Outils d'aide à l'annotation prosodique de corpus

Elisabeth Delais-Roussarie, Geneviève Caelen-Haumont, Daniel J. Hirst, Philippe Martin, Piet Mertens

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 qu'en 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 séq.), Simon (2003), 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 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.

[*] 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 (KU), Belgique

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

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 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 efficace. 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) ;

[1] Cette section a été rédigée par Philippe MARTIN. Le logiciel WinPitchPro peut être téléchargé sur le site http://www.winpitch.com

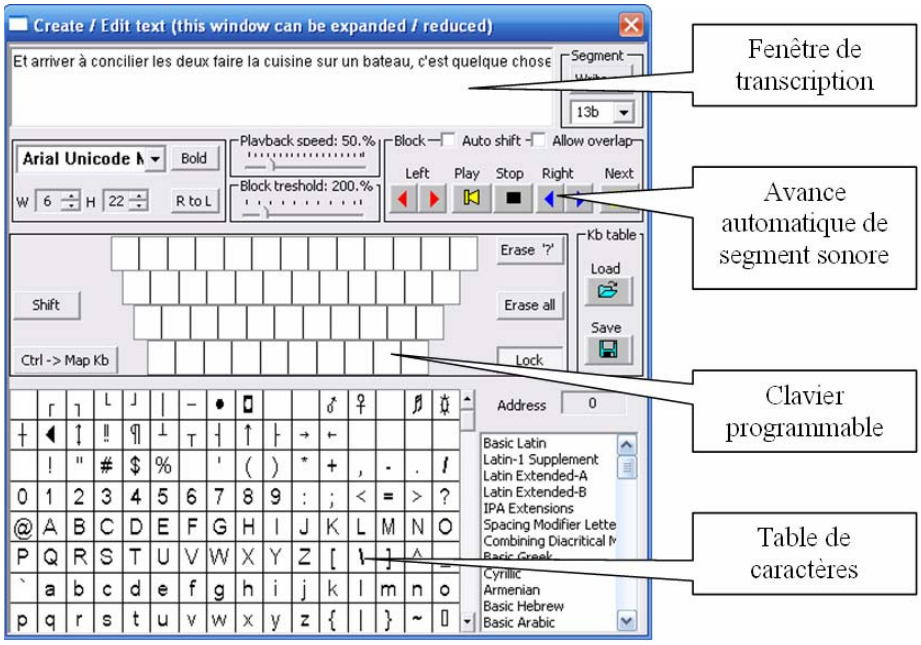


Figure 1 : **Fenêtre de transcription et d'alignement sous WINPITCH PRO**

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 tout 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®;

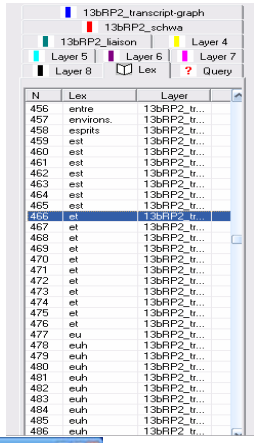
15. Lecture de fichiers multimédia dans la plupart des formats y compris les formats image (wav, aiff, au, mp2, mp3, mp4, mpeg, mpg, wma, etc.). Conversion des formats et des fréquences d'échantillonnage intégrés ; Transcription et alignement de fichiers multimédias ;

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) ;

