

SOMMAIRE

Jacques Durand, Bernard Laks, Chantal Lyche	
Introduction	3
Jacques Durand, Chantal Lyche, Bernard Laks	
Synopsis du projet PFC, la Phonologie du Français Contemporain : Usages, Variétés et Structure	5
Jacques Durand, Chantal Lyche, Bernard Laks	
Protocole d'enquête (mai 2002)	7
Elisabeth Delais-Roussarie, Jacques Durand, Chantal Lyche, Abderrahim Meqqori et Jean-Michel Tarrier	
Transcription des données : outil et conventions (15 avril 2002)	21
Jacques Durand, Bernard Laks, Chantal Lyche	
Directions d'analyse (juin 2002)	35
Jacques Durand, Bernard Laks, Chantal Lyche	
Format des rendus 2002 et 2003	71

INTRODUCTION

Le but de ce bulletin est de permettre de mieux coordonner le projet PFC en parallèle avec l'utilisation de notre site internet, du courrier électronique, des réunions et des colloques que nous organisons. Dans ce premier numéro du bulletin, le lecteur trouvera tous les textes fondamentaux pour les enquêtes et les analyses PFC : le protocole d'enquête, les conventions pour les transcriptions sous PRAAT, les directions d'analyse et le format des rendus.

Le volume 2 sera consacré aux outils PFC et aux manuels qui les accompagnent : le comparateur, le classeur-schwa, le classeur-liaison, Transpraat. Il inclura une copie CD de ces utilitaires (pour utilisateurs PC et MAC).

Les volumes suivants seront principalement consacrés aux résultats préliminaires des enquêtes. Nous ne visons pas à remplacer des publications dans des supports traditionnels (revues à comité de lecture, ouvrages dans des maisons d'édition). Il faut cependant reconnaître que dans une enquête aussi importante que la nôtre les stades intermédiaires sont essentiels. Nous accueillerons donc avec intérêt des contributions sur les enquêtes en cours chaque fois que l'avancement des travaux en justifiera la diffusion.

Jacques Durand, Bernard Laks, Chantal Lyche

Synopsis du projet PFC

LA PHONOLOGIE DU FRANÇAIS CONTEMPORAIN :

Usages, Variétés et Structure

« La phonologie du français contemporain (PFC) : usages, variétés et structure » est un projet international sous la direction de Jacques Durand (ERSS-UMR5610, Université de Toulouse-Le Mirail), Chantal Lyche (Université d'Oslo) et Bernard Laks (Université de Paris X).

PFC est un projet qui vise à décrire la prononciation du français dans sa diversité géographique, sociale et stylistique. A partir d'un protocole d'enquête uniforme, un groupe international d'une trentaine de chercheurs et leurs étudiants est impliqué dans la constitution d'un vaste corpus de français parlé à travers le monde. Grâce à des méthodes d'analyse et à des outils communs, le projet a pour ambition d'offrir une vision globale de la phonologie du français.

Le projet PFC part de la constatation qu'il est nécessaire de poursuivre le travail de description entrepris depuis au moins un siècle par tous les spécialistes de la communication parlée pour (1) fournir une meilleure image du français parlé dans son unité et sa diversité ; (2) mettre à l'épreuve les modèles phonologiques et phonétiques sur le plan synchronique et diachronique ; (3) favoriser les échanges entre les connaissances phonologiques et les outils du traitement automatique de la parole ; (4) permettre la conservation d'une partie importante du patrimoine linguistique des espaces francophones du monde, et ce en contrepoint aux corpus déjà constitués ; (5) encourager un renouvellement des données et des analyses pour l'enseignement du français.

Pour réaliser ces différents objectifs, les diverses équipes de PFC cherchent à couvrir un minimum de 30 points d'enquête dans l'espace francophone international. Chaque enquête est réalisée à partir d'un protocole commun soumis à des groupes d'environ une douzaine de personnes qui sont enregistrées de façon à mettre en évidence divers paramètres de variation sociolinguistique (lecture de mots, lecture de texte, conversation guidée et conversation spontanée). Pour chaque point d'enquête, les enquêtés sont répartis en deux groupes égaux d'hommes et de femmes, ils doivent être bien enracinés dans la communauté linguistique décrite, et ils correspondent à deux tranches d'âge au minimum. Toutes les enquêtes nous fournissent donc trois paramètres de base (lieu, sexe, âge) auxquels viennent s'ajouter d'autres paramètres incontournables (en particulier, profession et niveau d'éducation). Les enquêteurs doivent bien connaître les locuteurs pour leur demander de se plier aux « exercices » que comporte l'enquête : lecture à haute voix de la liste de 94 mots, lecture du

passage (une page), entretien guidé (une vingtaine de minutes par locuteur) et conversation libre (une trentaine de minutes par locuteur). L'enquêté doit être prêt à nous accorder entre une heure et une heure et demie de son temps.

Nous sommes heureux d'accueillir des projets d'enquête (maîtrises, DEA, doctorats, chercheurs) dans le cadre de PFC. Nous fournissons aux enquêteurs un ensemble d'outils pour les guider : un protocole détaillé et conçu pour une utilisation par des étudiants sans formation linguistique poussée, une introduction à nos techniques de transcription à l'aide d'un logiciel convivial Praat (mis au point par Paul Boersma, Université d'Amsterdam), et des directions d'analyse. Pour chaque point d'enquête et pour chaque locuteur, nous transcrivons *orthographiquement* une dizaine de minutes de conversation formelle et une dizaine de minutes de conversation informelle. Cette transcription orthographique nous fournit un point de départ utile pour les analyses plus techniques effectuées par la suite.

Dans le cadre du projet, nous privilégions dans un premier temps la description des systèmes phonémiques, la liaison et le 'e' muet. Nous avons mis au point divers utilitaires visant à aider le travail d'analyse : en particulier *Transpraat*, un logiciel qui convertit les fichiers de Praat en fichiers texte, un *Classeur-schwa* et un *Classeur-liaison* qui aident à comptabiliser et classer les (non-)occurrences de schwas/liaisons à partir d'une notation numérique mise au point à partir de la transcription orthographique et un *Compareteur*, outil qui permet de comparer la prononciation de tel ou tel mot de la liste par les locuteurs des diverses enquêtes.

Il est important de souligner que nos enquêtes se prêtent à des études phonologiques et phonétiques dans des domaines autres que ceux que nous privilégions comme base de départ (par exemple, accentuation, intonation, phénomènes phonétiques fins comme les assimilations de voisement ou la durée de segments phonétiques donnés). Rien n'empêche un chercheur de compléter l'enquête par d'autres éléments (par exemple, liste complémentaire de mots ou tests en laboratoire si les enquêtés s'y prêtent). Par ailleurs, nos conversations permettent une utilisation dans des champs autres que la phonologie : par exemple, syntaxe du français oral, étude de particules énonciatives, tours de parole.

La volonté du projet PFC est de créer une base de données comportant entre 300 à 500 locuteurs étudiés à partir d'une méthodologie commune et permettant des comparaisons intéressantes sur un échantillon représentatif de variétés du français contemporain.

Jacques Durand, Bernard Laks, Chantal Lyche

PROTOCOLE D'ENQUÊTE (mai 2002)

Jacques Durand, Chantal Lyche, Bernard Laks

Table des matières

0. Introduction

1. Matériel

2. Note sur les enquêteurs

3. Choix des informateurs

4. Procédure

4.1. Discussion libre

4.2. Entrevue guidée et fiche

4.3. Liste de mots

4.4. Texte

5. Références

6. Documents

6.1. Fiche d'enquête

6.2. Thèmes de discussion suggérés

6.3. Liste de mots

6.4. Texte

7. Questions éthiques

0. Introduction

Le protocole d'enquête présenté dans ce document est celui qui est adopté dans notre projet d'étude des variétés du français : *La phonologie du français contemporain : usages, variétés et structures*. PFC part de la reconnaissance que le français est trop souvent représenté comme une langue homogène et, à tous les niveaux de la description et de l'analyse, une conscience de la variété s'impose. En effet, le français n'est pas uniquement celui de la bourgeoisie

parisienne, croyance trop largement répandue, c'est aussi celui de Bruxelles, Lafayette, Genève, Moncton, Québec, Toulouse, Tours, et des autres régions francophones du monde.

Ce protocole est aussi peu technique que possible pour être utilisable par les étudiants qui participent au projet. Il doit être fidèlement reproduit dans toute enquête PFC. L'utilisation des mêmes méthodes et des mêmes textes permet un travail comparatif fiable. Il est donc préférable d'ajouter des éléments que de changer le texte (voir 6.4) et la liste de mots (voir 6.3). L'enquêteur n'est évidemment pas limité à ces données et, s'il en éprouve le besoin, il proposera un texte ou des mots supplémentaires. Nous lui demandons cependant de nous faire parvenir ce nouveau texte et/ou ces mots afin de les rendre accessibles aux autres enquêteurs.

1. Matériel

Il est préférable d'avoir un magnétophone de haute qualité. Nous-mêmes, nous avons utilisé un DAT, facilement transportable ainsi qu'un magnétophone de type minidisc. Les données sont digitales dans les deux cas et la qualité du son est tout à fait comparable, même si les magnétophones de type minidisc présentent une plus grande compression des données. Dans la mesure où les minidisc sont beaucoup moins chers que les DAT, certaines équipes leur ont accordé leur préférence mais il faut signaler que leur sortie est habituellement analogique.

Pour un enregistrement de dialogues, deux micros reliés par une prise Y peuvent être utilisés. Afin de garantir la qualité du son, on évitera d'effectuer les enregistrements près d'une fenêtre, dans une pièce où de nombreux bruits de fond créent des interférences. Il faut impérativement s'abstenir de bouger le micro lors de l'enregistrement. Un micro-cravate peut se révéler utile. Pour une bonne écoute, il peut être utile de brancher une paire de baffles directement sur le DAT ou de relier le DAT à une chaîne hi-fi. Il est évident que l'enquêteur doit être familiarisé avec tout son équipement avant de commencer les interviews.

Nous recommandons aux novices de se renseigner auprès du centre audiovisuel de leur université qui, de surcroît, sera souvent à même de prêter du matériel pour une durée limitée. Soulignons qu'il est possible d'effectuer de très bons enregistrements avec des magnétophones classiques qui sont souvent plus robustes que le matériel digital. Un complément d'information sur ces questions sera inclus sur le site internet PFC.

2. Les enquêteurs

Dans la mesure du possible, nous conseillons aux enquêteurs de travailler en équipe de deux dont l'un des membres sera un proche de l'enquêté, l'autre membre étant introduit comme un ami, un ami d'un ami, d'un parent, etc. Voir à ce sujet Milroy (1980). Lorsque vous fonctionnez à deux enquêteurs, l'entrevue « guidée » sera dirigée par l'enquêteur le plus éloigné de l'informateur et une période de 'familiarisation' est recommandable avant d'installer le magnétophone. En revanche, c'est l'enquêteur proche des enquêtés qui prendra la relève lors des discussions informelles. Une solution possible lorsque l'enquêteur travaille seul avec un groupe donné est de mener les discussions « guidées » et de laisser les enquêtés

seuls avec le magnétophone pour les conversations libres. Noter cependant qu'il faut très bien connaître la variété étudiée puisqu'en étant présent lors des discussions on facilite la compréhension des échanges à transcrire. Dans le cas, où l'enquêteur est seul et impliqué dans les deux types de conversation, il marquera une pause entre les deux phases de manière à bien les différencier.

Qu'il y ait un ou plusieurs enquêteurs, il est impératif que ce(s) dernier(s) ai(en)t une très bonne connaissance du terrain d'enquête sans quoi le protocole risque d'être inutilisable.

Tout enregistrement pose des questions éthiques importantes que nous abordons en 7.

3. Choix des informateurs

Les études préliminaires sont maintenant achevées. Dans cette phase finale, nous visons à un quadrillage aussi large que possible des variétés du français. Mis à part des sous-projets spécifiques qui peuvent avoir une couverture plus large ou plus restreinte, nous demandons aux enquêteurs de constituer des groupes d'une douzaine de locuteurs. Rappelons que l'enquêteur principal doit être très à l'aise dans le milieu où il opère. Soumettre les enquêtés à une épreuve de lecture exige une grande confiance de la part de ces derniers.

Nos moyens ne nous permettent pas une enquête sociolinguistique au sens fort du terme pour chaque point étudié. Le premier critère est que les locuteurs aient un bon ancrage dans la communauté de référence. Toutes choses égales par ailleurs, si l'on étudie Bordeaux, on choisira de préférence des locuteurs qui ont vécu toute leur vie ou presque dans cette ville. L'idéal est que tout locuteur sélectionné ait fait ses études primaires dans la région où se déroule l'enquête. En pratique, nous sommes conscients du fait que le travail avec un réseau dense peut présenter des sujets ne correspondant que partiellement à cet idéal. Sur le plan social, les tests de lecture exigent des locuteurs relativement bien scolarisés ce qui exclut certains milieux défavorisés. Si l'enquêteur a la possibilité d'obtenir une différence sociale à l'intérieur de son groupe, ce qui n'est pas toujours possible, il saisira cette opportunité. En revanche, il est crucial que le groupe étudié comporte un nombre égal d'hommes et de femmes répartis en deux tranches d'âge au minimum.

4. Procédure

Le but de l'enquête est de rassembler un échantillon de variétés de français et la procédure détaillée ci-dessous nous permet d'avoir accès à la variation individuelle. L'ensemble des enregistrements permet en effet une étude de plusieurs registres chez le même locuteur :

- dans les entretiens, un français soutenu et parfois un français familier
- dans les dialogues, un français familier
- dans la lecture, un français très soutenu

Il s'ensuit que les trois volets de la procédure d'enquête doivent impérativement être respectés : dialogue, lecture de texte, lecture de la liste de mots. Le dialogue lui-même se scinde en deux parties : une discussion libre de groupe et une entrevue guidée (plus fiche à remplir). L'ordre dans lequel se succèdent ces diverses tâches dépend de la disponibilité des enquêtés et ne peut pas être fixé *a priori*.

4.1 Discussions libres

Pour avoir accès au vernaculaire de nos sujets et pour tenter d'atténuer les effets de ce que Labov appelle le " paradoxe de l'observateur ", nous recommandons de jouer sur des liens de connaissance qui permettent d'étudier des réseaux denses sur le plan interpersonnel. Nous faisons nôtre le conseil suivant que donne Bourdieu dans un ouvrage connu :

On a ainsi pris le parti de laisser aux enquêteurs la liberté de choisir les enquêtés parmi des *gens de connaissance* ou des gens auprès de qui ils pouvaient être introduits par des gens de connaissance. La proximité sociale et la familiarité assurent en effet deux des conditions principales d'une communication "non violente". (Bourdieu, 1993 : 1395)

Cette méthode a été celle suivie par Labov dans certains de ses travaux sur le Lower East Side de New York et par les Milroy à Belfast. Nous recommandons autant que possible le travail à deux : un chercheur qui connaît le groupe de façon intime et un autre enquêteur qui le connaît moins bien ou pas du tout et dont la distance par rapport au groupe pourra dans certains contextes provoquer un style plus soutenu (voir entrevue guidée). Le travail à deux est un garde-fou utile. On travaille d'autant mieux sur des enregistrements qu'on a été présent et les deux enquêteurs se compléteront utilement pour les transcriptions et la phase d'analyse.

On notera qu'il est désirable d'avoir entre 20 mn et 30 mn de tours de parole spontanée par enquêté puisque nous demandons des fichiers et une transcription orthographique d'un minimum de 20 mn par locuteur correspondant aux fichiers audio pour chacun des deux styles conversationnels (10 mn entrevue guidée, 10 mn conversation en groupe).

L'expérience du projet nous enseigne qu'au-delà de trois personnes, les enregistrements sont difficilement exploitables. Les conversations libres doivent donc se limiter à deux ou trois personnes au plus. Ce qui différencie la discussion libre et la discussion guidée c'est que dans le premier cas, il n'y a aucune asymétrie de rôle. L'enquêteur, s'il est présent, est un simple membre du groupe. Dans le deuxième cas, c'est un enquêteur.

4.2. Entrevue guidée et fiche

L'entrevue guidée met en présence l'enquêteur et l'enquêté. Sa fonction est double. Elle permet d'abord la collecte d'informations sur les enquêtés, d'où la liste de questions qui aide l'enquêteur à compléter la fiche signalétique qui sera remplie après l'entretien. Elle fournit

également des enregistrements de bonne qualité dans un registre plus soutenu qu'une conversation à deux ou à trois entre amis. Pour aider l'enquêteur, nous suggérons en 6.2 des thèmes de discussion possibles et quelques techniques simples pour éviter des réponses monosyllabiques. Signalons qu'il y a des informations essentielles qu'il faut absolument obtenir au sujet des enquêtés à travers la discussion guidée et la fiche (par exemple, le locuteur a-t-il fait de longs séjours en dehors de sa région, ses parents étaient-ils de langue française, etc.).

4.3 Liste de mots

La liste de mots que nous proposons contient 94 mots et doit être commune à tous les groupes. Chaque enquête ajoutera les exemples qui semblent essentiels pour la variété étudiée. Attention, néanmoins, des listes trop longues peuvent décourager les volontaires.

Cette liste, qui sera reproduite avec une police facilement lisible, comprend deux parties : une liste où l'ordre des mots est aléatoire et une dizaine de paires minimales. On demandera à l'informateur de lire aussi le chiffre précédant le mot. Cette lecture est assez fastidieuse, mais aucun de nos informateurs n'a réclamé de pause alors que nous avons suggéré une interruption possible tous les 25 mots. Les dix derniers mots de la liste sont des paires minimales et cette phase de la lecture éveille la conscience linguistique des témoins. Il est alors recommandé de laisser le magnétophone tourner pendant que le locuteur s'exprime sur sa façon de parler, sur la norme d'usage, etc. Par ailleurs, il est utile pour éviter des froissements de feuille de mettre les documents à lire sous plastique.

4.4 Texte

On proposera à l'informateur une lecture silencieuse avant la lecture à haute voix afin qu'il se familiarise avec le texte et ne trébuche pas sur les mots. Il va sans dire que l'enquêteur ne présentera cette tâche qu'à un sujet raisonnablement scolarisé. Le texte sera reproduit avec double espace et une police de taille égale ou supérieure à 14 afin de faciliter la lecture. L'enquêteur devra veiller à ce que l'alignement à droite du texte n'exerce pas une influence sur la présence ou l'absence de liaisons et de schwas.

Pour ce qui est de la lecture du texte et de la liste de mots, l'enquêteur devra s'assurer avant l'enquête que cet exercice ne posera pas de problème.

5. Références

Ces quelques ouvrages permettront une familiarisation de l'enquêteur avec le travail qu'il est sur le point d'entreprendre.

- Pour la méthodologie de l'enquête, les ouvrages dans la collection 128 permettront d'éviter les pièges les plus courants. (Voir aussi Milroy 1980).

Blanchet, A. et A. Gotman. 1992. *L'enquête et ses méthodes : l'entretien*. Paris : Nathan Université.

Singly, F. de. 1992. *L'enquête et ses méthodes : le questionnaire*. Paris : Nathan Université.

Nous conseillons aussi la lecture des Documents (chap. 5) de Walter 1982

Walter, H. 1982. *Enquête phonologique et variétés régionales du français*. Paris : PUF.

- Dans la mesure où la première étape de nos travaux est une duplication de l'enquête menée par Durand, Slater et Wise en 1986, nous suggérons que l'enquêteur se familiarise avec les résultats de ce travail.

Durand, J., C. Slater et H. Wise, 1987. Observations on schwa in Southern French. *Linguistics* 25.5, 983-1004.

- Notre travail se situe dans le cadre des travaux de Labov et de Milroy et nous conseillons la lecture de :

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia : University of Pennsylvania Press. Traduction française, *Sociolinguistique*. Paris : Editions de Minuit, 1976.

- Avant d'entreprendre l'enquête, la lecture de l'ouvrage suivant est indispensable :

Milroy, J. 1980. *Language and social networks*. Oxford : Blackwell.

- Le lecteur trouvera une synthèse de diverses approches possibles en sociolinguistique et dialectologie dans :

Durand, J. et Lyche, C. (1999) La variation et le linguiste : méthodes et analyses. Domaine français et anglais. *Carnets de grammaire* 4, 63 p.

6. Documents

6.1 Fiche d'enquête

6.2 Questionnaire

6.3 Liste de mots

6.4 Texte

6.1 Fiche d'enquête

Fiche signalétique

NOM, prénom(s) :

Date de naissance :

Lieu de naissance :

Domiciles successifs (en nombre d'années)

.....

Domicile actuel :

.....

Professions successives :

.....

