Le projet PACOMUST, un corpus de PArole COntinue MUltiSTyle : objectifs et choix méthodologiques

Corine Astesano, Roxane Bertrand, Martin Brousseau, Michel Chafcouloff, Albert Di Cristo, Alain Ghio, Daniel J. Hirst, Sophie Lapierre, Pascale Nicolas, Pascal Roméas, et al.

TRAVAUX DE L'INSTITUT DE PHONETIQUE D'AIX

Volume 16, 1995, pp. 9-38

LE PROJET PACOMUST, UN CORPUS DE PAROLE CONTINUE MULTISTYLE OBJECTIFS ET CHOIX METHODOLOGIQUES

Corine ASTESANO
Roxane BERTRAND
Martin BROUSSEAU
Michel CHAFCOULOFF
Albert DI CRISTO
Alain GHIO
Daniel HIRST
Sophie LAPIERRE
Pascale NICOLAS
Pascal ROMEAS
Frédéric SABIO
Magali VINCENT

LE PROJET PACOMUST, UN CORPUS DE **PA**ROLE **CO**NTINUE **MULTIS**TYLE : OBJECTIFS ET CHOIX METHODOLOGIQUES

Résumé

Nous présentons, dans cet article, le cadre général d'un projet de recherche collectif: le projet PACOMUST, dont les buts principaux, à court et à long terme, sont la constitution d'une base de données de parole multistyle et la description des caractéristiques segmentales et prosodiques des corpus de cette base. Compte tenu de ces objectifs, nous insistons plus particulièrement sur la présentation et la discussion des aspects théoriques et méthodologiques qui président à la réalisation de la base et à son analyse. Dans cette perspective, notre étude comprend trois parties. La première est consacrée à la présentation de la typologie des interactions verbales, sur laquelle nous nous fondons pour classer les échantillons de la base. La seconde concerne l'organisation pratique de la base de données proprement dite. Enfin, nous exposons dans la dernière partie notre méthode d'étiquetage et de transcription prosodiques en justifiant les raisons qui ont guidé nos choix.

Abstract

The overall framework of the project PACOMUST is presented. The main objective of the joint research group is to collect large quantities of data on various speaking styles in French and to describe the segmental and suprasegmental characteristics of this speech data. In this paper, emphasis has been laid on theoretical and methodological choices.

The first part of the article outlines the typology of verbal interactions we use to classify speech data. The second part concerns the practical organisation of the data, followed by an account of the labelling and prosodic transcription method.

LE PROJET PACOMUST, UN CORPUS DE PAROLE CONTINUE MULTISTYLE : OBJECTIFS ET CHOIX METHODOLOGIQUES

Corine Astésano, Roxane Bertrand, Martin Brousseau, Michel Chafcouloff, Albert Di Cristo, Alain Ghio, Daniel Hirst, Sophie Lapierre, Pascale Nicolas, Pascal Roméas, Frédéric Sabio, Magali Vincent
(par ordre alphabétique)

Institut de Phonétique d'Aix-en-Provence
Laboratoire "Parole et Langage" URA 261, CNRS
29, Av.R.Schuman, 13621 Aix-en-Provence, FRANCE
email : dicristo@univ-aix.fr {phonetic, hirst, pascal}@univ-aix.fr

1. Introduction

1.1 Objectifs

Le projet PACOMUST (Parole Continue Multistyle), dont nous présentons ici les objectifs et les principales options méthodologiques, s'inscrit dans le cadre des activités du Groupe Aixois de Recherches en Prosodie (GARP). Ce Groupe, qui a été créé en 1994 au sein du Laboratoire "Parole et Langage" de l'Institut de Phonétique d'Aix, à l'initiative d'Albert Di Cristo, est composé de chercheurs en poste et de doctorants dont l'intérêt scientifique commun réside dans l'étude des caractéristiques phoniques, segmentales et prosodiques du français dans des situations de communication diverses et donc des phonostyles différents.

Les objectifs à long terme du Groupe sont, en premier lieu, la réalisation, la gestion et l'analyse d'une base de données de parole multistyle pour le français. En effet, notre intention est de constituer un capital de travail qui soit réutilisable pour des études descriptives et perceptives de la parole, à la fois par les chercheurs du Groupe et par ceux qui souhaiteraient utiliser ce matériau. Dans cette perspective, nous pensons plus particulièrement à la transcription et à l'étiquetage manuels des corpus qui représentent des opérations fastidieuses et coûteuses en temps.

D'autre part, nous avons entrepris d'effectuer une description des faits segmentaux et prosodiques (accentuation, organisation temporelle, rythme et intonation) fondée sur l'analyse pluriparamétrique des corpus qui constituent la base de données. Cette description est destinée à alimenter la base de connaissances sur la prosodie du français que nous nous efforçons de constituer depuis plusieurs années au sein de l'Institut de Phonétique d'Aix.

A court terme, le but principal du Groupe est de définir les principes méthodologiques qui seront mis en oeuvre pour atteindre les objectifs que nous venons de préciser.

1.2. Problématique

Le projet de recherche dont nous venons d'exposer brièvement l'orientation, nous paraît être légitime à plus d'un titre. Un consensus s'est établi depuis peu dans la communauté internationale sur la nécessité d'avoir recours à l'étude de vastes corpus, afin d'extraire de la multiplicité des données observables les informations linguistiques qui permettent d'établir la spécificité du système prosodique d'une langue ou de comparer diverses langues entre elles, plusieurs dialectes d'une même langue, voire plusieurs usages situationnels d'une même variété linguistique. Ce type de corpus, qui doit être organisé et comporter, pour être efficace et crédible, de nombreux locuteurs et contenir des échantillons de parole représentatifs de la langue actualisée en discours, demeure aujourd'hui encore assez rare pour ce qui concerne le français. On peut cependant faire référence aux corpus "Sankoff-Cedergren" (1971), "Montréal 84, pour la sociolinguistique[1] et l'analyse du discours ainsi qu'au corpus "GARS-GEDO" (1995), pour l'étude de la syntaxe du français parlé. Dans notre perspective, nous considérons que la démarche adoptée il y a une quinzaine d'années par Brown & al. (1980) est exemplaire, puisqu'elle consistait à fonder l'étude prosodique sur l'analyse de plus d'une centaine d'interviews de locuteurs originaires de la région d'Edimbourg.

Dans le cas d'un corpus étendu et organisé comme celui que nous entreprenons de constituer, il est souhaitable que l'extraction de l'information linguistique dont il a été question plus haut puisse être effectuée au moyen de procédures automatiques, ce qui n'est pas monnaie courante, bien que plusieurs tentatives prometteuses aient déjà été effectuées dans cette voie (Faraht, 1992: Dalsgaard, 1991 pour le segmental; Wang & Hirschberg, 1992; Wightman & Ostendorf, 1994; Campbell, 1994 pour la prosodie). Nous avancerons cependant l'idée que l'extension de la base (ou des bases) de données devrait permettre d'atteindre à plus ou moins long terme cet objectif (Hirst, 1994). En effet, il est permis de penser que le développement d'une base de données contribue à étendre notre capacité d'effectuer des prédictions empiriques sur les données de la base même, ce qui permet de franchir progressivement des étapes décisives en vue de l'élaboration de traitements entièrement automatiques.

Outre l'apport au développement technologique, la confrontation des prédictions et des données observées conduit à formuler des hypothèses de plus

[1]Ce corpus est constitué d'entretiens en vue d'études de linguistique variationniste. Le corpus "Montréal 84" est la prolongation du premier corpus en ce sens que la moitié des locuteurs du corpus Sankoff-Cedergen ont été enregistrés de nouveau en 1984. Pour plus de détails, voir Thibault et Vincent, 1990.

en plus contraignantes sur la nature des représentations linguistiques que l'on vise à établir. On relève plusieurs propositions récentes de modélisation de la prosodie du français, dont certaines ont été présentées par des chercheurs de l'Institut de Phonétique d'Aix (Hirst & Di Cristo, 1984; Di Cristo & Hirst, 1993, 1994; Rossi, 1994). Toutefois, les tentatives de ce type sont le plus souvent fondées sur l'observation de corpus relativement peu étendus ou limités à un style de parole particulier. Il est à noter, cependant, que ces travaux ont contribué au développement d'outils d'analyse performants (Hirst & Espesser, 1993) ainsi qu'à l'élaboration d'un cadre théorique et méthodologique applicable à l'étude de la prosodie indépendamment de la variation situationnelle (Hirst & Di Cristo, à paraître). Ces deux aspects nous ont paru constituer les prérequis nécessaires pour entreprendre dans de bonnes conditions l'analyse de plus vastes corpus réunissant des exemples de styles de parole différents. Du même coup, nous estimons que l'application à ces corpus des méthodes d'investigations et des principes d'analyse que nous avons déjà testés dans d'autres circonstances devraient nous permettre de vérifier dans quelles limites les prédictions des modèles théoriques sont compatibles avec la description de la diversité des styles et d'établir, en conséquence, quels sont les développements dont ces modèles devraient bénéficier (notamment en ce qui concerne la nature et le nombre de paramètres à considérer) pour être en mesure de rendre compte de cette forme de variabilité. C'est également la poursuite de ces objectifs qui a motivé la mise en oeuvre du projet que nous présentons dans cette étude.

