

Transcription Prosodique Normalisée au sein du Projet PFC (Phonologie du Français Contemporain) : l'Etat d'un Chantier

Anne Lacheret-Dujour¹, Chantal Lyche²

¹Institut Universitaire de France, Paris,
Laboratoire CRISCO, Université de Caen 14032 Caen Cedex

Anne.lacheret@crisco.unicaen.fr

²Universitetet i Oslo/Tromsø
chantal.lyche@ilos.uio.no

1. Introduction

L'objectif de notre communication est d'exposer l'approche retenue pour le codage prosodique de la base de données PFC (Phonologie du Français Contemporain), amorcé en décembre 2003. Après avoir présenté le projet dans ses grandes lignes, et défini le domaine de la prosodie sous l'angle phonétique, phonologique et symbolique, il s'agira de discuter les limites tant théoriques que pratiques de ce codage, d'expliquer pourquoi les secondes découlent en partie des premières, illustrant ainsi le fait qu'il n'existe pas d'outils et d'instrumentation utilisables et robustes sans un minimum épistémologique commun, cela quel que soit le niveau de traitement linguistique concerné, en l'occurrence le niveau suprasegmental. En pratique, nous essaierons de comprendre pourquoi ce minimum commun est particulièrement difficile à stabiliser quand il s'agit de travailler sur des données prosodiques, données effectivement particulièrement résistantes à l'approche structuraliste du linguiste d'une part, à l'instrumentation d'autre part, enfin au traitement automatique. Dans ce contexte, nous présenterons les objectifs actuels du codage prosodique dans PFC, discuterons des logiciels utilisables pour ce codage, pour finir : notre méthodologie de transcription sera présentée. Pour clore cette présentation, nous conclurons sur nos perspectives de recherche après avoir résumé les hypothèses fortes qui sous-tendent notre démarche.

2. Le projet PFC

Le propos de ce paragraphe est de rappeler les objectifs du projet PFC (voir aussi Durand & Lyche 2003) et le domaine de la prosodie sous l'angle substantiel (phonétique), formel (phonologique) et fonctionnel (fonctions syntaxiques et pragmatiques de l'accent et de l'intonation).

2.1. Les objectifs du projet

Le projet *Phonologie du français contemporain : usages, variétés et structures* (<http://www.projet-pfc.net>) se situe dans la tradition des grandes enquêtes

phonologiques et vise à constituer sur 50 points d'enquête un grand corpus oral d'environ 500 locuteurs, constitué de données strictement comparables. Tous les enregistrements sont effectués sur la base d'un protocole unique avec la prise en compte de quatre registres (liste de mots et texte lus, conversation dirigée et libre). L'objectif premier de PFC est de nature phonologique. Le corpus constitué doit permettre des analyses de phénomènes typiquement variables, et ce faisant de tester les modèles phonologiques et phonétiques qui accordent une large place à la variation. Dans un premier temps l'accent est mis sur le système phonémique du locuteur, sur la prononciation variable du 'e' caduc (schwa) et de la liaison. Pour chaque locuteur, le texte ainsi que quelques minutes d'entretien sont codés sous PRAAT, sur deux tires traitant respectivement le schwa et la liaison. Le codage, effectué sur la base de la transcription orthographique, est alphanumérique, son rôle est d'offrir un premier balayage des données pour les analyses ultérieures. Des principes de codage équivalents ont été envisagés pour la transcription de la prosodie tout en sachant que celle-ci, très lourde à mettre en œuvre et à exploiter réclame également ses propres contraintes de codage (cf. *infra*, 3.3).

2.2. Pourquoi la prosodie dans le projet PFC ?

D'une manière générale, la transcription prosodique poursuit deux pistes dont seule la première concerne les auteurs du papier : (i) mettre en valeur le lien entre deux phénomènes hautement variationnels (le schwa et la liaison) et la prosodie, et (ii) dégager les caractéristiques prosodiques générales de chaque variété de français étudiée dans un but purement descriptif et comparatif. Dans cette communication, nous nous centrerons exclusivement sur l'interaction schwa/prosodie.