Profession actuelle :

Situation familiale (marié, célibataire, etc.) :

 enfants, âge, scolarisation :

.....

Études (préciser jusqu'à quel âge et quel type d'études) :

.....

Langues parlées :

.....

Père de l'informateur, année de naissance :

 lieu d'origine :

 profession :

 études :

 langues parlées (étrangères ou régionales) :

Mère de l'informateur, année de naissance :

lieu d'origine :

profession :

études :

langues parlées (étrangères ou régionales) :

Époux, épouse, autre :

lieu d'origine :

profession :

études :

langues :

Personnes ayant joué un rôle important au moment de l'apprentissage du français par l'enquêté (grands-parents, nourrice...) :

Type de logement de l'enquêté (maison, appartement...) :

Intégration dans le quartier, relations de voisinage :

Activités culturelles, loisirs, voyages :

Autres informations :

Informations sur l'enquête

Nom de l'enquêteur (entretien guidé) :

Nom de l'enquêteur (entretien libre) :

Date de l'enregistrement, durée :

Lieu de l'enregistrement :

Localité :

Lien enquêteur/enquêté :

Professionnel :

Amical :

Familial :

Autre :

Observations sur l'enregistrement :

Informel : qui étaient les locuteurs présents ?

Ordre des situations dans l'enregistrement (ex. guidé, liste de mots, texte, libre) :

.....

Principaux thèmes abordés :

.....

Qualité de l'enregistrement :

Autres observations : (interventions d'autre personnes, interruptions téléphoniques notables...)

.....

.....

Enregistrement de groupe

Établir une fiche signalétique pour chaque participant et préciser les relations à l'intérieur du groupe.

6.2 Thèmes de discussion suggérés

Nous rappelons que ces thèmes ne sont donnés qu'à titre d'exemple. Avant toute chose, rappelons que les questions doivent être posées de façon à occasionner des développements par les enquêtés. Il faut à tout prix éviter les discussions du type suivant :

A : Êtes-vous né dans la région ?

B : Oui.

A : Vous vous plaisez ici ?

B : Oui.

A : Vos parents étaient-ils originaires de la région ?

B : Oui.

En général, ce sont des questions fermées qui entraînent les non dialogues. On remplacera donc de telles questions par des questions du type :

A : Est-ce que vous voulez bien me parler de votre enfance / des origines de vos parents ?

A : Pourriez-vous me raconter des anecdotes précises de vos premières années à l'école ?

La conversation doit être ouverte. L'utilité de certains thèmes est qu'ils nous aident à dresser un portrait sociolinguistique de la personne. Il est néanmoins important de s'assurer qu'à travers la fiche et l'entretien on obtienne des informations précises sur la date d'arrivée dans la communauté de l'enquêté et de sa famille immédiate, le niveau scolaire des enquêtés et de leurs parents, les longs séjours en dehors de la communauté linguistique visée, les langues étrangères parlées par l'enquêté, ses parents et ses grands-parents.

L'entretien guidée est un exercice périlleux car il doit à la fois laisser le témoin s'exprimer mais aussi permettre de dresser un portrait sociolinguistique détaillé. Les thèmes suivants gagnent souvent à être abordés :

- la scolarité (bons souvenirs, mauvais souvenirs, matières préférés, professeurs qui ont marqué)
- la famille (occupation des parents, des enfants, nombre de frères et sœurs, contacts avec les grands-parents, ancrage de la famille dans la ville ou le village)
- les autres langues parlées par le témoin (niveau de compétence du sujet et de sa famille, attitude à l'égard de ces dernières)
- travail (métier exercé, problèmes économiques, relations avec les collègues de travail)
- déplacements (séjours importants dans d'autres villes, à l'étranger)
- lieu de vie actuel (résidences primaires, secondaires)
- attachement à la région
- loisirs
- événements marquants dans la vie de la personne

6.3 Liste de mots

La liste de mots que nous utilisons comporte 84 mots en ordre aléatoire (mais qui doit être respecté) suivis de paires minimales (de 85 à 94). On ne fera pas de pause entre les mots en ordre aléatoire et les paires minimales.

Liste en ordre aléatoire

Attention de faire précéder la lecture de chaque mot par celle du chiffre placé devant.

1. roc
2. rat
3. jeune
4. mal
5. ras
6. fou à lier
7. des jeunets
8. intact
9. nous prendrions
10. fêtard
11. nièce
12. pâte
13. piquet
14. épée
15. compagnie
16. fête
17. islamique
18. agneau
19. pêcheur
20. médecin
21. paume

22. infect
23. dégeler
24. bêtement
25. épier
26. millionnaire
27. brun
28. scier
29. fêter
30. mouette
31. déjeuner
32. ex-femme
33. liège
34. baignoire
35. pêcheur
36. socialisme
37. relier
38. aspect
39. niais
40. épais
41. des genêts
42. blond
43. creux
44. reliure
45. piqué
46. malle
47. gnôle

48. bouleverser
49. million
50. explosion
51. influence
52. mâle
53. ex-mari
54. pomme
55. étrier
56. chemise
57. brin
58. lierre
59. blanc
60. petit
61. jeûne
62. rhinocéros
63. miette
64. slip
65. compagne
66. peuple
67. rauque
68. cinquième
69. nier
70. extraordinaire
71. meurtre
72. vous prendriez
73. botté

74. patte
75. étriller
76. faites
77. feutre
78. quatrième
79. muette
80. piquais
81. trouer
82. piquer
83. creuse
84. beauté
85. patte
86. pâte
87. épais
88. épée
89. jeune
90. jeûne
91. beauté
92. botté
93. brun
94. brin

6.4 Texte

Le Premier Ministre ira-t-il à Beaulieu ?

Le village de Beaulieu est en grand émoi. Le Premier Ministre a en effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux olympiques de Berlin en 1936, et plus récemment, son usine de pâtes italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand honneur ? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits habituels qui tournaient toujours autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde".

Le maire de Beaulieu - Marc Blanc - est en revanche très inquiet. La cote du Premier Ministre ne cesse de baisser depuis les élections. Comment, en plus, éviter les manifestations qui ont eu tendance à se multiplier lors des visites officielles ? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère. D'un autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout est fait pour le protéger. Or, un gros détachement de police, comme on en a vu à Jonquière, et des vérifications d'identité risquent de provoquer une explosion. Un jeune membre de l'opposition aurait déclaré : "Dans le coin, on est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous avons le soutien du village entier." De plus, quelques articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des activistes des communes voisines préparent une journée chaude au Premier Ministre. Quelques fanatiques auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus à quel saint se vouer. Il a le sentiment de se trouver dans une impasse stupide. Il s'est, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.

7. Questions éthiques

Toute utilisation d'un corpus oral pose des questions éthiques et l'enquêteur ne peut se permettre de prêter son corpus ni d'en publier des extraits sans en avoir demandé l'autorisation à ses locuteurs. L'enquêteur devra s'assurer que le locuteur donne son accord à une utilisation dans un but scientifique, des conversations enregistrées. Il précisera que l'anonymat du locuteur est toujours conservé dans d'éventuelles publications. Cela se fera de préférence en deux temps : premièrement, on demandera au locuteur si les personnes présentes peuvent se servir des enregistrements dans d'éventuelles publications (y compris de matériel sonore) et, deuxièmement, si ces enregistrements peuvent être communiqués à d'autres chercheurs pour leurs propres recherches. Nous recommandons de poser ces questions en fin d'interview lorsque l'enquêté connaît l'ampleur de ses confidences et a acquis une certaine intimité avec le chercheur. Sur un plan strictement légal, des formulaires écrits sont nécessaires pour protéger les chercheurs et les enquêtés. Un modèle possible est suggéré ci-dessous.

Consentement de participation

NOM :

Prénom :

M./Mme X vous propose de participer à une enquête dans le cadre du projet universitaire : “ La phonologie du français contemporain : usages, variétés et structure ”. Ce projet s’appuie sur des enregistrements effectués sur des échantillons de francophones de divers pays du monde.

Ces enregistrements sont faits dans un but de recherche scientifique. Les résultats (y compris les enregistrements) peuvent en être publiés aussi bien dans des revues scientifiques que dans des ouvrages commercialisés (par exemple, manuels pédagogiques). Ils pourront éventuellement être intégrés à des systèmes de production et de perception de la parole. Dans tous les cas de publication, les informations vous concernant seront exploitées de manière anonyme. Vous pourrez par ailleurs avoir accès à toute publication éventuelle si vous en faites la demande.

Votre consentement ne décharge pas les enquêteurs de leur responsabilité. Vous conservez tous vos droits garantis par la loi.

Si vous acceptez de faire partie de l’échantillon d’études dans les conditions énoncées ci-dessus, veuillez signer ce document.

SIGNATURE

TRANSCRIPTION DES DONNEES : OUTIL ET CONVENTIONS (15 avril 2002)

Elisabeth Delais-Roussarie, Jacques Durand
Chantal Lyche, Abderrahim Meqqori et Jean-Michel Tarrier

Table des matières

1. PRAAT : un outil de transcription et d'analyse
 - 1.1 Introduction
 - 1.2 Présentation générale de PRAAT
 - 1.3 Transcrire des données avec PRAAT
 - 1.4 Transpraat
 - 1.5 Gestion des tires sous Praat
2. Transcription orthographique des données
 - 2.1 Transcription orthographique
 - 2.2 Conventions de transcription

Ce document est conçu comme un guide d'aide à la transcription dans le cadre du projet PFC. Il comprend deux volets :

- la présentation de PRAAT, logiciel que nous avons choisi pour la transcription des données ;
- les conventions de transcriptions PFC.

1. PRAAT : un outil de transcription et d'analyse

1.1 Introduction

PRAAT est un programme informatique multi-plateforme qui permet de mener des analyses phonétiques, de faire de la synthèse de la parole et de manipuler des données (analyses statistiques, construction de grammaires, etc.). Notre présentation n'a pas pour but de fournir une description détaillée du logiciel *PRAAT*, mais essentiellement d'aider le néophyte à se familiariser avec cet outil pour la transcription de données orales rassemblées dans le cadre du projet PFC.

1.2 Présentation générale de PRAAT

Le logiciel PRAAT a été développé par Paul Boersma et par David Weenink de l'Institut de Phonétique d'Amsterdam. Il peut être rapatrié à partir du site suivant : <http://fonsg3.let.uva.nl/praat/>. Nous nous référons ici aux versions **3.9.9**, **4.0.5** et **4.0.1**. de ce logiciel.

Avec PRAAT, il est possible :

- d'enregistrer des fichiers audio qui pourront ensuite être analysés ;
- de transcrire, d'étiqueter et de segmenter des données audio (que les enregistrements aient été effectués sous PRAAT ou qu'ils proviennent d'autres fichiers, au format WAV par exemple) ;

- d'effectuer des analyses phonétiques et acoustiques au niveau segmental (spectrogramme, analyse de formants, sonagramme, etc.) et au niveau suprasegmental (pitch [courbe de Fo], intensité et durée) ;
- de manipuler et modifier le signal de parole (utilisation de filtres ; modification des contours intonatifs et de la durée, etc.)
- de faire de la synthèse de la parole (créer des stimuli audio, synthèse articulatoire, analyse-synthèse de données modifiées, etc.) ;
- de construire des outils d'apprentissage (Réseau de neurones et élaboration de grammaires dans le cadre de la théorie de l'optimalité (OT, Optimality Theory) ;
- de faire des analyses statistiques à partir des études phonétiques (analyses de covariances, etc.).

Des informations supplémentaires et un tutorial peuvent être obtenus à partir des sites WEB suivants :

- <http://www.fon.hum.uva.nl/praat/>
- <http://www.ling.lu.se/persons/Sidney/praat/>

Le tutorial est également disponible dans le logiciel lui-même (voir rubrique HELP). Il est imprimable et exportable à volonté.

1.3 Transcrire des données avec PRAAT

1.3.1 Premiers pas sous PRAAT

Bien qu'il soit possible d'effectuer directement des enregistrements avec PRAAT, nous nous attachons ici à décrire comment effectuer la transcription de fichiers déjà numérisés. Il faut savoir que PRAAT accepte divers formats de sons (WAV, AIFF, etc.). Celui pour lequel nous avons opté dans le cadre du projet PFC est le format WAV¹. Les fichiers à traiter seront en général sur votre disque dur ou sur un CD-ROM.

A l'ouverture du logiciel PRAAT, deux fenêtres apparaissent à l'écran :

- *PRAAT Objects* est la fenêtre dans laquelle apparaîtra la liste des objets ou fichiers à partir desquels s'effectueront les analyses.
- *PRAAT Picture* sert à créer des images qui pourront être insérées dans les documents Word (ou autres) : courbes de Fo, Sonagramme, transcription et étiquetage, etc.

Pour effectuer la transcription d'un fichier audio, voici comment procéder :

1. **Ouverture d'un fichier audio :** Pour ouvrir le fichier audio (.wav) dont vous voulez faire la transcription, cliquez sur [READ] dans la fenêtre « Praat Objects ». Dans le menu déroulant, sélectionnez [OPEN LONG SOUND FILE]. Une boîte de dialogue vous sera présentée et vous y indiquerez le nom et l'emplacement du fichier audio à ouvrir. Le nom du fichier audio sera alors indiqué dans la fenêtre « Praat Objects » et y apparaîtra en surbrillance.

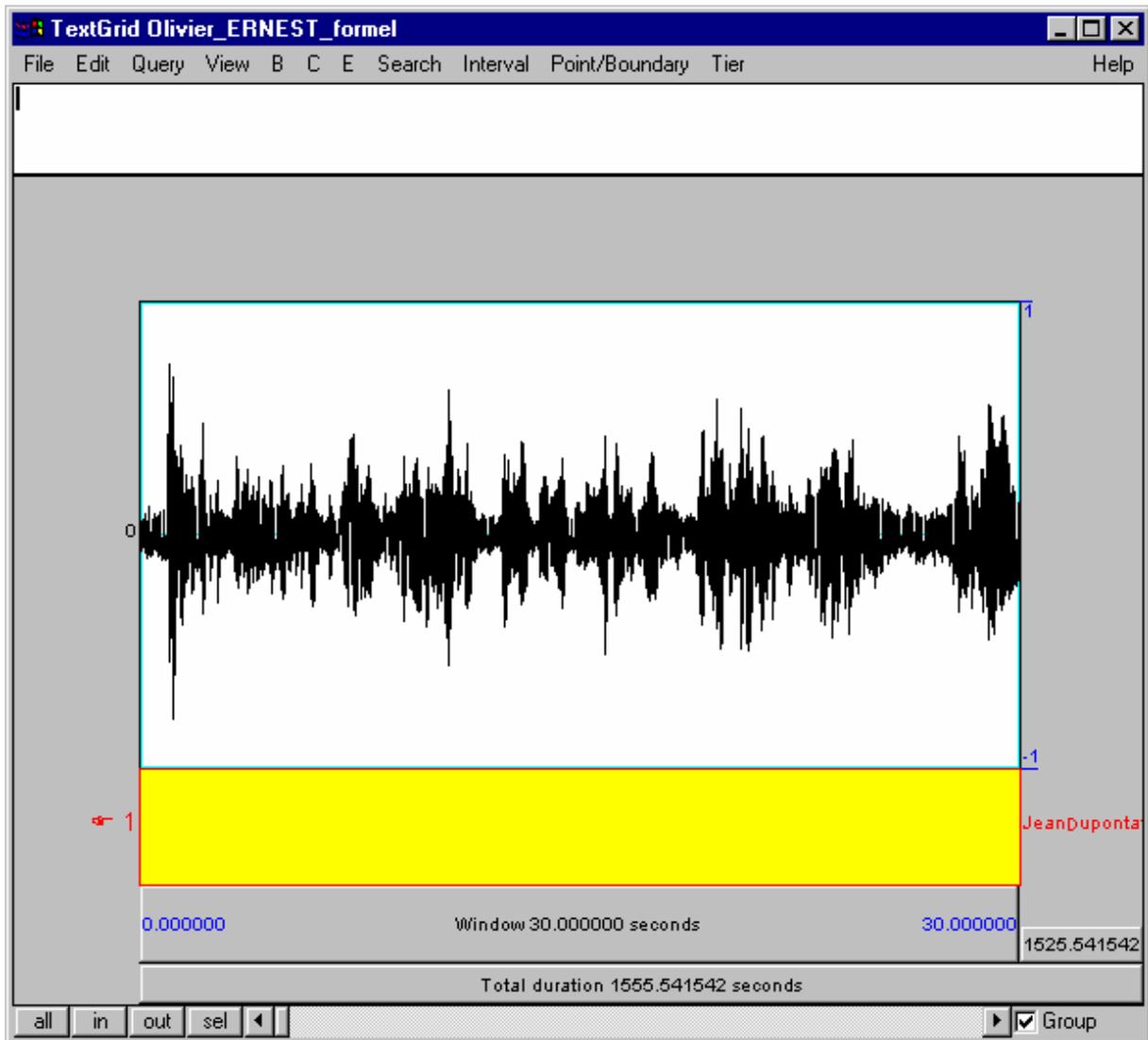
¹ Différents logiciels permettent de numériser les enregistrements audio sous le format WAV : par exemple COOL EDIT ou WAVE STUDIO, ce dernier étant fourni avec les cartes son Sound Blaster. Nous prenons ici le terme numériser au sens large qui inclut soit le transfert de fichiers numériques obtenus à partir de DAT, soit la conversion de fichiers son analogiques obtenus à partir de bandes magnétiques.

2. **Création d'un fichier texte pour les annotations/transcription** : Pour créer un fichier texte qui sera lié au fichier son ouvert, cliquez sur le bouton [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 indiquant [BELL]. 'MARY' 'JOHN' et 'BELL' correspondent aux noms attribués à trois 'tiers', parmi lesquelles on peut distinguer 'les intervals tiers' (MARY et JOHN) et un 'point tier'. Les 'intervals tiers' permettent de segmenter le signal en intervalles et d'annoter ou de transcrire le contenu des intervalles. Les 'points tiers' sont utilisés pour effectuer des annotations qui reposent sur la détermination et l'identification de points spécifiques (marquage des points cibles et des tons pour une représentation de l'intonation, etc.). Pour effectuer la transcription orthographique, nous nous intéresserons aux seuls « intervalles tiers ». Le champ [TIER NAME] sera donc le seul qui nous intéressera. Il nous permet de spécifier le nombre de *tiers* (tires ci-après) ou espaces dans lesquels on peut saisir des données). On pourrait, par exemple, spécifier trois tires correspondant à chacun des locuteurs en présence. Dans notre projet, pour effectuer la transcription orthographique des enregistrements, nous créerons une tire que nous pourrions appeler 'transcription orthographique', et cela quel que soit le nombre de locuteurs en présence². Dans le champ, c'est le nombre de mots séparés par un espace qui permet de spécifier le nombre de tires. Comme nous ne voulons dans un premier temps qu'une seule tire, nous devons lui donner un nom constitué de caractères sans espaces entre eux. Par exemple, si l'on transcrit une conversation libre [ou entretien non directif] entre un enquêteur et le locuteur Jean Dupont, on pourra nommer notre tire [31ajd1], en référence aux normes d'étiquetage des fichiers expliqués dans le document PFC *format des rendus*³. Comme il s'agit de la transcription orthographique du fichier audio, nous pouvons aussi indiquer dans le champ [TIER NAME] '*Transcript-graphe*' sans séparer cette séquence du nom (exemple : 31ajd1_transcript-graphe). En effet, tout espace correspond à une nouvelle tire.
3. Une fois le nom de la 'tire' attribué, Cliquez sur le bouton [OK]. On se retrouve dans la fenêtre « Praat Objects », avec le nom du son ouvert précédemment plus le textgrid qu'on vient de créer ; c'est ce dernier qui est maintenant en surbrillance.
4. Cliquez sur l'un des deux items et glissez vers l'autre en gardant cliqué de manière à les sélectionner tous les deux. A cette étape, nous avons les deux éléments en surbrillance.
5. Cliquez sur le bouton [EDIT]. Apparaîtra alors à l'écran une fenêtre ayant la forme de la figure suivante (si ce n'est que les versions les plus récentes intègrent un sonagramme entre le signal et la tire jaune) :

² Nous serons amenés à créer d'autres tires pour les codages schwa et liaison comme nous l'expliquerons plus loin.

³ Le format des rendus assigne à chaque locuteur un identifiant unique fait de six caractères : les trois premiers correspondent au lieu de l'enquête, les trois derniers aux deux initiales de l'enquêté suivies d'un chiffre représentant son rang alphabétique. Ainsi, le locuteur Jean Dubois (jd1) enregistré à Toulouse ville (31a) sera : 31ajd1. Il serait utile d'employer ces identifiants au sein même des transcriptions sous praat à la place des simples initiales des enquêtés ; cela uniformiserait tous les documents PFC.

