

Carnets de grammaire

Rapports internes de l'ERSS

Rapport n° 10 - octobre 2002

Constituer des corpus oraux : *Méthodes et Outils*

Elisabeth Delais-Roussarie*

* Equipe de Recherche en Syntaxe et Sémantique
UMR 5610
CNRS & Université de Toulouse-Le Mirail
email : Elisabeth.Roussarie@wanadoo.fr

Constituer des corpus oraux : Méthodes et Outils

ELISABETH DELAIS-ROUSSARIE

AVANT-PROPOS

Ce document est composé de trois chapitres qui vont paraître dans l'ouvrage édité par E. Delais-Roussarie, J. Durand et L. Labrune « Corpus et Variation en Phonologie : Méthodes et Analyses ».

Ces chapitres ont pour objectif :

- de présenter les différentes recommandations qui ont été faites par le TEI et le groupe d'experts européens EAGLES en matière de constitution et d'annotation de corpus oraux. Ces recommandations ont pour but de favoriser l'échange et la réutilisation des corpus à travers la communauté scientifique ;
- de donner un aperçu des différents problèmes que posent la constitution et la mise en forme des corpus oraux : choix des locuteurs et des situations d'enregistrement, mise au point des procédures de transcription des données, etc. ;
- de présenter plusieurs outils informatiques d'aide à la transcription des données. Une attention particulière est accordée à PRAAT, dans la mesure où ce programme est actuellement utilisé dans de nombreux projets français et étrangers.

CHAPITRE 1

Introduction	1
1. Constituer et documenter un corpus	2
1.1 Des données au corpus : quelques mises au point	2
1.2 Protocole, échantillonnage et typologie	4
1.3 Documenter un corpus	5
2. Transcrire les données	7
2.1 Qu'est-ce que transcrire	7
2.2 La transcription orthographique : Conventions et pratiques	8
2.3 Vers une uniformisation	12
3. Enrichir les annotations : conventions et outils	14
3.1 Etiquetage morphologique	14
3.2 Etiquetage syntaxique	16
3.2.1 Méthodes d'étiquetage syntaxique	17
3.2.2 Les niveaux d'étiquetage définis	17
3.2.3 Les recommandations formulées	18
3.3 Annotations prosodiques	20
3.3.1 Annotation prosodique : définition et niveaux	20
3.3.2 Approches utilisées	21
3.3.3 Vers une automatisation de la tâche	23
Conclusion	24
Grille de Biber	29
Grille élaborée par Sinclair et repris par Eagles	29
Recommandations d'EAGLES	30

CHAPITRE 2

Introduction	32
1. Outils d'aide à la transcription sans éditeur intégré	33
1.1 SoundScriber	33
1.2 VoiceWalker	35
1.3 Synthèse	37
2. Outils d'aide à la transcription avec éditeurs intégrés et alignement sur le signal	37
2.1 Soundwalker	37
2.2 Transcriber	40
3. Outils d'aide à la transcription permettant d'effectuer des analyses et/ou des requêtes	43
3.1 CLAN et le système CHILDES	43
3.1.1 Les modes d'édition	47
3.1.2 Les modes de fonctionnement de l'éditeur	48
3.2 AG Toolkit	50
3.2.1 Les graphes d'annotation	51
3.2.2 TableTrans	52
Conclusion	54

CHAPITRE 3

Introduction	58
1. Commencer avec PRAAT	59
1.1 Présentation générale	59
1.2 La prise en main	60
1.3 Effectuer des enregistrements ou saisir des données audio sous PRAAT	61
2. Annoter et transcrire des données avec PRAAT	63
2.1 Qu'est-ce que transcrire ou annoter et comment commencer ?	63
2.2 Définir les intervalles et leur assigner une étiquette ?	67
2.3 Sauvegarder ou modifier une annotation ou une transcription ?	67
3. Enrichissement et gestion des annotations	69
3.1 Créer, enrichir, fusionner et extraire des tires.....	70
3.1.1 Créer, modifier, extraire et fusionner des tires dans la fenêtre « Praat Objects »	70
3.1.2 Créer, modifier, extraire des tires dans une fenêtre « Textgrid »	73
3.2 Que peut-on faire d'autre avec les fichiers d'annotation ?	74
3.2.1 Faire des requêtes et extraire des informations à partir des fichiers d'annotation	75
3.2.2 Eclater les fichiers à partir des segmentations en intervalles	76
Conclusion.....	79

Constitution et annotation de corpus : Méthode et Recommandations

Introduction

La demande pour les corpus est de plus en plus importante, et cela dans de nombreux domaines : traitement automatique de la parole, traitement automatique des langues naturelles, didactique des langues, linguistique descriptive, psycholinguistique, sociolinguistique, etc. Mais derrière le terme « corpus » utilisé dans ces demandes se cachent plusieurs réalités : les uns sont en quête de corpus écrits, d'autres recherchent des corpus oraux, d'autres encore des corpus parallèles ou des corpus de parole. Bien que de nombreux projets visent actuellement à la constitution et à la diffusion de corpus, l'offre et l'état d'avancement en matière de disponibilité des données sont loin d'être identiques pour ces différents cas. Alors qu'il existe un nombre assez important de corpus écrits disponibles et exploitables pour travailler sur l'anglais et sur le français, la situation est nettement moins positive en matière de corpus oraux, et dans une moindre mesure, de corpus de parole. Pour mener des recherches sur le français parlé, il existe très peu de corpus oraux alignés, exploitables et accessibles.

La faiblesse de l'offre dans certains domaines peut paraître d'autant plus étonnante que nous vivons à une époque où il est apparemment facile d'obtenir des données numériques : de nombreuses données textuelles ont été saisies sur des traitements de texte ou des éditeurs et circulent sous format numérique sur la toile, des documents sonores sont souvent sauvegardés dans un format numérique (wav, MP3, etc.). Mais constituer un corpus ne signifie pas « regrouper » des fichiers ou des documents informatiques contenant des données langagières et provenant de sources variées (cf. Habert et al. (1997a et b)). La création d'un corpus linguistique repose sur un travail qui s'étend sur trois périodes :

- le travail préparatoire qui se fait avant la phase de saisie ou d'enregistrement des données ;

- le travail de saisie ou d'enregistrement des données proprement dit ;
- le travail de mise en forme des données afin de les rendre exploitables (transcription et annotation).

Dans ce chapitre, nous tentons de présenter certains aspects de ces phases de travail, en insistant surtout sur la première et la dernière phase. Certains problèmes relatifs à l'enregistrement des données audio sont traités dans le chapitre de J. M. Tarrier. Nous nous intéressons principalement à la constitution de corpus oraux et de corpus de parole, de nombreux ouvrages traitant déjà des problèmes de constitution des corpus écrits (cf., entre autres, Habert et al. (1997a et b)). Dans une première partie, après quelques rappels terminologiques, nous présentons en quoi consiste le travail préparatoire à la constitution d'un corpus oral. Puis, dans un second temps, nous nous intéressons à la tâche de transcription des données qui est un préalable à toute exploitation et diffusion d'un corpus oral. Pour finir, nous expliquons en quoi consiste l'annotation d'un corpus, mais seuls les étiquetages morphologiques, syntaxiques et prosodiques seront pris en compte. Dans l'ensemble de ce chapitre, nous essaierons de présenter les travaux de recherche menés par des groupes d'experts, notamment européens, afin de faciliter l'échange de corpus grâce à une « uniformisation » ou une standardisation des procédures.

1. Constituer et documenter un corpus

Avant de se lancer dans l'enregistrement des données qui composent un corpus, le linguiste doit réfléchir afin de déterminer :

- quels types de données il souhaite rassembler ;
- qui il veut enregistrer ;
- dans quelles situations doivent s'effectuer les enregistrements, etc.

Les réponses à ses questions dépendent en partie de l'utilisation qui sera faite ultérieurement du corpus, et donc des objectifs visés. Aussi est-il nécessaire de savoir quel type de corpus on veut constituer. Dans une première partie, nous ferons un rappel sur ce que sont les types de corpus et de documents constitutifs de corpus. Puis, nous tenterons de présenter les différentes questions et interrogations auxquelles il faut répondre avant de se lancer dans la collecte des données. Pour finir, nous expliquerons ce que signifie « documenter un corpus » et pourquoi cette tâche est importante pour toute utilisation ultérieure des données.

1.1 Des données au corpus : quelques mises au point

Comme nous l'avons mentionné dans l'introduction, un corpus n'est pas un ensemble de données variées, mais se définit d'après Sinclair (1996) comme :

« une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon de langage ».

Traduction de Sinclair (1996),
reprise dans Habert et al (1997b)

A côté de cette définition très restrictive qui peut conduire à se demander si de nombreux corpus oraux existant sont bien des corpus (corpus constitué dans le cadre de PFC, corpus du GARS, etc., base de données BREF, etc.), le groupe d'experts européens travaillant sur les corpus oraux et de parole propose une autre définition plus souple (cf. Gibbon et al (1998)) :

“A corpus is any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by people in others organisations.”

Dans le cadre de ce travail, nous retenons plutôt cette seconde définition qui tient compte de certaines spécificités des corpus oraux : la saisie des données peut se faire sur de nombreuses années. L'unité vient alors dans la méthodologie utilisée pour collecter les données, les transcrire et les annoter, cette dernière étant clairement explicitée dans des documents joints aux données.

A côté de ces définitions assez générales, une typologie des corpus a été proposée (cf., entre autres, Eagles (1996a), Habert (1997a)) et permet de distinguer :

- *les corpus écrits* qui regroupent des documents (ou textes) écrits sélectionnés selon des critères variés. Appartiennent par exemple à cette catégorie les corpus littéraires ;
- *les corpus oraux* qui regroupent des documents correspondant à des transcriptions, alignées ou non sur le signal, de productions orales (entretiens libres, entretiens guidés, exposés, etc.). Ces données peuvent servir pour mener des recherches sur l'oral (syntaxe de l'oral, prosodie, etc.) et ont souvent été collectées dans des situations diversifiées ;
- *les corpus de parole* qui regroupent des enregistrements de parole effectués dans des conditions expérimentales. Les données sont de nature très variée (lecture de logatomes, lecture de mots, lecture de phrases isolées, lecture de textes, etc.). En outre, la taille globale du corpus peut être assez limitée.
- *les corpus parallèles* qui sont composés de documents, traduits dans une ou plusieurs langues. Certains documents sont donc la traduction littérale d'autres documents du corpus ;

Par ailleurs, il est possible de faire une distinction entre *corpus de référence* et *corpus de spécialité*. Dans ce cas, le critère décisif pour classer les corpus est leur composition interne. Dans un corpus de référence, les documents constitutifs (ou « *text* » dans la terminologie anglo-saxonne) doivent être suffisamment

représentatifs pour donner une image de la langue telle qu'elle est dans toute sa diversité. Les données doivent provenir de différentes situations de communication, représenter plusieurs styles de documents écrits, etc. Un corpus de spécialité regroupe au contraire des données linguistiques sélectionnées pour représenter une dimension particulière : un domaine (français médical, etc.), un thème, une situation de communication, etc.

Pour constituer un corpus, que ce soit un corpus de référence ou un corpus de spécialité, les linguistes doivent établir une liste de critères pertinents pour sélectionner les données. Ce travail doit permettre de parvenir à une représentativité. La méthodologie utilisée pour effectuer cette tâche va être abordée dans la section suivante.

1.2 Protocole, échantillonnage et typologie

Pour constituer un corpus, le linguiste doit définir une méthodologie et élaborer une liste de critères. Cela permet de sélectionner les données et de décider des enregistrements qui doivent être faits. Parmi ces critères peuvent être mentionnés :

- *les objectifs de recherche visés* : si l'utilisateur veut construire un corpus pour étudier le vocabulaire des jeunes, il ne choisira pas les situations d'enregistrement et les documents à retenir de la même façon que s'il veut travailler sur des discours politiques ;
- *le type de corpus* : si le but est de constituer un corpus de référence, plusieurs critères interviennent en parallèle pour obtenir une meilleure représentativité possible. Il faut en effet tenir compte des différences socio-économiques et géographiques entre les locuteurs, des différentes situations de communication, des nombreux types de documents écrits (textes littéraires, rapports professionnelles, articles de presse, etc.). En revanche, pour créer un corpus de spécialité, il faut réfléchir aux paramètres qui permettent de bien choisir des données représentatives pour tel ou tel aspect de la langue.
- *les modalités d'enregistrement des données* : si le corpus est construit pour travailler sur les caractéristiques acoustiques d'un son, il est important que les enregistrements soient faits dans des situations expérimentales optimales (utilisation d'une chambre sourde, enregistrement numérique, etc.). Si, au contraire, le linguiste veut avoir une idée de la langue parlée dans certains quartiers populaires, il est préférable qu'il enregistre les locuteurs dans leur milieu, voire avec un informateur.

Dans un corpus de référence ou un corpus de spécialité, qu'il s'agisse d'un corpus oral ou d'un corpus écrit, la sélection des données (et des locuteurs) repose sur des procédures d'échantillonnage et de nettoyage. Pour la partie orale du BNC (British National Corpus) qui comprend 10 millions de mots, deux composantes ont été définies pour procéder à la collecte des données (cf. Aston & Burnard (1998), Crowdy (1995)) :

- une composante démographique qui est intervenue dans l'enregistrement des conversations « informelles ». Les locuteurs ont été

choisis en fonction d'un échantillonnage socio-démographique et géographique (cf. Aston & Burnard (1998)) ;

- une composante contextuelle qui a permis de sélectionner les données plus formelles (enregistrements de conférences, de débats publics, de séminaires, de meetings, d'émissions radiophoniques, etc.). Une classification par topique et par type d'interaction a été faite afin de ne pas sur-représenter certains sous-langages ou styles de communication.

Les tâches préparatoires visent à établir la liste des types de données qu'il faudra obtenir pour satisfaire les objectifs de constitution du corpus. Un protocole d'enquête peut ensuite être rédigé afin de clairement spécifier la démarche qui doit être utilisée pour collecter les données. Pour le projet *Phonologie du Français Contemporain*, un protocole a, par exemple, été remis à tous les participants afin que les enquêtes soient menées de la même façon. Ce document fournit également tous les éléments qui se retrouvent généralement dans les enquêtes sociolinguistiques, à savoir une liste de mots et de paires minimales que tout interviewé doit lire, un texte qui a été élaboré afin de permettre l'étude de différents points de phonologie (cf. Durand et al. (ce volume)).

Pour sélectionner les types de documents qui doivent être représentés dans le corpus, le linguiste peut également s'appuyer sur les travaux de typologie (cf. Biber (1994), EAGLES (1996f)). Bien que les classements à partir des notions de genres (genres littéraires, etc.) et de registres ne soient pas toujours convaincants, ils peuvent servir de points d'appui. Il est évident que des critères de catégorisation plus intuitifs reposant sur les thématiques (...) interviendront pour faire ces sélections (cf., entre autres, Habert et al. (1997a et b)).

1.3 Documenter un corpus

Après avoir sélectionné les données ou documents qui sont intégrés dans un corpus, et effectué, le cas échéant, les enregistrements, le corpus n'est pas directement utilisable. Pour les corpus oraux, c'est assez évident : les données doivent encore être numérisées (ou mises sous format numérique), transcrites au moins orthographiquement, voire annotées. Pour les corpus écrits, nous pourrions penser que l'essentiel est fait. Mais, dans un cas comme dans l'autre, il faut impérativement documenter le corpus, c'est à dire élaborer une documentation dans laquelle seront exposés :

- les choix faits pour sélectionner les documents ;
- les principes d'annotation utilisés ;
- le contenu de chacun des documents ou enregistrement.

Cette tâche est importante car, comme le dit si bien Habert et al. (1997b) :

« Sans une documentation jointe, un corpus est mort-né. L'un des dangers de la facilité actuelle de rassembler des textes électroniques est précisément que les objectifs de regroupement ainsi que ceux des annotations effectuées ne soient pas enregistrées : le corpus cesse d'être utilisable dès que se perd la mémoire de ces choix.

La documentation doit couvrir deux volets distincts : les sources utilisées et la responsabilité éditoriale de constitution du corpus d'une part, les conventions d'annotation d'autre part. »

Habert et al. (1997b), p. 156

Dans Habert et al. (1997a), deux grilles sont proposées et peuvent servir de point d'appui pour documenter un corpus, et surtout décrire les documents qui le constituent. L'une est inspirée des travaux de Biber et l'autre est élaborée par Sinclair et est reprise par Eagles (1996f). Elles sont toutes deux présentées en annexe (cf. § 7). Celle de Sinclair est d'une certaine façon plus complète et plus facilement adaptable pour les corpus oraux.

Pour chaque enregistrement, plusieurs informations doivent en général être fournies. Elles concernent :

- le locuteur : origine géographique, âge, sexe, niveau socio-culturel, langue première, etc.
- la situation d'interlocution : informations sur les relations existants entre le locuteur et ses interlocuteurs (qu'ils parlent ou soient simplement présents);
- la situation d'enregistrement : description des tâches à effectuer et des consignes données, lieu et cadre dans lequel se sont faits les enregistrements ;
- le support et le matériel utilisé.

Lorsque le corpus sera mis en forme, ses informations seront souvent données dans les en-têtes (cf. les recommandations du TEI, Sperberg-McQueen et al. (1999)).

Pour les corpus oraux, la documentation doit aussi intégrer des recommandations sur la mise en forme et la normalisation des données. Les points suivants doivent être abordés :

- les techniques d'enregistrement ;
- le choix des supports (cassette, DAT, Minidisc, etc.) ;
- les techniques de numérisation et le choix des codages (fréquence d'échantillonnage, etc.).

Cette tâche de documentation doit être faite pour permettre la diffusion et l'exploitation des corpus. Elle se situe à la frontière entre le travail préparatoire à la phase d'enregistrement et celui qui la suit : la documentation explicite en effet les choix retenus pour constituer le corpus, mais elle intègre également les décisions prises en matière de numérisation des données, de transcription et d'annotation.

2. Transcrire les données

Dans le chapitre précédent, nous avons tenté de présenter de façon synthétique les différents points auxquels il faut réfléchir pour collecter les données constitutives d'un corpus. Dans le cas des corpus oraux ou de parole, la collecte s'effectue sous la forme d'enregistrements. Même lorsque ces derniers ont été numérisés en respectant des formats de codage audio standardisés (wav., etc.), ils ne sont pas exploitables. En effet, des données audio brutes ne peuvent pas être analysées sans un travail préalable de transcription et de segmentation. Dans cette section, nous allons donc présenter en quoi consiste la tâche de transcription. Une attention particulière sera accordée à la transcription orthographique.

2.1 Qu'est-ce que transcrire

Transcrire des données sonores consiste à fournir une représentation symbolique du signal. Cette représentation n'est pas équivalente au signal, dans la mesure où elle est le résultat d'une analyse, ou plutôt d'une abstraction, des données réelles. Mais plusieurs types ou niveaux de transcription peuvent être définis (cf., entre autres, Gibbon et al (1998) ; Eagles (1996e)). Nous allons en présenter les caractéristiques ci-après :

- la transcription à partir des scripts d'enregistrement : si les enregistrements ont été effectués à partir de la lecture de scripts, le niveau de transcription le plus aisé et le plus facile à fournir est de « reprendre » intégralement les scripts.
- la transcription orthographique consiste à encoder sous forme orthographique standard les éléments contenus dans le signal. Ce niveau est très souvent utilisé pour transcrire les corpus oraux de taille assez importante. A partir de la représentation orthographique, il est possible d'obtenir une représentation phonémique à l'aide de systèmes de conversion graphème-phonème. Nous reviendrons sur ce niveau de transcription dans la suite de cette section.
- la transcription correspondant à une représentation phonémique de citation (« *citation-phonemic representation* ») : à ce niveau, tous les items linguistiques (ou mots) sont transcrits sous la forme d'une « représentation phonologique sous-jacente », qui correspond à la façon dont chaque item est prononcé en isolation.
- la transcription phonétique large : bien que la transcription reste phonémique à ce niveau, certains phénomènes de sandhi ou de jonction sont notés (liaison, insertion de schwa, « intrusive r », etc.), alors qu'ils ne l'étaient pas au niveau précédent.
- la transcription phonétique étroite : elle consiste à transcrire tous les sons qui ont été produits, en tenant compte des variantes allophoniques. Ce niveau de transcription ne repose pas nécessairement sur une segmentation et une analyse du signal. En

revanche, il est surtout réservé à des corpus de taille limitée et nécessite l'utilisation de symboles phonétiques (API, SAMPA, etc.).

- la transcription acoustique-phonétique : ce niveau de transcription repose sur une analyse et une segmentation précise du signal de parole, dans la mesure où il encode des informations aussi précises que les phases de fermeture et d'explosion lors de la production des plosives, les dévoisements, etc. Il est surtout utilisé pour transcrire des corpus courts et servant à des travaux de recherche précis en phonétique ou en traitement du signal.

Ces différents niveaux de transcription ne fournissent pas les mêmes informations. Pour choisir celui qui sera utilisé pour transcrire un corpus, plusieurs éléments entrent en ligne de compte : la taille du corpus, le degré de détails recherché et les objectifs visés. Si le corpus a été construit pour étudier les modalités des interactions et des prises de parole (chevauchements, silences, etc.), une transcription orthographique est suffisante. En revanche, si le but est d'étudier les variantes de prononciation pour certains termes, une transcription phonétique étroite est recommandée, au moins pour les parties pertinentes du corpus.

Il faut aussi noter que la difficulté de la tâche de transcription dépend bien entendu du niveau choisi, mais également du type de documents. Transcrire un texte n'est pas la même chose que transcrire de la parole spontanée, surtout s'il s'agit d'une conversation entre plusieurs personnes. Dans la section qui suit, nous allons nous intéresser plus particulièrement à la transcription orthographique, afin d'en décrire les difficultés et les contraintes.

2.2 La transcription orthographique :

Conventions et pratiques

Comme nous l'avons mentionné dans le paragraphe précédent, transcrire orthographiquement un document sonore consiste à représenter symboliquement, sous forme orthographique, ce qui est contenu dans le signal. Ce niveau de transcription est très souvent retenu pour les corpus oraux (cf. corpus du GARS, partie orale du British National Corpus, corpus de PFC, etc.). Mais, bien que des outils informatiques aient été développés pour aider le transcripateur dans sa tâche (cf. Delais-Roussarie, ce volume), celle-ci est souvent rendue difficile pour plusieurs raisons. De nombreux problèmes auxquels doit faire face tout transcripateur sont clairement exposés dans Benveniste et Jeanjean (1987). Parmi ceux-ci, nous pouvons mentionner :

- les problèmes liés à la difficulté d'écoute, d'autant que les enregistrements et l'écoute monophonique sur magnétophone provoquent des déformations ;
- les difficultés résultant de la reconstruction perceptive : en écoutant les données sonores, le transcripateur peut avoir tendance à reconstruire certains éléments inexistant sur le signal (mots erronés, etc.) ;

- le poids des préjugés, notamment linguistiques : le transcripateur, en s'appuyant sur sa compétence linguistique et sa représentation de la langue, voire du « bien parler », peut être amené à modifier ce qu'il entend. Ainsi, par exemple, le « ne » de négation, bien que souvent absent à l'oral, peut être inséré dans des transcriptions ;
- les ambiguïtés propres au code oral : dans de nombreux cas, une même séquence sonore peut correspondre à plusieurs interprétations à l'écrit. Considérons les deux exemples suivants :

Séquence sonore	Interprétation et encodage
[askəʃse:imãzsu:vã:oresto:]	A ce que je sais <u>il mange</u> souvent au resto.
	A ce que je sais <u>ils mangent</u> souvent au resto.
[lapatri:snepatu:ʒu:rlãdrwaulõne] ¹	La patrie ce n'est pas toujours l'endroit <u>où l'on est</u> .
	La patrie ce n'est pas toujours l'endroit <u>où l'on naît</u> .