2.3. Le domaine de la prosodie : substance, forme, fonction

Tenter de saisir la portée définitoire du domaine n'est pas une mince affaire (Lacheret & Beaugendre 1999), tant les définitions sont multiples, plus ou moins opaques et circulaires, centrées sur différents niveaux d'analyse maladroitement amalgamés : paramètres phonétiques actualisés dans la substance, formes prosodiques rattachées au système de la langue, fonctions linguistique et expressive de la prosodie. Dans ce contexte, il nous semble raisonnable d'envisager une définition modulaire du domaine, de ses caractéristiques substantielles aux fonctions que la prosodie assume dans le langage, même si en pratique, les différents niveaux de traitement interagissent étroitement. La figure 1 ci dessous présente pour chaque niveau (phonétique vs. phonologique) les objets prosodiques manipulés ; les événements phonétiques (en premier lieu fréquence fondamentale et durée) servant d'indices au repérage des structures accentuelle et intonative qui font l'objet de modélisations phonologiques. La figure 2 illustre concrètement l'évolution des paramètres prosodiques au cours du temps. La figure 3 enfin résume les fonctions susceptibles d'être véhiculées par la prosodie dans les langues à accent.

Prosodie : formes-substance

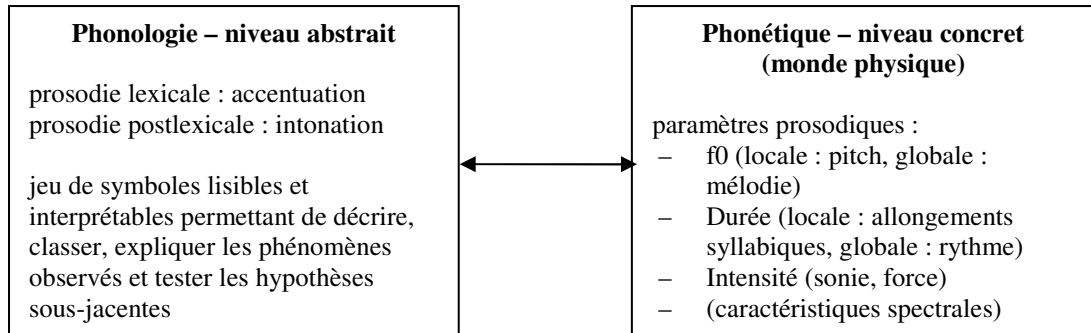


Figure 1 : les domaines phonologique et phonétique de la prosodie

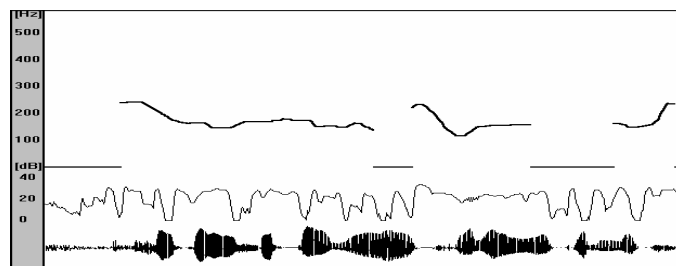


Figure 2 : représentation tridimensionnelle des paramètres prosodiques avec de bas en haut, respectivement, le signal acoustique, l'intensité en décibels, et la fréquence fondamentale en hertz, en repérant par des valeurs nulles les zones non voisées ; sur l'axe des abscisses le temps en millisecondes.