Fig. 1



La fenêtre ci-dessus 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. Immédiatement en dessous se trouve, en jaune, la tire proprement dite dans laquelle apparaîtra le texte tapé dans le champ de saisie. Sous la tire en jaune, se trouvent deux bandes horizontales 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 et 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 à droite ou à gauche du son.

6. **Sélection des intervalles et Création des marques :** Vous sélectionnerez alors des extraits du signal en cliquant puis en glissant (« click and drag ») dans la fenêtre donnant la représentation acoustique du son. Un carré rouge délimitera alors la portion sélectionnée. Vous pourrez écouter cette dernière 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.
7. Lorsque vous aurez identifié un segment que vous voulez transcrire, vous placerez des marques (voir plus loin) 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 vous permettra de constituer la frontière d'un nouvel intervalle.
8. Pour créer des marques, il faut cliquer sur le signal. Faire un simple clic sans glisser la souris, sinon, le mouvement est interprété comme étant une sélection et non une création de marque. Apparaît alors une « marque » rouge sur le signal et grise sur la tire. Attention, tant que vous n'avez pas appuyé sur la touche retour chariot (entrée), la marque n'est pas prise en compte par le logiciel et le fait de cliquer ailleurs la fera se déplacer. Avant de transcrire, assurez-vous que vous avez cliqué sur la portion de la tire concernée, laquelle deviendra jaune. PRAAT vous place automatiquement dans le champ d'édition en haut de la fenêtre. Tout ce que vous y tapez est répercuté dans la portion jaune (sélectionnée). Tout ajout ou effacement s'effectuera dans ce champ qui fonctionne comme la fenêtre d'un traitement de texte.
9. **Déplacement et effacement de marques déjà créées :** 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). On va alors au menu Point/Boundary où le choix Remove efface la marque sélectionnée. Le texte qui précédait la marque et le texte qui la suivait éventuellement sont fusionnés dans le même intervalle.
10. Quand vous avez fini votre séance de travail, vous pouvez fermer la fenêtre en cliquant sur la croix en haut à droite ou en choisissant [CLOSE] dans le menu [FILE].
11. Attention, ne pas quitter sans avoir préalablement sauvegardé votre travail, car cela ne se fait pas automatiquement. Des deux fichiers en surbrillance, il faut sélectionner uniquement [TEXTGRID] et pas le fichier son. Pour enregistrer, choisir [WRITE TO TEXT FILE] dans le menu [WRITE].
12. Si vous avez besoin de reprendre 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 aura déjà été créé. Après avoir ouvert le son avec le menu [OPEN LONG SOUND FILE], au lieu de créer ce fichier avec le bouton [TO TEXTGRID], il faut ouvrir celui qu'on a créé auparavant avec le menu [READ] et le sous menu [READ FROM FILE]. Attention, ne pas choisir [OPEN LONG SOUND FILE]. Ce dernier est, comme son nom l'indique, destiné à ouvrir un fichier son, et [READ FROM FILE] est destiné à ouvrir un fichier [TEXTGRID] créé dans une séance de travail antérieure.

1.3.2 La transcription sous PRAAT proprement dite :

On rappellera que PRAAT n'impose pas un style particulier de transcription. Les normes adoptées pour PFC sont décrites en 2. Quelques particularités cependant s'appliquent à la

transcription sous PRAAT dans notre projet, d'autant plus que nous n'utilisons qu'une seule 'tire' pour la transcription orthographique, quel que soit le nombre de locuteurs.

Toute intervention, que ce soit une prise de parole, un chevauchement ou une interruption, doit être précédée par les initiales du locuteur suivies de deux points et d'un espace ; par exemple, chaque fois que l'enquêteur intervient, on notera [E :]. Si l'enquêteur intervient pendant la prise de parole de 31ajd1 par exemple, on notera :

31ajd1 : Je n'aime pas la choucroute <E : Je comprends.> parce qu'il y a du porc⁴.

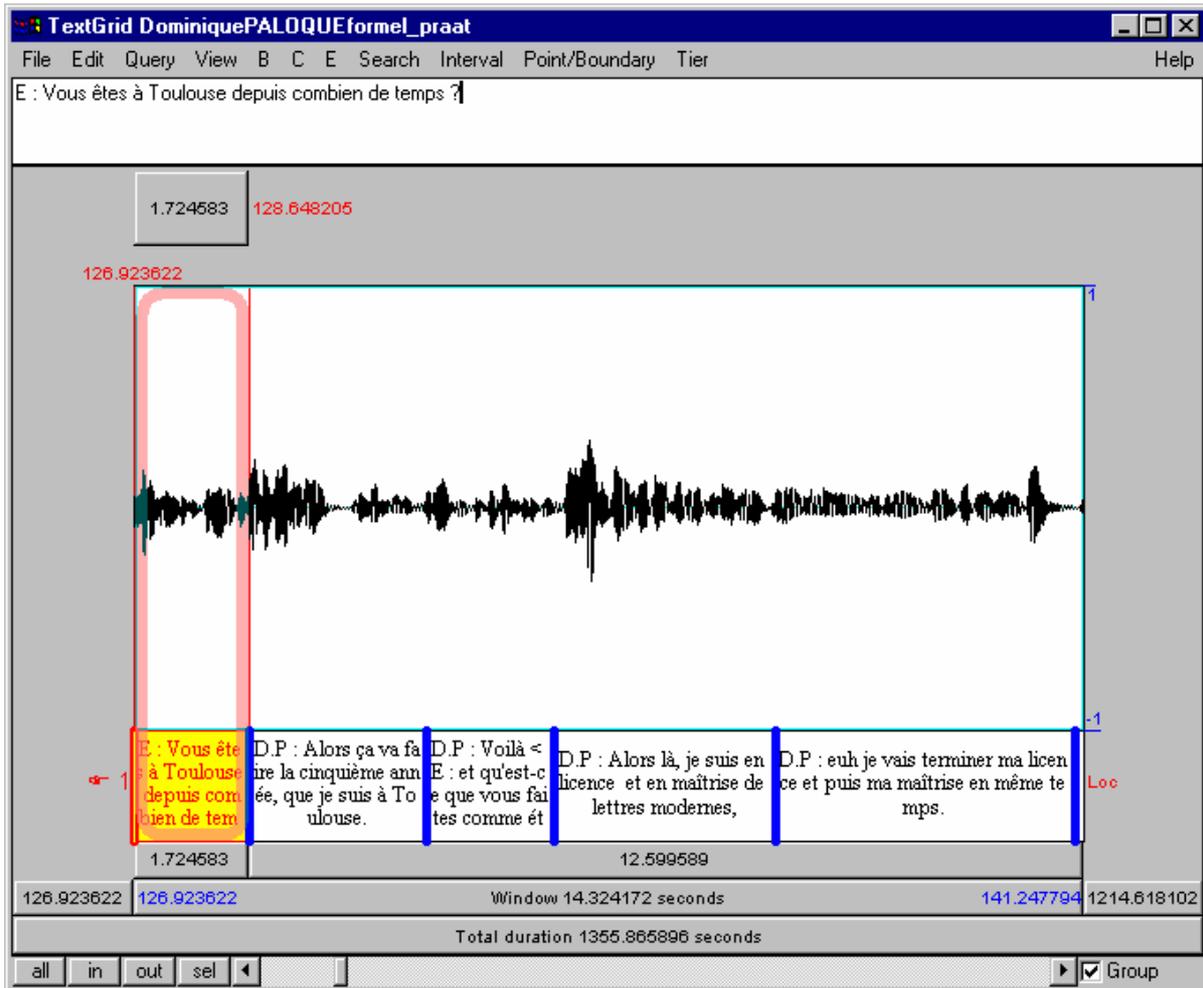
On remarquera que toutes les interventions commencent par une majuscule et finissent par une marque de ponctuation **sans retour chariot**.

Normalement, les intervalles sont créés en fonction des prises de parole. Chaque fois qu'on change d'interlocuteur, on crée un intervalle. Il arrive cependant qu'une personne prenne la parole et la garde assez longtemps. Pour que l'alignement se fasse sur des portions de signal de durée limitée, il est absolument nécessaire de créer un ou plusieurs intervalles au milieu de l'intervention. Il suffit, pour continuer à transcrire l'intervention, de reprendre les initiales du locuteur suivies de deux points et d'un espace.

On trouvera ci-dessous (fig. 2) un bref exemple de transcription tel qu'il apparaît dans Praat. On notera les deux dernières répliques qui correspondent au même tour de parole. On remarquera la partie sélectionnée dont on connaît la longueur 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. On peut constater que la transcription, dont on ne voyait qu'une partie dans le carré jaune est maintenant visible dans sa totalité dans le champ de saisie sous la barre des menus. Le schéma est suivi par la partie correspondant à ce passage dans le *textgrid* (rappelons que le *textgrid* est le fichier texte issue d'une transcription).

⁴ Nous rappelons ce que nous avons dit en note plus haut. Il serait utile d'employer ces identifiants au sein même des transcriptions sous praat à la place des simples initiales des enquêtés ; cela uniformiserait tous les documents PFC.

Fig. 2



Partie d'un fichier *Textgrid* correspondant à la figure 2

```

intervals [26]:
  xmin = 126.92362249533851
  xmax = 128.64820506304423
  text = "E : Vous êtes à Toulouse depuis combien de temps ?"
intervals [27]:
  xmin = 128.64820506304423
  xmax = 131.31852645820149
  text = "D.P : Alors ça va faire la cinquième année, que je suis à
Toulouse."
intervals [28]:
  xmin = 131.31852645820149
  xmax = 133.26563580883698
  text = "D.P : Voilà < E : et qu'est-ce que vous faites comme études ? >
"
intervals [29]:
  xmin = 133.26563580883698
  xmax = 136.60353755278356
  text = "D.P : Alors là, je suis en licence et en maîtrise de lettres
modernes,"
intervals [30]:
  xmin = 136.60353755278356
  xmax = 141.10970490711142
  text = "D.P : euh je vais terminer ma licence et puis ma maîtrise en même
temps."

```

1.4 Transpraat

Le contenu d'un textgrid est un texte balisé où il est très difficile de retrouver les dialogues patiemment transcrits. Transpraat, petit utilitaire informatique, a été développé au sein de notre équipe afin de remédier à cet inconvénient et nous l'envoyons sur simple demande. Transpraat permet de transformer ce format balisé en fichier *text* exploitable sur un traitement de texte (par exemple Microsoft Word). Le résultat de la transformation est immédiatement lisible dans la mesure où il a la forme d'un enchaînement naturel des dialogues. On notera que, dans la philosophie de PFC, le *textgrid* ou le texte transformé peuvent servir ultérieurement à divers traitements spécialisés. Il faut néanmoins être prudent car le texte obtenu avec *Transpraat* ne garde pas de liens avec le *textgrid* ou avec *PRAAT*. S'il y a une correction à faire sur le document, il convient de la faire sur le *textgrid* ou mieux en ouvrant *PRAAT*. En effet, on peut obtenir un texte lisible à partir d'un *textgrid* mais le contraire est impossible puisqu'on aura perdu le balisage qui permet l'alignement du signal au texte, spécificité de *PRAAT*.

Voici le texte issu de Transpraat et qui correspond à la partie du *textgrid* vue plus haut :

E : Vous êtes à Toulouse depuis combien de temps ?

D.P : Alors ça va faire la cinquième année, que je suis à Toulouse. Voilà < E : et qu'est-ce que vous faites comme études ? > Alors là, je suis en licence et en maîtrise de lettres modernes, euh je vais terminer ma licence et puis ma maîtrise en même temps.

On remarquera que les quatre derniers intervalles ont été fusionnés dans la mesure où il s'agit d'un seul tour de parole produit par le locuteur DP.

1.5 Gestion des tires sous Praat

Objectifs :

1. dupliquer une tire à l'intérieur d'un fichier de transcription (Textgrid)

Pour rappel, les codages du schwa et de la liaison se font sur des tires indépendantes de la tire de transcription orthographique. Il est donc nécessaire de créer des tires propres à ces codages. Leur création se fait par la duplication de la première tire orthographique.

2. réunir au sein d'un même fichier des tires appartenant à différents fichiers textgrid

Cette seconde manipulation est susceptible d'être opérée dans le cas où les codages du schwa et de la liaison sont effectués pour un même fichier par des personnes distinctes et de manière simultanée.

1. Création d'une nouvelle tire par duplication de la tire de transcription orthographique

1. Dans la fenêtre « TextGrid... » préalablement ouverte, cliquer tout d'abord sur la tire que l'on souhaite dupliquer (ici la tire de transcription orthographique),
2. puis choisir [Duplicate tier] dans le menu [Tier].
3. Une boîte de dialogue « Duplicate tier » apparaît alors, elle comporte deux champs [position] et [name].

Dans le champ [position] on mettra le nombre correspondant à l'ordre que l'on souhaite voir occuper par la tire créée. Par défaut la tire de transcription orthographique a la position « 1 », on choisira alors « 2 » pour la « position » de la tire de codage schwa et donc « 3 » pour la « position » de la tire de codage de liaison.

Dans le champ [name] on mettra le nom de la tire que l'on souhaite créer ; par exemple, si la tire de la transcription orthographique s'intitule [31ajd1_trans-orthe], on pourra appeler la tire de codage schwa [31ajd1_schwa], et la tire de codage liaison [31ajd1_liaison].

4. *Quand cela est fait, cliquer sur « OK ». On a alors créé une nouvelle tire dupliquant la première tire de transcription orthographique. On travaillera dans cette nouvelle tire afin d'y annoter les codages désirés.*

2. Réunir dans un même fichier TextGrid plusieurs tires provenant de différents fichiers

Lorsque plusieurs personnes travaillent séparément sur une même transcription, l'une codant par exemple le schwa et une autre la liaison, elles travailleront chacune dans un fichier TextGrid où elles auront créé pour leur besoin de codage (schwa ou liaison) une nouvelle tire. La base commune de ces fichiers, à savoir la tire de transcription orthographique, sera bien entendu identique et devra le rester (pour travailler sur des fichiers TextGrid ayant une même base commune il suffira de copier le premier fichier TextGrid dont les « codeurs » auront ainsi une même version).

Une fois les codages terminés, l'un des codeurs - par exemple le « codeur de liaison » - procédera à la manipulation suivante :

- 1- Après avoir cliqué sur la tire du codage de liaison, il choisira [Copy tier to list of objects] dans le menu [Tier].
- 2- Un objet [IntervalTier liaison] apparaît alors dans la fenêtre « Praat objects », objet que l'on sauvegardera (pour rappel, choisir [Write to text file] dans le menu [Write] de la fenêtre "Praat objects"), cet objet se verra automatiquement attribué une extension « .IntervalTier ».

Ce fichier .IntervalTier sera alors donné au détenteur de l'autre fichier .TextGrid (celui qui contient la tire de transcription orthographique et la tire de codage schwa) qui procédera alors à la manipulation suivante :

- 3- Les deux fichiers .TextGrid et .IntervalTier sont ouverts dans la fenêtre « Praat Objects ».
- 4- Cliquer sur l'un des deux items et glisser vers l'autre de manière à les sélectionner tous les

deux ; ils apparaîtrons alors en surbrillance.

- 5- Cliquer sur le bouton [Append].
- 6- Un nouveau fichier TextGrid est alors créé et apparaît dans la fenêtre « Praat objects ». Ce fichier TextGrid final, qui devra être sauvegardé, contient les trois tires : transcription orthographique, codage schwa et codage liaison.

2. Transcription orthographique des données

Les conventions présentées ci-dessous complètent la section précédente mais en sont indépendantes.

2.1. Transcription orthographique

La transcription d'un corpus oral facilite son utilisation et est préalable à un étiquetage phonologique, morphologique ou syntaxique. L'objectif que nous nous sommes fixé est d'aligner aussi fidèlement que possible les données orales au texte écrit. Lors du passage d'un corpus oral à un texte écrit le transcripteur est confronté d'emblée à la question suivante : le texte doit-il refléter le caractère oral du corpus ou être une sorte de traduction normée du dialogue enregistré ? Si le caractère oral doit être maintenu, une transcription phonétique fine s'impose. Si d'un autre côté, on souhaite mettre l'accent sur l'écrit, la transcription sera une traduction de l'oral et la norme écrite primera. Nous excluons immédiatement une transcription phonétique fine, peu lisible tout en étant beaucoup trop exigeante et trop facilement sujette aux erreurs. L'autre alternative n'est pas non plus retenue puisque notre objectif n'est pas la création d'un texte 'littéraire' mais l'alignement Texte/Parole. Entre ces deux extrêmes, de nombreuses alternatives sont possibles, et nous opterons pour une transcription orthographique standard, y compris une ponctuation standard, qui reflétera fidèlement le discours du témoin. Le texte ne sera pas normalisé et les éléments absents ne seront pas réintroduits, ainsi :

(1) [ivwapasakəmsa] sera transcrit : 'Il voit pas ça comme ça.'

(2) [japadkwa] sera transcrit : 'Il y a pas de quoi.'

La graphie standard du pronom personnel est maintenue, mais la particule de la négation absente de l'oral n'est pas introduite alors que la norme graphique l'exigerait.

En cas de doute, il peut s'avérer utile de consulter un dictionnaire : par exemple, l'adverbe *bien* a aussi une forme *ben* qui est acceptable. Ce mode de transcription supprime cependant une foule de renseignements : l'absence de schwa dans (2) n'est pas indiquée, de même que la chute de la liquide dans le pronom. Rappelons que toute transcription est une interprétation et un début d'analyse et notre but ici est de minimiser cette interprétation et cette analyse. En proposant un tel système de transcription, nous espérons réduire considérablement les erreurs dues à une écoute imparfaite et nous n'influons pas le phonologue ou le linguiste qui reviendra sur les données pour l'analyse de tel ou tel phénomène.

Soulignons enfin que tous les chiffres sont transcrits sous leur forme graphique, ceci pour ne pas interférer avec le codage de schwa ou de la liaison.

2.2. Conventions de transcription

Les conventions proposées doivent beaucoup au travail effectué autour du corpus du GARS (Aix-Marseille) et de VALIBEL (Louvain-la-Neuve). Nous avons aussi tenu compte des instructions de Gjert Kristoffersen (Bergen) pour la transcription d'une banque de données sur les dialectes norvégiens. Nous demandons à tous les transcrip-teurs de respecter scrupuleusement ces consignes.

2.2.1 *Tours de parole*

Chaque locuteur est désigné par ses initiales, l'enquêteur par 'E' ou E1, E2, etc. s'il y a plusieurs enquêteurs. Les initiales sont suivies de deux points (:), et sont reprises à chaque intervalle même s'il n'y a pas de changement de locuteur. Il n'y a aucun paragraphe, ni retour à la ligne. Rappelons que chaque tour de parole sera maintenu aussi bref que possible pour assurer l'exactitude de l'alignement et que si un locuteur s'exprime longuement, le signal sera coupé à intervalles réguliers. Une pause brève (ou un mouvement mélodique indiquant une continuation) est indiquée par une virgule, une pause plus longue (ou une fin d'énoncé marquée mélodiquement) par un point. Une question est signalée par un point d'interrogation.

2.2.2 *Chevauchements*

Lorsqu'un locuteur L1 parle et qu'un autre locuteur L2 se manifeste uniquement par des appréciations en arrière-plan comme oui, non, hum, pff, ces remarques discursives sont ignorées et ne sont pas transcrites. Par contre, si L2 interrompt véritablement son interlocuteur en produisant un énoncé, son intervention sera indiquée entre chevrons à l'intérieur du discours de L1 si ce dernier continue à parler. S'il y a véritable interruption et que cela mène à un changement de locuteur, cela sera indiqué par un changement d'intervalle. Dans les deux cas, le locuteur qui cause le chevauchement sera indiqué de façon univoque : <L : sans espace entre les différents symboles, mais avec un espace après les deux points.

- **Exemple 1** : L1 conserve la parole, malgré l'intervention de L2

L1 : Tu as vu Pierre ces derniers <L2 : Pierre, non.> temps ? Il est parti à Paris.

L2 interrompt L1 qui garde le tour de parole.

- **Exemple 2** : L1 conserve la parole, même si L2 intervient plus longuement.

L1 : Tu as vu Pierre ces derniers < L2 : Pierre, non. Je suis plus en contact avec lui.> temps ? Il est parti à Paris.

Cependant, si l'intervention est vraiment trop longue, ne pas hésiter à créer un nouvel intervalle.

- **Exemple 3** : L2 interrompt L1

Ici nous adopterons la convention suivante avec minuscules pour la prise de parole de L2.