Dans certains cas, ces ambiguïtés peuvent être levées par le contexte linguistique ou communicatif, mais dans d'autres, le transcripateur ne peut pas décider et doit donc retenir les différentes possibilités.

Outre ces difficultés, la tâche de transcription relève d'un paradoxe :

« Transcrire de la langue parlée tient un peu du paradoxe : garder dans une représentation écrite certaines caractéristiques de l' « oralité » ; faire le « rendu » de la chose orale tout en restant dans des habitudes de lecture établies depuis longtemps pour la chose écrite... On va se trouver tiraillé entre deux exigences : la fidélité à la chose parlée et la lisibilité de son rendu par écrit. »

Blanche-Benveniste & Jeanjean (1987), p. 115

Pour essayer de représenter au mieux l'oral sous la forme d'une transcription orthographique, les transcripateurs doivent décider de deux choses :

- ce qu'il faut noter : dans une transcription orthographique, on peut se limiter à encoder ce qui est dit par les locuteurs sur le signal sonore, mais on peut aussi noter les changements de locuteurs et les tours de parole, les faux départs, les hésitations, les chevauchements, les événements non linguistiques, qu'ils soient sonores ou non (aboiments d'un chien, toux, rires, hochements de tête, etc.), communicatifs ou non (pires, pleurs, étouffements, etc.).

¹ Cet exemple est extrait de Benveniste et Jeanjean (1987).

- comment il faut encoder les informations : les transcripteurs doivent définir comment ils souhaitent utiliser le code orthographique (respect ou non de l'orthographe standard, utilisation de la ponctuation, etc.), comment ils vont représenter typographiquement certains phénomènes (chevauchements, faux départs, etc.) et comment les transcriptions vont être mises en forme (représentation sur plusieurs lignes ou sur une seule, etc.)

Les choix faits doivent généralement être explicités sous forme de conventions et respectés par tous les transcripteurs du corpus. Ils doivent impérativement être diffusés avec les données pour en permettre leur réutilisation.

En général, si nous comparons plusieurs corpus oraux contenant le même type de documents (conversations guidées ou non guidées, par exemple), il existe une sorte de consensus dans la définition de ce qu'il faut encoder. Les éléments suivants sont très souvent retenus : les changements de locuteurs et les tours de parole, les chevauchements, les hésitations, les faux départs, les mots incomplets, les phénomènes tels que les rires, les pleurs, les toux, les bruits extérieurs. En revanche, des divergences apparaissent pour encoder de façon détaillée des entités telles que les pauses (voire leur longueur relative), les allongements, les phénomènes de sandhi (liaison, insertion de schwa) : certains choisissent de mentionner ces phénomènes, tandis que d'autres ne les traitent pas.

En ce qui concerne la façon de transcrire, de nombreux désaccords existent dans la pratique. Parmi les points de divergences, nous pouvons en citer deux :

- l'utilisation de l'orthographe standard : bien que toute transcription orthographique doive respecter l'orthographe standard, l'étude de plusieurs transcriptions montre que ce n'est pas toujours le cas. Le GARS refuse tout trucage orthographique et respecte l'orthographe standard, les grands dictionnaires de langue française servant de référence pour les mots courants, les noms propres, les onomatopées, etc. Si certaines prononciations particulières doivent être notées, elles le sont sous forme de notes². En revanche, dans les corpus traduits conformément aux conventions de Valibel³, l'orthographe standard n'est pas toujours respectée, notamment suite à l'utilisation de « jeux » typographiques : des majuscules sont insérées en milieu de mots afin de noter les réalisations des schwas ou des segments flottants, des dièses sont insérés à la place d'un graphème si ce dernier n'est pas réalisé (« t#' as vu » pour « tu as vu »). Considérons également l'extrait de transcription ci-après :

² Les conventions du GARS peuvent être consultées à l'adresse suivante : <http://www.up.univ-mrs.fr/delic/corpus/index.html>.

³ Valibel est le Centre de recherche sur les Variétés linguistiques du français en Belgique. Il a mis au point des conventions de transcription de l'oral qui peuvent être consultées et téléchargées à l'adresse suivante : <http://valibel.fltr.ucl.ac.be/>.

L0 c' est ça // alors **i-y-a i-y-a** quelques années tu as trouvé chaussure à ton pied tu t' es mariée // euh est-ce-que est-ce-que ton mari t' accompagne pour faire les courses ?
 L1 jamais c' est même moi qui vais chercher ses vêtements // **i** m' a encore dit tantôt demain tu dois aller me chercher une paire de chaussures de travail parce-que les miennes // sont usées // jamais **à-part** quand **i** faut un veston ou quoi je ne sais pas l' essayer à sa place // mais / généralement comme pour lui je vais toujours dans la même boutique / pour acheter ses pantalons // et **bè** je vais là je vais en chercher deux / ou trois // il les essaie à la maison je reporte celui qui ne me ceux qui ne vont pas

Extrait de transcription selon les conventions de Valibel

La transcription de plusieurs items est loin de respecter l'orthographe standard : **i-y-a** pour « il y a », **i** pour « il(s) », etc. Ces modifications visent à mieux rendre compte de l'oral, mais elles peuvent aussi conduire à des manques de consistance et à des erreurs (cf. Bilger (2000)).

- l'utilisation de la ponctuation : dans les conventions du GARS et de VALIBEL, l'utilisation de la ponctuation est rejetée, mais les pauses sont notées. Une distinction est même faite entre pauses longues et pauses courtes. Pour transcrire la partie orale du BNC, en revanche, il a été décidé d'utiliser la ponctuation. French (1992) explique en ces termes comment le transcripteur doit s'y prendre :

« Try to be guided by intonation – the rises and falls in the voice – as well as by the words themselves. If it sounds as though someone has finished a sentence and gone to another (their voice drops, they take a breath and start on a higher note), then it's probably safe to start a new sentence.»

Payne (1995) revient également sur ce problème et décrit de la façon suivante « ces phrases » de l'oral :

« The resulting « functional sentence » is perhaps difficult to define precisely in linguistic terms, but as an essentially practical unit creates few problems for transcribers, as it is using their intuition about when speakers are starting, continuing and completing what they are saying on the one hand, and when they are abandoning incomplete utterance on the other.»

Payne (1995), p. 204

Comme nous le voyons, les divergences sont beaucoup plus importantes dès qu'est abordée la question du « comment transcrire ? ». Il existe d'ailleurs d'autres différences si l'on compare les pratiques, notamment en ce qui concerne les choix typographiques : les uns indiquent les pauses par des barres obliques (« / »), tandis que d'autres utilisent des tirets (« - »), les crochets ont

des significations différentes chez les uns et les autres. Ce manque d'uniformité n'est pas très grave, dans la mesure où il repose sur des points de forme et non de fond. Il est d'ailleurs facilement dépassé si un document joint au corpus indique clairement les choix typographiques faits.

2.3 Vers une uniformisation

Les divergences de pratique que nous avons mentionnées dans la section précédente reposent parfois sur des conceptions différentes de ce que doivent être des représentations orthographiques de données sonores. Elles peuvent donc former un obstacle à l'échange et à la réutilisation des corpus oraux. Afin de pallier ces difficultés et de parvenir à une « standardisation », les experts européens (EAGLES) ont proposé une série de recommandations (cf., entre autres, Gibbon et al. (1998) et Eagles (1996e)). Pour faire ce travail, ils se sont appuyés sur plusieurs éléments :

- les travaux du *TEI* (Text Encoding Initiative) sur les documents oraux ;
- les besoins des différentes communautés qui utilisent des corpus oraux ou de parole : pour ce faire, les pratiques et les niveaux de transcriptions choisis dans la communauté de la parole (traitement automatique de la parole, phonétique, etc.) et dans la communauté linguistique (linguistique de corpus, linguistique descriptive, etc.) ont été comparés afin d'établir un minimum commun. La transcription orthographique a été retenue comme minimum commun, dans la mesure où elle peut permettre l'accès au « niveau phonémique de citation ».

Nous allons exposer les grandes lignes de ces recommandations, en indiquant à la fois ce qui doit être transcrits et comment cela doit être faits. Une traduction plus précise de certaines des recommandations est proposée en annexe (cf. § 7). Le lecteur intéressé peut également consulter Eagles (1996e) et Gibbon et al. (1998).

En plus des énoncés produits par les différents interlocuteurs, plusieurs informations doivent être encodées. Elles peuvent être classées en plusieurs catégories :

- **les informations relatives aux locuteurs en présence :**
 - l'identité des locuteurs doit être indiquée dans une « en-tête ». La description des locuteurs doit comporter leur nom (et la façon de le coder), mais également leur âge, leur origine géographique, etc. ;
 - les tours de parole ou changement de locuteur ;
 - les chevauchements.
- **les informations concernant les productions du locuteur, à savoir :**
 - les omissions lors de la lecture des textes,

- les réparations ou reprises, les répétitions, etc.
 - les fragments de mots, si ces derniers ne sont pas intégralement prononcés.
 - Les fragments inintelligibles. Dans ce cas, une distinction entre 'devinés', 'incertains' et inintelligibles' peut être faite.
- **les informations renvoyant à des phénomènes sonores :**
- les événements vocaux et semi-lexicaux, c'est à dire les pauses remplies (« filled pauses »), les hésitations. Pour ces événements, l'orthographe retenue doit être unifiée sur l'ensemble du corpus ;
 - les événements vocaux non lexicaux, c'est à dire les étouffements, les rires, la toux, les soupirs, etc. ;
 - les événements sonores non communicatifs et non vocaux, c'est à dire les bruits produits par les auditeurs, les claquements de porte, le téléphone, les sonneries, etc.

En ce qui concerne la façon d'encoder orthographiquement ces éléments, Eagles a proposé une série de recommandations qui reprend dans les grandes lignes les propositions faites par French (1992). Les voici :

- **les mots produits (y compris les onomatopées) sont transcrits en utilisant l'orthographe standard** : pour les variantes régionales qui ne sont pas nécessairement dans les dictionnaires « standards », il est possible de construire une liste annexe. Ce document sera alors fourni avec le corpus. Cette pratique est utilisée par Valibel pour certains termes belges (cf. <http://valibel.fltr.ucl.ac.be/>) ;
- **seules les formes contractées admises dans le dictionnaire sont utilisées** (« *gonna* » et « *wanna* » pour l'anglais, etc.) ;
- **les frontières de phrases ou d'énoncés sont indiquées par des ponctuations fortes** (point, point d'interrogation et point d'exclamation dans les ordres) ;
- **aucune ponctuation (virgule et point virgule) n'est utilisée à l'intérieur des énoncés** ;
- **les citations de textes lus ou les discours directs sont indiqués par des guillemets simples** ;
- **les sigles et acronymes sont transcrits tels qu'ils ont été produits** : en général, si un sigle est lu, il est noté comme une suite de majuscules ([krif] est noté CRIF⁴ et [pyf] PUF⁵). En revanche, s'il est épilé, un point est noté après chaque graphème ([seɛnɛrɛs] est noté C.N.R.S⁶) ;

⁴ CRIF : Conseil Représentatif des Institutions Juives de France.

⁵ PUF : Presses Universitaires de France.

⁶ CNRS : Centre National de la Recherche Scientifique.

- **les chiffres sont transcrits orthographiquement en tenant compte des variantes de prononciation** : le chiffre 90 est transcrit « *nonante* », s'il est prononcé [nonãt] et « *quatre-vingt dix* », si le locuteur dit [katʁəvẽdis].

Dans ces recommandations, aucune indication n'est donnée sur la forme que doit prendre les transcriptions. Les typographies utilisées pour représenter les chevauchements, les mots incompréhensibles et les fragments de mots sont choisies pour chaque projet de constitution de corpus, mais doivent être consignés dans un document. De même, peu de recommandations concernent la représentation spatiale des transcriptions : faut-il utiliser une tire par locuteur ou non ? où doivent être notées les difficultés rencontrées ? peut-on faire des appels de note pour décrire les réalisations phonétiques observées, etc. En fait, un seul point est mentionné :

« pour les transcriptions de dialogue entre plus de deux locuteurs, utiliser une notation sous forme de « partitions musicales ».

D'une façon générale, les recommandations faites par le TEI et Eagles en matière de transcription orthographique de documents oraux sont assez intéressantes, dans la mesure où elles proposent un cadre qui peut être aisément suivi, tout en laissant une assez grande liberté aux transcrip-teurs en ce qui concerne la forme des transcriptions (choix typographiques, mise en forme, format informatique de sauvegarde, etc.).

3. Enrichir les annotations : conventions et outils

Dans le paragraphe précédent, nous nous sommes surtout intéressée à la transcription orthographique d'un document sonore. Mais il est également possible de fournir d'autres informations sur le contenu linguistique des documents sonores : une catégorisation morpho-syntaxique des différents mots produits, une analyse syntaxique des énoncés, une analyse prosodique, etc. En général, ces annotations se font à un niveau distinct de la transcription orthographique. Dans cette section, nous allons étudier ces trois types d'annotation. Il est intéressant de noter que l'étiquetage morpho-syntaxique et l'analyse syntaxique se font à partir de la transcription orthographique, tandis que l'annotation prosodique est effectuée à partir du signal sonore (annotation prosodique). Pour chacun de ces niveaux, nous tenterons de voir en quoi consiste la tâche d'annotation, quelles recommandations ont été formulées et dans quelle mesure des outils peuvent faciliter le travail.

3.1 Etiquetage morphologique

Etiqueter morphologiquement un corpus consiste à assigner une étiquette grammaticale à chacun des mots produits. Ce niveau d'annotation peut être effectué automatiquement avec des étiqueteurs grammaticaux (tagger). Même pour les corpus oraux, les résultats des annotations automatiques sont satisfaisants (cf., entre autres, Valli et al (1999 et 2000), Véronis (1998 et 2000)). Mais, un travail de pré-traitement est souvent nécessaire pour les corpus oraux, les étiqueteurs ayant souvent été développés pour le code écrit.

Dans bien des cas les erreurs et les difficultés rencontrées lors de l'étiquetage de documents oraux sont comparables à celles observées dans le cas des corpus écrits. Elles résultent de deux types d'éléments :

- quel est le nombre d'étiquettes retenues, et donc quel est le degré de « détails » recherché lors de la tâche d'étiquetage ;
- quel dictionnaire est utilisé pour effectuer la tâche d'annotation. Dans certaines langues ou sous-langues, il est important de construire des dictionnaires particuliers qui regroupent des formes spécifiques. Pour traiter l'oral, il est fondamental d'inclure dans le dictionnaire les différentes interjections, onomatopées, etc. qui apparaissent dans les transcriptions. En français, il est nécessaire d'utiliser un lexique de formes composées (« *après que* », « *au fur et à mesure* », « *salle de bains* », etc.).

De même que pour les transcriptions orthographiques, des recommandations ont été formulées par EAGLES pour la tâche d'annotation morphologique (cf. Eagles (1996b)). Elles portent principalement sur le choix des étiquettes et reposent sur une distinction entre étiquettes obligatoires, étiquettes recommandées et extension particulière.

Les étiquettes qui doivent obligatoirement être assignées lors des tâches d'annotation morpho-syntaxique correspondent à celles représentant les catégories majeures (Nom, Verbe, conjonction, adjectif, etc.). Ces catégories sont données dans le tableau ci-dessous (extrait de Eagles (1996b)) :

1.	N [nom]	2.	V [verbe]	3.	AJ [adjectif]
4.	PD [pronom/determinant]	5.	AT [article]	6.	AV [adverbe]
7.	AP [préposition]	8.	C [conjonction]	9.	NU [numeral]
10.	I [interjection]	11.	U [unique/ non assigné]	12.	R [résiduel]
13.	PU [ponctuation]				

L'étiquette « unique » va être utilisée pour les items appartenant à des catégories très fermées avec peu de membres (les particules négatives, par exemple). L'étiquette « résiduel », elle, est assignée aux termes qui ne rentrent pas dans les catégories usuelles (mots étrangers, formules mathématiques, etc.).

Les étiquettes recommandées correspondent à celles qui sont fréquemment utilisées dans les descriptions grammaticales traditionnelles. Elles donnent des indications plus précises tels que le nombre et le genre pour les noms et les adjectifs, le temps, le mode et la personne pour les verbes, le degré pour les adjectifs et les adverbes.

Les extensions spéciales regroupent des étiquettes optionnelles qui ont été choisies pour répondre à deux types de besoins :

- la volonté de fournir des informations génériques supplémentaires. Pour ce faire, des étiquettes ont été définies pour donner des indications d'ordre plutôt sémantique (nom comptable ou non comptable, etc.) ;
- le traitement de langues européennes particulières.

En conclusion, les recommandations faites par EAGLES pour tenter d'uniformiser la tâche d'annotation morpho-syntaxique sont assez souples. Les seules étiquettes obligatoires sont en effet celles minimalement attendues pour ce niveau d'annotation.

3.2 Etiquetage syntaxique

L'annotation syntaxique de corpus consiste à ajouter des informations syntaxiques telles que le parenthésage en groupes (ou constituants), le marquage des relations de dépendance, etc. Cette tâche, bien que liée d'une certaine façon à l'étiquetage morphologique (qui constitue souvent un prérequis aux opérations d'étiquetage syntaxique), est beaucoup plus difficile à mener à bien. De ce fait, il existe très peu de corpus arborés ou annotés syntaxiquement (cf., pour le français, Habert et al. (1997b), Abeillé et al. (2000)). Ces difficultés s'expliquent de différentes façons (la liste n'étant pas exhaustive) :

- **la complexité de la tâche** : alors que dans l'étiquetage morphologique les unités à étiqueter sont prédéfinies (à l'exception peut-être des mots composés), elles ne le sont pas dans le cas de l'étiquetage syntaxique. Quelle que soit la méthode utilisée, il faut donc élaborer une segmentation de l'énoncé, puis assigner des étiquettes aux segments, et souvent, déterminer les relations qu'ils entretiennent entre eux, etc.
- **l'absence de consensus** : même parmi les linguistes, il n'existe pas de consensus sur ce que sont les unités ou segments syntaxiques (à l'exception sans doute des parenthésages en syntagmes nominaux et en syntagmes prépositionnels), sur les étiquettes à assigner à ces unités et sur les relations que les segments entretiennent entre eux. Les différentes approches théoriques qui sont utilisées pour décrire les phénomènes syntaxiques en témoignent (grammaire de dépendance, grammaire générative chomskienne, etc.).
- **la définition de ce qu'est la syntaxe** : il est également difficile de définir clairement ce qu'il faut attendre d'un étiquetage syntaxique. Les distinctions entre forme et fonction, les phénomènes à l'interface entre syntaxe et sémantique et entre syntaxe et discours doivent-ils être pris en compte et comment ? Autant de questions qui accentuent l'absence de consensus.

Malgré ces difficultés, plusieurs approches ont été utilisées pour annoter syntaxiquement un corpus. Nous allons donc présenter les méthodes

fréquemment utilisées, les niveaux d'étiquetage définis et les recommandations formulées.

3.2.1 Méthodes d'étiquetage syntaxique

L'étiquetage syntaxique de corpus peut se faire de trois façons distinctes :

- manuellement : dans ce cas, la tâche est effectuée par des linguistes. Cette méthode offre de bons résultats, bien qu'elle soit lente et coûteuse. Elle a été utilisée pour étiqueter syntaxiquement des corpus anglo-américains (*Susanne*, par exemple. Cf. Habert et al. (1997b)).
- automatiquement : l'annotation est alors faite par des analyseurs automatiques. Bien que les résultats des analyses effectuées automatiquement s'améliorent, elles doivent souvent être revues ultérieurement par des linguistes afin de corriger les erreurs. Le corpus anglo-américain *Penn treebank*, tout comme le corpus arboré du français (cf. Abeillé et al. (2000)), ont été étiquetés par des analyseurs automatiques.
- de façon mixte : dans ce cas, l'outil informatique aide le linguiste à effectuer la tâche d'étiquetage. Le travail se fait de façon interactive. Certains systèmes informatiques sont développés dans ce but (cf. le système « *Tosca* »).

Pour choisir d'utiliser une méthode plutôt qu'une autre, plusieurs éléments entrent en ligne de compte :

- les objectifs visés et le niveau d'étiquetage recherché ;
- la taille du corpus ;
- le style des documents : s'il s'agit par exemple d'un corpus oral, il est sans doute plus difficile d'obtenir une analyse syntaxique satisfaisante à partir des analyseurs automatiques existant, ces derniers ayant surtout été conçus pour traiter l'écrit.

3.2.2 Les niveaux d'étiquetage définis

Comme nous l'avons mentionné, la tâche d'annotation syntaxique est d'autant plus difficile à mener à bien qu'elle peut être appréhendée de différentes façons et fournir plusieurs types d'informations. Une synthèse des niveaux d'annotation est proposée par les experts d'EAGLES (cf. Eagles (1996c)) :

- découpage en groupes ou constituants : cette tâche est obligatoire si l'on veut étiqueter syntaxiquement un corpus. Mais, même si tout le monde s'accorde pour reconnaître un tel niveau, le consensus est loin d'exister sur le parenthésage lui-même.
- étiquetage des groupes ou constituants : ce niveau consiste à assigner des étiquettes aux groupes précédemment définis. Il est généralement considéré comme faisant partie de toute tâche d'annotation syntaxique.

- marquage des relations de dépendance : cette tâche consiste à indiquer les relations entretenues entre les groupes ou constituants. Bien que ce niveau soit représenté dans les modèles syntaxiques théoriques, il ne l'est pas toujours lors de l'étiquetage de corpus.
- étiquetage fonctionnel : ce niveau consiste à indiquer quelles fonctions grammaticales ou syntaxiques sont remplies par les groupes (*tête, modificateur*, mais aussi *sujet, verbe, objet*, etc.). Ce niveau d'étiquetage n'est pas toujours utilisé dans les tâches d'annotation syntaxique.
- indications des sous-classifications : ce niveau consiste à fournir des informations supplémentaires qui entrent en ligne de compte dans les phénomènes d'accord, de flexion, etc. (indication du genre, du nombre, du temps, etc.).
- marquage du rang : ce niveau consiste à noter le niveau d'enchâssement des éléments. Il est fortement dépendant du cadre théorique choisi pour représenter syntaxiquement les énoncés.

Parmi ces niveaux, seuls les deux premiers se retrouvent dans toute annotation syntaxique.

3.2.3 Les recommandations formulées

Tout comme pour la transcription de corpus oraux et l'annotation morphologique, les experts européens du groupe EAGLES ont proposé des recommandations en vue d'une « uniformisation » des tâches d'étiquetage syntaxique (cf. Eagles (1996c)). Mais, dans ce cas, ils ont dû tenir compte des particularités propres à l'objet : manque de consensus, diversité des cadres théoriques et des catégories, etc.

