

Figure 5. F₁-F₂ diagram for Speaker 3.

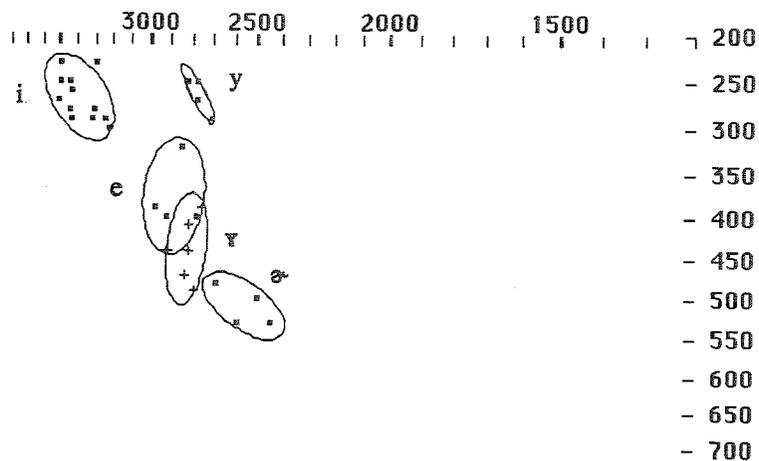
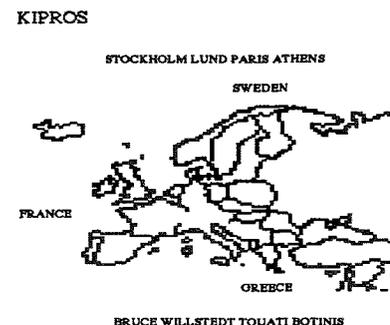


Figure 6. F₁-F₃ diagram for Speaker 3.

Lund University, Dept. of Linguistics
 Working Papers 35 (1989), 203-214



De la Prosodie Française du Dialogue Rapport du Projet KIPROS

Paul Touati

INTRODUCTION

Cet article présente une recherche en cours consacrée à la description de la prosodie du français telle qu'elle est employée dans le cadre de l'interaction verbale constituée par un dialogue. Cette étude de la prosodie dialogale du français est effectuée dans le cadre du projet KIPROS, acronyme basé sur le titre suédois du projet *Kontrastiv Interaktiv Prosodi* ou en français 'Prosodie Interactive Contrastive' (cf. Bruce et al. 1988). Le projet est soutenu financièrement par la Fondation du Tricentenaire de la Banque de Suède. L'objectif de ce projet est double. Il s'agit d'une part d'étudier dans une perspective contrastive la prosodie interactive du suédois (suédois standard et scanien), du grec et du français, langues dont la prosodie présente, on le sait, des différences structurelles intéressantes. D'autre part, le projet a pour ambition finale de développer un modèle générale de prosodie interactive.

En ce qui concerne notre contribution, disons que son effort principal a porté jusqu'à présent sur la mise au point d'une méthodologie susceptible de répondre à trois questions concernant l'analyse prosodique des dialogues du français.

La première question concerne la modélisation des données acquises pendant l'analyse acoustique ou plus particulièrement la manière dont sont mises en correspondance, à travers l'espace d'un modèle, les représentations linguistiques et les représentations acoustiques.

La seconde question concerne la possibilité d'exploiter des résultats obtenus sur un corpus lu en laboratoire (cf. Touati 1987) pour l'analyse des données acquises à partir d'un corpus spontané. A ce sujet, nous avons en préparation une étude sur l'interaction tonale entre l'accent démarcatif et l'accent focal dans le cadre d'énoncés lus, de texte lu et de dialogues spontanés (Touati en préparation).

Enfin la troisième question concerne l'intégration dans notre modèle d'analyse prosodique de catégories linguistiques spécifiquement interactives telles le tour de parole, la prise de parole, l'acte de parole, l'initiative, la réponse, le découpage en 'topique vs sous-topique' etc. Des modèles linguistiques d'interaction tels que celui de l'Initiative-Réponse (Linell & Gustavsson 1987) ou celui des jeux dialogals (Anward 1986) sont actuellement testés par mes collègues de la contribution suédoise (Bruce et al. en préparation) et celui du découpage en unités 'topique vs sous-topique' par la contribution grecque (Botinis 1989a,b).