L1 : Tu as vu ces derniers <L2 : Pierre, non.> temps ? Il est parti à Paris et il devait rentrer la semaine passée <L2 : Pierre.>

L2 : est vraiment un chic type. Et d'ailleurs, je te dirais que ...

L2 interrompt L1, 'Pierre' est articulé en même temps que 'passée' et L2 garde la parole.

2.2.3 Mots tronqués

La troncation d'un mot sera indiquée par une barre oblique suivie d'un espace.

Exemple 1 :

L1 : Il m'a pro/ pro/ mis de revenir vite.

L1 : Oui, et main/ maintenant il a enfin compris.

S'il y a une pause à l'intérieur d'un mot, mais que le locuteur, après cette pause, termine le mot sans répéter de syllabe, cela sera indiqué par une barre oblique immédiatement suivie d'un trait d'union.

Exemple 2 :

L1 : Il m'a pro/-mis de revenir vite

Les mots répétés eux sont repris et séparés par une virgule.

L1 : Il lui, lui, lui a dit de revenir vite.

2.2.4 Sigles

Les sigles épelés sont transcrits avec chaque lettre séparée par un point, alors que les sigles lus sont transcrits comme un mot. La prononciation inattendue d'un sigle lu peut être indiquée entre parenthèses en SAMPA⁵.

Exemple 1 :

L1 : La S.N.C.F. est à nouveau en grève.

NB : ici le sigle est prononcé [laesenseɛf] (SIL IPA) [laesEnseEf] (SAMPA), ce qui est attendu et n'est donc pas noté.

L1 : Il a été admis à l'I.U.T. de Caen.

De même, le sigle est prononcé [liyte] (SIL IPA) [liyte] (SAMPA).

Exemple 2 :

L1 : Le CNET [snet] (SIL IPA) [snEt] (SAMPA) embauche du personnel.

⁵ Pour des informations sur SAMPA, se reporter au document PFC « codage liaison ».

Dans ce dernier exemple, la personne a prononcé ce sigle d'une façon inhabituelle, d'où l'utilisation de la transcription SAMPA entre parenthèses.

Si la personne parle d'une organisation, d'une association ou d'un parti politique qui peut être dénommé par un sigle ou par une suite de mot (PS et Parti Socialiste), la transcription indiquera ce que le locuteur a réellement produit.

Exemple :

L1 : Il travaille au parti socialiste.

pour [*itravajopartisosjalist*](SIL IPA) [itRavajopaRtisosjalist] (SAMPA)

L1 : Il travaille au P.S.

pour [*itravajopees*](SIL IPA) [itRavajopeEs] (SAMPA)

2.2.5 Hésitations, bruits, onomatopées

Les hésitations seront transcrites euh. A la finale des mots il est parfois difficile de décider s'il y a hésitation ou prononciation d'un schwa. Dans TOUS les cas, la transcription sera euh.

Certaines productions du locuteur peuvent être transcrites de façon naturelle par des graphies conventionnelles admises dans le dictionnaire : pfft, psit, tss-tss, tsoin-tsoin. Dans tous les autres cas noter les bruits entre parenthèses, exemple (onomatopée chchchch).

Exemple :

L1 : Ben, je veux dire euh, que, euh, je fais plus de ski maintenant.

2.2.6 Le discours rapporté

Le discours rapporté sera signalé par des guillemets simples ('') en début et en fin de discours.

Exemple :

L1 : Il m'a dit 'tu es complètement idiot d'avoir accepté ce boulot', et je crois bien qu'au fond il avait raison.

2.2.7 L'utilisation des parenthèses

Les syllabes incompréhensibles seront indiquées entre parenthèses et un 'X' correspond à une syllabe.

Exemple 1 :

L1 : Paul s'en va et juste à ce moment, (XXX) et il tombe sur lui.

Les parenthèses seront aussi utilisées pour tout commentaire :

- commentaires sur l'enregistrement

Exemple 2 :

(rires), (bruit), et..

- commentaires sur la transcription et transcription approximative.

Exemple 3 :

L1 : Paul s'en va et juste à ce moment (bruit) (catastrophe) et il tombe sur lui.

2.2.8 Résumé :

- Les seules marques de ponctuation dans le texte sont la virgule, le point et le point d'interrogation.
- Les chevrons signalent un chevauchement et commencent toujours par identifier le locuteur.
 - Les guillemets simples signalent un discours rapporté.
 - Les parenthèses ont pour fonction d'incorporer des commentaires.
- La longueur des intervalles sous PRAAT doit être le plus bref possible, surtout, ne pas hésiter à fractionner aussi souvent que nécessaire. Ne pas oublier de reprendre les initiales du locuteur à chaque intervalle.

2.2.9 Exemple de transcription :

E : Eh, Eh, vous avez des, gardé des bons, vraiment un très bon souvenir de l'école.

OB : Ah oui l'école. Très, très bon, oui. Ah oui.

E : Il y a jamais eu des jours où.

OB : Pourtant, il fallait faire le ménage le matin.

OB : Fallait allumer le poêle, même en arrivant de la Chauterie là haut. (rires).

OB : Il fallait quand même euh faire <E : chercher le.> faire le / épousseter les bureaux, les tables. (rires)

E : Donc vous êtes allée à l'école jusqu'à quel âge ?

OB : Bien, douze ans et demi. J'avais douze ans au mois de janvier.

OB : J'ai passé mon certificat au mois de juin, je pense. Mai, juin.

OB : J'avais douze ans et demi.

E : Et après, qu'est-ce que vous avez fait après l'école, en fait ?

OB : Et bien après l'école, j'ai travaillé, avec mon père, avec mes parents.

OB : Je suis rentrée à la mairie, en quarante-six. J'ai laissé l'école en quarante-trois. (musique en bruit de fond).

OB : Je suis rentrée à la mairie en quarante-six pour les tickets d'alimentation.

OB : Et puis, j'y suis restée jusqu'en quatre-vingt, fin quatre-vingt-dix, mille neuf cent quatre-vingt-dix.

OB : Et voilà. Je me suis mariée en quarante-sept.

E : A Valeuil aussi ?

OB : A Valeuil aussi.

E : Donc. (hésitations et pauses) A Valeuil vous avez fait quoi comme métier, en fait, dans votre vie ?

On rappellera ici que TRANSPRAAT fusionne automatiquement les interventions fractionnées d'un même locuteur. TRANSPRAAT, appliqué à la transcription ci-dessus, donnera le résultat suivant.

E : Eh, Eh, vous avez des, gardé des bons, vraiment un très bon souvenir de l'école.

OB : Ah oui l'école. Très, très bon, oui. Ah oui.

E : Il y a jamais eu des jours où.

OB : Pourtant, il fallait faire le ménage le matin. Fallait allumer le poêle, même en arrivant de la Chauterie là haut. (rires) Il fallait quand même euh faire <E : chercher le.> faire le / épousseter les bureaux, les tables. (rires)

E : Donc vous êtes allée à l'école jusqu'à quel âge ?

OB : Bien, douze ans et demi. J'avais douze ans au mois de janvier. J'ai passé mon certificat au mois de juin, je pense. Mai, juin. J'avais douze ans et demi.

E : Et après, qu'est-ce que vous avez fait après l'école, en fait ?

OB : Et bien après l'école, j'ai travaillé, avec mon père, avec mes parents. Je suis rentrée à la mairie, en quarante-six. J'ai laissé l'école en quarante-trois. (musique en bruit de fond) Je suis rentrée à la mairie en quarante-six pour les tickets d'alimentation. Et puis, j'y suis restée jusqu'en quatre-vingt, fin quatre-vingt-dix, mille neuf cent quatre-vingt-dix. Et voilà. Je me suis mariée en quarante-sept.

E : A Valeuil aussi ?

OB : A Valeuil aussi.

E : Donc. (hésitations et pauses) A Valeuil vous avez fait quoi comme métier, en fait, dans votre vie ?

DIRECTIONS D'ANALYSE (juin 2002)

Jacques Durand, Bernard Laks, Chantal Lyche

0. Introduction

Ce document présuppose que les transcriptions des données ont été effectuées sous Praat (voir le document *PFC : Transcriptions*) et consigne les trois grands volets de l'analyse phonologique requise pour chaque locuteur enregistré. Nous explorons en (1) les grands axes qui ont guidé aussi bien l'élaboration de la liste de mots que celle du texte qui, comme nous l'avons souligné dans le protocole d'enquête, sont lus par tous les témoins. Les résultats de ces deux exercices de lecture revêtent de ce fait une importance considérable car non seulement ils nous permettent d'avoir accès à plusieurs registres du locuteur, mais surtout ils constituent le fondement de l'étude de la variation inter-dialectale. C'est pourquoi liste et texte ont été minutieusement conçus avec pour objectif, dans un premier temps, de dégager le système phonémique du locuteur et d'obtenir un ensemble de données sur le schwa et la liaison. Il semble alors raisonnable que la première tâche du transcripateur consiste à dresser l'inventaire phonémique du locuteur (section 2) sur la base de la liste et du texte. La section (3) est consacrée au codage du schwa et (4) à celle de la liaison et dans les deux cas, l'analyse ira au-delà de celle du texte et portera aussi sur des passages de conversation guidée et de conversation libre. Deux utilitaires informatiques *classeur-schwa* et *classeur-liaison* fourniront sur la base de ces codages des données chiffrées. Ces utilitaires ainsi que leur guide d'utilisation sont donnés dans un document séparé.

1. Quelques principes

1.1. Liste de mots

Dans la mesure du possible, tous les mots proposés font partie du vocabulaire courant. Sont donc bannis tous les termes techniques du type *allopathie* ou *pithécanthrope*. La quasi-totalité des exemples sont repris de la littérature phonétique et des grandes enquêtes déjà réalisées. Enfin, nous avons repris diverses paires minimales testées dans le passage à lire. Du point de vue phonologique, nous avons favorisé les oppositions vocaliques, surtout en position accentuée. Un examen plus complet des oppositions en position inaccentuée et de phénomènes comme l'harmonie vocalique ou le relâchement aurait exigé une liste bien trop longue. En ce qui concerne les consonnes, nous avons cherché à vérifier le statut de <gn>, la réalisation du <r> dans divers contextes, le groupe consonantique /ks/ et quelques assimilations de voisement.

Les résultats de la liste de mots ne peuvent être considérés comme fiables. L'informateur lira sans doute comme on lui a appris à lire. La liste de mots, comparée à la grille des phonèmes établie indépendamment, donnera cependant des indications sur l'attitude de l'informateur par rapport à l'orthographe et à la norme. Les réactions verbales des témoins à cet exercice sont une source précieuse d'information quant à leur positionnement sociolinguistique. D'autre part, même si un locuteur fait des distinctions artificielles, un ensemble de phénomènes échappe à sa vigilance. Dans nos enregistrements, nous avons ainsi remarqué le dévoisement de /r/ en position finale, une semi-vocalisation proche de la parole spontanée, la réduction de groupes consonantiques, des assimilations de voisements (*slip* prononcé [zlip]).

Pour éviter que l'informateur suffisamment scolarisé repère très vite ce que l'enquêteur recherche dans la liste de mots, il est essentiel que l'ordre des mots soit aléatoire. Nous reprenons en fin de liste quelques paires minimales afin de tester la conscience linguistique de l'informateur et de comparer les résultats avec ce qui aura déjà été produit. Nous avons été confrontés à ce problème de transparence dans la liste de phrases donnée par Morin et Kaye (1982) dans leur travail sur la liaison. Nous avons donné cette liste de 66 contextes de liaison à notre premier locuteur qui a vite compris que l'on testait les sites de liaison et qui faisait des commentaires sur les liaisons recommandées ou non. Une étude de la liaison et de schwa ne pourra se faire que lorsque l'informateur n'est plus sur ses gardes, c'est-à-dire lors d'une lecture d'un texte suivi ou d'une conversation.

1.2 Texte

La lecture du texte ne fournit aucune donnée pour une étude de la parole spontanée, mais vu la diversité des phénomènes en jeu, l'informateur est beaucoup moins conscient de sa performance linguistique dans cet exercice. On obtient alors un style soutenu certes, mais dans certains cas un style moins surveillé, voire naturel.

Le texte a été établi avec soin en termes de composition lexicale et grammaticale. Il a la forme d'un petit article dans un journal régional et ne pose aucune difficulté de compréhension. Tous nos informateurs ont d'ailleurs souligné la simplicité de la tâche. Si le style est qualifié de 'mauvais' ou de très peu littéraire, nous y voyons non pas une critique mais plutôt un compliment. Nous avons voulu éviter un style littéraire ou même le style d'un bon journalisme. Ce texte a été construit afin d'éliminer tout ce qui pouvait être ressenti comme trop régional ou même trop français. Tous les noms propres devraient être suffisamment neutres pour ne surprendre aucun locuteur, qu'il appartienne ou non à l'Hexagone. Le texte garde cependant un aspect culturel 'français', mais nous ne pensons pas que ce reste culturel constitue un frein à la lecture de locuteurs belges ou canadiens par exemple. Si dans un point d'enquête particulier il apparaît nécessaire de soumettre un texte supplémentaire aux locuteurs, nous conseillons vivement que ce texte soit, tout comme le nôtre, axé sur la simplicité lexicale et grammaticale.

Lors de la fabrication du texte, nous avons tenté d'inclure non seulement la majorité des segments vocaliques et consonantiques du français, mais aussi tout un ensemble de phénomènes phonologiques typiques du français. Le texte soumis permet de dresser l'inventaire des phonèmes du locuteur, il reprend certains mots et paires minimales repris

dans la liste. Le texte permet de tester deux phénomènes en détail : la prononciation du schwa et la liaison (y compris les phénomènes connexes de nasalisation et de 'h' aspiré. D'autres phénomènes sont néanmoins présents : par exemple, la palatalisation, les glissantes et diverses assimilations.

2. Système phonémique

Le système phonémique du locuteur sera établi dans un premier temps sur la base de la lecture des mots. Cette première ébauche sera ensuite validée par la lecture du texte, en particulier par l'intermédiaire des mots repris de la liste, et enfin par les entretiens. En termes non techniques, voici quelques-uns des axes organisateurs qui ont déterminé la sélection des mots utilisés ainsi que quelques questions qui permettront une première description.

2.1 Axes organisateurs de la liste

SONS FAMILLE A

(+ Schwa position finale cf. mal-malle-mâle)

mal

mâle

malle

pâte

patte

ras

rat

SONS FAMILLE é/è

épais

épée

épier

étrier

étriller

liège

lierre

pêcheur

pêcheur

piquais

piqué

piquer
piquet
fête
fêtard
fêter
faites

SONS FAMILLE EU

jeune
jeûne
creuse
creux
dégeler + différence schwa - EU
déjeuner
des genêts
des jeunets
feutre
meurtre
peuple

SONS FAMILLE O

beauté
botté
paume
pomme
rauque
roc
rhinocéros

SCHWA

FINAL : cf. mal-mâle, roc-rauque, etc.

INTERNE

bêtement
bouleverser

INITIAL

chemise

petit

GLISSANTES

miette

mouette

muette

cinquième

fou à lier

nous prendrions

vous prendriez

million

millionnaire

niais

nièce

nier

quatrième

relier

reliure

scier

influence

trouer

NASALES

blanc

blond

brin

brun

CONSONNE <gn> nasale palatale ou nasale + semivoyelle ?

agneau

compagne

compagnie

gnôle

baignoire

GROUPES CONSONANTIQUES FINAUX

aspect

infect

intact

GROUPES CONSONANTIQUES INTERNES

explosion

extraordinaire

ex-femme

ex-mari

ASSIMILATION VOISEMENT

islamique

slip

socialisme

médecin

Noter que les paires minimales sont en fin de liste, à savoir :

85. patte

86. pâte

87. épais

88. épée

89. jeune

90. jeûne

91. beauté

92. botté

93. brun

94. brin

2.2 Questions pour l'analyse

Nous proposons ci-dessous quelques questions générales qui peuvent servir comme point de départ pour l'établissement de l'inventaire phonémique du locuteur. On notera les différences observées entre la lecture de la liste de mots et la lecture du texte.

A des fins de comparaison, nous donnons ci-dessous deux transcriptions de la liste de mots et une du texte.

2.2.1 Liste de mots

Nous reprenons ici les transcriptions du dictionnaire *Le Petit Robert* 'PR' (édition de 1993) et celle de Martinet et Walter ('M&W') 1973, *Dictionnaire de la prononciation française dans son usage réel*.

	PR	M&W
1. roc	[Rɔk]	[rɔk]
2. rat	[Ra]	[ra]10, [ra]7
3. jeune	[ʒœn]	[ʒœn]16 [ʒœnə]1
4. mal	[mal]	[mal]17
5. ras	[Ra]	<i>chef éthiopien :</i> [ra]12[ra]3[ras]1 ?1 <i>adjectif :</i> [ra]15 [ra]2 [Ra]
6. fou à lier	[fualje]	--
7. des jeunets	[dezœnɛ]	<i>des :</i> [de]14 [dɛ]4 <i>jeunets :</i> [ʒœnɛ]9 [ʒœnɛ]7 [ʒœnɛ]1
8. intact	[ɛ̃takt]	[ɛ̃takt]
9. nous prendrions	[nuprãdrijõ]	[nuprãdr ??]
10. fêtard	[fɛtar]	[fɛtar]11 [fɛtar]4 [fɛtar]1 [fɛ:tar]1
11. nièce	[njɛs]	[njɛs]
12. pâte	[pat]	[pat]17 [pat]3
13. piquet	[pikɛ]	<i>pieu :</i> [pikɛ]15 [pikɛ]2 <i>jeu de cartes :</i> [pikɛ]12 [pikɛ]5
14. épée	[epe]	[epe]

15.	compagnie	[kõpaɲi]	[kõpaɲi] [kõpanji]
16.	fête	[fet]	[fet]13 [fɛ:t]4
17.	islamique	[islamik]	[islamik]
18.	agneau	[aɲo]	[anjo]9 [aɲo]8
19.	pêcheur	[peʃœR]	[peʃœr]11 [peʃœr]6
20.	médecin	[med(ə)sɛ̃]	[metsɛ̃]8 [medsɛ̃]7 [mɛdsɛ̃]2
21.	paume	[pom]	[pom]
22.	infect	[ɛ̃fɛkt]	[ɛ̃fɛkt]
23.	dégeler	[dez(ə)le]	[dezle]15 [dezəle]2
24.	bêtement	[betmã]	[betmã]15 [bɛ:tmã]2
25.	épier	[epje]	[epje]
26.	millionnaire	[miljõnɛR]	[miljõner]13 [miljõner]2 [mijõner]2 [mijõner]1
27.	brun	[brœ]	[brɛ̃]11 [brœ]5 [brɛ:]1
28.	scier	[sje]	[sje]15 [sije]2
29.	fêter	[fete]	[fete]14 [fete]3
30.	mouette	[mwet]	[mwet]
31.	déjeuner	[dezœne]	[dezøne]10 [dezœne]7 [dezne]1
32.	ex-femme	[eksfam]	<i>ex</i> : [eks][ɛks][ɛgs][ɛgs] <i>femme</i> : [fam]
33.	liège	[ljeʒ]	[ljeʒ]16 [lje:ʒ]1
34.	baignoire	[bɛɲwaR]	<i>baign-</i> : [bɛɲj]4 [bɛɲ]4 [bɛɲ]3 [bɛɲ]3 [bɛɲj]2 [bɛɲɲ]1 <i>-oir</i> : [war][war]
35.	pêcheur	[peʃœR]	[peʃœr]16 [peʃœr]1
36.	socialisme	[sɔsjalism]	<i>social-</i> : [sɔsjal] <i>-isme</i> : [ism] [izm]
37.	relier	[Rəlje]	[rəlje] [rəlije] [rəlie]
38.	aspect	[aspɛ]	[aspɛ]14 [aspe]3
39.	niais	[ɲje]	[ɲje]16 [ɲje]1
40.	épais	[epɛ]	[epɛ]
41.	des genêts	[dez(ə)nɛ]	<i>des</i> : [de]14 [dɛ]4 <i>genêts</i> : [ʒ(ə)nɛ]