Prosodie : formes-fonctions

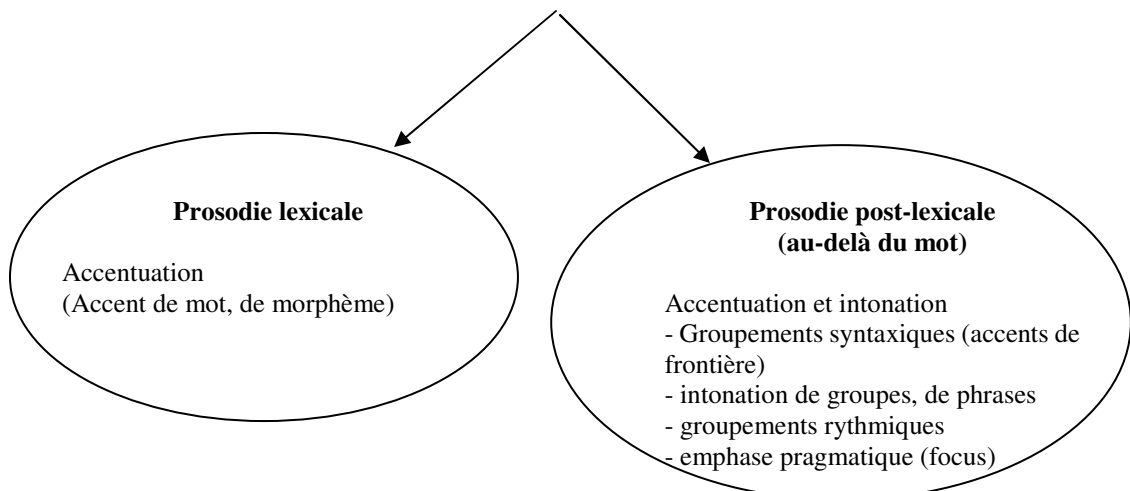


Figure 3. les fonctions lexicales et post-lexicales de la prosodie

Concernant la prosodie du français, la figure 3 appelle une remarque importante : le français faisant partie des langues à accent fixe, la fonction morphologique

éventuellement assignée à l'accent reste tenue dans cette langue. L'accent y assumant essentiellement une fonction postlexicale est donc étroitement associé à la structuration intonative qui en découle directement. Autrement dit, en français, l'accent primaire est un accent de groupe et non de mot, il est localisé obligatoirement sur la syllabe finale du dernier mot d'un groupe syntaxique, éventuellement sur la première syllabe de ce groupe, surtout quand ce dernier est long, il revêt ainsi, au sein du groupe accentuel, une fonction démarcative à droite (*le chat noir, prends-le*) ou à gauche (*Marie et Paul s'amusaient*). Outre l'accent démarcatif, on peut rencontrer des accents dit « secondaires », dont la position est fluctuante à l'intérieur des mots constitutifs d'un groupe accentuel (essentiellement la première syllabe d'un mot quand le contexte phonétique le permet). L'accent secondaire assume une fonction pragmatique (mise en valeur de la partie importante du message : emphase informationnelle ou expressive : *c'est un foulard que je lui ai offert, c'est merveilleux*) et souvent rythmique (équilibre syllabique entre deux sites accentuels).

3. Le codage prosodique proposé dans PFC

Une fois résolu, pour la tâche envisagée, le problème de la définition du domaine de la prosodie dans sa dimension phonétique, formelle et fonctionnelle, reste à formuler clairement et explicitement les hypothèses que le codage permettra de tester concernant l'interaction prosodie/schwa, le type de codage le plus optimal pour cette tâche et les outils logiciels associés, enfin la nature des données à coder sur le plan qualitatif et quantitatif. Par ailleurs, un travail de meta-analyse s'impose qui permettra de retenir les aspects les plus opératoires et innovants de la démarche envisagée.