Notre travail a suivi la démarche suivante. Au cours d'une première étape, il a été procédé d'une part à la constitution d'un corpus de base de dialogues enregistrés (émissions de radio ou programmes de télévision) et d'autre part à une analyse auditive et acoustique de fragments de dialogues (cf. Bruce et al. 1988). La seconde étape du projet a été consacrée à la modélisation des données acquises pendant l'analyse acoustique. Parallèlement à cette modélisation, une analyse par synthèse a été appliquée à quelques fragments choisis des dialogues afin d'estimer la validité perceptuelle en termes d'interaction de certaines configurations tonales (Touati & Bruce 1989).

Dans cet article, nous nous proposons de donner un premier aperçu sur la constitution du corpus et sur les analyses, sur la modélisation des données acoustiques et sur l'analyse par synthèse de fragments de dialogue. Chaque partie sera illustrée d'exemples accompagnés de remarques ponctuelles.

CORPUS ET ANALYSES

Corpus

Dans la constitution du corpus de base, les dialogues ont été choisis sur différents critères. Critère d'éclectisme: il s'agissait de ne pas clore le corpus en obéissant à un possible critère d'homogénéité risquant ainsi de retrécir

notre première approche à l'étude de la prosodie de la parole spontanée. Critère d'authenticité: engagés dans un dialogue radiodiffusé ou dans un débat télévisé, les locuteurs ne soupçonnaient pas que leurs productions orales seraient ultérieurement soumises à une analyse phonétique. Mais aussi et surtout critère de qualité technique: la qualité sonore des enregistrements était une condition nécessaire à une analyse auditive et acoustique satisfaisante des dialogues.

Dans son état actuel, le corpus est constitué par trois dialogues (précisons que le terme dialogue est générique et qu'il recouvre ici des formes d'interaction différentes). Une conversation amicale presque intime entre une adulte et une fillette d'environ cinq ans (D1=2 min 43 sec). L'adulte mène la conversation d'une manière suffisamment habile pour faire parler l'enfant d'un sujet bien grave, la mort. Ce dialogue représente un échantillon de prosodie interactive adulte-enfant.

Une interview entre un journaliste et une jeune comédienne (D2=2 min 34 sec). Le journaliste conduit l'interview. La relation entre les deux interlocuteurs est neutre. Ce dialogue représente un échantillon de prosodie fortement interactive où les interlocuteurs se doivent de coopérer.

Un débat politique entre un politicien de gauche et un politicien de droite au sujet du cumul des mandats (D3=7 min 48 sec). Ces deux politiciens chevronnés entretiennent une relation antagoniste. Une tierce personne, un journaliste, dirige les débats. Chaque politicien a un temps de parole minuté ce qui confère à chaque prise de parole l'allure d'un monologue ou d'un court discours politique, d'un texte. Ce dialogue semble constituer un bon échantillon de prosodie interactive avec des aspects fortement marqués de prosodie textuelle.

Analyses

La première analyse est auditive. Elle se traduit tout d'abord par une transcription orthographique où sont indiqués les hésitations, les rires, les passages à deux voix, les énoncés inaudibles surtout dans le cas de chevauchement de tours de parole etc. Outre le découpage en mot, elle comporte un premier découpage basé sur les pauses inter-locuteurs et intra-locuteurs.

Au cours de l'analyse auditive, un étiquetage prosodique sélectif du dialogue a été effectué. Une étiquette est assignée à une syllabe perçue comme porteuse de l'une des trois catégories prosodiques suivantes: accent démarcatif

final de groupe (étiquette: AD), accent 'focal' initial de mot ou de groupe (étiquette: AF) et jointure initiale de groupe (étiquette: JI)

L'analyse auditive soulève indirectement le problème de la transcription. Une transcription qui obéisse à la double contrainte de commodité de lecture et de pertinence d'actualisation écrite de l'oral serait en effet souhaitable. Malheureusement si la transcription orthographique s'avère le plus commode de lecture, elle n'est pas toujours la plus pertinente des transcriptions (cf. par ex. Gardès-Madray 1984). En ce qui nous concerne, nous avons décidé de lier transcription et niveaux d'analyse ou de modélisation.