42.	blond	[blɔ̃]	[blõ]	
43.	creux	[krø]	[krø]	
44.	reliure	[rəlɥyʀ]	[rəlɥyʀ] [rəliyr] [rəlijyr]	
45.	piqué	[pike]	[pike]	
46.	malle	[mal]	[mal]	
47.	gnôle	[ɲol]	[ɲol]9 [ɲol]8	
48.	bouleverser	[bulvɛʀse]	[bulvɛʀse]	
49.	million	[miljɔ̃]	[miljõ]16 [mijõ]2	
50.	explosion	[eksplozjɔ̃]	[eksplozjõ]10 [eksplozjõ]7	
51.	influence	[ɛ̃flyãs]	[ɛ̃flyãs]17	
52.	mâle	[mal]	[mal]17	
53.	ex-mari	[eksmaʀi]	<i>ex-</i> : [eks][eks][egs][egs] <i>-mari</i> : [maʀi]17	
54.	pomme	[pɔm]	[pɔm]	
55.	étrier	[etʀije]	[etʀije]17	
56.	chemise	[ʃ(ə)miz]	[ʃ(ə)miz]	
57.	brin	[brɛ̃]	[brɛ̃]	
58.	lierre	[ljɛʀ]	[ljɛʀ]	
59.	blanc	[blɑ̃]	[blã]	
60.	petit	[p(ə)ti] [pəti]16 [pti]2		
61.	jeûne	[ʒøn]	[ʒøn]13 [ʒœn]4 [ʒœ:n]1	
62.	rhinocéros	[ʀinɔsɛʀɔs]	[ʀinɔsɛʀɔs]15 [ʀinosɛʀɔs]10 [ʀinɔsɛʀɔs]2 [ʀinɔsɛros]1	
63.	miette	[mjɛt]	[mjɛt]	
64.	slip	[slip]	[slip]	
65.	compagne	[kɔ̃paɲ]	[kɔ̃paɲ]	
66.	peuple	[pœpl]	[pœpl]	
67.	rauque	[ʀok]	[ʀok]	
68.	cinquième	[sɛ̃kjɛm]	[sɛ̃kjɛm]	
69.	nier	[nje]	[nje]16 [nie]1	
70.	extraordinaire	[ekstʀaɔʀdinɛʀ]	<i>extra-</i> : [ekstʀa]9 [ekstʀa]4 [ekstʀa]4	

-ordinaire : [ɔrdinɛr]

71.	meurtre	[mœRtr]	[mœrtr]
72.	vous prendriez	[vupRãdRije]	[vuprãdr ???]
73.	botté	[bɔte]	???
74.	patte	[pat]	[pat]17
75.	étriller	[etRije]	[etrije]
76.	faites	[fɛt]	[fɛt]
77.	feutre	[føtr]	[føtr]
78.	quatrième	[katRijɛm]	[katrjɛm]16 [katrijɛ:m]1
79.	muette	[mɥɛt]	[mɥɛt] [myɛt]
80.	piquais	[pikɛ]	[???
81.	trouer	[trɥe]	[true]17
82.	piquer	[pikɛ]	[pikɛ]
83.	creuse	[krøz]	[krøz]
84.	beauté	[bɔte]	[bɔte]17
85.	patte	[pat]	[pat]17
86.	pâte	[pat]	[pat]17[pat]3
87.	épais	[epɛ]	[epɛ]
88.	épée	[epe]	[epe]
89.	jeune	[ʒœn]	[ʒœn]16 [ʒœnə]1
90.	jeûne	[ʒø̃n]	[ʒø̃n]13 [ʒœn]4 [ʒœ:n]1
91.	beauté	[bɔte]	[bɔte]17
92.	botté	[bɔte]	[???
93.	brun	[brœ̃]	[brẽ]11 [brœ̃]5 [brẽ:]1
94.	brin	[brẽ]	[brẽ]

2.2.2 Transcription du texte

La transcription du texte est donnée en version 'français standard', avec les conventions suivantes : espace entre les mots pour plus de lisibilité ; liaisons facultatives toutes notées ; consonne de liaison enchaînée sur la syllabe suivante ; schwas absents dans le contexte V # C.

læ pꝛœmje ministꝛ ira til a boljø

læ vilaz dœ boljø e tã gꝛã temwa

læ pꝛœmje ministꝛ a ã nefē deside d fœk etap dã set kœmyn o kuk dœ sa tuꝛne d la
kœzjō ã fē dane

zyskisi le sœl titkœ dœ glwæk dœ boljø ete sō vē blã sek se smiz zã swa œ fãrjō lœka
l dœ kuks a pje lwi gare katkijem o zø zolērik dœ bevlē ã mil nœf sã trãt sis e ply r
esamã sō nyzin dœ pat

zitaljen

kœskja dœk valy a boljø sœ gꝛã tœnœk

læ azak tu betmã kak læ pꝛœmje ministꝛ lase de sikki zabitqel

ki tuꝛne tuzuk zotuk de mem vil vø dekuvrik sœ kil apel la kãran

prœfōd

læ mœk dœ boljø mak blã e tã kœnãf tœ zēkje

la kœt dy pꝛœmje ministꝛ nœ ses dœ bese dœrqi le zeleksjō

kœmã ã plys evite le manifestajō ki ō ty tãdãs a s mytlije rakfwa ruk kjē lœk de v
izit zofisjel

la kot eskakre dy mō sē rjœk ki men o vilaz kœne de bakaz jak fwa kœ le zœpozã dœ
tu le bœk manifest lœk kœlœk

dœ notkœ kote a jak vwajaz dy pꝛœmje ministꝛ læ guvœnœmã pã kœtakt avek la p
œfektyr la ply pœf e sasyr kœ tu te fe ruk læ

pœteze

œk œ gꝛo detafmã dœ rœlis kœm ō nã na vy a zœkjek e de

vevifikasjō didãtite viskœ dœ pœvœke yn eksplozjō

œ zœn mãbvœ dœ lœpozisjō œkœ mem deklœkœ

dã l kwē ō ne zalu dœ notkœ libœkte sil fo mœtœ pat blãf ruk

sikyle nu n kœpœdō pa dœ la vœaksjō de zã dy pei

nu zavō læ sutjē dy vilaz œtje

dœ plys kelkœ zavtikl raky dã la depef dy sãtk lœksrœes læ nuvel

œpsœvatœk e wœst libœkte œdikœ kœ de zaktivist de kœmyn vwazin pœerak tyn zyk
ne sœd o pꝛœmje ministꝛ

kelkœ fanatik œkœ mem œtame œ zœn pœolœze dã legliz dœ sē

makē vil

lœ sēpatik mɛɤ dœ boljø n sɛ ply a kɛl sē sœ vwe
il a lœ sātīmā dœ sœ tɤuve dā zyn ēpas stypid
il sɛ tā dezɛspwɛ dœ koz deside a ekvɛk o pɤœmjɛ ministɤ
pɤɤ vɛɤifjɛ si sō vilaz etɛ vrɛmā yn etap nesɛsɛr dā la turnɛ prevy
boljø prefɛɤ ētɤ ēkɔny e tɤākil plyto kœ d sœ tɤuve o miljø dyn
bataj pɔlitik dō pɛ la televizjō sɛɛ temwē de miljō delektœɤ

2.3. Inventaire phonémique

Les réponses aux questions ci-dessous seront consignées dans le document *Commentaires* qui accompagnera le rendu de chaque locuteur.

2.3.1 Inventaire phonémique segmental (voyelles, semi-voyelles, consonnes)

Voyelles : le locuteur fait-il les oppositions suivantes ?

i y u : OUI/NON

e / ɛ

a / ɑ

o / ɔ

ø / œ

œ̃ / ē̃

ã / õ

Autres oppositions notées :

Longueur vocalique ? spécifier :

Le schwa est-il réalisé au niveau phonétique comme un [ø] ou un [œ] ?

Glissantes

Le locuteur fait-il les oppositions de surface suivantes : [j / w / ɥ] ?

SINON : quel est l'inventaire ?

Consonnes

Est-ce que l'inventaire est du type suivant : p, t, k, b, d, g, f, s, ʃ, v, z, ʒ, m, n, ŋ, j, l, R

SINON :

Consonnes supplémentaires :

Consonnes manquantes :

A l'écoute, y a-t-il un [ɲ] clair ou plutôt [nj] ?

Inventaire allophonique segmental (principaux allophones)

OBSERVATIONS :

2.3.2 Glissantes

Le locuteur fait-il systématiquement la synérèse dans le contexte C-voyelle fermée-voyelle ?

OUI/NON - COMMENTAIRES

2.3.3 Consonnes

Avez-vous noté des assimilations qui vous paraissent surprenantes ? Observez-vous une palatalisation sensible dans la phrase du texte *Nous avons le soutien du village entier* ?

3. Codage de schwa

3.1 Introduction

Dans un premier temps, l'enquête nous fournira des résultats d'une grande généralité sur la présence ou l'absence de schwa. Les données que nous recueillerons sont de deux ordres : les données sur schwa dans le texte lu et les données dans la parole spontanée. Ces données seront d'une telle ampleur qu'il est à espérer qu'elles feront progresser notre compréhension du phénomène. Un des objectifs majeurs du projet PFC, nous l'avons vu, est d'envisager de façon plus approfondie deux aspects de la phonologie du français : le schwa et la liaison. C'est pourquoi nous proposons un codage de ces deux phénomènes. De tels codages ne constituent qu'une étape préliminaire de toute analyse mais nous considérons que cette étape est nécessaire. Dans les deux cas, il s'agit de balayer largement le corpus et de recueillir des données statistiques qui permettront quelques grandes généralisations et qui serviront d'assise à des analyses plus fines. Une fois le corpus constitué et étant donné la rigidité de notre protocole qui exige l'enregistrement d'un texte unique, d'une liste de mots unique et de deux registres, nous serons en mesure de comparer les prononciations d'un très grand nombre de locuteurs sur des bases similaires. La mise au point d'utilitaires informatiques dédiés à des tâches de comptage et de classement de ces phénomènes permet d'automatiser et d'uniformiser les observations. Les variétés de français n'ont jamais été abordées de la sorte

et, pour la première fois, nous pouvons espérer obtenir des données vraiment complètes et comparables.

Le codage sera effectué sur la base de la transcription orthographique effectuée sous PRAAT. Il se fera sous une tire spécifique qui sera intitulée *codage-schwa*. Pour ce faire, on utilisera la fonction 'duplicate tier' qui donnera une réplique exacte de la transcription orthographique et on codera le schwa dans cette nouvelle tire. Une fois le codage terminé, notre utilitaire *classeur-schwa* effectuera un comptage et un classement des présences/absences de schwa.

3.2 Domaines de codage

Le codage des schwas qui a été mis au point a pris comme assise les traitements traditionnels, en particulier Dell (*Les règles et les sons*, Paris, Hermann, 1973/85), et a subi quelques modifications à la suite de commentaires faits par les transcrip-teurs qui l'ont testé.

Pour chaque locuteur il s'agit de :

- (i) coder le texte lu
- (ii) coder TROIS minutes de conversation guidée
- (iii) coder TROIS minutes de conversation libre

Ce codage demandera au transcrip-teur un gros travail d'écoute et de concentration et il est à prévoir que des erreurs se glisseront et que des sites seront omis. Nous sommes conscients de ces difficultés, mais nous sommes d'avis que, malgré les manques évidents, coder les sites de schwa est la seule possibilité dont nous disposons pour faire progresser nos connaissances à ce stade. Notons qu'un codage homogène permet une plus grande vérification des données, que la masse des observations compensera les erreurs éventuelles et que le décodage automatique qu'autorisent nos outils permettra de relever relativement aisément les sites oubliés. Un chercheur qui prévoit d'affiner l'analyse, de traiter le schwa de façon exhaustive sur son point d'enquête, multipliera sans aucun doute les contextes, mais son codage initial sera effectué dans les trois registres comme indiqué ci-dessus.

Nous partons en général du "e" graphique car il représente dans la plupart des cas une voyelle latente, c.-à-d. que l'on observe un lien direct entre la présence de cette voyelle dans la graphie et la possibilité d'une réalisation acoustique de la dite voyelle. Nous reprenons la définition de Malmberg (1976, *Phonétique française*, 5ème édition revue, Malmö) "Cet élément vocalique à prononciation facultative correspond au e de l'orthographe non suivi de consonne appartenant à la même syllabe". Autrement dit, notre point de départ est de noter tous les schwas graphiques, toutes les voyelles écrites "e" qu'elles soient prononcées ou pas. Cette prononciation peut être celle d'un schwa phonétique ou une voyelle moyenne antérieure, arrondie (la voyelle de *jeu*, ou de *peur*). Ce point de vue nous permet de rassembler un maximum de données qui seront soumises plus tard à analyse mais ne constitue en aucun cas une position théorique ni un début d'analyse. Dans le français du midi, pour les locuteurs les plus âgés, nous avons, par exemple, remarqué dans nos enregistrements que la première syllabe d'un mot comme *mener* (*menez*, *meniez*, etc.) est toujours stable et ne se distingue pas sur le plan phonétique de la première voyelle de *meunier*. Le codage de la première syllabe de

mener dans nos transcriptions ne nous force absolument pas à traiter cette voyelle comme un schwa phonologique (une voyelle latente) pour ces locuteurs. Ce qu'il permet d'établir, c'est la présence catégorique ou variable de la voyelle en question. C'est au phonologue, en fonction des observations et de ses orientations théoriques, de proposer une analyse. Si nous devons nous-mêmes caractériser le système de locuteurs pour lesquels ce "e" graphique correspond à une voyelle catégoriquement présente et phonétiquement identique à la voyelle <eu>, nous proposerions de traiter la première voyelle de *mener* comme une voyelle antérieure arrondie identique à celle de *meunier* et donc comme n'étant pas un schwa phonologique. Un autre phonologue qui voudrait mettre en avant les alternances du type *mener-mène* (dans l'esprit de Dell 1973/85) pourrait néanmoins proposer une autre solution.

Tous les schwas graphiques qui ne suivent pas une voyelle sont donc soumis au codage à deux exceptions près. Nous excluons les voyelles graphiques après voyelle, car à notre connaissance, elles ne sont jamais prononcées mais elles peuvent dans certaines variétés provoquer l'allongement de la voyelle précédente. Si tel est le cas, le transcripteur indiquera cet allongement dans ses commentaires généraux afin d'ouvrir la voie à une étude plus détaillée. Nous excluons aussi les schwas graphiques après un groupe obstruante+liquide (*grenade, mercredi*) à l'intérieur des mots mais on les codera en position finale (*entre*). Les schwas après obstruante+liquide à l'intérieur des mots sont généralement considérés comme lexicalisés, c.-à-d. qu'on leur donne le statut de voyelle pleine car ils ne sont jamais sujets à effacement. Par contre, un schwa graphique après tout autre groupe consonantique sera codé comme dans *gouvernement*. Seront aussi codés comme sites potentiels toutes les consonnes finales prononcées. En effet, un schwa final n'est pas exclusivement observé lorsque la graphie indique clairement la voyelle (*mon père*); il peut l'être aussi après une consonne graphique finale de mot comme dans *c'est pas net@*. Pour permettre de mieux comprendre ce phénomène tout comme celui de métathèse sur lequel nous reviendrons, nous proposons de coder non seulement les schwas graphiques mais aussi toutes les consonnes finales prononcées. De ce fait, dans *il va bien* prononcé [ilva] on codera le "l", mais si cette suite est prononcée *i' va bien*, la liquide ne sera pas codée.

En bref :

Le codage sur une tire spécifique dans Praat portera sur le texte, trois minutes de conversation guidée et trois minutes de conversation libre pour chaque locuteur.

Seront codés tous les schwas graphiques à l'exception des schwas qui suivent une voyelle et des schwas qui suivent un groupe obstruante+liquide à l'intérieur des mots.

Seront codées toutes les consonnes finales prononcées.

3.3 Le codage

Le codage comprend toujours quatre chiffres qui sont placés immédiatement après le "e" graphique ou avant/après la consonne concernée. Ces quatre chiffres ont pour objectif d'indiquer la présence ou l'absence de la voyelle et de délimiter les contextes de manière large. Ces chiffres vont de 0 à 5. Il n'est peut-être pas inutile de rappeler ici que dans les transcriptions orthographiques tous les chiffres devront être transcrits en graphie ordinaire,

ceci afin d'éviter toute confusion avec un codage quelconque. La date 2002 s'écrira donc *deux mille deux*.

Premier chiffre : présence / absence

Deuxième chiffre : position

Troisième chiffre : contexte gauche

Quatrième chiffre : contexte droit

3.3.1 Premier chiffre :

Schwa présent	1
Schwa absent	0
Schwa incertain	2

Les schwas de monosyllabes et ceux situés à l'intérieur des mots sont pour la plupart aisément codés et la difficulté réside souvent à déterminer de façon univoque si un schwa final est présent ou non. La valeur 2 correspond aux cas où l'écoute ne permet pas de décider de façon claire si on a un schwa ou non : par exemple, une hésitation à la fin d'un mot à finale consonantique ou une détente très vocalique de cette consonne peuvent faire hésiter le codeur. Attention, un abus de la valeur 2 indique soit une confusion dans le codage, soit la nécessité de recourir à des critères acoustiques plus fins.

3.3.2 Deuxième chiffre :

Monosyllabe	1
Première syllabe de polysyllabe	2
2ème syll. et suiv. de polysyllabe	3
Dernière syllabe de polysyllabe	4
Métathèse	5

Le deuxième chiffre a pour fonction d'explicitier la place du schwa à l'intérieur du mot. Les deux premiers contextes rendent compte des schwas souvent appelés rythmiques car leur présence/absence dépend du contexte qui les précède et du rythme. Ces schwas sont régulièrement analysés comme facultatifs et c'est à leur niveau que l'on attend la plus grande variation. Par première syllabe de polysyllabe, nous entendons des contextes comme *demain*, *petite*, *redevenir*. 3 et 4 représentent des contextes où l'absence de schwa est jugée catégorique dans la majorité des variétés observées, ce sont les schwas internes ou finaux. Le chiffre 3 marquera les schwas internes de *redevenir* ou *omelette*. Le chiffre 4 sera utilisé dans tous les cas en finale de mot : *mer_*, *mere*, *admire*, *pluriel_*, *redevenir_*.

Ce que nous appelons 'métathèse' est la présence observée d'un schwa à l'initiale pure d'un mot comme dans @j' pense au lieu de j@ pense. Ce phénomène est systématique en français cadien de Louisiane par exemple avec des mots qui commencent par "r", et ainsi revenir est très souvent réalisé @rv@nir. Certains ont proposé une analyse par métathèse car la consonne n'est jamais à la fois précédée et suivie d'un schwa et on n'observe jamais @r@v@nir ou @r@v'nir. Le fait de qualifier cet environnement de métathèse n'implique pas que nous donnions notre aval à cette analyse. Il se pourrait très bien qu'il s'agisse plutôt d'une épenthèse rythmique, et nous aurions tout aussi bien pu choisir le terme d'épenthèse. L'important ici est de distinguer ces schwas qui tendent à se multiplier dans le registre informel à l'aide d'un contexte spécial. Ainsi le début de l'énoncé isolé @j' pense pas sera codé 1532j0112 pense pas (voir plus loin pour les symboles en position 3 et 4).⁶

Dans tous les cas la métathèse doit être indiquée uniquement lorsqu'aucune autre alternative ne s'offre au transcritteur. Soit la suite *ils risquent de provoquer...* prononcée *i' risqu-@- d' provoquer*, deux 'analyses' sont disponibles : soit on indique un schwa final prononcé (*risqu@d' provoquer*), soit il y a métathèse dans le monosyllabe (*riqu' @d' provoquer*). Si le transcritteur est incapable à l'écoute d'opter pour l'une ou l'autre variante, et uniquement dans ce cas, il choisira la notation suivante : *risque2422nt 1542de...*

Notons enfin que le mot *parce que* sera analysé comme un mot simple et non pas comme un polysyllabe suivi d'un monosyllabe : le schwa de *parce* sera dans le contexte 3 alors que celui de *que* sera dans le contexte 4.

3.3.3 Troisième chiffre :

Voyelle à gauche (V(#)C@)	1
Consonne à gauche (C(#)C@)	2
Début de groupe intonatif	3
Schwa incertain à gauche	4
Groupe consonantique simplifié	5

Toutes les analyses de schwa donnent le contexte gauche de cette voyelle comme déterminant pour son maintien ou sa disparition et le troisième chiffre a pour effet d'explicitier ce contexte.

⁶Les autres cas de « métathèse » interne, comme ceux qu'on observe en Normandie où *grenier* peut être prononcé *gu@rnier* et *mercredi merc@rdi* peuvent être codées dans notre système (*ces gu@rniers = gu1212r0312niers*), bien que la notation *gu1512r0312niers* puisse également se défendre. Dans la mesure où ces exemples contiennent peut-être des voyelles stables (non assimilables en dernier ressort à des schwas), nous laissons la décision finale sur le codage de ces cas géographiquement restreints aux spécialistes des variétés en question.