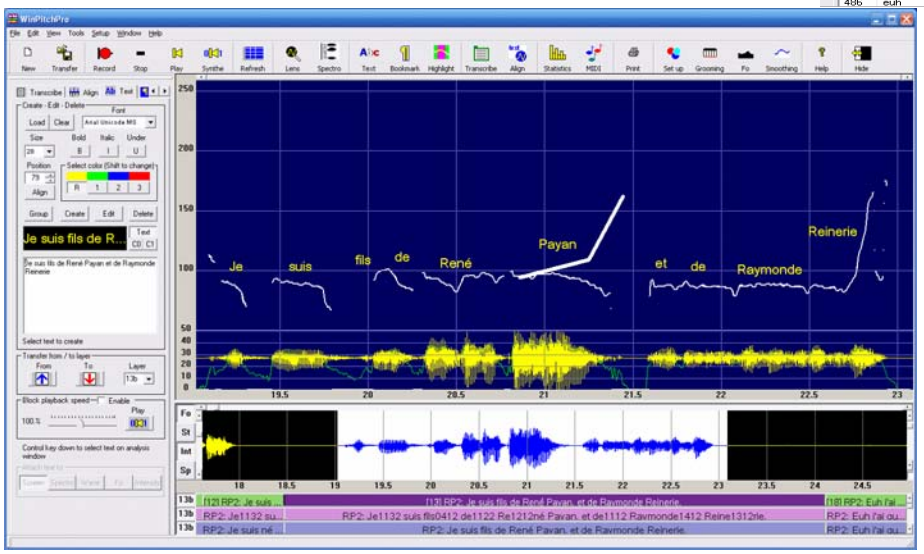


Figure 2 : **Ecran de WinPitch : fenêtres de commande (à gauche), de navigation (en bas à droite), d'analyse (en haut à droite).**

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 de 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 au format 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 des mesures précises telles que la durée du VOT des plosives, les valeurs des différents formants d'une voyelle, etc. ;
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

[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 :

FORMAT POSSIBLE 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

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 1).

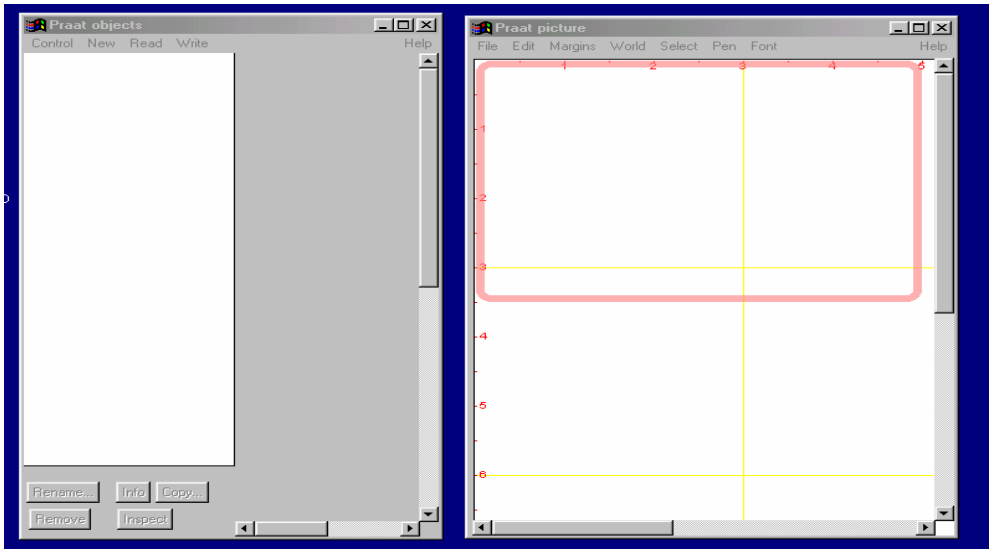


Figure 3 : **Ecran à l'ouverture de PRAAT**

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]	Edition 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, Editeurs, 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 :

