HAL
open science

# INTSMEL : un outil pour l'analyse des contours proéminents de F0

Geneviève Caelen-Haumont, Cyril Auran

## To cite this version:

Geneviève Caelen-Haumont, Cyril Auran. INTSMEL : un outil pour l'analyse des contours proéminents de F0. Bulletin PFC (Phonologie du Français Contemporain), 2004, n° 3, pp.115-125. hal-00256394

**HAL Id: hal-00256394**

**https://hal.science/hal-00256394v1**

Submitted on 15 Feb 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# INTSMEL : un outil pour l'analyse des contours proéminents de F0

Caelen-Haumont Geneviève[1], Cyril Auran[1, 2]
[1]Laboratoire Parole et Langage
Université de Provence, 29 av. R. Schuman – 13621 Aix-en-Provence Cedex 1
[2] IUFM d'Aix-Marseille
Tél.: ++33 (0)4 42 95 36 39 - Fax: ++33 (0)4 42 59 50 96
Mél: gcaelen@lpl.univ-aix.fr
cauran@wanadoo.fr

## 1. Introduction

L'objectif de cet article est de présenter un nouvel outil d'analyse des contours proéminents de F0 (fréquence fondamentale). En effet si nous disposons bien d'outils spécialisés dans le suivi de la courbe mélodique et donc de l'intonation, nous n'avons pas jusqu'à présent d'outil dédié à l'analyse des configurations locales. Les systèmes décrivant l'intonation définissent en fait les valeurs quantitatives non pas en termes de valeurs absolues, mais relatives. Or pour l'étude d'un contour mélodique précis, nous avons besoin d'une part de référer les contours en termes de valeurs absolues, et d'autre part de définir des configurations structurelles en termes de tons (de surface).

## 2. Intonation et mélisme

Caelen-Haumont et Bel [4] proposent de distinguer deux strates dans la prosodie, l'une qui serait la partie linguistique et conventionnelle, correspondant à l'intonation (en relation précisément avec la structure linguistique), et l'autre qui serait la partie exprimant l'affectivité (« émotion ordinaire » ou plutôt « l'émotion-racine », à la source des états affectifs de l'individu), et qui serait exprimée par le mélisme, d'expression locale, souvent mono-lexicale. Le mélisme serait ainsi caractérisé d'une part par une amplitude de F0 très importante, ce qui implique des valeurs maximales de F0, et serait d'autre part souvent lié à une rupture mélodique ou prosodique de la trame linguistique, rupture intervenant donc en dehors des frontières de syntagmes.

Etablie sur des données prosodiques, cette partition est par ailleurs appuyée par le contexte des recherches actuel dans le domaine psycholinguistique et le domaine neuronal des aires cérébrales. Tout d'abord, comme on le sait, Scherer [14] a testé deux modèles, covariance et configuration, le premier stipulant que l'information sur l'émotion et l'information sur le contenu linguistique fonctionnent de manière indépendante, ce qui implique que les traitements linguistiques et paralinguistiques puissent s'effectuer en parallèle. Les indices de F0 associés à cette approche sont le F0 global et l'amplitude de F0. Le second pose que le type de contour de F0 est un élément linguistique, ce qui induit deux types de variations de F0, linguistiques et paralinguistiques, et sous-entend que l'information liée au contour et à l'émotion n'est accessible que via les traits linguistiques de l'énoncé. L'indice de F0 correspondant est le type de contour. Au terme d'expérimentations perceptuelles, il apparaît que les deux approches sont pertinentes, la covariance semblant plus adéquate pour décrire la parole affectée par les facteurs biologiques (i.e. émotionnelles), la configuration, pour la parole affectée par les conventions linguistiques et socioculturelles.

Cette dichotomie des fonctions est encore reprise par le modèle de Scherer "pull" et "push" [13], deux processus qui rendent compte des expressions de la parole émotionnelle sous contrôle cognitif. Il y aurait ainsi deux types d'effets au niveau des expressions de la parole émotionnelle, les effets

“push” produits par des “déterminants internes au sujet” causant des modifications physiologiques en réponse à une activation émotionnelle, et les effets “pull” produits par des “déterminants externes au sujet” en relation avec le contexte environnemental et social où se trouve le locuteur.

Par ailleurs au niveau des aires cérébrales, il semble attesté [Pell, 10] d’une part que l’hémisphère gauche soit plutôt spécialisé dans le traitement de la linguistique et celui de la prosodie linguistique (niveau représentationnel), alors que l’hémisphère droit aurait plus d’habilité à traiter la prosodie émotionnelle, avec les indices de F0, véhicules attestés des intentions liées aux émotions.