En choisissant de constituer une base de connaissance multistyle, nous nous inscrivons dans un courant de recherche qui, sans rompre de façon définitive avec l'analyse traditionnelle de la parole dite "de laboratoire", concentre ses efforts sur l'étude de la parole en usage dans les diverses formes d'interaction qui représentent les contextes dans lesquels la prosodie exerce ses fonctions majeures (Tench, 1990; Pierrehumbert & Hirschberg, 1992; Bruce & al., 1994; Swerts, 1994).

A ce propos, il serait erroné de prétendre que l'analyse de la parole spontanée est contemporaine des travaux de phonétique les plus récents. Nous nous bornerons à citer ici deux exemples particulièrement significatifs: l'ouvrage de Crystal (1969), qui demeure une référence pour les chercheurs qui s'intéressent à la prosodie de l'anglais et la thèse de Zwanenburg (1964), en ce qui concerne le français, ces deux études étant comme on le sait entièrement fondées sur l'analyse de corpus de parole spontanée.

D'autre part, il faut bien reconnaître que les sociolinguistes et les spécialistes de pragmatique ont été les premiers, avec quelques années d'avance sur les phonéticiens, à mettre l'accent sur la nécessité de décrire les faits de langue dans toute leur complexité et donc dans leur variabilité. Dans cette perspective nouvelle, qui se démarque de la quête traditionnelle de l'invariance phonétique, il est vite apparu que la prise en considération des diverses sources de variabilité s'avérait incontournable. Les variations sociolectale et dialectale, par exemple,

sont à l'évidence constitutives de la langue et elles ne doivent plus de ce fait être considérées comme des écarts relatifs à une norme (dans le sens qu'attribuaient naguère à ce terme les adeptes d'orthoépie). La diversité des situations de communication et les différentes formes d'interactions verbales constituent également des sources de variabilité qui ne peuvent être plus longtemps ignorées. Les retombées phoniques de cette pluralité déterminent ce qu'il a été plus ou moins convenu d'appeler le style (Eskenazi, 1993, Eskenazi & Lacheret-Dujour, 1991). A quelques rares exceptions près (Fónagy, 1976; Duez, 1978; Lucci, 1984; Guaïtella, 1991), nous ne disposons pas pour le français de connaissances étendues sur les particularités prosodiques des différents styles. Le projet PACOMUST ambitionne également de contribuer à la constitution de cette base de connaissances.

La base de données PACOMUST est enrichie progressivement à partir d'une centaine d'heures d'enregistrement de parole continue que nous avons déjà sélectionnés. Ces derniers comprennent notamment des textes lus, des conférences des récits, des extraits d'interviews, de pièces de théâtre et de dialogues de films, ainsi que des dialogues simulés. Certains de ces corpus sont déjà à l'étude dans plusieurs thèses et les premiers résultats feront l'objet de présentations lors du prochain Congrès International des Sciences Phonétiques de Stockholm. Compte tenu de l'importance que revêtent les choix méthodologiques dans cette entreprise, nous avons décidé de consacrer le présent article à cette problématique. Nous exposons donc dans la première partie les critères que nous avons retenus pour classer les échantillons de notre corpus et nous proposons à cet effet une typologie des interactions fondée sur les recherches les plus récentes dans ce domaine. Nous décrivons ensuite dans la deuxième partie les principes d'organisation de la base de données proprement dite. Enfin, nous présentons dans la dernière partie les méthodes d'étiquetage et de codage prosodiques que nous avons choisi d'appliquer dans cette recherche.

2. Critères de classification des corpus et typologie des interactions verbales.

2.1. Critères de classification.

Comme nous l'avons précisé antérieurement, la base de données d'échantillons de parole que nous cherchons à constituer ambitionne d'être multistyle. Ce terme, qui a parfois été utilisé avec des significations différentes, fait en réalité référence, de façon implicite ou explicite, à plusieurs formes de variabilité qui concernent en particulier le mode de production de la parole (ex. l'opposition parole lue/ parole spontanée), la nature de l'interaction (ex. interaction à structure d'échange ou sans structure d'échange) et la situation dans laquelle l'échantillon de parole a été enregistré (ex. corpus radiophonique, télévisé, enregistrement en studio ou à l'extérieur, etc.). Il est donc indispensable que toutes ces informations soient prises en compte et puissent figurer de manière

non ambigüe dans les entrées de la base de données. Ce point sera particulièrement développé dans la section suivante.

L'explicitation du concept de parole multistyle passe également par la signification que l'on attribue au terme d'interaction. A ce propos, nous adoptons le point de vue selon lequel tout message est adressé et tout type de discours constitue une forme d'interaction, y compris le monologue (Bakhtine, 1977). L'adhésion à ce principe nous a donc amené à nous interroger en premier lieu sur les critères à retenir pour l'élaboration d'une typologie des interactions.

2.2. Cadre typologique.

Le cadre typologique que nous allons présenter et tenter de justifier est adapté du modèle récent de Vion (1992), qui est issu du courant de recherche de la linguistique interactionnelle et qui représente l'aboutissement d'une reflexion prenant appui sur les théories et les modèles antérieurs proposés dans ce domaine. A l'instar des linguistes qui se sont intéressés à la problématique de l'interaction (Roulet, 1985; Kerbrat-Orecchioni, 1989, 1990), Vion propose un modèle de structure hiérarchique de l'interaction composé de six unités formelles. L'unité maximale est *l'interaction*, proprement dite, qui est définie par "l'ensemble de ce qui se produit entre deux ou plusieurs sujets, du début à la fin de leur rencontre". Au niveau immédiatement inférieur se situe le *module*, qui constitue une unité homogène sur le plan discursif, plusieurs modules pouvant être constitutifs d'une même interaction (cf. ci-après). Au-dessous du module se trouvent d'abord la *séquence*, qui correspond à une unité fonctionnelle ou thématique, puis *l'échange*, qui représente l'unité minimale de dialogue. Enfin, aux niveaux inférieurs de la hiérarchie se situent *l'intervention*, qui est définie comme la plus petite unité monologale, souvent assimilée au tour de parole, et *l'acte de langage*, qui apparaît comme l'unité fonctionnelle minimale (une question suivie d'une réponse étant considérée comme deux actes de langage consécutifs).

L'unité linguistique de base que nous retenons pour catégoriser les éléments constitutifs de la base de données, notamment pour la constitution des fiches signalétiques (cf. la section suivante) est le *module*. Un corpus est constitué soit d'un ou de plusieurs modules en fonction du caractère homogène ou hétérogène du discours. Ce dernier cas correspond, par exemple, au passage de la lecture à l'interview (ou vice versa) au sein d'un même échantillon de corpus.

2.3. Traits définitoires du module.

Nous définissons le module dans une double optique qui se rapporte, d'une part, aux conditions de production du discours et, d'autre part, aux divers types d'interaction. Pour ce qui relève des conditions de production, nous retenons les deux distinctions suivantes: *spontanée/non spontanée* et *parole publique/parole*

privée. Le terme "spontané" sert à désigner les productions orales dans lesquelles le sujet élabore tous les aspects de son message à l'instant où il le produit. Sous la qualification générale de "non spontané", nous réunissons donc plusieurs formes de production orales comme la lecture, la présentation des journaux radiodiffusés ou télévisés (dans ce cas, on parle de "lecture interprétée"), la récitation, etc. En ce qui concerne la seconde distinction, nous appelons "parole publique" celle qui est adressée à un auditoire et "parole privée" l'usage qui en est fait à des fins personnelles, comme par exemple dans une conversation qui se déroule dans un lieu privé.

Conformément à l'analyse de Vion (1992), nous retenons ici deux catégories majeures d'interactions: celles "*à structure d'échange*" et celles qui ne possèdent pas cette spécificité. Les interactions "à structure d'échange" se caractérisent par le fait que les participants ont la possibilité de devenir énonciateurs à tout moment (énonciateurs potentiels), ce qui n'est pas le cas pour les interactions "sans structure d'échange", comme les commentaires sportifs, les présentations de la météo, les bulletins d'information, les revues de presse, les cours magistraux, etc.

Afin d'établir une classification plus précise des interactions auxquelles nous nous intéressons, nous retiendrons les quatre critères proposés par Vion (1992). Selon nous, ces critères présentent une pertinence particulière en ce qui concerne les interactions à structure d'échange. Plus précisément, il semble qu'ils aient été constitués au départ à partir de l'observation d'une interaction particulière (la conversation) puis adaptés ensuite aux autres types d'interaction. Ce point permettrait de justifier le fait qu'ils ne sont peut-être pas toujours pleinement satisfaisants pour catégoriser les diverses interactions. Malgré tout, ils ont l'avantage d'objectiver notre classification linguistique.