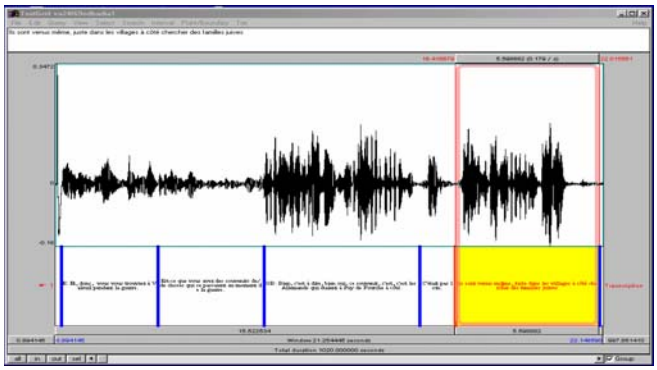


Figure 4 **: Fenêtre de segmentation, de transcription et d'alignement**

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."
        intervals [5]:
            xmin = 15.034119251845809
            xmax = 16.416678650055765
            text = "C'était pas loin."
```

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.

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

Fonction évaluée	WINPITCH	PRAAT
Généralités		
Plateforme acceptée	Windows	Windows, Unix, Mac. Logiciel multiplateforme
Ergonomie	Logiciel avec une interface utilisateur très conviviale	Interface utilisateur déroutante au premier abord.
Possibilités et fonctionnalités	Logiciel dédié à l'annotation de corpus et à l'étude des phénomènes prosodiques : - annotation de corpus ; - étude des paramètres prosodiques ; - étude des phénomènes segmentaux ; - requêtes contextuelles sur des données alignées.	Logiciel offrant un large panel de possibilités : - annotation de corpus ; - étude des paramètres prosodiques ; - étude des phénomènes segmentaux ; - construction de grammaire OT ; - 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 se fait 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 offre les possibilités suivantes : - aligner et annoter sur plusieurs tires ; - 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 prend 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

