

*Script***O**ralia

Romanistische Korpuslinguistik Romance Corpus Linguistics

Korpora und gesprochene Sprache
Corpora and Spoken Language

herausgegeben von
Claus D. Pusch und Wolfgang Raible

Jacques Durand (Toulouse), Bernard Laks (Paris),
Chantal Lyche (Oslo / Tromsø)

***La Phonologie du français contemporain (PFC) :* usages, variétés et structure**

We present a project aiming at a comprehensive socio-phonological investigation of varieties of French spoken in France and outside of France. The work described is based on a common methodology and will therefore provide comparable data for about 350 speakers. It is carried out by seven universities in France and four universities outside of France. We provide the historical context for this project and outline some of the procedures adopted. After a few remarks on the phonological exploitation of the corpus, we move on to explain some options made available by its automatization.

1. Introduction

Si la phonologie moderne tend 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. La variation, qu'elle soit géographique, sociale, stylistique ou tout simplement inhérente au système, y est omniprésente et son traitement représente l'un des grands enjeux de la linguistique moderne. Le projet que nous décrivons brièvement ici s'inscrit dans cette perspective et part de la constatation qu'il est nécessaire de poursuivre le travail de description entrepris depuis au moins un siècle par tous les spécialistes de la communication parlée pour (1.) fournir une meilleure image du français dans son unité et sa diversité, (2.) mettre à l'épreuve les modèles phonologiques sur le plan synchronique et diachronique. Après avoir situé notre projet dans un contexte historique assez large, nous envisagerons les différents volets de l'enquête avant de considérer quelques-uns des aspects phonologiques. En guise de conclusion, nous proposerons nos premières réflexions sur l'automatisation du corpus.

2. Perspective historique

Les grandes enquêtes qui se sont multipliées en France depuis le début du 19^e siècle ont été dans un premier temps de *type phonologique*. La première de grande notoriété est celle dite *de l'Empire* (Coquebert de Mombret 1807-1812), enquête par correspondance avec questionnaire, où il était demandé aux témoins de traduire dans l'idiome local la Parabole de l'enfant prodigue. Le but de cette enquête était en premier lieu d'ordre statistique puisqu'il s'agissait à terme de déterminer le nombre de locuteurs pour chaque idiome décrit. La technique du questionnaire utilisée ici a été systématisée dans l'enquête de Martinet (1945), enquête réflexive où le témoin devait répondre à des questions sur son propre usage, comme par exemple "prononcez-vous tel mot comme ceci ou comme cela ?". Même si dans ce dernier cas, le questionnaire était fouillé et la perspective phonologique claire, du fait de son manque de fiabilité, la méthode indirecte a largement été supplantée dans les autres grandes enquêtes par la méthode directe. L'une des plus célèbres (1897-1901) reste sans nul doute celle du Suisse Gilliéron dont les résultats ont permis l'élaboration de l'*Atlas Linguistique de la France* (1902-1907). Le travail consistait à vérifier pour chaque point d'enquête la forme lexicale et la prononciation d'une liste de 1 500 mots. Dans ce cadre, la primauté est donnée au fond de carte et chaque item est replacé sur une carte de France. En conséquence, les parlars ne sont plus observés comme des unités, ils sont déstructurés et le locuteur perd son identité (Bergounioux 1992). Les *Atlas régionaux* qui, sous l'impulsion de Dauzat (en 1939), ont été élaborés à travers la France entière, adoptent la même méthode mais à l'encontre de l'ALF, ils envisagent des aires dialectales bien définies.

Ces deux méthodes (indirectes et directes), qui ne permettent pas de dépasser le niveau du mot et du syntagme isolé, ont été réutilisées sous une forme amplement modifiée dans les enquêtes de Martinet et Walter (1973) et de Walter (1982). La première, engagée pour l'élaboration d'un dictionnaire de la prononciation, se concentre sur un point géographique bien circonscrit, Paris, à travers les interviews de dix-sept locuteurs appartenant "aux milieux cultivés de la bourgeoisie". La seconde, par contre, reprend la grande tradition des atlas en couvrant 35 points d'enquête, mais elle n'a pas dépassé le stade de la préenquête qui devait permettre la rédaction d'un questionnaire plus précis sur la base des oppositions relevées.

Les limites de ces enquêtes sont multiples aussi bien dans la méthode, comme nous l'avons remarqué plus haut, que dans le choix des locuteurs. Un seul locuteur est le plus souvent jugé représentatif pour un point d'enquête donné et, de surcroît, dans de nombreux cas, ces locuteurs sont des personnes âgées. Les études conduites à partir de ces enquêtes l'ont été, pour la plupart, dans le cadre structuraliste et elles s'efforçaient, pour chaque variété étudiée, de définir le système phonologique de base. La variation est ainsi laissée pour compte et comme Bergounioux (1992 : 106) remarque à propos du structura-

lisme en France, "on a l'impression que c'est un structuralisme qui refoule et la variation, et le corps." Dans le cadre d'enquête de type phonologique, le travail de Gougenheim *et al.* (1956) se distingue nettement des travaux déjà cités aussi bien dans les objectifs que dans la méthode. Ici, un ensemble de documents sonores déjà existants a été consulté et suppléé par quelques enregistrements dans un but pédagogique, l'élaboration du *Français Fondamental*. L'étude d'enregistrements de conversation libre se retrouve de façon systématique dans les enquêtes de *type sociolinguistique* dont la plus célèbre reste l'*Etude socio-linguistique sur Orléans* (ESLO).