3.1. Hypothèses sur l'interaction schwa/prosodie

Les hypothèses, telles que nous les avons formulées, s'articulent autour de trois thématiques, la première concerne l'interface entre le niveau phonologique profond (input) et le traitement phonologique de surface (output), la seconde s'inscrit dans une problématique rythmique et psycho-acoustique concernant le rôle des constructions prosodiques dans l'intégration perceptive des groupes intonosyntaxiques, la troisième concerne le conditionnement pragmatique des formes intonatives, *i.e.* l'interaction formes/fonction communicative. Pour ce qui est de la première thématique, la question récurrente posée en phonologie du schwa concerne la façon dont le lien entre forme profonde et forme de surface peut s'établir : la réalisation de surface (dans la parole) correspond-elle dans la structure profonde à un schwa encodé dans le lexique mental, pouvant être éliminé en surface ou s'agit-il d'un segment inséré, *i.e.* qui n'a aucune réalité dans la structure profonde ? L'hypothèse que nous défendons est la suivante : si nous sommes face à un processus d'élimination, alors la prosodie en porte les traces en surface dans la mesure où la voyelle avant et/ou après le segment éliminé est systématiquement différente (durée et/ou patron mélodique) de la même voyelle dans un autre contexte (*ex. coll'rette vs colmater*). Au sujet des problèmes d'intégration perceptive, la question est de savoir dans quelle mesure d'une part le nombre de syllabes d'un groupe joue un rôle dans la réalisation d'un 'e' muet (longueur du groupe), d'autre part la distance entre le schwa et l'accent primaire influence ou

non sa réalisation. L'hypothèse générique que nous posons repose sur le conditionnement rythmique du schwa, elle s'articule autour des deux points suivants : toute chose égale par ailleurs, plus un groupe est long, plus le 'e' aura tendance à tomber (*ex. portefeuille vs. porte-monnaie*), plus un 'e' est proche de l'accent, plus il aura tendance à être éliidé, en conséquence la position initiale est plus stable que la position médiane, qu'il s'agisse d'un mot (*secouer vs seulement*) ou d'un groupe (*je sais que c'est comme ça*). Concernant enfin le conditionnement pragmatique, deux questions se posent : (i) dans quelle mesure l'organisation topologique d'une séquence (début ou non d'un tour de parole, initiale ou non d'un groupe prosodique) influence la prononciation du 'e' ? (ii) l'organisation informationnelle joue-t-elle un rôle sur la distribution des 'e' réalisés ? A cet égard, les hypothèses que nous défendons se résument comme suit : (i) la position initiale de mot, tout comme le 'e' monosyllabique est une position de force de degré variable en fonction de sa distribution topologique dans le discours : le 'e' fréquemment éliidé au milieu d'un tour étant plus préservé à son initiale (*ça nous regarde pas vs. retrouver sa joie de vivre*). (ii) Un 'e' est plus facilement effacé à l'initiale d'un segment topical non informationnel que rhématique, porteur d'information, soit les segments suivants pour illustration : (a) *il vient d'main avec son frère* (b) *il vient demain avec son frère* en réponse respectivement aux questions : (a) *avec qui vient-il demain ?* (b) *quand vient-il avec son frère ?*

3.2. Logiciels de codage : quelques remarques

Le problème de fond, sans doute ici plus qu'ailleurs étant donné l'aspect intrinsèquement continu de la matière à décrire, est la quête d'un codage mutualisable, tel qu'il a pu être réalisé au sein de PFC pour le niveau segmental. Avant d'aborder la mutualisation du traitement prosodique en tant que tel (cf. *infra* 3.3), se pose la question de la mutualisation des outils. Nous avons rappelé en 1.1. que nous souhaitions continuer à utiliser PRAAT pour le codage de la prosodie, l'ensemble du codage (schwa, liaison, prosodie, transcription orthographique) étant ainsi disponible dans un et un seul format textgrid sur 4 tires de codage. Or, ce choix ne va pas sans problèmes, le principal étant lié aux fonctionnalités du logiciel : si PRAAT remplit parfaitement sa fonction de logiciel de traitement et d'analyse phonétique, en revanche il n'a pas été conçu pour servir de logiciel de catalogage et de codage, fonction que nous lui faisons porter rempli par l'utilisation de tires parallèles où figurent sur un même plan données et métadonnées. (Baude & al. 2005). Par ailleurs, le traitement ne peut être que manuel, ce qui suppose un temps de codage significatif et un gros nettoyage des données en aval (contrôle de la transcription) pour lancer les analyses automatiques qui suivront. En revanche, d'autres logiciels proposent déjà des outils d'aide au traitement, c'est le cas de WinPitch (Ph. Martin dév.) qui intègre une fonction d'aide à l'alignement, un ralentisseur de débit pour faciliter le travail de segmentation, la possibilité d'utiliser des balises XML permettant de structurer codage et catalogage, etc. Un transfert vers WinPitch paraîtrait donc souhaitable d'autant que l'interface PRAAT-WinPitch développée depuis peu par Ph Martin permet de récupérer les transcriptions déjà existantes réalisées sous PRAAT. Mais alors d'autres problèmes se posent découlant directement des caractéristiques du logiciel qui n'est toujours pas open source aujourd'hui, et qui reste payant pour les

