

Ch. 6 Le projet 'Phonologie du français contemporain' (PFC) et sa méthodologie*

Jacques DURAND
Chantal LYCHE

*Ce travail est la version préfinale d'un chapitre publié sous le même titre dans E. Delais-Roussarie et J. Durand (eds) *Corpus et variation en phonologie du français : méthodes et analyses*. Toulouse : Presses Universitaires du Mirail. Pour toute citation, se reporter à la version publiée.

1. Introduction

Si la phonologie moderne tend trop souvent à considérer le français comme une langue uniforme très standardisée, ce *français de référence* (Morin 2000), n'est unique que dans la représentation que s'en fait le phonologue qui le décrit.¹ La variation, qu'elle soit géographique, sociale ou stylistique est tout simplement inhérente au système, y est omniprésente et son traitement représente l'un des grands enjeux de la linguistique moderne. Le projet *Phonologie du français contemporain : usages, variétés et structures* (PFC ci-après), coordonné par Jacques Durand, Bernard Laks et Chantal Lyche,² s'inscrit dans cette perspective. Nous partons de la constatation qu'il est nécessaire de poursuivre le travail de description entrepris depuis au moins un siècle par les spécialistes de la communication parlée pour, entre autres, (a) fournir une meilleure image du français parlé dans son unité et sa diversité, (b) mettre à l'épreuve les modèles phonologiques sur le plan synchronique et diachronique, (c) constituer une base de données importante sur le français oral à partir d'une méthodologie commune, (d) favoriser les échanges entre les

¹ Nous remercions Bernard Laks et Jean-Michel Tarrier pour leurs conseils dans la préparation de cet article.

² Le projet PFC fédère au bas mot une trentaine de chercheurs et de nombreux étudiants : voir les Remerciements en fin d'article. On trouvera des informations de tout ordre concernant le projet à l'adresse internet suivante: <http://infolang.u-paris10.fr/pfc>.

connaissances phonologiques et les outils de traitement automatique de la parole, (e) élargir et renouveler les données pour l'enseignement du français et de la linguistique française.

Dans ce travail, nous n'essayerons pas de replacer le projet PFC dans son contexte historique ni de présenter et de discuter diverses approches théoriques de la variation (cf. Durand, Laks & Lyche 2002 et ce volume). Nous présenterons en revanche la méthodologie du projet PFC de façon détaillée en explicitant notre protocole d'enquête, nos techniques de transcription, de codage et d'analyse, et quelques-unes de nos ambitions théoriques.

2. La méthode d'enquête

L'intitulé complet que nous avons choisi pour notre projet *Phonologie du français contemporain : usages, variétés et structure* résume les principales orientations qui sont les nôtres. Le terme *Phonologie* y est pris dans un sens large et ouvert qui inclut la phonétique que nous ne voyons pas comme une discipline à part, encore moins annexe, mais comme une manière complémentaire d'envisager les structures cognitives et l'activité langagière dans sa dimension phonique (voir Durand & Laks 2002a, b). Le français qui nous intéresse est le français *contemporain*, non point comme objet normé et aseptisé, mais dans les *usages* que révèlent nos enquêtes. Ces enquêtes visent à l'élaboration d'un grand corpus oral de plus de 400 heures d'enregistrements comprenant quatre registres différents et une ample couverture géographique. Le français que nous cherchons à décrire n'est pas simplement celui de la bourgeoisie parisienne mais c'est celui de locuteurs de milieux et de lieux divers dans l'espace de la francophonie moderne. Faire une place centrale aux *variétés* du français et à la variation ne signifie néanmoins pas que tout se réduit à une fragmentation infinie. Nous croyons au contraire que derrière la diversité des usages, il y a de la *structure*. Si on ne se laisse pas abuser par la rhétorique, on peut soutenir que les modèles phonologiques, phonétiques et sociolinguistiques convergent vers la découverte d'invariants qui organisent les systèmes et les pratiques sociales (cf. Durand, Laks & Lyche, ce volume). En tout cas, telle est notre conviction même si on peut fort bien participer au projet PFC sans la partager.

Nous commencerons donc par le travail d'enquête qui est indispensable à la constitution de notre corpus. Pour tous les points d'enquête, le même protocole est utilisé par les enquêteurs qui sont choisis pour leur connaissance du terrain.³ Ce protocole comprend la lecture à haute voix d'une liste de mots, la lecture d'un texte, un entretien semi-directif que nous appelons "conversation guidée" et un entretien non directif que nous appelons "conversation libre". Les enquêteurs travaillent dans la mesure du possible en équipe de deux dont un chercheur connaît les témoins de façon intime. Le conseil de Bourdieu résume notre point de départ:

"On a ainsi pris le parti de laisser aux enquêteurs la liberté de choisir les enquêtés parmi les 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)

C'est aussi la méthode suivie par Labov (1972a, b) dans certains de ses travaux sur le Lower East Side de New York et par les Milroy à Belfast (Milroy 1980, 1987).⁴ Cette méthode a l'avantage d'atténuer le "paradoxe de l'observateur" et de fournir idéalement l'accès à deux registres conversationnels et, au minimum, de permettre des conversations aussi naturelles que possible. L'enquêteur familier du témoin peut participer à la conversation dite "libre" pour accéder au vernaculaire de ce témoin, alors que l'autre enquêteur dirigera la conversation dite "guidée". Ce travail d'équipe est aussi très bénéfique pour les transcriptions des enregistrements, les deux chercheurs s'aidant mutuellement et se complétant. Pour un certain nombre de points d'enquête, les conditions matérielles n'ont pas permis de constituer ces doublettes. L'enquêteur est alors seul. Mais dans tous les cas sa relation privilégiée avec les interviewés et sa participation au réseau dense constituent une condition préalable qui, à elle seule, a des effets d'atténuation du paradoxe de l'observateur.

³ Le protocole d'enquête PFC est décrit en détail dans Durand, Laks et Lyche (2002a). Voir Bergounioux (1992) pour un cadrage historique de l'enquête en linguistique française.

⁴ Voir Durand, Laks & Lyche, ce volume.

Une fois les équipes mises en place, le choix des témoins s'impose, mais quel est le portrait du témoin idéal ? Il appartient à un réseau dense dont, dans le cas idéal, l'un des enquêteurs est aussi membre. Milroy (1980) montre que, si tel est le cas, les sujets oublient très vite la présence du magnétophone alors qu'ils changent de registre dès qu'ils entrent en relation avec une personne qui n'appartient pas au réseau. D'un autre côté, la pression d'autres membres du groupe est telle que si, dans la conversation libre, un sujet se laisse intimider par la vue du micro, et adopte une langue plus standard et éloignée de son vernaculaire, ses intimes lui en feront immédiatement la remarque. Pour PFC, nous demandons de surcroît que le témoin soit bien ancré dans son environnement, que ce soit une personne qui ait vécu toute sa vie ou presque dans la communauté étudiée.

Sur le plan social, le sujet aura reçu un minimum de scolarisation et sera capable de lire à haute voix un texte et une liste de mots sans que cela ne devienne vite une tâche insurmontable. Cette exigence de scolarisation minimale exclut d'emblée divers groupes de locuteurs: jeunes en rupture forte avec le milieu scolaire, certains milieux très défavorisés, communautés francophones non-graphématisées comme c'est le cas en Louisiane par exemple. Si cela affaiblit évidemment la portée de la couverture sociale de notre projet, rien n'empêche un chercheur de compléter les enquêtes PFC par des études complémentaires ne faisant pas appel à la lecture à haute voix. Par ailleurs, il faut se méfier du préjugé selon lequel la lecture à haute voix ne serait possible que chez des sujets avec un bagage scolaire impressionnant. Bon nombre de nos enquêtés viennent de milieux modestes, parfois très modestes, et se sont parfaitement acquittés de cette tâche. La lecture à haute voix est tout à fait attestée dans l'intimité des familles (Qu'est-ce qu'elle raconte cette lettre qu'on a reçue des impôts? Qu'est-ce qu'ils disent dans le journal sur l'accident qui a eu lieu hier là en face?). Mais, en dehors de ces contextes, l'exercice évoque pour certains l'expérience scolaire avec son cortège d'échecs et de frustrations. Il faut donc que les enquêtés se sentent parfaitement à l'aise en présence de l'enquêteur et du magnétophone, d'où la nécessité de travailler avec des réseaux denses et de bien préparer l'enquête en rassurant les témoins sur le but de ces lectures.

En ce qui concerne la constitution du groupe, il est bien entendu que l'enquêteur s'efforcera d'obtenir à l'intérieur des

enquêtés une certaine différenciation sociale. Si la couverture sociale ne saurait être complète, il est cependant crucial que le groupe étudié comporte un nombre égal d'hommes et de femmes et idéalement les témoins seront répartis en plusieurs tranches d'âge : par exemple, 20-30 ans ; 30-50 ans ; 50-65 ans ; 65 +. Pour certains points d'enquête, il peut être difficile de satisfaire à ces exigences, mais un minimum de deux tranches d'âge est requis. Pour des raisons de stratégie générale, en particulier la lourdeur de la tâche de transcription et de codage, nous privilégions des groupes d'une dizaine de locuteurs par point d'enquête mais le protocole tel qu'il est conçu permet aussi bien d'étudier des locuteurs isolés que des groupes de centaines de locuteurs. Ainsi, en complément à Toulouse-ville, nous avons étudié avec la même méthode seize étudiants de l'Université de Toulouse-Le Mirail (huit garçons et huit filles) tous originaires de la région Midi-Pyrénées.

Il est important de signaler que PFC n'est pas une enquête *variationniste* au sens technique de grande enquête sociolinguistique. Les paramètres que nous venons de rappeler ne satisfont pas aux exigences d'un tel type d'enquête et encore moins à celles d'une enquête purement sociologique. Les échantillons retenus ne sauraient être véritablement représentatifs de l'ensemble des systèmes sociaux connus, mais dans la mesure où notre objectif est d'attester la diversité des usages phonologiques du français contemporain, nous pensons qu'une fois les enregistrements achevés, la diversité géographique couplée à une certaine diversité sociale nous permettra d'entreprendre des études sociolinguistiques plus fines et plus précises sur la variation phonologique. En tout état de cause, lorsque la base finale sera constituée, une véritable stratification sociolinguistique émergera de par la diversité des enquêteurs et de leurs réseaux personnels.

Les enregistrements⁵ sont effectués dans la mesure du possible chez les témoins et comprennent dans tous les cas les quatre volets mentionnés plus haut (lecture de mots, lecture de texte, conversation "guidée" et conversation "libre") sans qu'il y ait d'ordre prédéterminé pour les diverses tâches. Pour tous les locuteurs, l'enquêteur remplit une fiche d'information détaillée (voir Annexe 1) qui sous-tend notre classification sociolinguistique des témoins. Il demande à ces derniers de

⁵ Les questions liées à l'enregistrement sont abordées en détail dans Tarrier (ce volume).

signer un consentement de participation qui nous permet de disposer des enregistrements effectués à des fins de recherche ou à des fins pédagogiques (voir Annexe 2). Nous examinerons tour à tour ces quatre phases de l'enquête. avant de passer à nos techniques de transcription, de codage et d'analyse.

2.1 La liste de mots

La liste de mots (94 en tout) a été conçue de façon à établir un premier inventaire phonologique du témoin. Tous les mots font partie du vocabulaire courant et la quasi-totalité des exemples sont repris de la tradition phonétique et des grandes enquêtes déjà réalisées. Ceci garantit une comparabilité de PFC avec les enquêtes qui l'ont précédé et permet de donner aux analyses qui en découleront une profondeur historique certaine. Ceci permet également d'aborder la question du changement phonologique en temps réel. Nous avons favorisé les oppositions vocaliques en position accentuée. Un examen plus complet des oppositions en position inaccentuée et de phénomènes comme l'harmonie vocalique ou le relâchement des voyelles hautes aurait exigé une liste bien trop longue; mais certaines tendances peuvent déjà être dégagées. En ce qui concerne les consonnes, même si notre liste inclut tous les phonèmes consonantiques du français de référence, nous avons été assez restrictifs dans les phénomènes à examiner. En fin de liste, nous reprenons quelques paires minimales afin de tester la conscience linguistique du locuteur, mais comme le montre la liste ci-dessous, qui est celle que nous présentons à tous nos témoins, l'ordre des mots dans le corps de la liste (1-84) est aléatoire.⁶

⁶ La seule opposition classique en position accentuée qui n'est pas dans la liste de paires minimales est [ɔ]-[o]. Même si on peut le regretter, on notera néanmoins les mots suivants dans la liste en ordre aléatoire: (1) *roc*, (67) *rauque*, (21) *paume*, (54) *pomme* ainsi que la paire *cote-côte*. On se souviendra aussi que les paires minimales focalisent trop les témoins sur la norme. Nous avons jugé que 5 paires minimales en fin d'exercice suffisaient à établir non pas le système du locuteur mais son cantonnement quant à la norme.