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

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 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 de traitements complexes ou de les appliquer à un nombre élevé de données (fichiers).

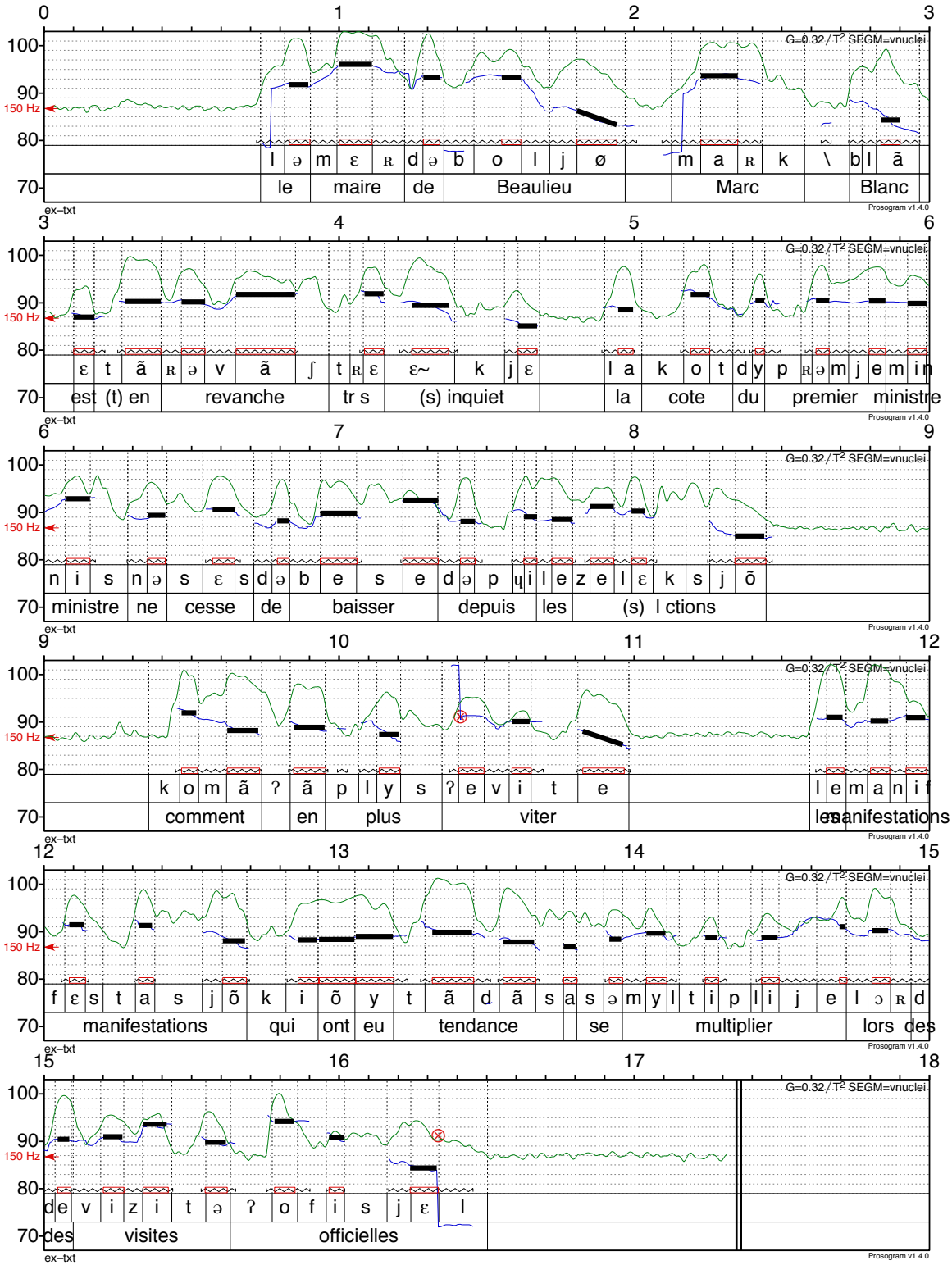
Des **outils annexes** permettent de faciliter l'analyse ou de l'appliquer à des données provenant de logiciels différents. Rappelons que l'outil de base permet déjà d'analyser plusieurs fichiers sons d'affilée, sans intervention de la part de l'utilisateur.

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

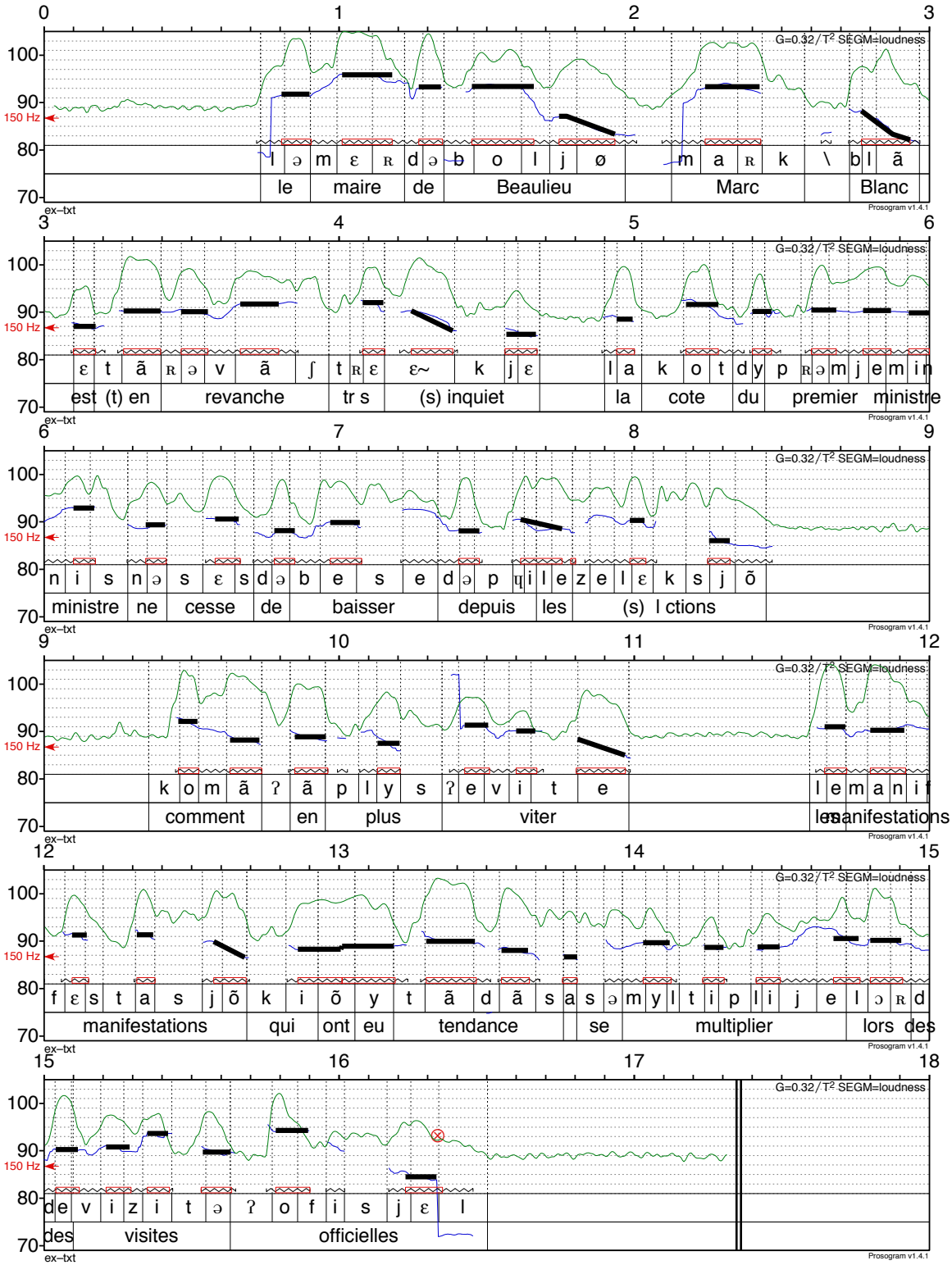
Parmi les extensions envisagées du prosographe, il y a la détection automatique des pauses et et le calcul de la proéminence syllabique.

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.

1. Prosogramme fait à partir d'un fichier segmenté manuellement :



2. Prosogramme fait à partir d'une segmentation automatique du fichier audio.



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 (1993), 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 dans la figure 5.

[6] Section rédigée par Elisabeth Delais-Roussarie, en collaboration avec D. Hirst (Intsint).