collectivités, ces deux caractéristiques étant totalement en contradiction avec une philosophie de mutualisation. Bref, la question suivante, récurrente en TAL et en linguistique de corpus, reste ouverte : comment trouver le meilleur compromis entre des objectifs de partage et de mutualisation d'un côté, un traitement plus pointu, intégrant des outils d'aide à la transcription de l'autre ?

3.3. Méthode et protocole de codage

Concernant la démarche méthodologique d'ensemble, si l'opposition entre méthode hypothético-déductive et stratégie inductive est souvent posée comme allant de soi, il s'agit dans la pratique d'un artefact. En effet, de même que l'éclairage des hypothèses par les données suppose un va et vient étroit entre hypothèses et données, une méthode uniquement inductive est illusoire : on ne cherche pas n'importe quoi, n'importe où, des hypothèses posées préalablement orientent drastiquement la recherche. Soit le cheminement suivant (figure 4):

- Dans une phase initiale, un premier jeu d'hypothèses va permettre de mettre en place un protocole de codage préalablement à la sélection des contextes à coder et à leur codage effectif. Une première exploration des données va nous amener à préciser nos hypothèses et à modifier notre protocole le cas échéant, notamment à restreindre ou étendre les champs de codage et les informations qui s'y déploient.
- Ce n'est que lorsque cette phase initiale est vraiment stabilisée que des requêtes automatiques peuvent être lancées afin de fournir des résultats statistiques qualitativement et quantitativement représentatifs pour l'analyse des données, conformément aux hypothèses stabilisées.

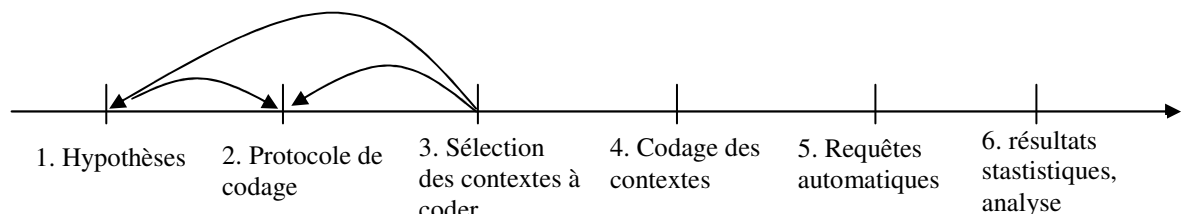


Figure 4 :

la démarche sur corpus, des hypothèses aux données, des données aux hypothèses

Nous ne reviendrons pas sur les hypothèses telles qu'elles ont été posées en 2.1, attardons-nous en revanche sur la stratégie de codage retenue : type et nature du codage, sélection des contextes à coder. Nous avons opté pour un codage générique quel que soit le type de production transcrite (lecture de texte, conversation dirigée, conversation libre) en sachant que certains marqueurs ne seront jamais utilisés dans des types de production et abondants dans d'autres (voir par exemple les marqueurs codant les pauses d'hésitations inutiles pour la transcription du mode <lecture de texte>).

