytɑ̃]
Une fois qu'ils sont dedans	[kisõ ddɑ̃]

En (5), seules les syllabes initiales des trois derniers exemples sont accentuées, proéminence qui pourrait être liée aussi bien aux groupes consonantiques occlusive+C qu'à la moindre fréquence lexicale des items en question. Une étude de quelques points d'enquête (Suisse, Vendée par exemple) montre en effet que *petit* est régulièrement prononcé sans le schwa initial et nous voyons dans *un petit peu* une expression suffisamment fréquente pour être figée dans sa prononciation (Bybee 2001). Dans les données dépouillées, les suites *#re-* (qu'il s'agisse d'un préfixe ou non) exhibent un comportement de monosyllabe, avec un schwa absent dans l'immense majorité des cas, même pour les mots peu courants comme *revitaliser*. Ces observations méritent d'être confrontées à des données plus vastes. Malécot (1976) et Walter (1977) par exemple ont soutenu que les préfixes *re* et *de* tendent à conserver le schwa alors que Hansen (1994) ne trouve pas de fondement empirique à ces affirmations. Dans son étude sur le français parisien, elle note une forte propension à la chute dans les mots à initiale en *re-*, mais suggère que l'effet conservateur de préfixe dont parlait Walter ne se réalise que pour les mots dont la fréquence lexicale serait faible. Dans ce cas, *revitaliser* devrait maintenir son schwa, or il le perd. La raison de la chute s'explique peut-être encore une fois par la longueur du groupe. Casali (1997), dans le cadre de la théorie de l'optimalité, propose une dominance systématique de PARSE(lex) sur PARSE(affixe)[28], dominance confirmée dans le cas du français où le préfixe *re* peut prendre la forme *ré-* ou *r-* devant voyelle. Nos données semblent cependant indiquer que toute initiale en *re-* se comporte comme un affixe, tendance qui demande à être vérifiée sur un corpus plus large.

[27] Nous sommes ici dans un contexte d'ouverture de segment rhématique (réponse à la question : *Quand vous-êtes-vous marié ?*).

[28] Par rapport à un input donné, tous les éléments de la racine sont maintenus plus systématiquement dans l'output que ne le sont ceux d'un affixe.

(6) Initiale en *re* : contexte V#*re*

Je suis revenu	[ʒsɥi ʁvəny]
Des relations	[de ʁlasjõ]
Je suis reparti	[ʒsɥi ʁpaʁti]
Si ça peut revitaliser un peu les campagnes	[sisapɲ ʁvitalize]
Ben, regarde Clément	[bɛ̃ ʁgaʁd klemɑ̃]

Dans l'étude des proéminences, on relève que les groupes consonantiques constitués de segments phonétiquement proches attirent régulièrement l'accent, mais que la force de cet accent est inversement proportionnelle au degré de sonorité de ces consonnes : occlusives sourdes < occlusives sonores < fricatives sourdes < fricatives sonores < nasales < /l/ < /r/[29]. Ainsi la proéminence est-elle nette dans *Ils vont recommencer* /rkɔ/ mais beaucoup moins dans *Des relations* /rla/. Dans tous les cas, la présence d'une coda semble accroître l'éventualité de perception accentuelle, comme dans *d'aller se colleter* [skɔl]. Ces facteurs interagissent de surcroît avec l'accent démarcatif de façon à ce que dans l'exemple *ils rechangent* [iʁʃɑ̃ʒ], la source exacte de la proéminence sur la deuxième syllabe reste délicate à déterminer.

Nous avons vu que la chute du schwa en syllabe initiale élargit considérablement l'éventail des groupes consonantiques autorisés en français. Cette floraison de possibilités a conduit certains (par ex. Pagliano 2005) à prétendre que le français, tout comme l'arabe, ne plaçait aucune restriction sur ses groupes d'attaque, alors que Turcsan (2005) conclut à partir des mêmes données sur la nécessité de distinguer pour l'analyse un niveau lexical d'un niveau post-lexical. Nous n'interviendrons pas ici dans le débat sur le nombre de niveaux exigés pour l'analyse, mais estimons que l'ensemble [groupe consonantique (à l'initiale d'un mot) + proéminence] a un statut perceptif particulier : c'est un indice net de la frontière gauche du mot prosodique[30]. La borne droite, quant à elle, est beaucoup plus fluctuante avec un groupe consonantique final le plus souvent réduit (Laks 1980). Même dans un registre plus soigné qui est celui de la lecture d'un texte, nous observons pour *premier ministre* une variété de réalisations comme [minist], [minis] ou même [minisʁ].

2.4. Consolidation de notre premier jeu d'hypothèses

À partir du tour d'horizon effectué dans les précédentes sections et en nous fondant sur notre étude des premiers résultats apportés par l'analyse des données de Brécey (Lacheret et Lyche 2005), nous avons développé un ensemble d'hypothèses centrées sur l'analyse des schwas monosyllabiques, initiaux et médians de polysyllabes à partir desquelles nous avons dégagé un premier jeu de requêtes informatiques. Nos deux hypothèses relèvent directement de l'interface prosodie-segmental et nous supposons pour la première que l'output est susceptible de nous renseigner sur l'input. Nous présumons en effet que si le schwa est présent de quelque manière que ce soit dans la structure sous-jacente, son élision modifiera le patron mélodique des voyelles environnantes. La deuxième hypothèse porte sur la

[29] Malécot propose également que la force (définie selon l'échelle de sonorité) des consonnes qui entourent le schwa influe sur sa chute. Dans la suite C1əC2, plus C1 est faible par rapport à C2, plus le schwa aura tendance à disparaître.

[30] Remarquons par ailleurs la tendance à transformer *ce, cet* ou *cette* en [stø] dans la parole spontanée, même dans le français du midi (*ce pull ?* [støpyl]) et ainsi à créer un groupe d'attaque qui aurait fort bien pu être évité.

distance accentuelle (DA) et la longueur des groupes (LG). Nous formulerons DA et LG sous forme de contraintes en (7) et (8).

(7) Distance Accentuelle (DA)
Plus le schwa est distant de l'accent terminal, plus il a tendance à rester.
Ex. *la r**e**lation étroite*

(8) Longueur des Groupes (LG)
Un schwa se maintient dans un groupe court mais pas dans un groupe long.
Ex. *port**e**-feuille*

DA repose sur l'hypothèse d'une analogie entre l'occurrence du schwa et celle de l'accent. De même que, dans la chaîne parlée, la syllabe accentuée se détache de son contexte par sa proéminence, constituant ainsi une figure par rapport aux syllabes inaccentuées qui l'entourent (Pasdeloup 1990), une syllabe dont le schwa est prononcé dans un environnement d'élision possible représenterait une figure par rapport à une syllabe dont le schwa a effectivement été effacé. Dans les deux cas, on a affaire à des syllabes marquées perceptivement. Deuxième hypothèse : une syllabe ne peut être doublement marquée (assumer deux fois le statut de figure), autrement dit, une syllabe dont le noyau est un schwa est nécessairement atone. Troisième hypothèse : pour qu'une syllabe assume ce statut de figure, il faut que les syllabes contiguës soient reléguées au rôle de fond[31]. Autrement dit, un schwa pré-accentuel serait rarement accentué pour éviter la contiguïté de deux figures qui occasionnerait l'équivalent d'un clash accentuel *(seulement)*.

Ces hypothèses posées sont riches de conséquences, elles méritent donc d'être vérifiées et appellent plusieurs questions :

- Observe-t-on une corrélation quantitativement représentative entre la distance accentuelle du schwa et son maintien ? Cette préférence dans la prononciation est-elle de nature post-lexicale (*i.e.* ne concernant que le mot prosodique ou un segment supérieur dans la hiérarchie prosodique), ou est-elle convoquée également dans des stratégies lexicales ? Dans le premier cas, le maintien du schwa s'expliquerait aisément : si un accent terminal, toujours présent en français, indique la frontière droite d'un mot prosodique, l'accent initial peut jouer le rôle de marqueur de frontière gauche (Lyche et Girard 1995). Schwa, par le biais de cette fonction de figure, serait habilité à remplir cette même tâche. Le maintien d'un schwa initial et/ou l'accentuation de la syllabe initiale s'expliqueraient par la tendance à éviter des problèmes d'intégration perceptive et à faciliter la segmentation, le traitement perceptif d'un mot prosodique long se faisant difficilement à l'aide d'un seul marqueur de frontière droite. Remarquons par ailleurs que dans le cas où le schwa est éliminé de la syllabe initiale, la syllabe suivante, promue à l'initiale suite à la disparition du schwa, porte le plus souvent un accent. Que cet accent provienne ou non de la présence d'un groupe d'attaque lourd (Basbøll 1988) et ne respectant pas toujours la hiérarchie de sonorité, sa fonction demeure inchangée, il contribue à signaler la frontière du mot prosodique.

- Cette corrélation concerne-t-elle le rapport entre le schwa et tous les sites accentuels (site démarcatif *vs*. site focal et/ou rythmique), ou n'est-elle liée qu'au repérage des démarcations de frontière ? Auquel cas, DA ne s'appliquerait absolument pas pour les accents internes qui seraient exclusivement du ressort de LG.

[31] C'est d'ailleurs dans ce sens qu'est formulée la règle de clash accentuel largement débattue dans la littérature (Dell 1984, Delais 1995, Lacheret et Beaugendre 1999).

- Dans le même ordre d'idée, quelle est la nature précise des contextes, *a priori* nombreux dans nos corpus, qui, malgré les déclarations de principe des phonologues, génèrent des clashs accentuels ? S'agit-il systématiquement de fonctionnalités accentuelles différentes ?

- Cette corrélation s'illustre-elle également pour une suite de schwas à l'intérieur d'un MP *(je me demande, il a redemandé du pain)*. Si ce phénomène est constaté, n'est-il pas modulé par des critères morphologiques (rôle de proforme du pronom, statut préfixal de la première syllabe d'un mot polysyllabique) ou non respecté pour des raisons articulatoires (respect de la contrainte de sonorité : *je te verrai plus tard*) ?

- Est-il vrai qu'un schwa prononcé alors qu'il pourrait être élidé ne cumule jamais le rôle de double figure ? Si, à l'inverse, on trouve des occurrences de schwa accentué, quels sont les processus autres que les mécanismes d'emphase qui pourraient expliquer de telles occurrences ?

3. Un protocole de codage stabilisé ?

À l'issue de l'atelier "prosodie" organisé à Caen en mai 2004, il nous est apparu d'abord qu'un protocole unique était souhaitable quel que soit le mode de production analysé (lecture, conversation libre vs. dirigée) mais qu'il se déclinerait en deux modes "étendu" *vs*. "standard". Pour ce protocole, nous nous sommes fixé la contrainte suivante : à chaque champ correspond une et une seule information (fonctionnelle, structurale, perceptive ou acoustique). Le protocole étendu nécessite une expertise phonétique et des logiciels dédiés, il est donc là en réserve pour être déployé le cas échéant en vue d'analyses complémentaires fines (en phonologie de laboratoire notamment). Nous décrirons donc celui-ci rapidement après nous être penchées plus spécifiquement sur le protocole standard qui, lui, est réalisable par un non spécialiste entraîné (environ 2 heures) et que nous avons mis à l'épreuve ces dernier mois pour la transcription de notre point d'enquête de Brécey.

3.1. Le protocole standard

L'architecture globale du protocole comprend 6 champs, les deux premiers fixant nos domaines de codage prosodique (la syllabe et le mot), les 4 autres fournissant les informations prosodiques exploitées pour les tests des hypothèses telles qu'elles ont été formulées précédemment. Rappelons que le signal correspondant à la suite sélectionnée pour analyse sera préalablement segmenté en syllabes et que chaque syllabe sera renseignée pour les 6 champs. Avant de présenter ce protocole, expliquons pourquoi nous avons finalement laissé de côté pour notre codage l'opposition mot-outil *vs*. mot lexical. En pratique, plus le codage est simple, moins les erreurs du codeur seront nombreuses. Or, est-il pertinent de demander au codeur de tenir compte d'emblée de l'opposition mots-outils *vs*. mots lexicaux en lui donnant par exemple une liste fermée de mots-outils, démarche qui va inévitablement augmenter sa charge de travail, pour au bout du compte être remise en question à tout instant, étant bien entendu que l'opposition mots-outils/mots lexicaux est loin de faire le consensus chez les chercheurs ? Illustrons notre propos en partant de l'opposition fonctionnelle suivante (R1):

- tout accent tombant sur la dernière syllabe d'un mot lexical est un accent primaire, i.e. de nature structurale, il a une fonction démarcative
- tout autre accent (interne de mot ou porté par la syllabe terminale d'un mot-outil) est doté d'une fonction secondaire (rythmique et/ou énonciative).

Cette règle est elle tenable bien longtemps ? Si oui, comment traiter alors les accents démarcatifs sur les mots-outils dans les occurrences suivantes : *prends-<u>le</u> avec toi, ni*

Robert, ni à mon avis, Marie, ne viendront au mariage de Simon), sans tenir compte pour le premier cas de la règle de report accentuel (Rossi 1979, 1980), pour le second de la nécessité de marquer le décrochement par l'incise ? Et à l'inverse, peut-on dire que l'accent est démarcatif dans *elle habite une grande maison* ? Évidemment non, bien que l'accent frappe effectivement une syllabe terminale d'un mot lexical (ici l'adjectif *grand*). Bref, pour que l'opposition MO/ML reste pertinente jusqu'au bout, il faut tenir compte de ces exceptions à R1 et de bien d'autres (voir en particulier Mertens et al. 2001, Lacheret 2003, Mertens 2004) ce qui n'est évidemment pas du tout cohérent avec l'optique d'un codage léger. Nous avons donc préféré restreindre nos domaines et nos champs associés dans une perspective de balayage large des données. À l'analyste de faire ses choix ultérieurement en fonction du modèle théorique qui sera le sien.

La démarche inverse nous a paru beaucoup plus pertinente : 1° faire réaliser un premier balayage large des données, 2° une fois les proéminences étiquetées sans parti pris de départ, l'analyste peut construire sa liste si nécessaire en fonction de ses critères et de ses objectifs particuliers et corréler le cas échéant certains items de sa liste avec certaines proéminences perçues.

En pratique, les 6 champs du codage standard se décomposent comme suit :

- Le **champ 1** indique le nombre de syllabes du mot en cours de traitement : 1, 2, 3 syllabes ou plus. Le codeur doit coder uniquement les syllabes perçues en tenant compte notamment de tout type de réduction vocalique. Ainsi *ministre* prononcé [ministʁə] est-il un mot de 3 syllabes ; prononcé [mnis], il devient monosyllabique.
- Le **champ 2** code le rang de la syllabe dans le mot : 1^{e} syllabe, syllabe interne, dernière syllabe de mot polysyllabique ou syllabe de mot monosyllabique.

Un code supplémentaire a pour finalité le repérage dans la tire prosodie d'un schwa final prononcé alors qu'il est *a priori* muet. Il s'agit pour le français du Nord d'une syllabe finale "inattendue" comme dans *bonjour-e*,ou dans *ministr-e*. Dans les variétés du Midi, un schwa final réalisé sera également codé afin de pouvoir différencier *lac bleu* [lakblɲ] de *laque bleue* [lakəblɲ].

- Le **champ 3** renseigne sur la perception de la proéminence. Une syllabe est dite proéminente si elle est perçue comme plus intense, plus forte que ses syllabes environnantes, elle se dégage ainsi comme une figure sur un fond dans le flux sonore.

L'objectif de ce champ est d'une part de renseigner sur la distribution de la proéminence afin de 1° distinguer différentes fonctionnalités accentuelles et 2° obtenir des résultats quantitatifs quant au statut accentué de la syllabe initiale ; d'autre part d'analyser une corrélation éventuelle entre la proéminence et l'élision vocalique (celle du schwa en particulier, cf. *supra* 1.1).

- Le **champ 4** indique des variations de durée syllabique éventuelles.

Ne disposant pas plus que quiconque ici de modèle de durée qui nous permette de quantifier objectivement l'allongement, ce dernier est calculé par rapport à la durée de la syllabe qui précède, considérée comme une durée de référence : pour qu'il y ait allongement, il faut que la durée de la syllabe en cours de traitement soit supérieure d'au moins 50% à celle qui précède[32].

[32] Étant donné l'instruction très générale donnée ici au codeur, il ne distinguera pas dans ce champ les allongements structuraux des allongements liés à la planification discursive et à la mise en mots (hésitations).

- Le **champ 5** donne des informations sur les occurrences des pauses produites et leur nature.

Étant donné notre domaine d'investigation et nos objectifs, ce champ reste relativement sommaire par rapport aux descriptions que l'on a pu proposer sur les types de dysfluences rencontrées dans le discours (Assié 2005). Nous distinguons les pauses silencieuses courtes et longues (seuil de référence = 300 ms), les *euh* d'hésitation et les répétitions syllabiques *(c'est in/incroyable).*

- le **champ 6** permet d'élargir le domaine de codage au sein du discours (syllabe initiale ou non de tour de parole).

Si l'on considère que la position initiale de mot est une position d'attaque, en se situant dans une perspective post-lexicale, on peut formuler l'hypothèse qu'il y a différents degrés d'attaque en fonction de la position du mot dans une séquence : initial de tour de parole, interne à un tour et initial de mot prosodique (précédé d'une pause longue de plus de 300 ms), interne de groupe. C'est donc le champ 6 qui nous permettra d'explorer cette hypothèse. En conséquence, on distingue dans ce champ 3 types de syllabes initiales : celles qui ouvrent un nouveau tour de parole (précédées d'une pause longue), celles qui introduisent un nouveau groupe prosodique et les autres.

Ouverture de tour : **#** ***[je*** *sais] [qu' c'est comme ça]*[33]
Initiale de mot prosodique : #*[je sais] [**qu'** c'est comme ça]*
Interne à un mot prosodique : [*il m'a dit] [qu'il **v**'nait demain soir]*

3.2. Le protocole étendu

Trois champs sont rajoutés dans le protocole étendu, fondés sur l'exploration de critères acoustiques. En pratique, il s'agit de tenir compte du niveau fréquentiel des noyaux syllabiques (champ 7), de la configuration mélodique des contours syllabiques (champ 8 et de l'intensité associée (champ 9).

Champ 7	**Proéminence F0 mesurée**
0	Nulle
1	Positive
2	Positive forte
3	Négative
4	Négative forte
Champ 8	**Contour F0**[34]
0	Plateau
1	Montant
3	Montant fort
4	Descendant
5	Descendant fort
Champ 9	**Proéminence intensité mesurée**
0	Nulle
1	Présente

[33] Où les syllabes accentuées sont soulignées. Nous envisageons un *continuum* d'élision : élision peu fréquente en ouverture de tour mais beaucoup plus en position interne.
[34] Pour les différents degrés de montée et de descente voir l'opposition classique continuatif (conclusif) mineur *vs*. majeur.

Pour ce codage, nous l'avons dit, il est nécessaire d'utiliser un logiciel dédié. Nous proposons le système PROSOGRAMME développé à l'université de Louvain par P. Mertens[35] qui, aujourd'hui, nous semble effectivement le plus abouti parmi les logiciels que nous connaissons pour ce genre de tâche. Rappelons que PROSOGRAMME repose sur un modèle psycho-acoustique de la perception mélodique et donc sur une méthode de stylisation automatique des variation de F0. Autrement dit, seules sont indiquées les variations psycho-acoustiques fonctionnelles, i.e. qui entraînent des réponses perceptives. En conséquence, l'analyste peut travailler sur des données propres et quantifiées (figure 2)[36].

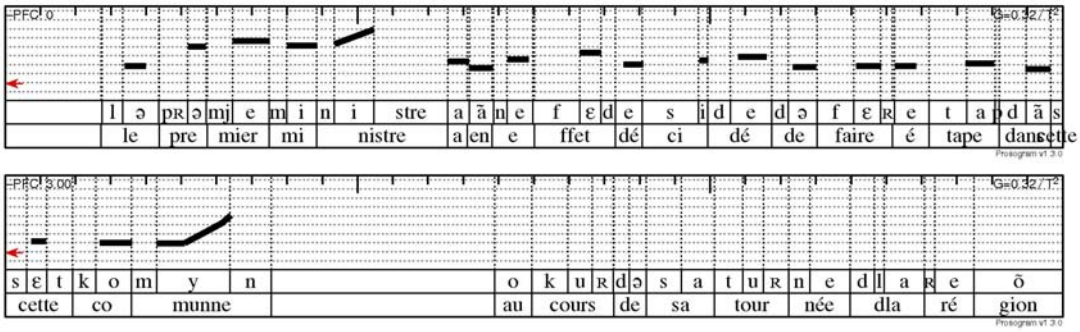


Figure 2. *Exemple de stylisation mélodique réalisée par le logiciel Prosogramme*

4. Conclusion

Nous avons souhaité ici faire le point sur l'état d'avancement de nos travaux depuis l'ouverture du chantier "prosodie" au sein du projet PFC en 2003. Première spécificité de ce codage : il ne s'agit pas d'un codage en soi et pour soi mais associé à une thématique ciblée qui se situe d'emblée dans une dimension multilinéaire : l'explicitation et la modélisation de l'interaction entre les variations de la prononciation du schwa et de la liaison, et les constructions prosodiques. Nous avons expliqué pourquoi nous avions finalement réduit cette thématique initiale en écartant pour l'heure de nos investigations la corrélation liaison-prosodie. Notre terrain ainsi balisé, nous avons commencé par situer notre problématique au sein des débats phonologiques classiques sur le schwa (représentation sous-jacente du schwa, insertion, élision ou segment flottant). Nous avons exposé ensuite nos hypothèses sur l'apport de la prosodie comme élément de réponse à des questions qui restent inéluctablement ouvertes faute de preuves, en insistant sur le phénomène de **traçage prosodique**. Autrement dit, la prosodie permettrait de confirmer certaines hypothèses au détriment d'autres en portant les traces de modifications de l'output par rapport à l'input (ex. traces d'élision manifestes), et cela pour différentes positions, en sachant que ces dernières (finale, interne, initiale de mot et monosyllabes) ne sont pas nécessairement assujetties aux mêmes règles et aux mêmes contraintes. A l'inverse la prosodie peut agir comme un déclencheur pour la réalisation du schwa (voir nos hypothèses sur la longueur des groupes et la distance accentuelle).

Mais pas de théorie sans méthodes et sans outils, surtout quand il s'agit d'explorer des données à grande échelle. Ce deuxième aspect qui sous-tend la qualité des résultats obtenus nous a amenées à exposer notre programme méthodologique autour de 3 points : la stratégie de sélection des contextes pour l'analyse, les différentes phases de notre codage et les critères retenus pour sa stabilisation (contraintes empiriques, et apport des premières données de terrain analysées), la mise en place d'un premier jeu d'hypothèses

[35] http://bach.arts.kuleuven.ac.be/pmertens/prosogram
[36] Pour une représentation détaillée du logiciel, voir en particulier Mertens (2004).

pour le développement d'une série de requêtes en cours d'implémentation et qui seront lancées dans une passe initiale sur 4 points d'enquête (la Réunion, Douzens, Vendée, Suisse) avec 5 locuteurs par point d'enquête.

Si nous attendons beaucoup de nos premiers résultats, en particulier pour la mise au jour de clusters invariants au sein de la francophonie qui viendraient confirmer nos hypothèses, et si un effort colossal a été effectué pour une sélection rigoureuse des contextes sur nos 4 points d'enquête, nous restons réalistes quant à leur représentativité tant d'un point de vue quantitatif que qualitatif : il est évident que pour l'ensemble des contextes dont nous souhaitons rendre compte, il y aura nécessairement des trous (voir la notion de données bavardes *vs*. silencieuses esquissée au § 2). Deuxième problème : celui de l'outillage. La tâche de transcription manuelle est un travail long et fastidieux, les erreurs sont inévitables. Des outils d'assistance au codage (syllabation semi-automatique, détection de proéminences) et de contrôle sont indispensables[37]. De la même façon, le développement d'un parseur/contrôleur pour la vérification des étiquettes est en cours de discussion avec l'équipe PFC du laboratoire Modyco (coll : A. Tchobanov, R. Walter). Troisième restriction : certaines mesures fines comme celle de l'allongement vocalique éventuel suite à la chute du schwa demandent un contrôle drastique des variables manipulées, d'où la nécessité de passer le relais à ce stade à la phonologie de laboratoire. C'est sans doute cette collaboration soutenue entre phonologie de terrain et phonologie de laboratoire qui permettra de reconfigurer le paysage de la phonologie contemporaine et d'alimenter de nouveaux débats sur des bases tout à la fois écologiques et scientifiquement contrôlées. En quelque sorte la première prépare le terrain de la seconde.

Références

Anderson, S. (1982). "The analysis of French schwa: or, how to get something for nothing". *Language* 58, 534-573.

Andreassen, H. (ce vol.) "Aspects de la durée vocalique dans le vaudois".

Angoujard, J.-P. (1997). *Théorie de la syllabe.* Paris : CNRS Éditions.

Assié, D. (2005). *Analyse syntaxique automatique de corpus oraux retranscrits.* Mémoire de DEA en Sciences du langage, Toulouse le Mirail.

Baude, O., M. Jacobson, A. Tchobanov et R. Walter (2005). "Interopérabilité des corpus sonores : le cas des corpus en français". *Phonological Variation : the Case of French,* University of Tromsø, Norway, 25-27 August 2005, http://www.projet-pfc.net/

Basbøll, H. (1988). "Sur l'identité phonologique du schwa français et son rôle dans l'accentuation et la syllabation". In P. Verluyten (ed.). *La phonologie du schwa français.* Amsterdam : Benjamins, 15-40.

Bibeau, G. (1975). *Introduction à la phonologie générative du français*. Montréal : Didier.

Bouchard, D. (1981). "A voice for the mute e". *Journal of Linguistics Research* 1(4), 17-47.

Bybee, J. (2001). *Phonology and Language Use.* Cambridge : Cambridge University Press.

Candea, M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits d'hésitation en français oral spontané.* Thèse de Doctorat, Université de Paris III.

Casali, R. (1997). "Resolving hiatus". ROA 215-0997.

[37] Dans cette optique, nous espérons beaucoup du logiciel PROSOGRAMME qui pourrait devenir un incontournable au sein de la communauté PFC.

Charette, M. (1991). *Conditions on Phonological Government.* Cambridge : Cambridge University Press.
Cotté, M.-H. (2001). "Le rôle de la perception dans la résolution des groupes consonantiques : le schwa français revu et corrigé". Paris : EHESS.
Delais, E. (1995). *Pour une approche probabiliste de la structure prosodique, étude de l'organisation prosodique et rythmique de la phrase française.* Thèse de Doctorat, Université de Toulouse-Le-Mirail.
Delattre, P. (1966). *Studies in French and Comparative Phonetics*. La Haye : Mouton.
Dell, F. (1984). "L'accentuation dans les phrases françaises". In F. Dell *et al.* (eds.). *Forme sonore du langage : structure des représentations en phonologie.* Paris : Hermann, 65-122.
Dell, F. (1985). *Les règles et les sons*. (Première édition 1973). Paris : Hermann.
Di Cristo, A. (1999). "Vers une modélisation de l'accentuation en français. Première partie : la problématique". *Journal of French Language Studies* 9, 143-179.
Durand, J. et J. Eychenne (2004). "Le schwa en français : pourquoi des corpus ?". *CORPUS* 3, 311-356.
Durand, J. et C. Lyche (2003). "Le projet 'Phonologie du Français Contemporain' (PFC) et sa méthodologie". In Delais-Roussarie, E. et J. Durand (eds.). *Corpus et variation en phonologie du français*. Toulouse : PUM, 213-276.
Durand, J., C. Slater, et H. Wise (1987). "Observations on schwa in Southern French". *Linguistics* 25(2), 983-1004.
Encrevé, P. (1988). *La liaison avec et sans enchaînement*. Paris : Seuil.
Fónagy, I. (1989). "Le français change de visage ?". *Revue Romane* 24(2), 225-254.
Fougeron,C. et D. Steriade (1997). "Does deletion of French schwa lead to neutralization of lexical distinctions?" *Eurospeech97*
Girard, F. et C. Lyche (2004). "PFC en Basse-Normandie: Brécey et le Domfrontais". *Bulletin PFC* 4, http://www.projet-pfc.net/**.**
Grammont, M. (1914). *Traité pratique de prononciation française*. Paris : Delagrave.
Hansen, A. B. (1994). "Etude du E caduc- stabilisation en cours et variations lexicales". *Journal of French Language Studies* 4, 25-54.
Hansen, A.B. (1997). "Le nouveau [ə] prépausal dans le français parlé à Paris". In Pérrot, J. (éd.). *Polyphonie pour Ivan Fónagy*. Paris : L'Harmattan, 173-198.
Klausenburger, J. (1994). "How abstract was/is French phonology ? A 25 year retrospective". In Lyche, C. (ed.). *French Generative Phonology : Retrospective and Perspectives.* AFLS/ESRI, 151-165.
Lacheret-Dujour, A. et Beaugendre, F. (1999). *La prosodie du français.* Paris : Éditions du CNRS.
Lacheret, A., (2003). *La prosodie des circonstants*. Paris-Louvain : Peeters.
Lacheret, A. et C. Lyche (2005). "Coding strategies in PFC: the prosody challenge". *Phonological Variation: the Case of French*, Tromsø, Norway, August 24-27, 2005, http://www.projet-pfc.net/.
Lacheret, A., C. Lyche et M. Morel (2003a). "Transcription prosodique normalisée : champ d'action et perspectives de travail pour le traitement d'une langue à accent libre comme le français". *Phonologie et phonétique du français : données et théories,* Paris, 11-13 décembre 2003.
Lacheret, A, C. Lyche et M. Morel (2003b). "Codage prosodique : lien prosodie - schwa/liaison". *Bulletin PFC* 3, 89-98.
Lacheret, A., C. Lyche et M. Morel (2004a). "Pour une transcription prosodique normalisée au sein du projet PFC (Phonologie du français contemporain) : champ d'action et

perspectives". *Les 25ièmes Journées d'Etude sur la Parole (JEP'04),* Fez, Maroc, 19- 22 avril 2004.
Lacheret, A., C. Lyche et M. Morel (2004b). "Propositions pour un codage sous Praat avec Prosogramme". *Atelier prosodie*, OFNEC, Université de Caen, 7-8 Mai 2004.
Lacheret, A. C. Lyche et M. Morel (2004c). "Éléments pour un traitement prosodique dans PFC". *Phonologie du français: Enjeux descriptifs et théoriques*, University of Calgary, Canada, 7-9 juillet 2004.
Lacheret, A., C. Lyche et M. Morel (2005). "Phonological Analysis of Schwa and Liaison within the PFC Project: how determinant are the prosodic factors?" *Interspeech 2005, 9th European Conference on Speech Communication and Technology,* Lisboa 5-8 septembre 2005.
Laks, B. (1980). *Différenciation linguistique et différenciation sociale: quelques problèmes de sociolinguistique française*. Thèse de Doctorat, Université de Paris VIII-Vincennes.
Léon, P. (1966). "Apparition, maintien et chute du "e" caduc". *La Linguistique* 2, 111-121.
Lowestamm, J. (1999). "The beginning of the Word". In J. R. Renison et K. Kühnammer (eds.). *Phonologika 1996: Syllables !?*, The Hague: Thesus 153-166.
Lucci, V. (1983). *Etude phonétique du français contemporain à travers la variation situationnelle : débit, rythme, accent, intonation, e muet, liaison, phonèmes.* Grenoble : Université de Langues et Lettres de Grenoble.
Lyche, C. et J. Durand (1996). "Testing Government Phonology ou pourquoi le choix du schwa". In Durand, J. et B. Laks (eds.). *Current Trends in Phonology,* Salford: European Studies Research Institute, 443-471.
Lyche, C. et F. Girard (1995). "Le mot retrouvé". *Lingua* 95, 205-221.
Malécot, A. (1976). "The effect of linguistic and paralinguistic variables on the elision of the French mute-e". *Phonetica* 33, 93-112.
Martinet, A. (1969). *Le français sans fard.* Paris : PUF.
Martinet, A. (1972). "La nature phonologique d'e caduc". In A. Valdman (ed.). *Papers in Linguistics and Phonetics to the Memory of P. Delattre.* The Hague: Mouton, 393-400.
Meqqori, A. et J. Durand (2003). "Manuel d'utilisation du classeur-schwa". *Bulletin PFC* 2 : 29-32. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.
Mertens, P. (2001, soumis pour publication). "A classification of French Adverbs based on Distributional, Syntactic and Prosodic Criteria". Manuscript, Departement of Linguistics, K.U.Leuven, pp. 21.
Mertens, P. (2004). "Le prosogramme : une transcription semi-automatique de la prosodie". *Cahiers de l'Institut de Linguistique de Louvain*, 30(1-3), 7-25.
Mertens, P., Auchlin, A., Goldman, J.-P., Grobet, A. et A. Gaudinat. (2001) "Intonation du discours et synthèse de la parole : premiers résultats d'une approche par balises". *Cahiers de Linguistique Française* 23, 189-209.
Montreuil, J.P. (2003). "La longueur vocalique en français de Basse-Normandie". In Delais-Roussarie, E. et J. Durand (eds.). *Corpus et variation en phonologie du français*. Toulouse : PUM, 321-348.
Neidle, C. (1979). "Syllable structure, markedness, and the distribution of the 'mute' e in French". *Phonology General Papers,* University of Trondheim.
Pagliano, C. (2005). "Initial consonant clusters in French: facts and consequence on the existence of the word in French". *Phonological Variation: the Case of French,* University of Tromsø, Norway, 25-27 August 2005, http://www.projet-pfc.net/.

Pasdeloup V. (1990). *Modèle de règles rythmiques du français appliqué à la synthèse de la parole.* Thèse de Doctorat, Université de Provence 1, Institut de Phonétique d'Aix-en-Provence.

Plénat, M. (1993). "Observations sur le mot minimal français". In Laks B. et M. Plénat (eds.). *De Natura Sonorum, Essais de phonologie*. Saint-Denis : Presses Universitaires de Vincennes.

Racine, I. et F. Grosjean (2002). "La production du E facultatif". *Journal of French Language Studies* 12(3), 307-326.

Rossi, M. (1979). "Le français, langue sans accent ?". In Fónagy, I. et P. Léon (eds.). L'accent en français contemporain, *Studia Phonetica* 15, Paris : Didier, 13-51.

Rossi, M. (1980). "Le cadre accentuel et le mot en italien et en français". In Léon P. et M. Rossi (eds.) *Problèmes de prosodie, Studia Phonetica* 17, Paris : Didier, 9-24.

Schane, S.A. (1968). *French Phonology and Morphology*. Cambridge, Mass.: MIT Press.

Schane, S.A. (1979). "The rythmic nature of English word accentuation". *Language* 55, 559-602.

Scheer, T. (2000). "L'immunité de schwa en début de mot". *Langue Française* 126, 113-126.

Tranel, B. (1984). "Floating schwas and closed syllable adjustement in French". *Phonologica 1984,* 311-317.

Tranel, B. (1987). "French schwa and nonlinear phonology". *Linguistics* 25, 845-866.

Tranel, B. (1998). "Optional schwa deletion. On syllable economy in French". In Authier, J.-M., B.E. Bullock, L.A. Reed (eds.). *Formal Perspectives on Romance Languages*. Amsterdam : Benjamins, 271-288.

Tranel, B. (2000). "Aspects de la phonologie du français et la théorie de l'optimalité". *Langue Française* 126, 39-72.

Turcsan, G. (2005). *Le mot phonologique en français du midi*. Thèse de doctorat, Université de Toulouse-Le Mirail.

Verluyten, S.P. (1982). *Recherches sur la prosodie et la métrique du français*. Thèse de Doctorat, Université d'Anvers. (Distribuée par University Microfilms International, Ann-Arbor, Michigan).

Verluyten, S.P. (1985). "Prosodic structure and the development of French schwa". In Fisiak, J. (ed.). *Papers from the 6th International Conference on Historical Linguistics.* Amsterdam : John Benjamins, 549-559.

Walker, D. (1996). "The new stability of unstable -e in French". *Journal of French Language Studies* 6, 211-229.

Walter, H. (1977). *La phonologie du français.* Paris : PUF.

Watbled, J.-Ph. (1991). "Les processus de sandhi externe en français de Marseille". *Journal of French Language Studies* 1(1), 71-91.

IVTS, un système de transcription pour la variation prosodique

Brechtje POST[1]
Elisabeth DELAIS-ROUSSARIE[2]
Anne Catherine SIMON[3]

1. Introduction

L'étude de phénomènes prosodiques comme l'accentuation ou l'intonation est souvent compliquée par le fait que ces phénomènes sont continus, et donc difficiles à représenter de façon discrète. Dès lors, tout système de transcription visant à encoder de manière discrète les événements prosodiques offre un intérêt dans de nombreux domaines comme la linguistique descriptive, la modélisation phonologique formelle ou le traitement automatique de la parole. Cela permet en effet de représenter les différents phénomènes, qu'ils soient locaux ou globaux, pour mener des études qualitatives (distribution des accents métriques, inventaire phonologique des contours intonatifs dans une variété donnée, réalisation de fonctions pragmatiques comme le contraste, le focus, etc.) ou quantitatives (nombre et fréquence des contours intonatifs, modalités de l'implémentation phonétique selon les styles, les dialectes, les tempos, etc.).

Bien qu'à l'heure actuelle les approches sur corpus jouent un rôle important tant en linguistique qu'en traitement automatique de la parole, il n'existe quasiment pas de systèmes de transcription des phénomènes prosodiques suffisamment flexibles pour encoder les connaissances prosodiques, en particulier dans des dialectes non encore décrits. Les systèmes de transcription les plus connus et les plus utilisés comme INTSINT (cf. Hirst et al (2000) ou Campione et al (2000)) ou ToBI (cf., entre autres, Beckman et al (2005)) offrent certaines limitations :

- INTSINT, en partant d'une analyse exclusivement acoustique, ne tient pas compte de données perceptives et linguistiques ;
- ToBI, de son côté, présuppose que l'inventaire phonologique des phénomènes intonatifs soit établi puisque le codage se fait directement au niveau phonologique ;
- INTSINT et ToBI privilégient nettement les phénomènes intonatifs aux dépens des phénomènes métriques, bien que les deux systèmes soient développés dans le cadre métrique-autosegmental ;
- ToBI et dans une moindre mesure INTSINT représentent le niveau local de la phrase ou de l'énoncé sans pouvoir rendre compte de phénomènes globaux au niveau du discours.

Afin de pallier ces difficultés, nous avons décidé de développer un système de transcription des phénomènes prosodiques en apportant quelques modifications au système IViE, un système de transcription qui était développé pour la transcription de variation intonative dialectale (voir par exemple Grabe et al. (2002) et Grabe & Post (2004) pour l'anglais ; et aussi l'adaptation d'Auer & Gilles (soumis) pour l'allemand. Ce choix se justifie par le fait que :

[1] Research Centre for English and Applied Linguistics, University of Cambridge, UK.
[2] CNRS, UMR 7110 / LLF (Laboratoire de Linguistique formelle), Université de Paris 7, France.
[3] Centre de recherche VALIBEL, Université catholique de Louvain, Belgique.

- IViE est un système qui est spécialement conçu pour traiter des variations dialectales, sans présupposer de connaissances phonologiques sur la variété étudiée ;
- IViE permet de traiter les phénomènes métriques, même s'ils ne sont pas réalisés par des variations mélodiques ;
- IViE peut être facilement adapté pour encoder les phénomènes non locaux comme les *downsteps*, les compressions et expansions de registres, etc. Cela est rendu possible par le fait qu'il a été développé dans un cadre métrique-autosegmental et qu'il a recours à une structure d'annotation plurilinéaire ;
- IViE propose une transcription sur plusieurs niveaux qui intègre des informations perceptives, phonétiques et phonologiques.

L'objectif que nous nous fixons dans ce papier est double :

- présenter les caractéristiques essentielles du système de transcription que nous nous proposons de développer et que nous appelons IVTS (pour "Intonational Variation Transcription System") ;
- expliquer pas à pas comment effectuer une transcription avec IVTS. Pour cela, nous nous appuierons sur quelques exemples précis extraits de données de la base PFC.

Dans un premier temps, une description générale du système IVTS sera proposée. Cela nous permettra de justifier les choix que nous avons faits, notamment dans la définition des différentes tires[4] de transcription. Puis, dans un second temps, nous expliquerons très précisément comment s'effectue une transcription sous IVTS. Bien que la majorité de nos exemples soit empruntés à différentes variétés de français, il est important de noter que IVTS n'est pas un système conçu pour une langue spécifique. L'architecture générale peut en effet permettre de transcrire n'importe quelle langue ou variété, pour autant qu'on prenne en compte certaines différences typologiques qui affecteront le choix des étiquettes disponibles pour une langue ou une variété particulière. Par exemple, la forme des étiquettes de la couche d'annotation phonétique locale dépend entre autres du caractère gouverné à gauche ou à droite du domaine d'implémentation tonale dans la langue en question (voir section 2.3 ci-dessous).

2. Description du système de transcription IVTS

Tout comme IViE, le système IVTS a recours à plusieurs tires d'annotation pour encoder les différentes informations prosodiques et linguistiques. Une transcription est organisée sur six niveaux, parmi lesquels quatre sont utilisés pour transcrire les phénomènes prosodiques. La transcription prend donc la forme suivante :

[4] Dans le logiciel d'annotation Praat, qui est utilisé dans le cadre du projet PFC, les couches d'annotation textuelle qu'on peut aligner sur le son sont dénommées *tiers*. Nous adoptons la version francisée de cette dénomination : tire.

Tableau1. *Six niveaux de transcription*

	Tire Commentaires (ou "comment tier")
Encodage des informations prosodiques	Tire Phonologie (ou "phonological tier")
	Tire Perception phonétique globale (ou "global auditory phonetic tier")
	Tire Perception phonétique locale (ou "local auditory phonetic tier")
	Tire Proéminences (ou "Rhythmic tier")
	Tire *Mots* (ou "Orthographic tier")

IVTS comporte un niveau de transcription qui n'apparaissait pas dans IViE (Grabe et al., 2002, 2004). Il s'agit de la tire de *Perception phonétique globale* qui sert à encoder les phénomènes phonétiques qui se produisent sur des empans temporels plus importants (phénomènes de compression ou d'expansion de registre, downsteps, etc.). Des exemples précis d'utilisation de cette tire sont proposés dans la section 3.4.

Le codage des informations prosodiques sur plusieurs niveaux offre de nombreux avantages, qui ont été mentionnés dans les publications consacrées à IViE (Grabe & Post, 2004). Tout d'abord, en utilisant explicitement quatre niveaux pour annoter les phénomènes prosodiques, le système permet de décrire les variations dialectales de nature taxonomique (Ladd, 1996). En effet, la variation (régionale) de la prosodie en français peut relever de différences qui affectent la distribution des syllabes proéminentes, la réalisation phonétique des contours intonatifs (implémentation phonétique), ou l'inventaire phonologique des contours présents dans une variété. Toutes ces différences peuvent être directement encodées dans la tire d'annotation dont elles relèvent. En outre, s'il existe des restrictions dans les possibilités d'agencement des contours, elles peuvent être dérivées à partir de la tire *Phonologie.*

Deuxièmement, le recours à plusieurs niveaux d'annotation prosodique permet d'obtenir plus d'uniformité et de transparence dans la réalisation des transcriptions. Comme les transcripteurs peuvent être amenés à analyser des variétés qui n'ont jamais été décrites auparavant, il est intéressant d'avoir accès aux différents types d'informations (perceptives, phonologiques ou rythmiques) qui ont conduit à faire telle ou telle analyse. Le processus dans son ensemble n'en est que plus transparent. De même, lorsqu'un inventaire phonologique des contours n'est pas encore établi, les transcripteurs peuvent d'abord travailler sur la tire *Perception phonétique locale* où ils encodent les différents mouvements mélodiques qu'ils perçoivent, et ensuite tenter de proposer une analyse phonologique. Ceci étant, ils auront toujours la possibilité de remettre en cause les hypothèses qu'ils auront formulées pour l'analyse phonologique.

À titre indicatif, nous proposons dans la figure 1 un extrait de transcription effectuée avec IVTS. Les différentes tires d'annotation sont alignées avec la courbe de fréquence fondamentale et le signal. Cet exemple a été fait sous PRAAT (Boersma et Weenink, 1996), mais d'autres logiciels d'analyse de la parole peuvent être utilisés pour faire une transcription avec IVTS (WinPitch, Waves, etc.).