La seconde analyse est acoustique. Les dialogues enregistrés ont été numérisés sur le VAX 11/730 du laboratoire de phonétique de Lund. L'extraction des paramètres pertinents (durée, intensité et fréquence du fondamental) a été réalisée à l'aide du programme API compris dans le système ILS. Le signal acoustique a ensuite été segmenté sur ordinateur (programme LABEL) en unités syllabiques à partir de tracés de l'onde sonore (oscillogramme), de l'intensité et de Fo. Au cours de la procédure de segmentation, l'identification de ces unités a été décidée à l'aide d'un feedback auditif. Aux unités ainsi identifiées ont été assignées les étiquettes prosodiques correspondant aux catégories de l'analyse auditive. Des profils temporels et des configurations tonales ont été dérivés des mesures obtenues au cours de la procédure de segmentation et d'étiquetage.

MODÉLISATION

Elaboré à l'occasion de recherches sur la prosodie du suédois, le modèle prosodique dit modèle de Lund (cf. Bruce & Gårding 1978, Gårding & Bruce 1981) a été également appliqué à l'analyse contrastive (pour une présentation du modèle et de ses concepts opératoires dans cette perspective et pour une application du modèle à l'analyse du suédois, du grec et du français cf. Gårding 1981 et Gårding et al. 1982 et pour application plus large du modèle à l'analyse du suédois et du français et pour certains remaniements du modèle cf. Touati 1987).

Le modèle tel que nous le concevons actuellement est présenté à la figure 1. Il comporte certaines reformulations et extensions à lumière d'une analyse prosodique préliminaire de fragments de dialogues (cf. Bruce et al. 1988 et Touati & Bruce 1989).

La correspondance entre les représentations linguistiques de l'input et les représentations acoustiques de l'output est assurée par trois modules. Insérés entre les représentations linguistiques et les représentations acoustiques, ces

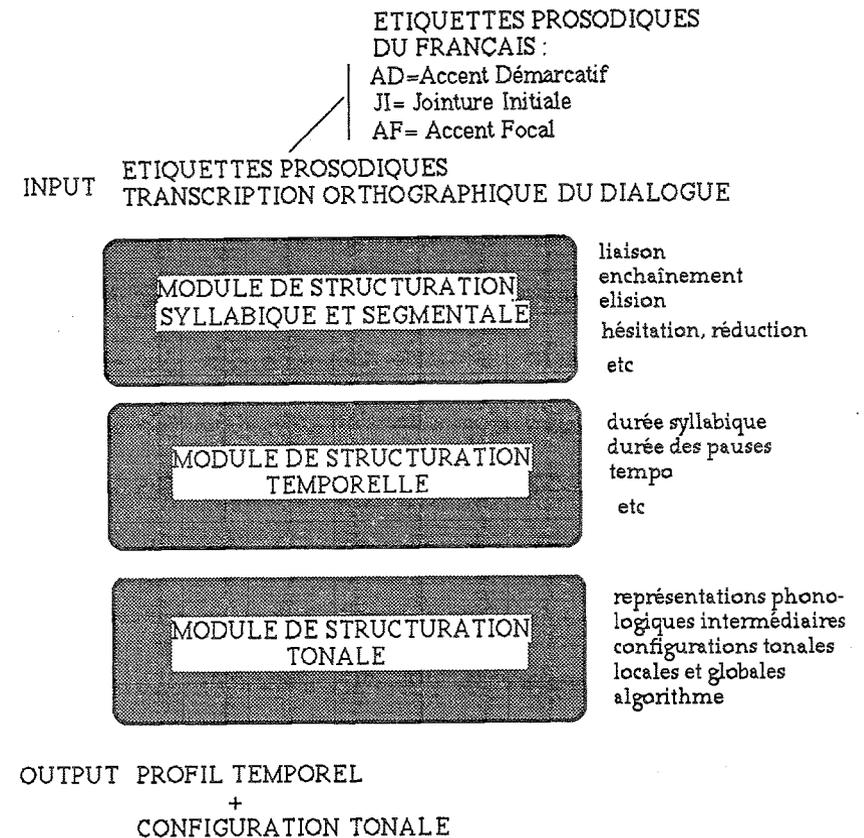


Figure 1. Modèle prosodique.