Ces critères figurent dans la fiche signalétique (cf en annexe) et se présentent le plus souvent sous la forme d'oppositions.

- *Symétrie / complémentarité*. La symétrie caractérise une forme de comportement en miroir auquel adhèrent les interlocuteurs, comme c'est le plus souvent le cas dans la conversation usuelle. En revanche la complémentarité implique une différence entre les interactants, liée par exemple à leurs statuts respectifs.

- *Coopération / compétition*. La coopération, comme dans un échange amical, suppose que les locuteurs sont sans cesse en quête d'un consensus général, alors que la compétition se définit par sa nature conflictuelle, à l'exemple de la dispute ou du débat.

- *Nature de la finalité*. La finalité de l'interaction constitue également un critère qui contribue de façon significative à son identification. La conversation, qui est fondée sur la relation entre les interactants et qui est centrée principalement sur le contact qu'ils collaborent à maintenir, possède une finalité interne. Par contre l'enquête, l'entretien et l'interview, qui sont motivés par la recherche d'informations ou de connaissances, ont une finalité externe.

- *Caractère formel /informel de l'interaction.* Il est fonction d'un certain nombre de conventions plus ou moins contraignantes liées par exemple au "lieu" dans lequel se déroule l'interaction.

Le choix des critères que nous venons de décrire sommairement nous a paru s'imposer dans la mesure où la constitution de la base de données est l'oeuvre de plusieurs collaborateurs, et où nous souhaitons éviter, autant que faire se peut, un mode de classement qui ferait la part trop belle aux décisions subjectives. Précisons, avant de conclure sur ce point, qu'il n'est pas nécessaire d'utiliser systématiquement les quatre critères retenus pour définir une interaction, que les termes des oppositions présentées ne sont pas toujours mutuellement exclusifs. Seule l'expérience et le dépouillement du corpus nous permettront de juger en toute objectivité de la valeur opératoire du modèle typologique que nous avons adopté.

3. L'organisation du corpus en base de données.

Nous allons présenter dans les pages qui suivent les procédures utilisées afin d'organiser les divers corpus en base de données. Cette partie de l'article est subdivisée pour des raisons pratiques en quatre sous-sections qui concernent, respectivement, des observations à caractère général (3.1, 3.2) et des remarques qui s'appliquent à notre projet (3.3., 3.4.)

3.1. Une organisation standardisée pour une meilleure diffusion des données

Il existe dans les laboratoires de recherche sur la parole et le langage de nombreux corpus qui présentent une grande richesse de contenu. Malheureusement, l'utilisation de ces corpus ne dépasse pas le plus souvent le stade de l'exploitation locale pour les raisons suivantes : le manque d'information et de diffusion entre instituts de recherche, l'importance de la masse de données à transférer et l'absence de standardisation des données sonores et linguistiques (étiquetage et transcription). Le premier problème est sur le point d'être résolu, grâce notamment à la constitution de groupes de recherche coordonnée et de réseaux thématiques communiquant par messagerie électronique. Le second problème, qui concerne la masse de données à traiter, s'estompe au rythme des améliorations technologiques: apparition à coûts modérés de disques durs de plusieurs GigaOctets, de système de sauvegarde sur bandes ou disques magnéto-optiques et, récemment, de dispositifs facilement accessibles de gravure sur CD-ROM. Le problème de la standardisation demeure cependant préoccupant et constitue, de nos jours encore, un obstacle majeur.

A titre d'exemple, la base de données américaine TIMIT[2] apparaît comme le résultat très positif d'un effort d'uniformisation qui rend possible une utilisation partagée. En Europe, le projet SAM (Fourcin et al., 1989) a abouti à la définition d'une norme qui a permis la constitution d'EUROM[3], un corpus de parole multi-lingue, ainsi que la création, en France, d'une base de données des sons du français, BD-SON[4], et d'un corpus de parole lue, BREF[5]. Dans le cas de ces projets, la standardisation s'est arrêtée au niveau de l'organisation des fichiers sonores et de leur descriptif correspondant (procédures d'enregistrement, description des locuteurs, transcription orthographique). Cet effort est essentiel et constitue une première étape.

3.2. Les principes de l'organisation en base de données

Comme on le sait, une base de données est une collection de documents que l'on peut accumuler puis extraire de façon sélective. Ces finalités nécessitent donc une organisation structurée et hiérarchisée des éléments de la base et des informations qui y sont associées, les deux phases de génération de la base et de consultation, étant bien évidemment distinctes.

La constitution d'une base de données de parole passe par différentes étapes, qui concernent:
- le choix du matériau linguistique,
- l'enregistrement sonore,
- la mise au format des données et leur informatisation,
- la transcription et l'étiquetage.

Dans notre projet le matériau linguistique est constitué d'un ensemble d'échantillons de parole continue multi-style, qui ont été enregistrés dans des situations discursives diverses. Etant donné que les échantillons proviennent de différentes sources (radio, télévision, locuteurs en situation...), il nous est difficile d'obtenir de façon systématique les renseignements relatifs aux conditions d'enregistrement, aux caractéristiques socio-géographiques des locuteurs, etc. Toutefois, ce manque partiel d'information ne nous apparaît pas constituer un handicap majeur, dans la mesure où les spécialistes qui constituent la base de données peuvent estimer et caractériser la nature de la variation sociolectale, dialectale et idiolectale des locuteurs, ainsi que la nature des interactions dans lesquelles ils se trouvent impliqués.

---

[2]TIMIT (Zue et al, 1990) est le fruit d'une collaboration entre Texas Instruments (T.I) et le M.I.T. 630 locuteurs ont été sollicités pour constituer cette base de données qui comporte au total 2342 phrases. Chacun des énoncés est transcrit et étiqueté.

[3]EUROM est une base de données multi-lingues issue d'un projet ESPRIT-SAM n°1541

[4]"BD-SON est une action menée sous l'égide du GRECO-PRC communication Homme-Machine à la demande des chercheurs de disposer d'une large base de sons, utile tant pour l'étude de la langue française que pour les recherches en traitement automatique de la parole." (Cervantès et al., 1986). Les données représentent un volume de 3,5 Giga octets, stockés sur 7 CD-ROM.

[5]Le corpus BREF, disponible sur CD-ROM , est le résultat d'une action menée par le LIMSI-CNRS sous l'égide du GRECO-PRC "Communication Homme-Machine" et de la C.E.E. Il contient plus de 100 heures de parole lue (extraits du journal "le Monde") provenant de 120 locuteurs.

La phase de structuration et d'informatisation des données peut s'effectuer de façon semi-automatique (ex: EUROPEC[6]). Toutefois, la complexité et la diversité du matériau linguistique sélectionné nous ont conduit à opter pour une organisation originale dont il va être question plus loin. Les problèmes relatifs à la transcription et l'étiquetage seront également évoqués dans la section suivante.

L'interrogation de la base de données doit pouvoir s'effectuer de façon simple par l'application d'un jeu de requêtes et de filtres. Ainsi, il doit être possible, par exemple, de demander la liste des reportages radiophoniques, avec ou sans structure d'échange, en parole spontanée ou non spontanée, ou encore de trouver tous les cas de consonnes occlusives placées en position initiale de mot ou d'énoncé. Il est nécessaire pour cela d'utiliser un système de gestion de base de données (SGBD).

3.3. Une base de données de parole continue multistyle (PACOMUST)

De nombreux SGBD ont été réalisés, parfois complexes, comme ORACLE, INGRES, KMAN[7] ou plus simples, comme PARADOX, ACCESS[8]. Par ailleurs, il existe des langages spécialisés pour l'interrogation de base de données, comme SQL (Sequence Query Language), qui permettent d'effectuer les types de requêtes auxquelles il a été fait allusion ci-dessus. GERSONS[9], par exemple, est un SGBD adapté à la parole. Il est conçu pour des corpus du type BD-SONS, EUROM, PSH[10] , BD-BRUIT[11] qui sont composés d'énoncés isolés, de phrases porteuses, de listes de nombres ou de logatomes enregistrés par un groupe de locuteurs rigoureusement sélectionnés.

Dans le cas de la parole continue multistyle, cette organisation hiérachique simple n'existe pas. En effet, les extraits que nous recueillons n'ont pas été produits pour devenir un matériau d'étude sur la parole, comme c'est le cas pour la "parole de laboratoire". Ils ne rentrent donc pas dans un moule parfait. Il est donc indispensable d'introduire une organisation adéquate. Notre objectif n'est pas de construire un SGBD ex nihilo mais plutôt d'utiliser les fonctionnalités existantes aptes à manipuler les objets complexes que sont les échantillons et les énoncés de parole continue multistyle.

---

[6]EUROPEC, qui est un logiciel d'enregistrement pour base de données de parole, a été développé dans le cadre d'un projet ESPRIT-SAM (n°2589)

[7]ORACLE, INGRES, KMAN sont des systèmes de gestion de base de données qui ont été testés dans le cadre d'un projet ESPRIT-SAM (rapport SAM, 1989). Ils fonctionnent aussi bien sur PC que sur stations de travail.