Les recommandations formulées reposent sur une distinction entre niveaux et étiquettes. En ce qui concerne le premier point, Eagles recommande d'annoter les corpus de façon à ce que les deux premiers niveaux soient représentés dans les annotations (segmentation en groupes ou constituants et étiquetage des constituants). Pour les étiquettes, une liste des étiquettes recommandées est fournie et doit autant que possible être respectée si une grammaire syntagmatique est utilisée. Voici la liste des catégories retenues :

- Phrase (Sentence)
- Proposition (Clause)
- Syntagme nominal
- Syntagme verbal
- Syntagme adjectival
- Syntagme adverbial
- Syntagme prépositionnel

Bien que ces catégories soient utilisées dans une très grande majorité de cas, elles ne font pas nécessairement référence à la même entité. Le syntagme

verbal ne se définit pas toujours de la même façon dans les grammaires syntagmatiques : dans certains cas, il coïncide avec le seul noyau verbal (verbe, auxiliaire et clitiques), dans d'autres il regroupe le noyau verbal et les arguments du verbe, etc. Ces recommandations peuvent donc être interprétées de plusieurs façons.

Pour le français, Abeillé et al (2000 et 2001) ont également proposé un cadre pour construire un corpus arboré. Les choix effectués sont intéressants et devraient pouvoir être utilisés même pour des corpus oraux (ce dernier point étant cependant à tester). En voici les grandes lignes :

- Etiquettes choisies : les étiquettes « syntagmatiques » retenues sont données dans le tableau ci-après.

AP	Syntagme adjectival
AdP	Syntagme adverbial
COORD	Syntagme (ou phrase) coordonné(e)
NP	Syntagme nominal
VN	Noyau verbal (verbes avec clitiques, auxiliaire, modaux, faire, etc.)
PP	Syntagme prépositionnel
SENT	Phrase indépendante (sont aussi notés ainsi tout fragment indépendants isolés par des ponctuations fortes ou des sauts de lignes)
VPpart	Proposition participiale
VPinf	Proposition infinitive (qui peut commencer par une préposition)
Srel	Proposition relative
Ssub	Proposition subordonnée (complétive, interrogative indirecte, etc.)
Sint	Proposition conjuguée interne (incise, discours direct, etc.)

- principes de découpage retenus : afin de résoudre certaines difficultés syntaxiques, plusieurs choix ont été formulés, parmi lesquels :
 - o l'absence de constituants discontinus ;
 - o la possibilité d'avoir des syntagme sans tête (les catégories vides ne sont donc pas utilisées) ;
 - o le refus des ambiguïtés résiduelles ;
 - o la limitation du nombre de syntagme unaire, c'est à dire composé d'un unique élément.

3.3 Annotations prosodiques

Nous allons maintenant nous intéresser à l'annotation prosodique des corpus. Alors que les tâches d'annotation morpho-syntaxique ou syntaxique s'effectuent à partir de la transcription orthographique des documents oraux, l'annotation prosodique repose sur une écoute, et parfois une étude phonétique, du signal. Dans un premier temps, nous allons expliquer ce qu'il faut entendre par annotation prosodique. Puis, nous présenterons plusieurs types d'approches utilisées pour annoter prosodiquement les corpus. Pour finir, nous mentionnerons certains outils qui peuvent aider le linguiste dans son travail d'annotation.

3.3.1 Annotation prosodique : définition et niveaux

Effectuer une annotation prosodique consiste à fournir une représentation symbolique et discrète de la prosodie observée dans un document sonore. Pour y parvenir, il faut donc savoir clairement ce qu'il faut entendre par prosodie. Or cela n'est pas si simple : une description prosodique peut reposer à la fois sur une définition des unités ou constituants prosodiques (syllabe, pied, syntagme phonologique, groupe accentuel, syntagme intonatif, etc.) et sur une analyse de phénomènes tels que le rythme, la mélodie, l'accentuation, etc.

En outre, la prosodie peut être appréhendée à différents niveaux (phonologique ou phonétique) et de plusieurs façons :

- en étudiant l'évolution des corrélats acoustiques associés aux phénomènes prosodiques (intensité, durée et fréquence fondamentale) ;
- en formalisant les rapports entre les unités prosodiques et les unités linguistiques et discursives (paragraphe, tours de parole, énoncés, topiques de discours, constituants syntaxiques).

D'une façon générale, les informations qui sont encodées dans les annotations prosodiques, et les niveaux requis pour effectuer la tâche d'annotation sont déterminés en fonction des objectifs visés :

- si le but est simplement d'avoir quelques indications prosodiques pour mener un travail sur la pragmatique du discours, il peut être suffisant de noter les découpages en groupes intonatifs, les pauses et les contours terminaux des groupes ;
- si, en revanche, le but est d'utiliser l'annotation prosodique pour entraîner des systèmes de reconnaissance de façon à ce qu'ils puissent fournir des découpages en groupes accentuels, des informations plus précises et reposant sur le niveau acoustique-phonétique sont nécessaires.

Avant d'aborder la tâche d'annotation proprement dite, il est donc important de déterminer quels informations ou phénomènes prosodiques doivent être encodés et quels niveaux doivent être utilisés.

3.3.2 Approches utilisées

Comme nous l'avons mentionné, annoter prosodiquement consiste à représenter symboliquement et de façon discrète les phénomènes prosodiques, voire les unités. Plusieurs systèmes de notation (ou de transcription) ont été développés. Ils peuvent être classés en deux catégories que nous allons présenter. Nous exposerons les caractéristiques essentielles de chaque type, en nous appuyant sur des systèmes particuliers. Comme notre exposé restera très limité, nous conseillons au lecteur intéressé de consulter, entre autres, Hirst et al. (1998a), Campione et al. (2000), Gibbon et al. (1998) et Klein et al. (1998). Voici donc ces deux catégories :

- les approches phonologiques ou top-down : l'annotation prosodique d'un corpus peut se faire à partir d'une écoute du signal. Durant cette tâche, le linguiste détermine les unités pertinentes (syllabes proéminentes, groupes accentuels et intonatifs, etc.) et tente de les « classer » en fonction de leur forme (types de contours associés, etc.). Ensuite, des représentations symboliques sont utilisées pour encoder ce qui a été observé. Ces représentations s'appuient généralement sur des analyses phonologiques ou linguistiques préalables qui ont permis de faire un inventaire des tons ou mouvements mélodiques et de proposer une représentation globale des phénomènes prosodiques (définition de tires, types de constituants, etc.). Ce type de systèmes d'annotation propose donc une représentation symbolique de niveau phonologique, sans nécessairement recourir à une analyse acoustique-phonétique. Parmi ces systèmes, nous pouvons mentionner le système anglais ou le système TOBI (« *Tone and Break Index* »). Leurs caractéristiques essentielles sont synthétisées dans le tableau ci-dessous :

Nom	Définition des unités	Phénomènes étudiés	Niveau d'alignement	Principes
« British School »	<u>Unités tonales</u> : - Majeure - Mineure	<i>Intonation</i> : les tons ou intonèmes sont décrits à partir de leur niveau et de la forme du contour.	Repose sur une segmentation en syllabes, les accents, les syllabes accentuées et les découpages en unités tonales étant notés.	Repose sur une détermination de l'accent nucléaire, des autres accents et des mouvements mélodiques qui leur sont associés.
TOBI	<u>Unités</u> : clitiques, mots lexicaux, syntagmes intermédiaires et syntagmes intonatifs.	<i>Intonation et accentuation</i> : Sont notés les accents de groupe, les « pitch accents », les tons frontières et les downsteps.	Les codages sont alignés sur les accents et les frontières de groupe.	S'inspire de la théorie métrique auto-segmentale et de l'étude de l'intonation faite par J. Pierrehumbert.

- les approches acoustiques-phonétiques ou bottom-up :

Certains systèmes d'annotation prosodique repose sur une stylisation de la courbe de fréquence fondamentale, et donc sur le niveau phonétique. Dans ce cas, la méthode utilisée n'est pas spécifique à une langue donnée. Parmi les systèmes de ce type, nous pouvons mentionner le modèle INTSINT développé à Aix en Provence (cf., entre autres, Hirst et al. (1993, 1994 et 1998b)) et le modèle d'IPO. Leurs caractéristiques sont présentées dans le tableau ci-dessous.

Nom	Définition des unités	Phénomènes étudiés	Niveau d'alignement	Principes
IPO	implicite	<i>Intonation</i> : Description des mouvements mélodiques, en tenant compte de leur direction, leur ampleur et leur timing, etc.	L'alignement se fait à partir des accents et des groupes prosodiques. S'appuie sur le signal et la segmentation syllabique	Repose sur une stylisation de F0 sous forme de droites, et sur une recherche de patrons ou mouvements récurrents, spécifiques à une langue.
INTSINT	implicite	<i>Intonation</i> : Transcription de la courbe de F0 à partir d'une détermination de points cibles, classés en fonction de leur réalisation (registres)	L'alignement s'effectue sur le signal.	Repose sur une stylisation par fonctions splines à partir des points cibles, et sur un inventaire de tons et une représentation des registres (hauteur relative vs. hauteur absolue).

Comme nous le voyons, il existe donc plusieurs systèmes capables de représenter la prosodie associée à une énoncé. Le choix s'effectue donc essentiellement en fonction des objectifs visés.

3.3.3 Vers une automatisation de la tâche

Nous avons vu précédemment que certaines tâches d'annotation peuvent être effectuées automatiquement : l'étiquetage morphologique à l'aide de « taggers » et l'étiquetage syntaxique à l'aide d'analyseurs. Pour l'annotation prosodique, il existe également quelques outils qui permettent d'automatiser (ou de « semi-automatiser ») la tâche. Ils offrent d'ailleurs certains avantages par rapport à l'annotation manuelle : plus de rapidité, d'exactitude, de précision et de cohérence (cf., entre autres, Campione et al. (2000 et à paraître) et Véronis (2000)). Cela s'explique par le fait que ces outils s'appuient essentiellement sur une analyse mathématique du signal pour déterminer les points cibles, calculer les registres et assigner les étiquettes.

Parmi ces outils, deux semblent particulièrement intéressants à utiliser :

- MOMEL et INTSINT (cf., entre autres, Campione et al. (2000)): cet outil s'appuie sur les algorithmes de stylisation de F0 développés à Aix en Provence (cf. Hirst et al. (1993 et 1994)) et sur le codage symbolique INTSINT. Il offre des résultats intéressants, mais nous regrettons qu'il ne fonctionne que sous UNIX ;
- PIT ou « Perception-Based Intonation Transcription) : ce modèle est distinct du précédent dans la mesure où il repose sur une stylisation de F0 sur base perceptive (cf., entre autres, Mertens et al. (1995 et 1997)). Il fonctionne à partir d'une segmentation syllabique du signal et d'une détermination des noyaux vocaliques. Il peut être utilisé sous PRAAT (cf. Delais-Roussarie et al. (ce volume)). Des informations sur cet outil peuvent être obtenues à l'adresse suivante : <http://bach.arts.kuleuven.ac.be/pmertens/prosody/pit/>

Conclusion

En conclusion, la création d'un corpus oral ou écrit repose sur une démarche méthodologique rigoureuse. Nous avons tenté d'en présenter certains aspects en insistant particulièrement sur deux phases :

- la phase préparatoire qui précède la saisie des données : durant celle-ci doivent être déterminés les types de corpus et de documents que l'on souhaite obtenir, les objectifs de recherches que l'on se fixe, la taille envisagée. A partir de ces différents éléments, il est possible de développer un protocole qui peut par la suite être joint au corpus pour le documenter.
- la phase de transcription et d'annotation : celle-ci est essentielle pour faire d'un ensemble de données saisies un corpus linguistique exploitable. Pour les corpus oraux, une attention particulière doit être accordée aux tâches de segmentation et de transcription, dans la mesure où les conventions et principes retenus imposent des restrictions sur les utilisations ultérieures. Pour les annotations linguistiques (syntaxiques, morphologiques et prosodiques), il est également important de bien choisir les étiquettes et la méthode utilisées. Dans bien des cas ces choix seront en partie guidés par les objectifs de recherche visés.

Dans l'ensemble de ce document, nous avons prêté une attention particulière aux corpus oraux, et cela pour deux raisons : i.) pour le français, il existe peu de corpus oraux exploitables à ce jour, ii.) peu d'ouvrages sont consacrés aux problèmes méthodologiques rencontrés lors de la création de tels corpus. En outre, nous avons essayé de présenter clairement les différentes recommandations européennes en matière de constitution et d'annotation de corpus oraux et de parole. Les suivre nous paraît essentiel si nous souhaitons garantir la diffusion et la réutilisation des corpus.

Références bibliographiques

Aarts, J., P. de Haan et N. Oostdijk (eds) (1993) *English Language Corpora : Design, Analysis and Exploitation*. Papers présentés à la 13th International Conference on English Language Research on Computerized Corpora. Nimègue, 1992. Rodopi.

Abeillé, A., L. Clément et A. Kinyon (2000) "Building a treebank for French", *Actes 2d Conférence on Linguistic Resources*, LREC, Athènes.

Abeillé, A., L. Clément et A. Kinyon (2001) "Building a treebank for French", in A. Abeillé (ed) *Treebanks*, Kluwer, Dordrecht.

Aston, G & L. Burnard (1998) *The BNC Handbook : Exploring the British National Corpus with SARA*. Edinburgh University Press. 256 pages

Bilger, M. (2000) « *Petite typologie des conventions de transcription de l'oral : aspects pratiques et théoriques.* » Manuscrit non publié. Université de Perpignan. 12 pages.

Blanche-Benveniste, C. & Jeanjean, C. (1987) *Le français parlé : transcription et édition*. Didier Erudition. Paris. 264 pages

Blanche-Benveniste, C.- Bilger, M.- Rouget, Ch.- van den Eynde, K. (1991) *Le français parlé : Études grammaticales*. Paris: Editions du Centre National de la Recherche Scientifique (collection Sciences du Langage)

Burnage, G et D. Dunlop (1993) "Encoding the British National Corpus" in J. Aart, P. de Haan et N. Oostdijk (eds). *Proceedings of the 13th International Conference on English Language Research on Computerized Corpora*. Nimègue, 1992. pp. 79-95

Burnard, L. (1995) "The Text Encoding Initiative : an overview" in G. Leech, G. Myers and J. Thomas (eds). Pp.69-81

Burnard, L. (1993) "The Text Encoding Initiative : a further report" in J. Aart, P. de Haan et N. Oostdijk (eds). *Proceedings of the 13th International Conference on English Language Research on Computerized Corpora*. Nimègue, 1992. pp. 37-45

Campione, E., & Véronis, J. (à paraître) "Semi-automatic tagging of intonation". In G. Sampson & D. McCarthy (Eds.), *Readings in Corpus Linguistics*. London: Continuum

Campione, E., & Véronis, J. (2001) « Etiquetage prosodique semi-automatique des corpus oraux » in *Actes de la Conférence Traitement Automatique des Langues (TALN'2001)* (pp. 123-132). Tours (France): ATALA.

Campione, E., & Véronis, J. (2001) « Semi-automatic tagging of intonation in French spoken corpora ». In P. Rayson, A. Wilson, T. McEnery, A. Hardie & S. Khoja (Eds.), *Proceedings of the Corpus Linguistics'2001 Conference* (pp. 90-99). Lancaster, U.K.: Lancaster University, UCREL.

Campione, E., Hirst, D., & Véronis, J. (2000) « Stylisation and symbolic coding of F0 : comparison of five models ». In A. Botinis (Ed.), *Intonation: Models and Theories* (pp. 185-208). Dordrecht: Kluwer Academic Publishers.

Crowdy, S. (1995) "The BNC spoken corpus". In G. Leech, G. Myers and J. Thomas (eds). Pp.224-242

Delais-Roussarie, E. (2002) « Protocole (I) : Enregistrement des données ». Ms non publié rédigé dans le cadre du projet d'ACI « jeune chercheur » n° 67031

Delais-Roussarie, E. (2002) « Protocole (II) : Annotation des données ». Ms non publié rédigé dans le cadre du projet d'ACI « jeune chercheur » n° 67031

Durand, J. ; B. Laks et C. Lyche (ce volume) « Méthodologie de l'enquête *Phonologie du français contemporain* » in E. Delais, J Durand & L. Labruno (eds). Ce volume.

French, JP (1992) *Notes and conventions for soundprint transcribers*. Unpublished papers. JP French associates. Cited by Payne (1995)

Dybkjær, L., Bernsen, N.O., Dybkjær, H., McKelvie, D. and Mengel, A. (1998) "The MATE Markup Framework". [MATE Deliverable D1.2](#), November 1998.

EAGLES (1996a) *Recommendations on corpus encoding. EAG-TCWG-CES/R-F*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

EAGLES (1996b) *Recommendations for the morphosyntactic annotation of corpora. EAG-TCWG-MAC/R*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

EAGLES (1996c) *Preliminary recommendations for the syntactic annotation of corpora. EAG-TCWG-SASG1/P-B*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

EAGLES (1996d) *Preliminary recommendations on corpus typology. EAG-TCWG-CTYP/P*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

EAGLES (1996e) *Preliminary recommendations on spoken texts. EAG-TCWG-SPT/P*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

EAGLES (1996f) *Preliminary recommendations on text typology. EAG-TCWG-TTYP/P*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

Eyes, E. et G. Leech (1993) "Progress in UCREL research: Improving corpus annotation" in J. Aart, P. de Haan et N. Oostdijk (eds). Proceedings of the 13th International Conference on English Language Research on Computerized Corpora. Nimègue, 1992. pp. 123-143

Gibbon, D. , R. Moore et R. Winski (eds).(1998) *Handbook of Standards and Resources for Spoken Language Systems*, Vol. 1: "Spoken Language System and Corpus Design", XV, 216 p.

Gibbon, D. , R. Moore et R. Winski (eds).(1998) *Handbook of Standards and Resources for Spoken Language Systems*, Vol 4: “Spoken Language Reference Materials”, XVI, 242 p.

Habert, B., C. Fabre et F. Issac (1997a) *De l'écrit au numérique : Constituer, normaliser et exploiter les corpus électroniques*. Masson, InterEditions. 320 pages.

Habert, B. , A. Nazarenko et A. Salem (1997b) *Les linguistiques de corpus*. Collection U, série « Linguistique ». Armand Colin. Paris. 240 pages

Hirst, D.J., Di Cristo, A. (1998a) “A survey of intonation systems”. In Hirst, D.J., Di Cristo, A. (Eds), *Intonation Systems: a Survey of Twenty Languages*. Cambridge : Cambridge University Press, pp. 1-44

Hirst, D.J., Di Cristo, A., Espesser, R. (1998b) “Levels of representation and levels of analysis for the description of intonation systems”. In Horne, M. (Ed.), *Prosody: Theory and Experiment*, Dordrecht: Kluwer Academic Publishers.

Hirst, D.J., N. Ide, et J Véronis (1994) “Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTTEXT project”, *Proceedings of the ESCA/IEEE Workshop on Speech Synthesis*, New York, September 1994.

Hirst, D.J., Espesser, R. (1993) “Automatic Modelling of Fundamental Frequency using a quadratic spline function”. *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, pp. 15, 75-85.

Johansson, S. (1995) “The approach of the Text Encoding Initiative to the encoding of spoken discourse” in G. Leech, G. Myers and J. Thomas (eds). Pp. 82-98

Klein, M., Bernsen, N.O., Davies, S., Dybkjær, L., Garrido, J., Kasch, H., Mengel, A., Pirrelli, V., Poesio, M., Quazza, S. et Soria, S. (1998) “Supported Coding Schemes”. [MATE Deliverable D1.1](#), July 1998

Knowles, G. (1993a) “From text to waveform : converting the Lancaster/IBM Spoken English Corpus into a speech database” in J. Aart, P. de Haan et N. Oostdijk (eds). *Proceedings of the 13th International Conference on English Language Research on Computerized Corpora*. Nimègue, 1992. pp. 47-58

Knowles, G. (1993b) “The machine-readable Spoken English Corpus” in J. Aart, P. de Haan et N. Oostdijk (eds). *Proceedings of the 13th International Conference on English Language Research on Computerized Corpora*. Nimègue, 1992. pp. 107-119

Leech, G., G. Myers et J. Thomas (Eds.) (1995) *Spoken English on Computer: Transcription, Markup and Applications*. Harlow: Longman.

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

Mertens, P., Beaugendre, F. et Alessandro, Ch. d' (1997) Comparing approaches to pitch contour stylization for speech synthesis. in Santen, J.P.H.

van Sproat Richard W., Olive, Joseph P. et Hirschberg, Julia (eds) (1997) *Progress in Speech Synthesis*. p. 347-363. N.Y.: Springer Verlag

Payne, J. (1995) "The COBUILD spoken corpus : transcription conventions" In G. Leech, G. Myers and J. Thomas (eds).

Peppé. S. (1995) "The Survey of English Usage and the London-Lund Corpus : computerising manual prosodic transcription" in G. Leech, G. Myers and J. Thomas (eds). Pp.187-207

Roach, P et S. Arnfield (1995) "Linking prosodic transcription to the time dimension". In G. Leech, G. Myers and J. Thomas (eds). Pp.149-160

Sinclair, J. (1995) "From Theory to Practice". In G. Leech, G. Myers and J. Thomas (eds). Pp.99-109

Sperberg-McQueen, C.M. et L. Burnard (eds.) (1999). *Guidelines for Electronic Text Encoding and Interchange*. TEI P3 Text Encoding Initiative. Revised reprint: Oxford May 1999

Sperberg-McQueen, C.M et Lou Burnard (eds) *TEI Lite: An introduction to Text Encoding for Interchange* (TEI Tutorial, fait pour répondre au besoin de 90% de la communauté participant au TEI dans 90% des cas.). Téléchargeable à partir de www.hcu.ox.ac.uk/TEI/Lite.

Valli, A. et Véronis, J. (2000) "Grammatical tagging of spoken corpora: an experiment" Paper presented at the Workshop "New methods and formalisms for corpus linguistics", Aix-en-Provence. Octobre 2000

Valli, A. et Véronis, J. (1999) « Etiquetage grammatical de corpus oraux: problèmes et perspectives ». *Revue Française de Linguistique Appliquée*, IV(2), 113-133

Véronis, J. (2000) « Annotation automatique de corpus : panorama et état de la technique ». In J.-M. Pierrel (Ed.), *Ingénierie des langues* (pp. 111-129). Paris: Editions Hermès

Véronis, J. (1998) « Annotation automatique de corpus: état de l'art », Colloque International " *Questions de méthode dans la linguistique de corpus*" (pp. 7-9). Perpignan (France).