LISTE DE MOTS PFC

| | | | | | |
|-----|--------------|-----|-------------|-----|----------------|
| 1. | roc | 33. | liège | 66. | peuple |
| 2. | rat | 34. | baignoire | 67. | rauque |
| 3. | jeune | 35. | pêcheur | 68. | cinquième |
| 4. | mal | 36. | socialisme | 69. | nier |
| 5. | ras | 37. | relier | 70. | extraordinaire |
| 6. | fou à lier | 38. | aspect | 71. | meurtre |
| 7. | des jeunets | 39. | niais | 72. | vous |
| 8. | intact | 40. | épais | | prendriez |
| 9. | nous | 41. | des genêts | 73. | botté |
| | prendrions | 42. | blond | 74. | patte |
| 10. | fêtard | 43. | creux | 75. | étriller |
| 11. | nièce | 44. | reliure | 76. | faites |
| 12. | pâte | 45. | piqué | 77. | feutre |
| 13. | piquet | 46. | malle | 78. | quatrième |
| 14. | épée | 47. | gnôle | 79. | muette |
| 15. | compagnie | 48. | bouleverser | 80. | piquais |
| 16. | fête | 49. | million | 81. | trouer |
| 17. | islamique | 50. | explosion | 82. | piquer |
| 18. | agneau | 51. | influence | 83. | creuse |
| 19. | pêcheur | 52. | mâle | 84. | beauté |
| 20. | médecin | 53. | ex-mari | 85. | patte |
| 21. | paume | 54. | pomme | 86. | pâte |
| 22. | infect | 55. | étrier | 87. | épais |
| 23. | dégeler | 56. | chemise | 88. | épée |
| 24. | bêtement | 57. | brin | 89. | jeune |
| 25. | épier | 58. | lierre | 90. | jeûne |
| 26. | millionnaire | 59. | blanc | 91. | beauté |
| 27. | brun | 60. | petit | 92. | botté |
| 28. | scier | 61. | jeûne | 93. | brun |
| 29. | fêter | 62. | rhinocéros | 94. | brin |
| 30. | mouette | 63. | miette | | |
| 31. | déjeuner | 64. | slip | | |
| 32. | ex-femme | 65. | compagne | | |

On demande au témoin de faire précéder la lecture de chaque mot du chiffre correspondant afin d'indexer plus facilement les productions, d'éloigner l'attention du mot testé, mais aussi pour rassembler des indications supplémentaires sur la prononciation des chiffres dans un style moins surveillé que la lecture des mots. Nous revenons sur cette question plus loin.

Les résultats phonologiques obtenus sur la base seule de la liste de mots ne peuvent évidemment pas être considérés comme totalement fiables mais cela ne signifie pas qu'ils soient sans intérêt. Même si l'informateur lit comme on lui a appris à lire, il n'en découle pas que ses réalisations soient aléatoires. En particulier, il existe une asymétrie dans la fiabilité des lectures à haute voix. La présence d'une opposition dans la lecture de la liste de mots (et en particulier dans les paires minimales), ne prouve pas, nous en convenons, que le locuteur fasse cette opposition dans la parole spontanée. En revanche, l'absence d'une opposition dans la lecture à haute voix est un indice très fort que l'enquêté ne pratique pas la distinction en question. Ainsi, la plupart de nos sujets méridionaux confrontés aux paires minimales 85 à 94 les lisent de manière identique (*patte* = *pâte*, *épais* = *épée*, *jeune* = *jeûne*, *beauté* = *botté*) à l'exception de la paire 93-94 *brun* vs. *brin*. Or, nous n'avons aucun exemple dans nos enregistrements où un des témoins en question, n'ayant pas fait les distinctions dans les paires minimales 85 à 92, les pratiquerait dans le reste de la lecture ou dans la parole spontanée. De même, si l'on examine les groupes consonantiques en position finale de mot (ex. /kt/ dans (8) *intact* ou (22) *infect*), on sait que, dans divers accents méridionaux (entre autres), ces groupes ne comportent pas (de façon variable ou catégorique selon les locuteurs) la coronale finale /t/ que l'on retrouve néanmoins dans la flexion ou la dérivation (cf *infect* [ɛ̃fɛk] vs *infecte* [ɛ̃fɛktə] ou *infecter* [ɛ̃fɛkte]). La prononciation [ɛ̃fɛk] avec une coronale dans la lecture de la liste de mots doit donc être confrontée à des données plus spontanées pour établir son statut. Mais, à nouveau, par le principe d'asymétrie, si un locuteur utilise [ɛ̃fɛk] dans la lecture à haute voix de la liste de mots, on est à peu près certain que ce trait de prononciation est profondément inscrit dans son système phonologique.

En combinant la liste de mot en ordre aléatoire, les paires minimales et la lecture du texte, on peut donc commencer à faire des hypothèses fortes sur la nature du système phonologique de chaque locuteur. Bien sûr, toute étude d'une variété donnée doit être préparée par des pré-enquêtes et confrontée aux travaux déjà effectués sur cette variété s'ils existent. On ne saurait trop insister sur la nécessité d'une connaissance intime du terrain pour toute

enquête si on ne veut pas se laisser abuser par certaines performances : le témoin peut très bien avoir mal identifié tel ou tel mot ou avoir des habitudes de lecture différentes selon la nature de la tâche. Ainsi, certains locuteurs languedociens lisent les verbes de la première conjugaison en *-er* en prononçant le <r> final: par ex. *chanter* [ʃãntɛʁ] (ou [ʃãntɛr] ou [ʃãntɛχ]). Mais cette prononciation, pour les locuteurs enregistrés à ce jour, ne s'applique qu'aux formes de citation et pas à ces mêmes mots dans des phrases lues à haute voix ou dans la parole spontanée : *Marie va chanter* [mãrivaʃãnte]. On ne pourra donc interpréter la prononciation [ʃãntɛʁ] que lorsqu'on aura examiné toutes les données et en écoutant attentivement les témoins en dehors des enregistrements quand on le pourra.

La phonologie des chiffres lu à haute voix est également très révélatrice. Mis à part les variations lexicales qui ont leur propre intérêt (par ex. *nonante* pour 90), l'ensemble des formes simples {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 20, 30, etc.) permet de balayer un grand nombre d'oppositions phonologiques du français. Ainsi, pour les voyelles nasales on pourra comparer les réalisations isolées ou en combinaison de *un, cinq, onze, vingt, trente*, etc. Divers processus phonologiques peuvent aussi se vérifier à partir des chiffres de la liste : par exemple, y a-t-il une règle de relâchement des voyelles hautes pour des mots comme *six, dix* ou *douze*, phénomène attesté au Canada mais aussi dans diverses autres régions francophones ? Si on s'intéresse au schwa en position finale, comment le locuteur prononce-t-il des mots comme *trente, quarante, cinquante* etc.? Dans le domaine des consonnes finales, quel sort est réservé aux consonnes finales graphiques de *cinq* ou de *vingt* prononcées isolément et dans des combinaisons comme *vingt-deux, vingt-trois, vingt-quatre*, etc. ?

Avant de clore cette section, il faut reconnaître que, si la liste vise à balayer les grandes oppositions vocaliques et consonantiques du français, elle ne saurait embrasser tous les phénomènes attestés dans la diversité des systèmes. Bien entendu, l'enquêteur peut compléter son étude par des listes ou des passages complémentaires si les témoins veulent bien se prêter à ces lectures supplémentaires. Pour le Canada,⁸ par exemple, Douglas Walker (Université de Calgary) a mis au point la liste complémentaire suivante qui permet de mieux cerner

⁸ Des listes complémentaires ont aussi été élaborées pour la Belgique, la Suisse, la Normandie, etc.

certains phénomènes qui jouent un rôle fondamental dans la prononciation canadienne comme le relâchement vocalique, les diphtongaisons ou le traitement des consonnes finales :

LISTE COMPLEMENTAIRE CANADA

| | | |
|---------------|----------------|--------------------------|
| 1. mettre | 41. équiper | 81. février |
| 2. maître | 42. député | 82. tabernacle |
| 3. écoeurer | 43. écouter | 83. coutume |
| 4. évêque | 44. professeur | 84. rendu |
| 5. calice | 45. piscine | 85. dire |
| 6. sable | 46. malle | 86. diète |
| 7. vite | 47. mâle | 87. duel |
| 8. libre | 48. Jacques | 88. tube |
| 9. juste | 49. cadenas | 89. tuile |
| 10. plume | 50. éclater | 90. pas d'idée |
| 11. couple | 51. éclat | 91. voûte immense |
| 12. pitoune | 52. il est là | 92. le prêtre |
| 13. ville | 53. voyage | 93. aveugle |
| 14. vire | 54. voyager | 94. convaincre |
| 15. juge | 55. mauvais | 95. vinaigre |
| 16. court | 56. jamais | 96. orchestre |
| 17. courte | 57. parfaite | 97. anglicisme |
| 18. boulevard | 58. toi | 98. debout |
| 19. filtrer | 59. bois | 99. pourrie |
| 20. abusif | 60. boivent | 100. ombre |
| 21. ministre | 61. noir | 101. épingle |
| 22. pilule | 62. boisson | 102. signe |
| 23. touriste | 63. voyons | 103. enseigner |
| 24. cuisine | 64. soirée | 104. champagne |
| 25. cuisiner | 65. doigt | 105. dehors |
| 26. pur | 66. avoir | 106. hâler |
| 27. rouge | 67. boire | 107. chercher |
| 28. neige | 68. froid | 108. arbre |
| 29. neutre | 69. crois | 109. plutôt |
| 30. chaude | 70. poigné | 110. il en a |
| 31. père | 71. bain | 111. c'est un niaiseux ! |
| 32. beurre | 72. quinze | 112. cent piastres |
| 33. port | 73. un | 113. sur la table |
| 34. part | 74. jungle | 114. dans la maison |
| 35. pâte | 75. crayon | 115. je les ai vus |
| 36. crainte | 76. honte | 116. sans les voir |
| 37. emprunte | 77. absent | |
| 38. honte | 78. il vente | |
| 39. lente | 79. reculer | |

| | | |
|-------------|--------------|--|
| 40. équiper | 80. brouette | |
|-------------|--------------|--|

2.2 Le texte

Le texte que nous soumettons à nos témoins a été créé pour les besoins de l'enquête et établi avec soin en termes de composition lexicale et grammaticale. Il a la forme d'un petit article dans un journal régional et n'a posé dans nos enquêtes aucune difficulté de compréhension. Nos informateurs ont le plus souvent 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 journalisme jugé de bon aloi. Ce texte a été construit afin d'éviter 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 vive ou non dans l'Hexagone. Le texte, que nous donnons ci-dessous, garde cependant un aspect culturel 'français', mais ce reste culturel n'a pas jusqu'à présent constitué un frein à la lecture de locuteurs belges ou canadiens par exemple.

Texte PFC (© Projet 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 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 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. "

Lors de la fabrication du texte, il fallait d'abord s'assurer que tous les segments vocaliques et consonantiques du français standard étaient représentés, tout en restant attentif à d'autres possibilités pour d'autres accents. Il fallait aussi inclure tout un ensemble de phénomènes phonologiques typiques du français. Sur le plan segmental, nous pensons avoir accompli cette tâche. Nous avons intégré à la construction du texte un ensemble de paires minimales potentielles comme *patte*, *pâte*, *jeune*, *jeûne*, *cote*, *côte*, *tourné* et *tournaît* qui permettent une comparaison systématique avec les mots de la liste. Nous avons aussi multiplié les exemples correspondant à chaque zone de la phonologie segmentale du français. Ainsi, le linguiste qui s'intéresse aux oppositions ou aux réalisations dans le domaine du 'A' trouvera déjà les exemples suivants dans le premier paragraphe : *à*, *village*, *émoi*, *a*, *étape*, *année*, *gloire*, *Garret*, *récemment*, *pâtes*, *italiennes*, *valu*, *hasard*, *car*, *lassé*, *habituels*, *campagne*. De même, pour les voyelles nasales et, en nous en tenant à nouveau au premier paragraphe, on repérera les mots : *grand*, *en*, *dans*, *région*, *fin*,

son, vin, blanc, un, champion, Berlin, récemment, son, donc, bêtement, profonde. Si on s'intéresse aux glissantes, on constatera que les trois segments [j w ʔ] sont bien représentés (même si tous les accents ne les opposent pas). Dans la mesure où ces trois glissantes sont en relation intime avec les voyelles hautes /i u y/ et alternent avec elles, il fallait inclure quelques contextes dans lesquelles on pouvait tester cette relation. Si on travaille sur les glissantes, on pourra donc examiner les exemples suivants :

[j, i]

Premier, région, champion, pied, quatrième, italiennes, qui a, inquiet, élections, manifestations, multiplier, officielles, Saint-Pierre, voyage, Jonquière, vérifications, explosion, opposition, coin, réaction, soutien, entier, vérifier, bataille, télévision, témoins, millions

[w, u]

émoi, gloire, soie, Louis, fois, voyage, Ouest, vouer, désespoir

[ʔ, y]

circuits, habituels, depuis

En ce qui concerne les phénomènes phonologiques, deux domaines ont surtout retenu notre attention dans la construction du texte: le schwa et la liaison (y compris dans son interaction avec la nasalisation et le *h* aspiré). Le schwa et la liaison sont deux questions incontournables de la phonologie du français. Elles engagent le domaine syntagmatique de façon forte et font intervenir plusieurs composantes (morphologie, syntaxe, sémantique, pragmatique). Nous passons ici de la phonologie du mot à celle des énoncés dans toute leur richesse rythmique et intonative. De plus, les réalisations du schwa et de la liaison sont fortement liées au positionnement sociolinguistique des locuteurs et permettent donc des comparaisons intéressantes du point de vue intra-locuteur, intra-dialectal et inter-dialectal. Ces deux questions reçoivent un traitement systématique dans le cadre du projet PFC et nous y revenons en §4.