-

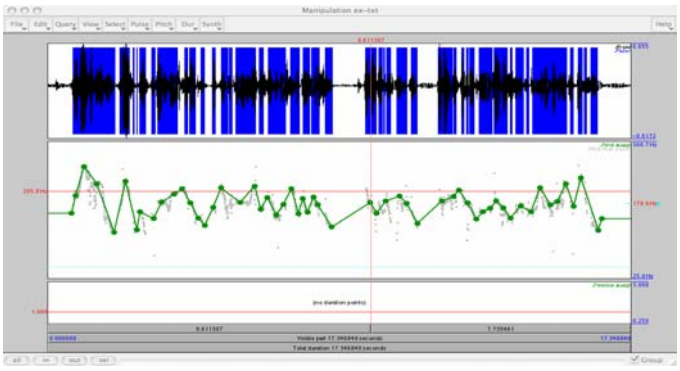


Figure 5 : **Manipulation de F0 par l'algorithme MOMEL**

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 (1999) ou 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.

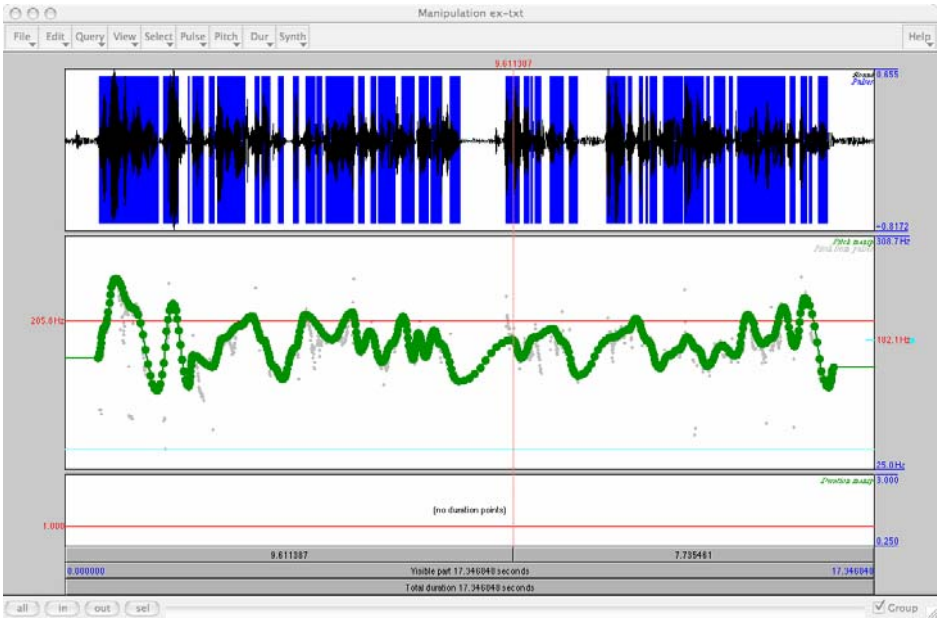


Figure 6 : **Courbe générée par Momel à partir des points cibles**

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).

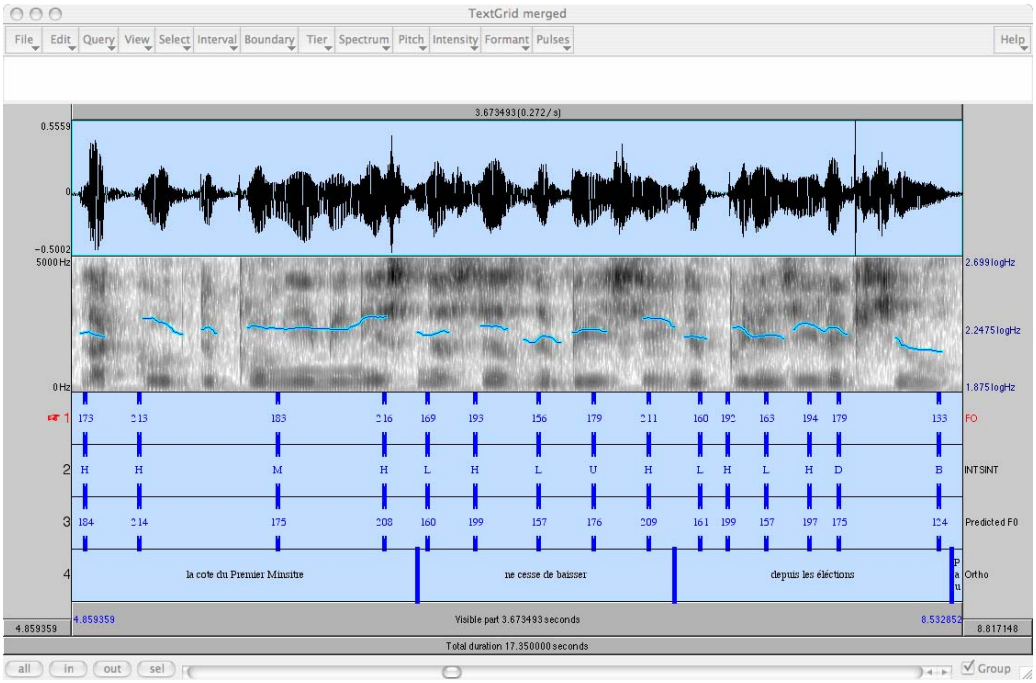


Figure 7 : **Codage symbolique INTSINT dans un TextGrid**

3. 3 MELISM[7]

MELISM est une procédure de codage discret de la mélodie qui a été recemment développée et est intégrée sous PRAAT sous la forme d'un script (G. Caelen-Haumont et C. 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,
- un stylisation de la courbe de fréquence fondamentale par détermination de points cibles effectuée par l'algorithme MOMEL (cf. § 3.2.1).

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

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 additonnant 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 ci-dessous) : 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).

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,
- 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.

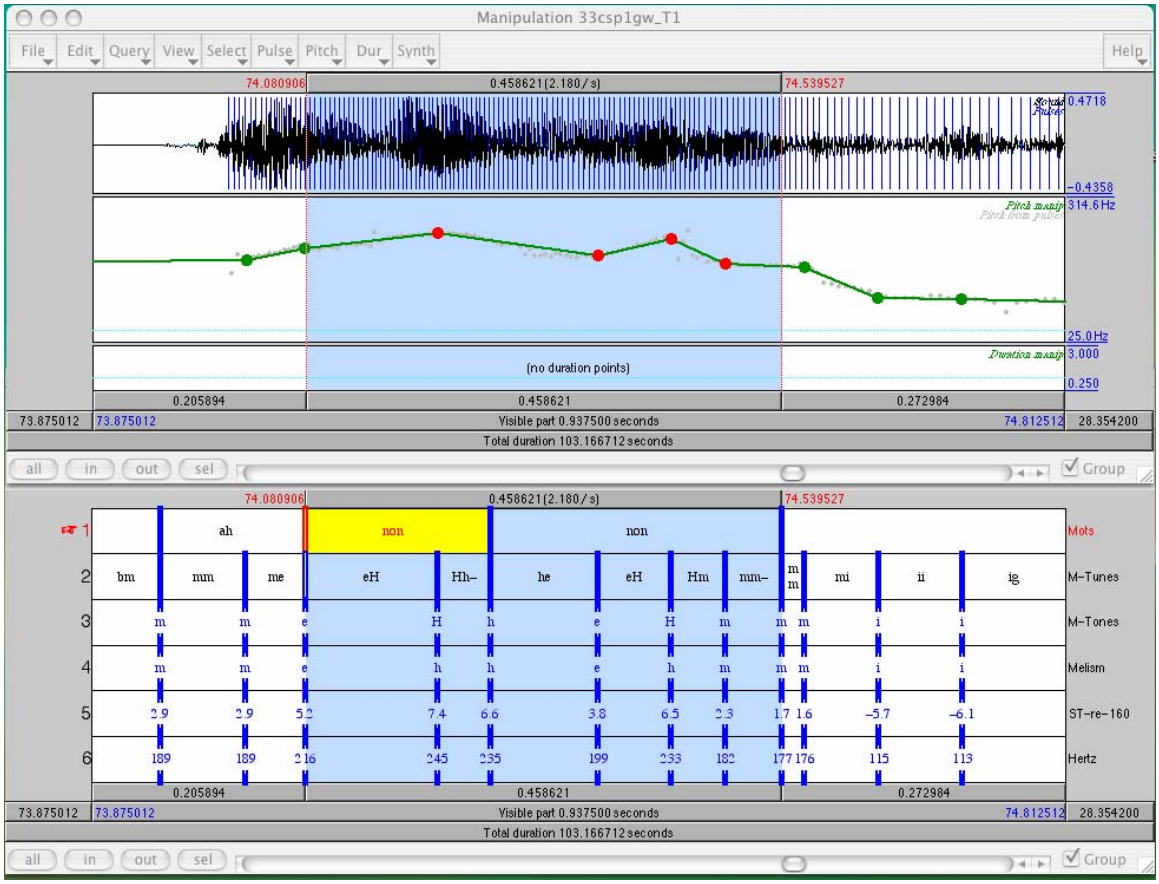


Figure 8 : **Codage MELISM et TextGrid**