Annexes

Grille de Biber

Les paramètres proposés par Biber sont repris dans Habert et al. (1997a) :

1. « canal : écrit/parlé/écrit lu
2. format : publié/non publié
3. cadre : institutionnel/autre cadre public/privé-interpersonnel
4. destinataire :
 - a. pluralité : non compté/pluriel/individuel/soi-même
 - b. présence : présence/absence
 - c. interaction : aucune/peu/beaucoup
 - d. connaissances partagées : générales/ spécialisées/ personnelles
5. destinateur :
 - a. variation démographique : sexe, âge, (...)
 - b. statut : individu/institution
6. factualité : informatif-factuel/intermédiaire/ imaginaire
7. objectifs : persuader, amuser, édifier, etc.
8. thèmes »

Habert et al. (1997a), pp. 43-44

Grille élaborée par Sinclair et repris par Eagles

Voici le détail de cette grille tel qu'elle est présentée dans Habert et al. (1997a) :

« - Caractéristiques externes :

1. Origine :
 - a. personnes impliquées : auteur, éditeur, traducteur, (...) avec lorsque cela est pertinent l'âge, le sexe, la ou les langues maîtrisées, le domicile ;
 - b. contexte
 - c. datation
2. Etat :
 - a. Mode de transmission :
 - i. Ecrits : imprimés(...), manuscrits.
 - ii. Oral : enregistrement à l'insu ou non des participants, lieu d'enregistrement (studio ou sur le terrain)

- b. données textuelles non liées : diagrammes, figures, etc.
3. Objectifs :
- a. Public visé, immédiat et plus lointain : taille, profil, relation auteur/public, etc.
 - b. Effet visé : information, discussion, recommandation, etc.(...)
- Caractéristiques internes :
1. Thème
 2. Style : les « niveaux de langue » peuvent être distingués (...). S'ajoutent le degré de préparation et l'interaction avec le public. »

Habert et al. (1997a), p. 44-45

Recommandations d'EAGLES

Voici la traduction de certains passages de la liste de recommandations proposées dans Gibbon (1998) :

« Liste de recommandations :

1. Pour la transcription de dialogues entre plus de deux locuteurs, utiliser une notation sous forme de partition musicale.
2. Pour les transcriptions orthographiques, utiliser autant que possible l'orthographe standard.
3. Indiquer les formes réduites des mots dans la transcription orthographique si a) ces formes apparaissent fréquemment, et b) si elles contiennent des effacements de syllabes.
4. Utiliser au moins deux types de syllabes « fillers » : une correspondant à une voyelle (*euħ*), et une correspondant à une nasale (*um*).
5. Les événements acoustiques non verbaux doivent être notés au bon endroit dans l'énoncé. Pour ce faire, transcrire d'abord les mots, puis indiquer quels mots sont prononcés en même temps que l'événement acoustique.
6. Si la transcription d'un corpus est faite orthographiquement, il est recommandé de générer une liste des mots utilisés à partir de la transcription. La forme orthographique des mots pourra alors être convertie sous forme phonémique par un système de conversion graphème-phonème. (...)
7. (...)
8. Il est recommandé aux transpositeurs de fournir des informations sur la tâche de transcription qu'ils avaient à effectuer et sur le fichier sonore qu'ils utilisaient. Certains locuteurs sont plus faciles à transcrire que d'autres. Cela dépend du débit, de la façon d'articuler, du nombre

d'hésitations, et du nombre de formes dialectales utilisés par les locuteurs.

9. Si la transcription est faite à plus d'un niveau (c'est à dire qu'elle contient une transcription orthographique avec des annotations prosodiques et l'indication d'hésitations), il est recommandé de ne travailler que sur un niveau à fois. Dans la vie de tous les jours, les auditeurs sont habitués à ignorer les hésitations, les faux départs, et autres imperfections, et aussi à ne pas prendre explicitement en compte la prosodie. Les transcrip-teurs doivent apprendre à entendre tous ces éléments. Il semble plus facile d'écouter les mots et de les transcrire dans un premier temps, puis d'effectuer les autres annotations.
10. Pour effectuer des transcriptions orthographiques, il n'est pas nécessaire de recourir à des transcrip-teurs expérimentés. En revanche, pour les transcriptions phonémiques ou phonétiques, il faut travailler avec des transcrip-teurs habitués à écouter le signal sonore de façon très précise et analytique.
11. Afin de fournir quelques indications sur le temps nécessaire pour effectuer une transcription, voici quelques exemples. Le temps nécessaire pour effectuer la transcription orthographique d'un document en parole spontanée est en moyenne équivalent à dix fois le temps réel du document. Le temps nécessaire pour effectuer la transcription orthographique d'un document composé de phrases lues est d'environ trois fois la durée du document, et pour la transcription orthographique d'un texte, il faut multiplier la durée du document par cinq.
12. Il faut toujours contrôler une transcription. Cela peut se faire de plusieurs façons.(...)
13. Pour sauvegarder le fichier de transcription, il est préférable d'utiliser un format compatible avec WAVES, cela permettant un meilleur échange de données.
14. (...)
15. (...)
16. (...)

Traduit de Gibbon et al. (1998, pp. 170-72)

Quelques outils d'aide à la transcription et à l'annotation de données audio pour constituer des corpus oraux

Introduction

Comme il a été mentionné au chapitre précédent, la demande pour des corpus oraux et des corpus de parole est très importante, et cela dans de nombreux domaines : traitement automatique de la parole (synthèse de la parole, reconnaissance automatique de la parole, identification de locuteurs, etc.), dialogue homme-machine, linguistique, didactique des langues, etc. Pour constituer et échanger de tels corpus ou collections de données, il ne suffit pas d'effectuer des enregistrements et de les numériser sous un format standardisé ; il est également nécessaire de segmenter le signal sonore et de transcrire au moins orthographiquement les informations qu'il contient (cf. chap. précédent). Bien que les tâches de segmentation et de transcription restent très lourdes, elles sont rendues plus aisées grâce au développement d'outils d'aide à la transcription.

L'objectif de ce chapitre est de présenter plusieurs de ces outils, en insistant à la fois sur les modalités de fonctionnement de chacun d'entre eux et sur leurs limites. Le nombre des outils de ce type étant important et sans cesse en évolution, nous ne pouvons pas viser à l'exhaustivité. Nous avons donc sélectionné plusieurs outils en nous appuyant sur les critères suivants :

- **fonctionnalités** : les outils présentés ici sont tous des outils d'aide à la transcription orthographique du signal sonore, si possible alignée.
- **accessibilité et coût des outils** : seuls des outils gratuits et en accès libre sur Internet sont décrits ;
- **plate-formes requises** : seuls les programmes fonctionnant sous Windows ont été pris en compte ;

- **interface, installation et utilisation** : les outils présentés peuvent être installés et utilisés facilement par des utilisateurs qui ne soient pas des informaticiens confirmés⁷. Cette facilité résulte généralement de la présence de plusieurs éléments : procédure d'installation facile et bien documentée, bonne documentation pour prendre en main le programme, etc.

Comme nous ne présentons que sept outils dans ce travail, nous renvoyons les lecteurs intéressés à différents travaux ou documents dans lesquels il est possible d'obtenir des informations et des évaluations complémentaires sur ce type d'outils (cf., entre autres, Isard et al (1998) ; Dybkjær et al. (2001) ; site Web : <http://www ldc.upenn.edu/annotation/>).

Dans un premier temps, nous décrivons deux outils d'aide à la transcription sans éditeur intégré. Ensuite, nous présentons deux outils avec éditeurs intégrés. Pour finir, trois outils permettant de faire non seulement des transcriptions alignées mais également des analyses et des requêtes sur les données transcrites sont présentés.

1. Outils d'aide à la transcription sans éditeur intégré

Les outils présentés dans cette section permettent une écoute « répétitive » des données qui peut être paramétrée de façon à faciliter la transcription. Mais la transcription et l'alignement sur le signal ne sont pas effectués directement dans un éditeur intégré au logiciel, mais peuvent l'être dans n'importe quel éditeur choisi par l'utilisateur : traitement de texte, éditeur SGML ou XML, etc.

1.1 SoundScriber

SoundScriber est un logiciel d'aide à la transcription fonctionnant sous Windows et développé par Eric Breck. Il peut être téléchargé à partir du site suivant : <http://www.lsa.umich.edu/eli/micase/micase.htm>. Il a été développé et utilisé dans le cadre du projet de constitution de corpus « MICASE » qui avait pour double objectif :

- l'étude des caractéristiques du parler utilisé dans le monde universitaire (sa grammaire, son vocabulaire, ses fonctions, etc.) ;
- l'étude des spécificités du parler académique en fonction des disciplines et sous-disciplines.

Le but du projet, lancé en 1997 par l'institut de langue anglaise de l'Université du Michigan (ELI : English Language Institute), était de collecter et de transcrire au moins 200 heures de parole (soit 1,7 millions de mots) dans le monde universitaire. En Avril 2002, l'ensemble des données collectées (200

⁷ Seuls les logiciels que nous avons réussi à installer sans l'aide d'un technicien informatique et avec lesquels nous avons pu effectuer la transcription intégrale d'un fichier son d'une durée de 5 minutes sont présentés.

heures) ont été transcrites orthographiquement et les transcriptions ont été revues et corrigés si nécessaire.

A l'ouverture du programme, la fenêtre suivante (cf. fig. 1) apparaît à l'écran :

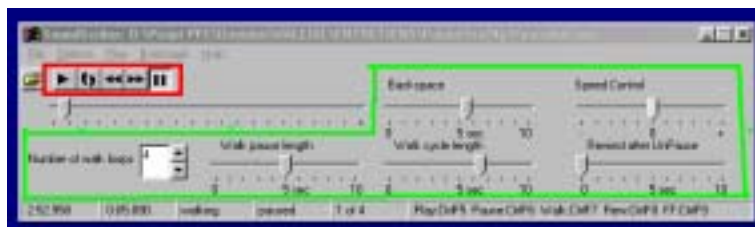


Fig. 1 : SoundScriber

Le menu déroulant [File] en haut à gauche permet d'ouvrir le fichier audio dont la transcription est à faire. Plusieurs formats audio sont supportés : WAV, AVI, AIFF, MP3, etc. Les touches dans le cadre rouge remplissent plusieurs fonctions, dont :

- les fonctions qui se retrouvent sur tous les magnétophones, à savoir : FF, RW, PLAY et PAUSE ;
- une touche représentant deux empreintes de pas et appelant une fonction spécialement conçue pour la transcription de données sonores : « WALK ». Cette fonction permet de jouer plusieurs fois un court extrait du fichier audio et d'avancer ensuite vers un autre passage qui chevauche le premier extrait. Un exemple précis de la façon dont cela se réalise est donné en [1].

[1] Réalisation de la fonction « WALK »

Soit le passage suivant :

OB: Bien c'est à dire bien oui ce souvenir c'est c'est les Allemands qui étaient à Puy de Fourche à côté. C'était pas loin. Ils sont venus même juste dans les villages à côté chercher des familles juives. (on) s'en rappelle. Et puis bon bien on (n') a peut-être moins souffert que dans les villes mais enfin on on avait quand même ressenti le danger de la guerre.

Avec la fonction « WALK », ce passage peut être joué de la façon suivante :

Bien c'est à dire bien oui ce souvenir
Bien c'est à dire bien oui ce souvenir
Bien c'est à dire bien oui ce souvenir
Bien c'est à dire bien oui ce souvenir
bien oui ce souvenir c'est c'est les Allemands
bien oui ce souvenir c'est c'est les Allemands
bien oui ce souvenir c'est c'est les Allemands

bien oui ce souvenir c'est c'est les Allemands
les Allemands qui étaient à Puy de Fourche
les Allemands qui étaient à Puy de Fourche
les Allemands qui étaient à Puy de Fourche
les Allemands qui étaient à Puy de Fourche
etc.

Les modalités de fonctionnement de WALK peuvent être paramétrées à l'aide des différents curseurs présents dans le cadre vert sur la figure. La paramétrisation permet de modifier :

- le nombre de répétitions ;
- la durée du passage joué ;
- la durée de la pause entre chaque répétition ;
- la durée du chevauchement lors du passage d'un extrait à l'autre,
- la vitesse de déroulement du fichier audio.

1.2 VoiceWalker

VoiceWalker est un logiciel d'aide à la transcription développé par Jack Du Bois de UCSB (University of California, Santa Barbara) dans le cadre du projet CSAE. Ce projet a pour objectif de constituer un corpus d'américain parlé appelé « the Santa Barbara Corpus of Spoken American English ». Les données ont été collectées sur la totalité du territoire américain et correspondent à des enregistrements de locuteurs parlant dans des situations diverses et sur des sujets variés.

Le programme peut être téléchargé à partir du site suivant : <http://www.linguistics.ucsb.edu/resources/resources.htm>.⁸ Il fonctionne de la même façon que SoundScriber, mais l'interface graphique est sensiblement différente.

A l'ouverture du logiciel apparaît à l'écran la fenêtre suivante (cf. fig. 2) :



Fig. 2 : VoiceWalker

Plusieurs fonctionnalités sont accessibles à partir des menus déroulants présents dans le cadre en bleu, à savoir :

⁸ Suite à une modification récente du site du département de linguistique de l'université de Santa Barbara (UCSB), certains liens sont erronés et rendent difficile l'accès aux programmes d'aide à la transcription VoiceWalker et SoundWalker (cf. § 3.1).

- l'ouverture d'un fichier audio (menu [File]) ;
- la lecture du fichier audio ouvert (menu [Play]) ;
- la sélection du nombre de répétitions (menu [Looping]) ;
- la paramétrisation de la fonction « Walk » s'effectue à partir du menu [Tools], en sélectionnant la touche [options] ;

Outre les touches présentes dans tout magnétophone (cf. dans le cadre jaune), une série de touches sont spécifiques à la fonction « Walk » et se situent dans le cadre rouge. Elles servent à :

- lancer la lecture en mode «Walk » ;
- rejouer le passage précédent ;
- jouer le passage courant ;
- jouer le passage suivant.

Par ailleurs, les informations temporelles sont données dans le cadre en vert.

Comme indiqué précédemment, la fonction « Walk » est paramétrable à partir du menu [Options]. S'ouvre alors la boîte de dialogue suivante (fig. 3) :

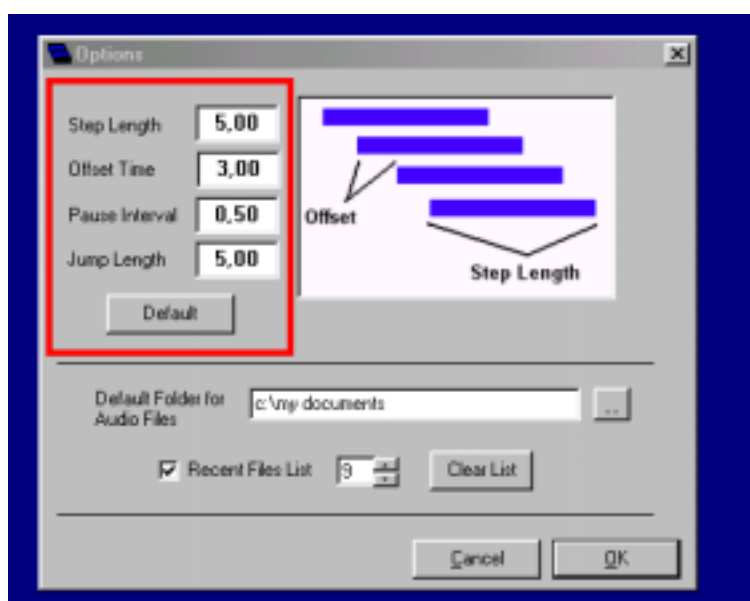


Fig. 3 : Modification des options

La modification des paramètres s'effectue par l'utilisateur en changeant les informations contenues dans le cadre en rouge. Les paramètres suivants peuvent être modifiés :

- la durée du passage ;
- la durée de non chevauchement ;
- la durée de la pause entre chaque répétition ;
- le nombre de répétitions.

1.3 Synthèse

Les logiciels SoundScriber et VoiceWalker sont très simples tant dans leur conception que dans leur fonctionnement, mais peuvent être d'une aide précieuse pour transcrire orthographiquement des données sonores. En outre, l'absence d'éditeur de texte intégré peut être vue comme un avantage : l'utilisateur est libre de choisir les conventions de transcription, les formats de saisies et de sauvegarde des transcriptions, etc. Ainsi, dans le cadre du projet « MICASE », les transcriptions des données ont directement été saisies avec des balises dans un éditeur SGML sous ce format. En revanche, ces outils ne permettent pas aisément d'aligner la transcription sur le signal. Même si des informations temporelles sont fournies dans les interfaces graphiques lors de l'écoute des données audio, elles ne sont pas directement accessibles lors de la saisie des transcriptions.

Ces outils peuvent être très utiles si l'utilisateur souhaite transcrire orthographiquement des fichiers audio et travailler ensuite sur ces seules transcriptions. En revanche, ils sont moins intéressants dès que l'objectif est de fournir des transcriptions alignées sur le signal.

2. Outils d'aide à la transcription avec éditeurs intégrés et alignement sur le signal

Dans ce chapitre vont être présentés deux outils d'aide à la transcription qui permettent d'effectuer assez facilement des transcriptions alignées sur le signal. Dans ces outils, l'alignement et la saisie de la transcription orthographique se font dans un éditeur intégré au logiciel.

2.1 Soundwalker

Soundwalker, tout comme VoiceWalker, est un logiciel d'aide à la transcription développé par Jack Du Bois de UCSB. Il a été utilisé pour transcrire les données collectées dans le cadre du projet CSAE (« The Santa Barbara Corpus of Spoken American English »). Il peut être téléchargé à partir du site suivant : <http://www.linguistics.ucsb.edu/resources/resources.htm>.

A l'ouverture du logiciel la fenêtre suivante (cf. fig. 4) apparaît à l'écran :

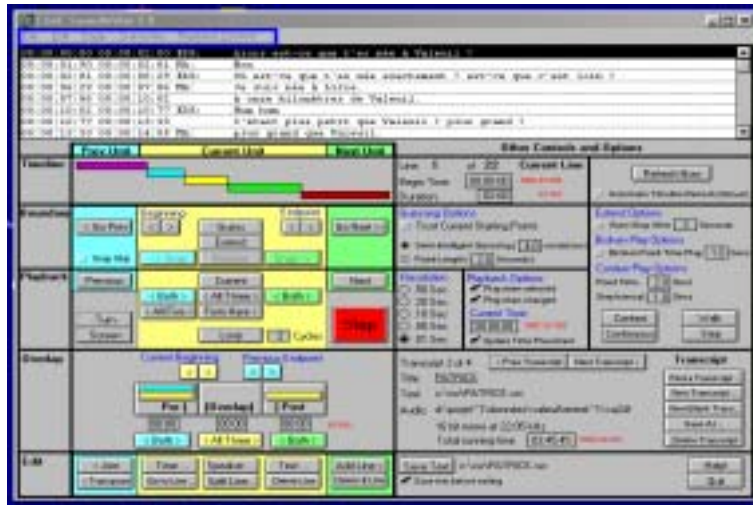


Fig. 4 : CSAE SoundWriter 2

Dans cette fenêtre, plusieurs zones sont présentes. En haut à gauche dans le cadre bleu se trouvent les menus déroulants qui sont utilisés pour :

- imprimer les écrans et les transcriptions (menu [File]) ;
- ouvrir des pages de transcription ou commencer une nouvelle transcription. Dans ce cas, il faut sélectionner [New Blank Transcript] dans le menu [Page]. S'ouvrira alors une boîte de dialogue qui demandera quel fichier WAV doit être associé à cette transcription ;
- configurer les paramètres qui seront utilisés pour jouer le fichier audio : nombre de réitérations ; durée des extraits, etc. (cf. menus [Transcribe] et [Playback Controls]).

Au dessous des menus déroulants se trouve l'éditeur de texte (cf. fig. 5).

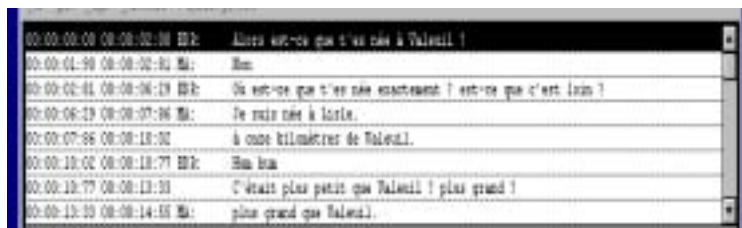


Fig. 5 : Editeur de texte de SoundWalker

Chaque ligne correspond à une unité intonative. Les temps de début et de fin de l'unité sont fournis. Apparaissent ensuite le nom du locuteur et la transcription de ce qui est dit.

Sous l'éditeur se trouvent deux zones différentes :

- sur la droite se trouve une zone dans laquelle plusieurs indications sont fournies et peuvent être modifiées, notamment les paramètres associés à la fonction « Walk », le titre de la transcription, le nom du fichier

texte dans lequel elle est sauvegardée, le nom et les caractéristiques du fichier audio, etc. ;

- sur la gauche se trouve une zone qui peut se diviser en 5 sous-parties (cf. fig. 6).

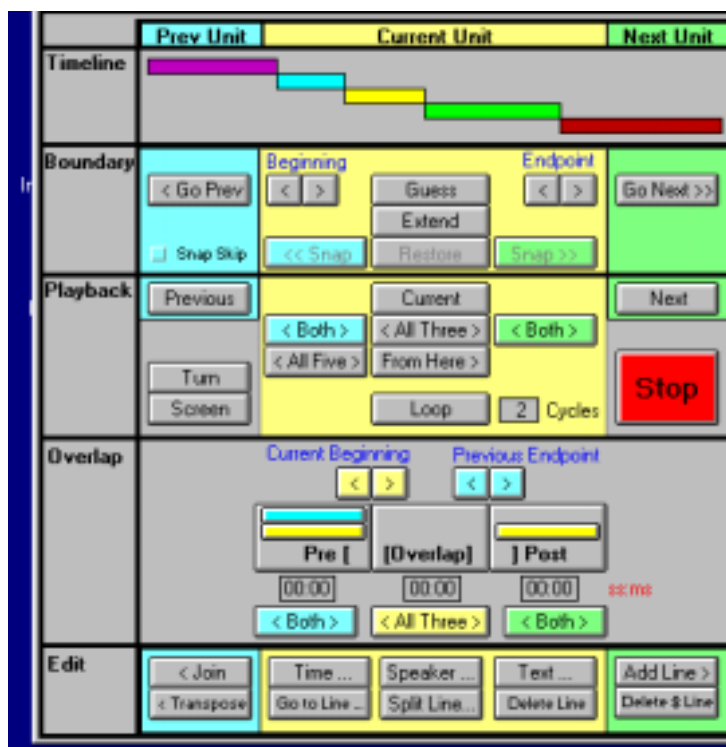


Fig. 6 : Fonctions des sous-parties

Les fonctions associées à chacune des sous-parties sont décrites dans le tableau ci-après :

Timeline	Dans ce cadre est fournie une représentation graphique de cinq unités intonatives consécutives, l'unité active étant représentée par le trait jaune. Les deux unités qui précèdent et les deux qui suivent étant représentées respectivement par les traits bleu et violet, et vert et marron. En cliquant sur ces traits, il est possible de jouer les unités ou intervalles.
Boundary	Cette section est utilisée pour déterminer les unités intonatives et positionner les frontières. Pour déterminer un intervalle, il faut appuyer sur la touche [Extend]. Le fichier audio est alors joué. La frontière droite de l'intervalle sera positionnée en appuyant à nouveau sur cette touche. Il est ensuite possible d'affiner le placement des frontières à l'aide des touches sous « Beginning » et « Endpoint ».
Playback	Cette zone contient les différentes touches permettant de jouer des portions du signal ou des unités intonatives afin, par exemple, de vérifier la qualité des transcriptions.