La lecture du texte ne fournit évidemment pas des données totalement fiables 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. On signalera d'ailleurs une transition stylistique vers un niveau plus relâché chez certains locuteurs dans la lecture de l'incise que nous avons intentionnellement incluse dans le texte: "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.'" Ainsi avons-nous noté chez certains sujets languedociens âgés la prononciation du mot *coin* comme [kwãŋ], prononciation récessive qui appartient à un registre plus familier que [kwɛŋ]. D'autres sujets du nord de la France effacent beaucoup plus de schwas ici que dans le reste du passage où ils adoptent un style moins allegro.

Sur le plan suprasegmental, le texte permet une exploration partielle du rythme particulier à la lecture et à son phrasé mais pas de l'intonation dans sa richesse énonciative. Dans les limites d'un texte qui devait être relativement court et facile à lire, nous n'avons pas établi une comparaison systématique entre divers types d'énoncés (assertions, questions, exclamations, ordres). Seules deux questions fermées (type QU-) ont été intégrées au texte: (a) Qu'est-ce qui a donc valu à Beaulieu ce grand honneur? (b) Comment, en plus, éviter les manifestations qui ont eu tendance à se multiplier lors des visites officielles ? Les participants au projet qui désirent plus particulièrement explorer l'intonation et comparer sous cet angle la lecture et le discours spontané doivent donc mettre au point un protocole supplémentaire suppléant ce manque.

2.3 Les conversations

Dans l'entretien guidé d'une durée de vingt à trente minutes, le locuteur est interrogé, par exemple, sur ses activités, son travail, ses projets, son enfance, ou sur des questions d'actualité. Nous ne donnons pas aux enquêteurs de directive précise si ce n'est d'éviter les questions qui amènent des réponses par oui ou par non, ou des réponses monosyllabiques. Le but de l'entretien est double : obtenir des renseignements sur le locuteur qui seront complétés par la fiche d'enquête (voir Annexe 1) mais aussi laisser le locuteur s'exprimer plus ou moins librement sur des sujets qui l'engagent. Si l'enquête est conduite par deux enquêteurs, c'est celui qui est le moins familier avec le témoin qui mènera

l'entretien de façon à garantir le côté semi-formel du dialogue. La dernière tâche du témoin est de participer à une discussion libre d'environ trente minutes. Les personnes en présence sont le témoin, l'enquêteur-ami et/ou un autre témoin, ami du premier et l'entretien est non directif. Autant que possible, nous limitons la conversation libre à trois locuteurs au maximum. Dans certaines enquêtes, nous avons dû enregistrer des conversations avec des groupes de plus de trois locuteurs. Si les résultats sont intéressants d'un point de vue sociolinguistique (tours de parole, par exemple), ils sont difficilement exploitables pour l'étude phonologique. Tout dialogue naturel de plus de trois personnes présente de nombreux chevauchements et conversations croisées qui créent périodiquement une véritable bouillie acoustique.

Le but de l'entretien libre est d'essayer d'approcher le vernaculaire de nos témoins et de diminuer autant que faire se peut l'effet du "paradoxe de l'observateur". Nous sommes conscients de l'aspect "artificiel" de la méthode : les sujets sont devant un microphone et cette présence même brouille les cartes. Pour nous assurer l'accès à un registre véritablement informel, il ne faudrait pas nous contenter d'une seule séance et prolonger l'enregistrement plusieurs heures. Il va de soi que cette procédure est exclue dans une enquête d'envergure comme la nôtre. Nous restons cependant convaincus que le respect des instructions décrites ci-dessus offre dans le meilleur des cas quatre registres bien définis. Nous tenons également à signaler que, pour des raisons déontologiques, nous sommes opposés à des enregistrements de locuteurs à leur insu. Outre qu'il est souvent difficile d'obtenir une bonne qualité d'enregistrement dans de tels contextes, il est capital de s'assurer que les enquêtés sont parfaitement au courant des buts du projet et de ses utilisations éventuelles (voir le formulaire *Consentement de participation* en Annexe 2).

2.4 Oral et écrit

Dans les paragraphes qui précèdent, nous avons souligné divers avantages de la lecture à haute voix pour l'étude de l'oral, tout en reconnaissant certaines limites au protocole PFC. En termes plus généraux, il n'est pas inutile de rappeler qu'on ne saurait traiter les sujets scolarisés dans les sociétés à forte tradition écrite comme des individus qui auraient passé toute leur existence dans une communauté non graphématisée. Les avis divergent sur le statut précis qu'il faut assigner à la

graphie au sein de la faculté de langage (cf. par exemple Durand 2000). Il semble néanmoins que les sujets alphabétisés et non alphabétisés n'ont pas le même accès psycholinguistique aux unités phonologiques fondamentales ni forcément les mêmes représentations mentales de ces dernières (cf. Morais et alii 1987, 1991, 1997). Même si l'on peut s'interroger sur l'influence de la graphie sur le système phonologique profond des individus qui est, on s'en souviendra, mis en place avant l'apprentissage scolaire, on ne saurait nier que divers domaines de la prononciation du français sont inextricablement liés à l'écrit. La liaison est un exemple frappant de ce genre de phénomène (voir Encrevé 1988). Même si l'on sait que la parole spontanée est moins riche en liaisons que la lecture à haute voix, l'effet de retour de cette dernière sur les usages (en particulier sur les liaisons dites facultatives) est probable et mérite d'être attentivement examiné. De même a-t-on pu signaler que certaines consonnes qui n'étaient pas prononcées ont été (ré-)introduites dans l'usage à partir de la graphie. Cette influence de la graphie sur la prononciation, dit 'effet Buben' en hommage au travail de Vl. Buben (1935) sur cette question, est bien attesté en ce qui concerne le 'r' des infinitifs du deuxième et troisième groupes (*sortir, vouloir*) qui n'étaient habituellement pas prononcés au XVIIIe siècle ou encore aux terminaisons en *-il* (*grésil, mil, gril*) dont la consonne est très souvent prononcées actuellement.⁹ Catach (1993: 67) déclare: "Notre langue, qui ne prononçait il y a deux siècles pratiquement aucun groupe de consonnes, si ce n'est avec les liquides (*r* et *l*), se surcharge actuellement, sous les actions combinées des emprunts, de l'école, de l'orthographe, etc., de nombreux groupes internes et finaux (*cheptel, dompter, sangsue*, etc.), ce qui selon nous est sans doute la modification la plus importante que l'on puisse noter en français depuis la disparition du *l* mouillé."

Mais si l'on ne saurait nier l'influence que peut exercer l'écrit sur l'oral (et vice versa), il faut bien reconnaître que bon nombre de questions n'ont pas été examinées en détail. Par exemple, s'il est vrai que des 'e' qui sont muets à l'oral, en dehors des accents du sud de la France, apparaissent dans la lecture de vers (ou certains types de chansons), est-ce vrai de la lecture à haute voix dans un style moins soutenu comme celui du texte PFC? Un des avantages de notre protocole est de

⁹ Sur l'effet Buben, voir Chevrot & Malderez (1999), Laks (2002).

permettre une mesure directe et précise de sous-aspects du rapport entre l'écrit et l'oral et donc de fournir, pour chaque locuteur, un index de cette corrélation.

3. Les conventions de transcription orthographique

L'une des conditions préalables à l'exploitation d'un corpus est la nécessité de conventions, documents et outils communs permettant de vérifier et de comparer données, analyses et résultats. La définition de ce qu'est précisément un corpus est objet de controverse. De façon minimale, nous accepterons la caractérisation suivante:

"A corpus is any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by people in other organisations" (Gibbon 1998, cité in Delais-Roussarie, ce volume).

Nous avons vu que le protocole d'enquête de PFC visait à constituer une base d'enregistrements sonores dans un format rigide donc strictement comparable. Une fois les enregistrements effectués, il sont numérisés et transférés sur des supports informatiques.¹⁰ Dans le projet PFC, nous avons opté pour des fichiers son au format standard .WAV. Nous avons également mis au point une structuration commune de tous les fichiers de données et d'analyse, ce qui est indispensable au travail scientifique dans un projet décentralisé.¹¹

La numérisation effectuée, se pose alors la question de la transcription de base des données sonores. Dans le projet PFC, nous adoptons comme "couche zéro" une transcription orthographique qui sert de point de départ à des annotations, des codages et des investigations phonétiques. Cette transcription se fait à partir d'un outil informatique, PRAAT, qui permet d'aligner les représentations graphiques sur le signal.

¹⁰ Nous utilisons ici le terme 'numériser' dans un sens non technique puisque les enregistrements à partir de DAT ou d'enregistreurs de type minidisk sont déjà, par définition, numériques. Quel que soit le support sur lequel les enregistrements ont été faits, il faut les rendre interprétables par un ordinateur. C'est le sens dans lequel nous prenons donc le terme 'numériser'.

¹¹ Voir Durand, Laks & Lyche (2002c) sur le format des rendus PFC.

Nous justifions et présentons ces conventions en 3.1 et 3.2 avant de nous pencher sur PRAAT en 3.4

3.1 Conventions orthographiques PFC

Lors du passage d'un corpus oral à un texte écrit, le transcripteur est confronté d'emblée à la question suivante: comment refléter le caractère oral du corpus?

Une première réaction, qui a souvent tenté phonologues et phonéticiens, consiste à proposer une transcription phonétique large ou étroite à partir des symboles que fournit un système comme l'Alphabet Phonétique International. Cette stratégie, qui est tout à fait motivée pour un projet à petite échelle, pose des problèmes épineux dans un projet décentralisé où l'on tient à analyser une masse de données importante. En effet, il est rare que les transcripteurs s'accordent sur la valeur des symboles et des diacritiques dès qu'on veut noter des phénomènes relativement fins. Mais au-delà du manque d'accord inter-subjectif, il est important de signaler qu'une transcription dite phonétique, comme premier niveau d'abstraction à partir du signal, pose de réelles difficultés. Si on adopte une transcription large, ou plus précisément phonémique, on met la charrue avant les bœufs : on suppose qu'on a déjà découvert le système qu'on cherche à établir à travers l'enquête. Si on privilégie au contraire des transcriptions étroites (de type allophonique), on se heurte à une difficulté de taille. Quel degré de finesse phonétique doit-on adopter ? En effet, au niveau phonétique, de nombreuses réalisations ne correspondent pas à des choix binaires mais à des valeurs sur des échelles continues. Ainsi la longueur des voyelles au sein des énoncés n'est pas simplement une opposition entre long (:) et bref, ni même entre long, demi long et bref, mais entre divers degrés de longueur que seule une étude acoustique précise peut résoudre. De même, l'assimilation de voisement n'est pas toujours un simple changement de valeur (voisé vs. non voisé) mais un phénomène scalaire et qui, selon divers spécialistes, entre en interaction avec d'autres paramètres phonétiques comme la dimension tendu/lâche. Ce travail ne peut être fait de manière fiable qu'en utilisant des outils de mesure acoustique. Mais s'engager dans une étude acoustique demande beaucoup de temps et d'énergie même pour un corpus de petite taille et exige de toute façon un déblayage du terrain. Pour ces diverses raisons et pour d'autres (cf. §3.2), nous avons privilégié un premier niveau de

transcription orthographique qui constitue la base du travail que nous effectuons sur le corpus. Cette couche zéro, du fait même qu'elle ne prétend pas refléter fidèlement la structure phonique, permet une exploration beaucoup plus rapide des enregistrements et en retour un travail plus précis sur les données sonores chaque fois que cela s'avère nécessaire.

Les conventions employées dans PFC doivent beaucoup au travail effectué autour du corpus du GARS à Aix (Université de Provence) et de VALIBEL (UCL, Louvain-la-Neuve).¹² Nous avons aussi tenu compte des recommandations du groupe européen EAGLES et des instructions de Gjert Kristoffersen (Université de Bergen) pour la transcription d'une banque de données sur les dialectes norvégiens. Nous avons opté dans le projet pour une transcription orthographique standard, y compris une ponctuation standard (mais plus réduite que la norme orthographique française).¹³ Nous commencerons par un exemple concret en 3.1.1. avant d'expliquer les principales conventions de 3.1.2 à 3.1.9. En 3.2, nous revenons sur quelques-uns des avantages du type de transcription orthographique proposée.

3.1.1 Exemple de transcription (conversation guidée)

Contexte: L'enquêtrice (E:) qui est originaire de Belgique interviewe une jeune locutrice (AB) de Pézenas (Hérault, 34120 France) qui lui parle de l'enseignement à l'université en France et des qualifications requises.

AB: Par rapport à des amis que j'ai, qui, <E: C'est vrai?> qui ont beaucoup de mal maintenant à enseigner sans l'agrégation à l'université, <E: A l'université?> ouais, <E: Ah bon?> alors qu'il y a cinq six ans en, en arrière,

¹² Pour les conventions du GARS, cf. Blanche-Benveniste & Jeanjean (1987), Blanche-Benveniste et alii (1991), Bilger (2000). Voir également le site suivant: <http://www.up.univ-mrs.fr/delic/corpus/index.html>. Pour VALIBEL, Centre de recherche sur les Variétés linguistiques du français en Belgique (dir. M. Francard), voir le site suivant <http://valibel.fltr.ucl.ac.be/>. Pour une discussion plus générique des questions de transcription, voir Delais-Roussarie (ce vol. a)

¹³ Les conventions de transcription PFC et leur intégration à PRAAT sont expliquées dans Delais-Roussarie, Durand, Lyche, Meqqori et Tarrier (2002).

c'était plus facile, apparemment, <E: Ah bon?> ouais. Peut-être, (rire) peut-être en histoire, c'est peut-être différent pour d'autres domaines. <E: Ça dépend peut-être des domaines ?> Hein, voilà, c'est ça, hein, peut-être que des profs d'histoire, il y en a tellement que, bon, euh, le nombre de jeunes gens qui ont des doctorats en histoire, que, bon, maintenant on peut pas être en fac sans l'agrégation.