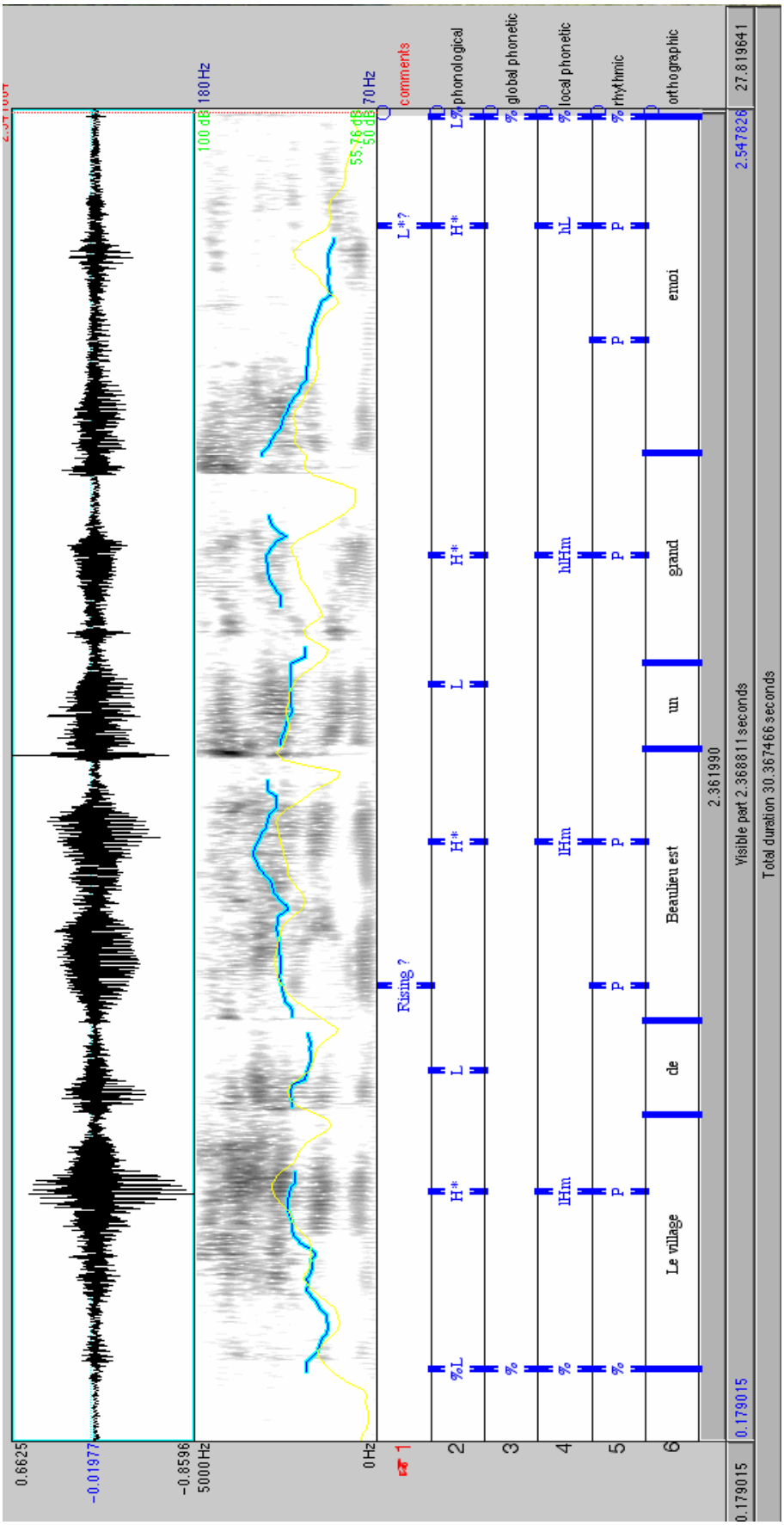


Figure 1. *Fragment d'une transcription IVTS faite sous PRAAT*

L'exemple donné est représentatif du protocole d'enquête du projet PFC. Il s'agit d'une phrase lue (extraite de la tâche de lecture d'un texte) par un locuteur belge issu de la région de Liège : "*Le village de Beaulieu est en grand émoi*".

La figure 1 permet de montrer que les mots prononcés sont transcrits dans la tire *Mots*, où chaque mot est aligné avec la portion de signal qui lui correspond. Les autres étiquettes sont fournies dans des *point tiers*. Elles sont alignées avec les frontières de groupe ou avec le noyau des syllabes proéminentes. Celles-ci sont notées dans la tire *Proéminences* à l'aide du symbole P. Les mouvements mélodiques locaux et globaux sont encodés dans les tires *Perception phonétique locale* et *Perception phonétique globale*. La catégorisation phonologique de ces mouvements est ensuite fournie dans la tire *Phonologie.* Les conventions et les symboles de transcription utilisés pour les tires prosodiques seront expliqués et motivés dans la section 3.

Pour permettre une comparaison entre ce système de transcription et IViE, la figure 2 fournit deux exemples de transcriptions faites dans IViE. Ces exemples proviennent du corpus IViE (Grabe et al, 2001). Comme nous pouvons le voir, le système de transcription est assez similaire mais les étiquettes utilisées pour encoder les informations phonétiques ou phonologiques sont différentes. Comme le français et l'anglais ont des systèmes intonatifs différents, l'ensemble des étiquettes retenues pour établir l'inventaire des catégories prosodiques est de fait différent. Les deux exemples dans la figure 2 permettent également de montrer comment les différences dialectales sont représentées, même si elles ne relèvent que de l'implémentation phonétique (cf. la tire *Perception phonétique locale*). Dans le cas présent, une même catégorie phonologique H*L (c'est-à-dire un mouvement descendant) est réalisée de deux façons distinctes dans les dialectes de Leeds et de Cambridge (cf. Grabe et al, 2000). Comme le montre la figure 2, il est possible avec IViE de préciser que cette différence entre les parlers de Cambridge et de Leeds se situe au niveau phonétique.

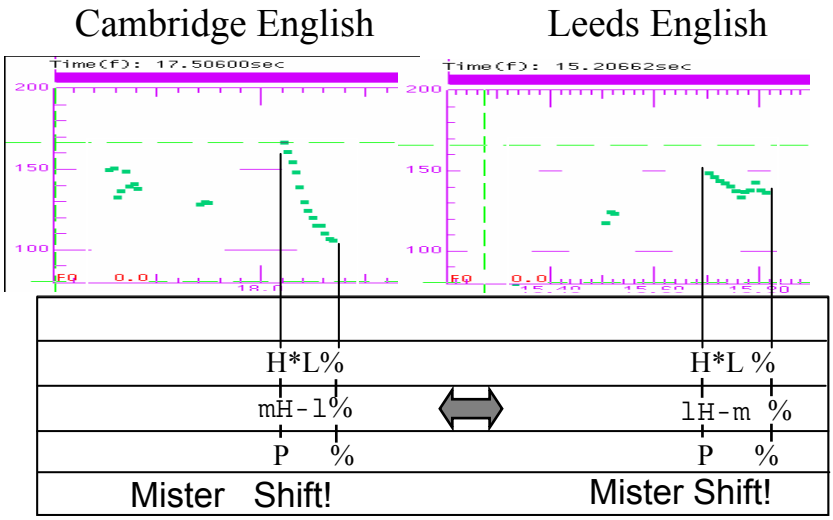


Figure 2. *Transcription avec IViE*

3. Comment effectuer une transcription sous IVTS

Les caractéristiques générales de IVTS ont été présentées dans la section précédente. Nous allons maintenant montrer pas à pas comment s'effectue une transcription avec ce système. Nous en profiterons pour expliquer quelles conventions et quels symboles de transcription sont utilisés à chaque niveau.

3.1. Construire la tire Mots

Comme le montre la figure 1, la tire *Mots* fournit la transcription orthographique des mots prononcés. Elle est construite à partir d'un alignement entre chaque mot et la portion de signal qui lui correspond. Dans un projet comme PFC, cette tire peut s'ajouter à la tire *orthographique* et se construit en ajoutant des frontières d'intervalles de façon à ce que chaque mot produit soit donné dans un intervalle. Afin d'effectuer cette segmentation, plusieurs choix doivent être faits, notamment en français, pour traiter des phénomènes comme l'enchaînement, la liaison et l'élision : soit les consonnes de liaison et d'enchaînement sont données avec le mot orthographique dont elles dépendent comme dans le tableau (2) a, soit elles sont fournies avec le mot sur lequel elles s'enchaînent comme dans le tableau (2) b.

Tableau 2. *Traitement des enchaînements, liaisons et élisions*

a. d'après la graphie

Segmentation retenue dans *Mot*	un [œ]	petit [pətit]	enfant [ãfã]	d' [d]	Allemagne [almaɲ]
Réalisation (API)	[œpətitãfãdalmaɲ]				
Énoncé	un petit enfant d'Allemagne				

b. d'après la prononciation (syllabation)

Segmentation retenue dans *Mot*	un [œ]	peti [pəti]	tenfant [tãfã]	d'Allemagne [dalmaɲ]
Réalisation (API)	[œpətitãfãdalmaɲ]			
Énoncé	un petit enfant d'Allemagne			

La seconde approche correspond en gros à une segmentation en mots prosodiques. Nous l'avons généralement adoptée comme en témoigne la segmentation fournie dans la figure 3, où *ira-t-il* est traité comme un seul mot.

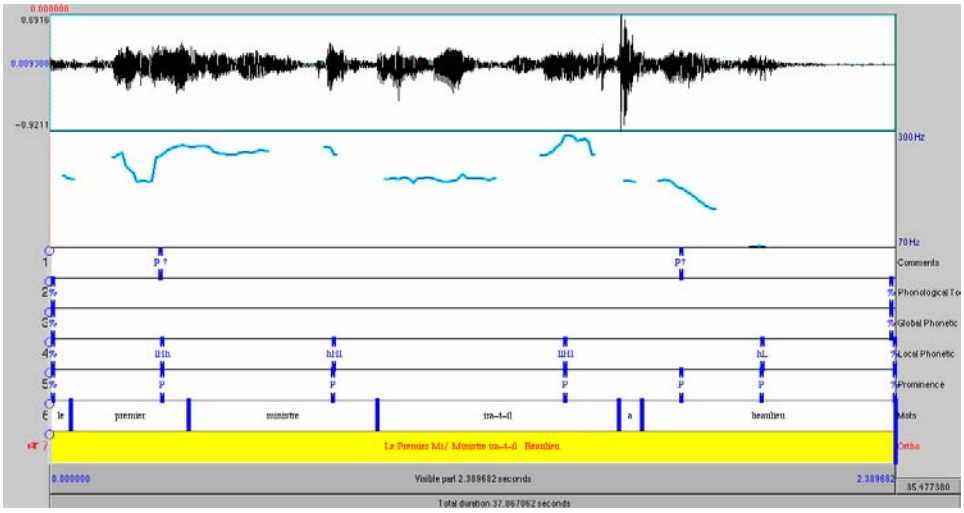


Figure 3. *Segmentation sur la tire* Mots

3.2. La tire Proéminences

La tire *Proéminences* est utilisée pour indiquer quelles syllabes sont perçues comme proéminentes par le transcripteur. De telles syllabes sont identifiées à l'aide d'une étiquette

P, alignée avec le noyau vocalique de chaque syllabe proéminente ou accentuée. Nous entendons ici par *proéminente* le fait que la syllabe soit perçue comme plus saillante que les syllabes qui l'entourent. Cela peut être marqué au niveau phonétique par un allongement de la durée, par un mouvement mélodique et/ou par un accroissement de l'intensité, et au niveau phonologique par un accent [mélodique]. En outre, deux autres types d'information figurent sur la tire *Proéminences* :

- les frontières de groupes intonatifs majeurs[5] sont notées par le symbole % ;
- les pauses, ruptures ou silences résultant d'hésitations ou d'auto-corrections sont indiqués par le symbole #.

Il est important de noter que l'utilisation de l'étiquette P repose **sur l'existence d'une réelle saillance phonétique**, et non sur une propriété structurelle abstraite du mot ou du groupe de mots comparable à la notion de *(lexical) stress*. En français standard, les syllabes proéminentes peuvent être réalisées sur la dernière syllabe métrique (c'est-à-dire dont le noyau n'est pas un schwa) des mots lexicaux (N, Adj, Adv, V), mais également sur les syllabes initiales ou pénultièmes de ces mêmes mots s'ils contiennent plus de deux syllabes (Pasdeloup, 1990; Delais-Roussarie, 1995 ; Post, 2000).

L'étiquette **P** ne distingue pas entre les accents en position finale de mots et les accents qui interviennent dans une autre position (initiale, pénultième, ou autre), même si la plupart des auteurs sont d'accord pour dire qu'une telle distinction existe en français standard, et qu'il semble fort probable que ce soit aussi le cas d'autres dialectes du français. La raison pour laquelle IVTS ne la reconnaît pas, du moins sur la tire *Proéminences*, est que la distinction se base surtout sur une analyse fonctionnelle des accents du français, plutôt que sur une simple perception de syllabes proéminentes (accentuées ou autres). Cependant, la position de l'accent est encodée de manière indirecte par l'alignement de l'étiquette P avec la tire *Mots* Mais comme on a remarqué que les variétés pouvaient diverger selon la fréquence des accents en position non finale (Carton 1989), le transcripteur peut noter la position de l'accent sur la tire *Commentaires* afin d'avoir accès à l'information d'une façon plus directe (en utilisant les étiquettes PF et PN, par exemple).

Pour effectuer une transcription des proéminences, le transcripteur écoute attentivement une portion de signal de quelques secondes correspondant à un ou plusieurs groupes intonatifs majeurs. Il assigne l'étiquette **P** aux noyaux des syllabes perçues comme plus proéminentes. Lorsque les saillances sont réalisées par une variation mélodique, des étiquettes alignées avec la proéminence sont ajoutées dans les tires *Perception phonétique locale* et *Phonologie*. Dans les cas où une proéminence résulte uniquement d'un allongement de la durée syllabique, et si elle n'est pas accompagnée de variation de fréquence fondamentale par rapport aux syllabes qui précèdent, aucune étiquette alignée avec la position **P** n'est fournie dans les tires *Perception phonétique locale* et *Phonologie*. Le phénomène peut cependant être signalé dans la tire *Commentaires.*

Nous proposons dans les figures 4 et 5 deux transcriptions d'un même énoncé "Le village de Beaulieu est en grand émoi" produit par deux locuteurs d'origine géographique différente. Dans la figure 4 qui correspond à la production d'un locuteur belge, deux syllabes proéminentes marquées P dans la tire *Proéminences* (appelée *Rhythmic* sur la figure 4) ne sont pas associées à des mouvements mélodiques. Elles sont respectivement réalisées sur la syllabe initiale de "Beaulieu" et sur la syllabe initiale de "émoi". Parmi elles, la seconde apparaît comme plus nette, puisque le transcripteur n'émet aucune réserve (cf. tire *Commentaires*, où figure l'étiquette "Rising ?").

[5] Ou "intonation phrase boundaries", voir par exemple Pierrehumbert (1980).

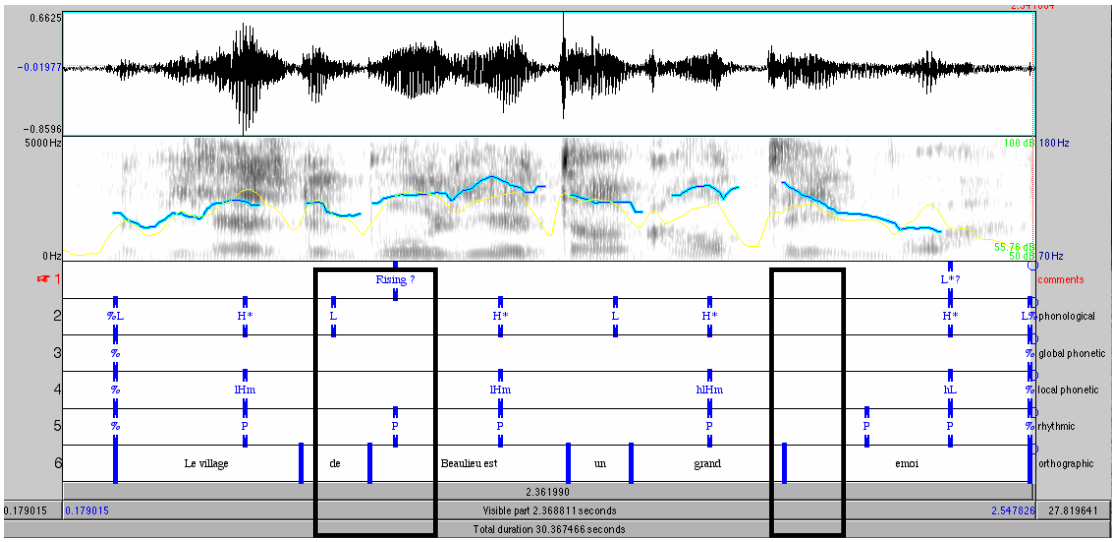


Figure 4. *Transcription d'un énoncé prononcé par un locuteur belge*

Comparons maintenant cette réalisation avec celle d'un locuteur marseillais, qui est représentée à la figure 5. Dans cette figure, les syllabes transcrites comme proéminentes sont généralement porteuses d'un mouvement mélodique puisqu'elles sont alignées avec des étiquettes dans la tire *Perception phonétique locale.* La seule exception est en fait la syllabe "veau" correspondant avec une amorce de mot (cf. notation de l'hésitation).

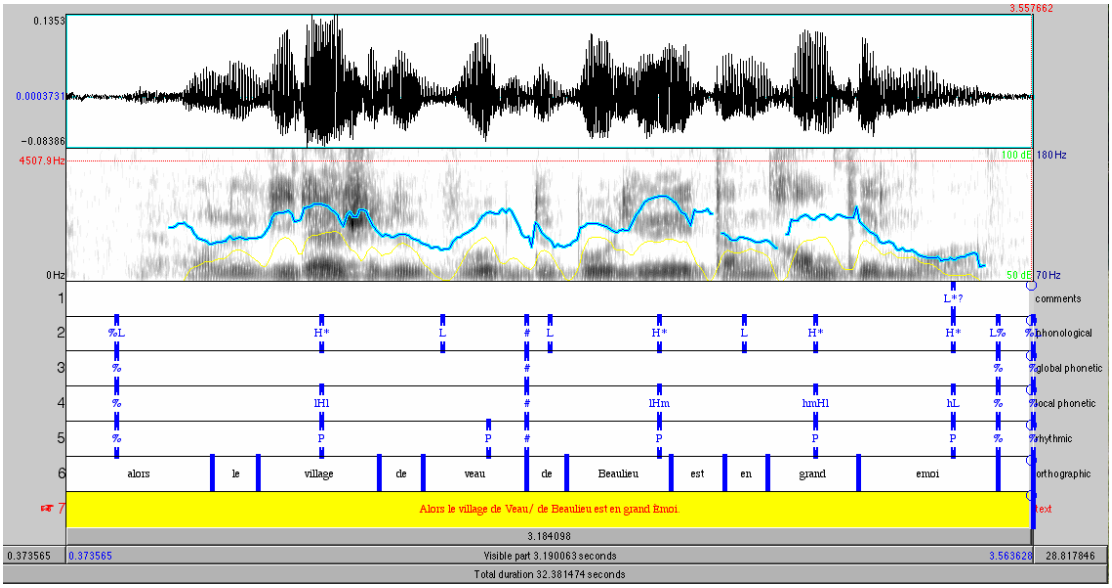


Figure 5. *Transcription d'un énoncé prononcé par un locuteur marseillais*

Bien que limitée, la comparaison que nous avons faite entre les transcriptions des productions des deux locuteurs peut amener à penser qu'une proéminence marquée par un allongement de la durée est généralement réalisée sur la syllabe initiale du dernier mot prosodique des groupes intonatifs dans une variété de français typique de la région de Liège (Belgique). En revanche, un accent de ce type est très rare en français standard ou dans d'autres variétés comme celle de Marseille**.**

3.3. La tire Perception phonétique locale

La tire *Perception phonétique locale* est utilisée pour transcrire la forme du mouvement mélodique réalisé sur toute syllabe proéminente ainsi que sur les syllabes adjacentes. Sont également notées sur cette tire les frontières des groupes intonatifs majeurs (avec le symbole %). L'accent est mis ici sur la configuration mélodique et les modalités d'alignement des mouvements intonatifs réalisés sur les syllabes proéminentes et sur celles qui les entourent. Les phénomènes mélodiques plus globaux tels que le registre ou le downstep ne sont donc pas encodés à ce niveau. On peut remarquer, et c'est une limite du système de codage, qu'aucune catégorie ne note explicitement les phénomènes d'allongement, qu'ils soient prédictibles (selon les règles de l'allongement en français standard présentées dans Hambye & Simon 2005) ou non prédictibles, et donc potentiellement marqués.

La transcription des mouvements mélodiques se fait sur des bases perceptives et auditives, et non à partir d'une analyse acoustique de la fréquence fondamentale. Elle s'effectue à partir d'une écoute attentive d'une portion de signal correspondant au **domaine d'implémentation accentuel** (noté ID, d'après le *domaine d'association tonale* de Gussenhoven, 1984). L'extension de ces domaines varie selon les langues. En Anglais, le ID correspond à un pied métrique auquel s'ajoute la syllabe précédente (cf. Grabe et Post, 2000). En français, en revanche, ce domaine se définit de façon différente dans la mesure où :

- la notion de *pied* n'est pas pertinente pour rendre compte de la prosodie de cette langue ;
- l'accentuation métrique du français repose sur une dominance à droite, au moins pour l'accent final (Verluyten, 1982; Dell, 1984; Di Cristo, 1999; Post, 2000; Delais-Roussarie, 2000).

La définition des domaines d'implémentation se fait à partir des syllabes proéminentes dont la proéminence est réalisée mélodiquement. En français, tout ID comprend (i) la syllabe proéminente notée P; (ii) toutes les syllabes qui la précèdent jusqu'à la syllabe proéminente précédente ou jusqu'à une frontière d' IP ; (iii) la syllabe qui la suit immédiatement. À titre indicatif les ID définis à partir de la réalisation de l'énoncé représenté dans la figure 1 sont listés sous (1).

(1) ID pour l'énoncé le village de Beaulieu (es)t en grand émoi.
Réalisation effective où les syllabes proéminentes sont en gras et soulignées
[l.vi.**laʒ**.d.bo.**ljø**.tɑ̃.**gʀɑ̃**.te.**mwa**]
Liste des ID
(le village de)
(de beaulieu est en)
((es)t en grand é)
((t)émoi)

Comme nous le voyons, deux ID consécutifs partagent généralement une syllabe. Ainsi la syllabe *de* dans *le village de Beaulieu* appartient aux deux premiers ID. Les ID situés en début et en fin d'IP ont respectivement une syllabe initiale et une syllabe finale qui n'apparaissent dans aucun autre ID.

L'annotation s'effectue en quelque sorte syllabe par syllabe au sein des domaines d'implémentation. L'étiquette retenue pour encoder un mouvement mélodique perçu sur une syllabe proéminente est choisie en fonction des mouvements mélodiques perçus sur les syllabes adjacentes **dans le même ID**. Cette notation n'est donc pas relative aux ID

successifs. Les niveaux retenus pour l'encodage mélodique sont haut (H ou h), moyen (M ou m) et bas (L ou l) et sont tous relatifs. En outre, ces niveaux sont notés en lettres majuscules dès lors que leur cible est alignée sur le noyau des syllabes proéminentes. La séquence "lHm" qui est réalisée sur les deux premiers ID de l'énoncé représenté sur la figure 1 indique un mouvement mélodique montant du niveau bas vers le niveau haut, et où le point H est atteint sur la syllabe accentuée. Ce mouvement est ensuite suivi d'une légère descente. À ce titre, il est important de noter que l'étiquette "moyen" (M ou m) ne correspond pas à un niveau absolu. Comme le montre la figure 1, les niveaux donnés par les étiquettes sont des **valeurs relatives** définies par rapport aux syllabes qui se trouvent à l'intérieur du même ID. Les niveaux ne tiennent pas compte des événements mélodiques réalisés dans des ID adjacents (ou autres), et ne correspondent pas à un niveau absolu défini en quelque sorte par le registre de la voix du locuteur. Ainsi, l'étiquette "moyen" n'apparaît que lorsque les étiquettes "bas" et "haut" ont déjà été utilisées pour annoter deux autres syllabes dans un même ID, lesquelles étant réalisées toutes les deux à des niveaux de hauteur différents.

D'une façon générale, les étiquettes attribuées aux différents domaines d'implémentation sont composées d'une séquence de symboles parmi lesquels celui du milieu est une majuscule qui prend la forme H, M ou L. Ceci étant, quelques cas diffèrent :

- lorsque la syllabe initiale du ID est accentuée, les étiquettes se composent de deux symboles dont le premier est une majuscule (alignée avec la syllabe proéminente). Des exemples d'étiquettes sont Lh, Lm, Ml, Mh, Hm, Hl ;
- de la même façon, lorsque la syllabe proéminente est la dernière d'un groupe intonatif majeur, l'étiquette s'achève par un symbole en majuscule (lM, lH, mH, etc.) ;
- lorsqu'un changement mélodique significatif se produit sur la syllabe proéminente initiale ou finale, il est transcrit au moyen d'une séquence de symboles en majuscules (LH pour un ton montant, HL pour un ton descendant, etc.).

Dans la figure 6, un ton composé (ou dynamique) de la forme LH est réalisé sur la syllabe [mã] de l'adverbe *récemment*. Le mouvement mélodique perçu sur l'ensemble du domaine d'implémentation auquel appartient cette syllabe est mLH.

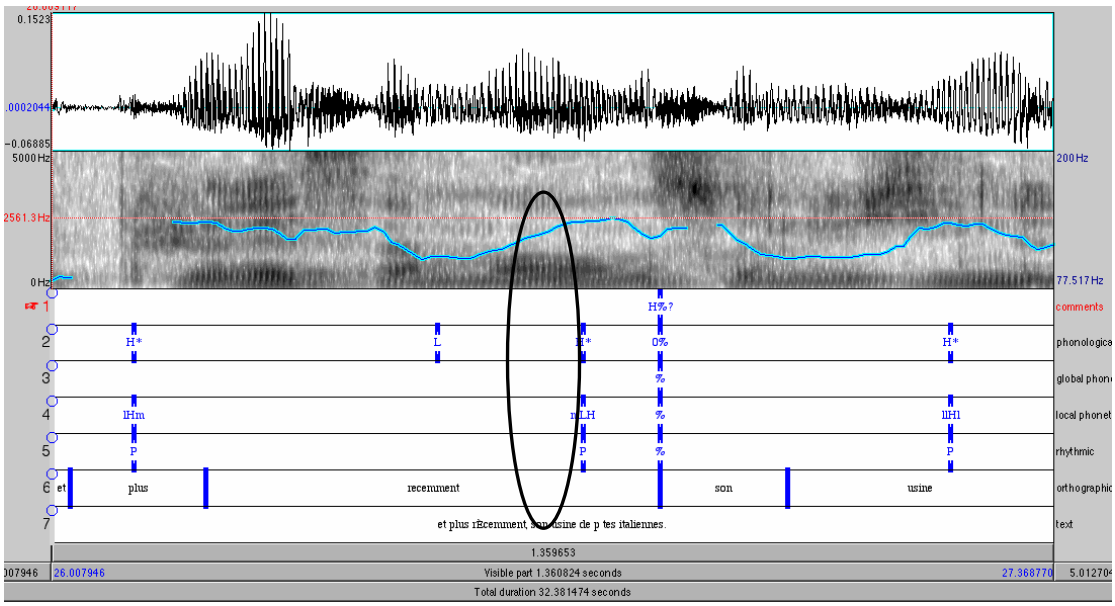


Figure 6. *Réalisation d'un ton composé sur une syllabe proéminente*

Si le contour associé au domaine d'implémentation avait été de la forme lmH, le mouvement du niveau moyen vers le niveau haut se serait produit sur deux syllabes : la syllabe proéminente où la cible H est atteinte et la syllabe qui lui est immédiatement adjacente sur la gauche (où le m est réalisé).

Dans la plupart des cas, les étiquettes assignées aux domaines d'implémentation sont composées de trois types d'éléments :

- un niveau mélodique noté en minuscule qui rend compte de la hauteur mélodique sur laquelle est réalisée la syllabe initiale du domaine d'implémentation ;
- un ton noté par une majuscule simple (H, L ou M) ou par une suite de majuscules (MH, LH, etc.), associé avec la syllabe proéminente ;
- un niveau mélodique noté en minuscule qui indique la hauteur de la dernière syllabe non proéminente du domaine d'implémentation.

Ceci étant, un symbole additionnel en minuscule peut être utilisé pour décrire le mouvement mélodique qui est réalisé sur la séquence de syllabes précédant la syllabe proéminente, si celui-ci ne peut pas être directement déduit d'une interpolation entre la cible initiale et la cible réalisée sur la syllabe proéminente. Ainsi, si un mouvement montant est réalisé sur la syllabe proéminente et est précédé d'un plateau bas, la transcription sera llHm. Dans ce cas, le premier l est associé à la syllabe initiale du ID, tandis que le second est associé à la syllabe précédant la syllabe proéminente.

Comme nous le voyons dans la figure 7, la syllabe qui précède immédiatement la syllabe proéminente ne doit pas nécessairement être associée à un symbole tonal. C'est ce qui se passe si le mouvement mélodique peut être interpolé à partir de la valeur de la cible initiale et de celle de la cible associée à la syllabe proéminente. D'une façon générale, la transcription mélodique perceptive doit se faire de manière à limiter autant que possible le nombre de symboles utilisés pour construire les étiquettes associées aux ID (ce principe est également suivi dans IViE).

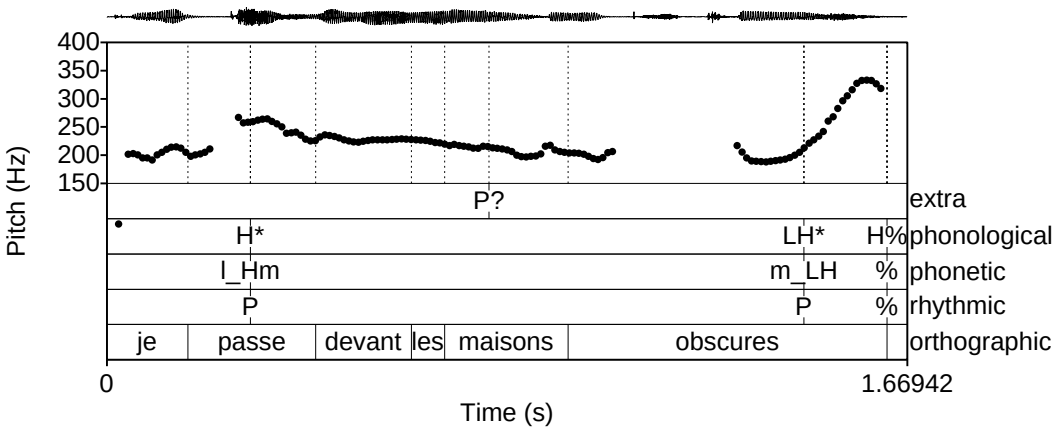


Figure 7. *Exemple d'interpolation dans le codage phonétique*

3.4. La tire Perception phonétique globale

Dans IVTS, nous avons développé la tire *Perception phonétique globale* qui n'existait pas dans IViE. Elle a pour but de rendre compte des phénomènes mélodiques qui sont réalisés sur un empan qui dépasse le cadre d'un ID (domaine d'implémentation), comme les *downsteps*, les *upsteps* ou les changements de registre. Alors qu'au niveau phonétique local, on encode la façon dont sont perçus auditivement les mouvements mélodiques

associés aux syllabes proéminentes, pour permettre une identification des accents mélodiques *(pitch accents)*, au niveau phonétique global, le transcripteur encode des phénomènes mélodiques qui se produisent sur plusieurs ID. Ces phénomènes modifient la réalisation d'un mouvement mélodique local dans un ID donné en fonction des mouvements mélodiques réalisés dans d'autres ID, ces différents ID se groupant à différents niveaux comme le syntagme, la phrase, l'énoncé, et même le paragraphe. Un exemple de codage est fourni dans la figure 8.

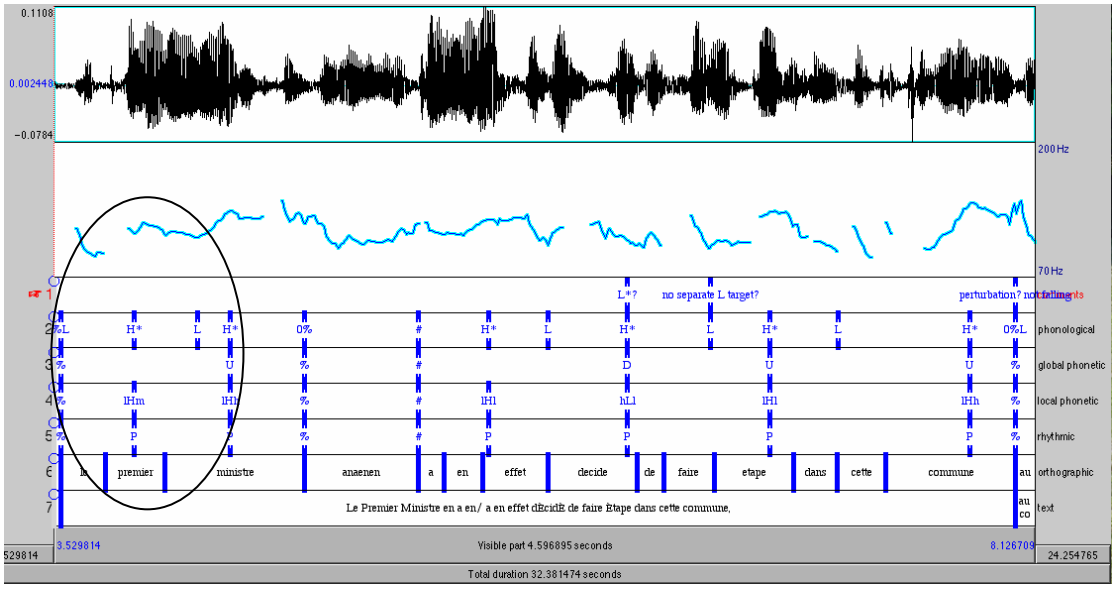


Figure 8. *Exemple de codage pour la tire* Perception phonétique globale

Dans cette figure, nous pouvons voir qu'une étiquette U est assignée à la syllabe [nistʁ] dans la tire *Perception phonétique globale*. Elle permet de rendre compte du fait que la cible mélodique H de cette syllabe (encodée dans la tire *Perception phonétique locale* à l'aide de lHh) est plus haute que la cible H associée à la syllabe [pʁə] de *premier*. Les étiquettes choisies à ce niveau d'annotation reflètent là aussi les phénomènes mélodiques perçus par le locuteur. Elles ne relèvent donc pas nécessairement de phénomènes phonologiques pertinents dans la langue étudiée. En français standard, par exemple, il arrive que dans un énoncé les successions de mouvements mélodiques montants soient réalisées avec un abaissement, mais cela résulte davantage de l'implémentation phonétique que de la phonologie (cf. Post, 2000), et sera donc encodé dans la tire Perception phonétique globale, et éventuellement dans la tire Perception phonétique locale. L'ensemble des étiquettes qui peuvent être utilisées pour encoder les phénomènes mélodiques non locaux dans la tire Perception phonétique globale est fourni au tableau 3.

Tableau 3. *Symboles utilisés pour la tire* Perception phonétique globale

Phénomènes	Descriptions	Symboles	Exemples stylisés
Upstep	Réhaussement local du niveau d'un mouvement mélodique montant.	U	
Downstep	Abaissement local du niveau d'un mouvement mélodique montant.	D	
Pitch span / Registre	Expansion ou compression du registre	S	
Reset / Register shift	Le registre est abaissé ou compressé sur l'ensemble de l'empan.	R	

Ces étiquettes peuvent être complétées par :

- d'autres symboles comme IP (pour groupe intonatif) ou UT (pour énoncé) afin d'indiquer l'empan sur lequel le phénomène non local se produit ;
- des indices (chiffrés) si le même phénomène se répète dans une séquence. Ainsi, par exemple, si l'énoncé le village de Beaulieu est en grand émoi (cf. (3)) est réalisé avec une succession d'abaissement des cibles hautes sur les syllabes [laʒ], [ljø], [gʀɑ̃] et [mwa], on peut assigner l'étiquette D1 à [ljø], D2 à [gʀɑ̃] et D3 à [mwa], les étiquettes étant alignées avec les positions étiquetées dans les autres tires.

Comme ce niveau d'annotation a été élaboré pour IVTS, il n'a pas encore été utilisé pour coder ce type de phénomènes sur des données variées. Aussi avons-nous l'intention d'évaluer la pertinence de cette tire de façon plus approfondie sur un grand nombre de données diverses. Cela permettra également de déterminer si les phénomènes ainsi transcrits et les étiquettes proposées sont satisfaisants.

3.5. La tire Phonologie

Les contours intonatifs sont catégorisés phonologiquement dans la tire *Phonologie*. Les symboles que nous utilisons sont liés à une approche métrique auto-segmentale, et ne sont donc pas compatibles avec n'importe quel modèle de l'intonation du français. En fonction de ses options théoriques, le chercheur pourrait adopter un autre ensemble d'étiquettes phonologiques. Les étiquettes utilisées pour effectuer cette catégorisation sont données dans le tableau (4). Dans les transcriptions, elles sont alignées avec les étiquettes fournies dans les autres tires (la tire *Proéminences* et les deux tires *Perception phonétique*).

Tableau 4. *Liste des symboles utilisés pour l'encodage dans la tire* Phonologie

Primitives tonales				
Tons		*Modifieurs*	*Tons de frontière*	
H*	H	^ : upstep	%H	H%
L*	L	! : downstep	%L	L%
+		>: propagation	%	%

Parmi ces différentes primitives,

- les **tons** servent à encoder des phénomènes mélodiques locaux dont la réalisation se produit sur une syllabe donnée ;
- les **modifieurs** sont utilisés en combinaison avec un ton pour indiquer que la réalisation de ce dernier résulte de modifications phonologiques plutôt globales comme celles qui sont généralement encodées dans la tire *Perception Phonétique Globale* (*uptsep, downstep*, etc.) (NB. Encodage phonologique seulement si la modification en question est de nature phonologique.) ;
- les **tons de frontière** sont associés aux frontières gauche ou droite des unités intonatives (syntagme intonatif, IP, groupes de souffle, etc.)[6].

Il existe cependant certaines règles dans l'utilisation de ces symboles :

- les tons de frontière sont toujours composés de l'élément **%** qui est placé à la gauche ou à la droite de l'élément mélodique (H, L, voire M ou 0) selon qu'il s'agisse respectivement d'un tons de frontière gauche ou d'un ton de frontière droite ;
- les accents mélodiques (ou *pitch accent*) sont marqués par le symbole ***** s'ils sont associés à des syllabes métriquement proéminentes ;
- Le diacritique + est parfois utilisé pour coordonner les différents tons qui composent un accent mélodique bi- ou tri-tonal (L+H* par exemple). Dans IViE et IVTS, le symbole + utilisé dans les accents mélodiques indique surtout que le ton non marqué (précédant ou suivant celui qui porte le symbole *****) fait partie de l'accent mélodique. Ce symbole ne marque donc pas l'existence d'une relation temporelle entre les tons (Ladd et al., 1999).

Les primitives tonales peuvent se combiner en accents mélodiques et en tons de frontière pour générer (ou construire) des contours intonatifs. L'étude des combinaisons permet alors d'établir l'**inventaire des contours tonals** dans une (variété de) langue particulière. Ainsi, de façon analogue à ce qui se passe au niveau segmental, l'organisation suprasegmentale est décomposée en plusieurs niveaux de représentation : les variations mélodiques continues et non-contrastives sont transcrites au niveau des tires *Perception phonétique locale* et *Perception phonétique globale*, et leur catégorisation comme phénomènes mélodiques contrastifs est encodée dans la tire *Phonologie*, où les différentes unités tonales contrastives (accents mélodiques et tons de frontières) sont répertoriées[7].

Comme IVTS est un outil de transcription qui peut être utilisé pour différentes langues ou variétés, il propose un large ensemble d'étiquettes qui ne seront pas nécessairement pertinentes pour décrire telle ou telle langue particulière. En français standard, par exemple, il ne semble pas nécessaire d'opérer au niveau phonologique une distinction entre trois tons de frontière gauche distincts (cf. Post, 2000) : ton de frontière

[6] Les unités intonatives majeures sont à distinguer des groupes prosodiques composés de syllabes consécutives dont la dernière porte un proéminence (et qu'on appelle Groupe Intonatif, Groupe Rythmique ou Syntagme Phonologique). Ces groupes intonatifs peuvent être construits sur des bases phonologiques, discursives ou syntactico-sémantique. Ils correspondent plutôt à une phrase noyau, c'est-à-dire à une structure [NP VP] sans extrapositions ou interruptions. Dans certaines grammaires phonologiques de l'intonation, le domaine d'ancrage des contours intonatifs est déterminé à partir d'informations sémantiques (cf. Delais-Roussarie, 2005). Dans ce cas, les tons de frontières sont réalisés sur les frontières des domaines.

[7] Établir une distinction entre les phénomènes phonétiques et les phénomènes phonologiques n'est pas une tâche facile dès lors qu'on étudie la prosodie et l'intonation ; il n'existe pas en effet d'outils ou de procédures suffisamment robustes pour permettre au chercheur de décider de la nature d'un phénomène. En outre, il n'existe pas non plus de cadre théorique globalement accepté pour rendre compte des variations intonatives (Ladd, 1996 et Gussenhoven, 2004 pour des discussions sur ce point).

haut (noté %H), ton de frontière bas (%L) et ton de frontière non spécifié (noté %). De la même façon, certains modifieurs peuvent rendre compte de phénomènes qui ont un statut phonologique dans une variété (ou une langue), mais qui, en revanche, relèvent de l'implémentation phonétique dans d'autres. En français par exemple, le downstep n'est pas un phénomène phonologique (cf. section 3.4), il ne sera donc pas encodé dans la tire *Phonologie,* mais dans la seule tire *Perception Phonétique Globale.*

Il est néanmoins important de noter que l'ensemble des étiquettes proposé sous (5) n'est pas théoriquement neutre. Il repose essentiellement sur une approche métrique-autosegmentale selon laquelle :

- il est possible de proposer une description phonologique des phénomènes intonatifs qui soit autonome de leur manifestation phonétique ;
- les contours intonatifs peuvent être analysés et représentés à partir d'accents mélodiques ("pitch accents") et de tons de frontière. Le terme accent mélodique ("pitch accent") est utilisé ici dans un sens purement formel, et non fonctionnel : les accents mélodiques sont réalisés sur des syllabes proéminentes (notées **P** dans la tire *Proéminences*) par un mouvement mélodique (déjà encodé dans la tire *Perception Phonétique locale*).

Une discussion s'impose à propos de l'interprétation non fonctionnelle des accents mélodiques ("pitch accents") que nous donnons ici. Dans certaines langues germaniques, par exemple, les accents mélodiques ont une fonction linguistique importante, puisque leur réalisation est associée à des phénomènes tels que la focalisation ou la topicalisation, fonctions que les "pitch accents" ne possèdent normalement pas dans une langue comme le français, qui utilise les mouvements mélodiques principalement pour démarquer les groupes de mots. La fonction fortement démarcative des mouvements mélodiques en français (et dans d'autres langues qui se comportent de la même manière) pourrait être avancée comme argument pour plaider contre une **distinction entre les accents mélodiques et les tons de frontière**. Cependant IVTS maintient cette distinction pour trois raisons :

1. afin que IVTS soit applicable à un large éventail de langues, les événements tonals localisés aux frontières et les syllabes métriquement fortes doivent être distingués ;
2. même au sein du système prosodique du français, la possibilité de généralisations à propos des réalisations de contours dans certaines variétés (par ex. les variétés méridionales qui autorisent l'accent en position pénultième) va se fonder sur une telle distinction ;
3. on argumentera même que cette distinction devrait aussi être faite en français standard dans la mesure où (a) elle rend compte de manière transparente et économique d'un fait de distribution : une plus grande variété de contours peuvent se produire à des frontières d'IP plutôt qu'à d'autres points de l'énoncé (cf. Post 2000) ; enfin, (b) cette distinction permet de faire l'hypothèse que la proéminence métrique est pertinente pour la réalisation des mouvements mélodiques (plutôt que la pure fonction démarcative), comme il a été démontré par Dell (1984) sur la base de la différence d'alignement des contours intonatifs dans des énoncés qui se terminent par une syllabe finale pleine et dans ceux dont la syllabe finale est réalisée par un schwa (cf. Dell 1984, aussi discuté dans Post 2000 : 177-179).

Notons pour finir que les différentes étiquettes proposées sont tout d'abord conçues comme un outil permettant de formuler des **hypothèses sur les contrastes**

phonologiques existant dans telle langue ou variété non encore décrite. Dans un second temps, lorsque le transcripteur aura transcrit et étudié des données produites par plusieurs locuteurs et dans différents contextes, les transcriptions effectuées dans la tire *Phonologie* pourront être utilisées pour établir l'inventaire des contours phonologiques ou des primitives tonales utilisés dans la langue considérée et pour construire une grammaire intonative de cette langue.

Comme on le voit, IVTS ne fournit pas l'inventaire phonologique de la langue étudiée ou transcrite. Cette tâche doit en effet être faite pas le transcripteur-chercheur. IVTS propose simplement une liste de **primitives tonales**, à partir desquelles seront construits les éléments contrastifs (inventaires des "pitch accents", inventaires des contours, inventaires des tons de frontières, etc.).

Dans certains cas, cela pourra aboutir au développement d'une grammaire intonative de la langue considérée ; mais un tel résultat ne sera pas toujours obtenu. Pour illustrer ce que peut être une telle grammaire on peut citer la grammaire intonative du français standard développée par Post (2000), et qui génère la liste suivante de contours intonatifs associés aux syntagmes intonatifs.

(2)

$$\left\{\begin{matrix}\%L\\ \%H\end{matrix}\right\} \quad (H^* (L))_0 \quad \left\{\begin{matrix}H^*\\ H+H^*\end{matrix}\right\} \quad \left\{\begin{matrix}L\%\\ H\%\\ 0\%\end{matrix}\right\}$$

Cette grammaire intonative du français standard pourrait servir de point de comparaison pour les grammaires qui seront découvertes pour d'autres variétés.