La valeur 1 indique l'absence de consonne finale dans la syllabe qui précède le schwa comme dans *Jean v'nait d' partir*, contexte où le schwa est très souvent absent. La valeur 2, par contre, correspond au cas où la séquence C@ est précédée par une consonne ou plus, soit à l'intérieur d'un mot (*gouvernement, arbre*), soit à la joncture de mots (*Luc venait de partir, pour venir*). Ce contexte, on le sait, favorise le maintien du schwa dans de nombreuses variétés du français. Nous avons cependant noté qu'il était inutile de coder un schwa à l'intérieur d'un mot s'il était précédé d'une obstruante suivie d'une liquide (cf. *brebis, écrevisse*). Attention, lorsque nous parlons de voyelle et de consonne à gauche ici, ce sont des concepts phonétiques et non graphiques qui sont en jeu. Par consonne, nous entendons toute unité non voyelle, plus précisément tout segment non syllabique (les glissantes sont donc des consonnes de ce point de vue).

Une précision s'impose ici. Dans l'exemple *pour v@nir*, avec prononciation du schwa, nous présumons qu'il n'y a pas de schwa épenthétique dans la deuxième syllabe potentielle de *pour* (*pou.r?*) ce qui correspond au codage *pour0412 ve1222nir* (où le quatrième chiffre est expliqué en 3.3.4 et où, pour simplifier, nous négligeons de coder la dernière syllabe de *venir*). Au cas où l'on observe la prononciation *pour@ v@nir*, le codage prendra la forme *pour1412 ve1212nir* puisque désormais la première syllabe de *venir* a une voyelle (à savoir un schwa) à sa gauche.

Le début de groupe intonatif correspond aux schwas après une pause marquée soit en début de phrase, soit à l'intérieur d'une phrase : *Venez ici tous* ou *Vous, venez ici*. Dans ce contexte, le schwa est traditionnellement considéré comme étant assez peu sujet à l'effacement mais néanmoins attesté. Enfin le troisième chiffre permettra d'indiquer les contextes où il y a hésitation sur la voyelle finale du mot précédent (schwa incertain à gauche = 4). Reprenons l'exemple *pour venir*. Dans le cas où le transcripateur hésite sur la présence ou l'absence d'un schwa dans *pour* ([pur@???]), le codage en tiendra compte dans le mot suivant : [pur@??? v@nir] = *pour2412 ve1242nir*. En indiquant clairement cette hésitation dans le contexte de la voyelle de *venir*, nous évitons de prendre des décisions arbitraires.

Notons aussi que très souvent les groupes consonantiques finaux sont simplifiés et ceci devra être reflété dans le codage ; c'est la fonction que remplit en partie le chiffre 5. On constatera qu'en différenciant l'endroit où l'on code dans la graphie on peut donner des informations précises sur les réductions phonétiques attestées. Soit les exemples suivants pour commencer :

- | | | |
|----|----------------------------|-----------|
| a) | Titre de gloire [titr@ d@] | titre1422 |
| b) | Titr' de gloire [titr d@] | titre0422 |
| c) | Tite de gloire [tit@ d@] | tit1452re |
| d) | Tit' de gloire [tit d@] | tit0452re |

On remarquera que nous adoptons exceptionnellement (et uniquement après le chiffre 5 en troisième position) le principe qu'un schwa final non prononcé suivant une consonne elle-même absente n'est pas codé (ex. *tit1452re, tit0452re*). Nous montrerons maintenant comment coder des exemples plus complexes de réductions finales :

e)	Ministre de [ministr@ d@]	ministre1422
f)	Ministr' de [ministr d@]	ministre0422
g)	Ministe de [minist@ d@]	minist1452re
h)	Minist' de [minist d@]	minist0452re
i)	Minise de [minis@ d@]	minis1452tre
j)	Minis' de [minis d@]	minis0452tre

a), b), e) et f) ne sont pas problématiques. Les autres exemples sont codés de façon redondante en utilisant le chiffre 5 et la position du code à l'intérieur du groupe consonantique pour préciser la réduction de ce dernier. Le chiffre 5 permettra au classeur et à l'analyste d'isoler plus facilement tous les exemples concernés.

3.3.4 Quatrième chiffre :

Voyelle à droite	1
Consonne à droite	2
Frontière intonative forte	
et/ou fin d'énoncé	3
Frontière intonative faible	4

Le quatrième chiffre précise le contexte droit du schwa. Ce contexte est généralement considéré, en français standard, comme n'influant pas sur la présence ou l'absence du schwa mais ce n'est pas vrai pour toutes les variétés du français. Dans le cas 1 (voyelle à droite), le schwa devrait être absent mais sa présence peut être occasionnée par un mot dit à 'h' aspiré : *un@ honte*. Étant donnée la multiplication des schwas finaux dans certains types d'accents, on veut également pouvoir noter une prononciation comme *un' laqu@ admirable* qui sera codée *laque1411 admirable*. La valeur 2 indique la présence d'une consonne après le schwa et nous ne différencions pas entre une ou deux consonnes. À notre connaissance, le nombre de consonnes qui suivent le schwa ne joue pas un rôle déterminant dans le maintien ou la chute de la voyelle. Une différenciation plus précise du contexte droit entraînerait un codage plus lourd. Nous laissons donc à l'analyse le soin d'établir le rôle que peut jouer le nombre de consonnes à droite comme à gauche, leurs traits distinctifs ainsi que leurs affiliations syllabiques.

Les valeurs 3 et 4 distinguent deux types de frontières : une frontière intonative forte et une frontière intonative faible. La frontière intonative forte est aisément discernable, elle se manifeste par une pause marquée. Une frontière intonative faible par contre, peut exiger des écoutes répétées et/ou des analyses phonétiques plus fines. Si l'enquêteur est entraîné à discerner ce type de frontières et qu'il pense qu'elles ont une incidence sur la présence/absence de schwa, il les inclura dans son codage. Dans tous les cas où une consonne s'enchaîne avec le mot suivant, il faudra indiquer une voyelle à droite. Soit, par exemple, un locuteur méridional qui prononce habituellement *laque* avec un schwa final et qui réalise *sa laque, il l'a trouvée* [sa.la.kil.la.tru.ve] avec (1) effacement du "e" final de *laque*, (2) frontière intonative marquée sur *laque* et (3) enchaînement du [k] sur l'attaque vide de *il*. Dans un tel

cas, on peut hésiter entre deux codages : l'effacement du schwa provient-il de la frontière intonative faible à droite (*Sa laque0414, il ...*) ou du contexte segmental droit (à savoir une voyelle : *Sa laque0411, il...*) ? On considérera que l'enchaînement de la consonne [k] donne priorité au contexte vocalique à droite sur la frontière intonative : *Sa laque0411, il...*).

Résumé

Premier chiffre

Schwa présent	1
Schwa absent	0
Schwa incertain	2

Deuxième chiffre

Monosyllabe	1
Première syllabe de polysyllabe	2
2e syll et suiv. de polysyllabe	3
Dernière syllabe de polysyllabe	4
Métathèse	5

Troisième chiffre

Voyelle à gauche (V(#)C@)	1
Consonne à gauche (C(#)C@)	2
Début de groupe intonatif	3
Schwa incertain à gauche	4
Groupe consonantique simplifié	5

Quatrième chiffre

Voyelle à droite	1
Consonne à droite	2
Frontière intonative forte et/ou fin d'énoncé	3
Frontière intonative faible	4

Exemple

Un énoncé transcrit :

« Je vois vraiment pas comment ce petit peut deviner si vite comment finir les montages en kit. C'est probablement parce qu'il a appris à l'école à se prendre en charge. »

et prononcé :

«§ J' vois vraiment pas comment ce p'tit peut deviner si vite // comment finir les montages en kit. // C'est probablement pa'c' qu'il a appris à l'écol' // à s' prend' en charge. // »

sera codé :

« Je0132 vois vraiment pas comment ce1112 pe0212tit peut de1212viner si vite0413 comment finir1412 les montage0411s en kit0413. C'est probablement parce0312 qu'il0411 a appris à l'école0413 à se0112 prend0451re en charge1423. »

Commentaires :

- *parce que* est analysé comme un mot simple. Comme la liquide n'est pas prononcée, le codage de *parce* indique une voyelle à gauche : *parce0312*.

Les "e" orthographiques apparaissant après un groupe obstruante+liquide ne sont pas codés à l'intérieur d'un mot (cf. *probablement*).

- le codage suit directement le schwa. Dans un pluriel, le codage sera indiqué avant le 's', et avant la terminaison pour les verbes : *ils mangent bien* prononcé *i' mang' bien* sera codé : *ils mange0412nt bien*. On notera que l'absence de la consonne 'l' dans la réalisation *i'mang'bien* a entraîné le non codage de 'il'. En revanche, la prononciation *ils mang' bien* sera codée *il0412s mange0412nt bien*.

Rappelons que les codes sont directement insérés dans les diverses positions sans espace entre les chiffres et les lettres qui précèdent/suivent.

3.4 Questions en suspens

Nous aborderons en guise de conclusion quelques problèmes liés à notre approche. Nous avons insisté plus haut sur la nécessité impérieuse de coder tous les 'e' graphiques (sauf après voyelle et après un groupe OL). Un transcritteur attentif pourra se poser la question suivante : que faire dans le cas de *peut-être* prononcé *p'têt'*, de *déjeuner* prononcé *déj'ner*, etc. ? Nous avons choisi de ne pas traiter ces élisions dans le cadre de schwa, et nous pensons que nous sommes ici dans la problématique des positions vides mais pas dans celui du schwa.

Notre codage, et nous insistons sur ce point, n'offre qu'un balayage préliminaire des données et ne nous donne par exemple aucun renseignement sur la chute itérative de schwa. Soit la phrase *J'ai pas de semaine de libre depuis six mois* prononcée *j'ai pas d' s'main' de lib' d'puis six mois*. Le codage de la première voyelle de *s'main'* (*se0222maine*) sera identique à celui de *d'puis* (*de0222puis*), et pourtant dans le premier cas, un schwa de monosyllabe est absent alors que dans le deuxième le contexte gauche est créé par l'absence d'un schwa final de mot. De la même façon, dans la suite *il va pas revenir* prononcée *i' va pas r'v'nir* le codage du deuxième schwa indiquera deux consonnes à gauche mais ne permet pas de spécifier que la

présence de ces deux consonnes est causée par l'absence du premier schwa. Il n'est cependant pas à exclure qu'un outil informatique plus développé ne nous permette de prendre en compte ces contextes. Dans le même ordre d'idées, rappelons que le codage ne donne aucun renseignement sur la qualité de la consonne ou des consonnes qui précèdent ou suivent le schwa. Et pourtant la nature de ces consonnes joue un rôle déterminant pour la présence ou l'absence de la voyelle : *c' type* est une forme courante alors que *t' casse pas la tête* l'est beaucoup moins. Le codage propose aussi une généralisation trop importante de la position du schwa à l'intérieur du mot. Un 3 en deuxième position indique un schwa situé dans la deuxième, troisième... syllabe d'un polysyllabe. Ce manque de précision ne nous permet pas de différencier entre *gouvern@ment* et *gouvern'mental* alors que le schwa est plus facilement absent dans l'adjectif que dans le substantif.

Finalement, il existe des cas où des transformations phonétiques postérieures à la chute du schwa entraîne des difficultés de codage. Ainsi, la prononciation [ʃɛ] pour *§je sais pas* peut être codée de deux manières. Soit on se réfère à la séquence lente de départ où une consonne suit le schwa sous-jacent. Cela donne le codage : *je0132 sais pas*. Soit on se réfère à la séquence de surface où le site du schwa effacé est devant une voyelle. On pourrait alors coder *je0131 sais pas*. Nous ne tranchons pas pour de tels cas qui mettent en jeu l'interaction entre des processus phonologiques et phonétiques complexes. Le codeur doit faire un choix et s'y tenir. Nous assumons toutes ces lacunes qui devront être éliminées lors de l'analyse phonologique du phénomène.

3.5 Codage du texte

Le codage du texte suivra la même procédure. Nous reproduisons ci-dessous le texte dans lequel nous avons souligné tous les schwas susceptibles d'être prononcés ou de provoquer l'allongement de la voyelle précédente. Toutes les consonnes finales prononcées sont aussi soulignées car, dans certains dialectes, elles peuvent entraîner l'apparition d'un schwa.

Le Premier Ministre ira-t-il à Beaulieu ?

Le village de Beaulieu est en grand émoi. Le Premier Ministre a en effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux olympiques de Berlin en 1936, et plus récemment, son usine de pâtes italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand honneur ? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits habituels qui tournaient toujours autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde".

Le maire de Beaulieu - Marc Blanc est en revanche très inquiet. La cote du Premier Ministre ne cesse de baisser depuis les élections. Comment, en plus, éviter les manifestations qui ont eu tendance à se multiplier lors des visites officielles ? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère. D'un autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout est fait pour

le protéger. Or, un gros détachement de police, comme on en a vu à Jonquière, et des vérifications d'identité risquent de provoquer une explosion. Un jeune membre de l'opposition aurait déclaré : "Dans le coin, on est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous avons le soutien du village entier." De plus, quelques articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des activistes des communes voisines préparent une journée chaude au Premier Ministre. Quelques fanatiques auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus à quel saint se vouer. Il a le sentiment de se trouver dans une impasse stupide. Il s'est, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.

On peut distinguer 10 grands contextes pour les schwas graphiques :

SCHWA INITIAL D'ÉNONCÉ

Le village	§ C@ # CV
Le Premier	§ C@ # OL
Le hasard	§ C@ # H
Le maire	§ C@ # CV
De plus	§ C@ # OL
Le sympathique	§ C@ # CV

SCHWA INITIAL DE MOT POLYSYLLABIQUE

Pas de codage après un groupe obstruante+liquide à l'intérieur d'un mot

Premier

Codage : deuxième chiffre = 2

ses chemises	V # C@
en revanche	V # C@
baisser depuis	V # C@
télévision, seraient	V # C@ ou V // C@ (// = frontière rythmique)

SCHWA INTERNE

Codage : deuxième chiffre = 3, quatrième chiffre = 2

bêtement	VC@CV
gouvernement	VrC@CV
détachement	VC@CV

indiqueraient VC@rV

SCHWA FINAL POSTCONSONANTIQUE DEVANT V

Codage : deuxième chiffre = 4

Ministre ira

Ministre a

fairere étape

commune au

chemise(s) en

coursere à

quatrième aux

pâte(s) italiennes

tendance à

visite(s) officielles

côte escarpée

mène au

proche et

comme on

Jonquièrere, et

une explosion

village entier

Francere et

prépare(nt) une

chaude au

fanatiques auraient

même entamé

une impasse

écrive au

village était

une étape

préfère être

être inconnue

SCHWA FINAL POSTCONSONANTIQUE DEVANT C

Codage : deuxième chiffre = 4

village de
étape dans
cette commune
titres de
gloire de
chemises/z/ en soie
olympiques de
usine de
pâtes/z/ italiennes
Ministre, lassé
mêmes villes
villes veut
appelle "la
campagne profonde
maire de
revanche très
cote du
Ministre ne
cesse de
visites/z/ officielles
Pierre qui
village connaît
barrages chaque
chaque fois
manifestent leur
autre côté
chaque voyage
voyage du
préfecture la
s'assure que
police, comme
risquent de
jeune membre
membre de
notre liberté

patte blanche
blanche pour
quelques/z/ articles
articles parus
Dépêche du
Centre, L'Express
activistes des
communes voisines
voisines préparent
préparent une
une journée
Quelques fanatiques
jeûne prolongé
l'église de
sympathique maire
maire de
impasse stupide
cause, décidé
Ministre pour
étape nécessaire
nécessaire dans
tranquille plutôt
centre d'une
une bataille
bataille politique
politique dont

SCHWA FINAL POSTCONSONANTIQUE D'ENONCE

Codage : deuxième chiffre = 4, quatrième chiffre = 3.

italiennes.
profondes."
officielles?
colère.
Ministre.
Martinville.

stupide.

SCHWA FINAL POSTVOCALIQUE MARQUEUR DU FÉMININ

Aucun codage dans ce cas.

tournée de

anne.

soie,

escarpe du

tournée prévue

prévue.

inconnue et

Question : PRESENT VIA LONGUEUR : OUI / NON

SCHWA DE MONOSYLLABE

Codage : deuxième chiffre = 1

village d <u>e</u> Beaulieu	C@#C@#CV
décid <u>e</u> faire	V#C@#CV
cours d <u>e</u> sa	C#C@#CV
tour <u>n</u> ée d <u>e</u> la	V#C@#CV
titres d <u>e</u> gloire	OL@#C@#CV (OLV = CV ici)
local d <u>e</u> course	C#C@#CV
olympiques d <u>e</u> Berlin	C@#C@#CV
son usine d <u>e</u> pâtes	C@#C@#CV
Qu'est-c <u>e</u> qui	V#C@#CV
Beaulieu c <u>e</u> grand	V#C@#CV
car l <u>e</u> Premier	C#C@#CV
découvr <u>i</u> r c <u>e</u> qu'il	C#C@#CV
mair <u>e</u> d <u>e</u> Beaulieu	C@#C@#CV
Minist <u>r</u> e n <u>e</u> cesse	OL@#C@#CV
à s <u>e</u> multiplier	V#C@#CV
fois qu <u>e</u> les	V#C@#CV
opposants d <u>e</u> tous	V#C@#CV
Minist <u>r</u> e, l <u>e</u> gouv	OL@ # (!/?) C@ # CV
s'assur <u>e</u> qu <u>e</u> tout	C@#C@#CV

pour le protéger	C#C@#CV
détachement de po	V#C@#CV
risquent de provoq	CC@#C@#CV
membre de l'opposi	OL@#C@#CV
Dans le coin	V#C@#CV
jaloux de notre	V#C@#CV
nous ne répondons	V#C@#CV
pas de la	V#C@#CV
avons le soutien	V#C@#CV
et Le Nouvel	V#C@#CV
indiqueraient que des	V#C@#CV
l'église de Saint	C@#C@#CV
maire de Beaulieu	C@#C@#CV
Beaulieu ne sait	V#C@#CV
saint se vouer	V#C@#CV
Il a le sentiment	V#C@#CV
désespoir de cause	C#C@#CV

SUITES DE SCHWAS

C@#C@#CV V#C@#CV

sentiment de se trouver	V#C@#C@#
sentiment de se trouver	V#C@#C@ #CV
plutôt que de se trouver	V#C@#C@ ...
plutôt que de se trouver	V#C@#C@#C@ ...
plutôt que de se trouver	V#C@#C@#C@ # CV ...

SITE FRÉQUENT D'INSERTION

Codage : deuxième chiffre = 4

Marc_Blanc

Ouest_Liberté

Pour illustrer le codage d'un texte, nous donnons ci-dessous une réalisation possible de notre texte où nous avons indiqué les pauses et où la présence de schwa est indiquée graphiquement. Ce texte est suivi de son équivalent codé.

Le Premier Ministr' ira-t-il à Beaulieu ?

//Le villag' de Beaulieu // est en grand émoi. // Le Premier Ministr' a, // en effet, // décidé de fair' étap' dans cett' comun' // au cours de sa tournée de la région en fin d'année. // Jusqu'ici les seuls tit' de gloir' de Beaulieu // étaient son vin blanc sec, // ses chemis' en soie, // un champion local de cours' à pied // (Louis Garret), // quatrièm' aux jeux olympiqu' de Berlin // en mill' neuf cent trent' six, // et plus récemment // son usin' de pâ't italienn'. // Qu'est-c' qui a donc valu à Beaulieu ce grand honneur ? // Le hasard, // tout bêt'ment, // car le Premier Ministr' // lassé des circuits habituels qui tournaient toujours autour des mêm' vill', // veut découvrir ce qu'il appell' "la campagn' profond(e)".//

//Le mair' de Beaulieu //- Marc Blanc - // est en revanch' très inquiet. // La cot' du Premier Ministr(e) ne cess' de baisser depuis les élections. // Comment, // en plus, // éviter les manifestations qui ont eu tendanc' à se multiplier // lors des visites officiell' ? // La cô't escarpée du Mont Saint Pierr' // qui mèn' au villag' // connaît des barrag' // chaqu' fois que les opposants de tous les bords manifestent leur colèr'. // D'un autre côté, // à chaqu' voyag' du Premier Ministre, // le gouvernement prend contact avec la préfetur' la plus proch' // et s'assur(e) que tout est fait pour le protéger. // Or, un gros détach'ment de polic', // comm' on en a vu à Jonquièr', // et des vérifications d'identité // risquent de provoquer un' explosion. // Un jeun' membre de l'opposition aurait déclaré : // "Dans le coin, // on est jaloux de notre liberté. // S'il faut montrer patt' blanch' pour circuler, // nous ne répondons pas de la réaction des gens du pays. // Nous avons le soutien du villag' entier." // De plus, // quelques articles parus dans La Dépêch' du Centre, // L'Express, // Ouest Liberté // et Le Nouvel Observateur // indiqueraient que des activist(e)s des comun' voisin' // prépar' un' journée chaud' au Premier Ministre. // Quelques fanatiqu' auraient mêm' entamé // un jeûn' prolongé dans l'églis' de Saint Martinvill'. //

// Le sympathiqu' mair' de Beaulieu // ne sait plus à quel saint se vouer. // Il a le sentiment // de se trouver dans un' impass' stupid(e). // Il s'est, // en désespoir de caus', // décidé à écri'r au Premier Ministr' // pour vérifier // si son villag' était vraiment un' étap' nécessair' dans la tournée prévue. // Beaulieu préfèr' êtr' inconnue et tranquill' // plutôt que d' se trouver au centre d'un' bataill' politiqu' dont, // par la télévision, // seraient témoins des millions d'électeurs. //

Texte codé :

Le1132 Premier Ministre0421 ira-t-il0411 à Beaulieu ?