E: Peut-être est-ce, peut-être est-ce, voilà une autre contrainte qu'ils imposent. Pour ça, je ne saurais pas le dire. Je ne saurais pas le dire, je sais pas du tout.

AB: Vous avez le même cursus, le même système universitaire en Belgique qu'en France, oui, c'est pareil?

E: Euh, ça s'appelle pas <E: C'est la même chose?> mais enfin c'est la même chose.

Les conventions adoptées dans le projet et illustrées en partie ci-dessus sont relativement simples :

3.1.2 Formes de mots

Au niveau lexical, les diverses réalisations d'un mot sont ramenées à la forme standard de ce mot en contexte. Dans la séquence parlée 'Je suis plus petite', que le locuteur ait dit [ptit] ou [ptitə] ou [pɛtit] ou [pʁit], etc., nous transcrivons 'petite'. Que le locuteur ait dit [ply] ou [py], nous transcrivons 'plus'. Qu'il ait dit [ʒə sɥi] ou [ʒsɥi] ou [ʃsɥi] ou [ʃɥi] etc., nous transcrivons 'je suis'. En revanche, nous ne réintroduisons pas des éléments lexicaux absents. Ainsi [ivwapasakɔmsa] sera transcrit 'Il voit pas ça comme ça' et pas 'Il ne voit pas ça comme ça' avec réinsertion normative de 'ne'.

3.1.3 Tours de parole et ponctuation

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. 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. Nous n'utilisons ni le point d'exclamation ni le point virgule.

3.1.4 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 entraîne un changement de locuteur, nous l'indiquons par un changement d'intervalle. Dans les deux cas, le locuteur qui cause le chevauchement est 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.

- *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.

3.1.5 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.

3.1.6 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 SAMPA¹⁴.

Exemple 1 :

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

NB : ici le sigle est prononcé [æsensɛɛf] (soit [laEsEnseEf] en 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] ([liyte] en SAMPA) et n'est pas noté.

Exemple 2 :

L1 : Le CNET (snET) embauche du personnel.

Dans ce dernier exemple, la personne a prononcé ce sigle d'une façon inhabituelle (à savoir, [snɛt]), d'où l'utilisation de la transcription SAMPA (snET) entre parenthèses.

3.1.7 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 les dictionnaires usuels : *pfft*, *psit*, *tss-tss*, *tsoin-tsoin*. Dans tous les autres cas, on note les bruits entre parenthèses, exemple (onomatopée *chchchch*).

Exemple :

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

3.1.8 Le discours rapporté

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

¹⁴ Rappelons que SAMPA est un système permettant de coder les symboles API standard à partir de symboles ASCII (voir Caelen-Haumont, 2002 et l'Annexe 3 ci-dessous).

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.

3.1.9 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 sont 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.

3.1.9 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.

3.2 Quelques avantages d'une transcription orthographique

Dans l'introduction de cette section, nous avons souligné les inconvénients d'une transcription phonémique ou phonétique si elle se donne comme principal point l'accès aux enregistrements. Au-delà des problèmes théoriques déjà soulevés, on remarquera qu'un des grands avantages d'une transcription orthographique est de permettre d'indexer rapidement toutes les occurrences d'une unité lexicale ou de séquences de mots données pour en examiner la prononciation. Si, pour rester concret, on s'intéresse de près aux prononciations de 'il y a', il est plus efficace de repérer tous les 'il y a' dans les transcriptions orthographiques et d'en examiner attentivement la prononciation (à l'aide d'un outil comme PRAAT, voir §4) que d'essayer de retrouver cette forme à partir

de séquences phonétiques du type [ilia, ilja, ija, ja]. Cette tâche inverse est infiniment plus complexe puisqu'elle présuppose qu'on ait établi un inventaire fini des séquences phonétiques pertinentes et qu'on puisse les séparer d'autres fragments de transcription auxquelles elles peuvent être accolées. C'est précisément pour cette raison que nous ne modifions pas la transcription orthographique de départ avec des conventions pseudo-phonétiques comme les suivantes: *J' crois qu'y a Paul qu'est v'nu pa(r)c' qu'i peut la supporter*. Dans PFC, nous respectons scrupuleusement l'orthographe standard. Nous écrivons, par exemple, 'je' que le locuteur ait dit [ʒə, ʒø, ʒœ, ʒ, ʒ, ʃ, etc], sauf évidemment devant voyelle (*j'ai acheté*). Contrairement aux apparences, ramener toutes les variantes attestées à 'j', comme on le fait souvent, revient à court-circuiter une analyse plus précise des données. Sans parler des ambiguïtés qu'on multiplie et du manque de cohérence dans l'adaptation des conventions de l'écrit aux méandres de l'oral, il devrait être clair que la diversité des réalisations phonétiques n'est pas directement répertoriables par des remaniements ad hoc de l'orthographe usuelle. De plus, toute transcription pseudo-phonétique fournit une image déformée de l'oral, le stigmatise et le confine dans un rôle de parent pauvre, en fait, fausse au départ l'analyse.

Il faut également rappeler que l'orthographe française dans la mesure où elle est en partie morphophonologique est une voie d'accès utile à de nombreuses alternances de surface du français. Il ne s'agit pas de s'engager ici dans le débat qui a pu opposer les linguistes qui ont considéré l'orthographe du français comme relativement motivée dans son rapport à la phonologie (Schane, 1968, Dell 1973) et ceux qui au contraire l'ont vu comme fondamentalement arbitraire (M. Durand 1936, Blanche-Benveniste & Chervel 1969). Tout ce que nous tenons à affirmer est qu'une annotation de l'orthographe standard est utile pour repérer et explorer divers phénomènes phonologiques en français. C'est en tout cas notre thèse en ce qui concerne le schwa et la liaison.

Enfin, si une transcription orthographique crée ici et là des problèmes spécifiques soulignés par Blanche-Benveniste et Jeanjean (1987),¹⁵ il ne faut pas oublier qu'il existe des phonétiseurs qui permettent de convertir des transcriptions

¹⁵ Voir Delais-Roussarie, ce volume: §3.2..

orthographiques en séquences de phonèmes standard. Ces dernières peuvent servir à leur tour de point de départ à une analyse plus conforme aux propriétés du système oral étudié pour les sous-parties du corpus que l'on tient à étudier de plus près.

4. PRAAT comme outil de transcription et d'analyse

L'une des difficultés majeures dans l'exploitation de nombreux corpus oraux est le fait que les transcriptions ne sont pas alignées sur le signal. On a deux objets que l'on peut exploiter séparément mais dont l'appariement exige des manipulations longues et complexes. De grands progrès ont été accomplis dans le domaine des outils de transcription et d'analyse (voir Delais-Roussarie, ce vol. (b)). Ces derniers se sont diversifiés et sont devenus beaucoup plus conviviaux que par le passé. Dans le projet PFC, nous avons opté pour PRAAT, un programme informatique gratuit qui a été développé par Paul Boersma et David Weenink de l'Institut de Phonétique d'Amsterdam.

Le logiciel PRAAT, que nous utilisons pour nos transcriptions et nos codages, offre de multiples avantages puisque, comme le soulignent Delais-Roussarie, Meqgori & Tarrier (ce vol.), il permet :

- d'enregistrer des fichiers audio qui pourront ensuite être analysés ;
- de transcrire, d'étiqueter et de segmenter des données audio (enregistrements effectués sous PRAAT ou autres fichiers au format WAV) ;
- 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.) .

Nous nous contenterons de parler ici de l'utilisation de PRAAT comme outil de transcription sans envisager toutes les facettes de cet outil très complet. Nous supposerons, comme nous l'avons dit plus haut, que les enregistrements ont déjà été numérisés et transférés sur un ordinateur sous forme .WAV.¹⁶

PRAAT est un outil particulièrement utile pour les transcriptions et annotations linguistiques car il permet de créer plusieurs 'tires' (angl. tiers) dans lesquelles on peut coder en parallèle des informations différentes. Dans le projet PFC, en ce qui concerne le texte et les deux types de conversation, nous utilisons la première tire pour la transcription orthographique et deux autres tires pour le codage du schwa et de la liaison. PRAAT n'impose pas un style particulier de transcription et les normes adoptées pour le projet PFC ont déjà été décrites. Même sans exploiter les ressources spécialisées de PRAAT, on s'aperçoit vite de l'avantage à tirer d'un tel outil. Il fonctionne, à tout le moins, comme un logiciel de traitement de texte combiné à un magnétophone de haute volée. En effet, il permet d'isoler par de simples clics des segments d'une longueur variable et de les écouter et réécouter autant de fois qu'on le désire. Un exemple de portion de signal avec sa transcription est fourni ci-dessous:

¹⁶ Nous utilisons ici le terme 'numériser' dans un sens non technique puisque les enregistrements à partir de DAT ou d'enregistreurs de type minidisk sont numériques. Quel que soit le support sur lequel les enregistrements ont été faits, il va néanmoins falloir transférer les enregistrements vers l'ordinateur sur lequel PRAAT est installé dans un format interprétable.

Exemple de Textgrid

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."

Tous les outils PFC sont conçus de manière à traiter directement ces textgrids dans leur format informatique. Néanmoins, pour l'utilisateur humain qui veut inspecter les transcriptions en dehors de PRAAT, ce format ne permet pas de retrouver facilement les dialogues et de les lire de façon confortable. Un petit utilitaire informatique 'Transpraat' a donc été développé au sein du projet par A. Meqqori afin de pallier ces inconvénients. Transpraat permet de transformer le format PRAAT balisé en fichier text exploitable sur les traitements de texte courants. Le dialogue enfoui dans le textgrid cité ci-dessous ressort sous la forme suivante:

Résultat de la conversion par Transpraat.

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.

4. Domaines d'analyse, codages et outils PFC

Nous avons souligné à plusieurs reprises l'importance d'une approche explicite dans le travail de constitution et d'exploitation d'un corpus. Comme l'explique si bien Habert (1997: 156):

"Sans une documentation jointe, un corpus est mort-né. L'un des dangers de la facilité actuelle de rassembler des textes électroniques est précisément que les objectifs de regroupement ainsi que ceux des annotations effectuées ne soient pas enregistrés : le corpus cesse d'être utilisable dès qu'on perd la mémoire de ces choix." (Cité in Delais-Roussarie, ce volume.)

Dans le projet PFC, nous avons choisi trois domaines obligatoires d'analyse tout en encourageant les chercheurs à explorer les domaines qui les concernent plus particulièrement. Le premier est l'inventaire phonologique de chaque locuteur, le deuxième le schwa et le troisième la liaison. Pour chacun de ces domaines, et en particulier pour le schwa et la liaison, nous adoptons des conventions de codage précises et des outils qui fournissent un socle empirique commun aux chercheurs impliqués dans le projet. Dans la mesure où ces choix structurent largement l'exploitation des enquêtes, nous les examinerons de manière détaillée.

4.1 Inventaires phonologiques et comparateur

La plupart des analyses structuralistes des systèmes phonologiques au niveau segmental établissent pour chaque variété les phonèmes, les allophones, les contraintes phonotactiques et l'incidence lexicale des phonèmes (au sens de Wells 1982). Par incidence lexicale (ou distribution lexicale), nous entendons ici le fait que deux locuteurs peuvent avoir le même ensemble de phonèmes mais une répartition différente et

aléatoire selon les mots. Ainsi, comme c'est souvent le cas, deux locuteurs peuvent distinguer /a/ et /ɑ/ mais ne pas s'accorder sur les mots dans lesquels on trouve l'un ou l'autre.¹⁷ Quelle que soit l'orientation théorique que l'on adopte, il est difficile de se dispenser d'un tel travail taxinomique préliminaire. Dans le projet PFC, nous ne prescrivons pas de format absolu pour l'étude des inventaires phonologiques mais nous supposons que pour chaque locuteur un portrait phonologique classique a été dressé. Nous fournissons à tous les enquêteurs une fiche phonologique d'orientation du type suivant:

Fiche d'orientation phonologique

A. Inventaire segmental (voyelles, semi-voyelles, consonnes)

(a) Voyelles : le locuteur fait-il les oppositions suivantes ?

i y u : OUI/NON

e / ε

a / ɑ

o / ɔ

ø / œ

ã / õ

ẽ / ĕ

Autres oppositions notées :

Longueur vocalique ? spécifier :

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

(b) Glissantes

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

SINON : quel est l'inventaire ?

(c) Consonnes

Est-ce que l'inventaire est du type suivant :

p, t, k, b, d, g, f, s, ʃ, z, v, ʒ, m, n, ɲ, l, r

SINON :

Consonnes supplémentaires :

Consonnes manquantes :

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

B. Inventaire allophonique segmental (principaux allophones)

Exemple /r/ = [ʁ] à la finale de mot sinon [r] par défaut

¹⁷ Pour une documentation précise de ce problème, cf. Martinet & Walter (1973), Walter (1976).

Observations :

Glissantes

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

OUI/NON - COMMENTAIRES

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 ?

En ce sens, notre travail initial se rapproche de celui d'Henriette Walter (1982) dans sa conception de l'enquête sur les variétés du français. Walter, dont nous partageons la méthodologie jusqu'à un certain point, plaide par exemple pour des enquêtes linguistiques et sociolinguistiques qui fassent une place fondamentale aux idiolectes et note que:

"William Labov, dans *Sociolinguistique*, indique bien en effet qu'il a commencé son étude par une enquête préparatoire de 70 interviews (p. 95), mais à aucun moment il ne donne de précisions sur les différents systèmes de ses informateurs. Le lecteur n'a connaissance de cette enquête préliminaire que dans la mesure où elle a permis à l'auteur d'isoler les quelques variables phonologiques pouvant être mises en relations avec la différenciation sociale." (1982 : 18).