4. Conclusion

Cet article présente les caractéristiques essentielles d'un système de transcription des phénomènes prosodiques et intonatifs, le système IVTS. Ce dernier propose de fournir une transcription sur plusieurs niveaux :

- le niveau des *proéminences* qui indique les syllabes perçues comme proéminentes par les transcripteurs, que ces proéminences se traduisent mélodiquement ou pas ;
- le niveau *phonétique local* dans lequel sont encodés les mouvements mélodiques réalisés sur des syllabes proéminentes. Cet encodage se fait sur la seule base perceptive ;
- le *niveau phonétique global* qui permet d'encoder les phénomènes mélodiques perçus non locaux (compression ou expansion de registre, etc.) ;
- le niveau *phonologique* qui permet au transcripteur de formuler des hypothèses sur les primitives et les catégories phonologiques utilisées dans la langue ou la variété qu'il étude.

Dans l'ensemble, nous avons essayé d'expliquer pas à pas comment s'effectue une transcription avec IVTS. Nous avons présenté pour chaque tire les différentes étiquettes qui peuvent être utilisées pour effectuer la transcription, en ajoutant autant que possible des exemples précis.

IVTS est, selon nous, un système de transcription suffisamment flexible pour être utilisé par des phonologues ou des spécialistes de la prosodie qui travaillent dans des paradigmes divers, et surtout pour tenter d'appréhender des variétés et des langues dont la description n'est pas encore faite. Reste donc maintenant à le tester et à l'améliorer, notamment à partir d'une étude des variétés de français !

Références

Auer, P. & Gilles, P. (soumis). "Prosodic Variation in German between Areality and Pragmatics: an Overview". Soumis pour publication à *Linguistics*.

Beckman, M. E., J. Hirschberg & S. Shattuck-Hufnagel. (2005). "The original ToBI system and the evolution of the ToBI framework". In S.-A. Jun (ed.) *Prosodic Typology* : *The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Boersma, P. & D. Weenink. (1996). "Praat, a system for doing phonetics by computer, version 3.4". *IFA rapport n° 132*.

Campione, E., D. Hirst & J. Véronis. (2000). "Automatic stylisation and symbolic coding of F0 : implementations of the INTSINT model". In A. Botinis (ed.), *Intonation. Research and Applications*. Dordrecht : Kluwer.

Carton, F. 1989. "La structuration temporelle dans les français régionaux du Nord-Est. Étude de phonétique expérimentale", *Mélanges de phonétique générale et expérimentale offerts à Péla Simon*, Strasbourg, Publications de l'Institut de Phonétique de Strasbourg, vol. 2, 215-230.

Delais-Roussarie, E. (1995). *Pour une approche parallèle de la structure prosodique: Etude de l,organisation prosodique et rythmique de la phrase française*, Thèse de Doctorat, Université de Toulouse-Le Mirail.

Delais-Roussarie, E. (2000). "Vers une nouvelle approche de la structure prosodique". *Langue Française* 126, 92-112.

Delais-Roussarie, E. (2005). *Phonologie et Grammaire : études et modélisation des interfaces prosodiques*. Mémoire d'HDR, Université de Toulouse-le Mirail.

Dell, F. (1984). "L'accentuation dans les phrases en français". In F. Dell, D. Hirst & J.R Vergnaud (eds). *Forme sonore du langage: structure des représentation en phonologie*, Paris : Hermann, 65-122.

Di Cristo, A. (1999), "Le cadre accentuel du français contemporain", 1e partie, *Langues* 2-3: 184-205 ; 2e partie, *Langues* 2-4, 258-267.

Di Cristo, A & J. Jankowski. (1999). "Prosodic organization and phrasing after focus in French". *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS)*, San Francisco, USA, 1565-1568.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German*. MPI Series in Psycholinguistics 7, Wageningen, Ponsen en Looien.

Grabe, E., B. Post, F. Nolan & K. Farrar. (2000). "Pitch accent realisation in four varieties of British English", *Journal of Phonetics* 28, 161-185.

Grabe, E., B. Post & F. Nolan. (2001). "Modelling intonational variation in English: The IViE system". In S. Puppel and G. Demenko (eds.), *Prosody 2000*, Poznan: Adam Mickiewicz University, 51-58.

Grabe, E. & B. Post. (2002). "Intonational Variation in English". In B. Bel & I. Marlin (eds), *Proceedings of the Speech Prosody 2002 Conference*, 11-13 Avril 2002, Aix-en-Provence: Laboratoire Parole et Langage, 343-346.

Grabe, E. & B. Post. (2004). "Intonational variation in the British Isles". In G. Sampson & D. McCarthy (eds). *Corpus Linguistics: Readings in a Widening Discipline*. Continuum International.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents.* Dordrecht: Foris.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press

Hambye, Ph. & A.C. Simon. (2006). “The production of social meaning via the association of variety and style: A case study of French vowel lengthening in Belgium”. *Canadian Journal of Linguistics* 49/3-4, 1001-1025.

Hirst, D.J., A. Di Cristo & R. Espesser (2000). “Levels of representation and levels of analysis for intonation”. In M. Horne (ed), *Prosody : Theory and Experiment*. Dordrecht : Kluwer.

Ladd, D.R. (1996). *Intonational Phonology*, Cambridge : Cambridge University Press.

Ladd, D.R., D. Faulkner, H. Faulkner & A. Schepman. (1999). “Constant ‘segmental anchoring’ of F0 movements under changes in speech rate”. *Journal of the Acoustical Society of America* 106, 1543-1554.

Pasdeloup, V. (1990). *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*, Thèse de Doctorat, Université d'Aix en Provence.

Pierrehumbert, J. (1980). *The phonology and phonetics of English Intonation*. Ph.D. Thesis, M.I.T.

Post, B. (2000). *Tonal and phrasal structures in French intonation*, The Hague : Holland Academic graphics.

Verluyten, S.P (1982). *Investigation on French Prosodics and Metrics*, Ph.D. Dissertation, Anvers.

La perception des proéminences et le codage prosodique

François POIRÉ[1]

1. Introduction

En mai 2004, à Caen, plusieurs participants et "amis" du projet PFC se sont réunis afin de discuter de l'ajout d'un codage prosodique à ceux déjà disponibles (schwa et liaison). Cette rencontre du groupe informellement appelé "Prosodie-PFC" faisait suite à plusieurs échanges de propositions de codage qui furent pour la plupart publiées dans le Bulletin PFC 3 (Eychenne et Mallet 2004) et dont certaines ont été l'objet de présentations particulières lors de cet atelier.

Au moins cinq grands principes presque unanimement acceptés ont façonné les échanges au cours de cette rencontre, à savoir qu'un codage prosodique :

1. doit, tout comme les codages schwa et liaison, être alphanumérique et être porté par la transcription orthographique ;
2. doit enrichir les codages schwa et liaison qui sont au cœur du projet PFC ;
3. doit être développé indépendamment d'un cadre théorique particulier ;
4. doit être facilement mis en application par les participants, qu'ils soient ou non spécialistes des études prosodiques ;
5. doit permettre la réalisation d'études prosodiques (intonation, accentuation, etc.).

Les discussions menées autour de ces questions n'ont malheureusement pas permis de dégager un consensus sur la meilleure manière d'intégrer à la base de données les informations de nature prosodique. Par contre, afin de faire progresser les discussions, il a été décidé de tenter une expérience de perception des proéminences. Étant à l'origine de cette proposition, j'ai donc été chargé d'élaborer un protocole de codage, de sélectionner des extraits et de compiler les résultats soumis par les participants.

Cet article rend compte de cette expérience. Dans un premier temps, je présente brièvement les raisons pour lesquelles l'identification des proéminences sur une base auditive pourrait contribuer à l'élaboration d'un système de codage de la prosodie dans le cadre d'un projet tel le PFC. Deuxièmement, la méthodologie retenue pour l'identification des proéminences et pour l'analyse des résultats est décrite. Finalement, les résultats eux-mêmes suivis d'une brève discussion viennent clore cet article.

2. La notion de proéminence dans l'étude de la prosodie

Un des principaux sujets sur lesquels les discussions ont buté lors de cette rencontre concerne le rôle de la syllabe dans le codage de la prosodie, plus précisément la quantité d'informations devant être incluse dans sa description ainsi que la manière dont ces informations doivent être obtenues. Par exemple, dans le domaine fréquentiel, doit-on seulement noter la présence d'un mouvement mélodique ou encore son ampleur ? Et comment doit-on juger d'un tel mouvement : sur une base auditive ou instrumentale ? Autre point litigieux : doit-on coder toutes les syllabes ou encore seulement celles jouant un rôle défini dans la mise en forme prosodique de l'énoncé, telles les toniques ? Dans ce dernier cas, que faire d'un accent d"emphase réalisé sur un monosyllabe clitique ? Mon propre système de codage de la prosodie (Poiré 2004) tentait de répondre à ces questions à partir

[1] The University of Western Ontario.

des deux propositions suivantes : (i) il existe un consensus assez général quant à la distribution des syllabes normalement accentuées en français (et donc proéminentes), à savoir la dernière syllabe d'"un mot plein ; en conséquence, toutes les syllabes toniques doivent être codées ; (ii) une syllabe non tonique mais proéminente (emphase, accent secondaire, allongement associé à l'hésitation) devrait être facilement remarquée et donc codée.

Le premier critère établit un parallèle méthodologique avec le codage du schwa et de la liaison. Les sites potentiellement porteurs d'une proéminence sont identifiables à partir de la transcription orthographique des fichiers sonores[2]. Le second critère repose sur l'idée qu'une proéminence prosodique moins prévisible que l'accent tonique sera facilement perçue. Dans les deux cas, l'identification des syllabes saillantes repose sur la définition générale de la proéminence :

> **"prominent** (adj.) A term used in AUDITORY PHONETICS to refer to the degree to which a sound or SYLLABLE stands out from others in its ENVIRONMENT. Variation in LENGHT, PITCH, STRESS and inherent SONORITY are all factors which contribute to the relative **prominence** of a UNIT. (…)". (Crystal 2003 : 375)

Le tableau suivant résume la portée de cette proposition. La zone ombragée correspond à ce qui serait effectivement codé dans le cadre de cette proposition.

Tableau 1. *Syllabes codées selon Poiré (2004)*

	+ proéminente	- proéminente
Dernières syllabes de mot plein	Accent tonique	Syllabe inaccentué ou désaccentuée
Autres syllabes	Emphase, accent secondaire, hésitation	Syllabe inaccentuée

Les syllabes visées par le codage prosodique étant identifiées, il ne resterait qu'à déterminer quelles informations devraient être retenues relativement à leur position dans le mot ou dans un autre domaine de référence, ainsi que celles ayant trait à la substance sonore exprimant la proéminence[3].

L'exercice peut être vu, je crois, comme une recherche de réponses aux trois questions suivantes, fort dépendantes les unes des autres :

1. Partageons-nous la même conception de la notion de proéminence ?
2. Est-il possible d'identifier les proéminences sur une base auditive seulement ?

[2] Dans la mesure bien sûr où les codeurs s'entendent sur la distinction entre mots pleins et mots de fonction (ou mots ± clitiques). Mertens (1993) et Mertens et al (2001) fournissent un cadre de référence fort utile à ce sujet. Mertens (2004) fournit même des indications pour l'accentuation de constructions syntaxiques particulières.

[3] Implicite dans ma proposition originale se trouve l'idée que l'accent français se manifeste principalement par la durée, qui est donc considérée comme le corrélat acoustique par défaut de la proéminence. Le codage proposé fait seulement référence de manière explicite à certaines variations de la fréquence fondamentale, dans l'esprit de Ladd (1996) : "*Distinction between pitch accent and stress*: pitch accents, in languages that have them, serve as concrete perceptual cues to stress or prominence. However, they are in first instance *intonational* features which are *associated* with certain syllables in accordance with various principles of prosodic organisation. The perceived prominence of accented syllables is, at least in some languages, a matter of *stress*, which can be distinguished from pitch accent". (Ladd 1996 : 42-43)

3. Différents codeurs peuvent-ils arriver de manière indépendante aux mêmes résultats quant à l'identification sur une base auditive des syllabes proéminentes ?

3. Méthodologie

Trois extraits d'environ trois minutes de locuteurs différents ont été sélectionnés et soumis aux participants à l'expérience. Tous n'ont pas codé les trois locuteurs, d'aucuns n'ont codé qu'une partie des extraits proposés. Je présente ici la synthèse des jugements posés à partir d'un extrait d'environ trois minutes d'un locuteur belge (conversation libre).

Huit personnes ont soumis un codage, dont deux qui ont préféré rendre une version commune. Nous traiterons donc les propositions de sept "codeurs" (C1…C7).

Le but de l'expérience n'étant pas d'expérimenter la proposition de Poiré (2004) mais bien d'explorer la possibilité d'utiliser la perception de la proéminence relative des syllabes à des fins de codage prosodique, une seule consigne a été fournie aux participants : à partir de la transcription orthographique, indiquer par un "1" toute syllabe perçue comme étant proéminente ("2" pour les cas incertains).

L'extrait en question prend la forme de 370 "données". Qu'est-ce qu'une donnée ? J'ai déterminé ces données de deux manières différentes :

1. Chaque fois qu'au moins un codeur codait une syllabe "1" ou "2", le mot contenant cette syllabe devenait une donnée. Si deux syllabes différentes d'un même mot étaient l'objet d'un jugement, le mot en question était répété deux fois dans la base de données.
2. Seules les syllabes perçues comme proéminentes (ou incertaines) ont été identifiées par les codeurs (ce qui correspond à la consigne). On peut donc déduire que les autres ont été perçues comme étant non proéminentes et auraient pu être codées "0". J'ai donc créé une donnée à partir de chaque mot "plein" dont la dernière syllabe n'a été codée par personne.

La raison pour laquelle j'ai tenu compte des mots pleins non codés est la suivante. Pour la plupart des chercheurs, la principale fonction de la proéminence est l'expression de l'accentuation[4]. Il me semblait alors raisonnable de penser que les syllabes finales de mots pleins recevraient une attention particulière. Dans cet esprit, j'ai aussi voulu distinguer les accents toniques des autres possibilités d'accentuation en assignant à chaque syllabe codée un des trois types suivants :

1. CLI pour "+clitique" : il s"agit des cas de mots clairement identifiables à cette catégorie (*et, la, de*…) qui ont été codés "1" ou "2" par au moins un codeur.
2. SEC pour "secondaire" : il s'agit des cas de syllabes non finales de mots de plus d'une syllabe qui ont été codés "1" ou "2" par au moins un codeur.
3. PRIM pour "primaire" : la dernière syllabe d"un mot "plein", codée ou non par au moins un codeur.

Aussi, deux cas ont été notés "AUT" pour "autre". Il est intéressant de noter par ailleurs qu'aucun accent d'emphase n'a été produit dans cet extrait.

Une dernière remarque concernant les données. Deux codeurs ont utilisé la valeur "3" signifiant "syllabe proéminente par allongement marqué". Afin de simplifier le traitement des données, j'ai remplacé "3" par "1". Aussi, un codeur a indiqué comme étant proéminents des "euh" d'hésitation, mais étant donné les consignes données, tous les autres

[4] Par exemple : "Proéminence : mise en valeur perceptive d"une syllabe, qui se manifeste par la perception d"un accent" (Lacheret-Dujour et Beaugendre 1999 : 283).

codeurs ont laissé de côté ces syllabes d'hésitation. J'ai donc retiré de la matrice ces “euh” codés.

Le début de la base de données est repris dans le tableau qui suit. Les mots pour lesquels au moins deux syllabes ont été jugées proéminentes sont marqués de l'astérisque.

Tableau 2. *Extrait de la base de données. La syllabe évaluée est toujours la dernière de l'extrait sauf dans le cas des accents secondaires*

no	type	TEXTE	C1	C2	C3	C4	C5	C6	C7
1	prim	Non	1	1	1	0	0	0	0
2	prim	enfin	0	0	0	0	1	1	1
3	prim	ça dépend	1	1	0	0	1	0	1
4	cli	de la	0	0	1	0	0	0	0
5	prim	De la masse	2	1	1	0	1	0	1
6	prim	du, du chargement	1	1	1	1	1	0	1
7	prim	ça dépend	1	1	1	0	1	0	1
8	prim	Si tu roules	1	1	1	1	1	2	1
9	prim	sur la voie	0	0	0	0	0	0	0
10	prim	publique	1	1	1	1	1	1	1
11	prim	euh . Ici dans le champ	1	1	1	1	1	1	1
12	prim	tu peux	0	0	0	0	0	0	0
13	prim	en faire	0	0	0	0	0	0	0
14	prim	ce que tu veux	1	1	1	1	1	0	1
15	prim	mais sur la route	1	1	1	1	1	2	1
16	prim	Normalement	1	1	0	0	1	0	1
17	prim	il y a.	1	0	1	0	0	0	0
18	prim	Ben c'est	0	0	0	0	1	0	0
19	prim	assez	0	0	0	0	0	0	0
20	prim	complexe	1	1	1	0	1	1	1
21	prim	ça dépend	2	1	0	0	1	0	1
22	cli	de la	1	0	1	0	0	0	0
23	aut	ben	2	0	0	0	0	0	0
24	prim	c'est pas	1	1	0	0	1	1	0
25	sec	*de la REmorque	0	0	0	1	0	0	0
26	prim	*de la remorque	1	1	1	1	1	2	1
27	sec	*c'est du CHARgement	0	1	0	0	1	0	0
28	prim	*c'est du chargement	0	1	0	0	0	0	0
29	prim	total	1	1	1	0	1	0	1
30	prim	tracteur	2	1	1	0	1	0	1

4. Résultats

4.1. Nombre de proéminences perçues

J'ai déjà mentionné que tous n'ont pas codé les extraits au complet. J'ai donc décidé de traiter les données en deux temps : un premier corpus couvrant la partie de l'extrait codée par les sept codeurs ; un second corpus couvrant une partie plus longue mais codée par seulement cinq participants. Le corpus 1 à sept codeurs comprend 170 données tandis que

celui à cinq codeurs (corpus 2) en compte 283[5]. Le tableau 3 montre la distribution des jugements selon les quatre types définis plus haut.

Tableau 3. *Distribution des proéminences perçues (avec ou sans certitude) par les sept codeurs, en fonction du type, corpus 1*

		Types de syllabes codées					
Codeurs	Codes	primaire	secondaire	clitique	autre	Total	Grand total
C1	1	59	0	11	0	70	97
	2	26	0	0	1	27	
C2	1	79	2	2	0	83	91
	2	8	0	0	0	8	
C3	1	76	4	1	0	81	81
	2	0	0	0	0	0	
C4	1	57	0	0	0	57	61
	2	4	0	0	0	4	
C5	1	57	1	2	0	60	60
	2	0	0	0	0	0	
C6	1	36	7	2	0	45	45
	2	0	0	0	0	0	
C7	1	31	0	1	0	32	39
	2	6	1	0	0	7	

En regardant la colonne “grand total”, on constate que le nombre de syllabes codées varie énormément dans une fourchette allant de 39 à 97 (pour environ deux minutes de parole). Une partie de cette variation est peut-être attribuable à la manière d'utiliser le code “2” (incertitude). Le tableau 3 montre clairement que plus de la moitié des cas incertains viennent du même codeur (C1), trois codeurs n'ayant jamais recours au “2”. Il est fort probable qu'une véritable procédure de codage prosodique viendrait restreindre l’utilisation de la valeur “2”. Par exemple, si l'on additionne les “1” et les “2” du codeur C1, on obtient à peu près le nombre de “1” des codeurs C2 et C3. Afin de mieux saisir l'ampleur de la variation entre les codeurs, c'est-à-dire en laissant de côté la question de l'interprétation du code “2”, j”ai rapporté dans le tableau suivant la proportion des codages certains (“1”) en fonction du nombre de données du corpus 1.

Tableau 4. *La proportion de syllabes codées “1” par rapport au nombre total de données du corpus 1 (n = 170)*

Codeurs	C2	C3	C1	C5	C4	C6	C7
%	48,82%	47,65%	41.18%	35.29%	33.53%	26.47%	18.82%

La fourchette demeure très large, de 18.82% à 48,82%.

Le tableau 3 montre aussi clairement que la proéminence de type “secondaire” n'est pas très populaire. Un seul codeur en note plus de cinq. La même remarque vaut pour le type “clitique”. Le codeur C1 note plus de la moitié des cas à lui seul.

Ces commentaires demeurent pertinents si l'on se tourne vers le corpus 2, avec seulement cinq codeurs (trois minutes de parole environ).

[5] Quatre codeurs ont codé l'extrait au complet pour un grand total de 370 données.

Tableau 5. *Distribution des proéminences perçues (avec ou sans certitude) par les cinq codeurs du corpus 2, en fonction du type.*

		Types de syllabes codées					
Codeurs	Codes	primaire	secondaire	clitique	autre	Total	Grand total
C1	1	92	3	12	0	107	152
	2	44	0	0	1	45	
C2	1	126	2	2	0	130	149
	2	18	0	1	0	19	
C3	1	117	5	1	0	123	123
	2	0	0	0	0	0	
C5	1	95	1	3	0	99	99
	2	0	0	0	0	0	
C6	1	58	9	2	0	69	69
	2	0	0	0	0	0	

La grande différence dans le nombre de proéminences perçues ne diminue pas vraiment même en ne tenant compte que des syllabes codées “1” (tableau suivant).

Tableau 6. *La proportion de syllabes codées “1” par rapport au nombre total de données du corpus 2 (n = 283).*

Codeurs	C2	C3	C1	C5	C6
%	44.98%	42.56%	37.02%	34.26%	23.88%

Une première constatation s'impose. Il semble que codeurs ne partagent pas la même définition de ce qu'est la proéminence.

4.2. Accord entre codeurs

4.2.1. Méthodologie

La section précédente semble indiquer que les codeurs ne s'entendent pas sur la définition même de la proéminence. Mais ils ont quand même codé ! Dans quelle mesure s'entendent-ils alors dans les cas codés? Cette seconde partie des résultats traite du degré d'accord entre les sept, puis les cinq codeurs des deux corpus. Je traite les données comme s'il s'agissait d'un accord inter-juges, même si nous savons tous que ce n'est pas le cas.

Afin de simplifier la manipulation des données, j'ai dû prendre une décision au sujet des syllabes codées “2” (proéminence incertaine). J'ai remplacé les “2” par des “0”. J'ai donc choisi le pire cas, c'est-à-dire diminuer le nombre de jugements positifs face à une syllabe donnée.

Dans le but d'évaluer le degré d'accord au sein de tout un corpus, j'ai réinterprété le total de proéminences perçues (ou non perçues) pour une donnée de la manière suivante : si tous les codeurs notent une syllabe “1” (proéminente), le degré d'accord est de 100% (total = 7 pour le corpus 1). Si aucun locuteur ne juge une syllabe donnée proéminente, le degré d'accord est aussi de 100% (total = 0, corpus 1). J'ai donc codé ce degré d'accord selon les critères du tableau suivant.

Tableau 7. *Échelle de degré d'accord entre les codeurs (7 codeurs)*

Total sur 7	Degré d'accord	Codage du degré d'accord
7 sur 7 ou 0 sur 7	Unanime	7
6 sur 7 ou 1 sur 7	Unanime sauf 1	6
5 sur 7 ou 2 sur7	Unanime sauf 2	5
4 sur 7	50% +1	4
3 sur 7	50% -1	3

La même logique a été appliquée au corpus 2 (cinq codeurs) : 5 sur 5 et 0 sur 5 sont codés 5, 4 sur 5 et 1 sur 5 sont codés 4 etc.

Le tableau suivant illustre cette transformation à l'aide d'un extrait de la base de données.

Tableau 8. *Extrait de la base de données*

Texte	C1	C2	C3	C4	C5	C6	C7	total	accord
Non	1	1	1	0	0	0	0	3	3
enfin	0	0	0	0	1	1	1	3	3
ça dépend	1	1	0	0	1	0	1	4	4
de la.	0	0	1	0	0	0	0	1	6
De la masse	0	1	1	0	1	0	1	4	4
du, du chargement	1	1	1	1	1	0	1	6	6
ça dépend.	1	1	1	0	1	0	1	5	5
Si tu roules	1	1	1	1	1	0	1	6	6
sur la voie	0	0	0	0	0	0	0	0	7
publique	1	1	1	1	1	1	1	7	7

4.2.2. Accord absolu

Le tableau 9 indique simplement dans combien de cas *n* codeurs sur 7 ont perçu une proéminence sur une syllabe donnée.

Tableau 9. *Distribution des données selon le nombre de codeurs ayant noté "1" une syllabe donnée, corpus 1.*

Nb de codeurs jugeant la syllabe proéminente	Nb de données	En % de 170	Nb de données codées "2" par au moins un codeur
7 sur 7	12	7.06%	0
6 sur 7	16	9.41%	4
5 sur 7	12	7.06%	4
4 sur 7	18	10.59%	8
3 sur 7	14	8.24%	5
2 sur 7	18	10.59%	6
1 sur 7	38	22.35%	5
0 sur 7	42	24.71%	10
Total	170	100%	42

Comment interpréter ce tableau? Notons simplement qu'il reflète bien le fait que le nombre de syllabes codées positivement (“1”) varie énormément d'un codeur à un autre. Ainsi, il est difficile d'espérer plusieurs “7 sur 7” ou “6 sur 7” et il est assez prévisible de retrouver un grand nombre de “1 sur 7” ou “2 sur 7”. Par contre, près de 25% des données sont unanimement reconnues comme étant non proéminentes (0 sur 7). Notons aussi qu'environ le quart des données comportent au moins un jugement incertain (42 sur 170). Regardons maintenant le “degré d'unanimité” entre codeurs.

Tableau 10. *Degré d'unanimité des codeurs, corpus 1*

Nb de codeurs ayant le même jugement sur une syllabe	Nb de données	En % de 170	Nb de données codées “2”par au moins un codeur
7 sur 7	54	31.76%	10
6 sur 7	54	31.76%	9
5 sur 7	30	17.65%	10
4 sur 7	18	10.59%	8
3 sur 7	14	8.24%	5
Total	170	100%	42

S'il s'agissait d'un véritable accord inter-juges, on obtiendrait un degré d'unanimité de 31.76% avec un seuil d'acceptabilité de sept codeurs sur sept. Si le seuil incluait les proéminences reconnues par six codeurs et plus, le degré d'unanimité passerait à 63.52%. À cinq codeurs et plus sur sept, on obtiendrait un degré d'unanimité de 81.17%. Ce dernier pourcentage correspond à peu près à la moyenne de la colonne “accord” du tableau 8 (5.68 sur 7, 81.17%). Notons que ces scores seraient probablement supérieurs selon la manière de traiter les proéminences incertaines (dernière colonne du tableau). Rappelons qu'ici, tous les “2” ont été remplacés par des “0”. Les mêmes constats peuvent être établis à partir du corpus 2 (cinq codeurs).

Tableau 11. *Distribution des données selon le nombre de codeurs ayant noté “1” une syllabe donnée, corpus 1*

Nb de codeurs jugeant la syllabe proéminente	Nb de données	En % de 283	Nb de données codées “2”par au moins un codeur
5 sur 5	36	12.46%	0
4 sur 5	30	10.38%	5
3 sur 5	36	12.46%	17
2 sur 5	30	10.38%	12
1 sur 5	59	20.41%	9
0 sur 5	92	31.83%	18
Total	283	100%	61

Il est normal, étant donné le plus petit nombre de codeurs, que les résultats soient plus “serrés”. Par exemple, à sept codeurs, le taux d'unanimité positive (7 sur 7, tableau 7) était de 7.06% tandis qu'à cinq codeurs, l'unanimité à 5 sur 5 passe à 12.46% (tableau 9). Cela se reflète aussi dans le tableau suivant.

Tableau 12. *Degré d'unanimité des codeurs, corpus 2*

Nb de codeurs ayant le même jugement sur une syllabe	Nb de données	En % de 283	Nb de données codées "2" par au moins un codeur
5 sur 5	128	44.29%	18
4 sur 5	89	30.80%	14
3 sur 5	36	12.46%	17
2 sur 5	30	10.38%	12
Total	283	100%	61

Dans le cas des cinq codeurs, la moyenne de la colonne "accord" s"élève à 4.11 sur cinq, soit 82.20%. Un seuil fixé à quatre codeurs et plus sur cinq donnerait 75.09% de degré d'unanimité.

4.2.3. L'accord par type

Les quatre tableaux qui suivent reprennent les données de la section précédente distribuées selon le type. Le tableau 13 montre clairement que si les proéminences étaient notées selon le principe que le jugement d'une majorité de codeurs prévaut, tous les SECondaires et les CLItiques (sauf peut-être un cas) seraient rejetés, même si le seuil à partir duquel une proéminence est acceptée était fixé à 4 sur 7.

Tableau 13. *Distribution des données par type selon le nombre de codeurs ayant noté "1" une syllabe donnée, corpus 1*

Nb de codeurs jugeant la syllabe proéminente	primaire	secondaire	clitique	autre
7 sur 7	12			
6 sur 7	16			
5 sur 7	12			
4 sur 7	17		1	
3 sur 7	14			
2 sur 7	12	3	3	
1 sur 7	21	8	9	
0 sur 7	40	1		1

En bref, la majorité de ces cas seraient unanimement rejetés même avec un degré d'unanimité relativement bas (5 sur 7 et plus par exemple).

Tableau 14. *Degré d'unanimité des codeurs, corpus 1*

Nb de codeurs ayant le même jugement sur une syllabe	primaire	secondaire	clitique	autre
7 sur 7	51	1	1	1
6 sur 7	37	8	9	
5 sur 7	24	3	3	
4 sur 7	17		1	
3 sur 7	14			

Ce constat vaut pour le corpus à cinq codeurs.

Tableau 15. *Distribution des données par type selon le nombre de codeurs ayant noté "1" une syllabe donnée, corpus 2*

Nb de codeurs jugeant la syllabe proéminente	primaire	secondaire	clitique	autre
5 sur 5	36	0	0	0
4 sur 5	30	0	0	0
3 sur 5	35	0	1	0
2 sur 5	24	3	3	0
1 sur 5	34	14	11	0
0 sur 5	89	1	1	1

Tableau 16. *Degré d'unanimité des codeurs, corpus 2*

Nb de codeurs ayant le même jugement sur une syllabe	primaire	secondaire	clitique	autre
5 sur 5	125	1	1	1
4 sur 5	64	14	11	0
3 sur 5	35	0	1	0
2 sur 5	24	3	3	0

5. Conclusion

Dans la section 1, je donnais une liste de trois questions auxquelles cet exercice apporterait peut-être des réponses, même partielles. Je les reprends ici :

1. Partageons-nous la même conception de la notion de proéminence?
2. Est-il possible d'identifier les proéminences sur une base auditive seulement?
3. Différents codeurs peuvent-ils arriver de manière indépendante aux mêmes résultats quant à l'identification sur une base auditive des syllabes proéminentes?

Il semble bien, qu'à la lumière de la compilation de ces propositions de codage, qu'un "non" constitue une réponse honnête aux trois questions. Mais j'aimerais quand même terminer par quelques commentaires qui redonneront un peu d'espoir à ceux qui, comme moi, croient que le recours à l'identification des proéminences sur une base auditive constitue un outil pratique pour le codage de la prosodie.

Premièrement, l'absence de consignes autres que "codez ce que vous percevez proéminent" rend l'exercice périlleux. Il est difficile de penser que l'analyse prosodique sans balises soit une approche fréquente. Le découpage d'un extrait en "domaines" à l'aide des pauses par exemple permet de mieux poser les questions suivantes : quelles sont les syllabes proéminentes de ce domaine ? Quel est le contour de ce domaine ? Un guide à l'usage des codeurs pourrait donc contenir un ensemble de propositions ou de consignes qui aideraient les chercheurs impliqués à mieux décrire la parole à l'étude. Un tel guide proposant un mode d'emploi incluant la prise en compte du découpage aurait peut-être permis d'éviter que dans le cadre de cette expérience certaines propositions de codage contiennent des énoncés complets sans aucune proéminence notée. En voici un exemple (en tout début d'extrait) :

(1) c'est du chargement total tracteur plus euh la remorque et tout ça enfin c'est assez compliqué quoi.

Sans même tenir compte des pauses ni de la déclinaison de F0, il semble assez logique de croire que cet extrait est “découpable” en au moins trois sous-parties et qu’au moins une syllabe par sous-partie sera accentuée et proéminente.

(2) c'est du chargement total tracteur plus euh
la remorque et tout ça
enfin c'est assez compliqué quoi.

Deuxièmement, il est certain que le recours raisonné à l'analyse instrumentale viendrait préciser la distribution des syllabes proéminentes (et des autres aspects de la prosodie). Je pense ici en particulier à l'utilisation systématique de l'affichage de F0. On peut aussi penser au recours ponctuel à la mesure de la durée dans une séquence donnée. Il faut, encore une fois, guider le codage et fournir des procédures, paraissant relativement simples pour le spécialiste mais utiles pour tout codeur.

Finalement, le fait qu'il s'agit de langue spontanée mérite que certains phénomènes reçoivent une attention particulière. Un exemple : ce que j'ai appelé “clitique” dans ma présentation des propositions de codages est souvent proéminent, en particulier devant une pause (“ça dépend de **la**”). Rappelons-nous que C1 a codé “1” la très grande majorité des cas identifiés et que les autres codeurs sont demeurés muets au sujet de ce phénomène relativement fréquent en langue spontanée.

Références

Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics*, Oxford : Blackwell Publishing.

Eychenne, J. et G. Mallet (eds.). (2004). *Du segmental au prosodique : protocoles, outils, extensions et travaux en cours. Bulletin PFC* 3. ERSS UMR 5610 CNRS & Université de Toulouse-Le Mirail.

Lacheret-Dujour, A. et F. Beaugendre (1999). *La prosodie du français*. Paris : CNRS Editions.

Ladd, R. (1996). *Intonational phonology*. Cambridge : University Press.

Mertens, P. (1993). *Accentuation, intonation et morphosyntaxe Travaux de Linguistique* 26, 21-69.

Mertens, P. (2004). Quelques allers-retours entre la prosodie et son traitement automatique *Le français moderne* 72(1), 39-57.

Mertens, P., A. Auchlin, J.-P. Goldman, A. Grobet, et A. Gaudinat (2001). “Intonation du discours et synthèse de la parole : premiers résultats d'une approche par balises”, *Cahiers de Linguistique française* 23, 189-209.

Poiré, F. (2004). “Le codage de la prosodie dans le cadre du PFC”, dans Eychenne, J. et G.-M. Mallet (dir.) “Du segmental au prosodique. Protocoles, outils, extensions et travaux en cours”, *Bulletin PFC* 3, CNRS et Université de Toulouse-Le Mirail.

La transcription des proéminences accentuelles : Mission impossible ?

Philippe MARTIN[1]

1. Introduction

L'étude de F. Poiré (dans ce volume), portant sur la transcription de la proéminence accentuelle d'un même corpus du projet PFC par un panel d'experts phonéticiens, montre un taux élevé de discordances dans les jugements des auditeurs. Dès lors, une question se pose : comment des experts, rompus aux techniques de transcription et spécialistes de prosodie, libres d'écouter et de réécouter à volonté les segments de parole dont ils devaient détecter les proéminences syllabiques, ont pu émettre des jugements aussi divergents ? La proéminence syllabique serait-elle un concept évasif ?

Notons pour commencer qu'une syllabe ne peut être proéminente à elle toute seule, la présence d'une ou de plusieurs syllabes voisines est nécessaire pour établir un contraste entre proéminence et non proéminence. Ainsi la syllabe unique d'un énoncé réduit à un seul mot comme *oui* peut-elle être proéminente ? Seul un raisonnement de nature phonologique pourrait répondre à cette question.

On sait depuis Saussure qu'un objet linguistique n'existe que s'il est doté d'une signification. La segmentation et la catégorisation des significations reposent sur un savoir *a priori* inconscient du sujet parlant. Ainsi, l'utilisateur de l'objet proéminence syllabique opère par un processus inconscient parmi les autres processus inconscients de perception de la parole. On peut donc dire que l'auditeur lambda ne se préoccupe pas consciemment de la nature ou de la fonction éventuelle des proéminences perçues. Le sémioticien et linguiste Luis Prieto (1975) dirait que l'auditeur utilise un système de catégorisation primaire élaboré lors de l'acquisition et par la pratique du langage.

Il en va tout autrement lorsque cet auditeur lambda devient conscient de l'opération, par exemple lorsqu'on lui demande un jugement sur la proéminence syllabique, en lui donnant des consignes telles que "pensez vous que cette syllabe est proéminente, accentuée, insistante ?". Le contenu de la consigne mettra en jeu une autre connaissance métalinguistique, impliquant des systèmes de catégorisations secondaires de l'auditeur.

Dès lors toute la procédure d'évaluation et de transcription est transformée, l'objet observé se trouve modifié (si tant est qu'il n'existe qu'au travers d'une pratique, et dans ce cas d'une procédure d'observation), tout comme dans le domaine de la physique des particules l'observation modifie le comportement de l'objet observé. Ceci mène à l'examen détaillé des conséquences.

2. Proéminences syllabiques et fonctions

Revenons à la signification ou aux significations de la proéminence syllabique. En français, comme dans beaucoup d'autres langues, la proéminence syllabique peut assumer trois fonctions (au moins) :

- une fonction démarcative (mais sur quel type d'unités ?) ;

[1] UFR Linguistique, Université Paris 7 Denis Diderot.

- une fonction indicatrice de la focalisation étroite (appelée traditionnellement accent d'insistance, et habituellement placée sur la première syllabe des mots pleins) ;
- une fonction indicatrice de la focalisation large (correspondant au découpage de l'énoncé en propos et thème).

La connaissance *a priori* de la fonction permet de classer les différentes manifestations acoustiques et perceptives de la proéminence. Ainsi la focalisation large identifiée comme telle permet de décrire la réalisation de ce type de proéminence par une chute mélodique à l'endroit de la syllabe normalement accentuée (par une proéminence syllabique relevant de la fonction démarcative). Cette chute mélodique est suivie d'un aplatissement des variations mélodiques sur les syllabes accentuées qui suivent. Notons en passant qu'une manipulation simple de *morphing* prosodique montre clairement que ces deux événements acoustiques doivent être concomitants pour assurer le codage et la perception de la focalisation large dans la phrase. La confusion avec une autre fonction qu'assumerait ce type de proéminence n'est donc pas possible, puisque la syllabe proéminente ne peut être, de par la nature même de la division de l'énoncé en propos et thème, en position finale.

La focalisation étroite porte souvent (mais pas uniquement) sur la première syllabe des mots pleins (noms, adjectifs, verbes, adverbes) focalisés. Si le mot plein en question a plus d'une syllabe, la position initiale de la syllabe proéminente permettra dans ce cas également à l'auditeur lambda d'attribuer une fonction adéquate à cette proéminence. Si le mot plein est monosyllabique (et se trouve en position finale du groupe accentuel, cf. ci-dessous), l'ambiguïté est possible entre les fonctions de focalisation étroite et la fonction démarcative, puisque dans un tel cas la proéminence d'insistance ne peut frapper une autre syllabe que la première et unique syllabe du mot plein.

3. Proéminence et accent démarcatif

Dans une perspective fonctionnaliste, il est naturel de forcer les marques phonologiques que l'on veut décrire à fonctionner, et c'est précisément ce principe qui est mis en œuvre par le test des paires minimales. En neutralisant tout contexte, [pjɛʁ] s'oppose à [bjɛʁ], alors qu'en réalité, même un phonème /b/ réalisé comme [p] ne perturbera sans doute pas outre mesure la compréhension de *je veux une bonne [pjɛʁ]* , du moins si le contexte implique le café du coin et non une carrière de marbre…

Mais comment faire avec la proéminence réalisant la fonction démarcative en français ? Des langues comme l'anglais ou l'italien présentent des "accidents" morphologiques permettant par exemple d'opposer *capitano* à *capitano* dans la phrase fameuse en italien *sono cose che **ca**pitano capi**ta**no* ("ce sont des choses qui arrivent capitaine"). La position presque toujours finale de l'accent lexical en français empêche évidemment l'apparition de telles curiosités morphologiques.

Si la perception et la transcription de l'accent dit d'insistance (accent de focalisation étroite), et de l'accent indiquant une organisation propos thème de l'énoncé (focalisation large) ne pose pas (trop) de problèmes pour un phonéticien entraîné, il n'en va pas de même pour l'accent dit lexical, qui se manifeste en première approximation en français sur la dernière syllabe prononcée. En effet, à la différence des autres langues romanes, cet accent "lexical" présente une caractéristique particulière, celle de pouvoir être éventuellement non réalisé. On parle alors, à propos des syllabes frappées par cet accent susceptible d'être réalisé, d'accentuabilité (Garde 1968).

Or la réalisation effective de l'accent lexical ou accent de groupe rythmique constitue un point central pour les modèles intonatifs dans la mesure où ces modèles

tendent à rendre compte des événements prosodiques remarquables, précisément à l'endroit des syllabes accentuées. Les modèles autosegmentaux ou phonosytaxiques utilisent des systèmes de transcription fort différents (ToBI par niveaux cibles, d'autres par contours) mais s'attachent tous deux à la description des réalisations mélodiques des syllabes accentuées. La confrontation avec les données présuppose évidemment que le caractère effectivement accentué des syllabes soit clairement établi, ce qui pose donc des problèmes particuliers pour des langues comme le français.

En première approximation, l'accent lexical frappe la dernière syllabe d'un mot de classe ouverte :

(1) Ju**liet**te Ju**liet**te s'en**nuie**

En fait l'accent lexical se place sur la dernière syllabe prononcée d'un groupe (groupe accentuel ou groupe rythmique) constitué d'un mot de classe ouverte (mots pleins : noms, verbes, adjectifs qualificatifs, adverbes) et de tous les mots de classe fermée (mots outils : conjonctions, prépositions, déterminants) qui sont dans une relation de sélection ou de solidarité (au sens distributionnel) avec lui. À cette catégorie s'ajoute celle des mots de classe fermée qui n'entretiennent pas de relation de dépendance avec une autre unité dans la phrase (par exemple le pronom *toi*, dans *toi tu penses que*…). Ce groupe est le groupe accentuel (Martin 1987), dont la définition évite la circularité encore rencontrée fréquemment dans beaucoup d'ouvrages.

Ainsi *tu le lui **don**nes* forme un groupe accentuel. Le mot de classe ouverte *donnes* est en relation de solidarité avec le pronom *je* et est sélectionné par les pronoms objets direct *le* et indirect *lui*. La dernière syllabe prononcée de ce groupe est donc accentuable, et également dans ce cas accentuée, puisque seule dans la phrase (toute phrase comporte au moins un accent lexical, indicateur de la modalité déclarative ou interrogative).

Dans le même exemple avec une modalité impérative, *donne-le **lui***, la syllabe accentuable placée sur *donne* se déplace sur la dernière syllabe prononcée du groupe, c'est-à-dire sur le pronom *lui*. En modalité interrogative, on a *le lui donnes-**tu***, avec cette fois l'accent de groupe placé sur *tu*, dernière syllabe prononcée du groupe.

Examinons à présent ce qui se passe lorsque plusieurs groupes accentuels apparaissent dans la phrase. Dans *l'amie de Juliette a vaincu son angoisse*, on distingue 4 groupes accentuels, tels que définis par des relations de dépendance entre mots pleins (classe ouverte) et mots outils (classe fermée) : | *l'amie* | *de Juliette* | *a vaincu* | *son angoisse* |.

Le problème ici vient du fait que, si la réalisation d'un accent de groupe accentuel est obligatoire sur la dernière syllabe prononcée de la phrase, donc sur la dernière syllabe de *angoisse*, il n'en va pas nécessairement de même pour les syllabes finales des autres groupes. En effet, on peut très bien, avec un débit très rapide, prononcer cette phrase avec un seul accent de groupe final, et ne pas réaliser les syllabes accentuables terminant les 3 premiers groupes**.** Du reste on peut montrer (Wioland 1985, Dell 1984) que les différentes possibilités combinatoires peuvent fréquemment correspondre à la structure syntaxique, la réalisation de 2 groupes accentuels | *l'amie de Ju**liet**te* | *a vaincu son an**gois**se* | étant privilégiée pour un débit de parole moyen.

Il est aisé d'établir que la réalisation effective des accents sur les syllabes accentuables dépend du nombre de syllabes de chaque groupe, donc en définitive de la durée d'énonciation des groupes. Si dans notre exemple on substitue à *l'amie de Juliette* la séquence *les hippopo**tam**es de Nabuchodono**sor***, il devient très difficile pour un locuteur de

ne pas accentuer les syllabes finales des groupes *les hippopotames* et *Nabuchodonosor*. Le nombre de syllabes minimal qui rend les syllabes accentuables effectivement accentuées semble tourner autour de 7. Ceci permet de rendre compte de l'accentuation d'exemples tels que *je ne lui ai **pas** pardon**né***, constitué d'un seul groupe accentuel de 8 syllabes, mais pour lequel un deuxième accent de groupe rythmique apparaît sur *pas* (la négation *ne…pas* appartenant à une classe fermée).

On a donc là un critère syntaxico-phonologique pour établir *a priori* l'accentuation des syllabes finales des groupes accentuels : les groupes sont définis par un mot plein, autour duquel gravitent des mots-outils en relation de dépendance ou de solidarité (c'est-à-dire de dépendance biunivoque) avec le mot plein (Martin 1987).