Le codage se subdivise en deux ensembles : un codage standard réalisable par des non spécialistes entraînés (2 heures d'apprentissage environ) et un codage étendu qui nécessite en partie une expertise phonétique et des logiciels dédiés. En pratique, l'architecture globale du protocole comprend 9 champs, les deux premiers

fixant nos domaines de codage prosodique (la syllabe et le mot), les 7 autres fournissant les informations prosodiques exploitées pour les tests de nos hypothèses. Précisément, les 6 champs du codage standard se décomposent comme suit : le **champ 1** indique le nombre de syllabes prononcées du mot en cours de traitement, le **champ 2** code le rang de la syllabe analysée dans le mot, le **champ 3** renseigne sur la perception de la syllabe : proéminente ou non, Le **champ 4** indique des variations de durée syllabique éventuelles, Le **champ 5** donne des informations sur les occurrences des pauses produites et leur nature. Précisons que, pour l'heure, ce champ reste relativement sommaire par rapport aux descriptions que l'on a pu proposer sur les types de dysfluences rencontrées dans le discours (pour une étude récente, voir Assié 2005). Nous distinguons les pauses silencieuses courtes et longues (seuil de référence = 300 ms), les 'euh' d'hésitation et les répétitions syllabiques (*c'est in/incroyable*). Le **champ 6** permet d'élargir le domaine de codage au sein du discours (syllabe initiale ou non de tour de parole). Le protocole étendu pour sa part renseigne sur les corrélats psycho-acoustiques des objets phonologiques dans les distributions observées. En pratique, il s'agit de tenir compte du niveau fréquentiel des noyaux syllabiques (champ 7), de la configuration mélodique des contours syllabiques, pente montante ou descendante, forte ou faible (champ 8) et de l'intensité associée (champ 9).

Concernant enfin la sélection des contextes, rappelons que le codage de la prosodie se distingue significativement du codage segmental dans la mesure où les unités à coder ne sont plus des lexèmes/graphèmes mais inévitablement des syllabes. La segmentation syllabique est en effet incontournable si on veut disposer d'informations prosodiques minimales certes mais suffisamment exhaustives pour pouvoir en tirer des informations exploitables. Un tel travail est une tâche lourde et fastidieuse, et, de fait, non dénuée de risques d'erreurs qu'il faut pouvoir, à terme, contrôler au maximum (cf. *infra*, conclusion). En conséquence, si plusieurs stratégies de sélection des passages sont *a priori* envisageables, peu s'avèrent tenir la route bien longtemps. En particulier, on pourrait souhaiter maintenir le parallélisme avec le codage schwa et le codage liaison en proposant la transcription aléatoire d'une portion du texte et d'une minute de chaque entretien par exemple. Cette procédure peut fournir des résultats fructueux à condition que les occurrences recherchées soient suffisamment fréquentes ce qui est souvent loin d'être le cas. Bref, il est vite apparu qu'une démarche non contrôlée ne donnerait pas de résultats pertinents, étant donné notre objectif particulier qui vise à mettre en relation la présence/absence du schwa avec la prosodie dans des contextes bien circonscrits. Une deuxième solution alors consiste à sélectionner attentivement les passages à coder en ayant constamment à l'esprit les types de requêtes que nous souhaitons effectuer par la suite, de manière à tester des hypothèses phonologiques clairement formulées au préalable. Bien entendu, nous avons opté pour cette seconde approche sans pour autant verser dans un optimisme débordant. La parole spontanée reste par définition spontanée sans garantie aucune de fournir tous les types de contextes que nous nous proposons d'envisager, d'où la nécessité d'expériences complémentaires en laboratoire (cf. *infra*, conclusion). Environ cinq locuteurs sont retenus pour chaque point d'enquête soumis au codage prosodique et la sélection s'établit sur la base de la qualité du signal, la présence de variations mélodiques bien nettes, enfin la richesse potentielle des contextes. Les passages à coder du

texte sont préétablis et favorisent l'étude des sites d'insertion attendue de schwa et les quelques instances de monosyllabes où l'absence de schwa est envisageable. Pour les conversations (dirigée et libre), nous préconisons le codage d'une minute de chaque type conversationnel pour chacun des témoins retenus. Pour finir, précisons que les passages sélectionnés sont discontinus et sont choisis essentiellement de façon à mettre en valeur les patrons prosodiques qui interpellent le codeur, *i.e.* des patrons que l'on qualifiera de « marqués » par rapport à l'intuition du codeur de ce que peut être une prononciation non marquée ou standard.