trois modules remplissent des fonctions distinctes selon la perspective qui relie ces représentations. Dans une perspective montante, c'est-à-dire qui va de l'output vers l'input (de la substance vers la forme), les modules fonctionnent comme des structures d'analyse qui sont à pourvoir en règles. Dans une perspective descendante (de la forme vers la substance), les modules peuvent, une fois fournis en règles, générer le signal acoustique. Dans l'état actuel des choses, nous travaillons essentiellement avec la perspective analytique montante, c'est-à-dire à la quantification des unités et à la production de règles.

Le module de structuration syllabique et segmentale

La fonction du module de structuration syllabique et segmentale est d'analyser les faits prosodiques qui permettent de pourvoir les modules suivants en support segmental, c'est-à-dire en lieux de segmentation pour les durées et en points de synchronisation pour les événements tonaux. Il est censé intégrer un ensemble de processus de syllabation et de transformations segmentales caractéristiques de la langue orale spontanée.

Le module de structuration temporelle

Le module de structuration temporelle a pour fonction principale d'analyser les faits prosodiques ayant trait au rythme (valeurs temporelles des syllabes inaccentuées et accentuées, des groupes), au débit (valeurs temporelles des hésitations et des changements de tempo), aux pauses (valeurs temporelles des pauses intra-locuteurs et inter-locuteurs).

Le profil temporel présenté dans la figure 2a propose une représentation des évolutions de la durée syllabique dans le temps. Ils nous renseignent sur la valeur temporelle de chaque syllabe à tout point du dialogue et sur la valeur des catégories prosodiques.

Le profil temporel de la figure 2b met par exemple en évidence la manière dont les changements de niveau de durée absolue d'une suite de syllabes répercutent les variations de tempo au cours du dialogue: *et tu sais qu'Hélène Keller a été prononcé plus rapidement que a existé.*

La figure 3 représente les valeurs moyennes de durée des différentes catégories prosodiques du dialogue D2 pour les deux locuteurs (ces valeurs moyennes ont été obtenues grâce à une extension utilitaire de calcul statistique intégrée au programme de segmentation LABEL).

A partir de notre corpus lu (Touati 1987), nous avons constaté un raccourcissement des syllabes initiales de groupes (JI) par rapport à la durée moyenne des syllabes inaccentuées (IN): ceci n'est pas le cas dans le dialogue étudié.

Cette figure nous renseigne également sur les valeurs moyennes de ces mêmes catégories pour les différents interlocuteurs (Lui vs Elle). La différenciation établie entre accent démarcatif interne à un tour de parole (AD) et en position final de tour de parole (ADF) révèle une claire hiérarchie catégorielle mais aussi une nette variabilité inter-locuteurs.