Comme dans les enquêtes de Walter dont nous poursuivons la tâche, nous constituons donc un inventaire de base des unités et des structures présentes dans le système de chaque locuteur. Nous nous séparons néanmoins de la méthode qu'elle suit sur plusieurs plans. Au niveau le plus général, nous ne pensons pas que les inventaires de type taxinomique suffisent à caractériser les systèmes phonologiques dans leur complexité. Tout d'abord, l'imbrication des réseaux de traits distinctifs et de phonèmes dans le système morphologique et syntactico-sémantique des langues nous semble incontournable. A la différence des approches en linguistique fonctionnelle, un phénomène comme la liaison fait pour nous fondamentalement partie du système phonologique du français (et pas simplement de sa morphologie ou de sa

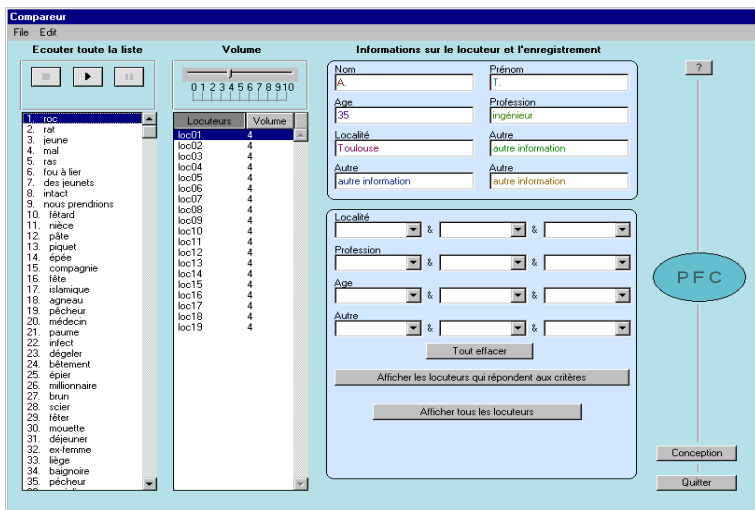
morphosyntaxe). Deuxièmement, à notre sens, les représentations phonologiques ne se limitent pas à de simples assemblages de phonèmes et de traits distinctifs, mais il faut intégrer à l'analyse un ensemble de concepts qui, pour la plupart, sont bien ancrés dans la tradition phonologique (more, syllabe, pied, mot phonologique, etc.) et ont été explorés en détail depuis plus d'un quart de siècle. Troisièmement, il nous semble qu'on ne peut écarter la formulation des 'généralisations' phonologiques par simple a priori méthodologique, quel que soit par ailleurs le formalisme de l'analyse: règles (strictes ou variables), contraintes (violables ou non), réseaux de neurones ou autre. Cela signifie que divers phénomènes (schwa, liaison, semi-vocalisation, etc.) exigent un outillage plus complexe que celui posé dans les études structuralistes classiques. Quatrièmement, sur le plan phonétique, si les études PFC sont dans un premier temps auditives, nous cherchons en collaboration étroite avec divers chercheurs, et en particulier le Laboratoire LPL d'Aix-en-Provence, à développer des outils permettant de construire automatiquement, pour chaque locuteur, des chartes formantiques caractérisant les systèmes des locuteurs d'un point de vue acoustique. Enfin, nous accordons une place centrale à la variation intra-locuteur, inter-locuteur et inter-dialectale. Or, l'outillage théorique que fournit la (socio)-linguistique moderne permet précisément de mieux penser ce domaine que les inventaires structuralistes classiques (voir Durand, Laks et Lyche, ce volume).

Ces remarques faites, nous ne prescrivons pas dans le projet PFC une méthode phonologique uniforme. Les chercheurs, pourvu qu'ils établissent un inventaire classique des systèmes en présence (dans le cadre des rapports à fournir en fin d'enquête) restent parfaitement libres de leur théorisation. Si Andreassen (2003), Durand et Lyche (1999, 2003), Lyche (2003) et Montreuil (ce vol.) offrent des systématisations de résultats dans le cadre de la théorie de l'optimalité, d'autres chercheurs suivent fort heureusement des orientations différentes (par exemple, Francard, Hambye et Simon dans le cadre des travaux de VALIBEL).¹⁸

¹⁸ Sans parler du fait que certains phénomènes traditionnellement traités comme phonologiques (par exemple, l'harmonie vocalique) reçoivent peut-être un meilleur traitement dans un cadre phonétique: voir Boula de Mareuil et Fagyal (2001) .

4.2 Le comparateur

Le projet PFC a développé un outil, mis au point par A. Meqqori, qui permet une comparaison simple des performances de lecture des mots de la liste pour un nombre quelconque de locuteurs. Le logiciel intègre un outil de recherche permettant une comparaison en fonction de divers paramètres tels que l'âge, la localité ou la profession des locuteurs. Dans sa version actuelle, l'outil est limité à la lecture de la liste (autrement dit celle des 94 mots précédés du numéro d'ordre). Une nouvelle version permettra notamment de comparer les mots de la liste et ceux du texte lu. Par la suite, le comparateur sera intégré à l'interface globale de visualisation de la base de données *PFC*. L'outil comprend un séquenceur qui permet d'aligner très précisément le mot sur sa réalisation et moteur de comparaison. Il permet au phonologue de pouvoir tester rapidement des hypothèses phonémiques et de balayer les principales réalisations allophoniques détectables à l'audition (comme par exemple les diverses formes de 'r' : [r R ʁ ʁ etc.]).



4. 3. Le schwa et son codage

L'enquête PFC livrera des résultats d'une grande généralité sur la présence ou l'absence de schwa. Les données que nous recueillons sont de deux ordres : les données sur schwa dans le texte lu et les données dans la parole spontanée (conversations guidée et libre). Nous procédons à des codages à partir de la graphie car cette dernière fournit des

renseignements précieux sur les sites possibles déclenchant la présence ou l'absence d'un schwa. De tels codages ne constituent qu'une étape préliminaire de toute analyse mais cette étape nous semble indispensable. 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, 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 ont rarement été abordées avec des principes de codage aussi explicites que les nôtres et nous pouvons donc espérer obtenir des données d'envergure comparables.

Le codage est effectué sur la base de la transcription orthographique dans PRAAT sur une tire spécifique. Une fois le codage terminé, notre utilitaire *classeur-schwa* permet d'effectuer un comptage et un classement des présences/absences de schwa (voir §3.4). Avant de passer aux conventions de codage, nous considérerons un bref exemple concret. Soit l'énoncé : "Je vois vraiment pas comment ce petit peut deviner si vite." Supposons qu'il est prononcé //§ J' vois vraiment pas comment ce p'tit peut deviner si vit' § // (dans une transcription informelle). Il sera codé : "Je0132 vois vraiment pas comment ce1112 pe0212tit peut de1212viner si vite0413".

L'avantage d'un tel codage est qu'il indique non seulement la présence ou l'absence d'un schwa mais aussi quels sont les contextes qui ont déterminé l'apparition ou la disparition de ce segment. Ainsi, le codage Je0132 indique que le schwa n'est pas prononcé (0), que le mot est monosyllabique (1), qu'il est situé en début de groupe intonatif (3) et que le mot suivant commence par une consonne (2).

4. 3.1 Le système de codage

Le système de codage de schwa a pris comme assise les traitements traditionnels, en particulier Dell (*Les règles et les sons*, Paris, Hermann, 1973/85). Pour chaque locuteur, nous codons le texte lu, trois minutes de conversation guidée et trois minutes de conversation libre.

Ce codage, que nous faisons sur une base auditive, demande au transcripteur un gros travail d'écoute et de concentration. Il est inévitable que des erreurs se glissent dans les annotations et que des sites soient omis. Malgré les manques évidents, coder les sites de schwa à grande échelle nous semble important 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 compense les erreurs éventuelles et que le décodage automatique qu'autorisent nos outils nous garantit un repérage simple des sites oubliés.

Nous partons en général du "e" graphique car il représente dans la plupart des cas une voyelle latente, c'est-à-dire 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): "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 consiste à annoter 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*).

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'est-à-dire 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 est codé comme dans *gouvernement*. Sont 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*[ə]. Pour permettre de mieux comprendre ce phénomène nous codons non seulement les schwas graphiques mais aussi toutes les consonnes finales prononcées.

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. Nous nous contenterons de résumer ici nos conventions. Nous donnerons plus loin comme exemple une partie du texte codée selon ces conventions

(a) *Premier chiffre :*

| | |
|-----------------|---|
| Schwa présent | 1 |
| Schwa absent | 0 |
| Schwa incertain | 2 |

(b) *Deuxième chiffre :*

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

(c) *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 |

(d) *Quatrième chiffre :*

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

Frontière intonative faible 4

4. 3.2 Le texte PFC et son codage

Le texte PFC a été construit avec soin en ce qui concerne le schwa. Nous avons intégré dans le texte tous les contextes recensés dans le travail fondamental de Dell (1973/1985) sur la question. Nous avons donc distingué 10 grands contextes dont nos codages tiennent compte :

(a) Schwa initial d'énoncé

Le village § C@ # CV

De plus § C@ # OL

(b) Schwa initial de mot polysyllabique

ses chemises V # C@

en revanche V # C@

(c) Schwa interne

bêtement VC@CV

gouvernement VrC@CV

(d) Schwa final postconsonantique devant V

Ministreira

communea

(e) Schwa final postconsonantique devant C

villagede

Ministrene

(f) Schwa final postconsonantique d'énoncé

colère.

Ministree.

(g) Schwa final postvocalique marqueur du féminin (présent via la longueur de la voyelle?)

tournee de

anne.

(h) Schwa de monosyllabe

décidé de faire V#C@#CV

Il a le sentiment V#C@#CV

(i) Suites de schwas

sentiment de se trouver V#C@#C@#

plutôt que de se trouver

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

(j) Sites fréquents d'insertion

Marc_Blanc

Ouest_Liberté

Pour illustrer le codage du texte PFC, nous donnons ci-dessous une réalisation possible de ce texte où nous avons indiqué par // des pauses ou des transitions rythmiques. Dans cette transcription informelle, l'absence de schwa est indiqué par une apostrophe; sa présence par un soulignement de la voyelle la présence de schwa est indiquée graphiquement. Ce texte est suivi de son équivalent codé.

Prononciation du texte

// 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 écrire' 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 sympathique0412 maire0412 de1122 Beaulieu ne1132 sait plus à quel0412 saint se1112 vouer. Il0411 a le1112 sentiment de1132 se1112 trouver dans une0411 impasse0412 stupide2413. Il0412 s'est, en désespoir0412 de1122 cause0413, décidé à écrire0411 au Premier Ministre0423 pour0412 vérifier si son village0411 était vraiment une0411 étape0412 nécessaire0412 dans la tournée prévue. Beaulieu préfère0411 être0421 inconnue et tranquille0413 plutôt que1112 de0112 se1122 trouver au centre1421 d'une0412 bataille0412 politique0412 dont, par0412 la télévision, se1232raient témoins des millions d'électeur0413s.

4.3.3 Le classeur Schwa

Cet outil, développé par A. Meqqori,¹⁹ accepte en entrée aussi bien des TextGrids que des sources reformatées par Transpraat. A partir de textes codés, cet utilitaire a pour fonction de classer et de comptabiliser toutes les occurrences

¹⁹ Voir Meqqori et Durand (2003) pour le manuel d'utilisation du classeur-schwa.

des codes PFC pour l'analyse du schwa. Il offre plusieurs modes de visualisation, de recherche et de sauvegarde qui accélèrent l'interprétation phonologique des résultats.

The screenshot shows the 'ClasseurSchwa' application window. It features a menu bar (Fichier, Edition), a toolbar with options like 'Traiter un texte (format textgrid)' and 'Traiter un texte (format texte seul)', and a main display area divided into several panes. On the left, there's a 'Jouer le son' section with a dropdown menu showing '0412' selected, and a table of occurrences. The main area is split into 'Texte codé', 'Texte recodé', 'Occurrences et contextes codés', and 'Occurrences et contextes recodés'. The 'Occurrences et contextes codés' pane shows a list of occurrences for the code '0412' with their respective counts.

| Code | Occurrences |
|------|-------------|
| 0112 | 08 |
| 0122 | 01 |
| 0134 | 01 |
| 0212 | 01 |
| 0312 | 03 |
| 0411 | 20 |
| 0412 | 57 |
| 0413 | 12 |
| 0414 | 08 |
| 0421 | 04 |
| 0422 | 03 |
| 0423 | 04 |
| 0432 | 02 |
| 1112 | 12 |
| 1122 | 16 |
| 1131 | 01 |
| 1132 | 07 |
| 1212 | 02 |
| 1214 | 01 |
| 1322 | 01 |
| 1412 | 01 |
| 1414 | 01 |
| 1422 | 11 |
| 1424 | 01 |
| 2412 | 01 |
| 2413 | 01 |
| 2422 | 01 |

4.3.4 Vers une théorisation

Nous avons souligné plus haut que notre approche nous permet de rassembler un maximum de données qui sont soumises à analyse mais ne constitue en aucun cas une position théorique forte et encore moins une véritable analyse. En effet, si on tient à être précis, tout ce que nous enregistrons dans divers contextes c'est la présence ou l'absence d'un segment vocalique dont l'interprétation précise appartient à la théorie phonologique. Nous avons utilisé le terme 'schwa' jusqu'ici pour ne pas trop alourdir la discussion mais le linguiste doit exercer beaucoup de prudence dans son interprétation des codages. Pour donner un exemple concret, dans de nombreuses variétés méridionales, nous avons remarqué dans nos enregistrements des locuteurs les plus âgés que la première syllabe d'un mot comme *mener* (*menez*, *meniez*, etc.) est toujours syllabe 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 (voir Durand, Slater & Wise 1987). 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.