La non-accentuation effective d'une syllabe accentuable (dans le cadre du groupe accentuel défini plus haut) est également un indicateur de cohésion syntaxique, et donc peut fonctionner comme indicateur de la structure prosodique de la phrase. Ainsi dans | *l'amie* | *de Juliette* | *a vaincu* | *son angoisse* |, on ne peut désaccentuer *de Juliette* sans également désaccentuer les groupes | *l'amie* | et | *a vaincu* |, ce qui donnerait **l'a**__**mie**__ *de Juliette a vaincu son an*__***goi***__***s****se*. Cette contrainte de désaccentuation est analogue à celle qui régit le recul d'accent lors d'une collision, c'est-à-dire d'une séquence de deux syllabes accentuées successives. Dans ce cas également, le recul ou non recul de la première syllabe de la séquence est corrélé à la structure syntaxique locale.

On a par exemple *Juliette aime le **cho**colat **chaud*** avec un recul d'accent sur *chocolat* lorsque les 2 unités *chocolat* et *chaud* sont dominés immédiatement par le même nœud (réponse à la question "*qu'est ce que Juliette aime ?*", et pas de recul *Juliette aime le choco**lat chaud*** lorsque 2 nœuds distincts dominent ces 2 unités (réponse à la question "*comment Juliette aime-t-elle le chocolat ?*").

4. Accent et frontière macro syntaxique

D'autres conditions, cette fois d'ordre macro-syntaxique, rendent également la réalisation de l'accent démarcatif obligatoire. On définit en général comme unité macro-syntaxique (un macro-segment) une séquence d'unités syntaxiques bien formées du point de vue de la syntaxe classique (générative transformationnelle ou autre). Un énoncé est alors constitué par un ou plusieurs macro-segments dont l'un d'eux, le noyau, peut apparaître seul à la place de tout l'énoncé, et reste bien formé tant du point de vue syntaxique que du point de vue prosodique. Il en résulte que le noyau est porteur de l'information de modalité (déclaration, interrogation, négation, etc.), et qu'il se termine nécessairement par un contour final de modalité (descendant et bas dans le cas déclaratif). Les macro-segments qui précèdent le noyau sont appelés préfixes, ceux qui le suivent sont soit des postfixes (cas des macro-segments placés après le contour de focalisation large), soit des suffixes (cas des compléments différés) (cf. Blanche-Benveniste 2000).

Dans l'énoncé oral spontané, de même que dans sa contrepartie graphique par la ponctuation, l'accent de groupe (i.e. l'accent de groupe prosodique) est essentiel pour indiquer les frontières des macro-segments, qui ne sont autrement indiquées que par l'absence de relation syntaxique entre les deux unités syntaxiques situées de part et d'autre de la frontière entre les macro-segments (à l'exception des compléments différés). Ainsi dans l'énoncé suivant :

(2) vous savez **pas** | tout le **temps**|on marche le soir comme **ça** | une pa**trouil**le | il nous **voit** | direct au con**trôl**e | on va au poste juste pour une carte d'identi**té** | t'as pas ta carte tu fais tes vingt quatre **heur**es | tu ressors t'as la haine encore **plus** | ça aug**ment**e | après t'as envie de tout ca**sser** | t'as envie **de** | tu **vois**. [Meaux, 2]

les frontières entre macro-segments, indiquées ici par le symbole "|", sont réalisées par une syllabe accentuée en finale de segments. L'accentuation de ces syllabes est réalisée par une chute importante de fréquence fondamentale sans augmentation de durée syllabique, caractéristique dans cet exemple du "parler des banlieues". La neutralisation de l'accentuation finale d'un macro-segment violerait nécessairement le principe de collision syntaxique, puisqu'il entraînerait la formation d'un groupe accentuel constitué d'unités appartenant à deux macro-segments distincts (Martin 1987).

5. Qui observe ?

Dans toute observation de données linguistiques, il faut se demander "qui observe" et "dans quel contexte". L'identification des proéminences accentuelles n'échappe pas à cette règle, et en outre met en jeu des paramètres tels que "quelles catégories morphologiques" sont mises en jeu, et "à quelle place dans la structure syntaxique".

5.1. L'auditeur naïf

L'observateur-transcripteur peut tout d'abord être un auditeur "naïf", qui dans la perception de parole réelle n'entend le flux de parole qu'une seule fois et n'a donc pas la possibilité de revenir en arrière pour modifier son jugement. Son système de perception est constitué par la grille de catégorisation primaire de son système linguistique, donc par sa connaissance, et partant de sa faculté de reconnaissance, du système accentuel de sa langue. Un tel observateur jugera donc de l'accentuation effective d'une syllabe accentuée selon ses propres critères, c'est-à-dire de la façon dont il aurait lui-même réalisé l'accent ou non, et ce aux endroits de la séquence syllabique où il s'attend à trouver un accent.

Un exemple typique porte sur la perception de la syllabe accentuée en espagnol par un auditeur francophone. Dans la phrase *es la hi**sto**ria de un delito feroz*, l'accent placé sur l'antépénultième dans *hi**sto**ria* ne sera pas reconnu au profit de la dernière syllabe au contour mélodique généralement montant. La conjugaison de la position finale dans le groupe et d'une caractéristique mélodique semblable au français convaincra l'auditeur "naïf" de la position finale de l'accent lexical.

5.2. Le phonéticien

Un observateur phonéticien va lui mettre en jeu ses systèmes de catégorisation secondaire, filtrant la réalité perceptive de manière plus fine que son système primaire établi par sa connaissance de la langue en question. Ce système secondaire aura été constitué par un apprentissage relatif aux caractéristiques plus générales des syllabes proéminentes, établies éventuellement par un travail effectué sur d'autres systèmes linguistiques que le sien, ou encore au travers d'une connaissance expérimentale des caractéristiques acoustiques de la proéminence.

5.3. Le phonologue

Le phonologue pour sa part mettra en jeu sa connaissance des propriétés phonologiques de la proéminence dans la langue, entre autres les positions attendues des syllabes accentuables, et n'aura pas de mal à distinguer les réalisations de l'accent lexical des accents de focalisation étroite ou de focalisation large, pour autant que les positions des syllabes impliquées ne soient pas ambiguës. L'accent de focalisation large se distingue facilement des autres accents lexicaux par leur nature acoustique (une chute important de fréquence fondamentale), alors que celui de focalisation étroite, généralement à mélodie montante et placé sur la première syllabe, pose problème si le groupe accentuel se réduit à une syllabe.

6. Dans quel contexte ?

Comme mentionné plus haut, la vitesse d'élocution va également jouer un rôle, et le ralentissement du débit obtenu par resynthèse (par exemple dans les logiciels PRAAT et WINPITCH), souvent utilisé par les transcripteurs pour faciliter leur jugement, modifiera également leur perception, en modifiant les conditions d'application du critère des groupes accentuels avec un maximum de 7 syllabes).

Les systèmes phonologiques des auditeurs et des locuteurs peuvent être les mêmes, mais leurs variantes dialectales différentes. La perception de la proéminence réalisée par un locuteur de Marseille ne sera pas nécessairement la même pour un auditeur fréquemment exposé à cette variante que pour un auditeur de Lille.

La fréquence d'occurrence pour une position spécifique dans la structure syntaxique jouera également un rôle. Ainsi, en position neutralisée, où l'auditeur ne s'attend pas à l'apparition d'une proéminence, un jugement positif aura-t-il plus de chances d'être retenu que par exemple en position de continuation majeure, et ce avec des traits prosodiques comparables.

7. L'analyse acoustique

Ailleurs dans cet ouvrage, Mertens propose un algorithme de transcription automatique de la prosodie, qui intègre les seuils de perception des glissandos, et dès lors prend également en compte les durées syllabiques (dans leurs parties voisées) et la vitesse de variation mélodique. Si *a priori* l'utilisation d'une telle approche semble ouvrir la voie à une transcription de la proéminence "neutre", c'est-à-dire dépourvue de facteurs subjectifs, la nature paramétrique du seuil de glissando rend cette solution évidemment dépendante du choix de l'utilisateur du logiciel, et renvoie en définitive au jugement d'auditeurs humains (par exemple en choisissant une valeur de seuil qui semble mieux correspondre à l'avis d'une majorité de transcripteurs).

Un autre problème ressort du processus utilisé dans l'algorithme de transcription, qui se base sur une segmentation syllabique fiable, au moins en ce qui concerne les syllabes *a priori* susceptibles d'être accentuées. Or d'une part une segmentation manuelle est sujette aux erreurs d'appréciation des frontières, même avec l'emploi de spectrogrammes, et d'autre part la segmentation automatique ne peut fonctionner de manière raisonnablement utile que pour des enregistrements avec très peu de bruit de fond, et de plus présentant peu d'écart par rapport aux prononciations standards qui ont servi à déterminer les paramètres nécessaires au moteurs de reconnaissance automatique.

Enfin la présence d'événements acoustiques non prosodiques ou extra-syllabiques (variations spectrales vocaliques, présence de pause succédant à la syllabe, etc.) constitue d'autres sources d'erreurs dans la qualification automatique de la qualité accentuée d'une syllabe puisque ne figurant pas parmi les paramètres pris en compte dans le calcul. Il s'agit dans ce cas d'une partie du contexte qui échappe au processus de jugement de l'accentuation.

Si la proéminence d'une syllabe se manifeste nécessairement par un ou plusieurs traits prosodiques qui la différencient d'une ou de deux de ses voisines (2 syllabes proéminentes peuvent être contiguës, mais pas 3), le paramètre de variation mélodique ne saurait suffire pour en déterminer la qualité d'accentuée, puisque, parmi bien d'autres raisons expérimentales, une syllabe peut être proéminente du seul fait de sa durée même avec un contour mélodique plat (par exemple dans la partie thème dans une division propos/thème ou dans un énoncé prononcé par un locuteur dépressif…). Un système automatique devra donc intégrer ces variations de réalisations, qui déterminent des zones non connexes dans un espace perceptif multidimensionnel.

Une solution possible pour résoudre au moins partiellement le problème consisterait à établir, à partir de mesures de traits prosodiques de durée, de variation mélodique, d'intensité, de qualité vocalique, de pause, etc., une hiérarchie de proéminence accentuelle, et ensuite déterminer une frontière au delà de laquelle les syllabes serait considérées comme proéminentes (cf. Martin 1979).

8. Conclusion

Un bouquet de raisons rend la perception et la transcription de la proéminence accentuelle en français des plus difficiles, jusqu'à l'apparenter à un art plutôt qu'à un processus rigoureux. Il n'est donc pas étonnant que les nombreux transcripteurs phonologues et phonéticiens cités dans le travail de Poiré (ce volume) s'accordent mal sur les données identiques proposées à leur expertise. Du reste, une expérience analogue avait été menée sur la cohérence de transcription ToBI par des experts de ce type de notation pour l'anglais et les résultats présentaient des dispersions similaires (Pirelli et al. 1994). À chaque fois l'observateur, du seul fait de son activité de perception, va juger de la proéminence en fonction de ses connaissances variées dans les domaines phonétiques, phonologiques, syntaxiques, etc. La probabilité d'occurrence d'une proéminence à tel ou tel endroit de la structure syntaxique des segments examinés sera déterminante en se conjuguant avec les habitudes de perception phonologique et phonétique de l'opérateur. La possibilité de réécoute des segments de parole, voire de ralentissement par resynthèse, viendra également influer sur les résultats.

Références

Blanche-Benveniste, C. (2000). *Approches de la langue parlée en français*, Ophrys, Paris.

Dell, F. (1984). "L'accentuation dans les phrases en français". In F. Dell, D. Hirst & J.R Vergnaud (eds). *Forme sonore du langage: structure des représentation en phonologie*, Paris : Hermann, 65-122.

Garde, P. (1968). *L'accent*, Collection Le Linguiste, PUF : Paris.

Martin, Ph. (1979). "Automatic Location of Stressed Syllables in French", *Proceedings of the International Congress of Phonetics*, Miami, 1091-1094.

Martin, Ph. (1987). "Prosodic and Rhythmic Structures in French", *Linguistics*, Vol. 25-5, 925-949.

Pirelli, J., M. Beckmann et J. Hirschberg (1994). "Evaluation of Prosodic Transcription labeling reliability in the ToBI framework", *ICSLP94* 1, 123-126.

Prieto, L. (1975). *Pertinence et pratique,* Éditions de Minuit : Paris.

Wioland, F. (1985). *Les structures rythmiques du français*, Slatkine-Champion : Paris.

Identification d'accents régionaux en français : perception et catégorisation

Cécile WOEHRLING
Philippe BOULA de MAREÜIL[1]

1. Introduction

On dispose aujourd'hui de nombreux enregistrements collectés en différents points de la Francophonie, dans le cadre du projet "Phonologie du Français Contemporain" (PFC). Ce projet (Delais-Roussarie & Durand 2003) s'inscrit dans le sillage de la grande enquête phonologique de Walter (1982), qui portait sur une trentaine de régions francophones d'Europe. Dans les données audio de PFC, recueillies sur le terrain, différents accents sont représentés, autant de déviations par rapport à une norme, repérables à certains traits phonétiques suffisamment saillants pour pouvoir être reconnus et caractérisés. Des auditeurs naïfs sont-ils capables d'identifier ces accents ? Les anecdotes sont monnaie courante en la matière. Avec quel degré de granularité, par exemple, des accents méridionaux du Sud-Est et du Sud-Ouest peuvent-ils être distingués ? Combien d'accents une oreille d'expert ou celle d'un non spécialiste est-elle à même de discerner ? Ces questions ne sont pas nouvelles en dialectométrie (Séguy 1973), même si ici, à la suite de Wells (1982), nous distinguons entre accents et dialectes traditionnels. Pour le linguiste Garde (2004), ce souci de tracer des frontières, de définir les limites entre le *même* et l'*autre*, est à relier à l'essor de l'idée d'état-nation au sens moderne et dans l'acception exclusive du terme. La frontière (du latin *frons*) est une ligne de démarcation, de délimitation, de séparation qui cristallise dans l'espace un fait social qu'elle ordonne, établissant des catégories — celles du proche et du lointain, du dedans et du dehors. En sciences du langage, la méthode des isoglosses a connu un certain succès : les isoglosses sont des lignes qui séparent des zones différant les unes des autres d'une certaine manière, par exemple par une prononciation différente d'un mot donné. On peut tracer des isoglosses selon plusieurs critères et obtenir des cartes en combinant les isoglosses obtenues. Toutefois, si on regarde des isoglosses tracées selon des critères différents, elles peuvent ne pas coïncider, et il est alors difficile de déterminer lesquelles utiliser de préférence. Une méthode alternative, dite des flèches, consiste à demander à des sujets de citer les lieux dont ils se sentent proches par la manière de parler et ceux qu'ils pensent être complètement différents (Preston 1989). À partir des réponses obtenues, il est possible de construire une carte, par exemple en reliant par des flèches les points désignés comme proches. Cependant, cette méthode ne fonctionne que pour des endroits proches géographiquement : pour des points éloignés, il est impossible de dessiner les flèches et, partant, de refléter leur proximité éventuelle dans la façon de parler. Comme nous nous intéressons à la fois à des variétés du Nord et du Sud de la France, une autre approche perceptive doit être appliquée.

La réponse aux questions posées ci-dessus dépend notamment de l'origine géographique des sujets interrogés. Il est donc intéressant de mener des tests perceptifs en différents lieux, pour confronter les résultats. Plusieurs tranches d'âge et "styles" de parole (celui de la lecture et celui de l'entretien notamment) étant représentés dans le corpus PFC, dans la lignée labovienne (Labov 1972), leur influence sur les performances demande également à être quantifiée. Tel est l'objet que ce travail se propose d'examiner.

[1] LIMSI-CNRS | BP 133 — F-91403 Orsay CEDEXh
{Cecile.Woehrling;Philippe.Boula.de.Mareuil}@limsi.fr

Clopper et Pisoni (2004) ont montré que, sans entraînement préalable ni retour (*feedback*), des auditeurs américains, invités à écouter des compatriotes de différents accents et à localiser leur origine géographique sur une carte des États-Unis, sont capables de distinguer trois grandes régions : Nouvelle Angleterre, Sud et Nord-Ouest. Une thèse sur les dialectes norvégiens et néerlandais notamment (Heeringa 2004) a également développé une cartographie des distances phonologiques et acoustiques qui existent au niveau lexical au sein d'un même ensemble dialectal. D'autres travaux encore ont été consacrés à l'identification de quatre variétés de néerlandais et de cinq variétés d'anglais par des auditeurs des pays concernés (van Bezooijen & Gooskens 1999). À notre connaissance, les résultats d'une tâche similaire de classification (*clustering*) perceptive n'existent pas pour le français, même si une distinction Nord/Sud semble évidente pour tout locuteur natif.

Depuis le début du XX^{e} siècle, à la suite des atlas linguistiques de Gilliéron et Edmont (1902-1920), la variation lexicale et phonétique a connu un grand intérêt. Une tentative de visualisation de l'Atlas Linguistique de la France (ALF, qui couvre plus de 600 localités) a récemment été entreprise (Goebl 2002). En sociolinguistique, les études se focalisent sur les représentations de variétés spécifiques, plus ou moins stéréotypées et éventuellement différentes des comportements en réaction à des échantillons de parole réels, lorsqu'on demande à des auditeurs d'évaluer des locuteurs, de les juger subjectivement en termes d'intelligence, d'amabilité, de virilité, de correction, etc. (Hintze *et al*. 2000). Mais la plupart des études sont descriptives, et ne permettent pas de prédire avec certitude les caractéristiques les plus discriminantes qui sont associées à un accent donné. La vaste base de données issue du projet PFC permet maintenant des études systématiques.

2. Expérience

2.1. Locuteurs

Un des objectifs du projet PFC est de couvrir un vaste territoire, à travers des informateurs ancrés géographiquement — étant nés et ayant passé une grande partie de leur vie en un même lieu. Notre propre expérience porte sur six régions francophones, correspondant à autant de points d'enquête PFC : Brécey (Normandie), Treize-vents (Vendée), le canton de Vaud (Suisse romande), Biarritz (Pays Basque), Douzens (Languedoc) et Marseille (Provence). Dans chacun de ces points d'enquête, nous avons sélectionné six locuteurs de niveaux d'étude variés, trois hommes et trois femmes, répartis en trois tranches d'âge : 15-30 ans (moyenne : 23 ans), 30-60 ans (moyenne : 47 ans), 60 ans et plus (moyenne : 73 ans).

Dans une expérience perceptive comme dans un sondage par téléphone ou autre, on est limité en temps : il faut donc faire des choix. Pour une tâche d'identification, comme chez Clopper et Pisoni (2004), une alternative entre six possibilités, avec un niveau de hasard de 17 %, nous a semblé un bon compromis. Comme on le voit dans la figure 1, nous avons retenu 3 points au Sud de la Loire : Sud-Ouest (des locuteurs bascophones), plein Sud (milieu rural de vignerons au parler conservateur) et Sud-Est (la grande ville de Marseille). Au Nord, deux points sont situés à la même latitude : Vendée et Suisse romande, le "parler de l'Est" étant incarné par le canton de Vaud. Un souci de symétrie voudrait que l'on choisisse un point en région parisienne. Mais comme Paris représente la norme (Calvet 1996) et que dans la conscience communément partagée "ceux qui ont un accent, ce sont toujours les autres", interroger des sujets sur "l'accent parisien" risquerait de les dérouter. Faut-il comprendre l'accent "parigot", l'accent "branché" ? Devant cette difficulté, nous avons plutôt opté pour un point situé au Nord-Ouest, sur le même axe que le Pays Basque et la Vendée, en Normandie.

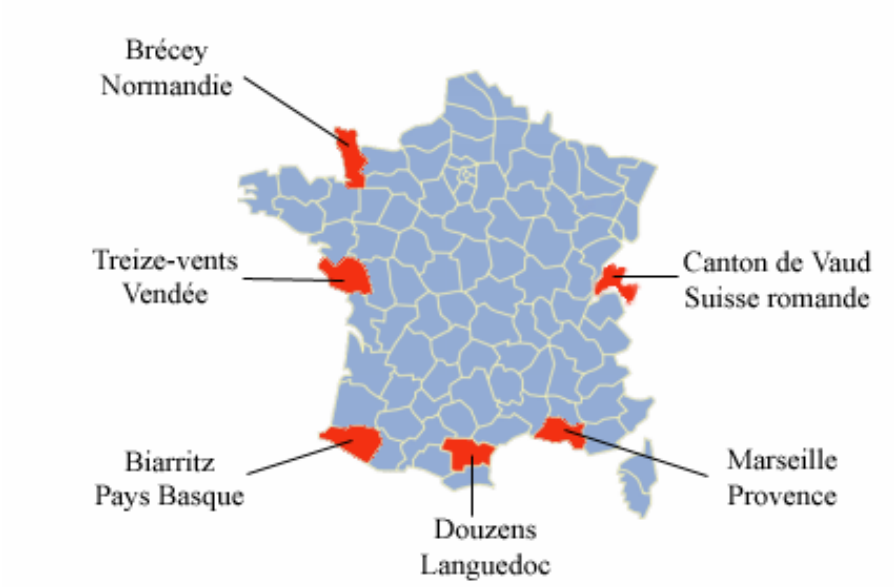


Figure 1. *Zones correspondant aux six points d'enquête retenus*

2.2. Stimuli

Pour chacun des locuteurs, nous avons choisi deux échantillons de parole. Le premier est une longue phrase lue (25 mots) tirée du milieu du texte PFC, identique pour tous : "La côte escarpée du mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère." Le second est un extrait de parole spontanée, tiré d'un entretien guidé — à caractère autobiographique. Il a été choisi d'après les critères suivants : énoncé assertif d'une durée équivalente à celle de l'extrait lu (8-9 secondes, *cf.* tableau 1), absence de référence à un lieu qui biaiserait l'identification, absence d'intervention de l'interviewer et peu d'hésitations de la part du locuteur. Avec en moyenne 33 mots par extrait, le débit de la parole spontanée (10,6 phonèmes/seconde) est comparable à celui de la lecture (10,4 phonèmes/seconde).

Tableau 1 : *durée des stimuli retenus pour l'expérience*

	Durée moyenne (s)	Durée maximum (s)	Durée minimum (s)
Lecture	8,30	12,44	6,2
Parole spontanée	9,39	10,96	7,4
Ensemble des *stimuli*	8,84	12,44	6,2

Au total, nous avons 3 tranches d'âge × 2 sexes × 6 points d'enquête × 2 styles de parole = 72 *stimuli*.

Les études antérieures reposent souvent sur la seule "parole lue", même si Clopper et Pisoni (2004) reconnaissent les problèmes que cela pose. L'utilisation de parole lue et de parole spontanée offre selon nous plusieurs avantages. D'abord, la présentation d'une même phrase lue permet de comparer toutes choses égales par ailleurs, et se prête mieux à des mesures acoustiques ; la parole spontanée quant à elle représente un registre de langue qui reflète mieux le vrai vernaculaire des locuteurs — leur façon naturelle de parler, dans un usage quotidien. En utilisant les deux types de parole, il est possible de vérifier si

l'un ou l'autre favorise la reconnaissance des accents, ou si les deux sont équivalents — permettant de nouvelles généralisations. Enfin, la présentation des deux types de *stimuli* dans un ordre aléatoire permet de ne pas lasser les auditeurs et de maintenir ainsi leur attention en éveil : les sujets n'ont pas à écouter toujours la même phrase.

2.3. Auditeurs et protocole

Le test perceptif a été soumis à vingt-cinq auditeurs résidant en région parisienne, tous de langue maternelle française. Sans problèmes d'audition connus et membres d'un laboratoire d'informatique (non rémunérés pour cette tâche), ces auditeurs, neuf femmes et seize hommes, étaient âgés en moyenne de 32 ans. Ils avaient passé en moyenne 21 ans en région parisienne. Ils se disaient quasiment tous familiers des accents de Marseille et de Suisse, quasiment tous non familiers des accents de Vendée, de Normandie et du Languedoc.

Le test se déroulait dans une chambre isolée, les auditeurs étaient munis d'un casque fermé du même modèle, le niveau d'écoute était confortable. Les *stimuli*, au format Wave, étaient échantillonnés à 22,05 kHz, 16 bits, mono. Leur niveau sonore avait été égalisé à l'aide du logiciel Goldwave[2], également utilisé pour la segmentation des stimuli.

Le test était réalisé à travers une interface conviviale (Vieru-Dimulescu & Boula de Mareüil, 2004), qui permet entre autres, en cliquant sur des boutons, d'entrer des informations sur la familiarité avec tel ou tel accent et de saisir les réponses. Tout d'abord, brève familiarisation, l'auditeur écoutait une fois la même phrase lue par un locuteur ou une locutrice (non utilisée par la suite) de chacune des six régions en question, qui était indiquée. Lors de la phase suivante, le test proprement dit, l'auditeur écoutait 74 *stimuli*, dont les deux premiers (phrases spontanées d'un locuteur du Nord et d'une locutrice du Sud) n'étaient pas comptés dans les résultats. Les 72 *stimuli* suivants, extraits lus ou spontanés mélangés, étaient présentés un par un dans un ordre aléatoire différent pour chaque auditeur. Après chaque écoute, l'auditeur devait préciser d'après l'accent l'origine du locuteur parmi les six possibilités déjà mentionnées : Brécey (Normandie), Treize-vents (Vendée), le canton de Vaud (Suisse romande), Biarritz (Pays Basque), Douzens (Languedoc) et Marseille (Provence). Il ne recevait aucune indication sur l'exactitude de ses réponses, et pouvait prendre le temps qu'il voulait pour répondre. Chaque stimulus pouvait être réécouté, mais il était impossible de revenir en arrière. L'expérience durait une vingtaine de minutes.

2.4. Analyse acoustique du corpus (phrases lues)

Des mesures acoustiques ont été menées, facilitées par l'alignement en phonèmes donné par la Reconnaissance Automatique de la Parole (RAP) : durée des schwas et des appendices nasaux éventuellement réduite à zéro (*cf*. tableau 2), extraction des formants de certaines voyelles (*cf*. tableau 3), etc. La procédure, fondée sur l'alignement dynamique de modèles de Markov cachés à densité continue, indépendants du contexte, est décrite dans Adda-Decker *et al*. (2005). Le décodeur produit la séquence de phonèmes réalisée la plus probable, laquelle est de taille variable puisque des variantes avec ou sans /ə/, avec ou sans /n/, etc. ont été prévues dans le dictionnaire de prononciation. Signalons ici que la RAP a déjà été exploitée pour l'identification de variétés de néerlandais (ten Bosch, 2000), et que la dégradation des performances pour la reconnaissance de variétés non prévues au départ a été à l'origine d'études acoustiques sur des variétés d'anglais (Yan & Vaseghi 2002 ; Hansen *et al.* 2004). Les systèmes de transcription automatique, en effet, sont des

[2] http://www.goldwave.com

révélateurs privilégiés de la variabilité, et peuvent être mis à profit pour affiner les descriptions.

Bien documentées, de grandes tendances caractérisent le français méridional par rapport au français standard (Walter 1982 ; Carton *et al*. 1983 ; Durand 1995 ; Binisti & Gasquet-Cyrus 2003) : d'abord, la réduction du nombre des oppositions de timbre semi-ouvert ~ semi-fermé pour les voyelles moyennes. Ensuite, là où le français standard utilise des voyelles nasales, le français méridional prononce souvent des voyelles partiellement nasalisées et suivies d'un élément consonantique nasal bien audible. Cet appendice s'articule au même lieu que la consonne suivante, si celle-ci existe, et se réalise souvent [ŋ] à la finale absolue. Enfin, le *e* caduc est presque toujours prononcé, ailleurs que devant voyelle. Nous avons cherché à le quantifier, en nous appuyant sur l'alignement automatique.

La phrase lue que nous avons retenue comprend les trois voyelles nasales /ɔ̃/, /ɛ̃/, et /ɑ̃/ du français, et compte dix schwas potentiels. Dans le tableau 2, les chiffres ont été calculés, une fois ramenées à 0 les durées des phonèmes inférieures à 30 ms. Au moment de la rédaction de ces lignes, nous n'avons traité que les points d'enquête des Languedoc, Pays Basque et Vendée. En moyenne, les locuteurs languedociens ont réalisé plus de 5 schwas sur 10 potentiels et 2 appendices nasaux pour 3 voyelles nasales (que le décodeur a détectées comme dénasalisées dans au moins la moitié des cas) ; les locuteurs basques ont réalisé 3 schwas et 1 appendice nasal, les locuteurs vendéens ont réalisé 2 schwas et 0 appendice nasal. Des mesures formantiques ont été prises au milieu des schwas, grâce à un script écrit pour le logiciel libre PRAAT[3]. Cependant, elles n'ont montré aucune différence sensible de timbre entre le Nord et le Sud.

Tableau 2 : *durées moyennes (en ms) des schwas et des nasales potentiels, calculées automatiquement par alignement sur la phrase lue dans le Nord (Vendée) et dans le Sud (Languedoc et Pays Basque).*

	côte	*Pierre*	*mène*	*village*	*barrages*	*chaque*	*que*	*de*	*manifest ent*	*colère*	*mont*	*saint*	*opposant s*
Languedoc	50	68	0	48	35	48	37	92	18	0	25	78	33
Pays Basque	28	0	20	38	5	15	33	57	13	0	0	50	8
Vendée	22	8	0	13	0	13	23	45	25	0	0	0	8

Le tableau 3 montre que le /O/ de *côte* est plus ouvert dans le sud que dans le nord, jusqu'à neutraliser l'opposition avec *cote*. Dans *connaît*, les /E/ sont comparables entre le nord et le sud, mais les /O/ sont différents, cette fois moins en termes d'aperture (F1) qu'en termes d'antériorité (F2). Dans le nord, le /O/ tend à être prononcé [œ], comme l'avait déjà remarqué Martinet (1969) dans un article célèbre. On peut également voir dans le cas présent un phénomène d'harmonie vocalique (Fagyal *et al*. 2002). Les valeurs des formants, arrondies à 50 Hz dans le tableau 3, ont été prélevées au milieu des voyelles en utilisant les options par défaut de PRAAT. D'après un test de Student bilatéral apparié, les différences nord/sud sont significatives pour les valeurs de F1 dans *côte* [$t(17) = -3{,}68$; $p < 0{,}01$] et les valeurs de F2 dans le /O/ de *connaît* [$t(17) = 8{,}37$; $p < 0{,}01$] ; elles ne sont pas significatives pour les autres mesures.

[3] http://www.fon.hum.uva.nl/PRAAT/

Tableau 3 : *fréquences formantiques des /O/ des mots côte et connaît dans la phrase lue par les hommes et les femmes des six points d'enquête*

	côte		*connaît*	
	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
Nord	450	1150	450	1500
Sud	550	1200	450	1100

Les parlers de Vendée ont également été étudiés (Léonard 1991), mais de l'aveu même de l'auteur, leur francisation s'opère à grande vitesse ; il en va de même en Normandie. Le français parlé en Suisse romande présente en revanche un certain nombre de caractéristiques notables (Hintze *et al*. 2000) : entre autres, l'opposition /a/~/ɑ/, l'utilisation de quatre voyelles nasales /ɔ̃/, /ɛ̃/, /ɑ̃/ et /œ̃/), ainsi que le maintien des oppositions de quantité vocalique. Même si nous n'avons pas pu mesurer ces faits dans notre corpus, il n'en reste pas moins que nos locuteurs suisses sont bien identifiables (*cf. infra*). En particulier, une diphtongaison plus ou moins subtile des voyelles moyennes peut être observée dans notre corpus, chez les locuteurs du canton de Vaud — par exemple, chez les plus vieux d'entre eux, à la fin du mot *escarpée*, où la différence de F1 donnée par PRAAT, entre le premier quart et le troisième quart du /e/, peut atteindre 100 Hz.

Non seulement les voyelles mais aussi les consonnes sont sujettes à la variation. En particulier, dans notre corpus, les locuteurs à l'accent prononcé du Languedoc — les plus âgés — produisent systématiquement des /R/ apicaux. Cette situation s'est perdue en Provence, tandis que nos locuteurs basques ont des *r* qui s'apparentent davantage à la *jota* espagnole [x] — à l'exception de la locutrice la plus âgée, qui "roule" les *r*. Quant au /R/ normand, soulignons qu'il est parfois très reculé ([ʕ] pharyngal). Mais nous n'avons pas quantifié ces différentes articulations du /R/. On sait que ce phonème est polymorphe, et dans certaines positions c'est la coarticulation qui joue le rôle principal (Autesserre & Chafcouloff 1999). L'identification du segment consonantique s'avère d'autant plus difficile que celui-ci peut présenter une structure formantique. Ce n'est pas un hasard s'il pose tant de problèmes en phonétique.

Nous n'avons pas analysé acoustiquement les énoncés spontanés, mais la prononciation [tja] pour *tu as* à Marseille, qui fonctionne presque comme un schibboleth, est ici illustrée dans une des phrases (*cf*. figure 2). Des expressions typiquement marseillaises, transcrites par exemple *ti es fada* (voire *ties fada*) dans des forums sur Internet, ainsi que les sketchs d'humoristes tels que Patrick Bosso suggèrent que ce trait est intégré dans la conscience linguistique des individus, et peut être simulé, imité, caricaturé. Plus généralement, /t/ et /d/ sont fortement palatalisés devant /i/ et /y/ à Marseille (Binisti & Gasquet-Cyrus 2003). Concernant nos auditeurs de région parisienne, les résultats montreront que l'énoncé spontané d'une jeune Marseillaise, avec cette forme de tutoiement en [tja], est nettement mieux reconnu que la phrase lue par la même locutrice.

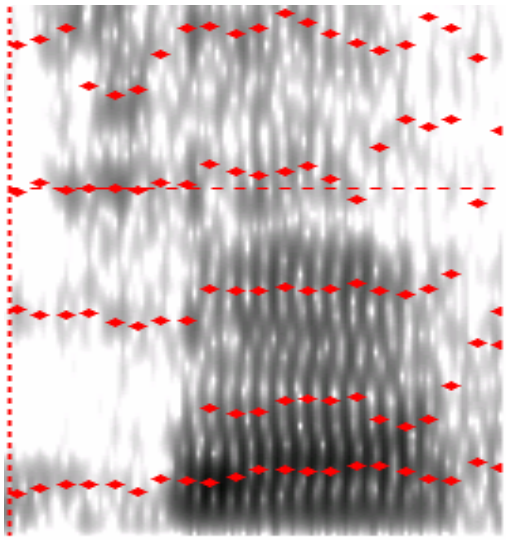


Figure 2. *Spectrogramme correspondant à la séquence* tu as*, prononcé par une jeune Marseillaise*

3. Résultats

Toutes les origines géographiques ont été reconnues significativement mieux que le hasard [$p < 0{,}01$ d'après des tests de χ^2 à 5 degrés de liberté]. Tous *stimuli* confondus, les auditeurs ont donné 42,1% de réponses exactes (*cf.* tableau 4, figure 3). Des analyses de variance (ANOVA) ont été conduites sur les réponses comptées comme correctes (1) ou incorrectes (0) avec le facteur aléatoire Sujet et les deux facteurs intra-Sujet Style et Âge. S'il n'y a pas d'effet du facteur Style (corrélation significativement positive des résultats [$T >> 1{,}64$] entre lu et spontané) ni d'interaction significative entre les deux facteurs, le facteur Âge se révèle avoir un effet majeur [$F(2, 48) = 0{,}08$; $p < 0{,}05$]. De fait, l'origine des locuteurs les plus vieux est mieux reconnue que celle des plus jeunes : on observe pour les catégories de locuteurs des plus aux moins âgés respectivement 46,3 %, 42,5 % et 37,5 % de réponses correctes. La différence est moins marquée entre la parole lue, donnant un score de 41,9 %, et la parole spontanée, donnant un score de 42,3 %.

De toutes les régions, la Suisse romande est la mieux identifiée. Elle est suivie par la Normandie et la Vendée, souvent confondues entre elles. Les locuteurs de ces deux régions ont été identifiés plus souvent comme normands que comme vendéens. Le taux de reconnaissance des locuteurs normands (supérieur à 50 %) tient donc d'une part à ce phénomène, et d'autre part au fait que les auditeurs hésitaient entre seulement deux réponses, une fois exclus la Suisse romande et le sud de la France — voir le bruit que représente l'identification à 48 % des locuteurs vendéens comme normands. Les chiffres sont par conséquent à prendre avec prudence.

Les trois régions du sud ont été confondues entre elles, de sorte que leur taux d'identification moyen est inférieur à 33 %. Vingt-deux des 25 auditeurs s'estimaient pourtant capables, avant le test, de reconnaître l'accent de Marseille parmi les six proposés. Contrairement à ce qui se passait pour le nord de la France, les auditeurs n'ont pas choisi l'une des trois régions beaucoup plus fréquemment que les autres — quelle que soit l'origine des *stimuli*. Fait intéressant, les *stimuli* du Pays Basque et de Provence ont majoritairement été reconnus comme languedociens, et ceux du Languedoc comme provençaux.

Ce sont ces locuteurs languedociens qui sont le mieux identifiés comme méridionaux — à 92 %. De façon symptomatique, leur accent est plus souvent associé à

Marseille que la cité phocéenne elle-même. Contraire à ce que suggèrent Coquillon *et al.* (2000) à un niveau prosodique, ce résultat est conforme à notre intuition : l'image qu'on se forge, au nord, de l'accent marseillais est en réalité un stéréotype méridional qui va bien au-delà de Marseille. Pour les locaux, d'ailleurs, il n'y a pas un, mais trois "accents de Marseille" (Binisti & Gasquet-Cyrus 2003) : celui des "quartiers nord" (correspondant à la façon de parler dans les quartiers difficiles), celui des "vrais" Marseillais (renvoyant à l'imaginaire pagnolesque) et celui de la "bourgeoisie marseillaise" (plus léger). D'après le niveau d'étude et la profession de nos locuteurs, ceux-ci appartiennent plutôt à la dernière catégorie, ce qui peut expliquer en partie les résultats.

Tableau 4 : *matrice de confusion sur l'ensemble des données pour 25 auditeurs de région parisienne (%). Les pourcentages sont donnés par rapport à 12 × 25 = 300 réponses*

Réponse / Origine	Vendée	Normandie	Suisse	Pays Basque	Languedoc	Provence
Vendée	**37**	48	4,7	2,3	6,7	1,3
Normandie	37	**50,7**	1,7	3,3	6,3	1
Suisse	10,7	14,7	**71**	0,7	2	1
Pays Basque	7	6	1	**36**	38,3	11,7
Languedoc	5,7	2	0,3	25	**29,3**	37,7
Provence	11,7	7,7	2	20	30	**28,7**

4. Clustering et scaling

Des analyses statistiques ont également été menées à l'aide du logiciel libre R[4], afin de fournir des traductions graphiques des résultats : sous forme de dendrogramme par des techniques de clustering hiérarchique, sous forme de plan à deux dimensions par échelonnement multidimensionnel (*multidimensinal scaling*). Les différentes analyses nécessitent de spécifier des mesures de distance. Nous en avons utilisé deux, relativement communes : la distance euclidienne (1) et la distance de Manhattan (2), encore appelée *city blocks*. Leurs expressions sont données ci-dessous pour deux éléments X et Y de taille *n*.

$$\delta(X,Y)=\sqrt{\sum_{i=1}^{n}(X_i-Y_i)^2} \qquad (1) \qquad\qquad \delta(X,Y)=\sum_{i=1}^{n}|X_i-Y_i| \qquad (2)$$

4.1. Clustering : principe

Le résultat d'un clustering hiérarchique est un dendrogramme, c'est-à-dire une représentation des données sous forme d'arbre dans lequel la distance verticale entre deux feuilles est fonction de la distance qui les sépare dans la matrice de confusion. Les dendrogrammes peuvent être produits de deux manières : de bas en haut, avec un algorithme agglomératif, ou de haut en bas, avec un algorithme divisif. Ces deux techniques sont présentées plus en détail par Manning et Schütze (1999).

- **Algorithmes agglomératifs.** Au début, chaque observation est une classe. Ces classes sont groupées petit à petit jusqu'à former une grande classe contenant toutes les observations. À chaque itération, les deux classes les plus proches sont regroupées pour n'en former plus qu'une.
- **Algorithmes divisifs.** Au début, toutes les informations sont regroupées en une seule grande classe. Les classes sont divisées petit à petit jusqu'à ce que toutes les classes ne contiennent plus qu'un seul élément. À chaque itération,

[4] http://www.r-project.org

la classe qui a le plus grand “diamètre” est sélectionnée. Le diamètre d’une classe est la distance maximum entre deux de ses éléments. Pour diviser une classe, l’algorithme cherche d’abord l’observation qui est la plus éloignée du groupe — qui a en moyenne la plus grande distance avec les autres observations du groupe. Cet élément constitue une nouvelle classe. L’algorithme cherche ensuite dans les observations de l’ancienne classe celles qui sont maintenant plus proches de la nouvelle, et les met dans la nouvelle classe. La classe d’origine est ainsi divisée en deux classes.

Ces algorithmes, contrairement par exemple à *k*-moyennes, ne dépendent pas de l’initialisation. Tous deux prennent en entrée une matrice d’observation, la matrice de confusion dans le cas présent. Pour chacun, il est possible de choisir une mesure de distance : euclidienne ou Manhattan.

4.2. Scaling : principe

L’échelonnement multidimensionnel est une technique qui permet la visualisation des distances entre données, dans notre cas grâce à leur représentation dans un espace à deux dimensions. R propose trois algorithmes d’échelonnement multidimensionnel : Classical Multidimensional scaling, Kruskal’s non metric Multidimensional scaling et Sammon’s non linear mapping. Ces trois algorithmes prennent en entrée une matrice de dissimilarité qui peut être calculée à partir des distances entre lignes de la matrice de confusion, ceci en utilisant l’un des deux types de calcul de distance mentionnés plus haut. Cette matrice de dissimilarité, comparable aux tables de distances entre grandes villes qu’on trouve dans certains agendas, s’obtient de la façon décrite dans la figure 3.

	Vendée	Normandie	Suisse	Pays Basque	Languedoc	Provence
Vendée	111	144	14	7	20	4
Normandie	111	152	5	10	19	3
Suisse	32	44	213	2	6	3
Pays Basque	21	18	3	108	115	35
Languedoc	17	6	1	75	88	113
Provence	35	23	6	60	90	86

	Vendée	Normandie	Suisse	Pays Basque	Languedoc
Normandie	22				
Suisse	398	416			
Pays Basque	444	442	484		
Languedoc	478	476	518	154	
Provence	410	410	456	138	82

Figure 3. *Passage de la matrice de confusion à la matrice de dissimilarité en utilisant la distance δ. Les chiffres correspondent aux résultats donnés en pourcentages dans le tableau 4*

- **Classical Multidimensional scaling.** Cet échelonnement, dit métrique, renvoie un ensemble de points représentant les données de telle sorte que les distances entre ces points soient les plus proches possibles des dissimilarités entre les données.
- **Kruskal's non metric Multidimensional scaling et Sammon's non linear mapping.** Ces deux algorithmes sont dits non métriques : au lieu d'essayer d'approximer les dissimilarités elles-mêmes, ils approximent une transformation monotone non linéaire de ces dissimilarités. La monotonie implique que des distances plus petites ou plus grandes sur la carte correspondent à des dissimilarités plus petites ou plus grandes, respectivement. Quant à la non linéarité, elle a pour conséquence d'éventuelles contractions ou expansions. Ce sont des algorithmes itératifs, dont la configuration initiale est donnée par l'échelonnement métrique. À chaque étape, les distances entre points sont comparées aux dissimilarités originales à l'aide d'une fonction de stress, le but étant que cette fonction prenne la plus petite valeur possible. Les deux algorithmes diffèrent par la formule de calcul du stress.

4.3. Clustering et scaling : résultats

Après cette présentation succincte des grands principes présidant aux algorithmes utilisés, revenons à une application à nos données. Dans presque toutes les configurations, le clustering donne une bipartition. L'algorithme divisif avec une distance euclidienne isole toujours la Suisse de la France. Avec la même métrique, l'algorithme agglomératif ne l'isole que pour le sous-ensemble des extraits de parole spontanée. Sur deux sous-ensembles des données, on a une tripartition Suisse romande, nord de la France et sud de la France (algorithme divisif avec distance de Manhattan sur les extraits de parole spontanée et les locuteurs d'âge moyen). Dans les autres cas, le nord est opposé au sud (*cf.* figure 4).

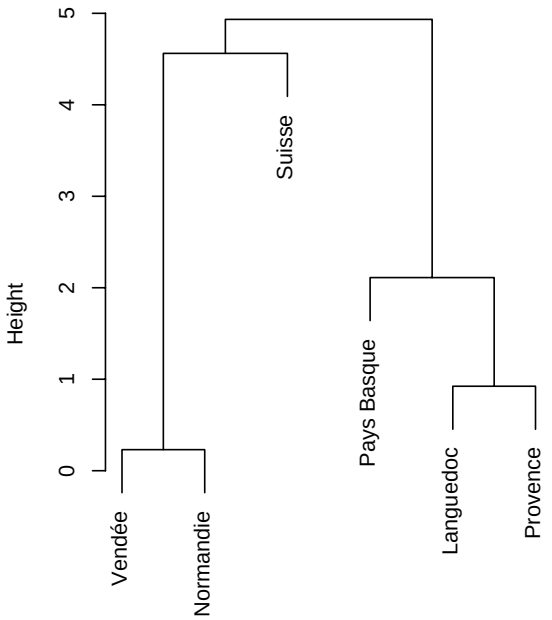


Figure 4. *Dendrogramme issu du clustering hiérarchique agglomératif utilisant une distance euclidienne*

Le scaling, pour sa part, fait apparaître trois groupes : nord de la France, sud de la France et Suisse romande. La première dimension correspond approximativement à l'axe est-ouest, la deuxième à l'axe nord-sud. Dans la figure 5, les axes ont été inversés : axe vertical orienté vers le bas pour placer le nord en haut, axe horizontal orienté vers la gauche pour placer l'ouest à gauche, conformément à une représentation géographique classique — ce qu'il est possible de faire car ce sont les distances entre les points qui sont importantes. Quels que soient l'algorithme et la distance utilisés, sur tout ou partie des résultats, les représentations graphiques sont comparables et les deux dimensions rendent comptent d'une bonne proportion de la variance (entre 91 et 98 % selon les cas).