4. Conclusion

Dans cette article, nous avons souhaité baliser les grands axes de ce que peut être une base de données phonologique dans sa composante prosodique en interaction avec la couche segmentale. La démarche unifiante de ce travail est la suivante : la prosodie permettrait de confirmer certaines hypothèses au détriment d'autres concernant la prononciation variable des segments, en l'occurrence le schwa (ou 'e' caduc ou encore 'e' muet). Elle constituerait notamment un indice précieux pour confirmer l'hypothèse d'un processus d'élision dans la parole d'un segment encodé dans le lexique mental des locuteurs ; à l'inverse, la prosodie peut agir comme un déclencheur pour la réalisation du schwa, il existe en particulier une relation entre les structures rythmiques et la prononciation du 'e' caduc. Evidemment, pas de théorie sans méthodes et sans outils surtout quand il s'agit d'explorer des données à grande échelle, c'est bien là le cœur du propos. Cet aspect fondamental nous a amenées à exposer notre programme méthodologique autour de 3 points : (1) la mise en place d'un premier jeu d'hypothèses pour le développement d'une série de requêtes en cours d'implémentation, (2) notre protocole de codage tel qu'il est stabilisé aujourd'hui après deux ans de réflexions et d'essais sur corpus, (3) la stratégie de sélection des contextes pour l'analyse. Si nous attendons beaucoup de cette première phase qui nous permettra de préciser les processus de variation dans lesquels interagissent effectivement les niveaux segmental et suprasegmental, et si un effort colossal a été effectué pour une sélection rigoureuse des contextes sur différents points d'enquête, encore une fois, nous restons réalistes quant à leur représentativité tant d'un point de vue quantitatif que qualitatif. En d'autres termes, il est évident que pour l'ensemble des contextes dont nous souhaitons rendre compte, il y aura nécessairement des trous, les données étant parfois étrangement silencieuses... Deuxième problème : celui de l'outillage. La tâche de transcription manuelle est un travail long et fastidieux, nous l'avons dit, les erreurs inévitables. Des outils d'assistance au codage (syllabation semi-automatique, détection de proéminences) et des outils de contrôle sont donc indispensables. Dans cette optique, le développement d'un contrôleur pour la vérification des étiquettes est en cours de discussion avec l'équipe PFC du laboratoire Modyco (coll : A. Tchobanov, R. Walter). Troisième restriction : certaines mesures fines comme celle de l'allongement vocalique éventuel suite à la chute du schwa demandent un contrôle drastique des variables manipulées, d'où la nécessité de passer le relais à ce stade à la phonologie de laboratoire. D'une manière générale, c'est sans doute cette collaboration soutenue entre approche de terrain et approche de laboratoire qui permettra de reconfigurer le paysage de la

linguistique contemporaine en proposant des outils pour tester les hypothèses des linguistes et alimenter ainsi les vieux débats, mais aussi pour en ouvrir de nouveaux.

Références

- Assié D.** 2005. *Analyse syntaxique automatique de corpus oraux retranscrits*. Mémoire de DEA en Sciences du langage, Toulouse le Mirail.
- Durand J., Lyche Ch.** 2003. 'Le projet 'Phonologie du Français Contemporain' (PFC) et sa méthodologie'. In E. Delais-Roussarie et J. Durand (eds.). *Corpus et variation en phonologie du français*. Toulouse : PUM, 213-276.
- Baude O. M. Jacobson, Tchobanov A. Walter R.** 2005. 'Interopérabilité des corpus sonores : le cas des corpus en français'. *Phonological Variation : the Case of French*, University of Tromsø, Norway, 25-27 August 2005, <http://www.projet-pfc.net/>.
- Lacheret A., Beaugendre F.** 1999. *La prosodie du français*, Paris, Editions du CNRS.