On aura compris qu'un traitement adéquat du schwa ne peut se faire que dans un cadre explicite. De nombreux membres du projet ont déjà contribué aux débats théoriques dans ce domaine (voir par exemple Durand & Laks 2001, Jetchev ce volume et les références qu'il fournit). Il est néanmoins évident que, pour avancer dans le traitement du schwa, il faut aller vers des données quantitatives qui permettent de régler certains problèmes. Par exemple, il est beaucoup question de la stabilisation du 'e' muet mais ces réflexions ne peuvent avancer qu'en observant le comportement des 'e' en contexte.

4.4. La liaison et son codage

Comme pour le 'e' muet, la notation de la liaison adoptée dans PFC vise, d'une part, à permettre une transcription aussi rigoureuse que possible des liaisons observées dans l'analyse auditive des enregistrements et, de l'autre, à favoriser une étude quantitative par extraction automatique des données. Comme c'est le cas pour le *classeur-schwa*, le *classeur-liaison* (§4.3) donne des résultats organisés et multifactoriels à partir desquels le chercheur peut commencer à tester des hypothèses fortes sur le fonctionnement de la liaison dans la diversité des usages.

Comme pour le schwa, le codage de la liaison est appliqué à la transcription orthographique de base dans une tire spécifique sous PRAAT. Il est utilisé pour tous les locuteurs dans le texte lu, cinq minutes de conversation guidée et cinq minutes de conversation libre. Le codage se veut aussi simple que possible car il 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). Le principal 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*), et dans quels contextes est-elle absente ou quasi absente (liaison *erratique*)?

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

Est 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 sont aussi notés tout comme les cas (rares) de liaison devant initiale consonnantique *il peut venir* (*il peut* [t] venir). Nous les avons appelés liaisons épenthétiques ci-dessous.

Notre approche, à ce stade, consiste à repérer manuellement les contextes de liaison. Nous nous appuyons ici sur les travaux classiques et, en particulier, les recensements de Delattre (1951) qui offre un tableau détaillé des cas classés comme liaisons obligatoires, facultatives et interdites. Parmi les contextes de liaison dite interdite, nous en relevons deux que nous avons choisi d'exclure de notre prérepérage: un substantif singulier +....., ainsi que la conjonction *et*+...: *un soldat/afghan*, *un soldat/arrive*, *et/il arrive*. Ces liaisons sont néanmoins notées si elles sont attestées. Dans des travaux en cours, nous explorons un prérepérage des contextes de liaison à partir d'analyseurs syntaxiques. Cela nous permettra une plus grande systématité et en retour un meilleur traitement des facteurs syntaxiques impliqués dans la liaison.

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 vérifier l'ampleur actuelle du phénomène 'liaison' selon les styles (lecture vs. conversation) et les variétés du français.

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, 4} où :

- 0 = absence de liaison ;
- 1 = liaison enchaînée ;
- 2 = la liaison non enchaînée ;
- 3 = incertitude;
- 4 = liaison "épenthétique"

Les autres symboles proposés plus bas signalent la nature de la consonne de liaison, l'existence de pauses-hésitations et/ou coups de glotte, les cas de liaison inattendue (par rapport à la graphie) et les liaisons avec consonnes nasales. On notera que dans tous les codages l'ordre des symboles est fixe.

Avant de présenter les codages en détail, nous donnerons un court exemple concret pour indiquer l'esprit dans lequel ils sont conçus. Soit le tout début du texte PFC où la lecture des cas de (non-)liaison est notée en transcription phonétique large entre crochets: "Le maire de Beaulieu - Marc Blanc - est [ɛ̃ a] en revanche très [ʁɛ zɛ̃kjɛ] inquiet." Le codage sera: "Le maire de Beaulieu - Marc Blanc - est10 en revanche très11z inquiet". L'interprétation des codes est la suivante:

est10 en = monosyllabe (1) + absence de liaison (0)

très11z inquiet = monosyllabe (1) + liaison enchaînée (1) en [z]

4.4.1 Le système de 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 codage PH est signalé par la lettre **h** en minuscule. Nous commençons par le nombre de syllabes du mot liaisonnant (4.4.1.1), nous passons ensuite à l'absence de liaison et aux cas d'incertitude (4.4.1.2), puis au cas de liaison par une consonne non nasale (4.4.1.3) en séparant les liaisons enchaînées (4.4.1.3.1), des liaisons non enchaînées (4.4.1.3.2). Nous passons en 4.4.1.4 aux liaisons en consonne nasale (normalement [n]). Enfin, en 4.4.1.5, nous offrons un codage pour les cas rares où la liaison est observée entre un mot liaisonnant et un mot à initiale consonantique ainsi que pour les liaisons ne correspondant pas à une consonne graphique.

4.4.1.1 Longueur du mot liaisonnant

Ce paramètre correspond au premier symbole de la séquence de codes:

- 1 = une syllabe
- 2 = deux syllabes ou plus

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

Exemples:

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

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

Toutes les descriptions traditionnelles incorporent dans leurs descriptions la longueur du mot liaisonnant. Selon Delattre (1951), une liaison est obligatoire après un invariable monosyllabique (*dans un an*), mais facultative après un invariable polysyllabique (*pendant un an*). Même si cette caractérisation est trop forte,²⁰ le facteur 'longueur du mot' joue un rôle non négligeable dans les enquêtes en cours d'étude. Ce n'est pas un hasard, si la liaison dite 'catégorique' correspond à des déterminants ou pronoms personnels, etc., qui sont tous des monosyllabes. Le codage de la longueur du mot liaisonnant nous permettra d'évaluer de façon plus précise l'importance de ce paramètre.

4.4.1.2 Absence de liaison et incertitude

L'absence de liaison correspond aux trois codages suivants:

²⁰ Dans notre corpus, *dans* ne déclenche pas obligatoirement la liaison.

4.4.1.2.1 Pas de liaison, pas de rupture

Codage: **0**

Exemple: *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 finale fixe s'enchaîne, on utilisera cette même notation : *pâtes italiennes* [pʰitaljen] se notera <pâtes10 italiennes>.

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

Codage: **0h**, où **0** = absence de liaison et **h** = hésitation

Exemple: *petit écrou* [pti PH ʔekru] → <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 <petit10h écrou> les trois réalisations suivantes [ptiʔekru], [pti PH ʔekru] ou [pti PH ekru]. Cela ne signifie pas que toutes ces réalisations ont nécessairement la même valeur mais qu'on laisse l'examen acoustique de ces cas à une étape ultérieure.

Il faut signaler que la notation 0h est également utilisée pour les cas où le mot liaisonnant se termine par une consonne qui n'est pas enchaînée :

Exemple: *pâtes italiennes* [pat ʔitaljen] → <pâtes10h italiennes>

4.4.1.2.3 Incertitude

Codage: 3

Exemple: *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.

4.4.1.3 Liaison réalisée

4.4.1.3.1 Liaison enchaînée

Codage: **1K(h)**, où **1** = liaison réalisée et où **K** est une variable correspondant à la valeur phonémique de la consonne de liaison (ex. z pour *mes amis* [mezami]). Le **h** entre parenthèses, dénote la présence ou l'absence de pause-hésitation

Exemples:

petit écrou = [pti tekru] → <petit11t écrou>

pâtes italiennes [patzitaljen] → <pâtes1 lz italiennes>.

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 ou coup de glotte)

Codage: **1K**

petit écrou signal auditif : [pti tekru] → codage <petit1 lt 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] → <trop1p 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 très exceptionnellement [k] - *long été*), il faut prévoir que certaines variétés puissent fonctionner d'une autre manière. On autorisera donc toutes les lettres consonantiques avec leur valeur phonétique transcrites en SAMPA : <b c d f g h j k l m n p q r s t v w x z > = IPA standard [b c d f g h j k l m n p q r s t v w x z].²¹

(ii) Liaison enchaînée mais avec pause-hésitation

Codage : **1Kh**

petit écrou [pti PH tekru] → codage <petit1 lth 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>

²¹ 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.

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. Le terme "épenthétique" est utilisé en référence à la graphie. Il ne faut y associer aucune interprétation théorique.

4.4.1.3.2 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* [tropete] 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.

4.4.1.4 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* → <bon11n appareil> Liaison enchaînée en [n].

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

2n : *bon appareil* → <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*, soit une voyelle nasale (type *mon été* 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* → <bon11nVO été>. Liaison enchaînée en [n] ; la dernière voyelle du mot liaisonnant est orale d'où VO pour Voyelle Orale.

1nVN : *bon été* → <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 est facultative pour les dix minutes de parole spontanée qui sont codées dans le projet mais elle est utilisée pour le texte.

4.4.1.5 Liaison devant un mot à initiale consonantique

Codage: **1tC**

on peut s'opposer → 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.

4.5. Exemple de codage

Un exemple d'application des codages à une lecture possible du début texte est la suivante. (Les (non-)liaisons sont notées en transcription phonétique large entre crochets.)

"Le maire de Beaulieu - Marc Blanc - est [ɛ tã] en revanche très [trɛ zɛkʃjɛ] inquiet. La cote du Premier Ministre ne cesse de baisser depuis les [lɛ zeleksjɔ] élections. Comment [kɔmã ā], en plus, éviter les manifestations qui ont [õt ?y] eu tendance à se multiplier lors des visites [vizit ʔfisijɛl] officielles ?"

Le codage sera:

"Le maire de Beaulieu - Marc Blanc - est11t en revanche très11z inquiet. La cote du Premier Ministre ne cesse de baisser depuis les11z élections. Comment20, en plus, éviter les manifestations qui ont1 eu tendance à se multiplier lors des visites20h officielles?"

L'interprétation des codes est la suivante:

- *est11t en* = monosyllabe (1) + liaison enchaînée (1) en [t]
- *très11z inquiet* = monosyllabe (1) + liaison enchaînée (1) en [z]
- *les11z élections* = monosyllabe (1) + liaison enchaînée (1) en [z]
- *Comment20, en* = polysyllabe (2) + absence de liaison (0)
- *Ont12th eu* = monosyllabe (1) + liaison non enchaînée (2) en [t] avec pause-hésitation et/ou coup de glotte (h)
- *visites20h officielles* = polysyllabe (2) + absence de liaison (0) + non enchaînement avec pause-hésitation et/ou coup de glotte (h).

4.4.3 Le classeur liaison

Le travail de codage n'est efficace que dans la mesure où des outils permettent d'extraire les informations, de les classer et de les croiser. Comme pour le schwa, A. Meqqori a mis au point pour le projet un outil, appelé Classeur-liaison qui permet une fouille systématique des codages effectués sous PRAAT. Cet outil convivial, qui est disponible sur le site WEB dePFC, fonctionne sur plusieurs plateformes et préfigure les outils de recherche qui permettront une exploitation de la base PFC sous sa forme finale.

4.4.4 Des observations à la théorisation

La liaison, on le sait, est un domaine très fréquenté à la fois sur les plans théoriques, descriptifs et pédagogiques. Malgré la multiplicité des travaux sur ce thème, nous sommes pourtant loin d'avoir des données quantitatives et qualitatives comparables sur la question.

Une des difficultés majeures est de séparer le prescriptif du descriptif dans de nombreux travaux publiés sur la liaison au cours du vingtième siècle. Le plus souvent, les milieux sociaux et les styles décrits ne sont ni identifiés de façon précise, ni stables. En décrivant la liaison, des phonéticiens dont on admire l'objectivité dans d'autres domaines, sont pris dans la tenaille de la norme et ne peuvent cacher leur mépris ou leur irritation à l'usage de diverses classes sociales. Passy (1913) et Martinon (1913) ne sont ni les premiers ni les derniers à nous rappeler en termes quasiment identiques que:

"L'emploi des liaisons varie considérablement selon le style et selon les personnes. Dans le langage littéraire on lie beaucoup plus que dans le style familier ; mais ce

sont surtout les instituteurs, les professeurs de diction, et encore plus les personnes peu instruites essayant de "parler bien", qui introduisent des liaisons en masse. Parfois alors elles se trompent et emploient mal à propos (z) ou (t) comme son de liaison [*des cuirs, des velours*](1)."²² (Passy 1913 : 117)

"D'ailleurs un grand nombre de liaisons sont facultatives et dépendent du goût de chacun [...]"²³ D'une façon générale, les professeurs en font plus que les gens du monde, à cause de l'habitude qu'ils en ont; les instituteurs en font trop, non pas tant peut-être en parlant qu'en enseignant à lire car ils ne savent pas toujours que, même en lisant, il y en a qu'on ne fait pas. Mais les acteurs en abusent étrangement, soit sous prétexte de correction, soit parce qu'ils s'imaginent qu'ils se font mieux comprendre, et cela à la Comédie Française comme ailleurs..." (Martinon 1913 : 356-357)

Cet entrelacement du prescriptif et du descriptif signifie qu'on n'a pas toujours la profondeur historique des observations qu'on pourrait souhaiter. De nombreuses descriptions s'accordent pour souligner que "la liaison est en perte de vitesse", ce qu'affirmait déjà Grammont (1914: 132). Mais comment vérifier cette assertion en l'absence d'études véritablement empiriques?