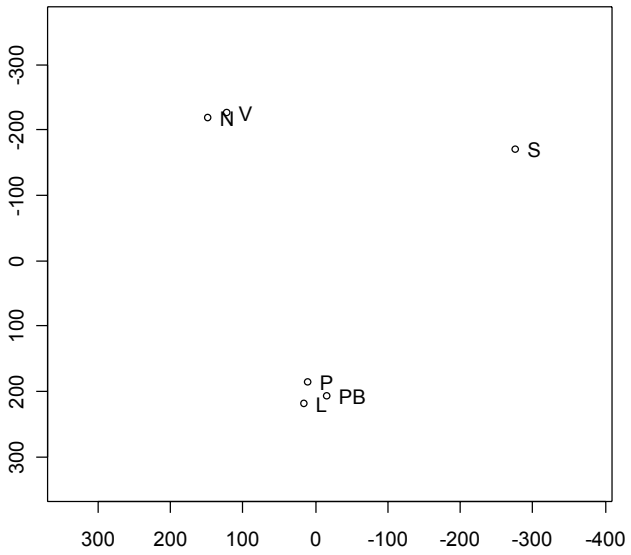


Figure 5. *Résultat de l'échelonnement multidimensionnel avec l'algorithme de Kruskal et la distance de Manhattan — N = Normandie, V = Vendée, S = Suisse, PB = Pays Basque, L = Languedoc, P = Provence*

5. Conclusion et perspectives

Les accents étant des marqueurs d'identité, la question de leur identification nous semble particulièrement importante. Dans cette étude, nous avons fait appel à la perception, à l'instar de Clopper et Pisoni (2004), à des techniques d'analyse de données comme l'échelonnement multidimensionnel, à l'instar de Heeringa (2004), et à la reconnaissance automatique de la parole, à l'instar de ten Bosch (2000) ou encore Hansen *et al*. (2004), Yan et Vaseghi (2002). À partir de 36 locuteurs de 6 régions francophones, nous avons dégagé trois accents, nord, sud et est, sans que le style (lecture formelle ou parole spontanée) ne semble affecter les résultats pour des auditeurs de région parisienne. Ce travail encore préliminaire ne nous permet pas de dire avec précision où passent les lignes de fracture, dans l'hypothèse où l'on voudrait découper ainsi le territoire. Comme l'écrivait de Saussure (1915 : 278–279), cette quête est d'ailleurs vaine dans des zones de transition qui peuvent être ouvertes, perméables et mouvantes. Face à un *continuum* linguistique qui partage une importante base, où tout n'est que fluctuation autour d'un socle commun, le concept de frontière est à manipuler avec précaution, et ne peut qu'aboutir à une partition insatisfaisante. C'est pourquoi nous en sommes restés aux grands traits. D'autre part, nous devons et allons répliquer cette expérience auprès d'auditeurs de Toulouse et d'Aix/Marseille.

Dans leur ouvrage, écrit il y a plus de 20 ans, Carton *et al*. (1983) considèrent une quinzaine d'accents en français, encore davantage subdivisés par Walter (1982). Ceci vaut-il encore aujourd'hui ? Si les accents de Provence et du Languedoc étaient distingués il y a une génération, notre expérience perceptive suggère qu'il n'en va plus de même en ce début de XXI^{e} siècle : une forme nivelée de français méridional émergerait. Cependant, les accents de Picardie, d'Alsace, d'Auvergne et de Corse, parmi d'autres, manquent à notre étude. Des techniques de fusion de données sont envisageables, pour ajouter de nouveaux points dans les espaces fournis par l'échelonnement multidimensionnel.

La représentativité des sujets pose toujours des problèmes sociolinguistiques extrêmement complexes. Un prototype existe-t-il pour une variété donnée ? Si oui, est-il atteint ? Et comment mesurer une distance par rapport à ce prototype, qu'il soit réel ou simplement imaginé ? Un accent peut être stigmatisé, dévalorisé et générateur de

ségrégation, ou au contraire revendiqué, brandi comme un drapeau pour se démarquer, affirmer son identité, son intégration à un groupe, à une communauté (Léon 1993). L'appréciation d'un accent est dictée par le statut social (notamment rural/urbain) auquel celui-ci est attaché, d'où les difficultés. On se heurte alors aux questions du prestige et du capital de sympathie de telle ou telle variété, de telle ou telle forme de la langue. Aussi un accent peut-il être plus ou moins léger, plus ou moins marqué (ce qui reste à mesurer sur une échelle) ; et un mouvement de standardisation peut affecter les couches les plus jeunes de la population. Dans ce contexte, le traitement automatique de la parole peut nous donner des indications précieuses de par son objectivité. La reconnaissance de la parole, en particulier, à travers le taux d'erreurs qu'elle fournit, et les proportions de réalisations de variantes de prononciation que permet de dégager l'alignement automatique, peut être vue comme un outil de mesure de distance par rapport à une norme.

Notre propre expertise scientifique et nos compétences techniques permettant de transcrire, segmenter et aligner en phonèmes de gros corpus de parole, des études résolument expérimentales ont été entamées sur la liaison, le schwa, l'harmonie vocalique et d'autres phénomènes engendrant une restructuration syllabique. Afin de mieux modéliser acoustiquement des parlers non standard et d'améliorer les performances des systèmes de RAP, plusieurs verrous technologiques sont à lever, ce qui revêt un intérêt à la fois pratique et théorique. À partir du corpus PFC, on peut étudier l'impact des accents régionaux sur les performances d'un système de transcription généraliste, et comment adapter un système de transcription à un accent régional. Il pourra être utile d'introduire des xénophones comme /x/, /ʁ/ ou /ŋ/. Parallèlement, il convient d'établir dans quelle mesure un auditeur peut identifier l'origine géographique d'un locuteur à partir d'un échantillon de parole. Ce sont dans ces deux directions que nous nous engageons, le point de départ consistant à cerner ce qu'il est raisonnable de tenter d'identifier automatiquement. Des corrélations entre les réponses des sujets et des traits phonétiques que la RAP permet d'extraire automatiquement seront ensuite calculées. La présence de schwas, notamment, dans les monosyllabes et en positions initiale, interne ou finale de polysyllabes, sera confrontée dans un avenir proche avec les résultats de l'expérience perceptive aussi bien qu'avec l'annotation impressive faite dans PFC.

6. Remerciements

Ce travail a été réalisé dans le cadre du projet VarCom, programme interdisciplinaire TCAN financé par le CNRS. Nous exprimons notre gratitude à Bianca Vieru-Dimulescu pour l'interface expérimentale et les analyses statistiques, à Martine Adda-Decker pour l'alignement automatique, à Noël Nguyen, Julien Eychenne, Jean-Michel Tarrier, Richard Walter, Philippe Hambye pour les données PFC et à tous les sujets que nous avons sollicités pour le test perceptif.

Références

Adda-Decker, M., P. Boula de Mareüil, G. Adda et L. Lamel (2005). "Investigating syllabic structures and their variation in spontaneous French", *Speech Communication* 46(2), 119-139.

Autesserre, D. et M. Chafcouloff (1999). "Étude expérimentale du rôle de l'organisation syllabique dans la prédiction des variantes de /R/ en français", *2es Journées d'Études Linguistiques*, Nantes, 154-159.

Binisti, N. & M. Gasquet-Cyrus (2003). "Les accents de Marseille", *Cahiers du français contemporain* 8, 107–129.

Calvet, L.-J. (1996). *Les politiques linguistiques*, Presses Universitaires de France, Paris.

Carton, F. M. Rossi, D. Autesserre et P. Léon (1983). *Les Accents des Français*, Hachette, Paris.
Clopper, C.G., et D.B. Pisoni (2004). "Some acoustic cues for the perceptual categorization of American English regional dialects", *Journal of Phonetics*, 32**,** 111-140.
Coquillon, A., A. Di Cristo et M. Pitermann, (2000). "Marseillais et Toulousains gèrent-ils différemment leurs pieds ? Caractéristiques prosodiques du schwa dans les parlers méridionaux", *Journées d'Étude sur la Parole*, Aussois, 89-92.
Delais-Roussarie, É. et J. Durand (2003). *Corpus et variation en phonologie du français : méthodes et analyses*, Presses Universitaires du Mirail, Toulouse.
Durand, J. (1995). "Alternances vocaliques en français du midi et phonologie du gouvernement", *Lingua* 95, 27–50.
Fagyal, Z., N. Nguyen et P. Boula de Mareüil (2002). "From *dilation* to coarticulation: is there vowel harmony in French?", *Studies in the Linguistic Sciences* 32(1), 1-21.
Garde, P. (2004). *Le discours balkanique. Des mots et des hommes*, Fayard, Paris.
Gilliéron, J. et E. Edmont (1902-1910). *Atlas linguistique de la France*, Champion, Paris.
Goebl, H. (2002). "Analyse dialectométrique des structures de profondeur de l'ALF", *Revue de linguistique romane* 66(261-262), 5-63.
Hansen, J.H.L., U. Yapanel, R. Huang et A. Ikeno (2004). "Dialect analysis and modelling for automatic classification", *INTERSPEECH / ICSLP*, Jeju, 1569-1572.
Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenstein distance*, Thèse de doctorat, Rijksuniversiteit, Groningen.
Hintze, M.-A., T. Pooley et Judge, A. (eds) (2000). *French accents: phonological and sociolinguistic perspectives*, AFLS/CiLT, London.
Labov, W. (1972). *Sociolinguistic patterns*, University of Pennsylvania Press, Philadelphia.
Léon, P. (1993). *Précis de phonostylistique. Parole et expressivité*, Fernand Nathan, Paris.
Léonard, J.L. (1991). *Variation dialectale et microcosme anthropologique : l'île de Noirmoutier*, Thèse de doctorat, Université de Provence.
Martinet, A. (1969). "C'est jeuli le Mareuc", *Le français sans fard*, PUF, Paris.
Manning, C.D. et H. Schütze (1999). *Foundations of statistical natural language processing*, The MIT Press, Cambridge.
Preston, D.R. (1989). *Perceptual dialectology*, Foris Publications, Dordrecht.
De Saussure, F. (1915). *Cours de linguistique générale*, Payot, Paris.
Séguy, J. (1973). "La dialectométrie dans l'Atlas linguistique de la Gascogne", *Revue de linguistique roman,* 37, 1-24.
ten Bosch, L. (2000). "ASR, dialects and acoustic/phonological distances", *ICSLP*, Beijing, 1009-1012.
van Bezooijen, R. et C. Gooskens (1999). "Identification of Language Varieties. Contribution of Different Linguistic Levels", *Journal of Language and Social Psychology* 18(1), 31-48.
Vieru-Dimulescu, B. et P. Boula de Mareüil (2004). "Contribution de la prosodie à la perception de l'*accent* étranger : une étude à base de parole naturelle italienne/ espagnole modifiée", *Colloque MIDL*, Paris, 139-144.
Walter, H. (1982). *Enquête phonologique et variétés régionales du français*, Presses Universitaires de France, Paris.
Wells, C. (1982). *Accents of English*, Cambridge University Press, Cambridge.
Yan, Q. et S. Vaseghi (2002). "A Comparative Analysis of UK and US English Accents in Recognition and Synthesis", *IEEE ICASSP*, Orlando, 413-416.

Caractéristiques tonales du parler de la région marseillaise : approche globale

Annelise COQUILLON[1]

1. Introduction

La recherche sur les variétés régionales foisonne de travaux sur la description du matériau verbal (lexical, phonématique, syntaxique, etc.), négligeant trop souvent la composante prosodique. Nous savons, à l'instar de Léon (1971), que la prosodie véhicule à la fois des fonctions linguistiques, paralinguistiques et extralinguistiques. Au delà de sa fonction impressive, intentionnelle et consciente, il semble important de considérer sa fonction identificatrice (extralinguistique), qui permet de caractériser le sujet parlant, généralement à son insu. Nous savons par ailleurs que la prosodie est une composante essentielle dans la caractérisation linguistique des langues naturelles et qu'elle contribue largement à différencier les langues entre elles, ainsi que les différentes variétés régionales, sociolectales et situationnelles d'une même langue.

Nous introduirons en premier lieu dans cette étude le corpus constitué dans le cadre du projet PFC (*Phonologie du Français Contemporain, usages, variétés, structure* ; Durand & Lyche, 2003). Nous procèderons ensuite à un tour d'horizon des travaux portant sur la caractérisation du parler de la région marseillaise, tant d'un point de vue segmental que prosodique.

Nous présenterons ensuite les premiers résultats de nos investigations sur la caractérisation prosodique, au niveau intonatif, d'une partie du corpus PFC Aix-Marseille. L'analyse porte sur une étude du registre tonal d'un point de vue diachronique (sur six locuteurs : deux pour chaque génération). Nous avons restreint notre domaine d'analyse aux passages transcrits et codés selon le protocole PFC pour la conversation guidée, cette dernière nous semblant idéale pour un traitement prosodique, en cela qu'elle représente de la parole spontanée dans une même situation de communication pour tous les informateurs.

2. Le point d'enquête Aix-Marseille

Le point d'enquête Aix-Marseille porte sur 8 locuteurs d'une même famille à travers trois générations, et qui comprend 3 femmes et 5 hommes. Il s'agit d'une famille originaire de Septèmes-les-Vallons (commune de Marseille) dont certains membres ont emménagé dans des villes avoisinantes (Aix-en-Provence, à 20 km) ou dans le centre de Marseille. Tous n'ont que peu voyagé, et lorsque c'est le cas, toujours dans un cadre de séjours touristiques (inférieurs à 1 mois).

Le corpus comprend ainsi trois tranches d'âge et les locuteurs sont issus de milieux socio-économiques moyens ou ouvriers. Le tableau 1 ci-dessous présente brièvement les membres de cette famille.

[1] Laboratoire Parole et Langage, Université de Provence - Aix-en-Provence.
annelisecoquillon@lpl.univ-aix.fr

Tableau 1. *Code, sexe, âge et profession des locuteurs du corpus Aix-Marseille*

1e génération		2e génération		3e génération	
AA1	f. 82 ans Secrétaire retraitée	RP2	h. 45 ans Chef cuisinier restauration collective		h. 27 ans Technicien vacataire sécurité environnement
RP1	h. 82 ans Ouvrier d'usine retraité	MA1	f. 53 ans Infirmière	FA1	h. 31 ans Responsable d'une boite de transport / livraisons
		JC1	f. 57 ans Agent administratif		
			h. 58 ans Informaticien		

Afin d'équilibrer les locuteurs dans cette étude, nous avons sélectionné 2 locuteurs dans chaque tranche d'âge : les deux disponibles pour les 1e et 3e générations, ainsi que RP2 et PA1 pour la 2e.

3. Quelques caractéristiques du parler de la région marseillaise

Nous exposerons brièvement dans ce chapitre les caractéristiques parmi les plus remarquables du français parlé dans la région marseillaise telles qu'elles ont été décrites dans la littérature. Les marques régionales de cette variété de français se retrouvent à plusieurs niveaux : lexical, syntaxique, phonologique et phonétique. Nous ne préciserons pas ici les particularités lexicales et syntaxiques, trop nombreuses, mais nous noterons néanmoins qu'elles sont pour la plupart issues du provençal. L'influence du substrat provençal sur le français de Provence, largement étudiée (notamment aux niveaux lexical, morphologique et phonologique), n'est plus aujourd'hui à démontrer. Nous introduirons alors quelques caractéristiques phonologiques et phonétiques du parler de la région marseillaise, sachant qu'il s'agit de caractéristiques générales.

Toute variété dialectale présente une grande variabilité tant au niveau social (sociolectes), stylistique et situationnel (registres de langue), qu'individuel (idiolecte). L'on assiste alors à une gradation des usages ("degré d'accent" régional) qui va du plus dialectalisé au plus neutre, et qui s'écarte plus ou moins de la norme même de la variété régionale. Il apparaît dans de nombreuses études, dont plusieurs réalisées dans le cadre de PFC (voir les différents *Bulletins PFC*), une tendance générale de neutralisation du degré d'accent régional, notamment chez les sujets les plus jeunes.

Les caractéristiques que nous introduirons ici seront donc générales, et pourront être confirmées ou infirmées par les points d'enquêtes de PFC.

3.1. Le système phonologique

Entre le parler de la région marseillaise et un français plus neutre régionalement, les différences phonologiques résident principalement au niveau des systèmes vocaliques. D'une manière générale, la durée ne joue aucun rôle phonologique (pas de distinction voyelle longue/brève) et les voyelles ouvertes, bien que présentes au niveau phonétique, n'entrent pas dans le système phonologique du français de Provence qui se trouve ainsi "réduit" par rapport à celui du français général. Notons que ceci vaut pour le français méridional dans son ensemble.

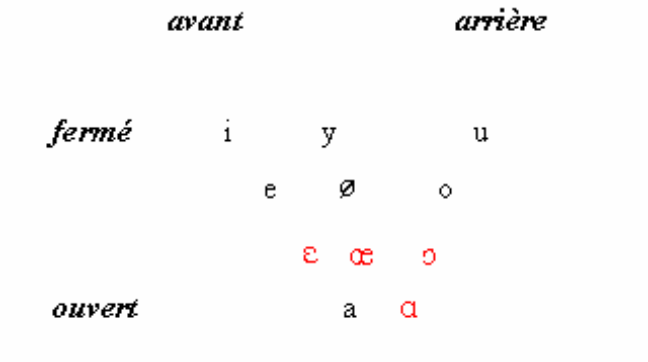


Figure 1. ***Trapèze vocalique des français de Provence (en noir) et de référence (gris), API***

À l'exception du /A/, tous les phonèmes vocaliques du français général se retrouvent au niveau phonétique en français méridional au niveau phonétique, mais pas en tant que phonème : /œ/ ; /ɛ/ et /ɔ/ sont respectivement des variantes combinatoires ou allophones des phonèmes /ø/ ; /e/ et /o/. La réalisation de surface de ces phonèmes dépend du contexte et par conséquent de règles phonologiques spécifiques ou "loi de position" (Durand 1976). Cette loi précise que la réalisation d'un segment vocalique sera ouverte dans le contexte d'un segment dépendant à sa droite, et fermé dans le cas inverse. Autrement dit, les voyelles fermées ne sont réalisées qu'en syllabe ouverte (lorsqu'il n'y a pas de consonne entre la voyelle et la frontière de la syllabe, cette situation incluant bien entendu les frontières de mot ou de phrase) et les voyelles ouvertes n'apparaissent qu'en syllabe fermée ou lorsque la syllabe qui suit contient un schwa (segment consonantique ou syllabe atone dépendante à la droite du segment).

Indépendamment des différences du système vocalique exposées ci-dessus, le méridional présente également des caractéristiques propres quant à la réalisation de certains de ses phonèmes. Nous présenterons ici les principales.

3.1.1. Les voyelles nasales

L'articulation des voyelles nasales est une des principales caractéristiques du français méridional. Ces segments sont généralement plus longs que ceux d'autres variétés de français, du fait qu'elles sont composées de plusieurs périodes : une partie orale, une partie de nasalisation (transitoire) et un segment consonantique nasal (Clairet 1997). Ce segment est variable en fonction du contexte, généralement homorganique (même lieu d'articulation) à la consonne subséquente : [m] après une consonne labiale, [n] devant une labiodentale, [ɱ] devant une fricative et [ŋ], [ɲ] ou [ɴ] devant une vélaire ou une pause.

La partie orale conserve le timbre de la voyelle ouverte correspondant à la nasale : [ɔ] devant [ɔ̃] ; [ɛ] devant [ɛ̃] ; [ɑ] devant [ɑ̃)] et [œ] devant [œ̃]. Un item comme *bon* en position finale sera donc généralement prononcé [bɔɔ̃ŋ]. Le fait qu'elles correspondent certainement aux voyelles orales peut avoir participé au maintient de la nasale /œ̃/, comme dans *un* ou *brun*, par opposition à /ɛ̃/ (*brin*), là où le français de référence tend à prononcer [ɛ̃] dans les deux cas.

En français méridional, il n'y aurait pas à proprement parler de voyelles nasales mais plutôt des voyelles orales suivies d'un appendice consonantique nasal (Durand, 1988), le système phonologique du provençal ne comportant pas de voyelles nasales.

3.1.2. Les voyelles orales

Elles n'ont encore que peu été étudiées. Outre le fait, comme nous l'avons vu plus haut, que le /A/ ne soit jamais réalisé et qu'il n'existe aucune distinction phonologique entre les voyelles ouvertes / fermées, la durée ne joue aucun rôle phonologique : il n'existe pas de distinction voyelle longue / brève.

Au niveau acoustique, il apparaît que les voyelles orales sont pertinentes dans l'identification des locuteurs méridionaux, comme l'a montré Chaumette (1997). Cette étude a également révélé un phénomène de diphtongaison des phonèmes [o] et [œ], caractérisé acoustiquement par un accroissement du premier formant, qui ne se retrouve pas dans un français "standard". D'autre part, il semble que les réalisations [ɛ] méridionales et non méridionales s'avèrent assez différentes.

Pour ce qui est du schwa, dont le maintien (notamment en finale) est une des caractéristiques du français méridional, nous le décrirons plus en détails dans le paragraphe suivant, étant donné son statut particulier ainsi que sa nature complexe.

3.1.3. Le schwa

Une des caractéristiques les plus prégnantes du français méridional est le maintient de la réalisation du schwa, notamment en position finale de mot. Selon Watbled (1995), le schwa est le segment correspondant à tout "e" graphique qui n'est pas susceptible d'être interprété comme /e/ ou /ɛ/. Ajoutons cependant qu'il peut apparaître à l'oral alors qu'il n'est pas présent dans la graphie, bien que cela semble relativement rare, surtout après une voyelle. Représenté par /ə/, l'ensemble de ses réalisations se répartit entre [ə], [ø], [a] et [ɔ].

Le schwa, appelé communément "e muet", a pour caractéristique principale d'être optionnel. Sa prononciation n'est en effet pas obligatoire ni fonction de règles phonologiques et sa présence n'apporte en français général aucun indice lexical. Notons toutefois qu'en français méridional, le schwa joue un rôle morphologique important, particulièrement comme marqueur du genre féminin, comme pour distinguer *tel* /tɛl/ de *telle* /tɛlə/, etc. (Durand & al., 1987) ou comme élément distinctif entre des mots tels que *mal* /mal/ et *malle* /malə/, qui ne sont pas des paires minimales en français.

Ce phonème peut ainsi être délimité par un critère de prononciabilité dont le maximum de réalisations effectives est tout près d'être atteint dans le parler méridional.

Une autre de ses caractéristiques est qu'il n'est pas accentuable et entraîne ainsi la réalisation de paroxytons, rares en français (en exceptant l'emphase), langue à accent fixe et à droite. Le schwa, lorsqu'il est effectivement réalisé, introduit une syllabe non accentuable et dépendante phonologiquement de la syllabe tonique qui précède, formant ainsi une unité supra syllabique : le pied prosodique, représentant l'unique cas en français où le pied sera composé de plus d'une syllabe. Le groupement de syllabes à l'intérieur du pied prosodique implique des relations spéciales entre elles, qu'elles n'ont pas avec les autres dans la phrase, comme la distribution temporelle (Coquillon et al. 2000).

3.1.4. Les consonnes

Elles ne semblent pas avoir fait l'objet d'études approfondies à ce jour. Toutefois, l'on relève dans l'article de Carton et al (1983) que les consonnes dites allongeantes du français ne remplissent pas ici leur fonction, notamment lorsqu'elles sont suivies de la voyelle /ə/. Seul le R est allongeant, mais uniquement lorsqu'il est sonore et qu'il n'est pas suivi de /ə/. D'autre part, les groupes consonantiques sont parfois simplifiés par la chute de la première consonne du groupe : *exemple* prononcé /ɛzɑ̃plə/.

3.2. L'accent tonique et l'intonation

Précisons que la littérature sur ces questions est particulièrement restreinte. Comme nous l'avons vu plus haut, une des particularités de l'accent tonique en méridional est qu'il peut tomber sur l'avant-dernière syllabe, notamment lors de la réalisation de schwas finaux. C'est également le cas lors de la prononciation de certains mots issus du substrat provençal tels que *aïoli* : /aj'ɔli/ et non /ajɔ'li/. Notons que la voyelle pénultième s'ouvre alors, selon la loi de position exposée plus haut (la dernière syllabe se retrouve en position atone et devient dépendante de la pénultième).

Ceci peut entraîner des schémas mélodiques particuliers au français provençal où, dans les mouvements ascendants, la courbe de f0 peut soit se terminer en forme de plateau (les deux dernières syllabes sont sur la même hauteur) ou encore le pic de f0 peut être réalisé tardivement (sur la syllabe post-tonique) (Carton et al., 1983), mais le /ə/ est moins intense. ; Coquillon, 2005). Dans les mouvements descendants, le schwa final d'unité intonative atteint généralement un niveau infra-bas dans le registre du locuteur ce qui entraîne une rupture dans la continuité de la courbe de f0, dans la mesure où la chute y est plus abrupte (empan mélodique et dynamique tonale plus marqués) (Coquillon, 2005).

D'autre part, nous avons mis en évidence (Coquillon, 2005) un patron mélodique particulier au parler de la région marseillaise : le contour en forme de chapeau mou.

4. Analyse prosodique : le registre tonal

Le paramètre tonal, souvent représenté au niveau acoustique par les modulations de la fréquence fondamentale (f0), est l'un des paramètres les plus utilisés dans les études sur la variation dialectale, que ce soit en ce qui concerne le registre tonal, les configurations tonales ou encore les phénomènes d'alignement. Pour cette raison, nous avons choisi d'analyser ce paramètre dans le corpus PFC : une approche globale nous permet d'évaluer le registre tonal utilisé par les locuteurs et ce, dans une perspective de comparaison diachronique (entre les trois générations du corpus).

Nous comparerons ainsi le registre tonal utilisé par les différentes générations de locuteurs de la région marseillaise. Ce paramètre s'est en effet révélé porteur d'information dialectale par de nombreuses études, telles que celles de Ménard et al. 1999 ; Paboudjian 2003 ; etc. Nous avons par ailleurs montré (Coquillon 2004, 2005) que les locuteurs de la région marseillaise présentent généralement un niveau tonal plus bas que les locuteurs non-méridionaux, ainsi qu'une tendance à utiliser une étendue tonale significativement plus importante. Ces résultats ont été validés statistiquement.

L'hypothèse de travail est ici que les diverses générations de locuteurs du corpus présentent des tendances différentes dans leur utilisation du registre tonal, ce paramètre se révélant alors distinctif. Certaines marques régionales tendant à se neutraliser chez les nouvelles générations de locuteurs, il est attendu que les plus âgés utilisent un registre tonal plus marqué régionalement que les plus jeunes.

4.1. Définition

La notion de registre employée ici au niveau tonal ("*pitch rang*" ou "*pitch register*") est à distinguer de celle renvoyant aux propriétés stylistiques du langage parlé, à savoir le registre langagier, ou niveau de langue, associé à la situation de communication (registres formel / informel ou soutenu / familier, etc.).

En prosodie, le registre tonal permet d'évaluer les caractéristiques de la f0 d'un locuteur à un niveau global ou, plus localement, d'un contour intonatif donné. Le registre est généralement défini par deux concepts distincts mais néanmoins étroitement liés : le niveau ("*overall level*") et l'étendue ou amplitude ("*span*") tonals. Le premier caractérise la

hauteur à laquelle le locuteur place sa voix et le second détermine l'espace tonal ou empan (large ou étroit) des fréquences qu'il utilise (Patterson & Ladd 1999). D'une manière plus générique, le registre est parfois considéré comme l'empan mélodique relatif utilisé par un locuteur dans un énoncé ou un corpus donné, en référence à la tessiture, empan mélodique absolu qui définit l'étendue mélodique susceptible d'être parcourue par la voix d'un locuteur (Rossi 1999). Le locuteur utiliserait essentiellement, en conversation, la plage centrale (registre usuel) des fréquences de sa tessiture.

4.2. Méthodologie

Afin d'évaluer le registre tonal global utilisé par les différents locuteurs du corpus PFC, nous nous appuyons sur une représentation stylisée de la courbe de f0 en termes de cibles tonales selon l'approche MOMEL (Hirst, Di Cristo et Espesser 2000), que nous décrivons ci-dessous.

Étant donné que le corpus enregistré selon le protocole PFC présente sur la même piste sonore le(s) locuteur(s) et le(s) enquêteur(s), une première préparation du corpus est nécessaire. Il s'agit en effet de ne conserver pour l'analyse que les points cibles correspondant à la f0 du locuteur analysé. Pour cela, nous avons choisi de garder le fichier son intact et avons supprimé manuellement les points cibles correspondant aux interventions de l'enquêteur. Pour ce faire, nous avons utilisé, pour des raisons pratiques, le logiciel MES (Espesser 1996, environnement Unix), dans lequel la méthode MOMEL (MOdélisation de la MELodie), élaboré par Hirst, Di Cristo et Espesser (2000), est directement implémentée. Rappelons néanmoins que cette méthode est également disponible sous le logiciel PRAAT (Boersma et Weenink 1993-2003), utilisé dans le protocole PFC, grâce à un script élaboré par Cyril Auran[2].

4.2.1. L'approche MOMEL

Acoustiquement représentée par la courbe de fréquence fondamentale (f0), l'intonation est la combinaison d'une composante microprosodique, directement dépendante de la nature intrinsèque des phonèmes, non motivée linguistiquement, et d'une composante macro-prosodique, qui reflète les choix syntaxiques et pragmatiques du locuteur en termes de patrons intonatifs (Di Cristo & Hirst 1986). Afin de factoriser ces deux composantes et d'extraire ainsi la composante macroprosodique pertinente, l'algorithme MOMEL propose une stylisation de cette courbe de f0, par une fonction spline quadratique. Cette représentation de la f0 se situe à un niveau de représentation phonétique, première abstraction du niveau acoustique de surface. Il en résulte une séquence de points cibles correspondant aux changements significatifs de direction de la f0, interpolés par une courbe en arcs de parabole continue. En parole, la perception des variations de hauteur est en effet continue.

Les points cibles repérés par MOMEL sont définis par des valeurs en Hertz et leurs coordonnées temporelles correspondantes (Hz ; ms). Les points cibles peuvent alors être utilisés pour une reproduction par synthèse de la parole d'origine, quasiment sans perte d'information par rapport à la courbe de f0 originelle.

4.2.2. Méthode d'analyse

Étant donné qu'il n'est pas établi que la perception prosodique des variations dialectales (para et / ou extra linguistiques) se fait en termes de tons linguistiques, nous avons opté

[2] http://aune.lpl.univ-aix.fr:16080/~auran/ (sous : recherche/ressources)

pour une approche plus phonétique que phonologique, basée sur la répartition des points d'inflexions de la f0 dans l'espace tonal utilisé par chaque locuteur.

Pour le niveau tonal, nous présenterons ainsi les calculs sur l'ensemble des cibles du corpus, en Hz.

Pour l'étendue tonale, nous avons choisi d'isoler les valeurs extrêmes (points cibles inférieurs et supérieurs) en fonction de leur distribution en écarts inter-quantile. Les points cibles concernés ont été codés en points inférieurs (*pinf*) et supérieurs (*psup*). Nous avons ainsi été en mesure de valider ce codage directement sur la courbe modélisée. Il est ressorti de cette vérification que si le seuil de quantile est trop faible (5 et 10%), le codage semble omettre trop de valeurs de pics et de vallées et le nombre de mesures devient alors trop faible pour être validé statistiquement. À l'inverse, lorsque le seuil est trop élevé (20% et plus), le nombre de cibles considéré est trop important pour refléter le registre car l'on tend dans ce cas à comparer les distributions de quasiment toutes les valeurs autour de la moyenne (la valeur moyenne des écarts entre les points inférieurs et supérieurs tend alors vers 0). Nous avons donc principalement retenu pour les calculs du niveau et de l'étendue tonale un seuil de quantile à 15%.

Les valeurs des points cibles en Hz ont ensuite été normalisées en demi-tons, nous permettant notamment de comparer directement les voix de femmes et d'hommes constituant le corpus. Les fréquences ont ainsi été converties en demi-tons, à partir d'une valeur de référence, ce type de normalisation s'effectuant en termes d'écart entre deux valeurs. Cette valeur de référence correspond ici à la médiane des fréquences en Hertz utilisée par chaque locuteur de notre corpus, selon la formule suivante (Boersma & Weenink 1993-2003) :

$$(1)\ \text{Demi-ton} = 12 * \ln(f / f_{ref}) / \ln 2$$

où "*f*" est la fréquence en Hz à normaliser en demi-tons et "*f ref*" est la médiane en Hz de tous les points cibles d'un fichier son (par locuteur).

Enfin, les résultats ont été analysés à l'aide d'un modèle linéaire à effets mixtes[3]. Ce type de modèle est adapté aux plans d'expérience à mesures répétées, où les données sont structurées par groupes ou blocs, et donc caractérisées par la présence de corrélations entre observations à l'intérieur d'un même groupe. Dans notre cas, il s'agit de mesures répétées sur chacun des locuteurs. Nous avons considéré comme seuil critique la valeur de $p = 0,05$. Cette approche est facilement applicable à l'ensemble des corpus PFC, et se révèlerait particulièrement intéressante pour comparer différentes régions entre elles.

4.3. Résultats

Nous présentons ci-après les résultats de l'analyse sur le registre tonal des locuteurs sélectionnés dans le corpus PFC Aix-Marseille. Nous avons isolé le niveau tonal de l'étendue tonale.

4.3.1. Niveau tonal : caractéristiques tonales moyennes

Parmi les différents paramètres utilisés pour définir le niveau tonal, nous avons choisi de présenter dans ce paragraphe les mesures suivantes :

- La moyenne et la médiane des fréquences de tous les points cibles repérés par MOMEL, ainsi que les valeurs des points cibles supérieurs et inférieurs (max et min).

[3] Je tiens à remercier tout particulièrement Robert Espesser, avec qui nous avons mis en place la méthodologie ainsi que le modèle statistique pour cette étude.

Le tableau 2 ci-dessous introduit quelques caractéristiques du niveau tonal des locuteurs de notre corpus : la moyenne et la médiane des fréquences de tous les points cibles repérés par MOMEL, ainsi que les valeurs maximale et minimale de ces points, en Hz. Notons que la normalisation des données en demi-tons n'est pas possible pour ce paramètre. Cette échelle de valeurs se calcule en effet en terme d'écart entre deux fréquences, ce qui ne correspond pas ici au type de données. Les locuteurs sont présentés dans ce tableau par âge décroissant.

Tableau 2. *Caractéristiques tonales moyennes (Hz) pour chaque génération et chaque locuteur*

génération	locuteurs	moyenne		médiane		max	min
1e	AA1	172,18	190,12	169,25	187,32	334,72	54,41
	RP1		155,39		155,34	294,24	55,57
2e	PA1	98,07	104,71	91,66	97,94	215,18	63,1
	RP2		90,05		87,24	187,09	60,17
3e	FA1	103,56	103,92	101,65	103,23	167,67	54,72
	SA1		103,27		100,07	187,4	55,15

Nous pouvons observer dans ce tableau que les deux locuteurs de la 1e génération présentent des moyennes, médianes, et valeurs du point cible *maximum* supérieures aux autres générations. Des valeurs de f0 élevées étant caractéristiques de voix aiguës, cette différence pourrait être due au type de voix de ces deux locuteurs (facteur extra-linguistique). Rappelons que AA1 est une femme (la seule du corpus ici présenté), mais il reste que le locuteur RP1 semble plutôt présenter une voix grave. L'analyse d'un nombre plus important de locuteurs serait nécessaire pour confirmer ou infirmer cette tendance. Les données sur l'étendue tonale (voir infra) nous apporteront des indices supplémentaires quant au registre de ces locuteurs, grâce notamment à la normalisation des données en demi-tons. Il n'apparaît par ailleurs que peu de différences entre les locuteurs pour ce qui est de la valeur du point cible *minimum*.

L'analyse statistique par modèle linéaire à effet mixte indique globalement que le facteur "génération"[4] n'est pas significatif : $F(1,4) = 6,53089$; $p = 0,0629$. Ici, la variable dépendante testée est "valeurs des cibles en Hz", la variable indépendante étant "génération". L'effet aléatoire "locuteur" est modélisé par une constante.

4.3.2. Étendue tonale

À partir de la méthodologie que nous avons adoptée ici, nous retiendrons pour le paramètre d'étendue tonale les mesures d'écarts de fréquence observés entre les *extrema* de f0 (15% des fréquences inférieures et supérieures) et la médiane, valeur de référence. Les données sont ainsi introduites en demi-tons.

[4] Une analyse préalable avec "génération" en facteur ordonné a montré qu'il était plus convenable, dans un tel modèle, de considérer ce facteur en tant que variable numérique.

Tableau 3. *Écart à la médiane des extrema en demi-tons par génération et par locuteur : moyenne générale (valeurs absolues), moyennes des points inférieurs et supérieurs*

génération	locuteurs	Moyenne (valeurs absolues)		pinf		psup	
1e	AA1	6,751	7,479	-8,204	-8,933	5,298	6,025
	RP1		6,07		-7,521		4,619
2e	PA1	4,242	4,779	-2,782	-3,012	5,702	6,545
	RP2		3,59		-2,502		4,678
3e	FA1	3,311	2,891	-3,023	-3,087	3,599	2,695
	SA1		3,648		-2,971		4,325

D'une manière générale, nous constatons que l'écart moyen à la médiane (en valeurs absolues) se réduit d'une génération à l'autre, dans l'ordre chronologique ($1^e > 2^e > 3^e$, avec un écart de 1,75 demi-ton en moyenne d'une génération à l'autre). L'analyse statistique semble indiquer que cette tendance est significative[5] : $F(1,4)= 19,1482$; $p= 0,012$.

Cette différence entre générations est de nouveau plus marquée entre la 1^e génération et les deux autres, qu'entre la 2^e et la 3^e.

Les valeurs des points cibles inférieurs et supérieurs confirment ces tendances. Nous pouvons voir que la 1^e génération semble utiliser un registre sensiblement plus étendu que les deux autres générations. En effet, la différence entre les moyennes des "*pinf*" et des "*psup*" est de 13,5 demi-tons, alors qu'elle est seulement de 8,5 pour la deuxième génération et de 6,6 pour la troisième. L'analyse statistique indique un fort effet d'interaction entre "génération" et "type inférieur et supérieur" : $F(1,1296) = 183,8447$; $p = <0,0001$. Le sens de variation d'une donnée à l'autre (pentes de *pinf* et *psup* par génération) est par ailleurs sensiblement différent.

Le détail par locuteur montre que cette tendance est peu, voire pas marquée pour les deux dernières générations, où SA1 (3^e génération) présente des moyennes (pour les 3 types de données) équivalentes à celles de RP2 (2^e génération), comme nous pouvons plus facilement l'observer sur la figure 2 ci-dessous. L'écart est donc plus net entre la première génération et les deux autres.

[5] Ce résultat a été obtenu sur l'inverse des demi-tons, la distribution des résidus étant plus proche d'une distribution normale, comme attendu dans ce type de modèle.

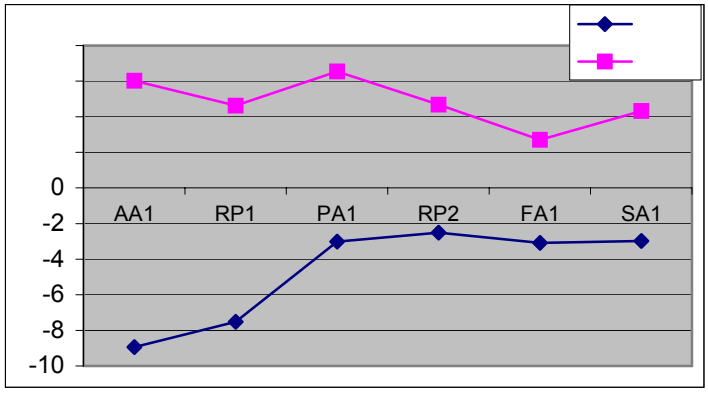


Figure 2. *Écart à la médiane des extrema (inférieurs et supérieurs) en demi-tons par locuteur*

Nous pouvons observer sur cette figure que la différence entre générations dans l'utilisation de l'étendue tonale est plus marquée par les valeurs des points *minima* que des *maxima*. Nous pouvons supposer que l'on assiste à une neutralisation chez les plus jeunes générations de l'étendue tonale en tant que marque régionale de l'accent de la région marseillaise. Il semble alors cohérent que ce soit les valeurs minimales qui soient les plus touchées par cette neutralisation : c'est le trait le plus marqué qui est affecté par cette variation.

5. Conclusion et perspectives

Dans cette étude, nous avons tenté d'évaluer le registre tonal de chaque génération de locuteurs d'une partie du corpus PFC Aix-Marseille. Il est apparu que ce paramètre, et notamment l'étendue tonale, semble se révéler discriminant entre les différentes générations de locuteurs du corpus. Cette tendance est plus marquée entre la première génération et les deux suivantes, et concerne tout particulièrement les points cibles inférieurs. Une étendue tonale importante, qui, comme nous l'avons montré (Coquillon 2004, 2005), est caractéristique du parler de la région marseillaise, semble être un trait que les générations plus âgées ont conservé. Cette marque régionale paraît en effet se neutraliser chez les jeunes générations, bien qu'elle reste néanmoins un des traits distinctifs par rapport à d'autres variétés de français. Les locuteurs utilisés dans l'étude de Coquillon (2005) appartiennent en effet à la tranche d'âge des 25-35 ans.

Cette première approche serait à confirmer sur un corpus plus étendu, et il serait également intéressant d'appliquer cette analyse à une comparaison entre différentes variétés de français, grâce aux nombreuses enquêtes réalisées dans le cadre de PFC.

En outre, une analyse auditive sur le degré d'accent des locuteurs serait à mettre en parallèle avec ce paramètre de registre tonal. Elle pourrait en effet éclairer les résultats à la lumière d'un indice important dans la variation régionale.

Références

Boersma, P. & D. Weenink (1993-2003). *PRAAT, a system for doing phonetics by computer*. University of Amsterdam (http://www.praat.org).

Carton, F., M. Rossi, D. Autesserre et P.Léon (1983). *Les accents des Français*. Coll. De bouche à oreille. Paris : Hachette.

Clairet, S. (1997). *Caractérisation des voyelles nasales en français méridional* : *essai méthodologique pour une analyse aérodynamique et acoustique.* Mémoire de DEA en phonétique (non publié), Université de Provence.

Chaumette, C. (1997). *Étude acoustique des voyelles orales en français méridional.* Mémoire de DEA en phonétique (non publié), Université de Provence.

Coquillon, A. (2005). *Caractérisation prosodique du parler de la région marseillaise*. Thèse de doctorat en phonétique, Université de Provence, 392 p.

Coquillon, A. (2004). "Contribution de la prosodie à l'identification du parler de la région marseillaise". *Actes du colloque 'Identification des langues et des variétés dialectales par les humains et par les machines' (MIDL 2004)*, 29-30 novembre 2004, Paris, 89-90.

Coquillon, A., A. Di Cristo et M. Pitermann (2000). "Marseillais et Toulousains gèrent-ils différemment leurs pieds ? Caractéristiques prosodiques du schwa dans les parlers méridionaux". *Actes des XXIII*^{èmes} *'Journées d'Étude sur la Parole' (JEP)*, 19-23 juin, Aussois, pp. 89-92.

Di Cristo, A. et D.J. Hirst (1986). "Modelling French micromelody: analysis and synthesis". *Phonetica* 43, 1/3, 11-30.

Durand, J. (1976). "Generative phonology, dependency phonology and southern French". *Lingua e stile*, Anno XI, No 1, 3-23.

Durand, J. (1988). "Phénomènes de nasalité en français du midi: phonologie de dépendance et sous-spécification". *Nouvelles Phonologies, Recherches Linguistiques de Vincennes* 17, Paris VII, 29-54.

Durand, J., C. Slater et H. Wise (1987). "*Observations on schwa in southern French*". Linguistics 25/5, pp.983-1004.

Durand, J. et Ch. Lyche (2003). "Le projet 'Phonologie du français contemporain et sa méthodologie'". In E. Delais-Roussarie et J. Durand (éds) *Corpus et variation en phonologie du français. Méthodes et analyses*, Presses Universitaires du Mirail, Toulouse, 213-278.

Espesser, R. (1996). "Mes: un environnement de traitement du signal". *Actes des XXIèmes 'Journées d'Étude sur la Parole' (JEP)*, 10-14 juin, Avignon, 447.

Hirst, D.J., A. Di Cristo et R. Espesser (2000). "Levels of representation and levels of analysis for the description of intonation systems". In M. Horne (eds), *Prosody : Theory and Experiment.* Dordrecht : Kluwer Academic Publishers, 51-88.