Overlap	Cette section permet de traiter les chevauchements entre plusieurs locuteurs. Des unités intonatives correspondant à chacun des locuteurs et indiquant les chevauchements peuvent être définies.
Edit	Dans cette partie sont regroupées les différentes touches nécessaires à l'édition et à la saisie de la transcription. Dans l'espace jaune clair se trouvent les touches à utiliser pour effectuer la transcription de l'unité intonative courante (saisie des informations temporelles ; saisie du locuteur ; saisie du texte prononcé, etc. Les touches dans l'espace bleu clair servent à modifier la transcription associée à l'intervalle précédent (concaténation d'unités, modification de l'ordre, etc.). Les deux touches dans l'espace vert clair servent à commencer la transcription des unités suivantes.

SoundWriter est un logiciel d'aide à la transcription très complet où la segmentation du signal audio, la transcription et l'alignement se font sur la base d'un travail en unités intonatives. Bien que le signal ne soit pas visualisable avec ce logiciel, le positionnement des frontières d'unités intonatives et la gestion des chevauchements peuvent se faire de façon très minutieuse grâce aux touches de défilement lent (cf. « Beginning » et « Endpoint » par exemple).

Ceci étant, plusieurs éléments sont à déplorer :

- l'absence de manuel et d'aide rendent l'utilisation du logiciel assez « hasardeuse » ;
- plusieurs « bugs » interdisent la sauvegarde du fichier de transcription dans un format *txt* et sa visualisation dans un programme externe (éditeur de texte, traitement de texte, etc.).

2.2 Transcriber

Transcriber est un logiciel d'aide à l'annotation et à la transcription manuelle de fichiers audio qui a été spécialement conçu pour la transcription orthographique d'émissions radiophoniques d'informations (cf. Barras et al. (1998, 2000a et b)). Il a été développé par Claude Barras et Edouard Geoffroy de la Direction Générale de l'Armement (DGA) et est également soutenu par le LDC (Linguistic Data Consortium). Ce logiciel peut être téléchargé à l'adresse suivante : <http://www ldc.upenn.edu/mirror/Transcriber/>. Il est écrit en C et en langage de script Tcl/Tk, et utilise la librairie Snack de gestion du son⁹. Il peut, de ce fait, gérer plusieurs formats audio, même si les fichiers sont d'une durée importante. Il fonctionne sur plusieurs plate-formes (Unix, Linux, Sun Solaris, Silicon Graphics et Windows).

⁹ Snack est un programme de gestion des fichiers audio développé par Kåre Sjölander et téléchargeable à l'adresse suivante : <http://www.speech.kth.se/snack/>.

A l'ouverture du programme apparaît à l'écran la fenêtre suivante (fig. 7) :



Fig. 7 : ouverture de Transcriber

L'utilisateur doit alors indiquer dans la boîte de dialogue le nom et l'emplacement du fichier audio dont il veut faire la transcription. S'ouvre alors une fenêtre de travail qui contient (cf. fig. 8) :

- une barre de menus déroulants (cf. cadre bleu) ayant plusieurs fonctionnalités décrites dans le tableau ci-après :

[File]	Création et ouverture de fichiers de transcription (et de fichiers audio associés), sauvegarde de la transcription au format XML, exportation de transcriptions dans un autre format (CHILDES, LDC) ;
[Edit]	Fonctions classiques d'édition (copier, coller) et recherche de topiques ou de locuteurs dans le fichier de transcription ouvert ;
[Signal]	Toutes les fonctions nécessaires à l'écoute du signal (pause, écoute, écoute avec pause, avance rapide, etc.) ;
[Segmentation]	Le menu regroupe les touches permettant une segmentation du signal en énoncés ([insert breakpoint]), en tours de parole ([create turn]) et en topiques de discours ([create section]). Les informations relatives aux locuteurs et aux topiques de discours peuvent être saisies dans des boîtes de dialogue accessibles à partir des touches [Edit turn attributes..] et [Edit section attributes] ;
[Options]	Ce menu regroupe l'ensemble des touches qui permettent de configurer les modalités de visualisation des données (résolution du signal audio, polices pour les transcriptions, etc.)
[Help]	Accès aux fichiers d'aide

- une zone de saisie de la transcription orthographique où peuvent également être indiqués les changements de topiques de discours et les tours de parole (cf. cadre rouge) ;
- une série de touches audio qui remplissent les différentes fonctions offertes par tout lecteur de fichier son (Pause, Play, FF, RW, etc.) ;
- une visualisation du signal et de la segmentation associée. Cette dernière est représentée sur trois tires qui correspondent aux trois niveaux de segmentation : topiques de discours, tours de parole (locuteurs) et énoncés avec leur transcription orthographique (cf. cadre jaune).

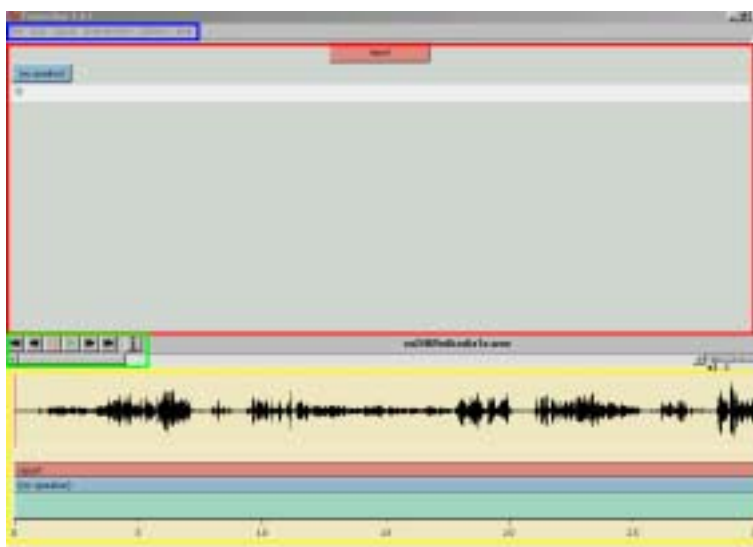


Fig. 8 : Fenêtre de travail de Transcriber

Après avoir créé la segmentation en topiques et en tours de parole, l'utilisateur saisit la transcription orthographique associée à chaque énoncé dans la fenêtre d'édition. Ces différentes informations sont représentées comme suit (cf. fig. 9) :



Fig. 9 : « Edition de la transcription »

Lorsque le fichier audio associé est joué, la partie de la transcription associée à l'énoncé entendu apparaît en sur-brillance (en blanc dans la fig. 9).

En conclusion, Transcriber est un outil intéressant qui offre de nombreux avantages :

- très bonne documentation ;
- interface graphique très conviviale ;
- sauvegarde des fichiers de transcription au format XML.

Mais plusieurs éléments contraignent fortement le « format des segmentations et des transcriptions », notamment :

- seule une segmentation à trois niveaux pré-formatés (topiques, locuteurs et énoncés) est possible ;
- les différents niveaux de segmentation sont structurés dans une hiérarchie respectant l'hypothèse de l'étagement strict (« strict layer hypothesis »).

En outre, le logiciel ne permet pas d'effectuer d'autres tâches telles que :

- l'extraction de fichiers audio correspondant à des énoncés ou à des tours de parole définis lors de la segmentation ;
- l'annotation des données sur la plan prosodique ou syntaxique en ajoutant de nouvelles « tires ».

3. Outils d'aide à la transcription permettant d'effectuer des analyses et/ou des requêtes

Les outils que nous avons présentés jusque là permettent principalement d'effectuer la transcription orthographique de fichiers sonores (cette dernière pouvant être alignée sur le signal) et de fournir des informations supplémentaires (segmentation en topiques de discours, changement de tours de parole, etc.). Nous allons maintenant nous intéresser à trois outils qui offrent aussi la possibilité de faire certains types d'analyses : PRAAT, CLAN et AG Toolkit. Dans ce chapitre, nous ne parlerons pas de PRAAT dans la mesure où un chapitre entier lui est consacré dans ce document (cf. chap. 3)

3.1 CLAN et le système CHILDES

Le système CHILDES (Child Language Data Exchange System) est un système qui a une double fonction :

- faciliter la constitution d'une base de données multilingues qui puissent être échangées et réutilisées ;
- aider à la recherche sur l'acquisition du langage et les interactions en fournissant des données et des outils.

Le système comporte trois composantes :

- un ensemble de principes et de conventions qui vise à faciliter la transcription et l'analyse : CHAT (Codes for the Human Analysis of Transcripts)
- un outil informatique d'édition des transcriptions et d'analyse automatique des données transcrites au format CHAT : CLAN (Computerized Language ANalysis)
- une banque de données qui regroupe des corpus enregistrés et transcrits de productions d'enfants dans une trentaine de langues.

Des informations sur ces différentes composantes et sur les modalités de participation au programme CHILDES peuvent être obtenues dans MacWhinney (2000a et b), Champaud (2001) et à partir du site WEB : <http://childes.psy.cmu.edu/>.

Dans cette section, nous présentons principalement la suite de programmes d'aide à la transcription et à l'analyse CLAN, mais nous serons amenée à exposer certains aspects de CHAT, dans la mesure où CLAN est dépendant des principes et conventions utilisés par CHAT.

Ces programmes ont été développés par Leonid Spektor de Carnegie Mellon University (Pittsburgh, USA) et peuvent être téléchargés à partir de l'adresse suivante : <http://childes.psy.cmu.edu/html/clan.htm>. Ils fonctionnent sur plusieurs plate-formes (Windows et Macintosh), la gestion des fichiers audio et vidéo sur lesquels sont enregistrées les productions se faisant grâce au logiciel QuickTime.

CLAN comprend deux composantes :

- une série de commandes ou de routines qui permettent d'analyser les données transcrites et codées au format CHAT (ou CA).
- un éditeur de texte qui sert en particulier à saisir les transcriptions.

Lorsque l'utilisateur ouvre le programme CLAN à partir du menu [Démarrer] ou en cliquant sur l'icône qui le représente, apparaissent à l'écran les fenêtres suivantes (cf. fig.10) :

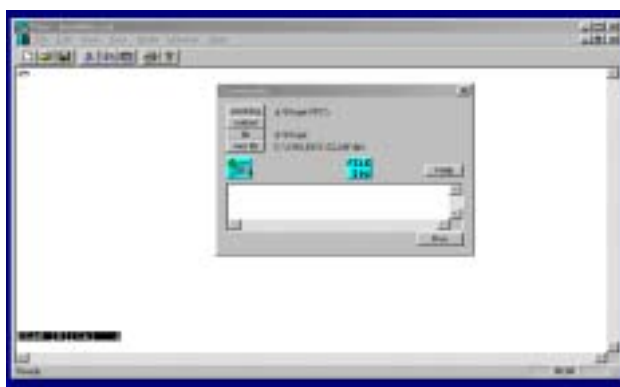


Fig. 10 : Ouverture de CLAN

Chacune des fenêtres correspond à une des composantes de CLAN :

- l'éditeur dont les modalités de fonctionnement seront exposées par la suite ;
- la fenêtre de commande (cf. fig.11) dans laquelle l'utilisateur indique le répertoire de travail (c'est à dire celui qui contient les fichiers de transcriptions à analyser) et saisit les commandes à effectuer.

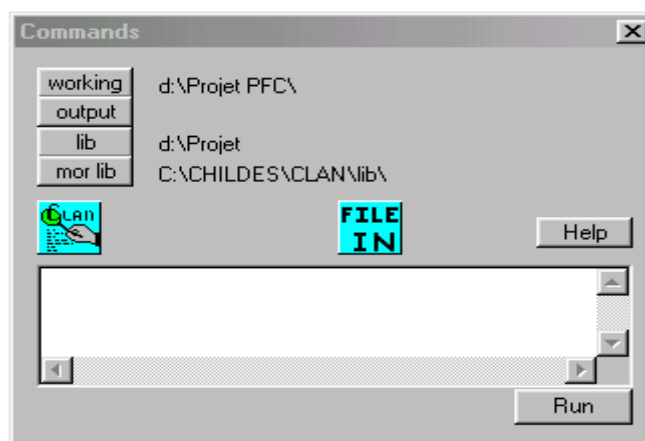


Fig. 11 : Fenêtre de commandes

Les répertoires de travail et de bibliothèques (qui, par défaut, sont les mêmes) peuvent être sélectionnés en cliquant respectivement sur les touches [working] et [lib]. Les commandes à effectuer sont, elles, directement saisies dans le cadre « blanc ». Dans le tableau ci-après, les commandes les plus fréquentes sont présentées :

Commandes	Description/ Fonction
CHECK	Vérifie la syntaxe des fichiers de transcription et signale les erreurs.
CHIP	Examine les répétitions et les expansions dans les interactions parent-enfant.
COMBO	Cherche des combinaisons complexes dans les énoncés.
COOCCUR	Cherche des patrons de cooccurrences de mots.
DATES	Calcule l'âge de l'enfant à partir de sa date de naissance et de la date du jour où l'enregistrement a été effectué.
FREQ	Calcule les fréquences des mots dans un ou plusieurs fichiers.
FREQMERG	Combine le résultat d'analyses faites avec FREQ.
GEM	Marque certains passages comme particuliers.
GEMLIST	Fait la liste des passages marqués par GEM dans un ou plusieurs fichiers.
KWAL	Cherche certains arrangements de mots.

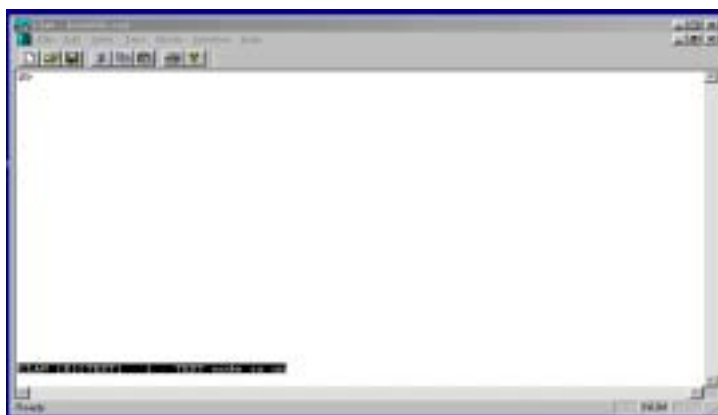
MAXWD	Recherche le mot le plus long d'un fichier.
MLT	Calcule la longueur moyenne des tours de parole.
MLU	Calcule la longueur moyenne des énoncés.
MODREP	Compare la phonologie de l'enfant aux modèles parentaux.
PHONFREQ	Calcule la fréquence d'apparition des phonèmes dans diverses positions.
TEXTIN	Convertit un texte normal au format CHAT.
WDLEN	Calcule les différentes longueurs d'énoncés en mots.

Pour lancer une commande ou une analyse, l'utilisateur doit :

- taper la commande dans la fenêtre de commande en respectant une syntaxe particulière. Pour, par exemple, calculer la fréquence d'apparition des mots produits par Antoine (noté ANT) dans le fichier *antoine01.cha*, il faut écrire : **freq *ANT antoine01.cha** ;
- appuyer sur la touche [Run] en bas à droite de la fenêtre de commandes.

Dans le cadre de ce chapitre, nous ne nous attarderons pas davantage sur les modalités de fonctionnements des différentes commandes. L'utilisateur intéressé trouvera la documentation nécessaire dans MacWhinney (2000a) et dans Champaud (2001).

Etudions maintenant comment une transcription alignée sur un fichier audio (ou vidéo) peut être faite à l'aide de CLAN. Dans ce cas, l'utilisateur doit se placer dans la fenêtre de l'éditeur (cf. fig. 12) :



Cette fenêtre contient les menus déroulants présents dans la majorité des éditeurs de textes :

- [File] pour ouvrir, fermer, sauvegarder ou imprimer des fichiers ;
- [Edit] pour effectuer des tâches telles que « copier », « coller », « rechercher », etc. ;
- [View] pour modifier l'apparence des textes saisis (tabulations, etc.) ;

- [Windows] pour activer ou désactiver des fenêtres ;
- [Help] pour accéder aux fichiers d'aide.

Outre ces menus « classiques » apparaît le menu [Mode] qui permet de sélectionner les modes selon lesquels vont s'effectuer les transcriptions et les annotations. Nous pouvons distinguer les modes d'édition des transcriptions et les modes de fonctionnement de l'éditeur.

3.1.1 Les modes d'édition

Les modes d'édition sont au nombre de trois : « mode texte », « mode CHAT » et « mode CA ». Ils sont sélectionnés dans le menu déroulant. Une fois le choix fait, le mode actif est indiqué en bas à gauche de l'éditeur (dans la fig. 12, le mode texte est actif). Les modes CHAT et CA doivent être sélectionnés si l'utilisateur veut ensuite lancer des commandes pour analyser les transcriptions.

Comme les modes CHAT et CA imposent le respect de certains principes dans la saisie des transcriptions, nous allons présenter quelques conventions propres au mode CHAT¹⁰ :

- Dans le mode CHAT, tous les fichiers doivent commencer par une en-tête qui a la forme suivante :

@Begin

@Participants : ANT Antoine, MAT Mathieu....

@Transcriber :

@Dependent :

Sur la ligne « participants », le transcripteur doit indiquer le nom de chacun des interlocuteurs précédé d'un trigramme. Sur la ligne « Transcriber » doit être donné le nom du transcripteur. Sur la ligne « Dépendent » sont indiqués le nom des tires qui sont utilisées en plus de la tire principale qui contient la transcription orthographique.

- chaque énoncé est transcrit comme une entrée séparée, et donc une ligne dans l'éditeur. Il a la forme suivante :

*ANT : tu as bien dormi ?

Une ligne principale commence par un astérisque *, qui est suivi d'un trigramme désignant le locuteur, de deux points, d'une tabulation et de l'énoncé. Tout énoncé doit être terminé par un signe de ponctuation forte et ne doit pas contenir d'autres signes de ponctuation.

- les lignes sont toutes numérotées, qu'il s'agisse des lignes principales ou des tires dépendantes.

La figure 13 représente une transcription courte au format CHAT.

¹⁰ Pour obtenir plus d'informations sur CHAT et CA, l'utilisateur peut consulter MacWhinney (2000a).

```

@Font:          Win95:Courier:-13:1
@Begin
@Participants:  MAR Marie, LYD Lydie Enquêteur
@Date:          03-NOV-1998
@Transcriber    Mathilde
@Dependent      Act
*LYD:          Alors est-ce que tu es née à Valueil ?
               %snd:"VA24F40MACADIA1.WAV"_122_1991
%act           Question yes/no
*MAR           Non.
               %snd:"VA24F40MACADIA1.WAV"_11000_13334
*LYD           Où est-ce que tu es née exactement?
               %snd:"VA24F40MACADIA1.WAV"_3372_4759
*LYD           Est-ce que c'est loin ?
               %snd:"VA24F40MACADIA1.WAV"_4904_5932
*MAR           Je suis née à Lisle à onze kilomètres de
Valeuil.
               %snd:"VA24F40MACADIA1.WAV"_6450_10032
*LYD           Mumun
%act           coopératif
*LYD           C'était plus petit que Valeuil plus grand.
               %snd:"VA24F40MACADIA1.WAV"_11000_13334
@End

```

Fig. 13 : Fichier de transcription CHAT

3.1.2 Les modes de fonctionnement de l'éditeur

Lorsque l'utilisateur effectue des annotations, il peut travailler sur les seuls fichiers de transcription dans l'éditeur, en ayant préalablement choisi les modes d'édition (CHAT ou CA). Mais, s'il veut faire des transcriptions alignées, il lui est possible de travailler en s'appuyant sur les fichiers vidéo ou audio. Dans ce cas, deux modes existent et peuvent être choisis dans le menu déroulant [Mode] :

- [Transcribe sound or movie] qui permet d'ouvrir le fichier audio ou vidéo dont on souhaite faire la transcription dans l'éditeur. Si l'utilisateur choisit un fichier audio, le signal s'affiche dans le bas de la fenêtre de l'éditeur. Si, en revanche, il sélectionne un fichier vidéo, celui-ci apparaît dans un écran de visualisation (cf. fig.14).



Fig. 14 : Editeur et visualisation de vidéo

- le mode « sonic » (sonic mode) qui permet de visualiser le signal sonore du fichier à transcrire, d'en sélectionner des extraits qui seront alors transcrits dans l'éditeur et d'associer ces extraits à la transcription à l'aide de la touche [s] sur la gauche de la ligne de visualisation du signal (cf. fig. 15)

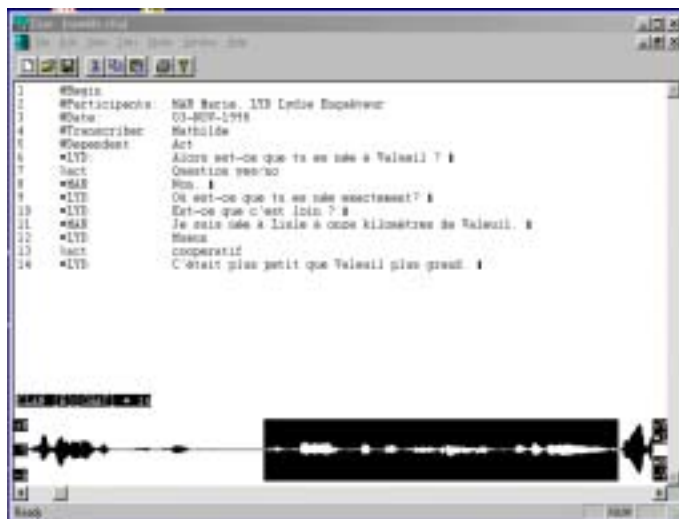


Fig. 15 : Edition en mode « sonic »

En cliquant sur la touche [S] dans la barre d'édition du signal, l'utilisateur insère des balises dans le fichier d'édition. Celles-ci permettent d'associer transcription et signal. Le résultat apparaît alors dans le fichier CHAT (cf. fig. 13). Le signal associé à ces énoncés peut ensuite être visualisé et étudié sous PRAAT grâce à la commande [Send to Praat] dans le menu [Mode].

En conclusion, le programme CLAN offre un certain nombre d'avantages :

- très bonne documentation ;
- plusieurs fonctions d'analyse intéressantes ;
- plusieurs modes de transcription ;

- compatibilité avec d'autres programmes, dont PRAAT.

Son utilisation ne peut être que recommandée dans le cas où les utilisateurs souhaitent effectuer des transcriptions de données sur l'acquisition du langage (langue première, langue seconde et langue étrangère). En revanche, ce programme impose le respect de plusieurs principes pour éditer les transcriptions (modes CA ou CHAT) qui sont assez lourds si les programmes d'analyse ne sont pas utilisés par la suite.

3.2 AG Toolkit

AG Toolkit (AGTK) est une boîte à outils qui permet de développer facilement des logiciels d'aide à l'annotation de données. Cet ensemble de programmes a été conçu par plusieurs chercheurs du « Linguistic Data Consortium » (LDC) dans le cadre d'un travail de recherche lancé dans la fin des années 90 et portant sur la représentation formelle des annotations de données sonores ou vidéo. Voyant que la diversité des logiciels d'aide à l'annotation et la multitude de formats de stockage constituaient un frein à l'échange et à la réutilisation des données, ces chercheurs ont proposé de recourir à des outils basés sur une architecture à trois niveaux, au lieu de deux (cf. Bird et al. (2000) et Ma et al. (2002)) :

- le niveau physique qui spécifie les formats de sauvegarde et de codage des annotations (XML, etc.) ;
- le niveau des applications qui renferme plusieurs composantes (les outils d'application, etc.) ;
- le niveau logique qui repose sur un formalisme linguistique capable de représenter lors d'une étape intermédiaire l'ensemble des informations saisies lors des tâches d'annotation et stockés dans un format particulier au niveau physique.