Dans les enquêtes PFC effectuées en France et déjà codées nous observons une stabilité de la plupart des liaisons présentées comme obligatoires ou interdites dans de nombreux travaux de référence: par exemple, DET + X (*les-enfants, ces-étudiants, nos-espoirs, un-autre, mon-ami*) ou PRON + X (*nous-avons, vous-autres*). De même, pour ce qui est des liaisons dites 'interdites', l'usage semble conforme aux attentes. Dans la lecture du texte, la liaison est effectivement absente dans:

²² Le renvoi (1) que fournit cette citation est tout aussi révélateur: "J'évite les liaisons. C'est prétentieux... et dangereux", Labiche, *La grammaire*."

²³ Et comme en écho, on peut citer Milner et Regnault (1987). Comme le souligne Laks (2002), ces derniers écrivent que la liaison est principalement affaire de goût (p. 59) et, notant que l'usage en la matière est historiquement fluctuant, ils citent (p 51) Abel Hermant qui, au début du siècle, fustigeait l'enflure bourgeoise en termes vifs: "Je vous accuserai d'aller à la bourgeoisie, si vous faites trop de liaisons. Nos pères en faisaient fort peu."

comment / en plus
 quelques fanatiques / auraient
 vraiment / une étape nécessaire

Les différences avec la "norme" se situent dans des cas déjà signalés dans d'autres travaux attentifs aux usages. Ainsi, la catégorie ADJ + N est loin de déclencher la liaison de façon automatique. Même dans la lecture du texte PFC, exercice censé provoquer la prononciation la plus formelle, les exemples de *gran(d) émoi* sans liaison sont fréquents et souvent accompagnés d'hésitation. Cela tendrait à démontrer que les locuteurs ne savent pas comment traiter de telles combinaisons qui ne font manifestement pas partie de leur lexique mémorisé. De même, certains mots grammaticaux monosyllabiques comme *pas* ou *dans*, souvent jugés comme déclenchant des liaisons obligatoires, se révèlent variables, parfois même au sein d'un même énoncé. Parmi les expressions fixes, certaines séquences comme *jeux olympiques*, présentées comme des exemples avérés de liaison obligatoire (*jeux-z olympiques*) par un Fouché (1959: 441) se révèlent variables.

La difficulté est que nous ne savons pas si la description des liaisons obligatoires, interdites et variables de la tradition correspondait en détail à l'usage dans des états de langue antérieurs. Alors que Passy (1913) classait 'pas' et 'est' avec les liaisons obligatoires, Grammont (1914: 132) signalait qu'on pouvait dire: *Mon père e(st) en ville*, ou *Mon père est en ville*; *Mon père n'est pa(s) en ville*, ou *Mon père n'est pa(s) z en ville*. Pour prendre un autre exemple, lorsque Passy (1913: 118) affirmait: "Dans la liaison, les plosives sont toujours soufflées, les fricatives toujours voisées: *un long hiver* (œlōkivær), *un gros homme* (œgrozom)", reflétait-il des observations assurées ou le "bon usage" tel qu'il le percevait, l'imaginait ou cherchait à le construire?²⁴ Martinon nous dit bien, au même moment, que "dans la liaison la gutturale douce devient forte" mais il signale que dans l'usage courant "le g ne se lie plus que dans l'expression *sang et eau*" (1913: 362). En vers ou dans le style oratoire, on observe, selon lui, un plus grand nombre de

²⁴ Affirmation que reprend Schane (1968: 127) en la liant au registre: "In some styles, the liaison consonants are voiced whereas the liaison stops (noncontinuant) are voiceless: e.g. *neuf heures* /nœvœr/, *grand ami* /grā tami/ 'big friend', *long été* /lōkete/'. L'inclusion de *neuf* est de toute façon discutable puisque la consonne finale de ce mot est prononcée en position prépausale et préconsonantale.

liaisons en [k] (comme *rang[k]élevé*) ; mais la note 3 à la même page révèle quelques doutes sur l'étendue du phénomène: "Je dois avouer d'ailleurs que dans les liaisons qui ne se font qu'en vers, comme celle de *long espoir*, il y a déjà tendance à conserver au *g* le son doux".

Le projet PFC ne peut prétendre combler les trous dans l'histoire de la liaison, même si la stratification en âge des enquêtés peut nous fournir des pistes intéressantes dans le domaine diachronique. En revanche, le projet fournira un référentiel important pour l'examen de la liaison dans ses dimensions individuelles et collectives.²⁵ On sait déjà que sur les points d'enquêtes étudiés, on a au bas mot une moyenne de 150 codages par locuteur. Cela signifie que pour une base exploitable de 300 locuteurs, le projet disposera de 45000 codages. Un tel chiffre nous placera dans une position nouvelle quant à l'étude de divers aspects de la liaison.

Bien évidemment, les observations que l'on peut extraire de nos codages ne fournissent pas une théorie. La liaison a été un lieu de bataille entre de nombreuses approches théoriques. Le générativisme s'y est opposé au structuralisme. A l'intérieur du générativisme, approches concrètes et abstraites, linéaires et non-linéaires, systèmes de règles et de contraintes, se sont disputés le terrain. Tout récemment, la montée de la linguistique de corpus a jeté de nouvelles hypothèses sur la question. Quel que soit le cadre théorique retenu, il devrait être clair que les données sont primordiales. Comment, par exemple, estimer le rôle du non-enchaînement dans la phonologie du français quand on va au delà des professionnels de la parole publique? Comment savoir quelle est la fonction précise des constituants syntaxiques si on se limite à des exemples construits? Comment savoir si l'espace francophone est relativement homogène ou non dans son traitement de la liaison? Le projet PFC, grâce à l'application d'un protocole uniforme et de techniques de transcription et de codage homogènes, fournira des réponses précises et quantifiables à ces questions. Divers collaborateurs au projet PFC, comme Durand (1986), Encrevé (1983, 1988), Girard & Lyche (1997), Laks (1980, 1983) et Walker (2001: 160-169), ont

²⁵ Dans le domaine de la liaison, Ågren, J. (1973) et De Jong (1993, 1994) fournissent des données importantes pour les comparaisons que le projet PFC permettra d'effectuer. Le travail sur l'acquisition de la liaison entrepris par des collègues comme J.-P. Chevrot et S. Wauquier-Gravelines est également précieux pour une meilleure compréhension de ce phénomène : voir Chevrot & Fayol (2000), Chevrot (2001), Wauquier-Gravelines (2002).

déjà offert des analyses de la liaison. On peut espérer que les nouvelles données vont permettre à ces chercheurs et à d'autres de faire avancer les débats et d'asseoir les constructions théoriques sur des observations plus fiables et plus fines.

5. Conclusion

Comme le notent Durand, Laks & Lyche (2003), dont nous reprendrons la conclusion, la description et l'analyse du français oral constituent un terrain scientifique ancien et très fréquenté. La description des processus phonologiques remonte aussi loin que l'institutionnalisation de la langue elle-même puisque, dès 1530, Palsgrave propose une analyse assez précise de la liaison, de l'élosion, de l'accentuation et du rythme. Les descriptions et les analyses, souvent basées sur l'intuition du linguiste ou sur sa reconnaissance de la norme académique, se sont depuis lors multipliées. A la fin du 19^{ème} siècle, les développements de la phonétique instrumentale et descriptive comme celle des premiers dispositifs d'enregistrement de la parole font naître un grand espoir dont Paul Passy en France et Daniel Jones en Grande-Bretagne se sont fait les chantres et les artisans. Esprit vite déçu puisque trop souvent la phonologie du français poursuivra imperturbablement son cours d'exégèse normative ou de contemplation théorique et que la phonétique du français continuera trop souvent elle aussi d'accumuler des descriptions éparses et ponctuelles, mais sans systématique et sans construction d'un cadre hypothético-descriptif précis.

Des atlas linguistiques de Gilliéron et Edmont (1902-1912) aux grandes enquêtes dialectologiques ou sociolinguistiques, l'histoire de la phonologie de corpus en France est ainsi celle d'un descriptivisme marginal sans grande influence sur le cours des phonologies analytiques ou prescriptives plus nobles parce que plus académiques. Pourtant, ce que nous apprennent l'histoire et l'épistémologie des sciences du langage c'est qu'il ne suffit pas de décrire et d'observer pour rendre compte d'un objet linguistique; il faut encore le construire et l'interroger dans une problématisation scientifique qui seule permet de le saisir comme objet. A l'inverse, toute l'histoire de la phonologie du français montre qu'il ne suffit pas de s'interroger sur les usages, voire d'exercer son introspection sur le sien propre, pour rendre compte de la langue dans sa diversité et sa variation. Sans protocole empirique et sans collecte systématique de données, la phonologie se trouve sans objet et c'est, comme l'on sait, la norme et le standard le plus reçu, et donc le moins fondé, qui

viennent alors, *volens nolemis*, occuper le vide empirique ainsi créé.

L'histoire des rapports entre phonologie et phonétique du français est ainsi celle d'un mariage annoncé mais sans cesse repoussé faute de premier rendez-vous. L'objectif premier du projet PFC est d'abord d'organiser la rencontre en rabattant sur l'enquête les problématisations théoriques de phonologie française dont elle a besoin pour construire son interrogation comme systématique et informative; c'est aussi de renvoyer à la phonologie du français le *datum* qui lui fait défaut, un *datum* mis en forme et problématisé, qui entre directement en résonance avec les questionnements théoriques qui sont les siens. Au-delà, l'objectif du projet PFC, en construisant une base de donnée représentative de la variété sociale, géographique et stylistique des usages phonologiques du français, est de constituer le référentiel contemporain nécessaire à la linguistique française. C'est assurément le rôle de la linguistique de corpus que de construire ce référentiel empirique, de développer les bases de données, de définir leur organisation, de développer les outils de fouille nécessaires à la linguistique. Mais ceci ne peut se faire hors des problématisations théoriques qui structurent les sciences du langage. On sait qu'il n'existe pas de corpus omnibus et qu'une base de données n'est pas un couteau suisse. C'est la raison pour laquelle le projet PFC, projet de phonologie de corpus se veut à la fois phonétique et phonologique, empirique et théorique, de terrain et d'élaboration conceptuelle. C'est à ce prix nous semble-t-il que nous comprendrons mieux ce qu'il en est à l'oral des usages du français.

REMERCIEMENTS

Un grand nombre de chercheurs et d'étudiants ont participé ou participent au projet PFC. Nombreux sont ceux qui nous ont apporté leur soutien intellectuel, moral et institutionnel. Sans leur aide, notre entreprise n'aurait aucune réalité. Nous renvoyons au site web PFC <http://infolang.u-paris10.fr/pfc> pour une liste des membres de notre projet. PFC a reçu le soutien financier de nombreux organismes parmi lesquels le CNRS, la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF), l'Institut de Linguistique Française (ILF), les régions Midi-Pyrénées et Midi-Aquitaine, le Ministère de la Recherche et les nombreuses universités des participants.

Références bibliographiques

Ågren, J. (1973). *Enquête sur quelques liaisons facultatives dans le français de conversation radiophonique: fréquence et facteurs/* Uppsala, Suède : Acta Universitatis Upsaliensis.

Bergounioux, G. (ed.)(1992). *Enquêtes, corpus et témoins en France hier et aujourd'hui. Langue française* 93: 3-23

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

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

Blanche-Benveniste, C. & A. Chervel (1969). *L'orthographe*. Paris: Maspero.

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

Boula de Mareüil, P. & Z. Fagyal (2001). Autour de l'harmonie vocalique en français. Aussois : *Actes des 23^e journées d'Etude sur la Parole*, 85-88.

Bourdieu, P. (1993). *La misère du monde*. Paris: Éditions du Seuil.

Buben, Vl. (1935). *Influence de l'orthographe sur la prononciation du français moderne*. Bratislava et Paris: Droz.

Caelen-Haumont, G. (2002). Quelques outils de transcription pour l'analyse de corpus oraux. *Quatrième Journées Phonologie du Français contemporain*, Paris X-Nanterre, 23-25 janvier 2002.

Catach, N. (1993). *L'orthographe*. (Cinquième édition.) Que sais-je? Paris: PUF.

Chevrot, J.-P. (2001). *Variation phonétique, développement, orthographe*. Dossier non publié présenté en vue de l'Habilitation à diriger des recherches. Vol. 1 : Synthèse. Université Stendhal – Grenoble 3.

Chevrot, J.-P. & M. Fayol (2000). Acquisition of French liaison and related child errors. In : M. Almgren et alii (eds)(2000), *Research on Child Language Acquisition*. Proceedings of the 8th Conference of the International Association for the Study of Child Language, vol. 2, Cascadilla Press, 760-774.

Chevrot, J.-P. & I. Malderez (1999). L'effet Buben: de la linguistique diachronique à l'approche cognitive (et retour). *Langue Française* 124: 104-125.

De Jong, D. (1993). Sociophonological aspects of Montreal French liaison. In : W. Ashby (ed.)(1993), *Linguistic Perspectives on Romance Languages. Proceedings of the Twenty-First Linguistic Symposium on Romance Languages*. Amsterdam : John Benjamins.

De Jong, D. (1994). Sociophonologie de la liaison orléanaise. In : C. Lyche (ed.)(1994), *Generative Phonology : Retrospective and Perspectives*, Middlesex : AFLS/ESRI, 95-130.