Ainsi selon nous, ces différentes études valident à propos de la parole affective, la distinction intonation / mélisme, établie à partir de la seule analyse des données, comme deux espaces spécialisés de l’information, en interaction, l’un lié plus spécifiquement à l’expression de la prosodie linguistique, l’autre à celle de la prosodie affective.

**3. Caractérisation prosodique du mélisme**

En l’absence de terme approprié, nous avons emprunté au domaine musical le terme de mélisme. Dans notre perspective, dans le cadre d’un item lexical, ou au plus d’une suite, ce terme s’applique à la forme acoustique et mélodique, désignant une structure phonologique de surface, avec une granularité adaptée à ses objectifs [Caelen-Haumont et Auran, 2].

Concernant cette granularité, la division du registre du locuteur selon 4 niveaux (cf Figure 1 ci-dessous, les niveaux S, E, C, I), usuelle depuis Delattre [1966], nous a paru insuffisante pour nos objectifs. Le codage que nous proposons est effectué de manière automatique (cf ci-dessous) sous Praat par la procédure INTSMEL. En accord avec la procédure INTSINT, nous avons en fait divisé chaque niveau en 3, le « cœur » représentant 50% du niveau, chacune des marges inférieure et supérieure, 25%, ce qui, hormis la première marge (25%) du registre le plus grave et la dernière (25%) du registre le plus élevé du locuteur (qui étant aux limites du registre n’ont pas de marge voisine), correspond à une succession de 7 plages de 50% (échelle logarithmique), comme le décrit la figure 1 ci-dessous :

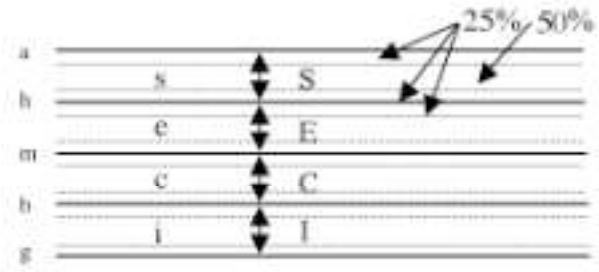


Figure 1 : Les 4 et les 9 niveaux du mélisme, avec aigu (a), supérieur (s, S), haut (h), élevé (e, E), moyen (m), centré (c, C), bas (b), inférieur (i, I), grave (g). Les niveaux S, E, C, I correspondent aux 4 niveaux de Delattre [1966].

Nous arrivons ainsi à 9 niveaux (cf table 1 ci-dessous) : aigu (a), supérieur (s), haut (h), élevé (e), moyen (m), centré (c), bas (b), inférieur (i), grave (g). Ces 9 niveaux en combinaison créent 81 tons bitonals. En fait ces 81 tons rendent compte de toute configuration mélodique du mot ou du groupe de mots, pour les langues tonales ou non tonales, mais pour notre analyse des mélismes, seulement 45 tons sont retenus.

Par ailleurs, pour une meilleure précision, nous avons fait une distinction entre les variations intra-niveau et les plateaux. Par opposition aux plateaux qui par définition sont non orientés (ex. : aa), les

variations intra-niveau sont toujours annotées avec le symbole +/- (ex. : mm+ ; ii-, etc.), qui code simplement l'existence d'une pente ascendante vs. descendante, déterminée par un seuil. Les niveaux sont naturellement gradués (a>g), mais dans une séquence, deux ou plusieurs cibles identiques de tons peuvent coexister, et dans ce cas on ne sait pas, quand elles ne sont pas adjacentes, où se trouve la F0 max. Une lettre capitale permet de lever l'ambiguïté : ainsi <Ss sa As ss> indiquerait que 1° Ss > ss, 2° As > sa, (et bien sûr que As > Ss).

| ton | aigu | supérieur | haut | élevé | moyen | centré | bas | inférieur | grave |
|---|---|---|---|---|---|---|---|---|---|
| | a | s | h | e | m | c | b | i | 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* |

Table 1 : Matrice des tons pour la description des configurations mélodiques des mots, en particulier les mélismes. Les caractères gras correspondent aux tons ascendants vs. descendants du mélisme, ceux en italiques aux plateaux.

Nous pouvons maintenant préciser que les corrélats prosodiques du mélisme sont :

- l'implication des niveaux les plus aigus, soit les niveaux a et s,
- ou une large excursion de F0 (interne ou externe, si une rupture de F0 intervient entre le mot courant et le mot précédent / suivant), atteignant au minimum le niveau h, à condition que l'amplitude soit suffisante (niveau inférieur au plus c, sinon b, i, g),
- l'un ou l'autre corrélat impliquant généralement un ralentissement sensible du débit et éventuellement une augmentation importante de l'énergie.