Les outils développés à partir de AG Toolkit sont tous basés sur cette architecture à trois niveaux, le niveau logique ayant pour fonction de représenter les informations d'annotation sous la forme de « graphes d'annotation » (cf. Bird et Liberman (2001), Maeda et al. (2001a) et Maeda et al. (2002)).

Les outils disponibles peuvent fonctionner sur plusieurs plate-formes (Unix, Linux, Windows) et sont téléchargeables à partir du site suivant : <http://agtk.sourceforge.net/>. Ils reposent tous sur une utilisation des logiciels suivants :

- les langages de script Tcl et Python ;
- le logiciel d'édition du signal sonore WaveSurfer ;
- le logiciel d'édition de vidéo QuickTime.

A l'heure actuelle, plusieurs outils d'annotation sont diffusés :

- TableTrans qui permet le codage d'observations faites lors de l'écoute du signal audio ou vidéo. Les informations codées sont représentées sous forme de tableau ;
- MultiTrans est un outil d'aide à la transcription d'interactions verbales enregistrées sur plusieurs pistes (une piste par locuteur). Le transcripteur peut transcrire chaque piste indépendamment des autres, cela permettant une bonne gestion des chevauchements. La fig. 16 représente la fenêtre de travail de MultiTrans avec une visualisation des signaux et des transcriptions associées ;

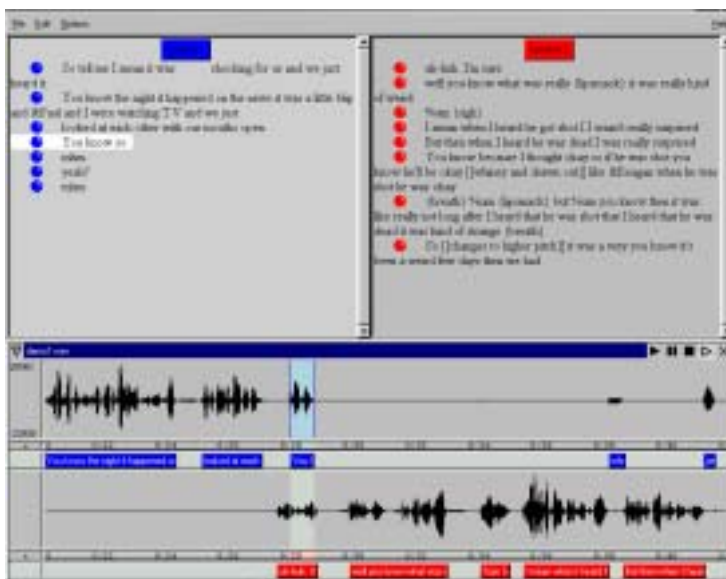


Fig. 16 : Fenêtre de travail de MultiTrans

- InterTrans qui permet de créer des annotations alignées sur le signal et correspondant à des translitérations ;
- TreeTrans qui facilite la création et la manipulation de représentations syntaxiques sous forme d'arbres.

Dans cette section, nous ne présenterons pas tous ces outils, mais nous expliquerons dans un premier temps ce que sont les graphes d'annotation, puis nous exposerons comment fonctionne TableTrans.

3.2.1 Les graphes d'annotation

La tâche d'annotation consiste toujours à assigner une étiquette à un intervalle, et cela quel que soit le niveau de segmentation ou le choix de l'intervalle : phonème, mot, syntagme syntaxique, groupe prosodique, énoncé ou tour de parole. Partant de cette constatation, Bird et Liberman (2001) ont développé un formalisme linguistique qui permet de représenter toute forme d'annotation, indépendamment du format physique dans lequel elle est sauvegardée (Childes, TIMIT, etc.). En effet, les annotations sont représentées sous la forme de graphes acycliques orientés appelés « graphes d'annotation »

où les étiquettes assignées aux intervalles sont encodées sur les arcs et où des indications temporelles optionnelles sont données dans les nœuds (cf. fig. 17).

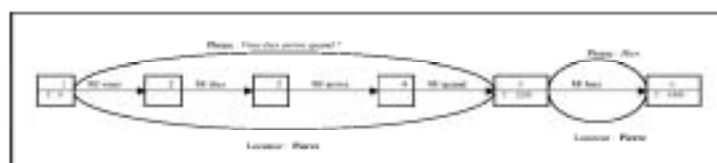


Fig.17: Graphe d'annotation

Dans cette figure apparaissent six nœuds qui permettent de définir cinq intervalles, mais parmi ceux-ci, seuls trois contiennent des informations temporelles : les nœuds 1, 5 et 6 qui coïncident avec des frontières d'énoncé (Phrase) et de tours de parole (Locuteur). Bien qu'une segmentation en mots soit effectuée, les intervalles ne sont pas alignés sur le signal. En outre, trois types d'annotation sont représentés dans ce graphe : l'étiquetage en mots (« M/vous », « M/êtes », etc.), le découpage en énoncés ou phrases (« Phrase : *Vous êtes arrivé quand?* », etc.), la segmentation en tours de parole (« Locuteur : Pierre », etc.).

Ce formalisme permet d'encoder toute forme d'annotation, sans imposer un niveau de segmentation, notamment pour l'alignement sur le signal. En outre, dans le cadre des travaux sur les graphes d'annotation, des langages de requêtes ont été développés et permettent d'extraire des informations des annotations ainsi formalisées.

3.2.2 TableTrans

TableTrans est un logiciel d'aide à la transcription qui a été développé à la suite des recherches sur les graphes d'annotation. Il permet d'effectuer une annotation alignée sur le signal de documents audio ou vidéo, en travaillant dans un tableau où les lignes correspondent à des portions du signal à annoter et les colonnes à des types d'annotation (Transcription, locuteur, identité de l'énoncé, etc.).

À l'ouverture du logiciel, la fenêtre suivante apparaît à l'écran (cf. fig. 18) :

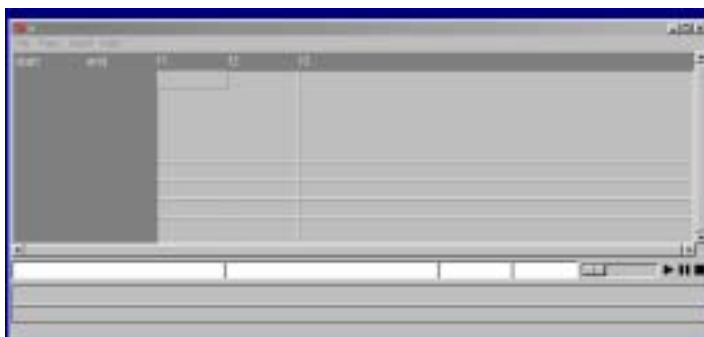


Fig. 18 : Ouverture de TableTrans

Les fonctions des différents menus sont synthétisées dans le tableau ci-après :

Menus	Description des fonctions
[File]	Permet la création et l'ouverture de fichiers d'annotation, l'ouverture de fichiers audio ou vidéo, la sauvegarde des tables d'annotation sous plusieurs formats (XML, Table, etc.).
[Trans]	Ce menu regroupe les fonctions classiques d'édition de texte (couper, copier coller), mais aussi les fonctions de classement et de visualisation des annotations par types et les fonctions de configuration des tables d'annotation (choix des colonnes et donc des tires, etc.).
[Sound]	Permet la sauvegarde d'une portion du signal correspondant à un intervalle, donc à une ligne dans le tableau.
[Help]	Accès à l'aide.

Après avoir ouvert le logiciel, le transcripteur doit ouvrir un fichier son (cf. menu [File]) et configurer le tableau d'annotation (cf. menu [Trans]) de façon à avoir autant de colonnes que de tires. Une fois ces tâches effectuées, le tableau d'annotation et la représentation graphique du signal sonore s'affichent à l'écran (cf. fig. 19) :

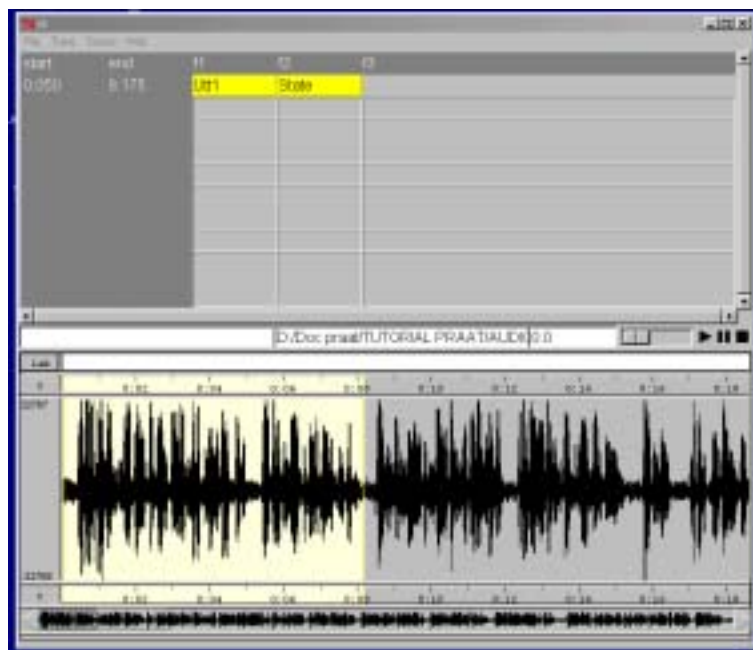


Fig. 19 : fenêtre de transcription

Dans cette fenêtre, le transcripteur peut sélectionner une partie du signal à l'aide de la souris sur la représentation graphique. La sélection apparaît alors en surbrillance. En appuyant sur la touche [Enter] de son clavier, l'utilisateur associe une ligne à cette sélection : les temps de début et de fin apparaissent alors dans les colonnes « start » et « end » du tableau. Ensuite, le contenu des autres colonnes (transcriptions orthographiques, identité du locuteur, etc.) est saisi. Afin de faciliter la transcription du passage, il est possible de rejouer l'extrait à l'aide des touches [Play], [Stop] et [Pause] situées au dessus de la représentation du signal. Après avoir achevé une ligne, le transcripteur peut

sélectionner le passage suivant et créer une nouvelle ligne avec la touche [Enter].

Les annotations ainsi effectuées peuvent être sauvegardées sous différents formats, mais surtout elles peuvent être représentées formellement sous la forme de graphes d'annotation. Des requêtes pourront alors être effectuées sur ces représentations.

D'une façon générale, cette suite d'outils est intéressante pour plusieurs raisons :

- l'idée de représenter toutes les annotations de la même façon afin de faciliter l'échange et la manipulation des informations est très attrayante ;
- les outils déjà construits sont flexibles et n'imposent pas un format particulier de codage des informations : le nombre des tires n'est pas limité, le niveau de segmentation pour l'alignement sur le signal peut être choisi par les utilisateurs.

En outre, dans le cadre de projets pluridisciplinaires, des outils particuliers peuvent être assez facilement développés.

Conclusion

Dans ce chapitre, nous avons présenté trois types d'outils d'aide à la transcription :

- les outils d'aide à la transcription sans éditeur intégré : ces programmes offrent les mêmes fonctionnalités qu'un lecteur de fichier audio classique (PLAY, PAUSE, STOP, FF, RW), auxquelles s'ajoute une fonction spécifique paramétrable qui permet une écoute réitérative et séquentielle du signal sonore. Cette fonction laisse à l'utilisateur le temps de saisir la transcription orthographique des séquences de signal jouées.
- les outils d'aide à la transcription avec éditeurs intégrés : ces logiciels permettent de segmenter le signal sonore et de saisir la transcription associée aux segments dans un éditeur de texte intégré. Les transcriptions ainsi obtenues sont alignées sur le signal. Les deux programmes que nous avons présentés dans ce cadre (VoiceWalker et Transcriber), bien qu'intéressants, contraignent fortement le format des annotations et des segmentations : le nom des tires et leur nombre sont en effet préformatés.
- les outils d'analyse et d'aide à la transcription : plusieurs outils peuvent être utilisés à la fois pour transcrire ou annoter un fichier audio (ou vidéo) et pour analyser les transcriptions et les fichiers audio, en extraire des informations, etc. Nous avons présenté deux de ces outils (CLAN et AG Toolkit) et mentionné un troisième dont les fonctionnalités sont décrites dans le chapitre suivant (PRAAT).

Bien que nous ayons proposé une « classification » des outils dans ce chapitre, ces programmes sont assez différents les uns des autres, et offrent des possibilités variées. Le choix d'un programme sur un autre doit être fait en fonction de plusieurs éléments :

- que souhaite faire l'utilisateur et pourquoi veut-il transcrire les données ? Les deux premiers outils sont intéressants, car simples d'utilisation, si l'objectif est de transcrire orthographiquement le contenu du signal et travailler ensuite sur les seules transcriptions. Si le but est de fournir une transcription alignée et de permettre une réutilisation des données, des outils comme Transcriber, Praat ou AG Toolkit sont préférables ;
- quel est le type des données à transcrire et quelles analyses seront faites sur les données ? Si le corpus à transcrire et à annoter est composé d'enregistrements de débats ou d'émissions radiophoniques, Transcriber est un très bon choix. S'il s'agit de données sur l'acquisition du langage, il est recommandé d'utiliser CLAN et de profiter des programmes d'analyses joints.

En conclusion, la diversité des outils d'aide à la transcription est intéressante, dans la mesure où chaque utilisateur n'a pas les mêmes objectifs et les mêmes attentes. Mais il faut éviter que la multitude des formats de sauvegarde et de codage deviennent un obstacle à l'échange des données. En ce sens, toute recherche visant à faciliter l'échange (cf. les graphes d'annotation) est bienvenue.

Références bibliographiques

Barras, C. , E. Geoffrois, Z. Wu et M. Liberman (1998) "Transcriber : a Free Tool for Segmenting, Labelling and Transcribing Speech" in *Proceedings of the First International Conference on Language Resources and Evaluation (LREC)*, pages 1373-1376, Grenade, Espagne.

Barras, C. , E. Geoffrois, S. Bird et Z. Wu (2000) "Transcribing with Annotation Graph" in *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC)*, pages 1517-1521, Mai 2000.

Barras, C. , E. Geoffrois, Z. Wu et M. Liberman (2001) "Transcriber : development and use of a tool for assisting speech corpora production" in *Speech Communication* "special issue on Speech Annotation and Corpus Tools" (S. Bird & J. Harrington eds) , Vol 33, No 1-2, Janvier 2001.

Bird S., K. Maeda, X. Ma, H. Lee, B. Randall et S. Zayat (2002) "TableTrans, MultiTrans, InterTrans and TreeTrans: Diverse Tools Built on the Annotation Graph Toolkit" in *Proceedings of the Third International Conference on Language Resources and Evaluation*, Paris: ELRA, 7 Pages.

Bird S. et M. Liberman. (2001) "A formal framework for linguistic annotation" in *Speech Communication* "Special issue on Speech Annotation and Corpus Tools" (S. Bird & J. Harrington eds) , Vol 33, No 1-2, Janvier 2001, 33: 23-60.

Bird, S., D. Day, J. Garofolo, J. Henderson, C. Laprun et M. Liberman (2000) "ATLAS: A flexible and extensible architecture for linguistic annotation" in *Proceedings of the Second International Conference on Language Resources and Evaluation*, pp 1699-1706.

Boersma, P. (2001) « Praat, a system for doing phonetics by computer » in *Glott International* 5 (9/10). Pages 341-345. Avec une préface de Vincent van Heuven.

Boersma, P. et D. Weenink (1996) «Praat, a System for doing Phonetics by Computer, version 3.4. » in *Institute of Phonetic Sciences of the University of Amsterdam, Report 132*. 182 pages. Téléchargeable à partir de www.praat.org.

Breen, J. (2001) *Multi-Channel Transcription Tool : User Manual*. Document HTML.

Champaud, C. (2001) « Une introduction au système CHILDES en français », Handout d'une présentation faite à Toulouse le 28 mars 2001. Université de Toulouse-Le Mirail. Téléchargeable à partir de : <http://childes.psy.cmu.edu/pdf/french.pdf>.

Dybkjær, L., S. Berman, M. Kipp, M. Wegener Olsen, V. Pirelli, N. Reithinger et C. Soria (2001) "Survey of Existing Tools, Standards and User Needs for Annotation of Natural Interaction and Multimodal Data" in *ISLE Natural Interactivity and Multimodality Working Group Deliverable D11.1*. Téléchargeable à partir de : <http://isle.nis.sdu.dk/>.

Dybkjær, L. et N.O. Bernsen (2000) "The MATE Markup Framework" in *Proceedings of the 1st SIGdial. Workshop on Discourse and Dialogue*, Hong Kong 7-8 October, 2000, pages 19-28.

Dybkjær, L. et N.O. Bernsen, (2000b) "Towards Corpus Annotation Standards – The MATE Workbench" in *proceedings of the COCODA Workshop 2000*, Beijing, China,.

Isard, A., D. McKelvie, B. Cappelli, L. Dybkjær, S. Evert, A. Fitschen, U. Heid, M. Kipp, M. Klein, A. Mengel, M. Baun Møller et N. Reithinger (1998) *MATE Deliverable D3.1 :Specification of Coding Workbench*. Téléchargeable à partir de : <http://mate.nis.sdu.dk/>.

Lieshout van, P. (2001) «PRAAT Workshop : A basic introduction ». Téléchargeable à partir de www.praat.org.

Ma X., H. Lee, S. Bird et K. Maeda (2002) "Models and Tools for Collaborative Annotation" in *Proceedings of the Third International Conference on Language Resources and Evaluation*, Paris: ELRA, 8 Pages

Maeda, K. (2001) *LDC TableTrans (Python version 1.0) : User Manual*. Document HTML.

Maeda K., S. Bird, X. Ma et H. Lee (2002) "Creating Annotation Tools with the Annotation Graph Toolkit" in *Proceedings of the Third International Conference on Language Resources and Evaluation*, Paris: ELRA, 8 Pages.

Maeda K., S. Bird, X. Ma et H. Lee (2001a) "A demonstration of Annotation Tools Based on Annotation Graphs" Ms. 6 mars 2001.

Maeda, K., X Ma, H. Lee et S. Bird (2001b) *The Annotation Graph Toolkit (version 1.0) : Application Developer's Manual*. Ms téléchargeable à partir de : http://www ldc.upenn.edu/AG/doc/tech_report/tr1.pdf.

Maeda K., S. Bird, X. Ma et H. Lee (2002) "Creating Annotation Tools with the Annotation Graph Toolkit" in *Proceedings of the Third International Conference on Language Resources and Evaluation*, Paris: ELRA, 8 Pages.

MacWhinney, B. (2000a) *The CHILDES project: Tools for analyzing talk*, Third edition. Volume I: Transcription format and programs. Mahwah, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-2995-4

MacWhinney, B. (2000b) *The CHILDES project: Tools for analyzing talk*, Third edition. Volume II: The database. Mahwah, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-3572-5

Annoter et segmenter des données de parole sous PRAAT

En collaboration avec Abderrahim MEQQORI
et Jean-Michel TARRIER

Introduction

Dans le chapitre précédent plusieurs outils informatiques d'aide à la transcription et à l'annotation¹¹ de données sonores ont été présentés et une classification a été effectuée sur la base des fonctionnalités des logiciels. Parmi les outils permettant à la fois la transcription, l'annotation et l'analyse de données sonores, le programme PRAAT a été mentionné. Dans ce chapitre, nous nous proposons de présenter plus en détails certaines des fonctionnalités de ce programme qui a été développé à l'Institut des Sciences Phonétiques de l'Université d'Amsterdam (Pays-Bas) par Paul Boersma et David Weenink.

Comme PRAAT a été retenu dans le cadre du projet « Phonologie du Français Contemporain » (cf. chapitre DURAND, LAKS ET LYCHE) pour transcrire orthographiquement et annoter les données, nous nous concentrons sur quelques fonctionnalités qui peuvent être intéressantes pour les tâches de transcription et d'annotation des corpus oraux, à savoir¹² :

- les tâches de saisie des données audio ;
- les tâches d'annotation et de transcription des données, et plus particulièrement la tâche de transcription orthographique alignée sur le signal ;

¹¹ Dans ce chapitre, nous appelons annotation la tâche qui consiste à assigner des étiquettes (morphologiques, syntaxiques, prosodiques, pragmatiques, etc.) à une portion du signal, ou plutôt au contenu linguistique ou extralinguistique de cette portion de signal. En revanche, nous entendons par transcription des données la tâche qui consiste à encoder orthographiquement ce qui est contenu dans le signal (transcription orthographique, indication des tours de parole et des changements de locuteurs, indication des événements sonores divers, etc.). Cf. Delais-Roussarie, ce volume pour plus de détails.

¹² Les indications et les saisies d'écran seront effectuées à partir de la version 4.012 du 3 avril 2002 fonctionnant sous Windows.

- les tâches de gestion des annotations et d'extraction d'information et de fichiers à partir des annotations ;

Dans cette présentation ce n'est pas une description détaillée du logiciel PRAAT qui est fournie, mais plutôt une aide permettant au néophyte de se familiariser avec cet outil pour la transcription et l'annotation de corpus oraux. Dans un premier temps, nous présentons rapidement le logiciel PRAAT et expliquons comment les données audio peuvent être saisies. Ensuite, nous expliquons de façon très détaillée comment elles peuvent être transcrites sous PRAAT : une attention particulière est donnée à la transcription orthographique. Puis, nous montrons ce qui peut être fait à partir des transcriptions et des annotations, notamment en matière d'extraction d'information et de segmentation des fichiers audio en sous-fichiers.

1. Commencer avec PRAAT

Avant d'expliquer comment débiter avec PRAAT et comment l'utiliser pour effectuer des enregistrements de données audio, nous allons rapidement mentionner certaines des possibilités offertes par ce programme.

1.1 Présentation générale

Le logiciel PRAAT, fonctionnant sur plusieurs plates-formes¹³, offre la possibilité d'effectuer de multiples tâches :

- enregistrement de fichiers audio qui peuvent ensuite être analysés sous PRAAT ;
- transcription, étiquetage et segmentation de fichiers audio dont la taille peut aller jusqu'à 2 Gigabytes, c'est à dire 3 heures d'enregistrement stéréo de qualité CD ou 16 heures d'enregistrement mono à 22 kHz. Ces enregistrements peuvent avoir été effectués sous PRAAT ou peuvent provenir d'autres fichiers audio ;
- analyses phonétiques et acoustiques au niveau segmental : le logiciel permet de mener des analyses spectrographiques et d'effectuer des mesures précises telles que la durée du VOT des plosives, la valeurs des différents formants d'une voyelle, etc. ;
- analyses des paramètres prosodiques (F0, durée et intensité] et modification par stylisation des courbes de fréquence fondamentale et d'intensité ;
- manipulation et modification du signal de parole (utilisation de filtre ; analyse-synthèse, etc.) ;
- construction d'outils d'apprentissage (Réseau de neurones et élaboration de grammaires dans le cadre de la théorie de l'optimalité (OT : Optimality Theory)) ;

¹³ Le programme PRAAT fonctionne sur les systèmes suivants : Windows (et plus précisément Win 95, 98, NT4, ME, 2000 et XP), Macintosh (systèmes 7.1 à 9.2), Linux, Sparc SOLARIS, HP Unix et Silicon Graphics IRIX (et plus précisément Indigo, Indy, O2, Onyx et autres).