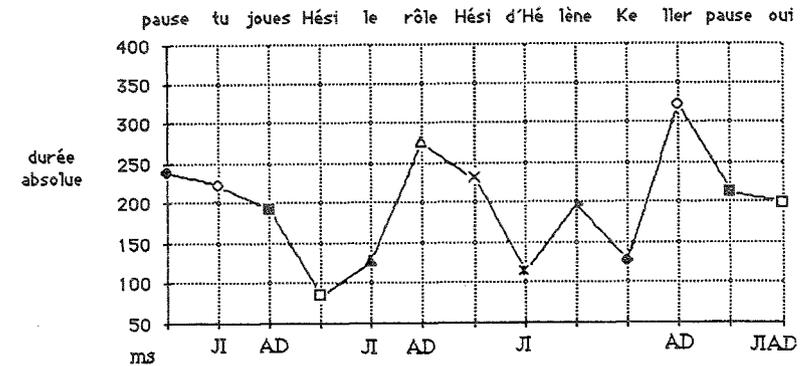


Figure 2a. Profil temporel du dialogue D2

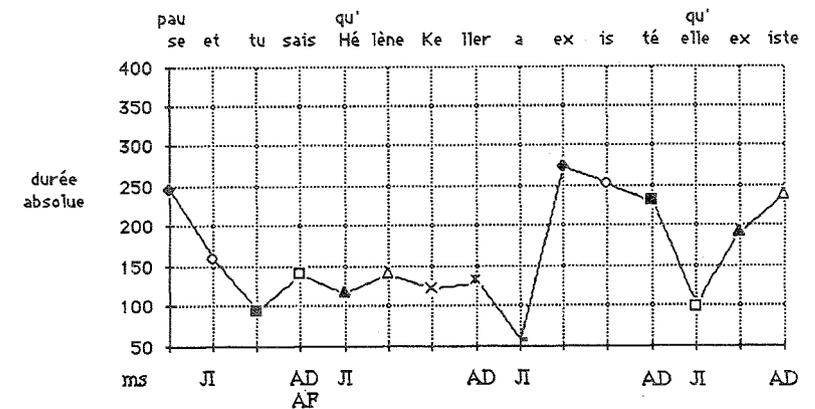


Figure 2b. Profil temporel du dialogue D2.

Le module de structuration tonale

Le module de structuration tonale a pour fonction d'analyser les faits prosodiques ayant trait à l'accentuation et à l'intonation. Les représentations phonologiques intermédiaires des catégories prosodiques AD, JI et AF sont traitées en termes de segments tonaux H (High/'Haut') et L (Low/'Bas') représentant des valeurs-cibles maxima et minima de Fo. Les configurations tonales locales et les configurations tonales globales y sont analysées qualitativement et quantitativement.

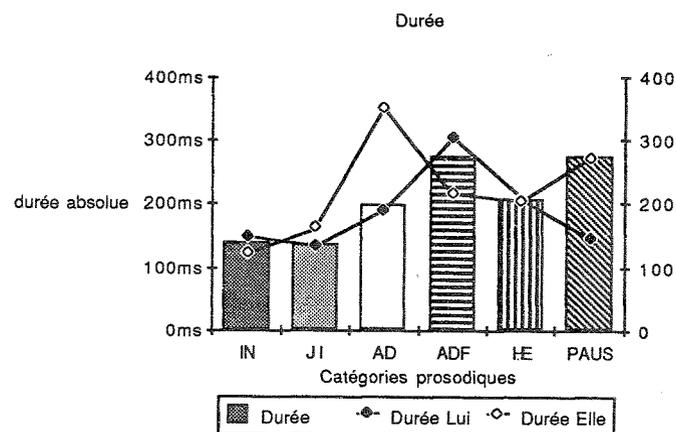


Figure 3. Durées moyennes des catégories prosodiques du dialogue D2.

Les configurations tonales locales sont essentiellement liées à l'accentuation. Nous avons en préparation une étude sur l'interaction tonale entre l'accent démarcatif et l'accent focal (Touati en préparation).

Un fragment du dialogue D3 (le débat politique au sujet du cumul de mandat) nous permettra d'illustrer la problématique des changements de registres tonaux dans leur fonction de structuration textuelle.

La figure 4 propose une transcription du fragment (input), situe le niveau d'analyse (module de structure tonale) et présente l'onde sonore et le tracé de Fo (output).

Nous pouvons observer que le locuteur emploie trois, voire quatre registres différents à la suite. Le premier registre, celui du segment de réplique (1), représente un registre tonal d'expansion 'moyenne' situé dans la zone centrale de la gamme tonale du locuteur. L'expansion du registre tonal du segment (2) se trouve réduite, avec deux niveaux de réduction, par rapport au registre du segment précédent. L'expansion du registre tonal du segment (3) se voit en revanche fortement amplifiée. Tant la réduction que l'amplification du registre tonal semble de toute apparence être obtenues par un ajustement (réduction vs augmentation) des valeurs des maxima de Fo, c'est-à-dire des valeurs tonales des points H (Haut).