Delais-Roussarie, E. (ce vol. a). Constitution et annotation de corpus : méthodes et recommandations.

Delais-Roussarie, E. (ce vol. b). Quelques outils d'aide à la transcription et à l'annotation de données audio pour constituer des corpus oraux

Delais-Roussarie, E., A. Meqqori & J.-M. Tarrier (ce vol.). Annoter et segmenter des données de parole sous PRAAT.

Delais-Roussarie, E., J. Durand, C. Lyche, A. Meqqori et J.-M. Tarrier (2002). Transcriptions des données : outils et conventions. *Bulletin PFC* 1 : 21-34. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Delattre, P. (1951). *Principes de phonétique française à l'usage des étudiants anglo-américains*. Middlebury College.

Delattre, P. (1966). *Studies in French and Comparative Phonetics*. La Haye: Mouton.

Dell, F. (1985). *Les Règles et les sons*. (Première édition, 1973.) Paris : Hermann.

Deyhime, G. (1967). Enquête sur la phonologie du français contemporain. *La linguistique* I: 97-108 et II: 57-84.

Durand, J. (1986). French liaison, floating segments and other matters in a dependency framework. In: J. Durand (ed.)(1986). *Dependency and Non-Linear Phonology*, Londres: Croom Helm, 161-201.

Durand, J. & B. Laks (2000). Relire les phonologues du français. Maurice Grammont et la loi des trois consonnes. *Langue Française* 126 : 29-38.

Durand, J. & B. Laks (eds)(2002a). *Phonetics, Phonology and Cognition*. Oxford: Oxford University Press.

Durand, J. & B. Laks (2002b). Phonology, phonetics and cognition. In : J. Durand & B. Laks (eds)(2002a), 1-50.

Durand, J., B. Laks & C. Lyche (2002a). (PFC) Protocole d'enquête. *Bulletin PFC* 1 : 7-20. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Durand, J., B. Laks & C. Lyche (2002b). Directions d'analyse. *Bulletin PFC* 1 : 35-70. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Durand, J., B. Laks & C. Lyche (2002c). Format des rendus. *Bulletin PFC* 1 : 71-74. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Durand, J., B. Laks & C. Lyche (2003). PFC : Un corpus numérisé pour la phonologie du français. In : G. Williams (ed.), *Les linguistiques de corpus*, Rennes: Presses Universitaires de Rennes (à paraître).

Durand, J. B. Laks & C. Lyche (ce vol.). Linguistique et variation : quelques réflexions sur la variation phonologique

Durand, J. & C. Lyche (1999). Regard sur les glissantes en français : français standard, français du Midi. *Cahiers de grammaire* 24 : 39-65.

Durand, J. & C. Lyche (2003). Structure et variations dans les systèmes vocaliques du français. In: C. Sanders, A. Coveney & M.-A. Hintze (eds.), *Variation et francophonie*, Paris : L'Harmattan (à paraître).

Durand, J., C. Slater & H. Wise (1987). Observations on schwa in southern French. *Linguistics* 25: 983-1004.

Durand, M. (1936). *Le genre grammatical en français parlé à Paris et dans la région parisienne*. Paris: D'Artrey.

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

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

Encrevé, P. (1983). La liaison sans enchaînement. *Actes de la recherche en sciences sociales* 46: 39-66.

Encrevé, P. (1988). *La liaison avec ou sans enchaînement. Phonologie tridimensionnelle et usages du français*. Paris : Éditions du Seuil.

Fouché, P. (1959). *Traité de prononciation française*. Paris: Klincksieck.

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

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

Gilliéron, J. & E. Edmont (1902-1912). *Atlas Linguistique de la France*. Paris : Champion.

Girard, F. & C. Lyche (1997). *Phonétique et phonologie du français*. (Troisième édition.) Oslo : Universitetsforlaget.

Hansen, A. B. (1990). Analyse sociolinguistique de deux évolutions linguistiques dans le français parlé à Paris. La stabilisation du 'e caduc' interconsonantique et l'apparition d'un /ə/ final. Mémoire de maîtrise, Université de Copenhague.

Jetchev, G. (ce vol.). La variabilité du schwa français vue à travers une approche syllabique.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, D.C. : Center for Applied Linguistics.

Labov, W. (1972a). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1972b). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Philadelphia Press.

Laks, B. (1977). Contribution empirique à l'analyse socio-différentielle de la chute des /r/ dans les groupes consonantiques finals. *Langue française* 34, 109-125.

Laks, B. (1980). Différenciation linguistique et différenciation sociale: quelques problèmes de sociolinguistique française. Thèse de doctorat, université de Paris VIII-Vincennes.

Laks, B. (1983). Langage et pratiques sociales. Étude sociolinguistique d'un groupe d'adolescents. *Actes de la recherche en sciences sociales* 46, 73-97.

Laks, B. (1992). La linguistique variationniste comme méthode. *Langages* 108: 34-50.

Laks, B. (2002). La description du français oral: la phonologie et la norme. *L'information grammaticale*, 5-10.

Le Roy, G. (1967). *Grammaire de diction française*. Paris: Edition de la pensée moderne.

Lieshout, P. van (2000). PRAAT workshop. A basic introduction. Version 1.5.

Lodge, R. A. (1993). *French : From Dialect to Standard*. London : Routledge.

Lyche, C. (1978). On the Optionality of French Schwa Deletion in Natural Generative Phonology. In : K. Gregersen (ed.), *Papers from the Fourth Scandinavian Conference of Linguistics*, Hindsø, 6-8 janvier 1978, Odense: Odense University Press, 441-446.

Lyche, C. (1979). French schwa deletion in Natural Generative Phonology. *Nordic Journal of Linguistics* 2, 91-111.

Lyche, C. (ed.).(1994). *Generative Phonology : Retrospective and Perspectives*, Middlesex : AFLS/ESRI.

Lyche, C. (1996). Genèse et traits caractéristiques du cadien : un aperçu phonologique. *Revue romane* 31, 29-49.

Lyche, C. (ce vol.). La loi de position et le français de Grenoble.

Malmberg, B. (1976). *Phonétique française*. (5ème édition.) Malmö.

Martinet, A. (1945). *La Prononciation du français contemporain*. Genève : Droz

Martinet, A. & H. Walter (1973). *Dictionnaire de la prononciation du français dans son usage réel*. Paris: France-Expansion.

Meqqori, A. & J. Durand (2003). Manuel d'utilisation du classeur-liason. *Bulletin PFC* 2 : 23-29 CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Meqqori, A. & J. Durand (2003). Manuel d'utilisation du classeur-schwa. *Bulletin PFC* 2 : 29-32. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Meqqori, A. & J. Durand (2003). Manuel d'utilisation du comparateur. *Bulletin PFC* 2 : 33-38 CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Meqqori, A. & J. Durand (2003). Manuel d'utilisation de Transpraat. *Bulletin PFC* 2 : 39. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Milner, J.-C. & F. Regnault (1987). *Dire le vers. Court traité à l'attention des acteurs et des amateurs d'alexandrin*. Paris: Editions du Seuil.

Milroy, L. (1980). *Language and Social Networks*. Oxford : Blackwell.

Milroy, L. (1987). *Observing and Analysing Natural Language*. Oxford : Blackwell.

Milroy J. & L. (1991). *Authority in Language. Investigating Language Prescription and Standardisation*. (Première édition, 1985.). Londres: Routledge.

Montreuil, J.-P. (ce vol.). Aspects de la longueur vocalique en français de Basse-Normandie.

Morais, J., S.L. Castro, L. Scliar-Cabral, R. Kolinsky & A. Content (1987). The effects of literacy on the recognition of dichotic words. *Quarterly Journal of Experimental Psychology* 39: 451-465.

Morais, J., S.L. Castro & R. Kolinsky (1991). La reconnaissance des mots chez les adultes illettrés. In : R. Kolinsky, J. Morais & J. Segui (eds.)(1991), *La reconnaissance des mots dans les différentes modalités sensorielles, études de psycholinguistique cognitive*, Paris: PUF, 59-80.

Morais, J., R. Kolinsky, P. Ventura & M. Cluytens (1997). Levels of processing in the phonological segmentation of speech. *Language, Speech and Hearing Services in the Schools* 24: 161-166.

Morin, Y.-C. (2000). Le français de référence et les normes de prononciation. *Cahiers de l'Institut de Linguistique de Louvain*, 21: 91-135

Palsgrave, J. (1530). *L'esclaircissement de la langue françoise, Composé par Maistre Jehann Palsgrave, Angloys, Natyf de Londres et Gradué de Paris*. Londres.

Schane, S.A. (1968). *French Phonology and Morphology*. Cambridge, Mass.: MIT Press.

Tarrier, J.-M. (éd.)(2003). *Enregistrement, prise de sons et outils PFC. Bulletin PFC* 2. 39 p. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Tarrier, J.-M. (ce volume). L'enregistrement et la prise de sons.

Walker, D. (2001). *French Sound Structure*. Calgary: University of Calgary Press.

Walker, D. (ce vol.). Aperçu de la langue français en Alberta.

Walter, H. (1976). *La dynamique des phonèmes dans le lexique contemporain*. Paris : France-Expansion.

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

Wauquier-Gravelines, S. (2001). Contraintes d'acquisition sur le traitement des segments flottants en français: le cas de la liaison. Manuscrit.

Wells, J.C. (1982). *Accents of English*. 3 volumes. Cambridge: Cambridge University Press.

ANNEXES**Annexe I : Fiche signalétique PFC**

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...)

Annexe 2 : 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) ou sur internet. 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:

Date:

Annexe 3 : SAMPA

Sampa est la transcription en codes clavier ASCII de l'alphabet standard API. Cette notation est accessible sur la toile à l'adresse www.sampa. Elle a été développée à l'origine par un groupe de phonéticiens européens dans le cadre du projet ESPRIT 1541 SAM (Speech Assessment Method, 1987-89). Elle a été appliquée vers 1989 à l'allemand, l'anglais, le danois, le flamand, le français et l'italien et depuis à un grand nombre d'autres langues. Il existe deux extensions à SAMPA (mises au

point par J.C. Wells, University College London) qui sont non standardisées mais néanmoins utiles: XSAMPA qui inclut tous les diacritiques API et SAMPROSA qui permet de coder la prosodie.

Comme pour l'API, une transcription en SAMPA est une chaîne de caractères (ou codes) écrits sans espace. Tous les symboles de l'alphabet latin ont été conservés en SAMPA. En ce qui concerne la transcription phonémique du français standard, on obtient les correspondances suivantes:

1. Consonnes

1.1 Plosives:

IPA /p b t d k g/=

SAMPA /p b t d k g/

| Symbole | Exemple | Transcription |
|---------|---------|---------------|
|---------|---------|---------------|

| | | |
|---|-------|-----|
| p | pont | po~ |
| b | bon | bo~ |
| t | temps | ta~ |
| d | dans | da~ |
| k | quand | ka~ |
| g | gant | ga~ |

1.2 Fricatives

IPA /f v s z ʃ ʒ/=

SAMPA /f v s z S Z/

| | | |
|---|-------|-----|
| f | femme | fam |
| v | vent | va~ |
| s | sans | sa~ |
| z | zone | zon |
| S | champ | Sa~ |
| Z | gens | Za~ |

1.3 Nasales

IPA /m n ɲ ŋ/=

SAMPA /m n J N/

| | | |
|---|---------|--------|
| m | mont | mo~ |
| n | nom | no~ |
| J | oignon | oJo~ |
| N | camping | ka~piN |

1.4 Liquides

IPA /l R/=
SAMPA /l R/

| | | |
|---|------|-----|
| l | long | lo~ |
| R | rond | Ro~ |

1.5 Glissantes

IPA /j w ɥ/=
SAMPA /j w H/

| | | |
|---|--------|------|
| j | pierre | pjER |
| w | coin | kwe~ |
| H | juin | ZHe~ |

2. Voyelles

2.1 Voyelles orales

IPA /i e ε a ɔ o u y ø œ ə/ =
SAMPA /i e E a A O o u y 2 9 @/

| | | |
|---|-----------|----------|
| i | si | si |
| e | ses | se |
| E | seize | sEZ |
| a | patte | pat |
| A | pâte | pAt |
| O | comme | kOm |
| o | gros | gRo |
| u | doux | du |
| y | du | dy |
| 2 | deux | d2 |
| 9 | neuf | n9f |
| @ | justement | Zyst@ma~ |

2.2 Voyelles nasales e~ a~ o~ 9~

IPA /ẽ ä õ œ/ =
SAMPA /e~ a~ o~ 9~/

| | | |
|----|------|------|
| E~ | vin | ve~ |
| A~ | vent | va~ |
| o~ | bon | bo~ |
| 9~ | brun | bR9~ |

Pour être précis, les valeurs API données ici pour les voyelles nasales /ẽ ä õ œ/ devraient être codées de la façon suivante: /E~ A~ o~ 9~/ . Tout dépend du degré de réalisme phonétique qu'on tient à assigner à la transcription phonémique des voyelles nasales. Dans la mesure où la plupart des systèmes du français décrits à ce jour n'opposent pas /e~/ et /E~/, /a~/ et /A~/, /o~/ et /O~/, on peut adopter une symbolisation phonémique

relativement lâche. Dans le cadre de PFC où nous sommes sensibles à la variation inter-dialectale, il semble plus utile d'adopter des transcriptions SAMPA en accord avec la réalité phonétique. Si le principal allophone de la voyelle de vin est un [ē], la transcription /E~/ est préférable à /e~/.