Léon, P. (1971). *Essais de phonostylistique*. Studia Phonetica n°4. Montréal, Canada : Marcel Didier.

Ménard, L., C. Ouellon et J. Dolbec (1999). "Prosodic markers of regional group membership: The case of the French of Quebec *versus* France". *Actes du XIV*^{ème} *'International Congress of Phonetic Sciences' (ICPhS)*, 1-7 août, San Francisco, USA, 1601-1604.

Paboudjian, C. (2003). "Intonation patterns as a mark of sociocultural identity. Observations from African-American English". *Actes du XV*^{ème} *'International Congress of Phonetic Sciences' (ICPhS)*, 3-9 août, Barcelone, Espagne, 1203-1212.

Patterson, D. et D.R. Ladd (1999). "Pitch range modeling: linguistic dimensions of variation" *Actes du XIVe International Congress of Phonetic Sciences*, San Francisco, USA, 1169-1172.

Rossi, M. (1999). *L'intonation, le système du français. Description et modélisation.* Paris : Ophrys.

Watbled, J.P. (1995). “Segmental and suprasegmental structure in southern French”. In J. C. Smith et M. Maiden (eds), *Linguistic Theory and the Romance Languages*, Amsterdam & Philadelphie, Benjamins, Current Issues in Linguistic Theory 122, 181-200.

Aspects de la durée vocalique dans le vaudois

Helene N. ANDREASSEN[1]

1. Introduction

La durée vocalique joue un rôle plus décisif dans le système phonologique du français suisse que dans celui du français de référence (désormais FR, *cf.* Morin 2000). Tout comme dans le FR, le suisse romand (désormais SR) est soumis à l'allongement conditionné soit par la bimoraïcité inhérente[2], soit par les consonnes dites allongeantes. Cependant, le phénomène saillant qui distingue le système vocalique du SR de celui du FR est la durée vocalique contrastive, où seule la quantité sert à différencier deux formes dans l'output. Dans ce travail, c'est notamment ce dernier type de durée qui va nous occuper, plus précisément, la façon dont ce phénomène est représenté dans la grammaire de la variété suisse du canton de Vaud.

Après avoir exposé notre corpus, nous présenterons dans une deuxième partie deux études antérieures portant sur le système vocalique romand. Nous considèrerons ensuite dans quelle mesure leurs observations sont conformes à nos propres données, et nous présenterons des résultats permettant d'identifier le statut de la durée contrastive dans le vaudois contemporain. Avant de conclure, nous discuterons enfin dans une dernière partie la façon dont la durée contrastive peut être intégrée dans la grammaire, une discussion menée dans le cadre de la théorie d'optimalité (Prince & Smolensky 1993). Nous nous concentrerons ici sur la durée observée dans des contextes morphologiquement féminins.

2. L'enquête PFC

2.1. "L'accent suisse romand"

Comment se situe "l'accent suisse" par rapport au FR ? En effet, quelques archaïsmes mis à part[3], le parler romand ne se distingue pas, d'après Knecht & Rubattel (1984), phonologiquement du FR. Selon eux, à part le lexique, c'est sur le plan de la réalisation phonétique que les divergences s'identifient[4]. Au plan phonologique, les aspects vocaliques qui caractérisent le suisse romand actuel ne sont que des réminiscences de la variété française répandue en Suisse romande. Il a été proposé que le français est devenu langue de conversation spontanée d'abord à Genève et à Neuchâtel, changement qui aurait eu lieu dans la deuxième moitié du XVII[e] siècle (Knecht 1985 : 145). A ce stade encore, la longueur fait partie intégrante du français parlé en France (et, partant, du français

[1] Universitet i Tromsø/CASTL, N-9037 Tromsø, Norvège | *helene.nordgard.andreassen@hum.uit.no*

[2] "The bimoraic vowels (tense [o, ø] and the nasal vowels) are long before all final consonants." (Féry 2001:16).

[3] Ces archaïsmes sont, entre autres, le maintien du contraste /œ ɛ̃/, l'opposition /o ɔ/ en position finale, ainsi que la durée vocalique contrastive.

[4] "Si la phonologie du français régional ne présente aucune particularité autre que le maintien de certains archaïsmes, la réalisation phonétique est, avec le lexique, le trait le plus saillant du français de la Suisse romande [...] a) **une résistance à l'accentuation de la dernière syllabe d'un syntagme (oxytonie)**, dont on peut vraisemblablement attribuer l'origine au substrat francoprovençal. [...] b) la variété de Suisse romande ne semble pas connaître de style *allegro* : non seulement **le débit est généralement assez lent**, mais un débit relativement rapide n'entraîne aucune modification phonologique (sinon la chute d'un plus grand nombre de *e* caducs). " (Knecht & Rubattel 1984 : 142-143, mis en italiques par les auteurs ; mis en gras par nous).

progressivement parlé en Suisse), et l'on constate également que ce trait linguistique disparaît petit à petit dans la plupart des variétés en France (*cf.* Girard & Lyche 2003 et Montreuil 2003 sur le maintien de la durée contrastive dans les variétés normandes).[5]

2.2. Les données

	sexe	*âge*	*lieu de naissance*	*domicile actuel*	*profession*
Loc1	F	30	Lausanne	Prangins	employée de bureau/mère au foyer
Loc2	F	31	Lausanne	Prangins	secrétaire dans des bureaux
Loc3	F	46	Nyon	Begnins	secrétaire municipale adjointe
Loc4	F	52	Nyon	Nyon	secrétaire
Loc5	F	65	Founex	Gland	mère au foyer
Loc6	M	31	Nyon	Prangins	ébéniste
Loc7	M	32	Nyon	Nyon	secrétaire municipal
Loc8	M	32	Prangins	Prangins	plâtrier peintre
Loc9	M	45	Nyon	Gland	employé de commerce
Loc10	M	56	Bex	Nyon	ingénieur chimiste
Loc11	M	59	Begnins	Nyon	fonctionnaire de police, retraité
Loc12	M	70	Vallorbe	Gland	docteur en science, retraité

Figure 1. *Le corpus*

Le corpus sur lequel ce travail se fonde a été construit dans le cadre du projet PFC[6]. Il consiste en l'interview de douze locuteurs, sept hommes et cinq femmes, dont l'âge varie de trente à soixante-dix ans. Ils sont tous originaires du canton de Vaud, domiciliés à présent à Nyon et dans ses environs (Begnins, Prangins et Gland), *cf.* la figure 2.

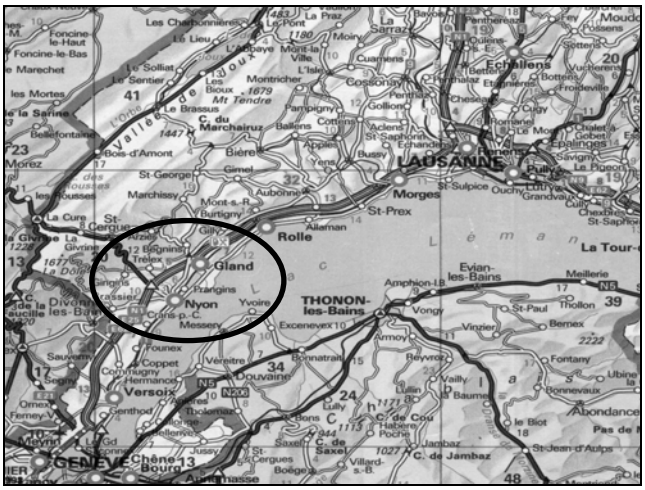


Figure 2. *Le point d'enquête*[7]

[5] "C'est vers le début du 19e s. aussi que disparaît la dernière trace de l'ancien *-e* féminin. Jusque-là les gens cultivés avaient encore distingué *aimée* de *aimé* par un allongement sensible de l'*é*." (von Wartburg 1993 : 229).

[6] *Phonologie du Français Contemporain : usages, variétés et structure*, *cf.* www.pfc-projet.net. Nous renvoyons à Andreassen (2003) et (2004) pour un traitement plus exhaustif de la phonologie du vaudois.

[7] Extrait de la *Carte touristique de la Suisse* publiée par la Société de Banque Suisse.

3. La durée vocalique

3.1. Études antérieures sur le système vocalique romand

À notre connaissance, peu de travaux ont envisagé la durée contrastive du SR. Cependant, en s'appuyant sur la méthodologie de Martinet (1945), les études de Métral (1977) et de Schoch et al (1980) présentent des résultats sur le jugement que divers locuteurs portent sur leur propre prononciation (et non pas sur leur usage réel), incluant dans le questionnaire des paires telles que armé-armée et nu-nue. Eu égard la grande diversité vocalique observée à travers les locuteurs interrogés, ils concluent qu'il est impossible de construire un système vocalique unique pour le suisse romand. Il en découle ainsi qu'il n'est pas certain qu'il existe une seule variété vaudoise non plus.

Ces deux études confirment le maintien de la durée contrastive, soulignant toutefois en même temps que cette distinction ne se maintient pas de manière stable à travers la Suisse romande. Comme le note Métral (1977 : 167) : "Si nous nous fondons sur les avis majoritaires, nous obtiendrons les caractéristiques de la *koinè*, mais comme ces dernières ne correspondent à la prononciation de tous les individus d'aucun canton, la *koinè* reste quelque chose d'hypothétique." Il continue cependant (p. 169) en disant qu'en ce qui concerne le canton de Vaud, "[la prononciation] s'accorde en général avec les caractéristiques de la *koinè*." Métral élabore alors un système vocalique pour le romand dans lequel il identifie une distinction de durée sur tous les niveaux dans l'espace vocalique.

	position finale			position intérieure		
-	i/iː	ü/üː	u/uː	i/iː	ü/üː	u/uː
-	e/eː	ö/öː	o/–	–/–	– / öː	–/oː
-	ɛ/ɛː	–/–	ɔ/–	ɛ/ɛː	œ / –	ɔ/–
-	a/–		ɑ/ɑː	a/–		ɑ/ɑː

Figure 3. *Le système vocalique romand sous l'accent*[8]

S'étant servis d'un type de questionnaire similaire, Schoch et al. (1980) confirment les résultats de Métral. Une différence d'importance entre les deux enquêtes concerne leur choix de locuteurs. Le corpus de Métral comprend 400 instituteurs tandis que le corpus de Schoch et al. comprend 513 collégiens dont l'âge moyen est de 15 ans. L'identité des résultats démontre alors que la tranche d'âge à laquelle appartient le locuteur n'a aucun effet sur son comportement linguistique, cf. par exemple le taux d'identification de la présence d'un *e* féminin, qui s'avère important à travers les différents points d'enquête.

[8] Métral (1977 : 168).

Identité :		NE	GE	VS	FR	JU	VD	SR
crue/cru	oui	2	5	20	32	0	4	63
	non	48	54	33	19	56	**124**	334
Si crue ≠ cru :								
durée		46	49	25	16	53	**94**	283
autre chose		2	-	7	2	2	7	20
durée + fermeture		-	3	-	-	1	9	13
le ə		-	1	-	-	-	**14**[9]	15
Identité :		NE	GE	VS	FR	JU	VD	SR
vie/vit	oui	0	2	3	22	0	0	27
	non	50	57	52	29	56	**128**	372
Si vie ≠ vit :								
durée		46	42	37	25	53	**76**	279
autre chose		2	2	15	3	1	8	31
vie = fille		2	5	-	3	2	13	25
le ə		-	6	-	-	-	**22**	28
durée + autre chose		-	2	-	-	-	8	10
Identité :		NE	GE	VS	FR	JU	VD	SR
armé/armée	oui	6	5	15	35	18	2	81
	non	44	54	40	16	38	**126**	318
Si armé ≠ armée :								
durée		40	40	29	16	35	**89**	249
autre chose		1	5	9	3	2	9	29
"arméie"		3	7	-	-	-	**27**	37

Figure 4. *Les oppositions /y ~ yː/, /i ~ iː/ et /e ~ eː/ dans Métral (1977 : 172-174)*

Prononcez-vous de façon identique :	*Total Genève*			*Total Lausanne*			*Total Moudon*		
V ~ Vː	OUI%	NON%	NR%	OUI%	NON%	NR%	OUI%	NON%	NR%
bout – boue	31.0	**66.2**	2.8	13.1	**86.9**		22.5	**76.1**	1.4
nu – nue	21.8	**73.9**	4.2	10.8	**88.5**	0.8	22.5	**77.5**	
lit – lie	14.8	**81.0**	4.2	4.6	**95.4**		9.9	**87.3**	2.8
armé – armée	10.6	**81.7**	7.8	9.2	**88.5**	2.3	9.9	**87.3**	2.8

Figure 5. *L'identification de la durée contrastive dans Schoch et al (1980:11)*

Considérons ensuite les voyelles de type /A/. Schoch et al. montrent que *mal* et *mâle* sont globalement distingués, et il faut probablement en rechercher la cause dans une différence de qualité et une différence de quantité saillantes. La distinction se fait beaucoup moins entre *mâle* et *malle*, ce qui indique une identité plus forte entre ces deux dernières formes qu'entre *mal* et *mâle*.

[9] Le taux d'identification d'un schwa post-vocalique démontre probablement la forte influence de la graphie. Dans le corpus d'Andreassen (2003), le schwa ne se réalise pas en position post-vocalique dans le parler spontané.

Prononcez-vous de façon identique :	*Total Genève*			*Total Lausanne*			*Total Moudon*		
a ~ ɑ	OUI%	NON%	NR%	OUI%	NON%	NR%	OUI%	NON%	NR%
mal – mâle	12.7	**85.9**	1.4	4.6	**93.1**	2.3	9.9	**88.7**	1.4
mal – malle	33.8	**57.0**	9.2	18.5	**80.0**	1.5	21.1	**74.7**	4.2
malle – mâle	56.3	**43.0**	0.7	63.1	**35.4**	1.5	74.7	**23.9**	1.4
rat – ras	73.9	24.6	1.4	52.3	46.2	1.5	45.1	52.1	2.8
pâte - patte	12.7	81.7	5.6	3.9	95.4	0.8	9.9	88.7	1.4

Figure 6. *L'intuition du locuteur quant à sa propre prononciation de /A/ (Schoch et al 1980 : 6)*

Nous identifions alors, dérivée de ces formes, toute une échelle de distinction : *mal-mâle >> mal-malle >> malle-mâle*. Il semble donc fort possible que l'on ait affaire à trois phonèmes différents au niveau sous-jacent, servant à distinguer les formes en question.

3.2. Résultats de l'enquête PFC

La méthodologie de PFC s'oppose à celle de Métral et de Schoch et al. en ce que l'on ne pose aucune question au locuteur sur sa prononciation. Cette méthodologie met par contre l'accent sur l'observation qui prend en compte quatre registres de parole. Pour notre étude PFC, il serait fructueux d'examiner comment nos données se situent par rapport aux résultats présentés dans la section précédente. Considérons dans un premier lieu les résultats des listes de mots et du texte PFC.

Concernant les voyelles de type /A/, la distinction de timbre et de quantité se fait à cent pour cent pour *mal-mâle* (de même que pour *patte-pâte*). Lorsque l'on introduit une troisième forme *malle*, elle se distingue de *mal* et *mâle* chez quelques-uns de nos locuteurs. Ce fait est conforme aux observations de Schoch et al. (op.cit.), où un nombre considérable de locuteurs sentent une différence entre les trois formes. Nous transcrivons le statut phonémique du A de *malle* comme ci-dessous tout en suivant l'observation de Métral (1977 :168) : *un* A antérieur (bref) apparaît en position non finale, tandis que le A postérieur connaît deux variantes dans ce même environnement, /ɑ/ et /ɑː/.

(1) mal mal mɑːl *mal - malle - mâle*

Si la distinction entre *mal* et *malle* porte uniquement sur la qualité, et celle d'entre *malle* et *mâle* sur la quantité, nous pouvons, d'après les résultats obtenus (*cf.* figure 6), avancer l'hypothèse que le caractère qualitatif en général est au niveau perceptif plus saillant que le caractère quantitatif (cf. Miller & Grosjean 1997 sur le rôle que joue la durée contrastive dans la perception).

Trois locuteurs ont lu une liste de mots supplémentaire qui *a priori* permettrait de confirmer l'existence des contrastes quantitatifs. Nous observons qu'un seul locuteur fait la distinction entre *renne* et *reine*. L'opposition /i–iː/ est également instable, différemment des paires *bout-boue* et *voie-voix* qui présentent une distinction de durée stable.

	voix-voie	*noie*	*renne-reine*	*veule-veulent*	*bout-boue*	*pile-île*
Loc7	wa waː	nwaj	ʁɛn ʁɛn	vøːl vøːl	bu buː	pil iːl
Loc9	wa waː	nwaː	ʁɛn ʁɛn	vøːl vøːl	bu buː	piːl iːl
Loc12	wa waː	nwaː	ʁɛn ʁɛːn	vøːl vøːl	bu buː	pil il

Figure 7. *La durée contrastive chez trois locuteurs vaudois*

Voie et *boue* font partie du groupe "mots à *e* féminin", des mots qui dans la graphie se terminent par un *e* post-vocalique, et qui dans la production souvent s'identifient par une voyelle longue finale. La présence de l'allongement vocalique dans des formes comme [ɑ̃tameː] *entamée*$_{\text{FÉM}}$ et [ʒɔliː] *jolie*$_{\text{FEM}}$, par opposition à l'absence d'allongement dans [ɑ̃tame] *entamé*$_{\text{MASC}}$ et [ʒɔli] *joli*$_{\text{MASC}}$, où le contexte structural est identique, indique que l'allongement en question ne saurait être conditionné par une contrainte purement structurale. La distribution de la longueur vocalique en position finale est essentiellement restreinte à des contextes qui en théorie sont morphologiquement féminins[10], facilement identifiables dans la graphie, cf. Voillat (1971 : 219) : "L'orthographe peut encourager la persistance d'éléments morphologiques dont le français se passe : c'est le cas de la marque du féminin conservée dans des mots en *-ée*, en *-e.* " Une étude du lexique montre que la marque du féminin se maintient également avec d'autres voyelles.

(2) *nu – nue*
bout – boue
bleu – bleue

Au niveau phonologique, l'allongement final dans les contextes féminins tels que *jolie* et *nue* est historiquement considéré compensatoire en ce qu'il résulte de la réduction graduelle d'un schwa post-vocalique, cf. l'évolution de la variété bourguignonne décrite dans Morin (1994 : 144) :

(3)	OFr *cru*	[kry]	>	*cru*	[krў]
	OFr *crue*	['kryə]	>	*crue*	[kryː]
	OFr *creu*	[kre'y]	>	*cru*	[kryˑ]
	OFr *creue*	[kre'yə]	>	*crue*	[kryː]

Le /e/ cependant se distingue des autres voyelles. Si on observe à travers le corpus une forte tendance à allonger le [e] en position finale, cet allongement va de pair avec la réalisation fréquente d'un [j] post-vocalique.

(4)	nyeːj	*nuée*
	aneːj	*année*
	kɔʁveːj	*corvée*
	tuʁneːj	*tournée*

[10] Soulignons que l'importance de la morphologie par rapport à cet allongement est mise en question par le fait que certaines formes masculines, parmi elles la forme *musée*, sont sujettes au même "procès" d'allongement que les formes morphologiquement féminines. Voir §5 pour une discussion de ce problème.

La réalisation [eːj] de *-ée* aurait son origine dans le "patois" puisque, le suffixe -ATA, qui en français s'écrit *-ée*, s'y prononcerait *-âye* (Knecht 1985 : 60). Il reste à expliquer son fonctionnement synchronique. En étudiant la figure 8 ci-dessous présentant le taux d'allongement dans la lecture des listes de mots, nous constatons que le [j] se réalise non seulement lors d'un [eː] final, cf. (4), mais également dans le cas où la voyelle finale est [iː]. Deux solutions se présentent désormais : soit le [j] déclenche lui-même l'allongement de la voyelle, soit la durée se réalise indépendamment du [j].

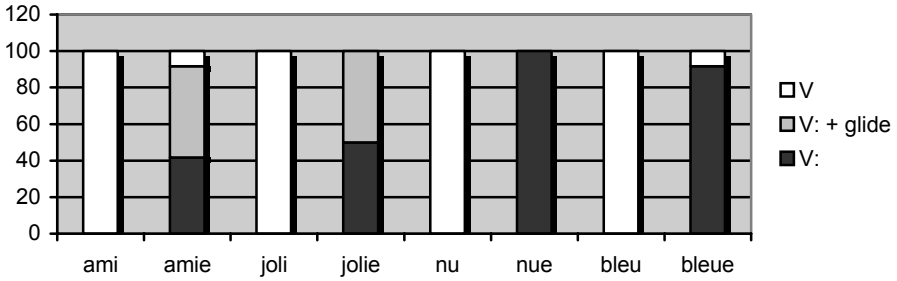


Figure 8. *Réalisation du marqueur féminin dans les listes de mots*

"… ou bien on considère que les *j* ne sont que des manifestations phonétiques facultatives (ce qu'elles sont d'ailleurs) et l'on oppose /e/ à /eː/ à la finale, à l'intérieur aucun de ces deux phonèmes ne se réalisant. […] [Cette] solution, moins économique il est vrai, est cependant celle qui emporte notre suffrage, car elle exprime mieux le fonctionnement solidaire du système : **toutes les graphies voyelle + *e* se traduisent phoniquement par un allongement vocalique accompagné facultativement d'une épenthèse consonantique."** (Métral 1977 : 162-163, mis en gras par nous).

Métral cherche à établir un système distributionnel des voyelles longues en position finale, un système qui fait du [j] le résultat d'une épenthèse post-vocalique. Reste cependant à découvrir le statut synchronique précis du [j]. On pourrait penser que le [j] est inséré pour briser un hiatus, fonctionnant ainsi comme une sorte de consonne transitoire. Ceci n'étant pas une solution courante dans le système phonologique du SR[11] ; l'insertion du [j] s'appliquerait non pas dans la phonologie "régulière", mais serait plutôt une tendance provoquée par des facteurs articulatoires. Une deuxième solution serait de le traiter comme faisant partie de la forme au niveau phonologique. Le premier facteur favorisant une telle analyse est que le [j] apparaît clairement quelle que soit sa position dans le syntagme phonologique.

(5)	en syllabe accentuée finale	la dikteːj	*la dictée*
		plydekõtɐ̯akteːj	*plus décontractées*
	en syllabe suivie de $_{\sigma}$[C(C)V…	rəkɔmãdejpaʁ	*recommandée par*
		aʁmejki	*armée qui*
	en syllabe suivie de $_{\sigma}$[V…	ʁɛstejã	*restée en*
		siɲalejuʁələveːj	*signalées ou relevées*

[11] Les [i, e] finaux de mot, sans [j] post-vocalique, sont grammaticaux dans le SR, *cf.* [amiãglɛ] *ami anglais.*

Deuxièmement, lorsque la voyelle est abrégée en position non tonique, par exemple dans [ynanejɑ̃nekɔs] *une année en Ecosse*, le [j] est le seul élément demeurant qui puisse réaliser la marque du féminin (rappelons que la durée n'est contrastive qu'en position tonique). A ce stade de notre analyse, nous savons que l'apparition du [j] doit avoir lieu au niveau lexical, puisqu'en position non tonique à l'intérieur d'un groupe rythmique, la voyelle longue est généralement réduite (*cf.* [pɑːt] *pâte* vs. [pɑtitaljɛn] *pâtes italiennes*). Dans le cas inverse, où le [j] serait inséré après la formation de groupes rythmiques, il s'attacherait à tous les (ou à aucun des) [i, e] finaux, qu'ils soient brefs ou longs lexicalement. Cette hypothèse découle du fait qu'après la formation de groupes rythmiques, le contraste quantitatif est neutralisé, éliminant ainsi le contexte d'insertion (voyelle longue). Mais l'image régulière dans le SR est cependant une réalisation de la glissante uniquement après les voyelles finales de mot qui s'allongent régulièrement en position accentuée. L'insertion du [j] doit, de ce fait, avoir lieu avant la neutralisation des valeurs quantitatives.

1.		*forme lexicale*	aneː
2.	évaluation phonologique	*forme d'output*	aneː(j)
3.	création de syntagmes		[il y a quelques années] [une année en Angleterre]
4.	création de groupes rythmiques/accentuation		(il y a quelques ann**ée**s)$_{GR}$ (une année en Anglet**e**rre)$_{GR}$
5.	réévaluation phonologique	*forme en position tonique*	ijakɛlkəzan**eːj**
		forme en position non tonique	ynan**ej**ɑ̃nɑ̃glətɛːʁ

Figure 9. *La dérivation de la voyelle lexicalement longue dans année*

Notons cependant qu'au premier regard, l'image inverse semble être le cas chez le Loc12, qui parfois insère un [j] dans d'autres contextes que purement "féminins", par exemple [etej œ] *été euh*… D'après cette observation, un traitement de la glissante comme une consonne insérée facultativement par une règle phonétique serait plus raisonnable. Cependant, en prenant en considération l'absence de [j] en contexte non féminin chez les onze autres locuteurs, il semble pertinent de proposer que la réalisation du [j] est le reflet d'une priorité donnée par la grammaire. Mais, l'appareil articulatoire et l'identité phonique entre [i, e] et [j] pris en compte, l'extension de la distribution du [j] post-vocalique pourrait s'élargir jusqu'à inclure tous les [i, e] phonétiques, ce qui semble être le cas du Loc12, un locuteur dont le système phonologique se distingue légèrement du système des autres Vaudois. Pour résumer, jusqu'au moment où de nouvelles données nous permettent de définir avec plus de certitude l'application du [j], on retient l'idée que ce segment remplit un rôle morphologique – il marque le féminin (cf. §4.4 pour une première analyse du phénomène).

À travers le corpus, trois locuteurs se distinguent des autres en ce qui concerne la réalisation du *e* féminin, i.e. Loc5, Loc11 et Loc12. Dû au cadre limité de ce travail, nous nous abstenons de rechercher l'importance d'éventuels facteurs sociaux, mais remarquons cependant brièvement que ces trois locuteurs sont les plus âgés du corpus, de même qu'ils montrent, tout au long de la conversation, un fort attachement à leur pays[12]. Les douze locuteurs réalisent tous le [j] "épenthétique" durant la lecture à haute voix, la variation inter-individuelle étant alors restreinte au parler spontané. Un objectif dans des recherches futures, suite à l'analyse de nouvelles données, sera donc de déterminer si les locuteurs plus

[12] Pour un traitement détaillé de la situation sociolinguistique en Vaud, nous nous référons à Singy (1996).

jeunes démontrent une variété de registres plus importante, à taux de réalisation de [i(ː)j/e(ː)j] divers, ou si l'on a tout simplement affaire à une certaine influence immédiate de la graphie. On pourrait s'attendre à ce que les traits régionaux se dégagent dans la conversation libre et non pas dans la lecture à haute voix. Comme les locuteurs plus jeunes varient dans leur réalisation du [j], on peut cependant avancer l'hypothèse que la palatalisation devient de plus en plus défavorisée dans le langage spontané, différent des locuteurs les plus âgés qui ont une grammaire dans laquelle cette consonne se réalise de manière plutôt stable. La forte réalisation du [j] en lecture semble conforter indirectement (par le biais de la graphie) l'hypothèse qu'il remplit un rôle morphologique.

Nous avons dans cette partie exposé nos résultats qui confirment le maintien de la durée vocalique en position finale. Tout comme le fait Morin (1994), entre autres, nous avons isolé deux facteurs qui contribuent à une identification première de la "distribution" de la durée contrastive ː 1) le contexte connaît la chute (historique) d'un schwa final, et 2) les formes soumises à l'allongement sont dans la majorité des cas morphologiquement féminines.

4. Discussion

Nous envisagerons dans cette partie un traitement théorique de la durée vocalique. Il s'agira tout d'abord d'explorer la façon dont la distinction quantitative survit dans les variétés romandes. Un deuxième point pertinent concerne la variation dialectale, c'est-à-dire en quoi la variété vaudoise diffère-t-elle du FR ? Selon la Théorie de l'Optimalité (OT), la grammaire est le résultat d'un conflit entre des contraintes de fidélité et des contraintes de marque, le premier type exigeant l'identité entre l'output et l'input, le deuxième type favorisant des structures non marquées dans l'output. Pour chaque input, le Générateur génère un nombre infini de candidats d'output, d'où le fait qu'un candidat unique est sélectionné par l'Évaluateur. Dans l'Évaluateur, des principes universels, connus sous le nom de contraintes, sont rangés de manière hiérarchique. Le candidat optimal est celui qui entraîne le plus petit nombre de transgressions des contraintes rangées dominantes. La variation linguistique des langues naturelles est expliquée par l'idée que les différences sont exclusivement dues à des ordres de contraintes différents.

A l'intérieur d'un tel modèle, la réalisation de la durée contrastive doit s'expliquer par un ordre de contraintes qui sélectionne comme output optimal un candidat ayant préservé la durée vocalique présente dans l'input. Contrairement au vaudois, la quantité distinctive n'est pas préservée dans le FR, et nous pouvons conséquemment avancer l'hypothèse que la contrainte interdisant la durée vocalique occupe une position moins importante dans la grammaire du vaudois que dans celle du FR. Il s'ensuit que la fidélité à l'élément dans l'input reflété par la longueur vocalique dans l'output domine, au moins dans le vaudois, la contrainte "pas de voyelles longues". Soulignons que toute analyse de la longueur contrastive doit finalement être compatible avec celle de la longueur contextuelle. Comme le cadre du présent article est volontairement restreint à l'aspect contrastif, nous nous abstiendrons dans ce travail d'aborder une analyse globale de la durée[13].

[13] L'analyse qui suit ne va en rien contre l'autorisation de l'allongement conditionné. Des contraintes plus spécifiques requérant l'allongement dans des environnements particuliers vont dominer la contrainte générale interdisant les voyelles longues (*cf.* Féry 2001 : 29 pour une grammaire partielle du français).

4.1. Représentation et évaluation de la durée vocalique

Montreuil (2003) propose une analyse dans le cadre d'OT réunissant la durée conditionnée et la durée contrastive. Dans son analyse, il propose que les voyelles longues soient spécifiées comme telles dans l'input. Dû à la position dominante qu'occupe FID-μ sur Vμ, les voyelles sortent longues dans l'output optimal.

/Eː, Œː, Oː, Aː/	Vμ & ATR-σ	FID-μ	RTR/C	Vμ	ATR-σ
e, ø, o, ɑ		*!			
ɛ, œ, ɔ, a		*!			*
☞ eː, øː, oː, ɑː				*	
ɛː, œː, ɔː, aː	*!			*	*

Figure 10. *Les voyelles moyennes tendues longues*[14]

Montreuil (op.cit.), tout comme le fait Morin (1994), présente la longueur morphologique comme un allongement compensatoire, soulignant qu'un modèle comme OT, basé sur l'output, *a priori* ne ferait pas de distinction entre la longueur lexicale et la longueur morphologique. L'élément impératif à retenir est que, quelle que soit la forme de l'input (voyelle longue ou brève, longueur lexicale ou morphologique), c'est *l'ordre des contraintes* qui ultérieurement vient décider l'output optimal. Cela découle de la notion de la Richesse de la Base : la grammaire est focalisée sur la forme d'output, ce qui signale que l'input peut *a priori* prendre une forme quelconque.

Évaluons dans un premier temps la paire [ny nyː] *nu – nue.* Pour la paire dans sa totalité, nous proposons que la seule différence entre les deux formes réside dans la spécification de la durée dans l'input. Rappelons que l'ordre des contraintes évaluant les deux formes n'alterne pas. Puisque le contexte structural de la voyelle est identique, le facteur distinguant les deux formes semble trouver sa source dans la fidélité dominante (la correspondance input-output)[15]. Peut-on cependant postuler de la structure autosegmentale au niveau sous-jacent ? Si oui, une solution serait que la marque du féminin est tout simplement représentée en tant que more supplémentaire dans l'input (cf. Montreuil op.cit.). Si la réponse est non, le marquage féminin doit se représenter sous une forme autre que purement morique au niveau sous-jacent.

Tentons alors une évaluation de la forme féminine avec deux formes d'input diverses, la première avec une voyelle longue, la deuxième avec la marque du féminin représentée par un schwa (le schwa historique, cf. Morin 1994). Nous notons que, pour [ny ː], un seul candidat sera optimal, quelle que soit la forme d'input sélectionnée. Une liste des contraintes utilisées dans la figure 11 est exposée dans (6).

---

[14] Vμ & ATR-σ pas de voyelle longue et relâchée
FID-μ le compte morique I-O reste inchangé
RTR/C pas de voyelle tendue devant C
Vμ les voyelles sont courtes (monomoriques)
ATR-σ les voyelles accentuées sont tendues (Montreuil op.cit. : 325-327)

[15] Quant à la marque, on considère universel l'ordre partiel Vμ/*V□ >> Vμμ/*V. Comme le note Montreuil (op.cit.), entre autres, aucune langue attestée n'autorise uniquement des voyelles longues, tandis que l'image inverse est souvent observée. La dominance de la fidélité semble ainsi être l'aspect crucial dans ce cas spécifique.

(6) $*_{\sigma}$[ə pas de schwa en position initiale de syllabe
FID-μ le compte morique I-O reste inchangé
Vμ les voyelles sont monomoriques
*V: pas de voyelles longues
Vμμ les voyelles sont bimoriques
*V pas de voyelles brèves

	ny:$_{FEM}$ nyə$_{FEM}$	$*_{\sigma}$[ə	FID−μ	Vμ/*V:	Vμμ/*V
a.	☞ny:				
b.	ny.ə	*!			*
c.	ny		*!		*
	ny				
a.	☞ ny				*
b.	ny.ə	*!	*		*
c.	ny:		*!	*	

Figure 11. *Évaluation de nu – nue dans la grammaire du vaudois*

La domination de FID-μ élimine le candidat c) des deux schémas puisqu'il ne maintient pas l'identité morique entre l'input et l'output. Le candidat b) contient uniquement des voyelles monomoriques, mais l'interdiction dans la phonologie du français de réaliser un schwa initial de syllabe s'impose de manière cruciale également dans l'évaluation de ces formes. Le candidat a) sort alors gagnant.

Jusqu'à maintenant nous avons favorisé une analyse qui donne priorité à la fidélité morique qui, à son tour, domine Vμ/*V:. La hiérarchie proposée dans la figure 11 se heurte cependant aux exemples [blø blø:] *bleu – bleue*, puisque ici il ne s'agit pas d'une forme qui sort bimorique dans l'output optimal, mais trimorique. Si l'on accepte l'idée de la bimoraïcité inhérente du [ø], il faut nécessairement ajouter un élément à sa variante longue. Il ne saurait s'agir d'une obéissance stricte à la contrainte FINAL LENGTH "PhP-final full vowels are bimoraic" (Féry 2001 : 18), car cela éliminerait plutôt la distinction entre *bleu* et *bleue*, rendant les deux voyelles bimoriques, voire identiques. Une solution serait de donner à BIMORE[16] moins de priorité dans la grammaire, acceptant qu'une voyelle finale puisse sortir trimorique. Une autre solution qui suit la même ligne est de séparer explicitement la moraïcité et la longueur, de sorte que l'on ait deux types de contraintes opérant visiblement dans la grammaire, Vμ >> Vμμ et *V: >> *V. Tandis que Vμ scanne les candidats en vue d'exclure les voyelles bimoriques, *V: cherche à pénaliser tout candidat ayant une voyelle allongée (par l'inclusion d'une more supplémentaire).

[16] BIMORE "Une syllabe est composée (tout au plus) de deux mores. " (Lyche 2003 : 355).

	blø: FEM bløə FEM	*σ[ə	FID−μ	*V:	Vμ	Vμμ	*V
a.	☞ blø:			*	*	*	
b.	blø.ə	*!			*	*	*
c.	blø		*!		*		*
	blø						
a.	☞ blø				*		
b.	blø.ə	*!	*		*	*	*
c.	blø:		*!	*	*	*	

Figure 12. *Évaluation de bleu/bleue dans la grammaire du vaudois*[17]

L'avantage immédiat de cette distinction entre moraïcité et longueur est, comme on le voit pour les candidats a) évalués, qu'elle permet de rendre compte du contraste entre le masculin et le féminin. Dans une perspective diachronique, cet ordre implique que *V: est la première contrainte qui va se déplacer dans la hiérarchie de façon à dominer FID-μ, un scénario vers lequel nous nous tournons désormais.

4.2. Les divergences entre les grammaires de SR et FR

Abordons maintenant le trait distinctif qu'exhibe la grammaire vaudoise par rapport à celle du FR. Comme il l'a déjà été souligné, les divergences inter-dialectales n'auraient pas leur source dans l'input, mais dans l'ordre des contraintes. Pour le phénomène de la durée vocalique, cette notion, fondamentale dans OT, implique que l'absence de voyelles longues finales dans le FR trouve sa solution dans une hiérarchie légèrement différente de celle du vaudois.

D'un point de vue diachronique, la disparition de la durée contrastive est provoquée tout d'abord par l'instabilité de réalisation des voyelles longues dans la langue ambiante, suivie d'une réorganisation des contraintes décisives dans la grammaire de l'enfant par rapport à celle des adultes. Suite à l'analyse dans la figure 12, les contraintes cruciales dans ce cas spécifique seraient FID-μ et *V:. Au niveau du lexique, on s'attendrait également à ce qu'un changement ait lieu. Nous proposons que, une fois la contrainte pénalisant les voyelles longues promue dans la grammaire, à une position dominante à FID-μ, petit à petit il n'y a plus aucune raison de différencier les inputs masculin et féminin. Cela suit directement de l'optimisation du lexique, cf. McCarthy (2002 : 77) : "If having a determinate underlying representation for [A] is important, then the choice can be made during language learning by a procedure that Prince and Smolensky call *lexicon optimization:* choose the underlying representation that gives the most harmonic mapping. The mapping /A/ → [A] is fully faithful, while the /B/ → [A] mapping incurs at least one faithfulness violation. Since the two mappings have otherwise identical violation-marks, the /A/ → [A] mapping is the more harmonic one."

Lorsque FID-μ perd de son importance et est désormais placée sous la dominance de *V:, le candidat à voyelle longue ne saurait être optimal. L'absence, de plus en plus régulière, de voyelles longues dans l'output optimal de la langue ambiante fait choisir à

[17] Le nombre de mores associé aux voyelles n'est pas montré dans le tableau. Notons cependant que [ø] est supposé bimorique, tandis que le [ø:] représente une voyelle trimorique.

l'enfant une forme d'input sans voyelle longue, un choix qui finalement entraînerait une restructuration plus globale de la structure lexicale des formes féminines[18].

	blø:$_{FEM}$	*$_{\sigma}$[ə	FID−μ	*V:	Vμ	*V
a.	blø		*!		*	*
b.	blø.ə	*!			*	*
c.	☞ blø			*	*	

Figure 13a. *Perte de la longueur dans le FR - fidélité aux voyelles longues*

	blø:$_{FEM}$	*$_{\sigma}$[ə	*V: ◄►	FID−μ	Vμ	*V
a.	☞ blø			*	*	*
b.	blø.ə	*!			*	*
c.	blø:		*!		*	

Figure 13b. *Perte de la longueur dans le FR - réorganisation de la grammaire*

	blø$_{FEM}$	*$_{\sigma}$[ə	*V:	FID−μ	Vμ	*V
a.	☞ blø				*	*
b.	blø.ə	*!		*	*	*
c.	blø:		*!	*	*	

Figure 13c. *Perte de la longueur dans le FR - changement d'input optimisé*

Une perte éventuelle de la longueur contrastive s'expliquerait par cette réorganisation des contraintes FID-μ et *V:, créant désormais une hiérarchie qui empêche toute voyelle spécifiée en tant que longue dans l'input de faire surface comme telle dans l'output. Pour résumer cette section, nous pouvons constater que le SR maintient l'information morphologique d'une manière plus générale que le fait le FR. Non seulement la marque du féminin y est gardée dans les formes à consonne finale, *cf.* [so] vs. [sɔt], mais elle l'est aussi dans les formes à voyelle finale. Nous pouvons alors proposer que le féminin est généralisé dans le lexique du SR.

4.3. Évaluation des modèles théoriques

Jusqu'alors nous avons expliqué la durée vocalique comme une fidélité au compte morique dans l'input, perspective soutenue par Féry (2001) et Montreuil (2003). Le modèle morique présenté dans Féry (op.cit.) comporte certaines faiblesses, et en particulier il est incapable d'expliquer la disparition des durées contrastives en finale absolue. Elle propose que la disparition de la durée constrastive est due au déplacement des consonnes finales dans la coda de la syllabe précédente, une resyllabation probablement causée par la réduction morique de la voyelle allongée.

"There is a tendency in French to eliminate the semisyllables and to integrate the final consonants into the coda of the preceding syllable. A word like *même* 'even', which is sometimes pronounced with a long vowel [mɛː.m], is often realized with a short one [mɛm], which allows the final [m] to be the coda of the syllable, instead of being relegated into a semisyllable. As a result, length contrasts like those illustrated in (31) are being lost." (Féry 2001 : 20).

[18] Soulignons que cette restructuration n'affecterait pas en premier lieu toutes les formes féminines. Il est évident qu'il faut, au niveau sous-jacent, distinguer les différents éléments réalisant le féminin (consonne additionnelle, durée vocalique etc.).

L'idée est intéressante, mais ce que néglige Féry est de montrer explicitement le facteur déclenchant ce développement. Est-ce l'élimination des demi-syllabes qui provoque la perte de la durée vocalique, ou est-ce au contraire l'élimination de la durée vocalique qui autorise le déplacement de la consonne en position de coda ? Par ailleurs, selon cette analyse, on s'attendrait à ce que la resyllabation de consonne soit observée à un taux plus important à travers toute la langue. Malheureusement, un tel type d'exemples manque entièrement dans son analyse.

Le point commun des analyses de Féry et de Montreuil par rapport aux paires minimales comme *pomme* vs. *paume* est la bimoraïcité spécifiée de la voyelle de *paume* dans l'input. Pour les paires incluant un *e* féminin, il s'agit plus spécifiquement d'une projection d'une more supplémentaire dans l'input, réalisée en tant que durée vocalique dans les contextes où la grammaire favorise la fidélité morique. Une approche autre que purement morique afin de traiter ce phénomène existe cependant. Elle découle de l'observation faite que la durée vocalique synchronique semble en partie réglée par un principe d'uniformité de morphème. Nous ne contestons pas le fait qu'il existe des formes à voyelle longue qui ne sauraient s'expliquer par des raisons purement morphologiques, mais le rôle synchronique de l'allongement constitue un aspect qui mérite d'être davantage étudié.

À un moment entre le XV^{e} et le XVII^{e} siècle, le schwa précédé par une voyelle tonique disparaît de la prononciation (Morin 1994 **:** 144). Ce schwa post-vocalique apparaît souvent dans des contextes dits féminins. Admettons que le schwa fasse difficilement partie de la représentation sous-jacente contemporaine de la forme féminine **:** quel est cependant l'élément qui fait que les adjectifs et les noms féminins gardent la longueur contrastive, s'il ne s'agit pas d'un élément sous-jacent que la grammaire exige réalisé dans l'output optimal ? Est-ce tout simplement une question de mores ? Il faut mettre en lumière le fait qu'en nous concentrant exclusivement sur une représentation et une analyse moriques, nous avons faussement réduit à néant le rôle explicite de la morphologie. Soulignons immédiatement que l'analyse du corpus vaudois élargi n'est pas terminée, et que de poser une analyse avec trop de détermination semble téméraire à ce stade. Nous tentons tout de même de proposer qu'une contrainte de nature morpho-phonologique, MAX-IO[FEM], est visiblement active dans la grammaire, exigeant la fidélité à un élément morphologiquement important dans la forme sous-jacente, le féminin (sa représentation d'input, more ou schwa, peu importe, cf. la Richesse de la Base). Dans un modèle comme OT, par lequel les contraintes sont considérées universelles, l'idée d'avoir des contraintes portant sur un morphème spécifique n'est pas idéale. Considérons cependant l'idée de Féry (2001 : 22) sur une contrainte FEM(C)[19] **:**

"At first sight, this constraint is too language specific, but if it is understood as a special case of a more general constraint positing that the feminine forms are phonologically more complex than their masculine counterpart, it becomes universal after all. The fact that the feminine is some kind of feature added to the unmarked gender – which is masculine – seems to be universal enough to elevate it to the status of a universal markedness constraint. "

Faute d'avoir à ce stade un formalisme qui peut rendre compte de l'influence exercée par la graphie sur la prononciation (cf. Laks 2005), une influence que nous ne pouvons occulter complètement, la vraie nature de la présente contrainte portant sur la marque du féminin reste à découvrir. L'existence même d'une telle contrainte semble

[19] FEM(C) exige que les adjectifs féminins se terminent par une consonne.

cependant justifiée en ce que l'on peut identifier à travers différentes langues la complexité particulière des formes féminines.

	bløəFEM blø:FEM	*σ[ə	MAX–IO[FEM]	*V:	IDENT–IO
a.	☞blø:			*	
b.	blø.ə	*!			
c.	blø		*!		*
	blø				
a.	blø:			*!	*
b.	blø.ə	*!			*
c.	☞ blø				

Figure 14. *Évaluation II des formes bleu/bleue*[20]

Dans certaines variétés du français, la longueur morphologique a été attestée non seulement dans le contexte féminin, mais également dans le contexte pluriel, cf. un [ami] *vs*. des [ami:] . Sur le plan phonologique, s'agit-il d'une représentation identique dans l'input pour les deux morphèmes, ou s'agit-il plutôt de deux phénomènes distincts ? Quant à la fidélité, faut-il inclure des contraintes séparées, l'une exigeant la réalisation du féminin, l'autre exigeant celle du pluriel ? Une preuve de l'existence de notre contrainte morpho-phonologique MAX-IO[FEM] serait un scénario dans lequel la longueur du pluriel[21] disparaît de manière stable tandis que la longueur du féminin persiste. Si les deux types d'allongement sont représentés de manière identique dans l'input, par une more supplémentaire, et que cette more est éliminée dans l'output optimal par une même contrainte structurale, il est difficilement justifiable de postuler la contrainte MAX-IO[FEM]. Si, au contraire, on attestait une baisse d'allongement dans les contextes pluriels et non pas dans les contextes féminins, l'existence de deux contraintes actives serait davantage plausible.