[8]PARADOX, ACCESS sont des logiciels standards de gestion de base de données dont l'utilisation est répandue sur ordinateurs personnels.

[9]GERSONS est le logiciel de gestion de la base de données des sons du français. Il a été développé à l'ICP à Grenoble avec le soutien du GRECO-PRC "Communication Homme-Machine".

[10]PSH/DISPE est une base de données sur CD-ROM de parole subaquatique et hyperbare. Elle a été issue de la collaboration entre le laboratoire "Parole et Langage" CNRS URA 261, Aix en Provence et l'Institut National de la Plongée Profonde de Marseille.

[11]la base de données BD-BRUIT est le fruit d'une action initiée par le GRECO-PRC "Communication Homme-Machine" en 1991. Elle est destinée à permettre l'étude générale des perturbations de la production de la parole par le bruit environnant (effet Lombard).

3.3.1. L'organisation des données de PACOMUST

Les corpus sources de PACOMUST sont appelés *corpus thématiques* (Fig.1) et sont identifiés par un nom générique (ex : le corpus "Sagan", qui est relatif à une émission radiophonique sur Françoise Sagan). Ces données sonores sont enregistrées sur cassettes audiophoniques et restent à la disposition d'éventuels utilisateurs. De chaque corpus thématique nous avons extrait des *passages* (Fig.1) représentant une grande diversité de thèmes et de styles (ex: présentation de "F.Sagan": reportage, "jeux olympique": interview). Ces passages, une fois sélectionnés, sont stockés sur cassettes audionumériques et archivés en une sonothèque, ce qui permet une diffusion des données et rend possible des études pluridisciplinaires, avec la perspective de réalisation d'un CD-Rom. Du point de vue discursif, chaque passage est composé d'un ou plusieurs *modules* (ex: dans le passage reportage sur F. Sagan, se succèdent un module "interview" et un module "lecture"). Le module présente deux caractéristiques essentielles: il possède à la fois une continuité temporelle et une homogénéité discursive. Il est, pour ces raisons, défini par un ensemble de descripteurs uniques. A un passage peut ne correspondre qu'un seul module (Fig.1, passage A1) ou plusieurs modules (Fig.1, passage A2 ou B1). Un *extrait* correspond à une sélection à l'intérieur d'un module. Il ne peut pas exister de frontières de modules au sein d'un extrait unique. En effet, une telle disposition empêcherait une catégorisation homogène du type de discours.

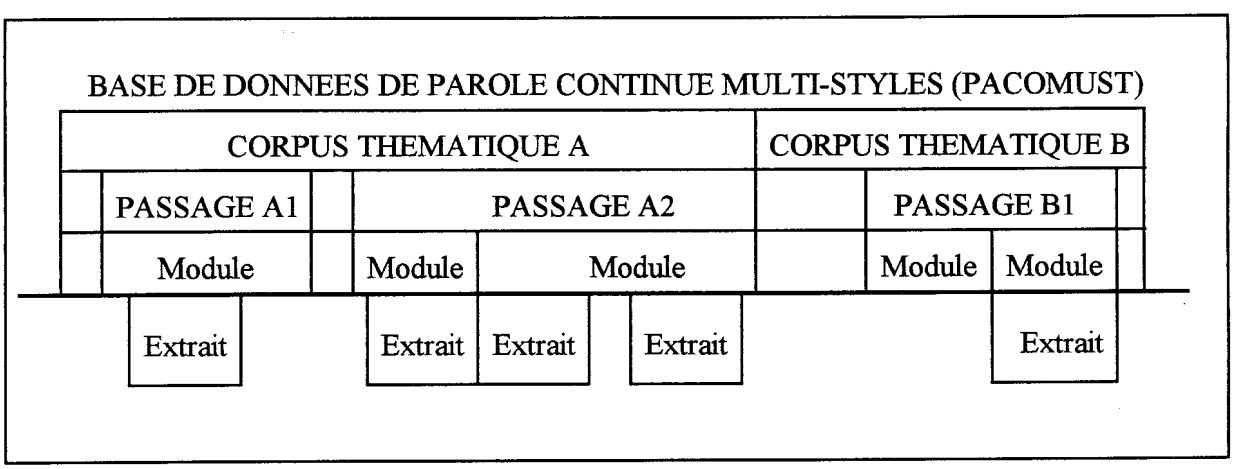

Figure 1 : structuration des données de la base du projet PACOMUST.

Dans une optique ascendante, un corpus thématique est finalement constitué d'une série d'extraits, objets observables pour nos études. Le regroupement des extraits en une interaction globale, ne s'effectue, si cela est nécessaire, qu'aux niveaux supérieurs, dans le cadre d'une étude comparative, par exemple.

3.3.2. Les descripteurs

Chaque extrait possède un ensemble de descripteurs qui contiennent:
- le lieu, la date, le matériel d'enregistrement, l'appréciation de la qualité sonore
- la transcription orthographique
- la spécification du type d'interaction, par référence à la typologie discursive utilisée (cf. ci-dessus, section 2)
- des informations relatives aux locuteurs
- des indications techniques relatives à son traitement informatique.

La structuration des données est fondée sur une hiérarchie stricte, ce qui signifie que les descripteurs d'un niveau supérieur restent valables pour les niveaux inférieurs (cf. la présentation de la fiche signalétique, § 3.4.). Il reste possible, via un utilitaire de gestion, d'extraire à volonté les modules possédant un descriptif précis. On interroge alors la base de données par un ensemble de requêtes.

3.3.3. Format des fichiers

D'un point de vue informatique, les extraits sont composés des fichiers suivants :
- un fichier de signal audio numérisé;
- un fichier contenant les descripteurs de l'extrait;
- un fichier contenant la transcription pseudo-orthographique de l'extrait
- des fichiers d'étiquetage (phonémique, syllabique, prosodique et pragmatique).

Dans une perspective de standardisation (rapport SAM, 1989), nous avons choisi de formater ces fichiers aux normes européennes (ESPRIT N° 2589-SAM), c'est à dire que les échantillons de signal numérisé sont mis dans un format binaire ne comportant ni entête, ni listing. De plus, le fichier de description est constitué de lignes ASCII commençant par un mot clé et comportant un certain nombre de champs. Les renseignements contenus dans ce type de fichier (cf. figure 2) peuvent être à la fois organisationnels (nom du corpus thématique, du passage, de l'extrait, etc.), linguistiques (type de discours, locuteurs, etc.) ou techniques (date, source d'enregistrement, fréquence d'échantillonnage, dynamique, etc.).

L'étiquetage s'effectue à plusieurs niveaux, chacun comportant un fichier d'étiquettes, toujours au format ASCII, où figureraient, non seulement l'étiquetage proprement dit, mais aussi des informations sur la personne qui a effectué la segmentation, le niveau de transcription, le type de segmentation, la date, etc. Cette organisation n'est pas actuellement en pratique mais reste à l'état de projet. Elle faciliterait toutefois l'échange de données car conforme aux normes européennes (ESPRIT N° 2589-SAM).

TYP: description	⇒ file type
DBN : PACOMUST_1	⇒ data base name
THC: sagan	⇒ thematic corpus
PAS: interview	⇒ passage name
EXT: a_propos_dernier_livre	⇒ extract name
INT : entretien/spontané/public/avec structure d'échange	⇒ interaction type
NSP : 2	⇒ number of speakers
SPK: 1/M. Domio/M/adulte/Fr/journaliste	⇒ speaker
SPK: 2/F. Sagan/F/adulte/Fr/écrivain	⇒ speaker
SRC: FRRA07B1.PFS	⇒ speech signal
RED: 06/12/94	⇒ recording date
REP: Aix - IPA	⇒ recording place
REO: L.S	⇒ recording operator
SAM: 16000	⇒ sampling

Figure 2 : exemple de fichier de description
(la typologie est en anglais pour être conforme aux normes européennes)

3.4. La fiche signalétique

La fiche signalétique synthétise l'ensemble des descripteurs d'un extrait. Elle permet la structuration en base de données. La recherche d'un extrait dans la base peut se faire, soit de façon descendante (choix d'un corpus thématique, d'un passage, d'un extrait puis consultation des descripteurs), soit par sélection préliminaire de descripteurs (ex: discours en français, du type entretien spontané public, avec structure d'échange, contenant au moins deux locuteurs ...), puis consultation des propositions. Un exemple de fiche signalétique est fournie en annexe.

4. Segmentation, étiquetage et codage du corpus

L'étiquetage d'un corpus pose un certain nombre de questions préliminaires qui concernent, notamment, la définition des niveaux, les conventions de transcription et les critères de segmentation adoptés. Nous n'avons pas la prétention de débattre ici ces diverses questions. Nous nous bornerons en conséquence à aborder les points qui paraissent essentiels à notre propos. Nous avons retenu deux niveaux d'étiquetage: phonémique et prosodique.

4.1. La segmentation et l'étiquetage phonémique

En ce qui concerne l'étiquetage phonémique, nous distinguons deux types de problèmes, liés, d'une part, aux frontières des unités segmentales et, d'autre part, à leur notation.