Le1132 village0412 de1122 Beaulieu est en grand émoi. Le1132 Premier Ministre0421 a, en effet, décidé de1112 faire0411 étape0412 dans cette0412 commune0413 au cour0412s de1122 sa tournée de1112 la région en fin d'année. Jusqu'ici les seul0412s titre0452s de1122 gloire0412 de1122 Beaulieu étaient son vin blanc sec0413, ses che1212mise0411s en soie, un champion local0412 de1122 course0421 à pied (Louis Garret), quatrième0411 aux jeux olympique0412s de1122 Berlin en mille0412 neuf0412 cent trente0412 six0413, et plus récemment son usine0412 de1122 pâte0411s italienne0413s.

Qu'est-ce qui a donc valu à Beaulieu ce grand honneur? Le hasar, tout bêtement, car le Premier Ministre lassé des circuits habituels qui tournaient toujours autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde".

Le maire de Beaulieu - Marc Blanc - est en revanche très inquiet. La cote du Premier Ministre ne cesse de baisser depuis les élections. Comment, en plus, éviter les manifestations qui ont eu tendance à se multiplier lors des visites officielles ? La côte escarpée du Mont Saint Pierre qui mène au village connaît des barrages chaque fois que les opposants tous les bords manifestent leur colère. D'un autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout est fait pour le protéger. Or, un gros détachement de police, comme on en a vu à Jonquières, et des vérifications d'identité risquent de provoquer une explosion. Un jeune membre de l'opposition aurait déclaré : "Dans le coin, on est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous avons le soutien du village entier." De plus, quelque article parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des activistes des communes voisines préparent une journée chaude au Premier Ministre. Quelque fanatique auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus à quel saint se vouer. Il a le sentiment de se trouver dans une impasse stupide. Il s'est, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver au centre d'une bataille politique dont, par la télévision, se seraient témoins des millions d'électeurs.

4. Codage de la liaison

4.1 Introduction

La notation proposée ici vise, d'une part, à permettre une transcription aussi rigoureuse que possible des liaisons observées dans l'analyse auditive d'un enregistrement (fût-il de lecture ou de parole spontanée) et, de l'autre, à favoriser une étude quantitative par extraction automatique des données. Le classement des données observées par grands types doit permettre en retour une analyse plus fine des phénomènes, éventuellement avec des outils d'analyse acoustique.

Comme c'est le cas pour notre utilitaire *PFC classeur-schwa*, le *classeur-liaison* donnera des résultats organisés et multifactoriels à partir desquels le chercheur pourra retourner aux

données. Il faut comprendre que la notation proposée ici se veut aussi simple que possible car elle doit permettre une approche globale des données. Ce n'est donc pas un accident si notre notation ne rentre pas dans des détails trop précis (par exemple, la qualité de la voyelle dans des liaisons du type (*trop*[p]*intéressant*), lesquels exigent un travail séparé de la part du chercheur concerné. L'objectif que nous nous fixons est une première approche des grands contextes de liaison : dans quels contextes la liaison est-elle toujours présente (liaison *catégorique*), dans quels contextes est-elle facultative (liaison *variable*), dans quels contextes est-elle absente ou quasi absente (liaison *erratique*) ?

Nous partons de l'hypothèse que tout comme pour le codage de schwa, la transcription de départ est orthographique sous une tire spécifique dans PRAAT. Le codage doit être effectué sous une autre tire que l'on intitulera *codage-liaison*. Nous rappelons qu'une nouvelle tire est créée à l'aide de la fonction 'duplicate tier' qui donnera une copie exacte de la transcription orthographique et que l'on codera la liaison dans cette nouvelle tire.

4.2 Domaines de codage

Le codage décrit ci-dessous sera utilisé pour tous les locuteurs dans :

- (i) le texte lu
- (ii) CINQ minutes de conversation guidée
- (iii) CINQ minutes de conversation libre

Le transcripteur décidera s'il désire ou non coder les mêmes passages pour le schwa et la liaison.

Par liaison nous entendons toute consonne latente prononcée lorsque le mot suivant a une initiale vocalique : *mes amis* (*mes [z]amis*), *petit ami* (*petit [t]ami*), *toujours ami* (*toujours [z]amis*). Il n'y a pas liaison dans *avec elle* (*avec [k]elle*) mais enchaînement tout comme dans *toujours amis* prononcé *toujours [r]amis*. La liaison peut-être enchaînée ou non. Elle est enchaînée lorsque la consonne apparaît dans l'attaque de la première syllabe du mot suivant et non enchaînée lorsqu'elle est prononcée en fin de dernière syllabe du mot qui déclenche la liaison (le 'mot liaisonnant' ci-après) :

<i>Il est ému</i>	<i>il est [t]ému</i>	liaison enchaînée
<i>Il est ému</i>	<i>il est[t] ému</i>	liaison non enchaînée

Sera donc codée toute consonne graphique finale susceptible d'être une consonne de liaison. Une consonne graphique n'est cependant pas toujours à l'origine d'une liaison et il y a aussi liaison dans *quatre enfants* (*quat' [z] enfants*), *il va à Paris* (*i' va [t]à Paris*). Ces cas de liaisons seront aussi notés tout comme les cas (rares) de liaison devant initiale consonantique *on peut venir* (*on peut[t] venir*). Nous excluons cependant du marquage quelques contextes où la liaison n'est jamais observée. Notre point de départ pour les contextes de liaison est le petit fascicule de Delattre (Delattre, Pierre 1951, *Principes de phonétique française à l'usage des étudiants anglo-américains*, Middlebury College) où un tableau détaillé liste les contextes de liaisons classées comme obligatoires, facultatives et interdites. Parmi les contextes de liaison dite interdite, nous en relevons deux que nous avons choisi d'exclure de notre précodage : un substantif singulier +..., ainsi que la conjonction *et*+... : *un soldat/afghan, un soldat/arrive, et/il arrive*. Ces liaisons seront néanmoins notées si elles sont attestées.

Tout comme pour le schwa, le codage que nous proposons ci-dessous ne permet qu'un premier balayage des données, il reste très "primitif" et ignore un ensemble de facteurs essentiels comme par exemple le contexte syntaxique. Nous pensons qu'il nous permettra cependant de confirmer que dans la parole spontanée, tout comme dans la lecture, très peu de liaisons apparaissent et que les classes de liaisons catégoriques sont peu nombreuses. Ces classes, du fait même de la faible ampleur du phénomène pourront alors aisément être identifiées.

4.3 Le codage

La notation proposée est la suivante : le mot liaisonnant se voit accoler un identifiant après la dernière lettre du mot. L'identifiant est une notation alphanumérique qui indique des valeurs où nous avons favorisé les réalisations par défaut. La complexité de la notation va donc croître avec la complexité des phénomènes phonétiques observés. Les crochets carrés représentent ci-dessous une transcription phonétique large, les chevrons la graphie. Dans les transcriptions phonétiques PH représente une pause et/ou une hésitation. Dans notre notation PH est codé par la lettre h en minuscule. Nous commençons par le nombre de syllabes du mot liaisonnant (4.3.1), nous passons ensuite à l'absence de liaison et les cas d'incertitude (4.3.2), puis au cas de liaison par une consonne non nasale (4.3.3) en séparant les liaisons enchaînées (4.3.3.1), des liaisons non enchaînées (4.3.3.2). Nous passons ensuite en 4.3.4 aux liaisons en consonne nasale (normalement [n]). Enfin, en 4.3.5, nous donnons une représentation pour les cas rares où la liaison est observée entre un mot liaisonnant et un mot à initiale consonantique ainsi que pour les liaisons sans consonne graphique.

Tous les codages comporteront au minimum deux symboles. Le premier chiffre indique le nombre de syllabes du mot liaisonnant :

1 = une syllabe

2 = deux syllabes ou plus

et le deuxième chiffre est tiré de l'ensemble {0, 1, 2, 3} où :

0 = absence de liaison ;

1 = liaison enchaînée ;

2 = la liaison non enchaînée ;

3 = incertitude ;

4 = liaison épenthétique.

On notera que dans tous les codages l'ordre des symboles est fixe.

4.3.1 Longueur du mot liaisonnant

1 = une syllabe

2 = deux syllabes ou plus

Dans tous les cas il s'agit de syllabes phonétiques et non graphiques.

les amis signal auditif = [lezami] ; codage = <les11z enfants>

beaucoup à voir [bokuavwar] → codage = <beaucoup20 à voir>

pâtes italiennes [pa titaljen] → codage = <pâtes10 italiennes>

Les valeurs en positions 2 et 3 seront explicitées plus bas. Dans notre dernier exemple, *pâtes* est prononcé en une syllabe et le codage prend en compte la prononciation.

Toutes les descriptions traditionnelles incorporent dans leurs descriptions la longueur du mot liaisonnant. Delattre (1951) par exemple, indique clairement qu'une liaison est obligatoire après un invariable monosyllabique (*dans un an*), mais facultative après un invariable polysyllabique (*pendant un an*). On sait aussi que la liaison est catégorique avec les pronoms personnels, articles, etc. qui sont tous des monosyllabes. Ce codage de la longueur du mot liaisonnant nous permettra de confirmer/infirmer l'importance déterminante de ce paramètre.

4.3.2 Absence de liaison et incertitude

L'absence de liaison donnera naissance aux codages suivants :

4.3.2.1 Pas de liaison, pas de rupture

Codage : 0

Il n'y a pas de liaison et les voyelles s'enchaînent sans rupture.

petit écrou : [ptiekru] → codage = <petit10 écrou>

Si le mot liaisonnant est un mot à consonne finale fixe au pluriel et que (a) le pluriel n'est pas réalisé et (b) la consonne fixe finale s'enchaîne, on utilisera cette même notation : *pâtes italiennes* [patitaljen] se notera <pâtes10 italiennes>. Evidemment, *pâtes italiennes* [patzitaljen] se notera <pâtes11z italiennes>.

4.3.2.2 Pas de liaison, pause-hésitation et/ou coup de glotte

Codage : 0h

0 = absence de liaison

h = hésitation

petit écrou [pti PH ekru] → <petit10h écrou>

Ce codage s'applique à un ensemble de cas où le locuteur hésite, fait une légère pause et/ou insère un coup de glotte. On notera donc : *petit écrou* [pti?ekru] ou [pti PH ?ekru] → <petit10h écrou>. Cela ne signifie pas que toutes ces réalisations ont nécessairement la même valeur mais qu'on laisse l'examen de ces cas à une étape ultérieure.

Il est important de signaler que cette notation est également utilisée pour le cas suivant où le mot précédent se termine par une consonne qui n'est pas enchaînée :

pâtes italiennes [pat PH italjen] ou [pat PH ?italjen] → <pâtes10h italiennes>

4.3.2.3 Incertitude

Codage : 3

3 = incertitude

petit écrou signal auditif = [pti ? ? ? ? ekru] → codage = <petit13 écrou>

Le chercheur n'a pas réussi à savoir si la liaison s'était effectuée ou non. Cas relativement rare mais dont la notation reconnaît la possibilité permettant éventuellement un retour ultérieur sur les données.

4.3.3 Liaison réalisée

4.3.3.1 Liaison enchaînée

Codage : 1K(h), où

1 = liaison réalisée

K = la valeur phonémique de la consonne de liaison (ex. z pour *mes amis* [mezami]).

(h) = présence ou absence de pause-hésitation

La liaison enchaînée peut avoir son origine dans une consonne orthographique ou non. Dans le premier cas, deux sous-cas sont envisagés : soit la consonne est à l'attaque sans pause-hésitation entre le mot liaisonnant et le mot déclencheur (cas par défaut) ; soit une pause-hésitation précède l'enchaînement de la consonne de liaison.

(i) Liaison enchaînée sans pause-hésitation

Codage : 1K

petit écrou signal auditif : [pti tekru] → codage <petit1 1t écrou>

1t = Liaison enchaînée en [t].

Dans tous les cas de ce type, on a le chiffre 1 suivi de la consonne de liaison, à savoir les consonnes orthographiques avec leur valeur API. Autre exemple : *il a trop été en Afrique* en [tro pete] → <trop1 1p été>.

Bien que les descriptions récentes du français ne mentionnent le plus souvent que les consonnes de liaison suivantes [t, z, n, g, p] (et marginalement [k] (*long été*)), il faut prévoir que certaines variétés puissent fonctionner d'une autre manière. On autorisera donc ici toutes les lettres consonantiques avec leur valeur phonétique SAMPA : <b c d f g h j k l m n p q r s t v w x z > = SAMPA [b c d f g h j k l m n p q r s t v w x z]. Un certain nombre de ces valeurs ne sont jamais utilisées à notre connaissance dans les variétés du français mais l'avantage de s'autoriser le plus de symboles possibles est que la notation peut éventuellement devenir allophonique si un chercheur le désire. L'inclusion de toutes ces lettres ne coûte absolument rien du point de vue de la conception de l'utilitaire informatique et de la complexité de la notation, d'où la position libérale choisie ici.

(ii) Liaison enchaînée mais avec pause-hésitation

Codage : 1Kh

petit écrou [pti PH tekru] → codage <petit11th écrou>

Liaison enchaînée en [t] mais où le [t] est précédé par une pause ou une hésitation (sans rentrer dans la finesse phonétique de ce qu'une pause-hésitation peut inclure : par ex. glottalisation).

(iii) Liaison épenthétique enchaînée

Codage 4K

Il va à Paris [ilvatapari] → codage <va14t à Paris>

Le chiffre 4 sera utilisé dans tous les cas de liaison épenthétique. Comme pour la liaison à origine graphique, la valeur phonétique de la consonne de liaison sera indiquée en SAMPA.

4.3.3.2 Liaison non enchaînée

Codage : 2K

petit écrou [ptit PH ?ekru] → codage <petit12t écrou>

2t = Liaison non enchaînée en [t] avec ou sans pause-hésitation et/ou coup de glotte perceptible à l'initiale du mot suivant. Ainsi *trop été* [trop?ete] et [trop PH ete] → <trop12p été>.

La liaison non enchaînée donne lieu à des ajustements phonétiques complexes entre le mot liaisonnant et le mot déclencheur. L'étude de ces phénomènes exige une étude auditive et acoustique fine une fois tous les cas repérés.

4.3.4 Liaison en consonne nasale

En ce qui concerne les liaisons en consonne nasale, le codage offre un choix. Soit l'analyse note simplement la présence d'une consonne nasale sans s'inquiéter de la qualité de la voyelle qui précède. Dans ce cas, nous utilisons la notation proposée jusqu'ici.

Exemples :

1n : *bon appareil* [bõ naparej] → <bon11n appareil> Liaison enchaînée en [n].

0 : *bon appareil* [bõaparej] → <bon10 appareil> Pas de liaison, pas de rupture

2n : *bon appareil* [bõn ?aparej] → <bon12n appareil> Liaison non enchaînée.

Cependant, les recherches sur la liaison en consonne nasale montrent que certaines variétés favorisent pour le mot liaisonnant soit une voyelle orale (type *bon été* [bõ nete]), soit une voyelle nasale (type *mon été* [mõ nete]) soit les deux types en question selon les mots et les contextes.

Nous offrirons donc pour ce domaine précis une notation facultative plus fine qui permettra de coder une différence entre VO (voyelle orale) et VN (voyelle nasale) dans le mot liaisonnant. Nous réserverons cependant cette notation aux seuls cas où la liaison en [n] est enchaînée :

Liaison en consonne nasale, version étendue :

1nVO : *bon été* [bɔ̃ nete] → <bon11nVO été>. Liaison enchaînée en [n] ; la dernière voyelle du mot liaisonnant est orale d'où VO pour Voyelle Orale.

1nVN : *bon été* [bɔ̃ nete] → <bon11nVN été>. Liaison enchaînée en [n] ; la dernière voyelle du mot liaisonnant est nasale, d'où VN pour Voyelle Nasalisée.

Bien évidemment la transcription d'un (sous)-corpus donné doit faire un choix dès le départ entre la notation large et la notation étroite avec VO/VN. Cette notation sera facultative pour les dix minutes de parole spontanée mais elle sera utilisée dans le texte (voir 4.4.2).

4.3.5 Liaison devant un mot à initiale consonantique

Codage : 1tC

on peut s'opposer [ɔ̃pøtsɔpoze] → codage <on peut11tC s'opposer>

1tC = Liaison inattendue en [t] ici entre un mot liaisonnant et un mot à initial consonantique. Comme dans les autres exemples, le [t] peut être remplacé par toute consonne liaisonnante. Le symbole C indique que le mot suivant est à initiale consonantique.

4.3.6 Liste des codes envisagés

En examinant les codages ci-dessous, on notera premièrement que l'ordre des symboles est strict. Deuxièmement, K est une variable pour une consonne de liaison. K est pris dans l'ensemble des symboles consonantiques du français écrit avec leur valeur phonémique : K ∈ {b,... z, B, ...Z}.

Premier symbole

- 1 Mot liaisonnant monosyllabique
- 2 Mot liaisonnant polysyllabique

Symboles 2 et suivants

- 0 Absence de liaison
- 0h Absence de liaison mais pause et/ou hésitation et/ou coup de glotte
- 1K Liaison enchaînée en K

1Kh	Liaison enchaînée en K avec pause hésitation
1KC	Liaison enchaînée en K suivie d'un mot à initiale consonantique
2K	Liaison non enchaînée
3	Incertitude
4	Liaison épenthétique

Facultatif dans les dialogues, obligatoire pour le texte :

1nVN	Liaison enchaînée en [n] précédée d'une voyelle nasale
1nVO	Liaison enchaînée en [n] précédée d'une voyelle orale

4.4. Comment procéder

4.4.1 Méthode

En premier lieu, l'analyse devra consister à repérer tous les contextes de liaison théoriquement possibles. Par 'théoriquement', nous entendons ici 'tous les contextes repérés où une consonne latente est suivie d'un mot à initiale vocalique. Ainsi, les séquences suivantes :

- (a) en allant
- (b) les enfants
- (c) ils étudient
- (d) grand émoi
- (e) jeux olympiques
- (f) circuits habituels
- (g) tellement imprévu
- (h) les enfants arrivent
- (i) projet identique
- (j) et avant de partir

seront toutes codées à l'exception des deux dernières (i) et (j). Les descriptions du français parlé contemporain ont en général signalé que la liaison était attestée dans tous les exemples (a)-(g). De plus, pour le français standard de France, les exemples (a) à (e) sont habituellement décrits comme exhibant des liaisons catégoriques :

- (a') en[n]allant
- (b') les[z]enfants
- (c') ils[z]étudient
- (d') grand[t]émoi
- (e') jeux[z]olympiques

En revanche, dans (f) et (g) la liaison est le plus souvent présentée comme variable :

- (f') circuits[z]habituels ou circuits//habituels
- (g') tellement[t]imprévu ou tellement//imprévu

Enfin, dans (h) et (i), (j), les spécialistes ont en général signalé qu'en français parlé contemporain la liaison était absente dans les exemples :

(h') les enfants // arrivent (SN pluriel + SV)

(i') projet // identique (N singulier + X)

(j') et // avant de partir (et +...)

Nous adoptons ici une approche aussi ouverte que possible en excluant un minimum de contextes possibles. Nous pensons que les liaisons en [z] sont les plus fréquentes et (h') est à notre avis concevable dans certains styles (voir plus bas la liste de Delattre (1951)). C'est pourquoi nous avons décidé d'exclure uniquement les contextes de (i') et (j'). Nous savons cependant que si les données sur le texte sont comparables puisque les contextes de liaisons auront été préfléchés, il est à craindre que dans la transcription de la parole spontanée, de nombreux sites potentiels soient omis. Le transcripateur n'entend pas la liaison et même s'il/elle relit le texte attentivement la possibilité de liaison lui échappant, il laissera passer un ensemble de contextes. Sera alors faussé le rapport entre les liaisons perçues et les liaisons potentielles. Ce rapport est cependant d'un intérêt divergent selon qu'il s'agit de liaisons considérées comme catégoriques ou erratiques.