Ce dernier scénario semble plus probable. L'extension d'une marque du pluriel visible est dans le SR marginale par rapport à celle de la marque du féminin, cette dernière existant partout dans la langue. Si l'on met l'accent sur la structure morphologique plutôt que sur la structure phonologique, on va simplifier la forme de l'input. En effet, la marque du féminin et du pluriel est spécifiée morphologiquement et préservée par les contraintes MAX-IO[FEM] et MAX-IO[PLUR], et non plus par la spécification de deux mores de la voyelle (cf. $blø_{\mu}$ $blø_{\mu\mu}$). L'input s'en trouve ainsi allégé[22]. La structure morique des divers candidats ne sera déterminée qu'une fois ils sortent du Générateur. La structure du candidat optimal est ensuite réglée par la grammaire.

4.4. La question du [j]

Le dernier point qui reste à discuter est l'apparition régulière du [j] en position finale des mots en *-ée*. Il faut rendre compte des deux facteurs suivants **:** 1) la réalisation d'un [j]

[20] A ce stade dans l'analyse, il n'est pas possible de proposer un ordre entre les contraintes IDENT-IO et *V:. Nous aurions pu enlever *V: sans que le candidat optimal change, cependant l'idée qu'une contrainte pénalisant des voyelles longues existe dans la langue fait que nous la gardons dans la hiérarchie.

[21] Pour la longueur du pluriel, *cf.* entre autres Montreuil (2003).

[22] Que la solution optimale soit un allongement de la voyelle et non pas d'autres solutions possibles pourrait s'expliquer par une double volonté : 1) maintenir le contraste masculin – féminin, et 2) réaliser un contraste minimal. L'allongement vocalique n'ajoute pas de nouveau segment à la forme de base et pourrait de ce fait être considéré causer un contraste minimal. A ce moment, les implications d'une telle analyse restent à découvrir.

lorsque la voyelle précédente est abrégée, et 2) l'absence d'insertion de [j] dans d'autres mots à [e] final. Ces deux observations intimement liées indiquent au premier regard que la réalisation du [j] est un résultat d'une priorité à la fidélité. Tentons alors une analyse où le [j] observé fait partie du candidat sélectionné par l'Évaluateur.

En position finale absolue, nous observons les trois possibilités [aneː aneːj anej]. Cette multitude de candidats optimaux est particulière pour le /eː/[23]. L'aspect à retenir est la réalisation stable de soit un [j], soit une voyelle longue (ou les deux à la fois). Il semble donc que la marque du féminin possède, dans le cas d'une suite *-ée* finale, deux moyens pour se réaliser à la surface. A l'intérieur d'un syntagme, la durée n'est plus une option et l'on recourt à la glissante seule. Ne négligeons pas qu'une glissante est absente dans les mots "féminins" se terminant par une autre voyelle que [e]. Il est vrai qu'une distribution symétrique exigerait la réalisation d'une glissante à la suite du pendant postérieur du [e], notamment le [o] dans le contexte */oːFÉM/. Cependant, comme le souligne Métral (1977 : 163), il n'existe pas d'opposition (graphique) **-(e)aue* vs. *-eau* en SR, reflétant l'asymétrie phonémique exposée dans la figure 3. Mais reste à expliquer pourquoi aucune glissante n'est autorisée (de manière stable) après /iː uː øː/. Sans tenter d'en faire une analyse trop détaillée, il semble exister dans la grammaire une contrainte interdisant [eː] en position finale. Si l'on considère l'inventaire vocalique de manière plus globale, la distribution des voyelles moyennes antérieures non arrondies se distingue des autres voyelles moyennes par le fait que la variante tendue ne se réalise pas normalement en syllabe fermée (cf. entre autres Féry 2001 et Lyche 2003). Que ce timbre ait une distribution particulière également en syllabe finale ouverte n'étonnerait alors aucunement. La divergence observée entre [eː] et les autres voyelles pourrait s'expliquer par la position dominante que tient une contrainte *eː]# par rapport à *Vː[24]. Le candidat ayant un [j] final ne transgresse pas cette contrainte spécifique. De plus, une fois la forme avec glissante sélectionnée, le [j] demeure partie du mot tout au long de la dérivation.

aneəFEM aneːFEM	*σ[ə	MAX-IO[FEM]	*eː]#	*Vː	NOCODA	IDENT-IO
a. ☹ anej					*	*
b. aneː			*!	*		
c. ☞aneːj				*!		
d. ane		*!				*
e. aneə	*!					

Figure 15. *Évaluation de la forme* année

Le candidat a) demeure un défi pour l'analyse proposée. Pourquoi la grammaire, si le [j] par lui-même suffit en tant que réalisateur de la marque du féminin, ne préfère-t-elle pas de manière plus stable [anej], sans voyelle longue, plutôt que [aneːj], candidat qui dans le SR est généralement la forme optimale en position finale absolue ? Métral (1977 : 162) observe les deux réalisations [eːj] et [ɛj], indiquant une corrélation étroite entre l'aspect qualitatif et l'aspect quantitatif. Sans avoir mené une analyse phonétique fine, il s'avère que, dans notre

[23] Différemment que pour le /eː/, cela est attesté pour le /iː/ avant tout en position finale absolue. Il est également à noter que la présence du [j] ne semble pas aussi saillante et catégorique suite à un [i(ː)] qu'elle l'est suite à un [e(ː)].

[24] L'ordre *eː]# >> *V: entre dans une relation Panini (Prince 1997) en ce que la contrainte plus spécifique est donnée plus de priorité que la contrainte plus générale. Soulignons cependant que dans le cadre d'OT, cela n'est pas une relation de dominance qui découle de manière automatique. Il n'existe *a priori* aucune restriction quant à la précédence des contraintes.

corpus également, le /e/ (abrégé) peut légèrement s'ouvrir dans le cas où il est suivi d'un [j]. Cela suit directement l'observation plutôt générale du français que seul le [ɛ] apparaît en syllabe fermée. La voyelle fermée longue, suite à la même règle, est donc préférablement reléguée aux syllabes ouvertes. Un autre aspect, mis à côté dans ce travail, est le timbre sous-jacent des voyelles moyennes. Il est généralement estimé que ces voyelles ne sont pas spécifiées pour la tension dans l'input[25], mais que leur timbre dans l'output découlerait automatiquement de l'évaluation grammaticale. Si une des forces conflictuelles opérant sur la syllabe fermée de *année* est la marque (dans ce contexte connue sous le nom de *Loi de Position*), qui choisirait [ɛ], une deuxième force conflictuelle pourrait trouver sa source dans les relations de correspondance, dans l'uniformité de morphème (le féminin en position finale est spécifié par une voyelle longue). Premièrement, la présence du [j] défavorise la voyelle longue marquée. Deuxièmement, le manque de timbre sous-spécifié indique qu'il ne s'agit pas non plus d'une fidélité qualitative. La voix décisive est celle de la hiérarchie. Comme nous notons une tendance à réaliser [eːj], il en découle que la grammaire donne priorité à la régularité morphologique (féminin = durée). Il faut enfin rappeler qu'une suite optimale *voyelle longue* + *[j]* n'est pas la pire solution. Tout d'abord, sa présence lors d'un [iː] montre que le phénomène allongement + fermeture n'affecte pas exclusivement le [eː] mais peut s'appliquer aux deux voyelles fermées non-arrondies dans les contextes féminins. Dans ce cas également, le candidat est [amiːj] plutôt que [amij], une voyelle longue suivie de la glissante où le [j] seul suffirait comme marque du féminin réalisée. Deuxièmement, déjà noté dans Montreuil (1995 : 82), il existe des variétés du français, plus conservatrices, qui incluent le [j] dans le groupe de consonnes allongeantes. Si la grammaire du SR autorise le [j] final à allonger la voyelle précédente de manière plutôt régulière dans des contextes autres que ceux reflétés dans la graphie par *-ée* et *-ie*, cela pourra mettre davantage en lumière le rôle plus global que joue la glissante [j] dans le système (morpho-)phonologique du français suisse.

5. Conclusion

Dans ce travail, nous avons présenté le statut actuel de la durée vocalique dans une variété du français suisse. Avec notre propre corpus PFC comme base, nous avons montré que la durée, en tant qu'élément distinctif, persiste en position finale absolue, contribuant à la création de paires minimales telles que /ami amiː/ et /blø bløː/. Dans une deuxième section, nous avons discuté la façon dont la durée contrastive est représentée dans la grammaire. La longueur contrastive persiste dans la grammaire sous la forme d'une spécification lexicale, fidèlement réalisée dans l'output. Les deux propositions théoriques présentées mettent l'accent sur la représentation morique et l'importance morphologique, respectivement. Nous avons opté pour une analyse "hybride" par laquelle nous acceptons que la marque du féminin au niveau sous-jacent soit représentée par un élément contribuant, dans ce cas, au poids prosodique de la voyelle dans l'output. En prenant en compte l'insertion du [j] suite au [e(ː)], nous suggérons cependant que la durée en tant que réalisateur du féminin ne soit pas la seule solution grammaticale.

Un premier aspect à discuter dans des recherches futures concerne la longueur morphologique dans sa totalité. Si le pluriel et le féminin sont représentés de manière identique dans leur forme sous-jacente, sans marquage individuel, on s'attendrait à un comportement et un développement plutôt identiques de la longueur des deux morphèmes. Si, au contraire, il s'agit de deux éléments séparés qui se développent de manière

[25] Cf. Féry (2001), Lyche (2003) et Montreuil (2003).

individuelle (taux de diffusion lexicale variables), nous pouvons avec plus de certitude avancer l'hypothèse qu'une contrainte spécifique au morphème existe dans la grammaire.

Une deuxième question à laquelle ce travail n'apporte pas de solution satisfaisante concerne la façon dont nous pouvons expliquer la voyelle longue dans *musée*, forme masculine sujette à la même règle d'allongement que d'autres formes en *-ée*. Montreuil (2003) en rend compte en proposant que la longueur ne soit pas exclusivement morphologique et que l'on doive prendre en considération le côté étymologique de certains marquages[26]. Cette observation a pour conséquence une image plus complexe de la durée finale, c'est-à-dire que le facteur morphologique n'est pas souverain en ce qui concerne ce phénomène. D'une part, la durée semble étymologique, découlant d'une généralisation des formes auparavant prononcées *-âye*. D'autre part, la durée est motivée de façon plus synchronique en ce qu'elle sert à distinguer deux formes morphologiquement distinctes.

La durée contrastive est un trait qui a disparu dans la majorité des variétés du français. D'après les résultats obtenus, il semble que ce phénomène ne soit pas en voie de disparition dans le vaudois. Cependant, pour obtenir une image plus complète de l'extension de la durée contrastive, un corpus représentant la Suisse romande dans sa totalité semble nécessaire, un objectif que l'on peut atteindre dans le cadre du projet PFC.

Références

Andreassen, H.N. (2003). Comment le schwa et la consonne de liaison vacillent et s'évanouissent dans le vaudois. Un traitement de la variation. Cand. philol. thesis., Universitetet i Tromsø.

Andreassen, H.N. (2004). "Une contrainte de fidélité flottante pour le traitement du schwa et de la liaison dans le canton de Vaud." Eychenne, J. et G. Mallet (eds.), *Du segmental au prosodique : protocoles, outils, extensions et travaux en cours. Bulletin PFC* n^0 3. ERSS UMR 5610 CNRS & Université de Toulouse-Le Mirail, 139-184.

Féry, C. (2001). "Markedness, Faithfulness, Vowel Quality and Syllable Structure in French." *Linguistics in Potsdam* 16, 1-31.

Girard, F. et C. Lyche (2003). "La Phonologie du Français Contemporain dans le Domfrontais : un français en évolution." *La Tribune Internationale des Langues Vivantes* 33, 166-173.

Knecht, P. (1985). "La Suisse Romande." R. Schläpfer (ed.) *La Suisse aux quatre langues.* Genève : Editions Zoé, 125-169.

Knecht, P. et C. Rubattel (1984). "A propos de la dimension sociolinguistique du français en Suisse romande." *Le Français Moderne* 52, 138-150.

Laks, B. (2005). "La liaison et l'illusion." *Langages* 158, 101-125.

Lyche, C. (2003). "La loi de position et le français de Grenoble." Delais, E. et J. Durand (eds.), *Corpus et variation en phonologie du français : methodes et analyses.* Toulouse : Presses Universitaires de Mirail, 349-372.

Martinet, A. (1945). *La prononciation du français contemporain.* Genève : Droz.

McCarthy, J. (2002). *A Thematic Guide to Optimality Theory.* Cambridge : Cambridge University Press.

Métral, J.-P. (1977). "Le vocalisme du français en Suisse romande. Considérations phonologiques." *Cahiers Ferdinand de Saussure* 31, 145-176.

Miller, J.L. et F. Grosjean (1997). "Dialect Effects in Vowel Perception : The Role of Temporal Information in French." *Language and Speech* 40 (3), 277-288.

[26] Ici encore le rôle de la graphie n'est probablement pas négligeable.

Montreuil, J.-P. (1995). “Weight and length in Conservative Regional French.” *Lingua* 95, 77-96.

Montreuil, J.-P. (2003). “La longueur vocalique en français de Basse-Normandie.” Delais, E. et J. Durand (eds.), *Corpus et variation en phonologie du français : methodes et analyses*. Toulouse : Presses Universitaires de Mirail, 321-348.

Morin, Y.C. (1994). “Phonological interpretations of historical lenghtening.” Dressler, W.U., M. Prinzhorn et J. Rennison (eds.), *Phonologica 1992. Proceedings of the 7th International Phonology Meeting*. Turin : Rosenberg & Sellier, 135-155.

Morin, Y.C. (2000). “Le français de référence et les normes de prononciation.” Francard, M., G. Geron et R. Wilmet (eds.), *Actes du colloque de Louvain-la-neuve 3-5 novembre 1999, Cahiers de l’Institut de Linguistique de Louvain* 26, 91-135.

Prince, A. (1997). *Paninian Relations.* Ms. UMass, Amherst 10/31/97. Téléchargé le 30 mai 05 de [www.ling.rutgers.edu/gamma/talks/umass1997.pdf]

Prince, A. et Smolensky, P. (1993). *Optimality Theory : Constraint Interaction in Generative Grammar.* Report no. RuCCS-TR-2. New Brunswick, NJ : Rutgers University Center for Cognitive Science.

Schoch, M., O. Furrer, T. Lahusen, et M. Mahmoudian-Renard (1980). “Résultats d’une enquête phonologique en Suisse romande.” *Bulletin de la Section de Linguistique de la Faculté des Lettres de Lausanne* 2.

Singy, P. (1996). *L’image du français en Suisse romande. Une enquête sociolinguistique en Pays de Vaud.* Paris : L’Harmattan.

Voillat, F. (1971). “Aspects du français régional actuel.” Marzys, Z. et F. Voillat (eds.), *Actes du Colloque de dialectologie francoprovençale (Neuchâtel).* Neuchâtel, Faculté des Lettres, Genève : Droz. 216-241.

von Wartburg, W. (1993). *Evolution et structure de la langue française.* (12^{e} édition). Tübingen et Basel : Éditions Francke.

Quatre générations de proéminences mélodiques chez une famille française de Cussac Fort-Médoc (33)

Geneviève CAELEN-HAUMONT

1. Introduction

Depuis les premiers travaux informatisés sur la parole, l'amplitude mélodique a été à l'origine de très nombreuses études en France comme à l'étranger. Une des toutes premières initiatives en matière de recherche a été ainsi de concevoir une grille de 4 ou 5 niveaux, vraisemblablement inspirée de la notation musicale, afin de quantifier les modulations de F0 et de repérer des contours caractéristiques.

Indépendamment des études proprement linguistiques (dans un premier temps dans les années soixante-dix, très orientées vers les aspects syntaxiques), le secteur de l'affectivité, des attitudes, des émotions a été également très tôt investigué, et d'emblée on y a pressenti l'importance de l'amplitude mélodique. Spécifiquement pour le français dans ce domaine, on peut citer parmi les premiers travaux Faure (1970, 1973), Fónagy et al. (1973), Léon (1970, 1971, 1976), Carton (1970, 1972). Ces études se caractérisent par la recherche de profils acoustiques distinctifs de l'émotion, des attitudes, ou sociolinguistiques, et s'appuient sur les valeurs pertinentes de F0, formes et patrons mélodiques, durées et débits.

Plus récemment, Ohala (1983) a proposé d'instancier un *Code biologique* de la fréquence fondamentale, *Frequency Code,* fondé sur le fait que les larynx de petite taille ont tendance à produire des notes plus élevées que ceux de plus grande taille. Il en découle que les fréquences élevées sonnent comme vulnérables et soumises, les fréquences basses comme protectrices et dominantes. Des expérimentations ont montré que pour le code fréquentiel, le registre fréquentiel F0 moyen et l'étendue de F0 étaient pertinents.

Dernièrement, on sait que Chen et al. (2002) ont proposé à leur tour deux autres codes, le *Code de Production* et le *Code de l'Effort*. Ce dernier est fondé sur le fait qu'un effort articulatoire plus intense a tendance à créer des réalisations phonétiques plus élaborées et plus explicites. Les codes biologiques se prêtent à deux types d'interprétations, information et affectivité, et dans le cas du Code de l'Effort, l'emphase répond à une information sur l'importance subjective que le locuteur accorde à son message, tandis qu'une émotion comme la surprise répond à celle de l'affectivité. Dans les deux cas, le locuteur dépense de l'énergie pour signaler ces fonctions. Trois indices de F0 manifestent cet effort et se révèlent pertinents pour l'interprétation des significations : la hauteur des pics de F0, le retard dans l'alignement des pics, et le registre de F0.

L'étude que nous proposons ici fait suite à de nombreux travaux que nous avons entrepris depuis la fin des années soixante-dix, portant sur les aspects sémantiques, pragmatiques et la subjectivité. Dans ces travaux, nous avons été amenée en particulier à tester différents modèles syntaxiques, sémantiques, pragmatiques, empruntés à la littérature ou originaux, ayant tous la propriété de quantifier, en prédiction, des valeurs mélodiques telles que l'amplitude ou les valeurs maximales (voir à ce sujet, Caelen-Haumont 1991, à paraître).

Dans l'étude présente, nous utilisons une grille d'analyse particulière en 9 niveaux, conçue pour analyser avec précision les modulations et l'amplitude de F0. Notre objectif est de contribuer, dans le cadre des données PFC, à la description du parler du français régional de Bordeaux, par le domaine de la prosodie. Nous focalisons notre étude

sur des éléments essentiels de la mélodie que sont ses proéminences. Pour ce faire, nous nous sommes intéressée aux productions d'une famille présentant des critères de très grande stabilité géographique, la famille n'ayant jamais bougé de Cussac. Ce village, situé au nord de Bordeaux, d'environ 1000 habitants, est situé en plein territoire des vins du Haut-Médoc, à quelques kilomètres de Margaux, Saint-Julien Beychevelle, Moulis... C'est dire que non seulement le vin y est bon (!), mais que la viticulture a toujours été depuis des dizaines ou des centaines d'années le secteur d'activités essentiel de ce village.

Les proéminences mélodiques ont été détectées de manière automatique par la procédure MELISM (Caelen-Haumont et Auran, 2004) intégrée comme MOMEL (Hirst et Espesser, 2000) sous PRAAT. La procédure MOMEL se trouve en amont de la procédure MELISM. La détection automatique a l'avantage de supprimer, comme la contribution de François Poiré en ouverture de ce volume nous le montre bien, les problèmes relatifs à la détection perceptive de telles structures.

L'étude est acoustique au sens où nous n'avons pas effectué de corrections perceptives sur les données brutes. Les corrections effectuées l'ont été de manière automatique par les procédures. Ces corrections concernent les variations micro-prosodiques et la stylisation de la courbe de F0 qui en résulte. D'autres corrections manuelles ont été également apportées, nous y reviendrons plus loin (cf 3., Procédure MELISM, méthode et unités).

Après avoir décrit le corpus et les locutrices, nous aborderons dans cet article la question de la méthode et des unités, puis présenterons ensuite les résultats selon différentes perspectives.

2. L'indice de confiance

Parmi les 16 enregistrements pratiqués à Cussac en juillet 2002, une famille de 4 générations a retenu particulièrement notre attention dans la mesure où ces 4 générations vivent ensemble, dans la même maison familiale attenante au cuvier et aux chais. L'indice de confiance que nous avions défini en 2003 avec la contribution de Joël Pynthe est donc pour ces 4 femmes de 100%.

Nous jugeons opportun de mentionner cet indice et son mode de calcul. C'est un indice de confiance destiné à évaluer empiriquement, dans le contexte national du locuteur, son niveau de représentativité dans la base de données PFC. Cet indice de représentativité est évalué par rapport au critère de "nomadisme" *vs*. stabilité du locuteur et prend en compte la durée du séjour à l'extérieur de son lieu de vie originel, mais aussi la durée de sa réinsertion sur le sol natal. Cet indice ne mesure pas l'écart linguistique, mais relativise simplement la "durée d'exposition linguistique extérieure" par la "durée de récupération" une fois le retour opéré. Cet indice est calculé par les fonctions mathématiques suivantes :

Ic = It - Im

Im = Ie - Ir

où Ic = indice de confiance relatif

It = indice de confiance maximale (ou totale)

Im = indice de métissage

Ie = indice d'exposition extérieure

Ir = indice de récupération

L'indice maximal (Im) est supposé atteint lorsque l'enfant a 15 ans et qu'il est toujours resté sur son lieu natal. L'unité de l'indice (s) est calculé par période de 6 mois : 15 ans équivalent ainsi à 30 s (semestres). Il correspond aussi à 100% de confiance.

L'indice d'exposition extérieur (Ie) est calculé de la même façon en fonction du nombre de semestres passés hors de la "région natale". Ainsi 8 ans passés hors de la région d'origine correspondent à 16 s. Cette région natale de stabilité est évaluée arbitrairement par un rayon de 30 km autour du lieu de naissance et de résidence. L'indice de récupération (Ir) est également calculé par la formule :

$$Ir = Ie \times \alpha \times In$$

où Ie = durée d'exposition linguistique extérieure

α = indice de "recouvrement linguistique"

In = durée dans la Rn depuis le retour.

L'indice α est une fonction linéaire évaluée subjectivement : en 10 ans, quel que soit Ie, la récupération est estimée complète.

	Ie	α	In	$\alpha \times$ In	Ir Ie × a × In	Im Ie - Ir	It 15 × 2	Ic It - Im	*Ic %*
Ie = 3, In = 2	**3**	**0.05**	**2**	**0.1**	**0.3**	**2.7**	**30**	**27.3**	***91***
Ie = 3, In = 3	**3**	**0.05**	**3**	**0.15**	**0.45**	**2.55**	**30**	**27.45**	***91***
Ie = 3, In = 4	**3**	**0.05**	**4**	**0.2**	**0.6**	**2.4**	**30**	**27.6**	***92***
Ie = 3, In = 5	**3**	**0.05**	**5**	**0.25**	**0.75**	**2.25**	**30**	**27.75**	***92***
Ie = 3, In = 6	**3**	**0.05**	**6**	**0.3**	**0.9**	**2.1**	**30**	**27.9**	***93***
Ie = 3, In = 7	**3**	**0.05**	**7**	**0.35**	**1.05**	**1.95**	**30**	**28.05**	***93***
Ie = 3, In = 8	**3**	**0.05**	**8**	**0.40**	**1.2**	**1.8**	**30**	**28.2**	***94***
Ie = 3, In = 9	**3**	**0.05**	**9**	**0.45**	**1.35**	**1.65**	**30**	**28.35**	***94***
Ie = 3, In = 10	**3**	**0.05**	**10**	**0.5**	**1.5**	**1.5**	**30**	**28.5**	***95***
	3	**0.05**	**14**	**0.75**	**2.25**	**0.75**	**30**	**29.25**	***97***
Ie = 3, In = 16	**3**	**0.05**	**16**	**0.80**	**2.4**	**0.6**	**30**	**29.4**	***98***
Ie = 3, In = 20	**3**	**0.05**	**20**	**1**	**3**	**0**	**30**	**30**	***100***
Ie = 10, In = 2	**10**	**0.05**	**2**	**0.1**	**1**	**9**	**30**	**21**	***70***
Ie = 20, In = 20	**20**	**0.05**	**20**	**1**	**20**	**0**	**30**	**30**	***100***

Tableau 1. *Exemples de calcul de l'Indice de Confiance (Ic)*

Ainsi le Tableau 1, en l'absence actuelle d'un script qui le fournirait automatiquement, permet de calculer facilement l'Indice de Confiance (Ic). Ici, des lignes 1 à 12, l'indice d'exposition (Ie) est fixé à 3s (soit 18 mois), et l'indice de durée de retour varie par pas de 2s (ou 1 an). Dans ce cas, l'indice de confiance (Ic), exprimé en pourcentages, correspond respectivement de 91% à 100%. Ligne 14, Ie étant fixé à 10s (60 semestres, 30 ans) pour un temps de retour de 2s (2 semestres, 1 an), Ic est de 70% (cf dernière colonne). Ligne 15, Ie étant de 20s (ou 120 semestres, 60 ans) pour un temps de retour de 20s (20 semestres, 10 ans), l'indice de confiance est de 100%.

Ce calcul est simple : il ne peut en effet être sophistiqué dans la mesure où il est empirique et relativement arbitraire, mais il donne une appréciation globale de la confiance que l'on peut accorder au locuteur potentiel, c'est-à-dire qu'il permet de vérifier rapidement si ce locuteur/locutrice est susceptible d'être retenu(e) on non pour les enquêtes PFC. C'est un point intéressant qui permet d'éviter de recenser des locuteurs non suffisamment représentatifs.

3. Les locutrices

La méthode viticole pratiquée est la méthode traditionnelle, et les connaissances accumulées depuis de nombreuses générations, le soin, la vigilance apportés, font que le vin proposé est de grande qualité : c'est la motivation profonde de ces femmes, sous-jacente à l'ensemble de leurs enregistrements. Comme les générations successives se sont renouvelées par des filles, ce sont 4 viticultrices qui ont été enregistrées. Il s'agit donc de :

- l'arrière grand-mère, HV1, viticultrice, née en 1915 à Cussac, 87 ans à la date de l'enregistrement,
- la grand-mère, SP1, viticultrice, née en 1935 à Cussac, 67 ans,
- la mère, LR1, viticultrice, née en 1961 à Cussac, 41 ans,
- la fille, LR2, en école d'agriculture, née en 1987 à Cussac, 15 ans.

Ces locutrices ont enregistré selon le protocole PFC, à savoir, un texte, une liste de mots, et des conversations, soit libres avec les participants, soit guidées par des questions que posait l'enquêteur. Pour ces locutrices, on dispose donc de 8 textgrids (4 "libres" et 4 "guidés"), qui ont été découpés chacun en 3, en raison de la taille trop importante des fichiers sons correspondants pour le traitement informatique. Ces 3 fichiers ont été identifiés par T1, T2, T3 (par exemple : 33chv1gw_T1.wav). On dispose donc en tout de 24 fichiers parole.

4. Procédure MELISM, méthode et unités

4.1. La procédure MELISM et les unités tonales

La procédure MELISM ayant été décrite précédemment dans le Bulletin PFC n°3 (Caelen-Haumont et Auran 2004a) et par ailleurs (Caelen-Haumont et Auran 2004b), je me contenterai de rappeler les principaux points.

Les proéminences mélodiques ayant été définies par notre communauté PFC comme une priorité de l'analyse prosodique, mon objectif est donc de les décrire de manière très précise dans le cadre d'une phonologie mélodique de surface, afin de dégager si possible des morphèmes et des patrons mélodiques représentatifs, voire invariants.

Plus précisément, le travail porte essentiellement sur les "mélismes[1]", terme repris au domaine du chant (Caelen-Haumont et Bel 2000) pour désigner, dans notre acception, la *forme* acoustique et mélodique d'un item, le plus souvent lexical, isolé ou non. Le terme est donc référencé par rapport au mot et à ses limites. Dans cette étude, le terme de "mélisme" (cf. la note 1 ci-dessous) prend l'acception générale de "motif mélodique plus ou moins complexe, ayant au moins une de ses cibles réalisées dans les niveaux aigus, *a, s, ou h"*.

Les mélismes se définissent par rapport à une échelle tonale, dont l'unité est le demi-ton (échelle logarithmique), leur extension maximale correspondant à l'étendue du registre du locuteur. Il est bien connu que les mots qui possèdent une proéminence mélodique présentent souvent des modulations très particulières en termes de montée, descente et/ou plateau. A ma connaissance le schéma mélodique de ces modulations n'a jamais été exploré, et il m'a semblé qu'il était intéressant de l'étudier. Pour ce faire il a fallu donc concevoir un outil de précision : c'est ce que réalise le logiciel MELISM qui structure le registre de tout locuteur en 9 niveaux, en fonction des points-clefs que sont le F0 moyen, le minimum et le maximum de F0 repérés dans l'ensemble de son enregistrement (ce qui suppose de les déterminer avec soin lorsque l'enregistrement se

[1] En fait, comme l'étude présente vise la description phonologique des proéminences mélodiques (sans visée sémantique et/ou pragmatique), nous avons élargi notre investigation aux proéminences dont la cible mélodique est a, s, h (sans restriction pour *h* comme c'est le cas pour les mélismes).

distribue en plusieurs fichiers son). De la sorte les modulations vont être interprétées en termes de séquences bitonales successives qui, situées dans les limites du mot, sont appelées par commodité “syllabes tonales”.

Le tableau 2 propose un parallèle entre les syllabes lexicales et les séquences mono- ou bitonales du mot *connu*, tel qu’il est réalisé figure 1. On voit clairement qu’il n’y a pas de correspondance entre les 2 syllabes lexicales et les 4 séquences mono- ou bitonales.

Tableau 2. *Syllabes lexicales et syllabes tonales*

mot	connu			
syllabes lexicales	/ko/		/ny/	
séquences mono- / bitonales ou “syllabes tonales”	bc	cH	Hh-	hm

Dans la mesure où ces séquences tonales décomposent le “mot tonal” comme les syllabes lexicales décomposent le mot lexical, et que le terme de “syllabes tonales” est à la fois plus clair et plus court que celui de “séquences mono- / bitonales”, par commodité dans le cadre de cette étude, nous optons pour lui, même si les fonctions morphologiques de ces deux types de syllabes ne sont en rien comparables.

Tableau 3. *Matrice des séquences tonales pour la description des configurations mélodiques des mots, en particulier les proéminences mélodiques subjectives (ou mélismes) sur fond jaune, ayant pour corrélat mélodique, les niveaux les plus aigus*

ton	Mélismes aigu a	supra s	haut h	élevé e	moyen m	centré c	bas b	infra c	grave g
a	*aa*	*as*	*ah*	*ae*	*am*	*ac*	*ab*	*ai*	*ag*
s	*sa*	*ss*	*sh*	*se*	*sm*	*sc*	*sb*	*si*	*sg*
h	*ha*	*hs*	*hh*	he	hm	*hc*	*hb*	*hi*	*hg*
e	*ea*	*es*	eh	*ee*	em	ec	eb	ei	eg
m	*ma*	*ms*	mh	me	*mm*	mc	mb	mi	mg
c	*ca*	*cs*	*ch*	ce	cm	*cc*	cb	ci	cg
b	*ba*	*bs*	*bh*	be	bm	bc	*bb*	bi	bg
i	*ia*	*is*	*ih*	ie	im	ic	ib	*ii*	ig
g	*ga*	*gs*	*gh*	ge	gm	gc	gb	gi	*gg*

La matrice (9 x 9) ci-dessus (cf. Tableau 3) donne une représentation de toutes les séquences bitonales possibles, les cases en grisé plus clair et en italiques correspondant à la définition tonale du mélisme au sens strict.

Les corrélats prosodiques du mélisme sont donc 1°) l’implication des niveaux les plus aigus, soit les niveaux *a* et *s,* ou 2°) une large excursion de F0, atteignant au minimum le niveau *h,* à condition que l’amplitude soit suffisante, excluant pour l’une des cibles de F0 les niveaux *h, e* et *m,* c’est-à-dire supérieure au tiers du registre global du locuteur. Comme nous venons de le préciser, étant donné que notre objectif est purement phonologique, nous n’avons appliqué dans le cadre de cette présente étude aucune restriction au niveau *h,* ce qui

permet de traiter les 3 niveaux mélodiques, *a, s et h,* dans les mêmes conditions. Ces traits sont accompagnés généralement d'un ralentissement sensible du débit et, éventuellement, d'une augmentation importante de l'énergie.

Dans la figure 1 ci-dessous, la fenêtre inférieure présente l'analyse obtenue par la procédure MELISM. Sous la tire de l'étiquetage (*"... qui l'ont connu, bon ..."),* se trouve représentée la tire des séquences bitonales. Le mélisme de *connu,* qui possède deux syllabes lexicales, se réalise en 4 séquences bitonales, c'est-à-dire 4 syllabes tonales (cf. ci-dessous).

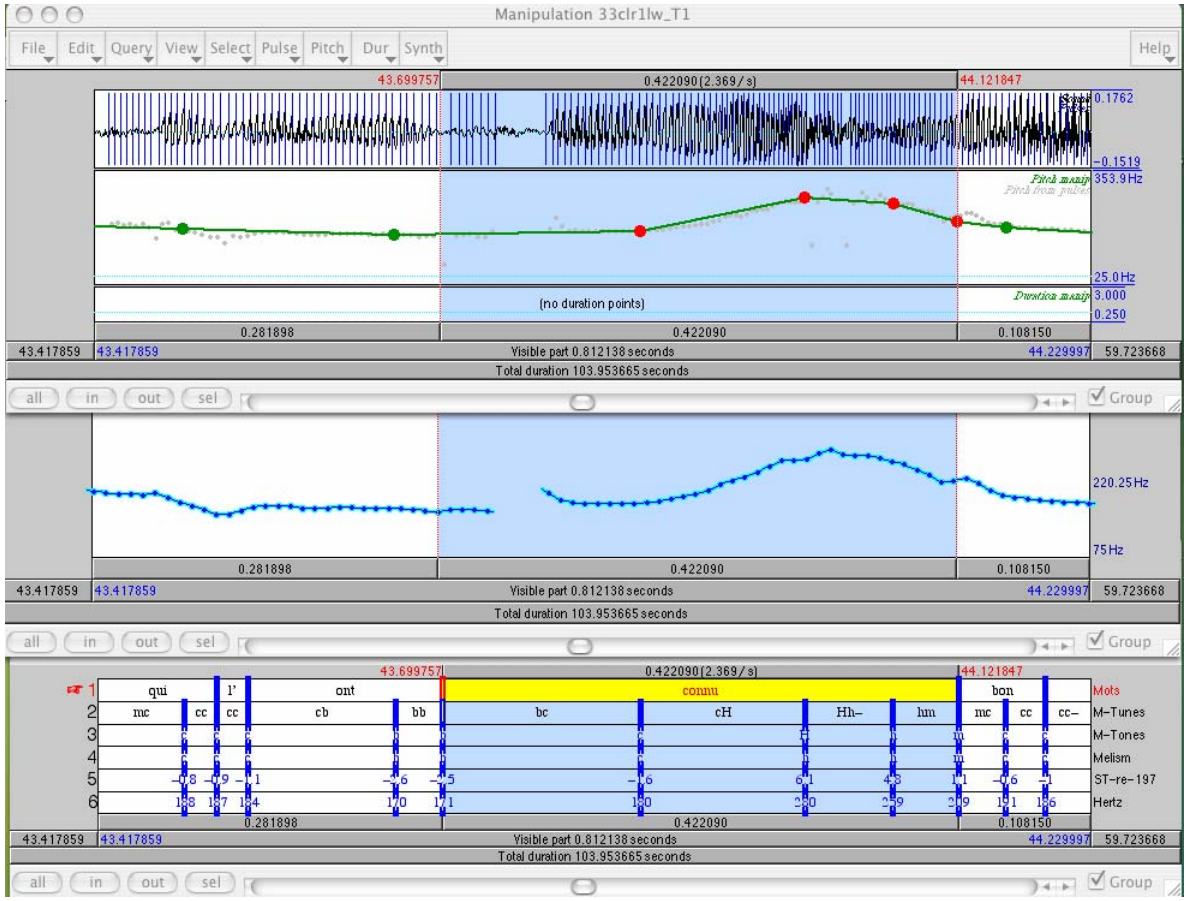


Figure 1. *Exemple de mélisme (/connu/, locutrice LR1_T1)*

De haut en bas : la fenêtre "Manipulation" qui fournit la courbe stylisée de F0, obtenue sous PRAAT *via* MOMEL, après filtrage de la micro-mélodie, avec les points-clefs de la structure du mélisme obtenus à chaque changement significatif d'orientation de la courbe de F0 ; en dessous, la courbe de F0 d'origine (PRAAT), et dessous encore, la fenêtre "MELISM" : de bas en haut, respectivement, les valeurs de F0 en Hertz (MOMEL), leur conversion en demi-tons dans le registre de la locutrice, le codage tonal (premier passage en dehors de toute segmentation), le codage tonal après segmentation et indication de la cible tonale la plus aiguë (en majuscules), et les séquences bitonales résultantes, appelées encore "syllabes tonales".

4.2. Obtention des données

Bien que le calcul des cibles tonales et le découpage en syllabes tonales soit automatique, toutefois, une série d'opérations manuelles reste à ce jour nécessaire :

1. en vue de l'harmonisation des différents fichiers son relatifs à une même séquence d'enregistrements pour le même locuteur, il convient en premier

lieu de repérer soigneusement F0 moyen (moyenne des F0 moyens), *minimum* (des *minima*) et *maximum* (des *maxima*),

2. et ce, sous certaines contraintes : le repérage des valeurs fausses de F0 (signal de mauvaise qualité surtout en fin d'émission, ou bruité, décrochage sur un harmonique inférieur ou supérieur, F0 résiduel dans les consonnes dévoisées ...).

Ces étapes, il est vrai, sont longues et fastidieuses. La méthode ne se satisfait pas en effet d'un demi-travail, mais cette exigence garantit par la suite des données d'excellente qualité, et très homogènes.

1. une fois nanti d'une courbe de bonne qualité, on procède à un premier passage de la procédure qui permet de cibler tous les pivots mélodiques significatifs.
2. on relève ensuite les passages possédant les valeurs qui entrent dans la définition du mélisme (niveaux *a, s et h* avec ou sans conditions de restriction), et l'on segmente le (ou la suite de) mot(s) mélismé(s) avec leur contexte.
3. on soumet alors une deuxième fois l'ensemble de ce fichier "parole" à la procédure, et l'on obtient alors le codage tonal dans le cadre de chaque mot.
4. dans le même temps s'affiche la fenêtre "Manipulations" qui permet de modifier la courbe de F0. Il s'agit à ce moment-là de supprimer les éventuelles cibles mélodiques superflues, et cette opération s'effectue en vérifiant à l'écoute que la F0 reste conforme à l'original.
5. un dernier passage de la procédure délivre les cibles et les syllabes tonales définitives. Chaque passage de la procédure est proche du temps réel.

Cette analyse porte uniquement sur le paramètre de la fréquence fondamentale. Toutefois le paramètre du temps, qu'il serait intéressant de joindre à cette étude, reste disponible dans la mesure où d'une part le codage dans le Textgrid est référencé dans le temps jusqu'au niveau même de la syllabe tonale, et d'autre part où les mélismes ont fait l'objet d'un relevé dans une banque de données (cf. Tableau 4), avec certains de leur critères comme le contexte avant/arrière, l'indication du temps aux frontières gauche et droite, la suite des syllabes tonales, la suite des cibles avec leur valeur correspondante en demi-tons.

- (y'avait)								
un		52.58	b -3					b -3.2
égrappoir	**M** **bc ch hH**	**52.65** **53.17**	**b** **-3.2**	**c** **-0.5**	**h** **5.6**			**H** **5.6**
dessus	**MF** **he eS Ss sh**	**53.17** **53.64**	**h** **5.6**	**e** **4.4**	**S** **8.5**	**s** **8.1**		**h** **5.9**
et	**Mc** **Hb bb**	**53.64** **53.77**	**h** **5.9**	**b** **-1.7**				**c** **-1.6**
(et les hommes								

Tableau 4. *Extrait de la banque de données, locutrice SP1. De gauche à droite, la liste des mots mélismés avec leur contexte, la catégorie de mélismes (ici M, mélisme interne, MF, mélisme final de groupe sans pause, Mc, mélisme par contact) et le mot tonal avec ses syllabes, les temps aux frontières des mots, la cible de gauche avec sa valeur en demi-ton, les cibles internes, et la cible finale*

5. Résultats

L'étude de 2004 comportait les données de 3 locutrices, enregistrements libres, et comprenait 84 items. L'étude présente analyse les données des 4 locutrices des enregistrements libres et guidés, soit 294 structures, qui comportent de 1 à 5 syllabes tonales.

5.1. Choix de la méthode d'analyse

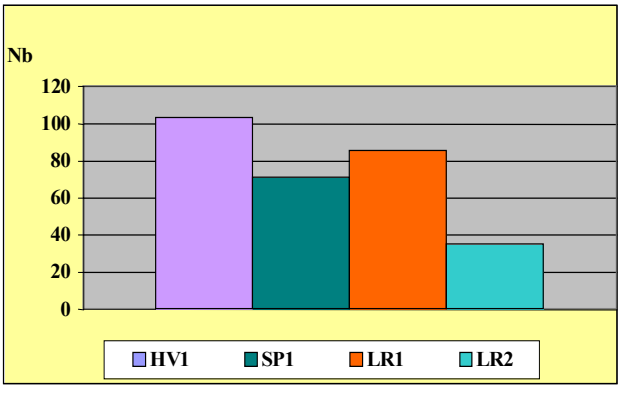
Devant la variété de ces structures, il est certain que plusieurs méthodes d'analyse sont possibles. On peut par exemple les étudier en fonction de leur forme globale : pente linéaire montante, descendante, pentes alternées (chapeau droit ou inversé). On peut aussi les analyser par leur degré d'amplitude : du plateau jusqu'à la pente la plus contrastive.

Pour ma part, j'ai préféré choisir une autre méthode qui recourt aux notions de "famille" et de "génération de structures". Il y a en effet plusieurs intérêts à procéder de cette façon. Tout d'abord elle intègre les méthodes précédentes. Par ailleurs elle a le grand avantage de réduire la variété des structures en les rattachant à des structures simples. Enfin elle privilégie deux points fondamentaux, les frontières gauche et droite du mélisme, qui ont des caractéristiques intéressantes : en effet ces points appartiennent à la fois à la courbe intonative (plan de l'énoncé, du groupe prosodique, de l'organisation syntagmatique), et à la fois au mélisme, c'est-à-dire à l'expression subjective, *hic et nunc*, du locuteur, laissée à sa liberté, fonction de son intention sous-jacente, consciente ou pas.

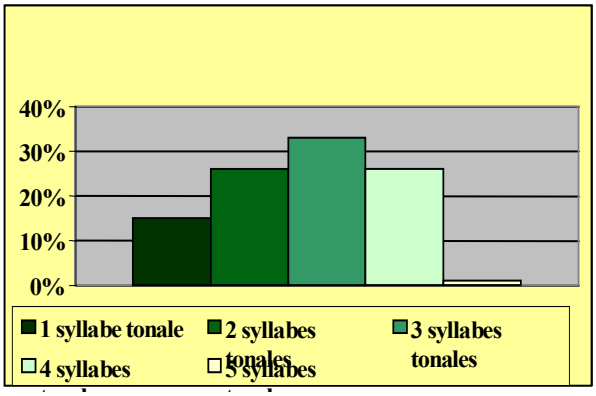
Nous aborderons l'analyse des familles tonales au paragraphe 5.7. Pour le moment, nous commençons par une description générale de ces structures.

5.2. Nombre de mélismes par locutrice

Le graphique 1 ci-dessous montre la répartition des 294 items par locutrice. On constate que le nombre d'items est nettement inférieur chez LR2, la locutrice la plus jeune, ce qui représente *a priori* un handicap pour mener des comparaisons entre interlocutrices. Une priorité de l'étude à venir est donc d'harmoniser le nombre de structures entre les locutrices, en débordant le cadre strict des codages phonologiques, liaison et /ə/, et en allant chercher d'autres séquences de parole dans les enregistrements libres et guidés. Le projet est de porter le nombre à 100 structures par personne.



Graphique 1. *Nombre de structures proéminentes par locutrice*



Graphique 2. *Pourcentages de structures proéminentes en fonction du nombre de syllabes tonales*

5.3. Pourcentages de structures en fonction du nombre de syllabes tonales

Le graphique 2 ci-dessus s'applique à l'ensemble des items, toutes locutrices confondues. On remarque une majorité de structures à 3 > 2 et 4 syllabes tonales. Le nombre des structures à 5 syllabes est infime (4 items).