4.1.1. Les problèmes de frontières segmentales

Il serait fastidieux d'évoquer à nouveau ici toutes les difficultés théoriques et pratiques auxquelles l'on est inévitablement confronté dans la réalisation d'une tâche de segmentation de la parole. En face de ces difficultés, la solution consiste parfois à ne pas prendre de décision sur l'assignation des frontières d'un segment et à se limiter à en signaler le centre. Nous n'avons pas retenu cette solution et nous avons choisi d'indiquer explicitement les frontières des unités segmentales.

Par ailleurs, il existe plusieurs systèmes de segmentation automatique ou semi-automatique de la parole (Faraht et al., 1992; Dalsgaard, 1991). Par rapport à une segmentation manuelle, ils possèdent l'avantage d'un découpage déterministe, non dépendant de la subjectivité d'un opérateur humain et surtout ils autorisent un gain de temps considérable. Ces systèmes, toutefois, manquent encore de robustesse et ils demeurent le plus souvent à l'état de prototypes. Nous avons donc opté pour un *étiquetage manuel*. Nous espérons par là-même étendre la base de connaissances sur les critères de segmentation et contribuer ainsi à l'amélioration des systèmes automatiques.

Dans le cadre d'une segmentation manuelle, nous sommes pleinement conscients du fait que les frontières- ou les discontinuités - susceptibles d'indiquer les limites d'une unité phonique sont souvent issues d'un choix arbitraire effectué sur la base des indices jugés pertinents par un expérimentateur (Meunier, 1994). En fait la détermination d'un segment est le fruit d'une décision et ne s'impose pas comme une vérité incontestable. Afin d'évaluer la part de subjectivité qui revient à l'expérimentateur humain et la variabilité qui en découle, nous avons réalisé au sein du Groupe une série d'expériences qui consistaient à comparer les résultats d'opérateurs humains dans une tâche de segmentation de corpus de parole continue spontanée. Ces résultats, qui feront l'objet d'une publication ultérieure, nous ont permis de constituer une typologie des difficultés d'ajustement des frontières et d'établir ainsi un ensemble de critères susceptibles de renforcer la cohérence des choix de segmentation au sein du groupe de travail.

4.1.2. Les problèmes de transcription phonémique

Outre la difficulté de l'ajustement des frontières, se pose également pour l'expérimentateur celui du choix de la transcription de l'élément segmental, qui introduit une nouvelle source de variabilité néfaste. Les cas les plus fréquents sont, par exemple, la confusion entre [e] et [ɛ], [ø] et [œ], [o] et [ɔ]. On relève par ailleurs des différences fréquentes entre transcripteurs dans le cas de segments "ambigus" comme un [b] partiellement dévoisé qui est transcrit soit [b], soit [p]. Le cas particulièrement épineux du "r" est aussi à considérer: doit-on effectuer une transcription phonologique /R/ ou phonotypique /ɾ/, /χ/ ou /ʁ/ ? Une liste de ces difficultés est recensée dans (Autesserre et al., 1989).

Afin de réduire les disparités perceptives entre opérateurs humains et de faciliter la tâche de segmentation, nous avons opté, tout d'abord, pour un niveau de transcription archiphonémique dans lequel /E/ est noté pour /e/ ou /ɛ/, /A/ pour /a/ ou /ɑ/, /&/ pour /ø/ ou /œ/, /O/ pour /o/ ou /ɔ/, /U~/ pour /ɛ~ / ou /œ~/. Le "r" est toujours transcrit par /R/. Cet étiquetage apparaît finalement comme la base de départ d'un étiquetage susceptible d'être perfectionné en fonction des besoins particuliers du chercheur. La possibilité reste ouverte au transcripteur de proposer un étiquetage phonotypique dans lequel il a loisir d'apporter une précision sur la nature exacte du segment et de contribuer ainsi à son identification.

Pour permettre une uniformisation de transcription, nous avons choisi d'utiliser l'alphabet phonétique international pour l'informatique: SAMPA (Fourcin et al., 1989).

4.2. Proposition d'un cadre pour l'étiquetage prosodique

Bien que des efforts d'harmonisation aient été accomplis dans cette voie, il n'existe à ce jour aucun standard pour l'étiquetage de la prosodie, comme il en existe un pour le segmental, sous la forme de l'Alphabet Phonétique International. Certes, à la suite de la Convention de Kiel, la dernière version de l'API propose d'inclure des symboles destinés à la notation des faits prosodiques (IPA, 1989). Cependant, ces nouvelles propositions n'ont pas reçu un accord unanime et n'ont pas fait l'objet d'une évaluation systématique (Bruce, 1989). Il n'en demeure pas moins que la nécessité de disposer d'un système de notation de la prosodie se fait de plus en plus pressante, à la fois pour les linguistes et les spécialistes du traitement automatique de la parole.

Dans le domaine des applications technologiques, la notation et l'étiquetage prosodiques sont susceptibles de fournir un grand nombre de données aisément utilisables, aussi bien en reconnaissance automatique qu'en synthèse de la parole. Ils peuvent également faciliter la diffusion des recherches plus fondamentales en prosodie et rendre plus accessible la comparaison de données issues de bases diverses. Enfin, ils représentent une étape incontournable en vue de la mise au point d'un système de transcription automatique. En ce qui nous concerne, l'étiquetage prosodique de la base de données constitue à l'évidence une étape cruciale pour établir la nature des relations entre les structures prosodiques et l'organisation discursive des échantillons de la base de données.

4.2.1. Problèmatique de l'étiquetage prosodique

L'étiquetage prosodique soulève de multiples problèmes qui concernent plus particulièrement:

- le choix des évènements à étiqueter,

- le support servant de référence à l'étiquetage (acoustique ou impression auditive),
- l'inventaire des symboles de notation proprements dits.

Parmi ceux qui se sont intéressés à cette problématique, il existe un consensus sur le premier point, dans la mesure où l'on s'accorde sur la nécessité de noter des faits majeurs comme les niveaux de frontières ou les degrés de rupture, les proéminences et certains évènement tonals, comme les accents mélodiques (pitch accents), les tons de frontière et les changements de direction de la courbe mélodique (Silverman et al., 1992; Bruce et al., 1994). Malgré ce consensus, l'étiquetage d'une base de données multistyle pose des problèmes particuliers relatifs à la notation de certains phénomènes, comme la toux, les raclements de gorge, les coups de glotte, le souffle, qui relèvent de l'étude des faits suprasegmentaux, mais qui se situent hors du champ de l'analyse prosodique proprement dite (Crystal, 1969). Le cadre que nous allons proposer demeure ouvert et laisse la possibilité d'introduire ultérieurement de nouvelles étiquettes en fonction des besoins qui peuvent se présenter.

Enfin, comme nous estimons qu'il est difficilement concevable d'envisager l'élaboration d'un système d'étiquetage prosodique en dehors de tout cadre théorique, c'est sur la base de considérations relatives à ce cadre que nous allons tenter de justifier notre approche.

Parmi les travaux récents (cf. Llisteri, 1994 pour une synthèse), le système d'étiquetage de la prosodie qui bénéficie actuellement de la plus grande notoriété est sans nul doute le système ToBI (Tones and Break Indices), qui a été conçu pour transcrire la prosodie de l'anglais américain (Silverman et al., 1992; Beckman et Ayers, 1994) mais qui ne semble pas, pour diverses raisons, être applicable en l'état à d'autres langues.

L'une des raisons, qui a suscité récemment diverses suggestions et critiques (Wightman & Campbell, 1994), est que l'utilisation de ToBI suppose que l'inventaire des schémas intonatifs de la langue concernée soit déjà connu, ce qui ne paraît pas être le cas pour la plupart des langues. Cette restriction étant précisée, il n'en demeure pas moins que le système ToBI a déjà été soumis à plusieurs évaluations (Pitrelli & al., 1994) et il constitue aujourd'hui une référence pour l'anglais américain.

Le système ToBI utilise deux types d'étiquettes pour indiquer, respectivement:

(i) le degré de rupture prosodique entre deux mots consécutifs, selon une échelle graduée de 0 à 4;

(ii) l'organisation tonale des énoncés, au moyen d'une série de symboles (H, L, *, %) qui permettent de représenter les patrons mélodiques qui sont associés aux syllabes accentuées et aux frontières des unités intonatives en anglais américain.

Une des particularités de ToBI réside dans le fait que la transcription des corpus s'appuie à la fois sur l'observation des données physiques (courbes de Fo) et sur l'analyse auditive des corpus. Cependant, tous les systèmes de codage n'adoptent pas la même démarche. Par exemple, l'approche du groupe de recherche suédois sur le dialogue (Bruce et al, 1994) fonde entièrement la transcription sur l'analyse auditive. Pour les auteurs, ce choix est justifié par le fait que l'objectif visé est une transcription phonologique des données d'observation plutôt qu'une transcription phonétique.