- écriture de script pour effectuer plus rapidement certaines tâches d'extraction d'information, d'édition, etc.

Dans cette présentation, nous ne nous intéressons qu'à certaines de ces fonctionnalités, mais des informations complémentaires peuvent être obtenues sur le site Internet <http://fonsg3.let.uva.nl/praat/>, dans des « tutorials » (cf. Lieshout (2001)) et dans les fichiers d'aide intégrés au programme.

1.2 La prise en main

Le programme PRAAT peut être téléchargé à partir du site Internet suivant : <http://fonsg3.let.uva.nl/praat/>. Une fois installé, il peut être exécuté de plusieurs façons :

- en double-cliquant sur l'icône PRAAT du répertoire « Program files » ;
- en double-cliquant sur un raccourci de PRAAT placé sur le bureau (Desktop) de l'ordinateur ;
- en sélectionnant PRAAT dans le menu « Démarrer », si un raccourci du programme y a été installé.

Lorsque le programme est lancé, deux fenêtres s'ouvrent comme le montre la figure 1 :

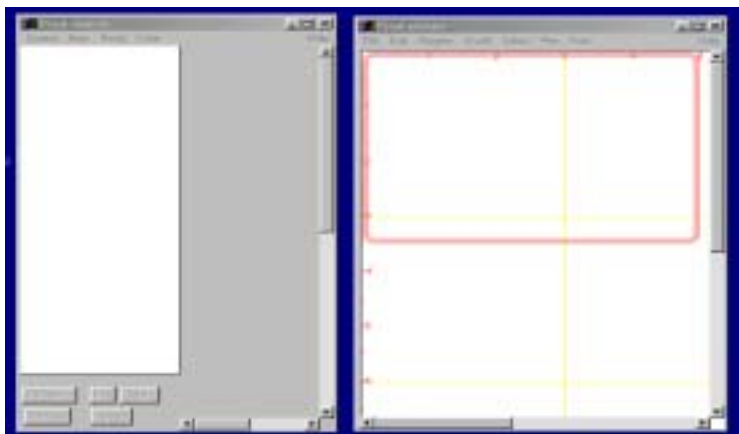


Fig.1 : Ouverture de PRAAT

La fenêtre de gauche est intitulée « **Praat objects** » et sert à « lister » les différents objets (fichiers sons, fichiers d'annotation, etc.) à partir desquels sont effectuées les analyses. La fenêtre de droite, intitulée « **Praat picture** », est utilisée pour reproduire des figures (sonagramme, courbe de F0, etc.) qui pourront être exportées vers d'autres logiciels (traitement de texte, etc.).

Dans la fenêtre « **Praat objects** », des menus déroulants peuvent être ouverts en cliquant sur [Control], [New], [Read], [Write] et [Help]. Une description synthétique des fonctions offertes dans les menus est proposée dans le tableau ci-dessous :

	Descriptif des fonctions accessibles à partir des menus déroulants
[Control]	Edition de scripts, Configuration (taille des buffers, etc.),
[New]	Création de fichiers divers (enregistrement audio, création de sons, fichiers d'annotation), création de grammaires OT, création de sons par synthèse articulatoire, etc.
[Read]	Ouverture de fichiers existants (annotation, fichiers audio, fichiers d'annotations effectuées sous XWaves, etc.)
[Write]	Sauvegarde des fichiers sous des formats particuliers (fichiers binaires, fichiers texte, etc.)
[Help]	Accès au manuel et à plusieurs tutoriaux (introduction à Praat, introduction à l'édition de scripts, etc.), accès à des informations pratiques (FAQ, Editeurs, etc.).

1.3 Effectuer des enregistrements ou saisir des données audio sous PRAAT¹⁴

Les données audio à enregistrer peuvent provenir de deux sources distinctes :

- d'une source externe (Platine K7, platine DAT, Minidisc, etc.) qui a été préalablement branchée sur une des lignes « entrée » de la carte son (SPDIF, Line IN, Optique IN, etc.) ;
- d'un microphone qui a également été branché sur la prise microphone de la carte son (ou de l'ordinateur). Dans ce cas, l'enregistrement est effectué directement sous PRAAT.

Une fois ces « branchements » faits, la saisie des données audio ou les enregistrements se font en cliquant sur la touche [New] dans la fenêtre « **Praat objects** » et en sélectionnant [Record mono Sound] ou [Record stereo Sound] dans le menu déroulant. S'affiche alors la fenêtre représentée dans la fig. 2.

¹⁴ Pour effectuer des enregistrements sous PRAAT, il faut impérativement que l'ordinateur (PC, Mac, etc.) ait une carte son.

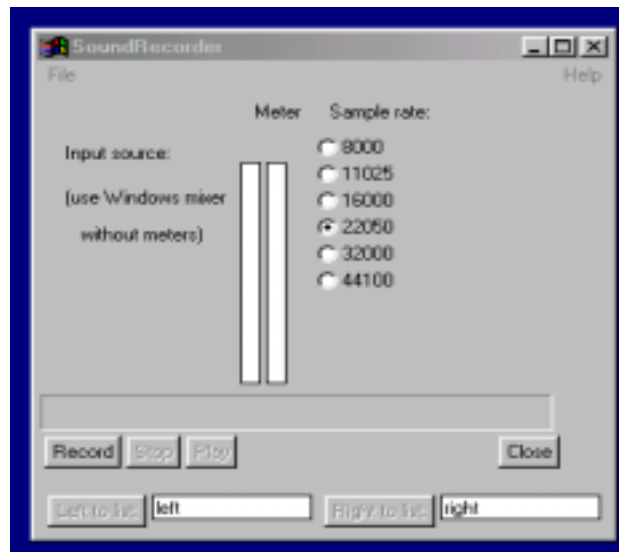


Fig.2 : Fenêtre d'enregistrement de données

Cette fenêtre contient :

- un vu-mètre qui permet de contrôler le niveau du volume d'enregistrement. Si le niveau n'est pas satisfaisant, les réglages sont à effectuer à partir de la fenêtre de commande de volume du système d'exploitation (Sous Windows, il s'agit de la fenêtre [Commande de volume] accessible à partir du menu « **Démarrer** », sous menu « **Divertissement** ») ;
- une liste des fréquences d'échantillonnage qui peuvent être choisies pour effectuer la numérisation des données audio. L'utilisateur y sélectionne la fréquence à l'aide de la souris ;
- des touches qui permettent de lancer un enregistrement (touche [Record]), de l'arrêter (touche [Stop]), de l'écouter (touche [Play]) ;
- une touche ([To list]) qui a pour but de confirmer la saisie des données sonores et leur affichage dans la liste des objets dans la fenêtre « **Praat objects** ».

L'enregistrement à proprement parler s'effectue en appuyant sur la touche [Record], le niveau du volume d'enregistrement s'affichant alors sur le vu-mètre. Une fois les données enregistrées, l'utilisateur appuie sur la touche [Stop]. S'il est satisfait de la qualité des données saisies, il appuie sur la touche [To list] pour que les données sonores enregistrées s'affichent dans la liste des objets dans la fenêtre « **Praat Objects** ». Ensuite, les données peuvent être définitivement sauvegardées sur le disque dur ou sur un périphérique de stockage. Pour ce faire, les données qui viennent d'être enregistrées doivent être sélectionnées dans la liste des objets (le fichier son apparaît alors en surbrillance) et l'utilisateur doit choisir l'une des touches permettant la sauvegarde d'un fichier audio dans le menu déroulant [Write]¹⁵. S'ouvre alors

¹⁵ L'enregistrement des données audio peut se faire sous plusieurs formats qui sont synthétisés dans le tableau ci-dessous :

une boîte de dialogue dans laquelle l'utilisateur indique le nom du fichier à sauvegarder et le répertoire dans lequel il veut effectuer cette sauvegarde.

2. Annoter et transcrire des données avec PRAAT

Bien qu'il soit possible de numériser des données et d'effectuer des enregistrements sous PRAAT (cf. § 2.3), il est également possible de travailler sur des données qui ont été numérisées avec d'autres logiciels et qui se trouvent sur le disque dur de l'ordinateur ou sur un périphérique de stockage (CD-Rom, Disquette, etc.). Avant d'effectuer la transcription et les annotations, il faut ouvrir le fichier audio que l'on souhaite transcrire ou annoter. Pour ce faire, il faut cliquer sur [Read] dans la fenêtre « **Praat objects** » et sélectionner dans le menu déroulant :

- [Read from file] si le fichier audio est de taille assez limitée (ou d'une durée n'excédant pas 3 minutes d'enregistrement en mono à 22 Khz) ;
- [Open long sound file] si le fichier est de taille plus importante.

Une boîte de dialogue s'ouvre alors et demande à l'utilisateur d'indiquer le nom du fichier audio à ouvrir (avec son extension) et le répertoire dans lequel il se trouve. Le fichier apparaît alors dans la fenêtre « **Praat objects** » et peut être sélectionné pour en effectuer une transcription ou pour l'annoter.

2.1 Qu'est-ce que transcrire ou annoter et comment commencer ?

Avant d'expliquer précisément comment procéder pour transcrire ou annoter des données sous Praat, il est important de réfléchir sur ce que supposent les tâches d'annotation et de transcription. En fait, annoter et transcrire des données audio nécessitent d'effectuer deux choses :

- de segmenter le signal audio : cela consiste à délimiter des intervalles ou à déterminer des points précis. La taille des intervalles et la nature des points varient en fonction des objectifs visés lors de la tâche d'annotation. Les intervalles peuvent correspondre à des tours de parole, à des énoncés, à des syntagmes syntaxiques ou prosodiques, à des mots, à des syllabes, à des phonèmes, etc. ;

FORMAT POSSIBLE POUR LA SAUVEGARDE (ET LA LECTURE) DES DONNEES AUDIO	TOUCHE A SELECTIONNER DANS LE MENU DEROULANT [Write]
AIFF	[Write to AIFF file...]
AIFC	[Write to AIFC file...]
WAV	[Write to WAV file...]
AU	[Write to Next/Sun file...]
NIST	[Write to NIST file...]
Kay Sound file	[Write to Kay Sound file...]
RAW 16 bits big Endian file	[Write to raw 16-bits Big Endian file...]
RAW 16 bits Little Endian file	[Write to raw 16-bits Little Endian file...]

- d'assigner des étiquettes aux points ou aux intervalles délimités. La nature des étiquettes est variée : il peut s'agir du nom des locuteurs, de la transcription orthographique ou phonémique de ce qui est dit sur la portion de signal segmentée, de la catégorie syntaxique ou grammaticale des éléments présents dans l'intervalle, etc.

Avant d'entreprendre des tâches de transcription et d'annotation, la détermination du niveau de segmentation et des étiquettes (ou des informations à encoder) est nécessaire. Mais il faut également savoir comment seront représentées ces différentes informations. Pour la seule tâche de transcription orthographique d'un court entretien entre deux locuteurs, plusieurs représentations peuvent être utilisées. Considérons l'exemple suivant :

(1) Dialogue :

Pierre : Allo Jean, c'est Pierre ! Comment vas-tu ?

Jean : Ca va. Et toi ?

Pierre : Ca va, mais j'ai beaucoup de choses à faire en ce moment. Je t'appelle parce que j'aimerais savoir si tu veux venir dîner samedi.

Jean : Oui, pourquoi pas ? Vers quelle heure ?

Pierre : Viens vers vingt heures. Ca ira ?

Jean : OK, pas de problème. A samedi donc.

Pierre : A samedi.

Sous PRAAT, les annotations et transcriptions sont représentées de façon pluri-linéaire : chaque tire est définie pour encoder certaines informations. Pour l'exemple (1), la transcription orthographique dans son ensemble peut être représentée sur une seule tire (cf. fig. 3) ou, au contraire, une tire peut être créée pour chacun des locuteurs (cf. fig.4).



Fig. 3 : Transcription sur une seule tire

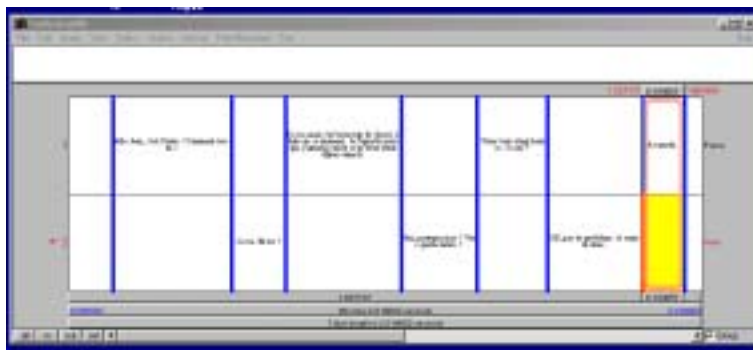


Fig. 4 : Transcription avec une tire par locuteur

Que le transcrip-teur choisisse de créer une seule tire pour l'ensemble de la transcription orthographique ou, au contraire, qu'il crée une tire par locuteur, il doit procéder de la même façon, à savoir créer un fichier texte qui sera associé au fichier audio et qui contiendra les transcriptions et annotations de chacune des tires. Pour y parvenir, il faut sélectionner dans la fenêtre « **Praat objects** » le fichier audio dont on veut effectuer la transcription ou l'annotation : celui-ci apparaît alors en sur-brillance. Ensuite, il faut sélectionner le menu [Label & segment] à droite de la liste des objets dans la fenêtre « **Praat objects** » et choisir [To TextGrid]. Une boîte de dialogue est alors ouverte avec deux champs : l'un intitulé [Tier names] et indiquant [MARY JOHN BELL], l'autre intitulé [POINT TIER] et contenant le mot [BELL]. Dans le champ [Tier names], le transcrip-teur doit indiquer le nombre de tires qu'il veut créer, sachant que c'est le nombre de mots séparés par un espace qui permet de spécifier le nombre de *tires*. Pour créer une seule tire pour l'ensemble de la transcription, il suffit de donner un nom à cette tire et d'indiquer ce nom dans la boîte de dialogue (par exemple : « orthographe », cf. fig. 3). En revanche, pour créer deux tires, une pour chacun des locuteurs en présence par exemple, le transcrip-teur doit donner deux noms et indiquer dans la boîte de dialogue les noms (par exemple : « Paul Pierre », cf. fig. 4). Dans le champ [Point tier], il faut indiquer le nom des tires qui sont composées d'une succession de points et non d'une succession d'intervalles. Dans le cas des transcriptions orthographiques ou phonétiques, les représentations symboliques du contenu du signal audio sont conçues comme des étiquettes assignées à des intervalles, en revanche, si une transcription prosodique est faite à partir de points correspondant à des maxima ou des minima de F0, cette dernière sera conçue comme une « *point tier* ».

Une fois la boîte de dialogue remplie, il faut cliquer sur le bouton [OK]. On se retrouve dans la fenêtre « **Praat Objects** », où apparaissent le nom du fichier audio ouvert précédemment et le fichier texte associé que l'on vient de créer et qui s'appelle « *TextGrid* », ce dernier étant alors en surbrillance. Il faut alors cliquer sur l'un des deux items et glisser vers l'autre de manière à les sélectionner tous les deux. A cette étape, nous avons les deux éléments en surbrillance.

Il faut ensuite cliquer sur le bouton [Edit]. Apparaîtra alors à l'écran une fenêtre :

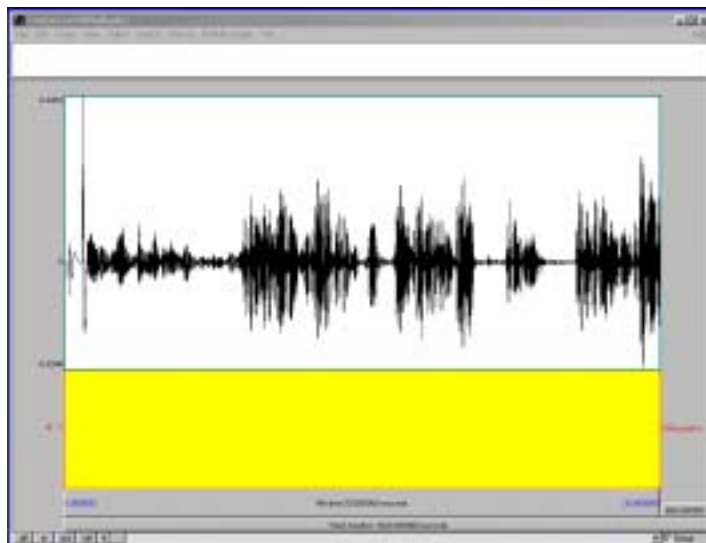


Fig. 5 : Fenêtre de création de « *textgrid* »

La fenêtre d'écran est composée de plusieurs éléments. En haut de la page, un champ de saisie immédiatement sous le menu. Sous ce champ de saisie apparaît une représentation graphique du son. Par défaut, apparaissent un sonagramme et une courbe de fréquence fondamentale, mais il est possible de modifier les paramètres afin de ne faire apparaître que le signal (cf. fig.5). Pour effectuer ces modifications, il suffit de sélectionner les représentations qui doivent apparaître à l'écran à l'aide de la touche [Show analyses...] dans le menu [View]. Immédiatement en dessous de ces représentations graphiques du son se trouvent les tires, dont le nombre variera en fonction des choix effectués par le transcripneur. Parmi ces tires figure en jaune la tire active, c'est à dire celle dans laquelle apparaîtra le texte tapé dans le champ de saisie. Sous les tires se trouvent deux bandes correspondant à la longueur du son. En bas à gauche de la fenêtre se trouvent quatre boutons qui vous serviront à visualiser le son à différentes échelles. Ils se présentent comme suit :

- Bouton [*all*] : il permet de visualiser l'ensemble de la séquence. En général, quand la longueur du signal atteint un certain seuil, le programme n'affiche pas la représentation acoustique.
- Bouton [*in*] : il permet d'agrandir la portion du signal à voir (*zoom in*) ;
- Bouton [*out*] : le contraire du précédent (*zoom out*) ;
- Bouton [*sel*] : il permet de ne visualiser que la sélection, autrement dit, quand vous avez sélectionné une partie du son, le fait de cliquer sur [*sel*] agrandit cette partie à la taille de toute la fenêtre ;

En plus de ces quatre éléments de navigation, on dispose d'un traditionnel ascenseur qui permet de se déplacer sur la droite ou sur la gauche.

Le travail de transcription s'effectue à partir de cette fenêtre et consiste à déterminer des intervalles et à leur assigner « des étiquettes ».

2.2 Définir les intervalles et leur assigner une étiquette ?

Pour définir les intervalles, il faut d'abord sélectionner des extraits du signal en cliquant puis en glissant (« *click and drag*») dans la fenêtre donnant la représentation graphique du signal sonore. Un carré rouge délimite alors la portion sélectionnée. Cette sélection pourra être écoutée en cliquant sur la bande correspondant à la longueur de la sélection (spécifiée en secondes) qui est apparue au-dessus et au-dessous de la fenêtre en question. Lorsque un segment à transcrire a été identifié, des marques (voir plus loin) doivent être placées pour créer des intervalles. La première marque créée délimite l'intervalle qui va du début du document sonore jusqu'à la marque. Par la suite, toute marque créée à droite permettra de constituer la frontière d'un nouvel intervalle.

Pour créer les marques, il faut faire un simple clic sur le signal sans faire glisser la souris, sinon le mouvement est interprété comme une sélection et non comme une création de marque. Apparaît alors une « marque » rouge sur le signal et grise sur la *tire* active (en jaune). Pour valider cette marque et donc la créer, il faut appuyer sur la touche « retour chariot » ([Enter]). Tant que la marque n'a pas été validée, elle n'est pas prise en compte par le logiciel et le fait de cliquer ailleurs la fera se déplacer. Avant de transcrire, il faut s'assurer que la marque (alors en bleu) et que l'intervalle ont été créés sur la *tire* concernée, laquelle est en jaune. Le logiciel PRAAT vous place alors automatiquement dans le champ de saisie en haut de la fenêtre. Tout ce qui y est saisi est répercuté dans l'intervalle créé alors en jaune. Tout ajout ou effacement s'effectuera dans ce champ qui fonctionne comme la fenêtre d'un traitement de texte. Dans cet éditeur, plusieurs caractères particuliers (certains caractères accentués, les symboles phonétiques, etc.) ne peuvent pas être saisis directement sur le clavier, mais nécessitent la saisie d'une suite de signes¹⁶.

2.3 Sauvegarder ou modifier une annotation ou une transcription ?

Quand une partie de la transcription est faite ou lorsque la séance de travail est finie, la fenêtre peut être fermée en cliquant sur la croix en haut à droite ou en choisissant [Close] dans le menu [File]. Mais, avant de quitter définitivement PRAAT, il faut sauvegarder le fichier de transcription (*TextGrid*) car cela ne se

¹⁶ Le détail des signes à taper pour saisir des symboles particuliers est fourni dans le fichier d'aide accessible à partir du menu [Help], en sélectionnant [About special symbols]. Le tableau ci-dessous indique les signes à saisir pour certains caractères français fréquents :

Symboles ou caractères	Séquences à saisir
â, î, ê, ô, û	\a^, \i^, \e^, \o^, \u^
À, Î, Ê, Ô, U	\A^, \I^, \E^, \O^, \U^
ä, ï, ë, ö	\a", \i", \e", \u"
Ä, Ì, È, Ö	\A", \I", \E", \U"
ù	\u`
ç	\c.
«	\<<
»	\>>

fait pas automatiquement. Pour ce faire, il suffit de le sélectionner dans la fenêtre « **Praat objects** » et de choisir [Write to text file] dans le menu [Write]. Apparaît alors une boîte de dialogue « Write Object(s) to text file » dans laquelle il faut indiquer dans quel répertoire et sous quel nom doit être sauvegardé le fichier. Par défaut, le logiciel PRAAT assigne à ce fichier le même nom que celui du fichier audio auquel il est associé, mais l'extension est différente, puisqu'il s'agit d'un fichier *.TextGrid*.

Pour reprendre par la suite une transcription déjà commencée lors d'une précédente séance de travail, la marche à suivre est la même, sauf en ce qui concerne le fichier *.textgrid* qui, dans ce cas, a déjà été créé. Après avoir ouvert le fichier audio avec les touches [Read from file] ou [Open long sound file], il faut ouvrir le fichier texte associé (*.TextGrid*) qui a été créé auparavant à l'aide de la touche [Read from file] dans le menu [Read].

A tout moment, les intervalles créés et les étiquettes assignées peuvent être modifiés par des opérations de déplacement ou d'effacement de marques :

- pour déplacer une marque déjà créée, il faut cliquer dessus au niveau de la *tire* et glisser en gardant cliqué. La marque suit le mouvement.
- pour éliminer une marque, il faut d'abord la sélectionner en cliquant dessus (ce qui la fait passer du bleu au rouge). Puis il faut sélectionner la touche [Remove] dans le menu déroulant [Point/Boundary]. Le texte qui précédait la marque et, s'il existe, le texte qui la suivait sont fusionnés dans un même intervalle.