Les études récentes sur la liaison concordent sur la régression du phénomène et nos premiers résultats montrent que certaines des distinctions classiques sont loin de se vérifier dans les faits. Par exemple, même dans la lecture du texte PFC, la liaison dans *grand émoi* n'est pas universelle et le mot composé *jeux olympiques* se révèle variable dans certains de nos enregistrements (au Québec par exemple nos enregistrements fournissent *jeux[z]olympiques* et *jeux//olympiques*). Nous pensons de ce fait que le transcripateur devra apporter une attention particulière aux sites souvent décrits comme obligatoires et facultatifs afin de confirmer ou d'infirmer la présence de liaison catégorique ou variable dans ces cas. Afin d'aider le transcripateur dans ce travail de pré-repérage, nous reprenons ci-dessous les contextes de liaison obligatoire et de liaison facultative selon Delattre (1951) :

OBLIGATOIRE

Art +

Les_enfants, un_enfant, aux_enfants

Adj +

Ces_enfants, son_ami, deux_amis, quelles_histoires, plusieurs_hommes, un_gros_arbre

Pronom personnel +

Vous_êtes, vous_y_êtes, allons-nous_en

Verbe + pron. pers.

Allez_y, vient_elle

...est (impersonnel) +

c'est_évident, il est_interdit

préposition, adverbe monosyllabique +

dans_un an, en_hiver, très_utile, bien_étrange

FACULTATIVE

Nom pluriel +

Des plats_exquis, tes parents_attendent, j'ai deux mots_à dire

Pronom (pas pers.) +

Plusieurs_écoutent, ce dont_elle parle, toutes_auront peur

Pronom pers. postposé *nous, vous* +

Qu'avez-vous_à faire, amusons-nous_un peu

Verbe +

Je suis_imbattable, il vivait_encore, ils le font_écrire, avez-vous bien_observé, nous sommes_allés

Préposition, adverbe polysyllabique +

Depuis_un an, avant_une semaine, jamais_à l'heure, tellement_occupé

Conjonction monosyllabique +

Mais_alors, puis_on partira

X + invariable

les maisons_et les prés, bons_ou mauvais, les miens_aussi, il s'attend_à partir, il a toujours_à redire

L'intérêt de ce rappel est d'aider le transcripneur à repérer des contextes potentiels mais seuls les résultats du codage sur plusieurs centaines de témoins nous guideront pour une nouvelle classification. On remarquera que certaines catégories émergeront de l'analyse sans avoir à être explicitement fléchées dans le pré-repérage. Ainsi, nous avons exclu du codage le contexte Nom singulier + adj. (en dehors des composés lexicalisés), une liaison dans ce contexte sera néanmoins notée. Ainsi, si le transcripneur venait à entendre *savant[t]authentique*, il notera <savant21t authentique>. Lorsqu'il consultera les listes de liaisons que fournira le compteur classeur-liaison, il pourra évaluer la force d'un tel exemple (est-ce le seul exemple de ce type ? est-ce un lapsus ? se pourrait-il que cette séquence soit un quasi-composé ou une séquence adjectif + substantif et non substantif + adjectif ? etc.).

4.4.2 Illustration : codage du texte

Le but de cette section est d'illustrer le codage du texte PFC. En 4.4.2.1 nous présentons un premier pré-repérage possible des contextes de liaison dans le texte suivi en 4.4.2.2 par les cas attestés dans un exemple construit pour favoriser l'éventail notationnel. Rappelons que dans le cas de liaison avec consonne nasale enchaînée, la qualité (orale ou nasale) de la voyelle précédente sera explicitée.

En 4.4.2.3, le classement qu'effectuera le compteur est fourni pour ce codage précis.

4.4.2.1 Pré-repérage. Echantillon texte PFC

Le Premier Ministre ira-t-il à Beaulieu ?

Le village de Beaulieu est*** en grand*** émoi. Le Premier Ministre a en*** effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises*** en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux*** olympiques de Berlin en mille neuf cent trente six, et plus récemment, son*** usine de pâtes*** italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand*** honneur ? Le hasard, tout bêtement, car le Premier

Ministre, lassé des circuits*** habituels qui tournaient toujours*** autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde".

Le maire de Beaulieu - Marc Blanc – est*** en revanche très*** inquiet. La cote du Premier Ministre ne cesse de baisser depuis les*** élections. Comment***, en plus, éviter les manifestations qui ont*** eu tendance à se multiplier lors des visites*** officielles ? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les*** opposants de tous les bords manifestent leur colère. D'un*** autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout*** est fait pour le protéger. Or, un gros détachement de police, comme on*** en*** a vu à Jonquières, et des vérifications d'identité risquent de provoquer*** une explosion. Un jeune membre de l'opposition aurait déclaré : "Dans le coin, on*** est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous*** avons le soutien du village entier." De plus, quelques*** articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des*** activistes des communes voisines préparent*** une journée chaude au Premier Ministre. Quelques fanatiques*** auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus*** à quel saint se vouer. Il a le sentiment de se trouver dans*** une impasse stupide. Il s'est***, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment*** une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver*** au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.

Tous les contextes sont identifiés avec leur classification selon Delattre (1951) dans la grille ci-dessous :

LIAISONS	
1. <i>est_en</i>	FAC
2. <i>grand_émoi</i>	OBL
3. <i>en_effet</i>	OBL
4. <i>chemises_en soie</i>	FAC
5. <i>jeux_olympiques</i>	OBL
6. <i>Berlin_en</i>	INT
7. <i>son_usine</i>	OBL
8. <i>pâtes_italiennes</i>	FAC
9. <i>grand_honneur</i>	OBL
10. <i>circuits_habituels</i>	FAC
11. <i>toujours_autour</i>	FAC

12. <i>est_en revanche</i>	FAC
13. <i>très_inquiet</i>	OBL
14. <i>les_élections</i>	OBL
15. <i>comment_en plus</i>	INT
16. <i>ont_eu</i>	FAC
17. <i>visites_officielles</i>	FAC
18. <i>les_opposants</i>	OBL
19. <i>un_autre côté</i>	OBL
20. <i>tout_est</i>	OBL
21. <i>on_en</i>	OBL
22. <i>en_a vu</i>	OBL
23. <i>provoquer_une</i>	FAC
24. <i>le coin_on</i>	INT
25. <i>on_est</i>	OBL
26. <i>nous_avons</i>	OBL
27. <i>quelques_articles</i>	OBL
28. <i>des_activistes</i>	OBL
29. <i>préparent_une j.</i>	FAC
30. <i>fanatiques_auraient</i>	FAC
31. <i>plus_à quels</i>	FAC
32. <i>dans_une</i>	OBL
33. <i>s'est_en</i>	FAC
34. <i>vraiment_une étape</i>	INT
35. <i>trouver_au</i>	FAC

Le lecteur attentif aura remarqué que nous avons exclu du repérage et du codage 6 et 24, deux instances de substantifs singuliers + X, domaine où la liaison est absente dans tous les styles et que nous avons décidé de ne pas coder à moins évidemment qu'une liaison ne se fasse entendre.

4.4.2.2 Annotation illustrative de la liaison

Pour le texte sera utilisée la version étendue de la liaison en consonne nasale (voir 4.3.4.).

Le Premier Ministre ira-t-il à Beaulieu ?

Le village de Beaulieu est11t en grand12h émoi. Le Premier Ministre a en11nVO effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises20 en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux11z olympiques de Berlin en 1936, et plus récemment, son11nVN usine de pâtes10h italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand11t honneur ? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits20 habituels qui tournaient toujours21z autour des mêmes villes, veut découvrir ce qu'il appelle "la campagne profonde".

Le maire de Beaulieu - Marc Blanc – est12t en revanche très11z inquiet. La cote du Premier Ministre ne cesse de baisser depuis les11z élections. Comment20, en plus, éviter les manifestations qui ont10 eu tendance à se multiplier lors des visites20h officielles ? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les11z opposants de tous les bords manifestent leur colère. D'un11nVN autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout11t est fait pour le protéger. Or, un gros détachement de police, comme on11nVO en11nVO a vu à Jonquière, et des vérifications d'identité risquent de provoquer23 une explosion. Un jeune membre de l'opposition aurait déclaré : "Dans le coin, on11nVO est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous11z avons le soutien du village entier." De plus, quelques21z articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des11z activistes des communes voisines préparent20 une journée chaude au Premier Ministre. Quelques fanatiques20 auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus10 à quel saint se vouer. Il a le sentiment de se trouver dans11z une impasse stupide. Il s'est12t, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment20 une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver20 au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.

4.4.2.3 Résultats escomptés du classeur liaison

Le classeur liaison permettra d'extraire les données de plusieurs façons en posant des questions diverses : liste de toutes les liaisons dans un corpus donné, liste de liaisons de divers types, nombre de liaisons enchaînées vs. non enchaînées, nombre de non-liaisons, nombre de liaisons en /z/, en /t/, en /p/. Par exemple, aux questions Q ci-dessous, il fournira des réponses R du type suivant :

Q : liste de tous les cas codés pour les liaisons dans le corpus :

R :

est11t en

grand12h émoi.

en1 1nVO effet
chemises20 en
jeux1 1z olympiques
son1 1nVN usine
pâtes10h italiennes.
grand1 1t honneur
circuits20 habituels
toujours2 1z autour
est12t en
très1 1z inquiet
les1 1z élections
comment20 en
ont10 eu
visites20h officielles
les1 1z opposants
D'un1 1nVN autre
tout1 1t est
on1 1nVO en

en1 1nVO a
provoquer23 une

on1 1nVO est
nous1 1z avons
quelques2 1z articles
des1 1z activistes
préparent20 une
fanatiques20 auraient
plus10 à
dans1 1z une
s'est12t
vraiment20 une
se trouver20 au

Q : Liste des liaisons avec 'est' dans le corpus

R :

est1 1t en

est12t en

s'est12t en

Q : Nombre de liaisons réalisées avec 'est' dans le corpus

R : 3

Q : Liste des liaisons non effectuées avec 'est' dans le corpus

R : 0

Le texte sera codé sur le modèle donné en 4.4.2.2 par tous les transcrip-teurs pour tous les témoins. De la même façon, seront aussi codées 10 minutes de dialogue (5 minutes de conversation guidée et 5 minutes de conversation libre). Bien évidemment, les chercheurs sont encouragés à dépasser ces limites pour leur propre exploration de la liaison.

FORMAT DES RENDUS 2002 ET 2003

Jacques Durand, Bernard Laks et Chantal Lyche

Le projet PFC est basé sur un ensemble d'enquêtes réalisées de façon décentralisée dans le monde francophone. Ces enquêtes doivent être assez cohérentes et substantielles pour permettre les recherches présentes et à venir. En termes scientifiques, nous nous sommes assigné deux phases principales : 2002-2003 (constitution et homogénéisation de la base de données et première exploitation) et 2004-2007 (exploitation à grande échelle). Même si le projet est à long terme, nous avons choisi décembre 2003 comme date-butoir pour la remise des résultats scientifiques pour chacune des enquêtes, résultats qui seront publiés sous la forme de plusieurs ouvrages sous la direction de Jacques Durand, Bernard Laks et Chantal Lyche. Les enregistrements, la numérisation et les codages doivent donc être terminés d'ici avril 2003, pour pouvoir être intégrés à la base de données commune et faire l'objet d'analyses préliminaires par les outils que nous développons.

Pour chaque témoin, depuis le début du projet, nous envisageons environ 1h d'enregistrement en tout, correspondant à la lecture de la liste de mots et du texte, à la conversation guidée (15-25 mn environ) et à la conversation libre (20-30 mn). Dans la structure fédérative que nous avons adoptée, il est nécessaire que les tâches soit partagées. Nous n'intégrerons à la base de données que les enquêtes qui sont numérisées, transcrites et codées pour les phénomènes communs au projet (inventaires phonologiques, schwa, liaison).

1. Rendu données

Pour chaque locuteur, il faudra fournir 4 fichiers .wav, une fiche d'information auxquels s'ajoutera un fichier de notes d'enquête.

1.1 Fichiers .Wav

Les 4 fichiers .wav correspondent à la liste de mots, au texte, à la conversation guidée et à la conversation libre.

Les paramètres d'enregistrement sont les suivants :

Canaux : Mono

Taux d'échantillonnage : 22 Khz (22050 Hz)

Taille d'échantillonnage : 16 bits

Lors de la numérisation de la liste de mots et du texte, il est impératif de supprimer les commentaires qui précèdent et qui suivent la liste ou le texte à proprement parler. D'autre part, nous n'intégrerons pas à la base de données commune les listes ou textes complémentaires spécifiques à certaines enquêtes.

Pour identifier les fichiers, nous avons mis au point un étiquetage où chaque locuteur et chaque activité sont identifiés de façon unique par une séquence de 8 symboles alphanumériques. Nous partirons d'exemples hypothétiques avant d'examiner les principes sous-jacents à ces noms :

31cmd1mw.wav : fichier son d'un enregistrement fait dans le département français 31 (Haute Garonne), au point d'enquête c (Toulouse banlieue), du témoin md1 (Marie + Dubois + 1) qui lit la liste de mots (m) sous sa forme son (w), information qui sera également codée par l'extension .wav.

sgabg3lw.wav : fichier son d'un enregistrement effectué en Suisse (s), à Genève (g), au point a, témoin bg3, Blanche Garou indice 3 car l'enquête inclut aussi Bernard Garou (bg1) et Bianca Garou (bg2) ; le fichier concerne la conversation libre (l) sous sa forme w, ce qui est également codé par l'extension .wav.

Les principes sont les suivants :

Position 1 et 2 : pour la France, le département (01, 31, etc.), pour les autres pays l'initiale b (Belgique), c (Canada), s (suisse), etc., suivie d'une initiale pour la ville ou la région.

Position 3 : l'indice du point d'enquête, à savoir la lettre a s'il n'y a qu'une enquête, sinon b à z pour les divers points d'enquête par département ou pays (après accord entre les enquêteurs et la coordination du projet pour éviter des homonymies).

Position 4, 5 et 6 : initiales du témoin prénom + nom + chiffre. Le chiffre est 1 s'il n'y a qu'un seul témoin avec les initiales en question. Au-delà de 1, les chiffres sont assignés en fonction de l'ordonnancement alphabétique des prénoms.

Position 7 : m (pour les mots de la liste), t (pour texte), g (pour conversation guidée), l (pour conversation libre).

Position 8 : w (pour wave), t (pour texte). Pour .wav et .txt, notre codage est redondant mais il ne l'est pas dans le cas de fichiers d'analyse puisqu'on trouvera plus bas g (pour textgrid).

1.2 Fiche d'information

Dans le protocole d'enquête nous demandons que pour chaque témoin soit remplie une fiche d'information. Cette fiche sera donnée sous format .txt (ou .doc). Cette fiche d'information a

un identifiant global en positions 7 et 8 : les deux lettres cv (mnémoniques pour curriculum vitae) qui reprend les mêmes principes que ceux des fichiers son. On aura donc par exemple :

cqajp1cv.txt : fichier concernant une enquête faite au Canada (c) dans le Québec(q), point d'enquête a (Montréal), témoin jp1 (Jean Paloque 1) et fournissant la fiche d'information cv sous format texte : .txt

1.3 Notes d'enquête

Pour chaque enquête, nous demandons également 1 fichier (.txt ou .doc) qui sera appelé : « Notes d'enquête ». Dans ce fichier seront consignées de manière synthétique toutes les informations pertinentes sur l'enquête (noms des enquêteurs, mode de sélection des enquêtés, relation entre les enquêtés, suggestions pour des enquêtes futures dans la même localité, etc.).

Ce fichier « Notes d'enquête » a un identifiant sous la forme suivante :

cqc00000.txt

31a00000.txt

à savoir les 3 symboles correspondant au point d'enquête (par exemple, c pour Canada, q pour Québec, point d'enquête c ou encore 31a pour enquête a en Haute Garonne) suivis de 5 zéros, suivis de l'extension .txt.

2. Rendus transcriptions

Pour chaque locuteur, il faudra fournir trois fichiers .TextGrid : conversation dirigée, conversation libre, lecture du texte.

Nous demandons des transcriptions orthographiques alignées sous PRAAT de 10 mn pour la conversation guidée et de 10 mn pour la conversation libre. Attention, ces fichiers ne doivent pas être découpés en sous-fichiers. La transcription orthographique se fera à l'intérieur de la première tire créée lors de la création d'un fichier texte. Rappelons que cette tire a la forme 31cmd1_transcript-graphe. Voir document PFC : Transcriptions des données : outil et conventions 1.2.2.

On notera qu'il faut également transcrire le texte lu sous Praat. En effet, le codage du schwa et de la liaison se fait sous PRAAT à partir de l'alignement texte lu/son sous Praat pour deux raisons : a) les lectures donnent lieu à des répétitions, des omissions ou des écarts qui sont importants pour les codages, b) un codage directement sur des fichiers .txt ou .doc ne permet pas une intégration et des révisions dans Praat.

Le rendu sera donc sous la forme de trois fichiers du type textgrid (sous Praat) signalés par la lettre g en huitième position des noms. Soit, par exemple, pour le témoin Marie Dubois (31cmd1) considérée plus haut :

31cmd1tg.TextGrid = transcription alignée du texte lu (t) sous forme textgrid (g) avec extension .TextGrid

31cmd1gg.TextGrid = transcription alignée de la conversation guidée (g) sous forme textgrid (g) avec extension .TextGrid

31cmd1lg.TextGrid = transcription alignée de la conversation libre (l) sous forme textgrid (g) avec extension .TextGrid

Le choix de la localisation des 10 mn à transcrire orthographiquement est laissé aux responsables des enquêtes. Toutes choses égales par ailleurs, commencez la transcription au début de l'enregistrement.

3. Rendus analyses

Il y a trois types d'analyses communes à toutes les enquêtes : l'inventaire segmental, la liaison et le schwa. Les analyses du schwa et de la liaison consistent dans un premier temps en des codages faits sous PRAAT sous deux tires différentes.

1) L'analyse de l'inventaire phonologique du système de chaque locuteur se fera à partir de la fiche établie à la section 2.2 des Directions d'analyse. Rappelons que, dans l'état actuel des choses, cette fiche est ouverte. Elle permet des développements brefs ou longs selon les observations des enquêteurs. Nous autoriserons deux types de rendus en ce qui concerne les notations phonétiques : soit avec SAMPA, soit avec la police phonétique de référence pour PFC : SIL DOULOS. On aura donc un fichier s (pour système phonémique) de type texte (t) (pour ceux qui travaillent avec SAMPA) ou de type doc (pour ceux qui utilisent SIL DOULOS). Si nous continuons avec le témoin Marie Dubois considéré jusqu'ici, nous aurons :

31cmd1st.txt = s pour système phonémique sous forme t avec extension .txt

ou bien

31cmd1st.doc = s pour système phonémique sous forme t avec extension .doc

2) L'analyse du schwa porte sur 3 enregistrements : le texte, la conversation guidée (portion de 3 MINUTES) et la conversation libre (portion de 3 MINUTES). Dans une première phase, la création de nouveaux fichiers avait été envisagée. Il est apparu plus commode que les codages se fassent dans le fichier textgrid principal sur des tires différentes. Pour effectuer le codage de schwa, il conviendra tout d'abord de créer une nouvelle tire (fonction 'Duplicate Tier') directement au-dessous de la tire Transcript-graphe. Voir document PFC :

Transcriptions des données : outil et conventions 1.5.0. Cette deuxième tire sera intitulée pour notre locuteur témoin : 31cmd1_schwa.

3) L'analyse de la liaison porte sur 3 enregistrements : le texte, la conversation guidée (5 MINUTES) et la conversation libre (5 MINUTES). Le codage de la liaison se fera sur une troisième tire créée comme précédemment à l'aide de la fonction 'Duplicate Tier'. Pour notre locuteur témoin cela donnera : 31cmd1_liaison.

Les codages pour schwa et la liaison se font sous PRAAT à partir de la transcription orthographique alignée. Garder l'intégrité des fichiers textgrid dans la mesure où les codages minimaux de départ seront sans doute élargis par la suite. Les passages codés pour le schwa et la liaison peuvent coïncider, mais cela n'est pas une nécessité.