Ce type d'annotation répond en fait à certains objectifs précis :

- fournir une information quantitative (valeur de F0) relatives aux cibles du locuteur,
- proposer un système d'annotation et une stylisation indépendants du groupe, de la phrase, de l'énoncé, du corpus, du locuteur, du sexe, de l'âge,
- fonder un codage à la fois sur des valeurs absolues de F0 pour ce qui concerne la linéarité du discours, mais relatives par rapport au registre du locuteur,
- décrire phonologiquement (en surface) la structure interne des mots mélismés,
- faciliter la comparaison des différents patrons mélodiques en fonction du statut syntaxique, sémantique et/ou pragmatique des items,

- faciliter la comparaison des différents systèmes mélodiques qui existent dans les différents langages, y compris les langues à tons,
- permettre de tester les hypothèses théoriques à l'aide de procédures expérimentales.

**4. D'autres systèmes d'annotation**

Dans le domaine de l'annotation prosodique, il existe en fait de nombreux systèmes. Dans ce paragraphe, nous ne prétendons pas décrire l'ensemble des systèmes existants, mais simplement nous focaliser sur quelques uns d'entre eux, de manière à mettre en lumière les particularités de INTSMEL.

Un des plus connus à l'échelon international est sans doute le système ToBI [Pitrelli et al., 11]. S'il procède bien d'un point de vue théorique, il dépend aussi des jugements empiriques de l'utilisateur. Il se présente comme un mixte de perception (les indices de coupure ou *break indices*), et de perspectives acoustique et phonologique (événements mélodiques tels que les *boundary tones* et les *pitch accents*). Le système ToBI est en fait lié à l'analyse linguistique d'un langage déterminé, puisque les symboles prennent en compte outre les syllabes accentuées (*stressed syllables*), les relations fin vs. début de groupes (*phrases junctures*), les frontières initiales et finales des phrases.

Le système PIT [d'Alessandro et Mertens, 5], s'inscrit dans la ligne de recherche initiée par l'Institut IPO [t'Hart et al., 6] en vue de la synthèse de la parole. Il a été développé dans l'objectif d'une analyse automatique de l'intonation du français utilisant un système de reconnaissance automatique de la parole (ASR). Il repose sur une première segmentation du signal de parole (et donc des contours de F0) en syllabes phonétiques, puis sur une deuxième segmentation en segments tonals successifs. Cette dernière inclut une intégration perceptuelle des variations du pitch à moyen et court terme, le seuil de glissando (c'est-à-dire le seuil de perception à partir duquel on perçoit une montée ou une descente de F0 comme telle dans une unité phonétique, fonction de la pente et de la hauteur), et le seuil de glissando différentiel. A cette étape, le système fournit des segments tonals, des contours stylisés (*tonal score*), qui sont les entrées du module de synthèse. Le système semi-automatique Mingus, qui s'appuie sur le modèle tonal de l'intonation du français [Malfrère et al., 9], est partiellement intégré dans le système automatique. Il propose un ensemble de 4 tons x 2 (syllabes accentuées et atones), qui prennent en compte les variations absolues de F0 plus grandes qu'une tierce majeure, et au sein de ces dernières, les variations relatives pour les tons *downstepped* et *upstepped* (chute vs. montée progressive et continue de F0, parfois fonctionnelle dans certaines langues). Tout comme le système ToBI, la procédure d'annotation PIT s'inscrit au sein d'une analyse linguistique du langage étudié et en particulier syntaxique.

Bien que les travaux de [Rossi et Chafcouloff 12] soient anciens, pour l'étude prosodique du français, ils n'en constituent pas moins cependant une des références importantes des systèmes de codage, tels par exemple qu'INTSMEL. Ce système bien connu repose sur la définition de 6 niveaux intonatifs, couvrant l'ensemble de la tessiture du sujet, en termes de valeurs quantitatives de F0, et leur écart-type. Ces 6 intonèmes sont ainsi définis par rapport au F0 moyen du locuteur, comme des formes (hauteur, orientation de la pente, amplitude de F0 spécifiques), associées à des contenus syntaxiques. Ces morphèmes tonals permettent d'établir des paires contrastives (question vs. assertion, continuation mineure vs. majeure de groupe, conclusive mineure vs. majeure de groupe), qui exercent des fonctions syntaxiques.