4.2.2. Présentation du cadre théorique

La méthode d'étiquetage et de transcription que nous nous proposons de mettre en oeuvre dans le cadre du projet PACOMUST se démarque des précédentes à plus d'un titre et s'inspire d'un modèle d'analyse prosodique développé à l'Institut de Phonétique d' Aix (Hirst, 1994, Di Cristo & Hirst, 1986; Hirst & Di Cristo, à paraître). Pour la compréhension de ce qui va suivre, il nous a paru utile de rappeler les grandes lignes de ce modèle avant d'en exposer les applications à la méthode de transcription et de codage des faits prosodiques.

Il paraît légitime d'affirmer qu'une théorie unifiée de la prosodie se doit d'intégrer tous les niveaux de représentation, depuis le niveau le plus concret de la représentation analytique des faits physiques (acoustique et physiologie), jusqu'au niveau le plus abstrait des représentations phonologiques de la langue (Figure 3). Il est généralement admis que ces niveaux extrêmes ne peuvent pas être associés de façon directe et que leur mise en relation doit s'effectuer, soit par la médiation de processus interprétatifs (Pierrehumbert & Beckman, 1988; Beckman, 1991), soit, comme c'est le cas dans notre approche (Hirst, 1994; Hirst & Di Cristo, à paraître), par celle de niveaux de représentation intermédiaires. Selon nous, ces niveaux doivent être doublement interprétables: au niveau immédiatement inférieur et au niveau immédiatement supérieur, afin de satisfaire à ce qu'il est convenu d'appeler une "condition d'interprétabilité". Entre le niveau le plus concret des faits physiques et le niveau le plus abstrait des représentations phonologiques (ou niveau phonologique profond), notre modèle intègre donc deux niveaux intermédiaires identifiés, respectivement, comme le niveau de la "représentation phonétique" et le niveau de la "représentation phonologique de surface" (Figure 3).

Si le niveau physique correspond à celui des représentations analytiques classiques sous la forme de courbes brutes d'intensité et de fréquence fondamentale, le niveau de la représentation phonétique est illustré par des configurations prosodiques "similaires" dans lesquelles ont été toutefois effacées les variations micromélodiques qui reflètent comme on le sait des contraintes de type universel (Di Cristo, 1978; Di Cristo & Hirst, 1986). Dans cette perspective, la représentation phonétique d'une courbe de Fo est celle d'une courbe lisse et continue, qui est modélisable par une séquence de points-cibles que relie une

fonction d'interpolation du type spline quadratique. Cette modélisation est réalisée automatiquement au moyen de l'algorithme MOMEL (Hirst & Espesser, 1993).

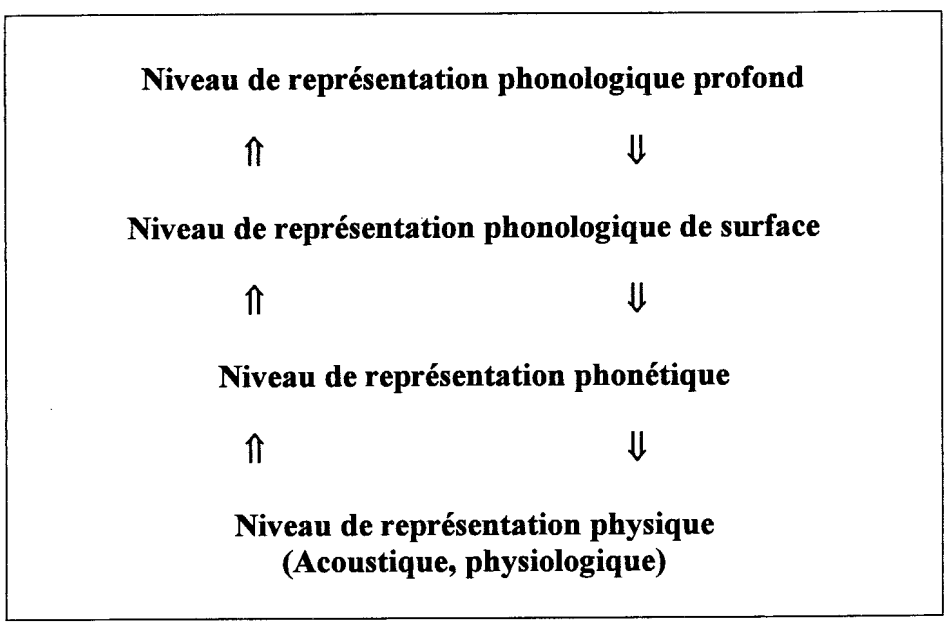

Figure 3. Niveaux de description et niveaux de représentation de l'analyse prosodique.

La distinction entre le niveau de la représentation phonétique et celui de la représentation phonologique de surface est fondée sur l'opposition traditionnelle entre fait continu et fait discret. Le passage du premier au second niveau s'effectue à l'aide du système INTSINT (INternational Transcription System for INTonation), qui a été développé comme une première approximation d'un système de transcription de l'intonation (Hirst & Di Cristo,à paraître). Ce système permet précisément d'effectuer le codage des points cibles obtenus automatiquement au moyen de l'algorithme MOMEL, grâce à un inventaire très réduit de symboles catégoriels et cela, quelle que soit la langue considérée, un peu à la manière dont procède l'Alphabet Phonétique International pour noter les caractéristiques des voyelles et des consonnes des divers idiomes. A la différence de ToBI, l'utilisation d' INTSINT ne présuppose pas que les schémas mélodiques de la langue concernée soient connus avant d'entreprendre la transcription de ses caractéristiques prosodiques. C'est la raison pour laquelle INTSINT nous paraît mieux adapté pour la description prosodique des échantillons de la base de données que nous sommes en train de constituer.

L'une des idées qui a servi à constituer le cadre de référence théorique d'INTSINT est que les valeurs des cibles de Fo sont programmées par les locuteurs selon deux modalités différentes (Hirst & Di Cristo, forthcoming,a): soit en termes de tons absolus {**T**op, **M**id, **B**ottom}, qui font référence à la tessiture du sujet dans le domaine de l'Unité Intonative, soit en termes de tons relatifs {**H**igher, **S**ame, **L**ower, **U**pstepped, **D**ownstepped} dont la valeur est déterminée

en fonction de la cible précédente. Une seconde distinction est établie entre tons non-itératifs { **H**, **S**, **L**} et tons itératifs { **U**, **D**}, ce qui permet de rendre compte de certains faits intéressants, notamment des phénomènes dits d'abaissement. Enfin, précisons qu'il est également possible d'utiliser à la place des symboles alphabétiques un jeu de symboles graphiques dont on indique ci-après les correspondances:

{ T [⇑], M [⇒], B [⇓], H [↑], S [→], L [↓], U [⌋], D [⌉] }.[12]

Il a été souligné, dans l'introduction du présent article, que le but de l'analyse prosodique est d'extraire l'information linguistique à partir de l'observation et de l'interprétation des données. Le codage symbolique des points-cibles constitue une étape capitale dans cette voie, car il permet d'étayer les inférences que l'on est en mesure d'effectuer sur la représentation des structures prosodiques de la langue au niveau le plus abstrait.

4.2.3 Procédures d'étiquetage et de transcription prosodiques

La méthode d'étiquetage et de transcription prosodiques que nous allons exposer succinctement dans les lignes qui suivent est à la fois plurilinéaire et pluriparamétrique. La démarche procède de la définition préliminaire de deux modules autonomes et homogènes quant à leur contenu : le niveau *subjectif interprétatif* et le niveau *acoustique* (Figure 4).

=>Le niveau subjectif interprétatif :

Le niveau subjectif interprétatif est considéré comme le niveau d'étiquetage des évènements auditifs et fonctionnels. A ce niveau, la tâche du transcripteur n'est pas assimilable à un filtrage de type psycho-acoustique, mais à une interprétation proprement linguistique. On notera que le codage des évènements tonals ne relève pas de ce module. En effet, comme l'inventaire exhaustif des schèmes tonals du français n'est pas suposé connu, l'intégration du codage des évènements mélodiques dans ce module reviendrait à introduire une tâche non-linguistique parmi un ensemble de tâches spécifiquement linguistiques, ce qui dérogerait au principe d'homogénéité que nous avons adopté.

L'étiquetage prosodique proprement dit suppose que l'étiquetage segmental ait été préalablement effectué. Dans l'attente de la mise en oeuvre d'une procédure de segmentation automatique fiable, ce dernier est réalisé manuellement. Nous avons retenu d'indiquer l'emplacement des pauses lors de l'étiquetage segmental.

Toutes les étiquettes prosodiques du niveau subjectif interprétatif sont également apposées manuellement. Elles concernent les évènements suivants :

---

[12] Tous ces symboles graphiques sont disponibles dans la police "symbol".

- les niveaux de rupture ou de frontière perçus, selon une echelle graduée.
- les proéminences, avec indication de leur nature (emphatique, non emphatique, etc)
- les hésitations et les faux départs.