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 Di Cristo, 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 séq.), 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

[8] Cette partie a été rédigée par Elisabeth DELAIS-ROUSSARIE.

modélisation sur des bases acoustiques. Dans certains cas, les résultats obtenus diffèrent comme l'a montrée Olivier (2005) ;

- 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 communication, 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.; Bel, B. Le caractère spontané dans la parole et le chant improvisés : de la structure intonative au mélisme. Revue PArole, no. 15-16. 2000, p. 251-302.

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

Caelen-Haumont, G.; Auran, C. INTSMEL : un outil pour l'analyse des contours proéminents de F0. Bulletin PFC n°3, 115-125. 2004 [ART/SCL]

Caelen-Haumont, G.; Auran, C. The phonology of Melodic prominence: the structure of melisms. Actes, Speech Prosody 2004 (2004 mars 23-26 : Nara, Japon). 2004, p. 143-146.

C-ORAL-ROM (2005) http://lablita.dit.unifi.it/coralrom.

Delais-Roussarie, E., A.Meqqori et JM 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, pp. 159-185.

Grabe, E. et B. Post (2002). "Intonational variation in the British Isles". In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 conference*, 11-13 April 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 & M.Huckvale (eds) *Improvements in Speech Synthesis*, Wiley.

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, pp. 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 Horne, M. (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*, pp. 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*, New York, September 1994.

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

Mertens, Piet (2004). “The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model”. In B. Bel & I. Marlien (eds.) *Proceedings of Speech Prosody 2004*, Nara (Japan), 23-26 March. (ISBN 2-9518233-1-2)

Mertens, P. et Ch. d’Alessandro (1995) “Pitch contour stylization using a tonal perception model”. *Proc. Int. Congr. Phonetic Sciences 13*, 4, 228-231 (Stockholm 1995)

Mixdorff, H. (1999). “A novel approach to the fully automatic extraction of Fujisaki model parameters”. ICASSP 1999.

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

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*, Tallinn, Estonia, 4-6 April, 2005, pp. 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. (2003). *Structuration prosodique du discours en français. Une approche multidimensionnelle et expérientielle*. Berne: Peter Lang.