INTSMEL se distingue de ce système d'une part par le nombre de ses niveaux, et surtout par l'absence d'une perspective syntaxique : en fait INTSMEL analyse des configurations de F0 au sein

de tout continuum, qu'il s'agisse de l'ensemble de l'énoncé ou du mot lexical. Pour notre part, l'étude est lexicale.

Il existe par ailleurs d'autres systèmes qui sont indépendants de toute analyse linguistique a priori, et qui de ce fait ont le grand avantage de pouvoir s'appliquer à toute langue, « langue à tons » ou non, et donc tout type d'organisation mélodique. C'est le cas en particulier d'INTSINT [Hirst et Di Cristo, 8]. Ce système d'annotation distingue entre :

- les tons absolus à savoir *Top (T), Mid (M), Bottom (B),* calculés sur l'amplitude mélodique maximale (min vs. max de F0),
- les tons relatifs, *Higher (H), Same (S), Lower (L),* calculés en fonction de la cible précédente, quelle qu'elle soit, relative ou absolue,
- et deux autres tons relatifs, *Upstepped (U), Downstepped (D),* qui caractérisent des variations dans un intervalle plus réduit.

Par contraste, INTSMEL, appliqué dans notre perspective dans le cadre du codage de séquences courtes (mot ou suite de mots), propose un codage des valeurs considérées de manière absolue. Ce type d'outil a été défini sur les lacunes des systèmes existants, qui se sont révélées à l'étude des contours de F0 proéminents dans le cadre du mot lexical (ou une suite courte), que l'analyse porte sur des mots isolés ou que le mot soit intégré dans un groupe. De la sorte INTSINT et INTSMEL offrent des perspectives d'analyses complémentaires, INTSINT s'attachant à décrire l'intonation, sans référence à une analyse linguistique quelconque, INTSMEL se spécialisant dans la description des contours mélodiques au sein des mélismes, et donc des unités lexicales, voire grammaticales. De fait la procédure INTSMEL par elle-même est également indépendante de la couche linguistique et fonctionne comme telle, mais l'usage que nous en faisons personnellement en l'appliquant sur les items segmentés et étiquetés, l'inscrit, comme on l'a déjà dit, dans le contexte du mot ou d'une suite. La différence entre ces deux systèmes repose sur deux dimensions : la valeur différente de la cible et la référence linguistique des items.

Ainsi dans le champ du codage de F0, on retrouve cette même dichotomie évoquée paragraphe 2, le système INTSINT codant l'intonation, le système INTSMEL, les mélismes.

## 4. INTSINT, INTSMEL et PRAAT

INTSINT et INTSMEL sont deux procédures automatiques [Caelen-Haumont 3], qui fonctionnent sous PRAAT. Elles se développent l'une et l'autre à partir de la procédure MOMEL [Hirst et al., 7] qui permet de calculer et même de reconstruire la valeur de F0 sous-jacente, y compris pour les segments non voisés, réalisant alors une courbe de F0 continue, proche de la perception auditive. MOMEL repère ainsi dans le continuum mélodique les valeurs min / max relatives, indépendamment de la structure linguistique. A la sortie de la procédure MOMEL, les systèmes INTSINT et INTSMEL, fournissent chacun une annotation de F0 automatique.

### 4.1. INTSINT et INTSMEL

INTSMEL constitue le composant principal final de la procédure MELISME qui consiste fondamentalement à chaîner les algorithmes MOMEL, QSP et INTSMEL. Deux outils mineurs (scripts en langage Perl) entrent aussi dans la procédure dans le cadre de la conversion des données aux formats TextGrid et PitchTier de Praat.

De manière plus précise, l'algorithme MOMEL a pour objectif de modéliser la courbe originale de F0 de sorte que toutes les caractéristiques micro-segmentales (le composant micro-prosodique) soient supprimées [3]. La courbe ainsi calculée est donc similaire à celle obtenue à partir d'une séquence de segments entièrement voisée et en constitue le composant macro-prosodique ([6], [7]). LA sortie de l'algorithme MOMEL consiste ainsi en une suite de point-cibles définis dans un espace temps/fréquence.

L'interpolation entre ces point-cibles, destinée à la visualisation et aux mesures de valeurs ponctuelles, est ensuite réalisée à l'aide de courbes splines quadratiques par l'algorithme QSP : les point-cibles MOMEL sont récupérés en entrée et des valeurs de F0 modélisée sont générées toutes les 10 ms pour la totalité du signal de parole.

La combinaison de MOMEL et QSP permet ainsi de traiter une séquence de point-cibles comme une représentation phonétique appropriée des courbes de fréquence fondamentale.