INPUT:

- (1) // → une telle loi à l'Assemblée Nationale //
- (2) // ↓ mais je vois très bien / ↓ que vous ne le ferez pas //
- (3) // ↑ je crois que nous sommes le seul pays au monde //



OUTPUT : CONFIGURATION TONALE

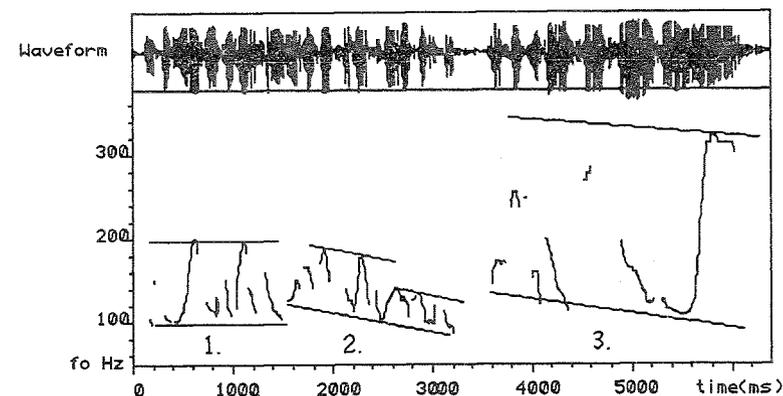


Figure 4. Input, module et configurations tonales du dialogue D3.

Trois questions se posent en relation avec ces changements de registres. La première concerne le nombre de registres tonaux que le locuteur a à sa disposition. Peut-on considérer que le locuteur possède un répertoire fixe de registres qui forment une classe paradigmatique et qui sont simplement agencés dans le temps ou bien doit-on considérer que le locuteur opère les changements de registres en appliquant une réduction ou une augmentation automatique des valeurs des maxima de Fo par rapport aux valeurs des maxima du segment précédent dans un pur contraste syntagmatique ?

La réponse à cette question détermine en quelque sorte la réponse à notre seconde question: comment transcrire les changements de registres? Dans la transcription que nous proposons, la flèche, placée au point de rupture tonale, représente un changement de registre syntagmatique: une flèche vers le bas signifie un registre réduit par rapport au registre précédent, une flèche vers le haut signifie un registre amplifié par rapport au précédent. Le choix de cette notation semble donc constituer en soi une réponse à notre première question. Il n'est pourtant pas définitif.

La troisième question concerne l'interprétation sémantique de ces changements de registres. Il paraît évident que le segment (1) pose dans un registre moyen un sujet (*une loi*) qui va faire l'objet de deux commentaires.

Le premier commentaire (segment (2)), adressé en aparté dans un registre réduit signifie trois choses: une interpellation indirecte de l'adversaire politique, une critique de la position idéologique de cet adversaire et une prise à témoin du public/électeurs qui se doit de connaître la vérité. Cette parenthèse a un effet interactif très marqué dans la mesure où elle simule une réplique possible de l'interlocuteur (ici: 'je ne le ferai pas').

Le second commentaire prononcé avec une registre très amplifié apporte un complément d'information où il est insisté sur la position unique de la France en ce qui concerne le cumul de mandats (segment (3)). Il montre par contraste tant tonal que sémantique l'incongruité de cette situation politique et de la position de l'adversaire. Ce registre tonal amplifié est maintenu pendant les apports d'information qui suivent constituant ainsi un cliché tonal lequel, repris à plaisir, donne l'impression que le discours politique s'est mis en branle et que l'interaction entre les deux interlocuteurs est minimale.

ANALYSE PAR SYNTHÈSE

L'analyse par synthèse a pour objectif de répondre à des questions générales qui ont été posées dans Bruce et al. 1988. Ici nous aimerions préciser ces questions de la manière suivante: y-a-t-il une correspondance biunivoque entre un signifié interactionnel et une configuration tonale? ou bien: un signifié interactionnel peut-il être porté par des configurations tonales différentes? ou encore: Y-a-t-il équivalence interactionnel entre des configurations différentes?

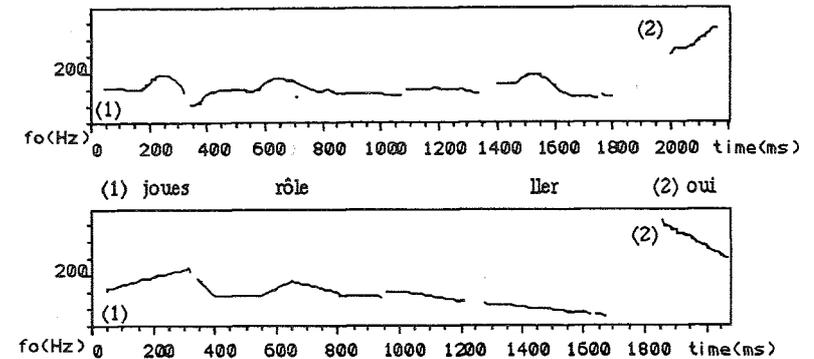


Figure 5. Configurations tonales originale (en haut) et synthétisée (en bas) du dialogue D2.