5.4. Corrélations

Ayant précédemment illustré le domaine des relations entre syllabes lexicales et syllabes tonales, nous précisons maintenant la relation quantitative entre elles (Tableau 5).

Tableau 5. *Corrélations*

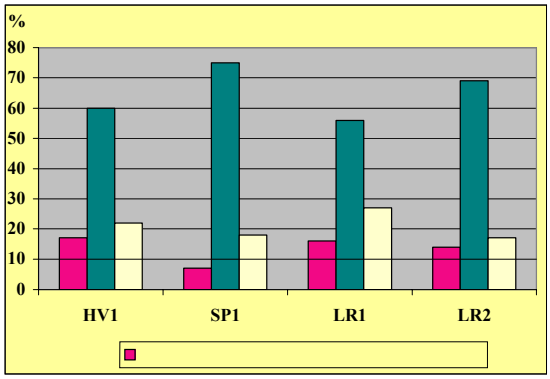
nombre de syllabes tonales / nombre de syllabes lexicales	0.50
nombre de syllabes lexicales / différence de niveaux mélodiques	0.40
nombre de syllabes tonales / différence de niveaux mélodiques	0.65

Nous précisons que la différence de niveaux mélodiques est calculée sur l'amplitude mélodique maximale dans le mot. Il apparaît dans ce tableau qu'il n'existe pas de corrélations entre les deux types de syllabes (0.50). La différence de niveaux mélodiques concerne le nombre de niveaux dans la structure. Il est intéressant de savoir si le nombre de niveaux est fonction du nombre de syllabes.

Concernant les syllabes lexicales, nous ne sommes pas étonnés de constater que ce nombre n'est pas corrélé (0.40) à celui des syllabes lexicales puisque les deux types de syllabes ne sont pas elles-mêmes corrélées. Parallèlement on constate qu'il existe un coefficient plus élevé (0.65) pour les syllabes tonales : en effet il semble vraisemblable que le nombre de niveaux augmente avec le nombre de syllabes tonales, mais ce n'est pas une corrélation très forte, ce qui indique sans doute que parmi les structures qui ont le plus de syllabes tonales, il en existe un certain nombre en plateaux.

5.5. Pentes et plateaux

Une étude simple consiste à répertorier au niveau des familles tonales (aux frontières des mots) le nombre de pentes ascendantes, descendantes et de plateaux, toutes locutrices confondues. Nous trouvons une majorité de pentes ascendantes (187, 64%), dépassant largement le nombre des descendantes (65, 22%), et les plateaux (42, 14%). Une analyse plus fine mène à la répartition par locutrice (graphique 3) :



Graphique 3. *Distribution des pentes et plateaux en fonction des locutrices*

Les pourcentages étant calculés par rapport au total de chaque locutrice, on constate tout d'abord que les résultats sont globalement homogènes entre les locutrices, et secondairement que SP1 a proportionnellement le plus de pentes ascendantes, LR1 le moins. Inversement, cette dernière possède le plus de pentes descendantes. Nous verrons plus loin si le nombre de pentes ascendantes *vs*. descendantes est en relation avec le nombre de pauses, comme ce résultat le suggère.

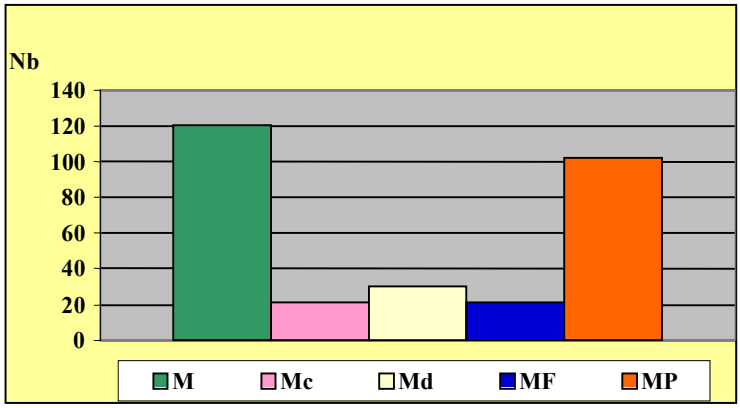
5.6. Types de structures

Les structures recouvrent en fait plusieurs catégories susceptibles d'avoir des comportements prosodiques, et en particulier mélodiques, différents. Nous avons donc distingué 5 catégories différentes qui tiennent principalement à la position dans le contexte :

- internes au groupe syntagmatique (M). Il s'agit ici de l'archétype du mélisme comme expression subjective du locuteur, car il n'est pas déterminé par des contraintes syntaxiques ou de voisinage.
- par contact (Mc) d'un mélisme antérieur ou postérieur (il s'agit souvent de mots grammaticaux antérieurs au mélisme, mais aussi postérieurs dans le cas d'un enchaînement avec un syntagme prépositionnel ou un SN2).
- de discours indirect (Md). Ces types de mélismes ont été réalisés par une seule locutrice, HV1.
- situé en fin de tour de parole (MF) sans pause, devant une pause bruitée *(euh),* ou encore en fin de syntagme enchaîné sans pause avec le suivant.
- situé devant pause silencieuse (MP).

Les deux derniers types sont déterminés le plus souvent par des contraintes syntaxiques, auxquelles d'ailleurs peuvent s'ajouter des contraintes subjectives. On peut repérer ce double processus par le biais d'une analyse sémantique et/ou pragmatique. Sur le plan quantitatif, il est bien difficile de démêler la quantité qui revient à chacun de ces plans. Il est probable d'ailleurs que ces quantités ne soient pas cumulatives, car l'un et l'autre plan sont susceptibles de produire des proéminences mélodiques dans les aigus. Ceci signifie, semble-t-il, que les MP à caractéristique subjective peuvent être étudiés sous le même rapport que les M, bien qu'ils possèdent le double statut.

Ces différents types n'ont pas tous la même proportion. Le graphique 4 ci-dessous en indique la population :

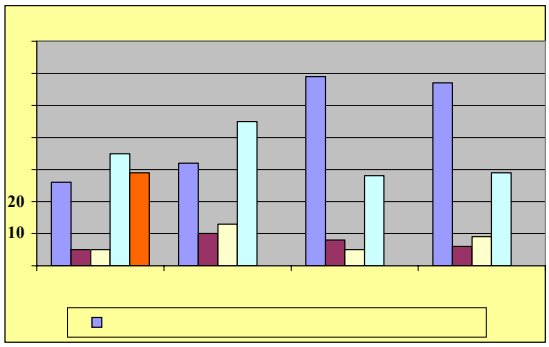


Graphique 4. *Nombre de structures proéminentes par type : M, interne au syntagme, Mc, par contact, Md, de discours indirect, MF, de fin de tour de parole sans pause, MP, devant pause*

Comme on le voit sur le graphique 4, les proéminences internes et devant pause sont de 4 à 6 fois plus nombreuses que les autres types, qui sont par ailleurs en nombre sensiblement égal (de 20 à 30 items). Ce sont les mélismes internes qui sont les plus nombreux. Il est probable que si nous avions affaire à de la parole lue, les proéminences mélodiques devant pause auraient été les plus nombreux.

L'étude ici étant purement descriptive des phénomènes mélodiques, c'est-à-dire sans analyse sémantique ni pragmatique, nous ne distinguerons pas les MP subjectifs des autres.

Le graphique 5 ci-dessous présente la répartition de ces types de structures par locutrice. La cinquième barre chez HV1 est celle des items de discours indirect (Md). Chez cette locutrice, on note que les Md sont plus nombreux que les M. Pour que la comparaison soit plus probante, nous avons ramené les pourcentages au total de chaque locutrice. Il apparaît par ailleurs que le rapport nombre de M / nombre de MP est inversé en fonction de l'âge : dans la limite de nos observations, il semble que chez les locutrices les plus âgées, HV1 et SP1, d'une part la disparité des effectifs est moins grande entre les M et les MP, d'autre part la différence se trouve au profit des MP, plus nombreux, alors que l'inverse est constaté chez les plus jeunes, LR1 et LR2, le contraste quantitatif entre M et MP étant plus important chez les plus jeunes au bénéfice des M. Une hypothèse qui pourrait rendre compte de cette inversion de tendance entre locutrices plus jeunes et plus âgées pourrait être une capacité respiratoire différente, ou une ressource énergétique plus ou moins forte, ou plus ou moins bien contrôlée. *Ainsi l'hypothèse d'un invariant relatif à l'âge pourrait être posée ici.*



Graphique 5. *Répartition des structures proéminentes par locutrice*

Il apparaît par ailleurs que proportionnellement SP1 possède le plus de MP, et inversement LR1, le moins. Mais cette dernière locutrice se distingue surtout par la proportion de ses M, et par rapport aux autres, et parce qu'elle maximise au mieux le rapport nombre de M / nombre de MP. Ceci va de pair avec la remarque précédente (§ 5.5) concernant le nombre de pentes ascendantes / descendantes chez SP1 et LR1 : si SP1 a le plus de pentes ascendantes, c'est parce qu'elle compte le plus grand nombre de MP (et inversement pour LR1).

Nous avons poussé l'analyse au niveau des types secondaires (Mc, MF, Md) pour trouver si leurs caractéristiques mélodiques pouvaient les rattacher à la catégorie soit des M, soit des MP. Une hypothèse est de rattacher a priori les MF aux MP, puisque la grande majorité d'entre eux se trouvent en fin de groupe comme les MP. En fait nous avons constaté que les Mc, Md et MF ne montrent pas de différence significative dans leur répartition dans les 9 niveaux mélodiques par rapport aux M. Tous montrent un étalement plus important. Inversement, ces 4 catégories s'opposent toutes à celle des MP qui montrent une concentration dans les zones les plus aiguës. *Ceci tendrait à signifier que ce n'est pas*

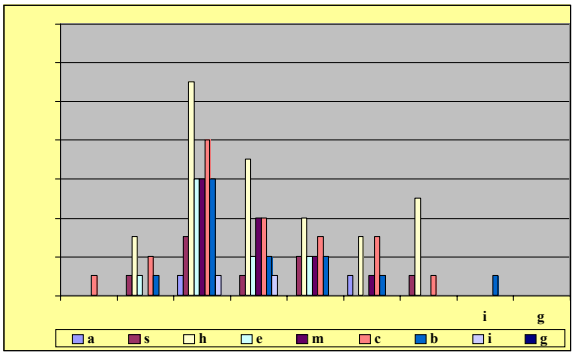
la catégorie linguistique qui détermine des corrélats mélodiques, mais simplement l'effet physiologique "pause".

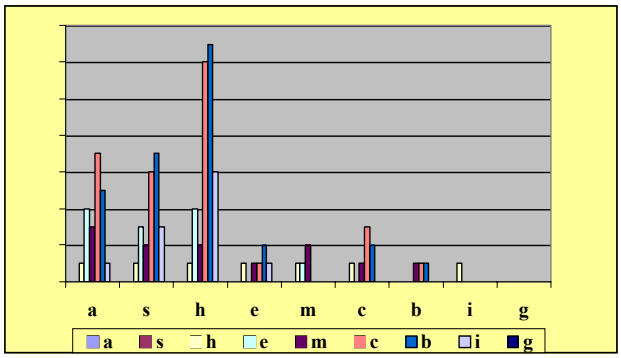
Par ailleurs nous avons cherché à savoir si la répartition différait encore en fonction des pentes ascendantes et descendantes, et les résultats ont montré que ce n'était pas le cas : en effet, d'une part il y a très peu de MP de pente descendante, et d'autre part les M ne montrent pas de différence dans leur répartition entre les niveaux mélodiques, qu'ils appartiennent à une pente ascendante ou descendante.

De ce fait il n'est pas utile de fragmenter l'analyse en catégories plus fines. Nous avons donc regroupé en 2 tableaux distincts les mélismes, selon qu'ils étaient réalisés avec pause (MP : 102 items, graphique 7) ou sans pause (M + Mc + MF + Md, 192 items, graphique 6).

On constate dans ces deux graphiques que la distribution est sensiblement différente : hors pause l'étalement est plus grand, les séquences tonales groupent davantage leurs effectifs de manière plus centrale *(h>e>m>c)* ; devant pause, les effectifs se concentrent davantage dans les régions les plus aiguës *(h>a, s)*. Dans tous les cas cependant, c'est la séquence tonale *h* (pente ascendante, descendante, plateau) qui a le plus d'effectifs.

Concernant les proéminences devant pause, la grande majorité (85%) est de type ascendant, les types descendants (8%) et plateaux (7%) équilibrant leurs effectifs. On retrouve les résultats depuis longtemps connus dans la littérature, à savoir que l'amplitude mélodique est généralement large (de fait 76% des MP ascendants partent des niveaux *b>c>i*, classés ici par ordre décroissant d'effectifs, pour atteindre les niveaux *h*, *s* et *a*).





Graphique 6 **Graphique 7**

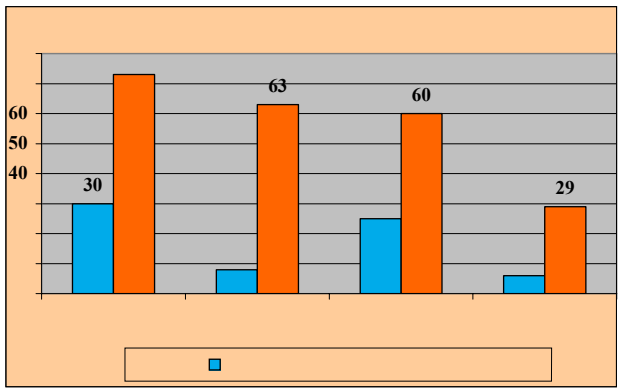
Pourcentages de structures dans les séquences bitonales (ou familles tonales) selon qu'elles ne sont pas suivies d'une pause (graphique 6), ou qu'elles se situent devant une pause (graphique 7).

Il est intéressant de constater toutefois un manque de symétrie concernant les MP descendants et en plateaux, car on n'a recensé aucune donnée dont le point de départ se trouvait dans les niveaux *a* et *s*. Chez ces derniers (MP), tout se passe entre les niveaux *h* et *c* pour les descendants, et *h* et *b* pour les plateaux : donc restriction de la plage mélodique et centrage vers les zones moyennes basses. *C'est une remarque intéressante pour l'invariant qu'elle suggère, et qu'il faudrait vérifier sur une étendue plus importante du corpus.*

5.7. Mots lexicaux et grammaticaux proéminents

Les proéminences se répartissent encore selon leur type morphologique : d'une part les mots lexicaux, de loin les plus nombreux (225 ML), et les mots grammaticaux (69 MG). Ces derniers désignent les déterminants, les pronoms personnels clitiques, les prépositions,

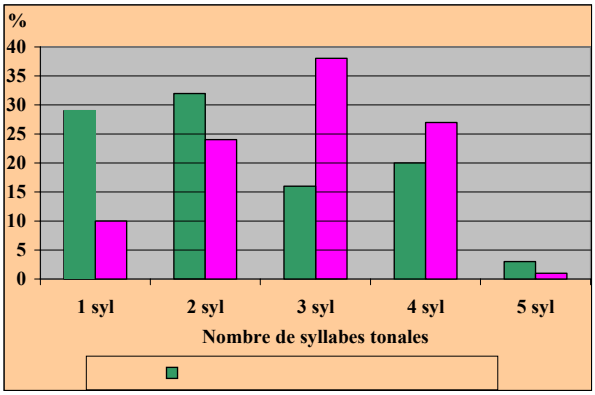
les conjonctions. Il est important de préciser que dans l'ensemble les mots grammaticaux sont proéminents par contact (catégorie Mc) avec un mot lexical.



Graphique 8. *Nombre de mots lexicaux et grammaticaux par locutrice*

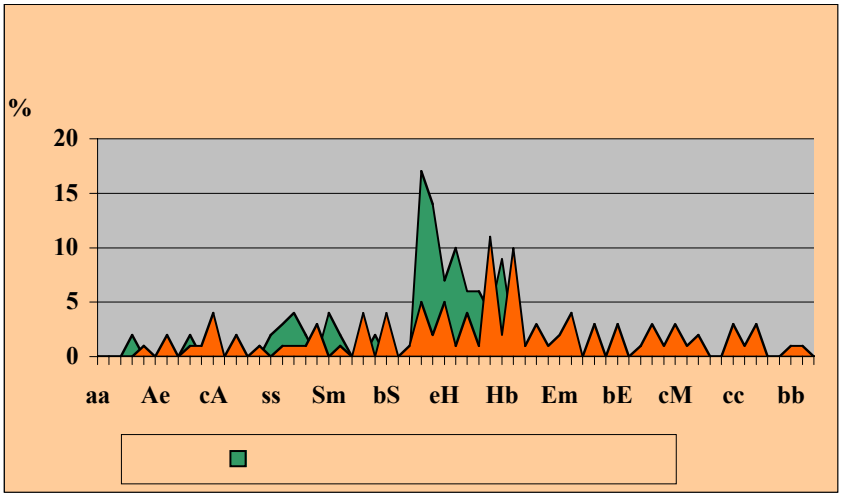
La répartition par locutrice (graphique 8 ci-dessus) montre une certaine disparité dans la répartition. Par exemple SP1 possède proportionnellement le plus grand nombre de ML (ML / MG, 89% / 11%), et inversement LR1 et HV1, à égalité, le plus de MG (MG / ML 29% / 71%). Ceci signifie que SP1 cible avec plus de précision ses proéminences que les autres locutrices qui, inversement, ont pour le traitement de ces proéminences une visée linguistique mélodiquement plus englobante. Il faudrait vérifier avec plus de données relatives si cette caractéristique de SP1 se confirme.

Dans le graphique 9 ci-dessous, nous proposons une répartition des ML et MG en fonction du nombre de syllabes tonales. Si les données se répartissent, pour ces deux types morphologiques, de 1 à 5 syllabes, les ML montrent une préférence pour 3 > 4 > 2 syllabes tonales. Les MG possèdent une distribution un peu mieux équilibrée, mais toutefois avec des effectifs plus importants en 1 et 2 syllabes. On voit bien sur le graphique l'alternance joliment symétrique qui joue entre les MG (1et 2 syllabes tonales) et les ML (3 et 4 syllabes).



Graphique 9. *Répartition des mots lexicaux (ML) et grammaticaux (MG) en fonction du nombre de syllabes tonales*

Une petite étude statistique permet de mieux caractériser ces deux types morphologiques. Le nombre moyen de syllabes tonales est respectivement, pour les MG et ML, de 2.36 et 2.85. Plus précisément, 61% des MG ont 1 ou 2 syllabes tonales, contre 34% pour les ML. Par ailleurs 86% des MG possèdent une cible *h,* contre 48% pour les ML.



Graphique 10. *Répartition des structures dans les catégories des mots grammaticaux et lexicaux par familles tonales. Pourcentages ramenés au total respectif des MG et ML*

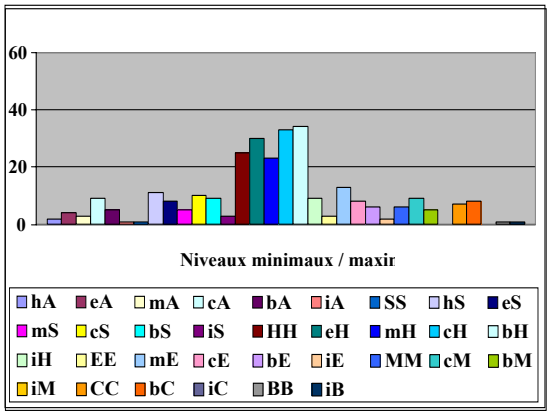
En effet, dans le graphique 10 ci-dessus, nous voyons que 1°) l'étalement de l'amplitude (décrite par les familles tonales) aux frontières dans les mots lexicaux est plus important que pour les mots grammaticaux, 2°) les mots grammaticaux sont davantage ciblés autour du niveau *h,* avec une amplitude faible, centrale, en particulier *hh* et *he,* 3°) au sein des mots lexicaux, le niveau *h* est également une cible un peu plus fréquente que les autres niveaux, mais la tonalité est plus grave en *ch* et *bh,* ce qui signifie en fait une amplitude plus importante.

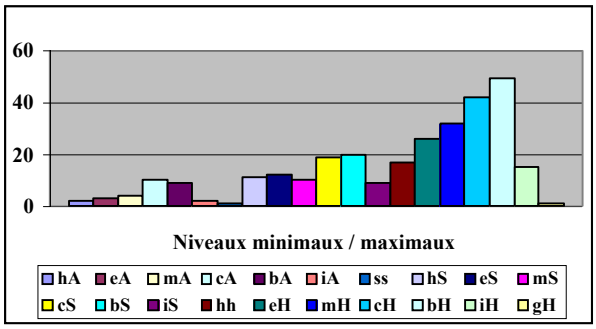
6. L'amplitude mélodique

Nous avons principalement examiné jusqu'à présent les familles tonales, c'est-à-dire les niveaux mélodiques aux frontières du mot, déterminant un type d'amplitude, relative à la courbe intonative. Comme pour les corrélations (paragraphe 5.4. ci-dessus), nous nous intéressons maintenant à l'amplitude mélodique maximale réalisée au sein de ces structures proéminentes. Cette amplitude n'a pas de localisation précise, elle est juste au sein de la structure, frontières éventuellement comprises.

Du fait que pour l'amplitude maximale, nous avons obligatoirement un des deux niveaux qui est *a, s* ou *h,* nous nous attendons à une amplitude plus grande.

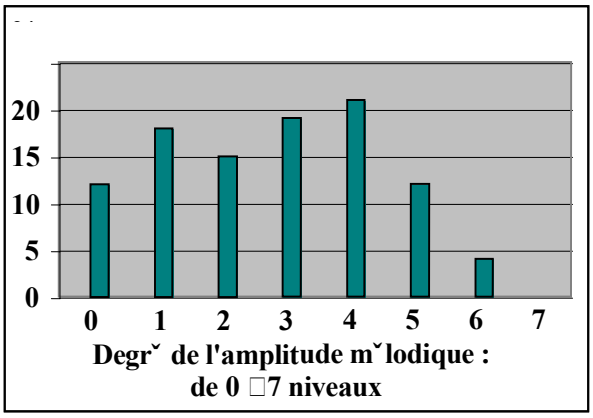
La question est donc la suivante : comment dans les plages communes (comprenant *a, s* ou *h*) se réalise l'étalement de l'amplitude juste aux frontières et au sein du mot ? Existe-t-il un créneau mélodique restreint commun, et si oui, dans quelle plage ?

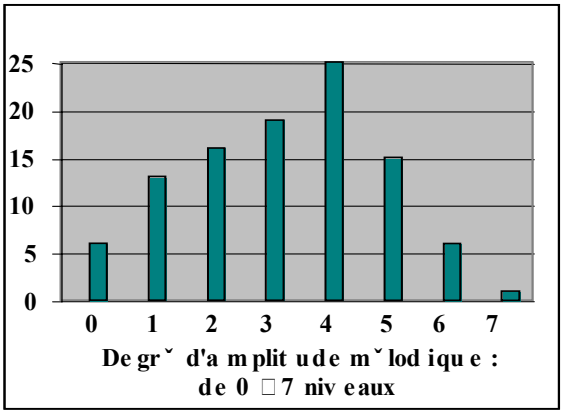




Graphique 11. *Amplitude aux frontières* **Graphique 12.** *Amplitude maximale*
Amplitude mélodique dans les 294 structures proéminentes, aux frontières (Graphique 10), maximale dans la structure (Graphique 11).

De l'examen des deux graphiques 11 et 12 ci-dessus, il ressort que 1°) l'étalement aux frontières est de fait un tiers plus grand (33 degrés d'amplitude *vs*. 20), 2°) la plage préférentielle de l'amplitude est strictement la même, celle du niveau *h,* réalisée majoritairement avec $e < m < c < b$, que la pente soit croissante ou décroissante, 3°) si les plages préférentielles sont identiques, cela signifie qu'il existe une première relation, relativement stable, entre *minima* et *maxima* mélodiques, mais aussi, semble-t-il, une deuxième relation entre la courbe intonative locale (amplitude aux frontières du mot) et le mélisme (amplitude maximale interne au mot). Cette relation est certainement liée au rapport efficace effort / discrimination, c'est-à-dire à l'appropriation que réalise le locuteur de sa langue. *Nous posons là une présomption d'invariant, sans doute purement articulatoire, mais nous ne pouvons pas préciser s'il est propre à cette famille, à la région ou au français.*





Graphique 13. *Amplitude aux frontières* **Graphique 14.** *Amplitude maximale*
Comparaison de la répartition des structures proéminentes en fonction du degré d'amplitude croissant, toutes données confondues : Graphique 13 (amplitude aux frontières, 225 items), Graphique 14 (amplitude maximale, 294 items)

Pour compléter ces informations, nous cherchons maintenant à comparer les degrés d'amplitude entre ces deux localisations, exprimés en écart de niveaux. Pour ce faire, nous envisageons la même plage mélodique, c'est-à-dire toutes les données qui comportent au moins un niveau *a, s* ou *h.* Ceci explique la disparité des effectifs car toutes les structures tonales ne possèdent pas de cible *a, s* ou *h.* Les graphiques 13 (amplitude aux frontières,

225 items) et 14 (amplitude maximale, 294 items) ci-dessus montrent de manière comparative la répartition, en pourcentages, des données en fonction de l'amplitude croissante. On observe dans les deux cas une large répartition au moins sur 6 niveaux, présentant graphique 14 une distribution en cloche, avec un pic pour les 4 niveaux d'écart, et graphique 13 une répartition globalement un peu plus forte dans les écarts plus faibles. Cependant on ne constate pas de réelle disparité entre les données des deux graphiques.

7. Les familles tonales et génération

Au paragraphe 5.1., nous avons précisé que ce serait les familles tonales qui nous fourniraient le cadre de l'analyse présente, et c'est bien ce qui a été réalisé. Nous voudrions maintenant revenir sur cette notion et étudier plus à fond ces familles.

Locutrices	Types	1 syll tonale	2 syll	3 syll	4 syll	5 syll	Mots
HV1	Mc	eH					la
HV1	Md	eH					ma
LR1	M		eH Hh				vrai
SP1	M		eH Hh				non
SP1	M		ee eH				aussi
SP1	M		ee eH				non
LR1	M		ee eH				alors que
LR2	M		eh hH				plus
LR1	M			ee eH Hh			non
HV1	MP			em mH HH			mari
LR1	MP			em mH HH			petite
SP1	MP				em mc cH HH		ˇgrappaient
SP1	M			eh hS Sh			(c'est-□-)dire
HV1	Md				ee eh hS Sh		dˇjeuneras
HV1	M				*ee em mS Sh*	ee ec cm mS Sh	caoutchouc
SP1	MP				eb bm mS Sh		ˇgrapper

Tableau 6. *Exemple d'une famille tonale.*
De gauche à droite, le code des locutrices, des types de proéminences, puis les structures tonales de une à 5 syllabes, et les mots correspondants

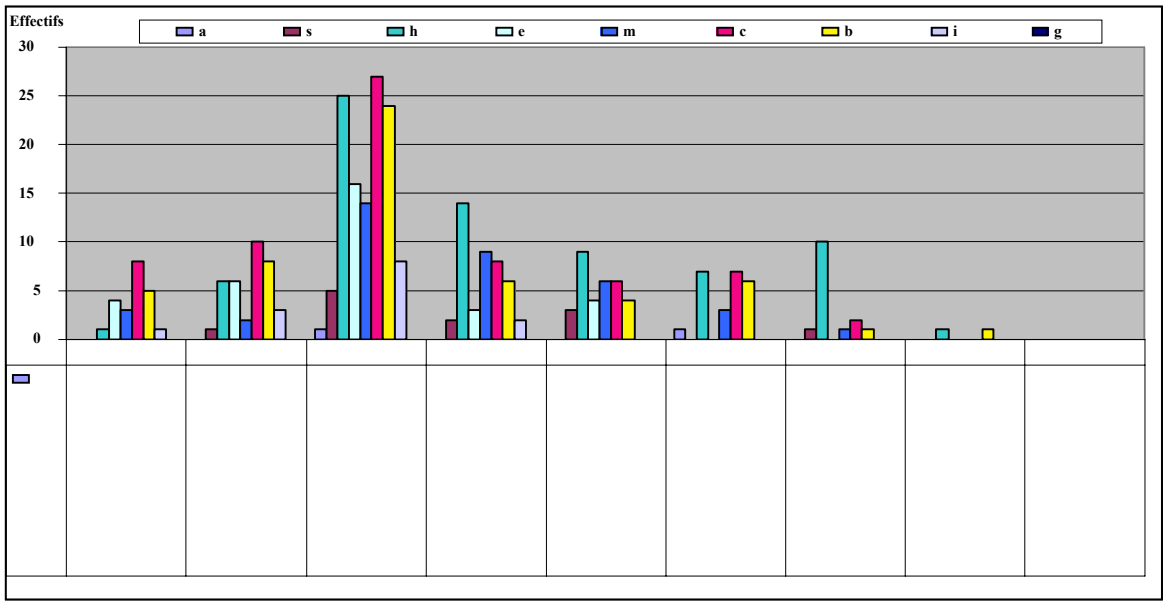
Le tableau 6 nous présente un exemple parmi d'autres de famille tonale (16 items). Comme toutes les familles suffisamment étoffées, cet exemple transcende respectivement, de la première à la dernière colonne, toutes les locutrices, les types de proéminences, les mots grammaticaux et lexicaux. Ainsi, depuis les tons réalisés aux frontières formant une "famille tonale", points appartenant, comme on l'a vu, à la fois à la courbe intonative et au mélisme, on a dérivé des structures par ajout successif d'une cible (en italiques), formant avec la cible précédente ou suivante une nouvelle syllabe tonale, obtenant ainsi une arborescence tonale comme suit :

eH		*ma*	HV1
e*h h*H		*plus*	LR2
e*m m*H HH		*petite*	LR1
	em m*c c*h HH	*égrappaient*	SP1

Les majuscules ne rentrent pas en compte dans la méthode, elles indiquent simplement la (ou les) cible la plus haute. La cible correspond au changement d'orientation de la courbe mélodique, après filtrage de la micromélodie par MOMEL. Les syllabes tonales (ex. : *eh, hH, mH, mc* ...) sont présentées dans l'exemple ci-dessus séparées par un espace symbolisant une frontière tonale. Ces cibles sont attestées dans notre corpus ou reconstituées en amont depuis une structure plus élaborée. Dans ce cas, tableau 6, la

structure présumée est en italiques. Le nombre de syllabes tonales détermine le niveau de profondeur : dans l'exemple de la famille *eH,* le niveau de profondeur est de 5.

Cette méthode permet en particulier de constituer des familles de structures translocuteur, d'objectiver des proximités acoustiques (et sans doute perceptives), de réduire et d'expliquer dans une certaine mesure, la variabilité, et de dresser à terme un inventaire caractéristique d'un locuteur, d'une famille, d'une région, d'une langue.

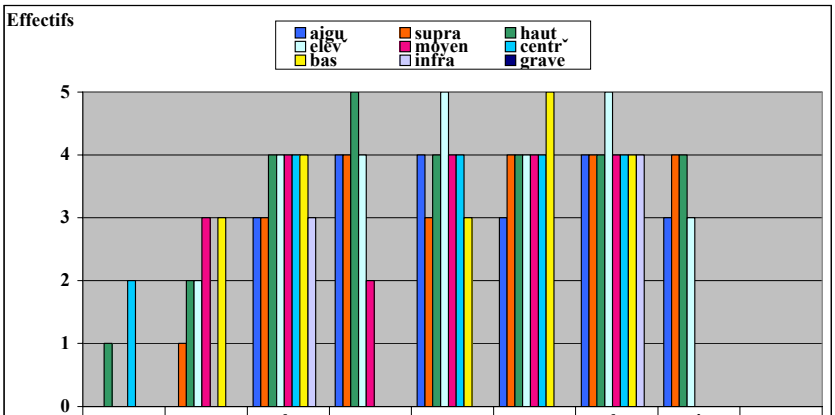


Graphique 15. *Effectifs des familles tonales sur 294 items, toutes locutrices confondues*

Le graphique 15, avec la matrice correspondante, présente une répartition des effectifs en fonction des familles tonales, c'est-à-dire, comme on le sait, en fonction des niveaux mélodiques réalisés aux frontières du mot. Nous rappelons aussi que toutes ces structures comportent au moins un des 3 niveaux mélodiques aigus, *a, s* ou *h,* et pas nécessairement aux frontières.

Il apparaît très nettement sur ce graphique que *le niveau h est à nouveau le plus représenté,* croisé avec les niveaux *h, c* et *b.* Par ailleurs une régularité se produit avec ces niveaux : les niveaux *c* et *b* possèdent en effet des effectifs un peu plus nombreux avec les niveaux aigus *(c>b* avec *a, s* et *h),* le niveau *h* avec les autres *(e, m, c, b),* le niveau *h* recueillant respectivement les effectifs les plus nombreux (avec *h, c, b*), à l'intersection d'une dynamique vers les niveaux aigus et d'une autre vers les plus graves.

Nous nous intéressons maintenant à la profondeur des familles tonales, c'est-à-dire à la répartition du nombre des syllabes tonales au travers de leur famille (dans les structures proéminentes).



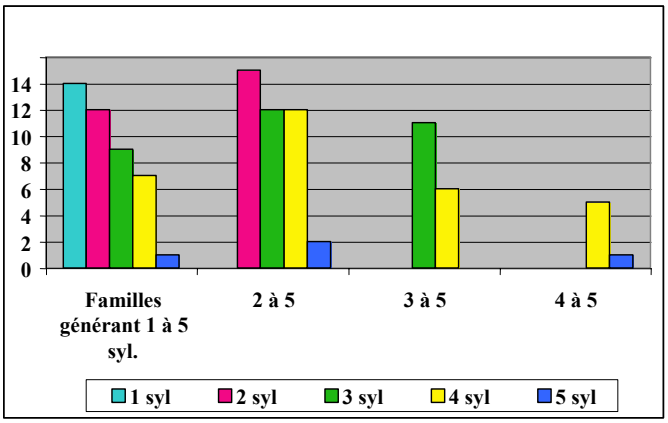
	a	s	h	e	m	c	b	i	g
aigu	0	0	3	4	4	3	4	3	0
supra	0	1	3	4	3	4	4	4	0
haut	1	2	4	5	4	4	4	4	0
elev˘	0	2	4	4	5	4	5	3	0
moyen	0	3	4	2	4	4	4	0	0
centr˘	2	0	4	0	4	4	4	0	0
bas	0	3	4	0	3	5	4	0	0
infra	0	0	3	0	0	0	4	0	0
grave	0	0	0	0	0	0	0	0	0

Graphique 16. *Profondeur des familles tonales (294 items), toutes locutrices confondues*

Comme on le lit sur le graphique 16, les familles les plus profondes se recrutent plutôt dans les plages du registre centrales et moyennes basses : eH, mE, Cb, bE. Ceci est à mettre au compte sans doute de plusieurs facteurs, à savoir que d'une part statistiquement, le locuteur a plus de chances de passer par les niveaux moyens de son registre que par les extrêmes, d'autre part que ces points correspondent aux points d'ancrage de la courbe intonative, globalement plus centrale, et enfin, en corrélation avec ce qui précède, que l'énergie la moins coûteuse se trouve dans le registre moyen du locuteur : ajouter un niveau dans son registre de base peut contribuer en effet à augmenter la complexité d'un motif mélodique.

Un dernier développement sur cet aspect génératif permet de comptabiliser le nombre de structures proéminentes en fonction de leur fécondité, c'est-à-dire leur capacité à générer, dans la limite de nos observations, un nombre de syllabes tonales. En clair la question que l'on pose est de savoir s'il existe plus de familles générant de 1 à 5 syllabes tonales, que de 3 à 5 syllabes par exemple. En effet le "chant" mélodique, plus ou moins étendu, complexe, peut caractériser par exemple un groupe social comme les premiers travaux référencés en début de cette présentation l'ont bien montré.

Les familles tonales sont au total au nombre de 46 (sur les 81 possibles de la matrice 9 x 9, cf. le tableau 2). Le graphique 17 donne une image représentative de la composition de nos données, exprimées en nombre de familles.



Graphique 17. *Effectifs de familles tonales en fonction de leur pouvoir génératif*

Comme on le constate sur ce graphique, les familles les plus nombreuses sont celles qui génèrent de fait tous les niveaux de profondeur, c'est-à-dire de 1 à 5 syllabes tonales. On peut ainsi comparer pour chacun des 4 groupes de familles du graphique, depuis la première colonne qui à chaque fois indique le nombre total de familles qui génère la totalité des structures de leur groupe, l'érosion relative de leurs effectifs.

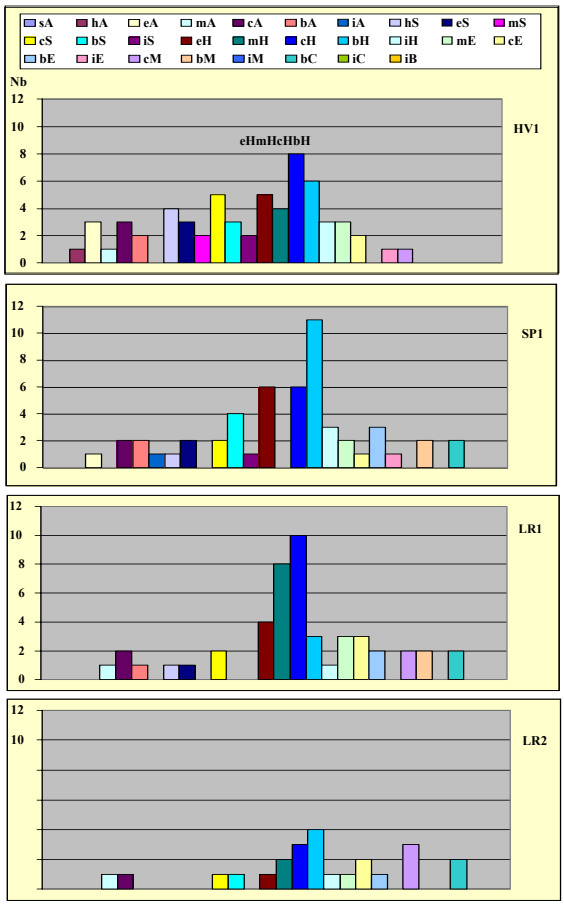
Par exemple le groupe 1 se compose de 14 familles qui génèrent au moins 1 syllabe ; parmi elles, 12 génèrent au moins 2 syllabes, au sein desquelles, 9 au moins 3, etc. Parmi les familles du groupe 2, différentes des précédentes, elles sont 15 à compter au moins 2 syllabes (jamais une famille dans ce groupe n'est attestée avec une seule syllabe), et puis parmi elles, 12 à compter au moins 3 syllabes tonales, etc. Le groupe 3 rassemble des familles qui n'attestent jamais 1 ou 2 syllabes dans nos données, elles sont 11 à compter au *minimum* 3 syllabes, puis 6, 4 syllabes, et aucune 5 syllabes, etc. De fait les familles 1 et 2 regroupent à elles deux 236 structures sur 294.

8. Pentes ascendantes

Pour terminer cette étude, nous comparons les pentes ascendantes des 4 locutrices. Ce type de pente est intéressant dans la mesure où il compte les effectifs les plus nombreux, recueillant 187 sur 294 structures. La comparaison (graphique 18 ci-dessous) porte sur le nombre de structures tonales (sur 294) réparties dans les familles tonales. L'échelle est la même pour les 4 locutrices.

Il est intéressant de constater que quelle que soit la locutrice, l'étalement est globalement le même. Par ailleurs un noyau central pour le début et fin du mélisme dans la structure intonative est constitué par les familles tonales eH, mH, et surtout cH et bH. *Les familles tonales cH et bH constituent donc, dans nos données, l'ancrage intonatif le plus fréquent pour les mélismes.*

Comparativement, la locutrice HV1 concentre le moins ses débuts et fins de proéminences mélodiques alors que sa petite fille, LR1, inversement, les concentre le plus autour notamment de mH et cH.



Graphique 18. *Effectifs de structures tonales (294) dans les familles tonales en fonction de chaque locutrice*

9. Conclusion

Concernant la méthode, cette étude a manifesté plusieurs intérêts. Tout d'abord, celui d'offrir une plus grande précision descriptive, permettant d'opposer des catégories (cf. les types de proéminences, les ML et MG, les locutrices ...), et de dégager certains invariants. Ensuite elle a permis de réduire la variabilité : les 294 structures se réduisent à 46 familles, et parmi elles, 14 familles se partagent les 2/3 de ces structures. Ces familles correspondent aux points d'ancrage du mot à la fois dans la courbe intonative et le mélisme en tant que mélodie subjective et affective. Enfin, elle permet d'extraire tout échantillon de parole segmenté, quels que soient sa taille et son statut linguistique, à des fins de comparaison intra- ou inter-locuteurs, quels que soient le dialecte, la langue ou le système prosodique (langue à tons ou non). En particulier ces familles tonales permettent de regrouper les données des 4 locutrices avec beaucoup d'homogénéité, ce qui donne la possibilité de poser l'hypothèse d'invariants propres à une famille, un village, une région etc.

En effet, au terme de cette étude nous constatons une assez grande homogénéité des résultats parmi les données des locutrices. Cette homogénéité est rendue possible par le protocole méthodologique qui cible très précisément les *extrema* mélodiques (*minimum* vs. *maximum*) du locuteur à travers les différents fichiers son, en évacuant aussi toutes les sources d'erreur de détection. On voit ainsi que malgré la disparité des effectifs des structures tonales parmi les locutrices, les familles tonales restent stables, bien alignées d'une locutrice à l'autre.

L'homogénéité des résultats se manifeste :

- par le regroupement des structures tonales autour des cibles mélodiques centrales, notamment la cible *h ;*
- quelles que soient les locutrices, les mots grammaticaux (MG) ont des caractéristiques proches, et de même pour les mots lexicaux entre eux (ML) ;
- il en va de même pour les proéminences hors pause *(M, Mc, Md, MF)* qui montrent une parenté nette dans leur comportement mélodique, quelle que soit la locutrice ;
- à cet égard il est intéressant de noter que les proéminences finales sans pause (MF) s'identifient davantage à des structures sans pause (archétype *M)* qu'à des structures finales (archétype *MP),* à qui elles s'opposent comme les autres.

Par ailleurs, nous souhaitons mieux harmoniser les effectifs des locutrices, portant nos observations et analyses par exemple à 100 structures tonales par locutrice. Ceci nous permettrait sans doute de mettre au jour des invariants familiaux ou régionaux, et aussi de développer dans de meilleures conditions une étude sur la fonction pragmatique du motif tonal dans les proéminences mélodiques, qui est un point capital. Il nous semble également important de susciter le même genre d'études à plus grande échelle, nationale ou internationale au sein de PFC, de manière à valider les processus mélodiques invariants comme définitivement représentatifs et identitaires.

Références

Caelen-Haumont, G. [à paraître]. *Prosodie et sens*, Marges Linguistiques.

Caelen-Haumont G. (1991). *Stratégies des locuteurs en réponse à des consignes de lecture d'un texte : analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodiques*. Thèse de doctorat d'état : Université d'Aix-en-Provence, 577 p.

Caelen-Haumont G., Auran C. (2004a). "The phonology of melodic prominence: the structure of melisms". *Proceedings of Speech Prosody 2004*, Nara, Japon, 143-146.

Caelen-Haumont G., Auran C., (2004b). "INTSMEL : un outil pour l'analyse des contours proéminents de F0", *Bulletin PFC* 3, 115-125.

Caelen-Haumont G., Auran C. (2005). *Manuel d'utilisation de la procédure MELISM sous Praat.*

Carton F. (1970). "Pente et rupture mélodique. Analyse instrumentale et fonctionnelle d'un trait prosodique régional", *Proceedings of the 6th International Congress of Phonetic Sciences,* Prague, 1967, Academia of Publishing House of Czechoslovak Academy of Prague- Munich, 237-241.

Carton F. (1972). *Recherches sur l'accentuation des parlers populaires de la région de Lille,* Publications de l'Université, Lille, 363 pages.

Chen A., Gussenhoven C., Rietveld A. (2002). "Language-specific uses of the Effort code", *Proceedings of the 1st International Conference on Speech Prosody*, SP 2002, Aix-en-Provence, Proceedings on line, http://www.lpl.univ-aix.fr/sp2002/.

Faure G.. (1970). "Contribution à l'étude du statut phonologique des structures prosodématiques", *Studia Phonetica*, 3, 93-107.

Faure G.. (1973). "Tendances et perspectives de la recherche intonologique", *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 5-29.

Fonagy I., Bérard E. (1973). "Questions totales et implicatives", *Studia Phonetica* 8, 53-98.

Hirst D., Di Cristo, A., Espesser R., (2000). "Levels of Representation and Levels of Analysis for the Description of Intonation Systems". Horne, M. (ed.), *Prosody:*

Theory and Experiment. Text, Speech and Language Technbology 14. Kluwer Academic Publishers, 51-87.
Léon P.R. (1970). "Systématique des fonctions expressives de l'intonation, Analyse des faits prosodiques", *Studia Phonetica* 3, 56-71.
Léon P.R. (1971). "Essais de Phonostylistique", *Studia Phonetica*, 4.
Léon P.R. (1976). "De l'analyse psychologique à la catégorisation auditive et acoustique des émotions dans la parole", *Journal de Psychologie* 3-4, 305-324.
Ohala J. J. (1983). "Cross-language use of pitch: an ethological view", *Phonetica* 40, 1-18.