Dans un troisième temps, l'algorithme INTSMEL, bien que générant une sortie visuellement proche de celle produite par INTSINT, diverge en fait de ce second algorithme tant dans ses objectifs que ses bases théoriques. En effet, comme nous l'avons vu, l'algorithme INTSINT code une séquence de point-cibles MOMEL à l'aide de l'ensemble de symboles tonals {M, T, B, H, L, S, U, D} dont certains sont absolus et d'autres relatifs ; *a contrario*, INTSMEL utilise un ensemble de neuf symboles absolus ({a, s, h, e, m, c, b, i, g}) correspondant à des fractions (sur une échelle logarithmique) du registre du locuteur. Les point-cibles ainsi codés sont ensuite regroupés automatiquement en tons de mélisme.

L'algorithme INTSMEL procède en une succession d'étapes :

- segmentation du registre du locuteur (afin d'obtenir les neuf niveaux détaillés plus haut) ;
- codage de chaque point-cible et de chaque frontière de domaine (empan d'un élément lexical dans le cadre de l'étude des mélismes) ;
- optimisation du premier codage par réduction de groupes « frontière – point-cible » au sein d'une fenêtre flottante de 40 ms ;
- marquage des variations intra-niveaux par opposition aux plateaux ;
- marquage des extrema locaux d'un domaine de codage (cf paragraphe 3 ci-dessus)

### 4.2. INTSMEL et PRAAT

L'ensemble des algorithmes associés à la procédure MELISME sont gérés manière modulaire et appelés à partir d'un unique script Praat nommé « melism.praat ». Le formulaire de paramètres de ce script permet ainsi de sélectionner un ensemble de fonctionnalités appliquées en un ensemble de fichiers son (traitement « batch »).

L'extraction de fréquence fondamentale est ainsi tout d'abord effectuée avec entre autres paramètres des extrema (et une valeur moyenne) soit définis par l'utilisateur (correspondant aux valeurs observées sur un ensemble de fichiers), soit fixés à des seuils classiques tels que 75Hz-500Hz (la valeur moyenne est alors calculée sur les valeurs de F0 ainsi obtenues). Les valeurs de F0 obtenues sont stockées dans un fichier au format ASCII destiné à constituer l'entrée de l'algorithme MOMEL.

Dans un deuxième temps, l'exécutable MOMEL calcule les coordonnées des point-cibles obtenus à partir des valeurs de F0 fournies. La sortie générée prend la forme d'un fichier ASCII converti au format PitchTier de Praat par un script Perl spécifique.

Troisièmement, l'exécutable QSP génère un autre fichier ASCII contenant les valeurs de F0 de la courbe modélisée par interpolation quadratique entre les point-cibles MOMEL. Ce fichier est ensuite converti au format PitchTier par un script Perl spécifique afin d'en faciliteer la visualisation et l'édition. Il est à noter que cette étape est devenue facultative depuis la version 4.1.17 de Praat, qui l'intègre à présent comme option d'interpolation entre points au sein d'un objet PitchTier. Les versions ultérieures n'intègreront donc plus cette étape.

La procédure INTSMEL, codée en langage script Praat, est ensuite appelée, impliquant les étapes détaillées en 4.1 ci-dessus et générant finalement un fichier au format TextGrid pour Praat.

Après la phase de segmentation (manuelle, semi-automatique voire automatique), et l'utilisation de la procédure MELISM intégrée dans PRAAT, chaque item segmenté est codé, de sa frontière gauche à sa frontière droite, à l'aide d'une ou plusieurs étiquettes phonologiques, les « tons » tels que définis par la procédure INTSMEL. Cette séquence peut être simple (seulement une pente, ascendante ou descendante, d'amplitude plus ou moins importante, ou un plateau), ou complexe (alternance de pentes opposées ou parallèles, et/ou de plateaux), mais par définition la structure du mélisme commence ou finit par un niveau aigu.

## 5. Exemple d'application à l'analyse pragmatique

Cette seconde partie concerne un exemple d'application de l'outil d'annotation INTSMEL à un corpus. Du fait que INTSMEL a été créé indépendamment de tout a priori théorique, il gagne en généralité d'utilisation, et peut donc être utilisé dans de nombreuses études, d'orientation phonétique, phonologique, syntaxique, sémantique, pragmatique. Dans le cadre du Projet PFC et son utilisation dans le domaine phonétique et phonologique par exemple, il permet de repérer des contours propres à une personne ou à une communauté linguistique. Pour notre part ici, nous donnons à titre d'exemple, une application dans le domaine sémantique et pragmatique.