A titre d'illustration, nous avons choisi deux répliques du dialogue D2:

		JI	AD	JI	AD	JI	AD
Le journaliste	(1)	//	tu	joues	le	rôle	de / d'Hélène Keller ? //
La comédienne	(2)	//	oui	//			

Dans ce court échange à deux termes, le terme (1) signifie une demande de confirmation d'information et le terme (2) une confirmation de cette information. Ajoutons que (1) sémantise une question et (2) une réponse. Comme on peut le constater à partir de la figure 5 (en haut), la configuration tonale globale de (1) est descendante et celle de (2) montante. De plus (1) montre des variations tonales locales sur les syllabes *joues*, *rôle* et *ller*.

Nous avons changé par étapes la configuration tonale de (1) à l'aide de la synthèse LPC en vue de produire un ordre en sémantisant un impératif du genre: *tu joues le rôle d'Hélène Keller!* Nous avons produit ainsi un certain nombre de stimuli synthétiques. Le stimulus terminal est présenté dans la figure 5 (en bas). Dans ce stimulus le profil temporel a été également rectifié.

La question qui se pose est de savoir si d'une part ce dernier stimulus est par exemple acceptable en tant qu'expression tonale d'un ordre et surtout si d'autre part il est interactionnellement acceptable lorsqu'il est mis en contexte dans la bouche du journaliste.

Une problématique similaire peut être posée à propos du changement de configuration du terme (2).

REFERENCES

- Anward, J. 1986. 'Emotive expressions'. *Papers from the Ninth Scandinavian Conference of Linguistics*, ed. Östen Dahl, 39-52. Dept. of Linguistics, University of Stockholm.
- Botinis, A. 1989a. *Stress and prosodic structure in Greek*. Lund: Lund University Press.
- Botinis, A. 1989b. 'Discourse Intonation in Greek'. *Working Papers* 35, 5-23. Dept. of Linguistics, Lund University.
- Bruce, G. & E. Gårding. 1978. 'A prosodic typology for Swedish dialects'. *Nordic Prosody*, eds. Eva Gårding et al., 219-228. Dept. of Linguistics, Lund University.
- Bruce, G., P. Touati, A. Botinis & U. Willstedt. 1988. 'Preliminary Report from the KIPROS Project'. *Working Papers* 33, 23-50. Dept. of Linguistics, Lund University.
- Gardès-Madray, F. 1984. 'Praxématique et interaction verbale'. *Langages* 74, 15-29.
- Gårding, E. 1981. 'Contrastive prosody: A model and its applications'. *Studia Linguistica* 35, 146-166.
- Gårding, E., A. Botinis & P. Touati. 1982. 'A comparative study of Swedish, Greek and French intonation'. *Working Papers* 22, 137-152. Lund: Dept. of Linguistics.
- Gårding, E. & G. Bruce. 1981. 'A presentation of the Lund model for Swedish intonation'. *Nordic Prosody II*, ed. Thorstein Fretheim, 33-39. Trondheim: Tapir.
- ILS. 'Interactive Laboratory System' V4 (1982) & V5 (1985). Signal Technology, Inc. Goleta, California.
- LABEL. 'Programme de segmentation et d'étiquetage'. L. Eriksson, Laboratoire de Phonétique, Université de Lund.
- Linell, P. & L. Gustavsson. 1987. *Initiativ och respons. Om dialogens dynamik, dominans och koherens*. Studies in Communication (SIC) 15. University of Linköping.
- Touati, P. 1987. *Structures prosodiques du suédois et du français*. Lund: Lund University Press.
- Touati, P. & G. Bruce. 1989. 'Report from the KIPROS Project'. *STL-QPSR* 1, 51-54. Stockholm: KTH, Speech Transmission Laboratory.