Dans la fig. 6 ci-dessous, un exemple de transcription est donné. Plusieurs éléments peuvent être remarqués :

- les deux dernières répliques correspondent au même tour de parole, dont le locuteur est l'enquêteur (noté E).
- la longueur de la partie sélectionnée est connue grâce aux inscriptions sur les bandes qui se trouvent au-dessus et au-dessous et qui permettent d'écouter la séquence par un simple clic.
- la transcription, dont on ne voit qu'une partie dans le carré jaune, est visible dans sa totalité dans le champ de saisie sous la barre de menus.

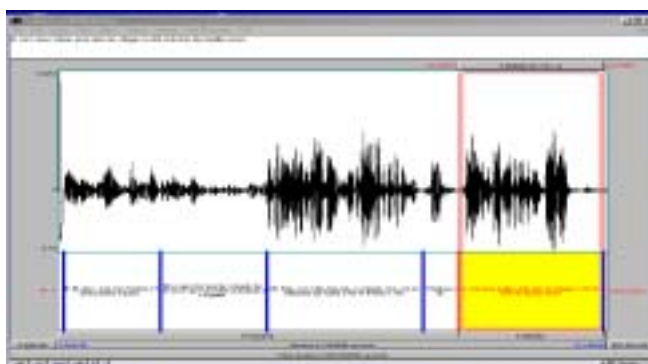


Fig. 6 : Fenêtre de création de « TextGrid »

Les transcriptions sont enregistrées dans un fichier appelé fichier *.TextGrid* où sont indiqués les temps de début et de fin de chaque intervalle créé, ainsi que l'étiquette associée à l'intervalle. Le fichier *.TextGrid* correspondant à la transcription présentée dans la fig. 6 est donné dans la figure 7 :

```
File type = "ooTextFile"
Object class = "TextGrid"
```

Fig. 7 : Fichier TEXTGRID

3. Enrichissement et gestion des annotations

Dans cette section, nous allons présenter comment effectuer plusieurs tâches importantes dans la gestion des annotations et des transcriptions, à savoir :

- Créer, enrichir ou modifier les tires existantes ;
- Effectuer des requêtes sur les tires et les annotations ;
- Segmenter les fichiers en sous-fichiers à partir des annotations.

3.1 Créer, enrichir, fusionner et extraire des tires

Nous avons vu dans la section 3 comment créer des tires dans lesquelles peuvent être encodées différents types d'informations : transcription orthographique, découpage prosodique, etc. Pour ce faire, il faut créer un fichier texte *.TextGrid* associé à un fichier audio. Ce fichier texte regroupera deux types d'informations :

- la segmentation en intervalles qui a été effectuée par l'annotateur ;
- le nom des étiquettes associées à chaque intervalle.

Une fois saisies, ces informations sont enregistrées par l'utilisateur dans le fichier *.TextGrid*. Ensuite, il est possible de modifier ce fichier texte après l'avoir ouvert à partir de la touche [Read from file] dans le menu [Read] de la fenêtre « **Praat Objects** ».

Les opérations de base sur les tires peuvent se faire de deux façons distinctes : dans la fenêtre « **Praat Objects** » ou à l'intérieur de la fenêtre de *TextGrid* (cf., entre autres, fig. 4, fig. 5 et fig. 6).

3.1.1 Créer, modifier, extraire et fusionner des tires dans la fenêtre « Praat Objects »

Le fichier *.TextGrid* que l'utilisateur souhaite modifier doit être sélectionné dans la fenêtre « **Praat Objects** ». Une fois qu'il apparaît en surbrillance, plusieurs touches apparaissent à droite de la liste des objets dans la fenêtre « **Praat Objects** ». Parmi celles-ci, trois touches jouent un rôle fondamental dans la création, l'extraction et la modification du fichier *.TextGrid* :

- la touche [Modify] donne accès à un menu déroulant qui offre plusieurs fonctionnalités dans la création et la modification des tires. Une description en est faite dans le tableau ci-après.

	Fonction	Mode d'emploi
[Insert interval tier...]	Permet la création d'une nouvelle tire formée d'intervalles	Après sélection de cette touche apparaît une boîte de dialogue dans laquelle l'utilisateur doit remplir deux champs : <ul style="list-style-type: none">- position de la tire à créer (position 1 représentant la tire qui est au sommet) ;- nom attribué à la tire à créer.
[Insert point tier...]	Permet la création d'une nouvelle tire formée d'une succession de points	Une boîte de dialogue apparaît pour demander la position et le nom de la tire à créer.

[Duplicate tier...]	Permet de créer une nouvelle tire en dupliquant une tire existante. La segmentation et les étiquettes de cette tire se retrouvent dans la nouvelle tire, mais il est possible de les modifier.	Une boîte de dialogue apparaît pour demander : <ul style="list-style-type: none"> - quelle tire doit être dupliquée ; - quelle position doit prendre la nouvelle tire ; - comment elle doit s'appeler.
[Remove tier...]	Permet de supprimer une tire existante.	Dans une boîte de dialogue, l'utilisateur doit indiquer le numéro de la tire à supprimer.

- la touche [Extract tier...] permet d'extraire une tire particulière d'un fichier *.TextGrid*. Des requêtes pourront alors être faites sur cette tire. Lorsque la touche [Extract tier...] a été sélectionnée, une boîte de dialogue apparaît pour demander le numéro de la tire à extraire. La tire extraite est ensuite automatiquement placée dans la liste des objets. Elle est nommée *IntervalTier x*, où *x* représente le nom qui a été attribué à cette tire lors de sa création (« *IntervalTier Ortho* », « *IntervalTier Jean* », etc.). Elle peut être sauvegardée en choisissant la touche [Write to text file] dans le menu [Write] de la fenêtre « **Praat objects** » et aura alors une extension *.IntervalTier*.
- la touche [Extract part...] permet d'extraire une partie du fichier *.TextGrid* en donnant des indications temporelles sur la partie à extraire (Début et fin).

Outre ces opérations d'ajout et d'extraction de tires, il est possible de regrouper des tires dans un même fichier *.TextGrid* plusieurs tires venant de fichiers différents. Cette opération peut être très utile dans certaines situations de travail. Admettons par exemple que plusieurs personnes travaillent séparément sur un même projet et ont chacune des tâches de transcription et d'annotation distinctes à partir d'un même fichier audio : un individu peut faire la transcription orthographique du fichier audio après avoir fait une segmentation en tours de parole, un autre peut annoter les découpages en groupes prosodiques et faire une segmentation en groupes intonatifs, etc. Ces différents fichiers d'annotation peuvent être sauvegardés de deux façons distinctes :

- soit comme un fichier *.TextGrid* ;
- soit comme un fichier *.IntervalTier*, si l'annotateur a extrait la tire d'un fichier *.TextGrid*.

Dans tous les cas, les fichiers séparés peuvent être réunis pour former un nouveau fichier *.TextGrid*. Pour ce faire, il faut ouvrir les fichiers à regrouper (*.TextGrid* ou *.IntervalTier*) avec la commande [Read from file] dans le menu

[Read] de la fenêtre « **Praat Objects** ». Ils apparaissent alors dans la liste des objets et peuvent être sélectionnés avec la souris. Plusieurs cas de figures sont alors à envisager dans les regroupements :

1. Si les deux fichiers sélectionnés et apparaissant en surbrillance sont des fichiers *.TextGrid*, le regroupement s'effectue avec la touche [Merge] qui apparaît à droite de la liste des objets dans la rubrique « Synthesize ». (cf. fig.8).



Fig. 8 : Regroupement de *.TextGrid*

2. Si les deux fichiers sélectionnés et apparaissant en surbrillance sont des fichiers *.IntervalTier*, le regroupement s'effectue avec la touche [Into TextGrid] qui apparaît à droite de la liste des objets dans la rubrique « Collect ». (cf. fig.9).



Fig. 9 : Regroupement de *.IntervalTier*

3. Si l'un des fichiers sélectionnés et apparaissant en surbrillance est un fichier *.TextGrid*, et l'autre un fichier *.IntervalTier*, le

regroupement s'effectue avec la touche [Append] qui apparaît à droite de la liste des objets (cf. fig.10).



Fig. 10 : Regroupement de fichiers

Le résultat des opérations de regroupement est toujours un fichier *.TextGrid* qui apparaît en surbrillance dans la liste des objets. Il peut être sauvegardé et renommé à l'aide de la touche [Write to Text file] dans le menu [Write].

3.1.2 Créer, modifier, extraire des tires dans une fenêtre « Textgrid »

Certaines des opérations présentées dans la section 4.1.1 peuvent être faites directement dans la fenêtre TextGrid, à partir du menu [Tier]. Ces opérations sont synthétisées dans le tableau ci-après :

Touche	Fonction	Mode d'emploi
[Copy tier to list of object...]	Permet d'extraire la tire active (donc en jaune) et de la faire apparaître dans la liste des objets avec le nom « <i>IntrevalTier X</i> »	-

[Add interval tier...]	Permet la création d'une nouvelle tire formée d'intervalles.	Une boîte de dialogue apparaît pour demander la position et le nom de la tire à créer.
[Add point tier...]	Permet la création d'une nouvelle tire formée d'une succession de points.	Une boîte de dialogue apparaît pour demander la position et le nom de la tire à créer.
[Duplicate tier...]	Permet de créer une nouvelle tire en dupliquant la tire active.	Une boîte de dialogue apparaît pour savoir quelle position doit prendre la nouvelle tire et comment elle doit s'appeler.
[Rename tier]	Permet de renommer une tire.	Une boîte de dialogue, indiquant le nom actuel de la tire à renommer, apparaît à l'écran et doit être remplie par l'utilisateur.
[Remove all text from tier]	Permet de supprimer l'ensemble du texte saisi dans la tire active, sans supprimer la segmentation en intervalles.	
[Remove tier...]	Permet de supprimer la tire active.	

3.2 Que peut-on faire d'autre avec les fichiers d'annotation ?

Il est possible avec le logiciel PRAAT d'effectuer d'autres tâches à partir des fichiers d'annotation, notamment :

- extraire des informations sur les étiquettes, les intervalles, etc. ;
- utiliser les segmentations pour créer des sous-fichiers correspondant aux intervalles.

Dans cette section, nous nous limitons à présenter les fonctionnalités directement accessibles à partir de la fenêtre « **Praat objects** ». Mais les utilisateurs peuvent aussi écrire des scripts pour effectuer certaines tâches (extraction de fichiers ou d'intervalles, etc.) ou extraire des informations. Nous n'aborderons pas ce point dans ce chapitre, mais renvoyons le lecteur intéressé au manuel Praat et au [Scripting tutorial] accessible à partir du menu [Help] de la fenêtre « **Praat objects** ».

3.2.1 Faire des requêtes et extraire des informations à partir des fichiers d'annotation

Lorsqu'un fichier *.TextGrid* est sélectionné et apparaît en surbrillance, il est possible d'effectuer plusieurs types de requêtes en cliquant sur des touches accessibles à partir du menu [Query] dans la fenêtre « **Praat objects** ». Le tableau ci-après présente ces requêtes :

Types de requêtes	Touches	Description de la fonction	Modalités
Requêtes sur les "tires »	[Get number of tier]	Recherche sur le nombre total de tires	Une boîte de dialogue s'affiche et demande à l'utilisateur de donner le numéro de la tire à considérer
	[Get tier name]	Recherche sur le nom d'une tire	
	[Is interval tier]	Recherche sur la nature d'une tire (Est-elle une tire d'intervalles ?)	
Requêtes sur les tires constituées d'intervalles	[Get number of interval]	Recherche sur le nombre d'intervalles d'une tire	L'utilisateur doit indiquer dans une boîte de dialogue le numéro de la tire à étudier.
	[Get starting point]	Recherche la valeur temporelle au début d'un intervalle	Il faut indiquer dans une boîte de dialogue le numéro de la tire et de l'intervalle à prendre en considération.
	[Get end point]	Recherche la valeur temporelle à la fin d'un intervalle	
	[Get label of interval]	Recherche sur l'étiquette assignée à un intervalle	
	[Get interval at time]	Recherche quel intervalle se situe à un moment précis	Une boîte de dialogue s'ouvre et demande d'indiquer le « moment »

Requêtes sur les tires constituées d'une succession de points	[Get number of points]	Donne le nombre de points présents sur une tire	Le numéro de la tire à étudier doit être donné dans une boîte de dialogue.
	[Get time at point]	Recherche la valeur temporelle correspondant à un point.	Les numéros de la tire et du point à prendre en considération doivent être indiqués.
	[Get label at point]	Recherche l'étiquette correspondant à un point.	
Requêtes sur des étiquettes	[Count label]	Décompte le nombre d'intervalles ou de points ayant une étiquette particulière.	L'utilisateur doit donner le nom de l'étiquette et le numéro de la tire à prendre en considération.

Les résultats des requêtes s'affichent toujours dans une fenêtre appelée « **Info** ».

3.2.2 Eclater les fichiers à partir des segmentations en intervalles

Les segmentations effectuées et les étiquettes assignées aux intervalles peuvent être utilisées pour découper le fichier sonore en plusieurs fichiers plus petits et plus faciles à manier pour des analyses acoustiques plus précises. Admettons par exemple que le but d'une recherche soit d'étudier la prosodie associée à certains types d'énoncés dans un corpus de parole spontanée. Lors de la phase d'annotation, les transcripateurs ont sans aucun doute créé une tire dans laquelle une segmentation en énoncés a été effectuée et une série d'étiquettes correspondant aux types d'énoncés a été assignée aux intervalles. Il est alors possible d'extraire du fichier audio plusieurs fichiers audio plus petits correspondant aux types d'énoncés à étudier. Avant d'expliquer comment procéder, il est important de noter que ces fonctionnalités ne sont accessibles que si le fichier audio a été ouvert à l'aide de la commande [Read from file], et non [Open long sound file]. Dans certains travaux de recherche à partir de corpus, il peut donc être préférable de segmenter les fichiers audio en sous-fichiers n'excédant pas trois minutes. Certaines tâches de segmentation pourront alors être faites plus aisément.

Lorsque le fichier son et le fichier *.TextGrid* qui lui est associé ont été ouverts à l'aide de la touche [Read from file], ils figurent dans la liste des objets de la fenêtre « **Praat objects** ». L'utilisateur peut donc les sélectionner avec la souris de façon à ce que les deux « objets » soient en surbrillance. Sur la droite

apparaissent alors deux touches (cf. fig.11) : [Extract all intervals...] et [Extract intervals...]

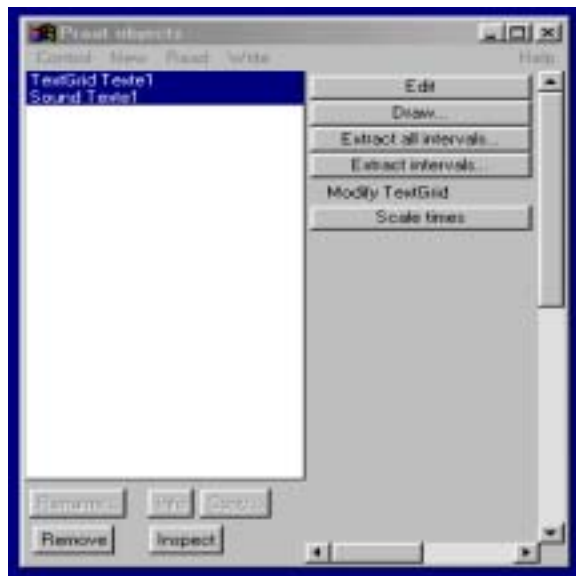


Fig.11 : Segmentation en sous-fichiers

Si l'utilisateur clique sur la touche [Extract all intervals], une boîte de dialogue s'affiche à l'écran et demande le numéro de la tire dont il faut extraire les intervalles. Une fois que l'utilisateur a rempli la boîte de dialogue, l'opération d'extraction est engagée : les fichiers son correspondant à chacun des intervalles de la tire sont placés dans la liste des objets de la fenêtre « **Praat objects** ». Ces fichiers peuvent ensuite être utilisés comme tout autre fichier audio pour effectuer des analyses. Ils peuvent également être sauvegardés.

Si la touche [Extract intervals...] est choisie par l'utilisateur, une boîte de dialogue s'affiche également (cf. fig. 12).

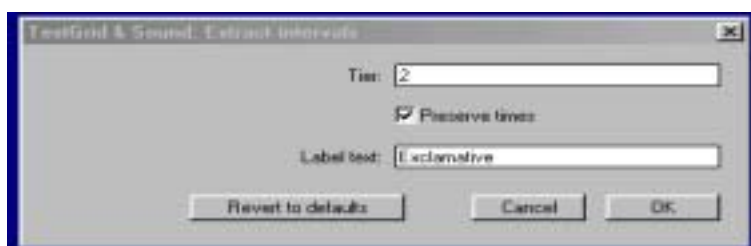


Fig. 12 : Boîte de dialogue pour extraire des intervalles

Dans cette boîte doivent être inscrits le numéro de la tire à laquelle appartiennent les intervalles et l'étiquette associée à ces intervalles. Dans la fig. 12, la boîte de dialogue a été remplie de façon à extraire tous les énoncés exclamationnels d'un fichier son, une tire « *Type d'énoncés* » ayant été créée par l'annotateur. Une fois la boîte de dialogue remplie et la touche [OK] sélectionnée, les intervalles sont extraits automatiquement et placés dans la liste des objets (cf. fig. 13).

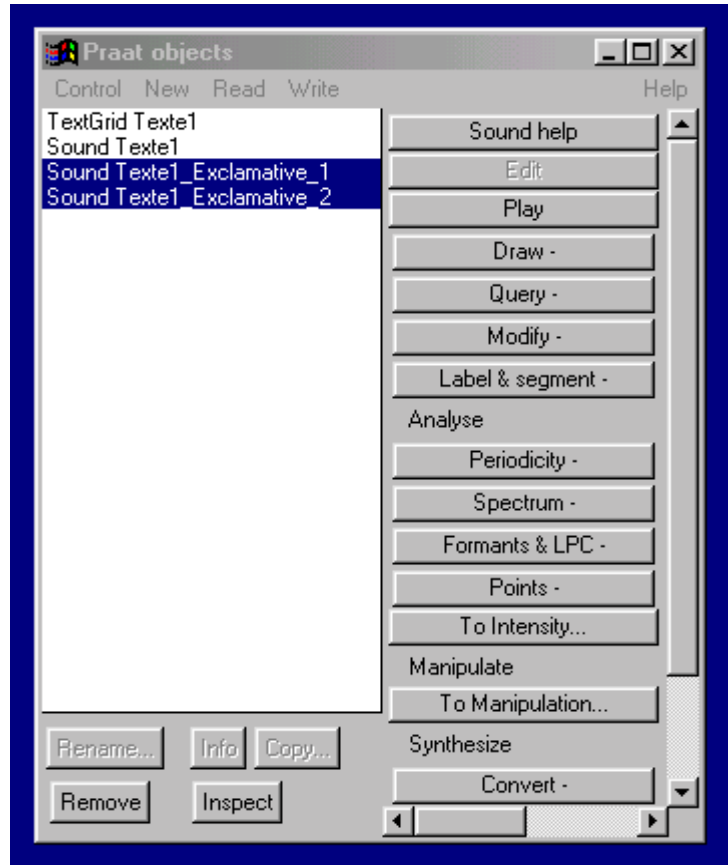


FIG. 13 : « Praat objects » après extraction des exclamatives

Les fichiers son des intervalles ayant pour étiquette « *Exclamative* » figure dans la liste des objets en surbrillance. Ils peuvent être sauvegardés (cf. menu [Write]) et analysés acoustiquement ou prosodiquement. Comme il s'agit des seuls fichiers « son », la transcription correspondant aux intervalles (et donc aux sous-fichiers) n'est pas donnée lors de l'édition du son. En revanche, elle peut être obtenue facilement à partir du fichier *.TextGrid* associé au fichier son dont ont été extraits les sous-fichiers. Pour ce faire, voici comment procéder :

1. Lors de l'extraction des intervalles, l'utilisateur doit impérativement sélectionner la case « *Preserve times* » (cf. fig. 12) ;
2. Après avoir sélectionné un des fichiers son extraits automatiquement, l'utilisateur peut obtenir le temps de début et de fin de ce fichier à l'aide des touches [Get starting time] et [Get finishing time] dans le menu [Query] sur la droite de la fenêtre « **Praat objects** ».
3. Une fois les temps notés, l'utilisateur peut sélectionner le fichier *.TextGrid* dans la liste des objets. Celui-ci apparaît alors en surbrillance. Ensuite, un extrait du fichier *.TextGrid* peut être obtenu en appuyant sur la touche [Extract part...] à droite de la fenêtre « **Praat objects** ». Une boîte de dialogue s'affiche et demande les temps de début et de fin de la partie que l'on souhaite

extraire. Il suffit d'y indiquer les temps obtenus lors de l'opération 2.

4. Cet extrait de *TextGrid* et le sous-fichier son peuvent alors être sélectionnés et édités, comme cela a été expliqué précédemment (cf. § 3). Pour uniformiser les informations temporelles, il peut être nécessaire de modifier le fichier *.TextGrid* extrait à l'aide de la touche [Scale times] qui apparaît sur la droite de la fenêtre « **Praat objects** ».

Les fichiers son extraits et les fichiers d'annotation qui leur sont associés, qui ont également été extraits, peuvent être sauvegardés (et renommés) comme tout autre fichier.

Conclusion

Dans ce chapitre, nous avons présenté plusieurs fonctionnalités de PRAAT qui sont utiles dans le cadre de tâches d'annotation et de transcription. Nous avons principalement insisté sur les fonctions suivantes :

- la saisie des données audio à partir du logiciel ;
- la création de fichiers de transcription et d'annotation. Pour ce point, nous avons principalement conçu les tâches de transcription et d'annotation comme des tâches de segmentation et d'assignation d'étiquettes à des intervalles ;
- la manipulation des annotations et des transcriptions, que ce soit pour modifier et enrichir les annotations, pour effectuer des requêtes sur les fichiers d'annotation, pour extraire des informations ou pour créer des sous-fichiers audio.

Le logiciel PRAAT offre bien d'autres possibilités que nous n'avons pas évoquées et qui peuvent être utiles, même pour les tâches d'annotation et de transcription. Pensons par exemple aux fonctions d'analyse de la fréquence fondamentale et de stylisation de la courbe : une annotation prosodique par points cibles peut en effet être élaborées à partir de ces fonctions. Mais, comme nous nous étions fixé pour objectifs la présentation détaillée de fonctions de PRAAT directement en relation avec des tâches d'annotation et de transcription, nous avons laissé de côté toutes les tâches d'analyse. L'utilisateur peut obtenir des informations sur ces points dans, entre autres, Lieshout (2001), Boersma (2001)

Références

Boersma, P. (2001) « Praat, a system for doing phonetics by computer » in *Glott International* 5(9/10). Pages 341-345. Avec une préface de Vincent van Heuven.

Boersma, P. et D. Weenink (1996) «Praat, a System for doing Phonetics by Computer, version 3.4. » in *Institute of Phonetic Sciences of the University of Amsterdam, Report 132*. 182 pages. Téléchargeable à partir de www.praat.org.

Lieshout van, P. (2001) «*PRAAT Workshop : A basic introduction* ». Téléchargeable à partir de www.praat.org.