### 5.1. Corpus et objectifs du dialogue

Inspiré du Corpus HCRC Map Task [Brown et al. 1], le corpus étudié est l'un des 8 dialogues enregistrés, d'une quinzaine de mn chacun, saisis dans de très bonnes conditions acoustiques, et avec pistes séparées pour chaque locuteur. Ce dialogue met en scène un jeu de rôles avec 2 locutrices, une « touriste » et une « employée de l'office de tourisme », avec des conflits d'objectifs (la touriste préférant les activités sportives, l'employée préconisant les activités culturelles), et des ambiguités spatiales (les plans n'ayant pas la même version de la ville). Il faut réactualiser le plan à la borne d'information, établir un programme, fixer l'itinéraire en fonction des nouveautés de la ville (travaux, sens circulatoire, noms des rues). L'étude porte sur les réalisations de la locutrice O4.

Nous formulons les hypothèses suivantes :

1° la locutrice s'implique subjectivement dans la tâche qui lui est demandée, à savoir dans son rôle d'employée du tourisme,

2° les indices de cette implication se trouvent au niveau mélodique dans les mélismes,

3° les items lexicaux « mélismés » correspondent aux objectifs du dialogue et aux motivations personnelles de la locutrice.

En fonction des objectifs généraux du dialogue mentionnés ci-dessus, nous posons que ces objectifs sont répertoriables grâce à 5 champs lexicaux des domaines sémantique et pragmatique :

- champ A : les coordonnées spatiales de la ville, des objets touristiques (noms et directions),
- champ B : les qualités descriptives des objets culturels (spatiales, temporelles, architecturales, économiques, etc.),
- champ C : les outils informatiques, manipulation, conseils, mise à jour du plan,
- champ D : l'appréciation subjective des objets touristiques, des tâches respectives, l'argumentation,
- champ E : l'interaction verbale, particulièrement les verbes et adverbes de l'interaction, les particules phatiques.

5.2. Mélismes et expérimentation

Pour cette expérimentation, nous n'avons retenu que les mélismes des items lexicaux atteignant exclusivement les niveaux *a* et *s*.

| A | B | C | D | E | F | Total |
|---|---|---|---|---|---|---|
| 17 | 14 | 3 | 19 | 16 | 2 | 71 |
| 23.9% | 19.7% | 4.2% | 26.8% | 22.5% | 2.8% | 100% |
| 53 74.6% | | | | 18 25.4% | | |

Table 2 : Pourcentages des mélismes en fonction des champs sémantiques / pragmatiques des mots lexicaux, A : coordonnées spatiales, B : qualité des objets, C : tâche technique, D : évaluation subjective et argumentation, E : interaction verbale, F : autres.

Sur les 1860 mots du corpus O4, 71 ont été l'objet de ce type de mélisme (en *a* et *s*), se distribuant en 35% à la finale d'un groupe prosodique, et 65% au sein du groupe. Après segmentation et étiquetage, les cibles ont été recalculées au sein des bornes des mots lexicaux pour définir leurs tons.

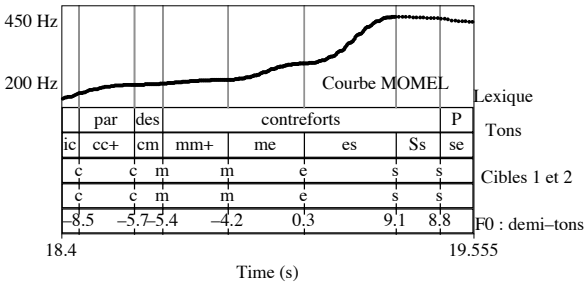


Figure 1 : exemple de mélisme sur le mot *contreforts,* avec de bas en haut, les valeurs de F0 en demi-tons, les cibles MOMEL (cibles 1 : valeurs directes, cibles 2 : recalculées), les tons (structure phonologique superficielle), le lexique, et enfin la courbe MOMEL en Hz.

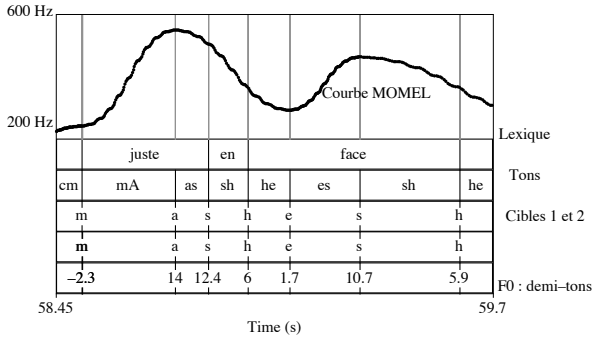


Figure 2 : Séquence de mots « momélisée » *juste en face*. Même légende que Figure 1.