Bien qu'ils ne relèvent pas strictement du niveau fonctionnel, nous indiquons également :

- les changements de tempo
- les variations de registre.

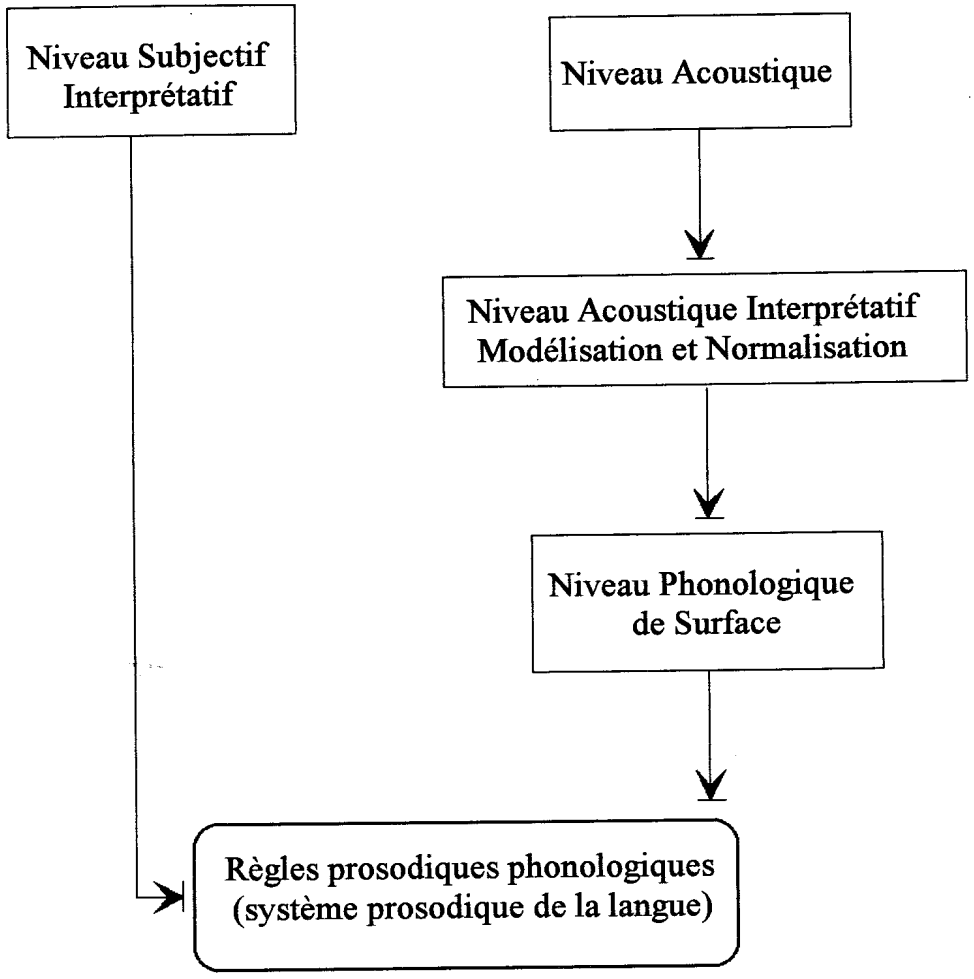

Figure 4 : Organigramme de la procédure d'étiquetage et de transcription prosodiques de la base de données

=>Le niveau acoustique :

Le second module, qui représente le niveau acoustique, concerne les faits de substance. Il constitue l'étape préliminaire d'alignement des symboles phonétiques ou orthographiques avec les différentes représentations analytiques de la Fo, de l'intensité et de l'évolution temporelle du spectre (spectrogramme).

La conversion du niveau acoustique brut en niveau acoustique interprétatif

comprend la mise à l'oeuvre de deux procédures qui correspondent, respectivement, à la modélisation et à la normalisation des données. La modélisation de la Fo est réalisée automatiquement par l'algorithme MOMEL qui convertit la courbe originelle en une courbe lisse et continue (représentation phonétique) sur laquelle figurent les localisations temporelles des points-cibles. En ce qui concerne la normalisation, nous utilisons l'échelle ERB pour la fréquence fondamentale (Hermes & Van Gestel, 1991) et une méthode inspirée des travaux de Campbell (1992), Barbosa & Bailly (1994), et de Bagshaw (1994), pour la normalisation temporelle. Une méthode de normalisation de l'intensité, fondée sur des travaux en cours au sein du laboratoire, est actuellement à l'étude.

Comme nous l'avons indiqué précédemment, le passage du niveau phonétique au niveau phonologique de surface constitue en fait une étape qui consiste à extraire du continuum sonore des évènements discrets. En ce qui concerne Fo, ce codage est, à l'heure actuelle, réalisé manuellement, à l'aide du système INTSINT. Toutefois, une procédure automatique est en cours de réalisation . De même, nous travaillons à l'élaboration d'un système de codage des variables temporelles et de l'intensité, car il n'existe pas à notre connaissance un équivalent d'INTSINT relatif à ces paramètres.

Ainsi que le montre l'organigramme de la figure 4, le dernier module est issu de la mise en correspondance des informations fonctionnelles du niveau subjectif interprétatif et des informations relatives aux formes prosodiques obtenues à la sortie du module identifié comme le niveau phonologique de surface. Cette mise en correspondance permet, par exemple, d'établir une relation entre la présence d'une frontière identifiée par le transcripteur et l'inventaire des formes prosodiques qui concourent à cette identification. Par là même, elle représente une étape fondamentale en vue de l'édification du système des règles phonologiques qui constituent le système prosodique de la langue.

Conclusion

Nous nous sommes limités dans cet article à présenter les grandes lignes d'un projet de recherche collectif qui apparaîtra particulièrement ambitieux si l'on considère l'étendue de la base de données que l'on se propose de constituer et la complexité des méthodes d'analyse que l'on envisage d'utiliser pour procéder à son étude. Toutefois, compte tenu, d'une part, de l'avancement des recherches effectuées à l'Institut de Phonétique d'Aix, notamment dans les domaines de l'analyse théorique des éléments prosodique et du développement des outils informatiques adaptés à leur modélisation et, d'autre part, du nombre de chercheurs qui se sont investis dans son élaboration, il est loin de nous sembler irréalisable.

Dans la présente étude, nous avons choisi de mettre l'accent sur les problèmes théoriques et la discussion des choix méthodologiques. C'est la raison pour

laquelle les deux premières sections de cet article sont consacrées, respectivement, à la présentation d'une typologie des interactions et aux modalités d'organisation d'un corpus en base de données. La dernière partie concerne plus particulièrement le domaine de la prosodie. Nous nous sommes particulièrement attachés à y montrer comment la méthode d'étiquetage et de transcription des éléments prosodiques que nous préconisons s'intègre dans le cadre des approches théoriques qui ont fait l'objet, au sein de notre Groupe, de plusieurs développements durant ces dernières années. Cette méthode, qui pose encore de nombreux problèmes, ne doit pas, selon nous, être considérée comme un ensemble figé, mais comme un système évolutif au sujet duquel nous sommes disposés à accueillir d'éventuelles suggestions. De même, nous envisageons dès à présent d'étendre la base de données à diverses variétés de français ainsi qu'à plusieurs langues. Dans cette perspective, nous demeurons ouverts à toutes les formes de collaboration. Les données actuelles de la base font déjà l'objet de plusieurs études dans le cadre d'une dizaine de thèses. Les thèmes de ces recherches concernent principalement: l'organisation prosodique du texte lu et de la narration, l'accentuation et le rythme dans divers phonostyles, les stratégies de segmentation intonative dans la parole spontanée et le rôle de la prosodie dans le discours interactif (dialogue et conversation). Nous attendons également que ces recherches contribuent à évaluer et à perfectionner la méthodologie dont nous avons exposé les fondements dans cet article.

Remerciements

Bibliographie

Autesserre D., Perennou G. and Rossi, M. (1989). Methodology for the transcription and labelling of a speech corpus, *Journal of the International Phonetic Association*, 19, 1-15.

Barbosa P. and Bailly G. (1994). Characterisation of rhythmic patterns for text-to-speech synthesis, *Speech Communication*, 15, 127-137.

Bagshaw, P. C. (1994). *AutomaticProsodic Analysis for Computer Aided Pronunciation Teaching*, Ph.D. thesis, University of Edinburgh. U.K.

Beckman, M. (1991). Metrical structure versus autosegmental content in phonetic interpretation. *Proceedings of the XIIth International Congress of Phonetic Sciences* (Aix-en-Provence), vol.1, 374-378.

Beckman, M.E. and Ayers, F. (1994). *Guidelines for ToBI labelling*, Ohio State University.

Beckman, M.E. & Hirschberg, J. (1994). The ToBI annotation conventions, *Document*, Ohio State University.

Brown, G., Currie, K.L. and Kenworthy, J. (1980). *Questions of Intonation.*

Croom Helm: London.

Bruce, G., (1989). Report from the IPA working group on suprasegmental categories, Lund University Department of Linguistics and Phonetics, *Working papers*, 35, 25-30.