La distribution des mélismes en fonction des champs lexicaux est présentée table 2 ci-dessus. On constate que 74.6% des mots lexicaux mélismés (*a* et *s*) correspondent aux champs lexicaux liés aux objectifs du dialogue (A, B, C, D), soit les 3/4. Le quart restant correspond très majoritairement au lexique de l'échange discursif (E, 22.5%). Ces résultats sont très proches de ceux de l'étude perceptive [Caelen-Haumont, 2] : en effet sur les 151 mélismes sélectionnés auditivement (soit le double), 71% correspondaient aux champs lexicaux des objectifs du dialogue, 28 % au lexique des échanges discursifs.

Les exemples ci-dessus (Figures 1 et 2) présentent des extraits de parole avec mélisme, le premier extrait de « ... on peut voir encore des contreforts ... », de forme ascendante puis plateau, avec 14.5 demi-tons d'amplitude, ainsi que la séquence de tons qui le décrivent. Le deuxième exemple (« juste en face [...] y'a l'marché ») est une séquence de 3 mots mélismés, présentant une suite de pentes montantes et descendantes, avec une amplitude extrêmement importante sur *juste* (16.3 demi-tons), qui arrive dans la zone maximale du registre de la locutrice (près de 600 Hz). Comme on le voit la procédure INTSMEL décrit avec précision la courbe de F0, ses formes et ses valeurs, distinguant bien notamment entre les variations inter-niveau (Figure 1, ex : mm+ ) et les plateaux (ex : sS). Dans la figure 2, inversement, les plateaux sont absents.

5.3. Valeurs, croyances, buts perlocutoires, fonctions prosodiques

Ces résultats nous permettent d'une part, par le compromis satisfaisant qu'il offre entre généralité et précision, de valider notre outil de sélection et d'annotation des mélismes, et de ce fait nous autorisent à poursuivre l'analyse dans le domaine pragmatique.

L'analyse retient tout d'abord que la locutrice s'est parfaitement investie dans sa tâche d'employée du tourisme. En effet si la majorité du vocabulaire mélismé se trouvait en dehors des champs lexicaux liés aux objectifs du dialogue, nous aurions pu en conclure que le locuteur n'avait pas été motivé par sa tâche, ce qui manifestement n'est pas le cas.

Mais sur quoi repose cet investissement ? Autrement dit quelles sont les valeurs qui le fondent ? En fait dans les mots mélismés, on trouve 2 types de valeurs, celles qui sont transmises par les consignes du dialogue, à savoir informer, établir un programme de visites culturelles, élaborer un itinéraire, mettre à jour le plan, et par ailleurs celles qui sont propres à la locutrice O4 : l'importance qu'elle accorde aux éléments naturels (les arbres du parc, la roseraie, les oiseaux ...), aux éléments traditionnels (marché, artisanat, costumes ...), aux éléments esthétiques des monuments anciens, aux éléments conviviaux (accueil et atmosphère des restaurants, amabilité des commerçants ...), qui en fait fonctionnent comme des arguments personnels pour atteindre les objectifs de la tâche. En réalité

ces différentes valeurs, d'ailleurs compatibles entre elles, sont plus profondément des croyances, véritable source d'énergie (cf l'état d'excitation, ou *arousal* en anglais, bien décrit dans la littérature), pour la parole et donc la prosodie. De ce fait ces croyances sont également de 2 types, celles qui sont issues de la tâche (et dans ce cas, tous les locuteurs des différents dialogues qui se sont investis, les partagent), et celles qui sont propres à la locutrice (variables en fonction des locuteurs).

Relativement à la tâche, c'est croire (et faire croire) à la nécessité de convaincre la touriste de se rendre dans les lieux culturels, de l'aider dans ses choix, l'informer sur le plan touristique, dans sa tâche technique, et de le faire sur le registre de l'amabilité et de la disponibilité. Relativement à la locutrice, c'est croire (et faire croire) qu'un objet touristique donné (château, église, restaurant traditionnel ...) nécessite le déplacement de la touriste, que les éléments et les repères de la ville (ronds-points, sens uniques ...) sont une connaissance indispensable pour atteindre mieux et plus vite le point de visite, qu'une caractéristique de cet objet (historique, esthétique, exotique, caractère exceptionnel ...) est susceptible de la séduire. En fait la croyance repose sur le sentiment que ces valeurs sont partageables et motivantes pour autrui.