Bruce, G., Granström, B., Gustafson, K., House, D. and Touati, P. (1994). Modelling Swedish prosody in a dialogue framework. *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP)*, Yokohama, S20-6.1, 1099-1102.

Campbell, N. (1992). *Multi-Level Timing in Speech*. Ph.D. Thesis. University of Sussex. U.K.

Campbell, N. (1994). Combining the use of duration and Fo in an automatic analysis of dialogue prosody. *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP)*, Yokohama, S20-9.1, 1111-1114

Cervantès, O., Serignat, J.F., Descout, R. et Carré, R. (1986). Définition et réalisation d'une base de données des sons du français., *15èmes Journées d'Etudes sur la Parole*, Aix-en-Provence, GALF, 213-216.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge University Press.

Dalsgaard, P., Andersen, O. and Barry, W. (1991). Multi-lingual label alignment using acoustic-phonetic features derived by Neural Network Techniques *Proceedings of ICASSP*, Toronto.

Di Cristo, A. (1978). *De la Microprosodie à l'Intonosyntaxe*. Thèse de Doctorat d'Etat. Université de Provence.

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

Di Cristo, A. and Hirst, D.J. (1993). Rythme syllabique, rythme mélodique et représentation hiérarchique de l'intonation du français. *Travaux de l'Institut de Phonétique d'Aix*, 15, 9-24.

Di Cristo, A. and Hirst, D.J. (à paraître). L'accentuation non-emphatique en français: stratégies et paramètres. *Hommages I. Fónagy*.

Duez, D. (1978). *Essai sur la Prosodie du Discours Politique*. Thèse de Doctorat. Université de Paris III.

Eskenazi, M. (1993). Trends in speaking styles research. *Eurospeech 93*, Berlin, 501-509.

Eskenazi, M. and Isard, A. (1991). Characterizing the change from casual to careful style in spontaneous speech. *Journal of the Acoustical Society of America*, 90 (4), 2363-2364.

Eskenazi, M. and Lacheret-Dujour, A. (1991). Exploration of individual strategies in continuous speech. *Speech Communication*, 10, 249-264.

Fahrat, A., Perennou, G. et Vigouroux, N. (1992) Segmentation en événements phonétiques et modèles markoviens HMM pour l'étiquetage phonotypique. *19èmes Journées d'Etudes sur la Parole*, Bruxelles.

Fónagy, I. et Fónagy, J. (1976). Prosodie professionnelle et changements prosodiques. *Le Français Moderne*, 44, 193-228.

Fourcin, A.J., Harland, G., Barry, W. and Hazan, V. (1989) *Speech Input and Output Assessment. Multilingual Methods and Standards*. Ellis Howood.

GEDO (à paraître). L*es Données Orales*. Numéro spécial de "Recherches sur le Français Parlé". Groupe d'étude sur les données orales (Aix-en-Provence).

Guaïtella, I. (1991). *Rythme et Parole: Comparaison Critique du Rythme de la Lecture Oralisée et de la Parole Spontanée*. Thèse de Doctorat. Université de Provence.

Hermes D. J. and Van Gestel J. C. (1991). The frequency scale of speech intonation, *Journal of the Acoustical Society of America.* 90 (1), 97-102.

Hirst, D.J. (1994). The symbolic coding of fundamental frequency curves: from acoustics to phonology. *International Symposium on Prosody* (sept. 1994). Yokohama, Japan.

Hirst, D.J. and Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, 15, 71-85.

Hirst D.J., Ide, N. & Veronis J. (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.

Hirst, D.J. and Di Cristo, A. (1984). French intonation: a parametric approach. *Die Neueren Sprachen*, 83 (5), 554-569.

Hirst, D.J. et Di Cristo, A. (1986). Unités tonales et unités rythmiques dans la représentation de l'intonation. *15 èmes Journées d'Etudes sur la Parole*, Aix-en-Provence, GALF, 93-95.

Hirst, D.J. and Di Cristo, A. (forthcoming a). *Intonation Systems: a Survey of Twenty Languages*. Cambridge University Press.

Hirst, D.J. and Di Cristo, A. (forthcoming b). Levels of description and levels of representation in the analysis of intonation. *Travaux de l'Institut de Phonétique d'Aix*.

IPA (1989). Report on the 1989 Kiel Convention. *Journal of the International Phonetic Association*, 19 (2), 67-80.

Kerbrat-Orecchioni, C. (1989). L'approche interactionnelle en linguistique, BUSCILA: *L'interaction*, 7-25.

Kerbrat-Orecchioni, C. (1990). *Les Interactions Verbales*, Tome 1, A. Colin: Paris.

Llisterri, J. (1994). *Prosodic Encoding Survey*, LRE Project 62-050 Multext, Deliverable 1.5.3.

Lucci, V. (1983). *Etude Phonétique du Français Contemporain à Travers la Variation Situationnelle*. Publications de l'Université de Grenoble.

Meunier, C. (1994). *Les Groupes de Consonnes: Problématique de la Segmentation et Variabilité Acoustique*. Thèse de Doctorat, Université de Provence.

Pierrehumbert, J. and Beckman, M. (1988). *Japanese Tone Structure*. Linguistic Inquiry Monograph 15. MIT Press.

Pierrehumbert, J. and Hirschberg, J. (1992). The meaning of intonational contours in the interpretation of discourse, in: Cohen, P.R., Morgan, J. & Pollack, M. (Eds). *Intentions in Communication*, MIT Press: 271-311.

Pitrelli, J.F.; Beckman, M.E. and Hirschberg, J.; (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japon, S05-5.1, 123-126.

Rossi, M. (1993). A model for predicting the prosody of spontaneous speech (PPSS model). *Speech Communication*, 13, 87-107.

Roulet, E. (1985). *l'Articulation du Discours en Français Contemporain*. Peter Lang: Berne.

SAM (1989), *Multi-lingual Speech Input/Output: Assessment, Methodology and Standardisation*. Extension phase, Final Report

Silverman, K.; Beckman, M.; Pitrelli, J.; Ostendorf, C.; Wightman, C.; Price,P.; Pierrehumbert, J. and Hirschberg, J. (1992). *ToBI : a standard for labeling English prosody. Proceedings of the 1992 International Conference on spoken language processing (ICSLP)*, Banff, Alberta, Canada, October 13-16, Vol.2, 867-870.

Swerts, M. (1994). *Prosodic Features of Discourse Units*. Doctoral Thesis. Eindhoven.

Tench, P. (1990). *The Roles of Intonation in English Discourse*. Peter Lang: Bern.

Thibault, P. et Vincent, D (1990). Un corpus de français parlé. *Recherches Sociolinguistiques* 1, 145 p.

Vion, R. (1992). *La Communication Verbale,Aanalyse des Interactions*. Hachette: Paris.

Wang, M. and Hirschberg, J. (1992). Automatic classification of intonational phrases boundaries. *Computer Speech and Language*, 6, 175-196.

Wightman, C.W. and Ostendorf, M. (1994). *Automatic labeling of prosodic patterns. IEEE Transactions on Speech and Audio Processing* (october)

Wightman, C.W. and Campbell, N. (1994). Improved labeling of prosodic structure, *IEEE Transactions on Speech and Audio Processing.*

Zue, V., Seneff, S. and Glass, J. (1990). Speech Database development at M.I.T: TIMIT and beyond. *Speech Communication*, 9 (4), 351-356.

Zwanenburg, W. (1964). *Recherches sur la Prosodie de la Phrase Française*. Thèse de doctorat, Université de Leyde.

Annexe

Exemple de fiche signalétique

Corpus Thématique

- nom : J.O. d'Albertville
- référence : FRRA0700
- durée : 25 min 30 sec.

- archivage sur DAT
  - lecteur DAT Sony
  - 10/09/1994
  - I.P. Aix-en-Provence
  - L.S.

- date d'enregistrement : 10/06/1994
- lieu d'enregistrement : ?
- matériel d'enregistrement : chaine HI-FI
- source : radio - RTL - journal des sports

Passage

- nom : interview Carole Merle
- référence : FRRA07B0
- durée : 1 min 15 sec.

- transcription orthographique
  - notation GARS
  - 08/10/1994
  - I.P. Aix-en-Provence
  - A.S.

- qualité sonore : bonne
- remarques : nombreux chevauchements

Module

- nom : interview
- référence : FRRA07B1
- durée : 55 sec.

- Typologie de l'interaction :

Conditions

☑ spontané		non spontané	☐
☐ parole publique		parole privée	☐

Structure de l'interaction

☑ à structure d'échange — sans structure d'échange ☐

☑ complémentaire	/	symétrique	☐
☑ coopératif	/	compétitif	☐
☐ finalité interne	/	finalité externe	☑
☑ formel	/	informel	☐

Types

théâtre/cinéma	☐
lecture	☐
récitation	☐
lecture interprétée	☐
conversation	☐
transaction	☐
consultation	☐
enquête	☐
entretien	☑
dispute	☐
débat	☐
discussion	☐
autre	☐