De ce fait, les mélismes qui mettent en relief par le registre aigu (niveaux *a* et *s* et/ou large excursion mélodique) tout ce vocabulaire attaché aux valeurs de la tâche et de la locutrice, exercent, indépendamment de la strate linguistique, et comme elle, une fonction perlocutoire par l'action sur autrui, volontaire, consciente ou pas. C'est sans doute grâce aux notes aiguës du mélisme que l'auditeur peut identifier, attachée à ces valeurs exprimées lexicalement, l'expression d'une vérité du locuteur, d'une subjectivité ou « émotion-racine », et qu'il peut y adhérer, ou au contraire la réfuter. L'espace lexical du mélisme est à ce titre un lieu de rencontre avec l'autre, une communication privilégiée de l'intersubjectivité et souvent du passionnel.

## 6. Conclusion

Nous avons défini dans ces pages un outil de description phonologique (de surface) des proéminences mélodiques des unités lexicales, ou mélismes, et son application à l'analyse sémantique et pragmatique de ces contours. Cet outil est adéquat dans la mesure où il simplifie la courbe mélodique naturelle de F0, et la structure sur les plans quantitatif et qualificatif, et rend l'analyse linguistique et pragmatique à la fois plus objective et surtout beaucoup plus rapide (intégré sous Praat, la procédure calcule tous les coefficients prosodiques, la courbe MOMEL, le codage INTSMEL, avec une durée proche du temps réel pour 70s de parole). L'analyse phonologique doublée d'analyses sémantique (champs lexicaux) et pragmatique, permet de caractériser les objectifs du locuteur et ses valeurs. Une prochaine application de cet outil dans PFC concernera les contours mélodiques propres aux communautés linguistiques du sud de la France.

Références

[1] G. Brown, A. Anderson, R. Shillcock, and G. Yule (1984).Teaching talk: Strategies for production and assessment. Cambridge University Press.

[2] G. Caelen-Haumont (2002). Perlocutory Values and Functions of Melisms in Spontaneous Dialogue. In Proceedings of the 1st International Conference on Speech Prosody, SP 2002, Aix-en-Provence, 195-198, on line <http://www.lpl.univ-aix.fr/sp2002 /papers.htm>.

[3] G. Caelen-Haumont, C. Auran (to be published).The Phonology of Melodic Prominence: the Structure of melisms. In the 2nd International Conference on Speech Prosody Proceedings, SP 2004.

[4] G. Caelen-Haumont, 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.

[5] C. d'Alessandro, P. Mertens (1995). Automatic pitch contour stylization using a model of tonal perception, *Computer Speech and Language,* 9, 257-288.
[6] J. t'Hart, R. Collier, A. Cohen, (1990). *A Perceptual Study of Intonation,* Cambridge University Press, Cambridge.
[7] D. Hirst, A. Di Cristo, R. Espesser (2000). Levels of Representation and Levels of Analysis for the Description of Intonation Systems. In Horne, M. (ed.): Prosody : Theory and Experiment. Text, Speech and Language Technbology, 14. Kluwer Academic Publishers, 51-87.
[8] D. Hirst, A. Di Cristo (1998). Intonation Systems: a survey of twenty languages. Cambridge University Press, Cambridge.
[9] F. Malfrère, T. Dutoit, P. Mertens, (1998). Fully Automatic Prosody Generator for Text-to-Speech Synthesis. *Proceedings of International Conference on Speech and Language Processing.* Sidney, Australia, 1395-1398.
[10] M. D. Pell (2002). Surveying Emotional Prosody in the Brain. In B. Bel, I. Marlien (eds.) Proceedings of the 1st International Conference on Speech Prosody, SP 2002, Aix-en-Provence, on line, <http://www. lpl.univ-aix.fr/sp2002/papers.htm>.
[11] J.F. Pitrelli, M.E. Beckman, J. Hirschberg (1994). Evaluation of prosodic transcription labeling relaibility in the ToBI framework, *Proceedings of ICSLP 94,* 123-126.
[12] M. Rossi, M. Chafcouloff (1972). Les niveaux intonatifs. *travaux de l'Institut de Phonétique,* 1, 167-176.
[13] K. R. Scherer (1989). Vocal correlates of emotion. In Wagner and Manstead Eds., Handbook of psychophysiology: Emotion and social behavior, London, Ed. Wiley.
[14] K. R., Scherer, D. R. Ladd, K. E. A. Silverman (1984). Vocal cues to speaker affect: Testing two models, J.A.S.A., 76 (5), 1346-1356.