Tout comme le projet du *Français Fondamental*, l'ESLO (1969-70) est une enquête à grande échelle entreprise dans un but pédagogique. Les enregistrements portent sur 144 sujets et offrent un total de 300 heures de discours. Les témoins sont sélectionnés par l'INSEE (*Institut National de la Statistique et des Etudes Economiques*) selon la méthode de l'échantillonnage au hasard sur la base de trois principes de classification : le sexe, l'âge et la catégorie socio-professionnelle. Les entretiens sont réalisés à partir d'un questionnaire sociologique fouillé où les sujets les plus divers (travail, loisirs, attitudes par rapport à la langue, etc.) sont abordés. Si la ville d'Orléans a été choisie pour des raisons historiques, géographique et dialectales, les instigateurs du projet n'ont aucunement pour objectif de décrire un français standard ou même neutralisé, pas plus que le français régional d'Orléans. L'intérêt des enquêteurs porte sur toutes les variétés de français parlées à Orléans à l'exception de la langue des immigrés, il s'agit de faire le "portrait sonore" d'une ville. Si les enregistrements n'offrent pas de représentativité géographique et focalisent peut-être trop sur les différences sociales, ils sont aussi difficilement exploitables d'un point de vue phonologique, en l'absence de tout test phonologique. Ces entretiens se prêtent néanmoins à des études portant sur l'interface phonologie-syntaxe, et De Jong (1994) a su tirer un large profit de ces données pour ses travaux sur la liaison.

A l'opposé de ce travail macro-sociologique, de nombreuses enquêtes réalisées en France sont plutôt de type micro-sociologique, comme celle de Laks (1977 ; 1980 ; 1983) qui se rapproche plus du travail effectué par Labov sur la langue des jeunes new-yorkais (Labov 1972). Laks se propose de

saisir un effet différenciateur, une marque phonique qui soit produite comme marque sociale et qui soit donc signe de position et d'origine sociale et, par là même, signe de différenciation en même temps qu'elle est aussi signe de connaissance et de reconnaissance sociale. (Laks 1977 : 110)

Cette enquête a été réalisée à Villejuif, commune ouvrière de la banlieue parisienne, sur un réseau relationnel serré (Milroy 1980). Elle vise à établir la grammaire de locuteurs individualisés et s'attache plus précisément à la chute des liquides finales de mot, à la simplification des pronoms sujets (*il / elle*) et à l'usage de *puisque / parce que / lorsque*. L'enquêteur s'est surtout intéressé à un petit groupe de locuteurs qui, *a priori*, constituaient un groupe naturel, six adolescents de 14-15 ans, enfants de petits salariés, non scolarisés dans l'enseignement long et habitant le même quartier. Pendant neuf mois, l'enquê-

teur a passé quatre après-midi par semaine dans la 'Maison de l'enfance – Maison pour tous', lieu de rencontre privilégié des jeunes. Il a ainsi pu s'intégrer au groupe et, en laissant tourner son magnétophone, a pu avoir accès au vernaculaire de ses témoins. Pour obtenir des échantillons plus formels, l'enquêteur a enregistré ces mêmes témoins en dehors de la 'Maison de l'enfance'. Il dispose alors de deux registres qu'il va pouvoir comparer.

Parmi les nombreuses enquêtes personnelles, nous citerons encore celle d'Agren (1973) pour ses études de la liaison, celles de Dausés (1973), Hansen (1990) et Durand / Slater / Wise (1987) pour l'étude du schwa. Dans cette dernière enquête, la méthode combinait une approche labovienne classique et une approche en termes de réseaux denses. Les huit témoins appartenaient tous à la famille proche de Durand et les enregistrements comprenaient la lecture d'une liste de mots, la lecture d'un texte et des conversations informelles. Toutes ces enquêtes personnelles se caractérisent par l'absence d'envergure géographique, elles sont limitées pour la plupart à la région parisienne et aucune ne prend systématiquement en compte la composante géographique, la composante sociale et la composante stylistique.

Le projet décrit ici ne s'inscrit pas de façon précise dans une démarche sociolinguistique. L'objectif premier de notre enquête reste phonologique et descriptif. Si la variation sociale, géographique ou stylistique est bien évidemment au cœur de notre problématique, qui vise à proposer une description aussi fine que possible de la phonologie du français contemporain, de ses variétés, de ses usages et de ses structures, nous n'avons pas pour objet de travailler la variation sociolinguistique à grande échelle, de la quantifier, de la pondérer par rapport aux différents paramètres sociaux ou même d'en donner des attestations parfaitement représentatives. Ce travail proprement sociolinguistique pourra être entrepris après que PFC a pu dégager les structures phonologiques, les variantes et les usages à analyser. Comme l'a souvent rappelé Labov, l'enquête phonologique descriptive est un préalable aux études variationnistes et ne se confond pas avec elles.

3. *La Phonologie du français contemporain (PFC)*

Ce projet se situe donc dans la continuité des grandes enquêtes phonologiques décrites ci-dessus dans son but affiché d'envisager le français dans son unité et sa diversité et de permettre ainsi la conservation d'une partie importante du patrimoine linguistique des espaces francophones du monde. Mais l'enquête aspire aussi à fournir des données qui iront au-delà d'une simple description. Il s'agit de proposer des analyses phonologiques de phénomènes typiquement variables et, ce faisant, de mettre à l'épreuve les modèles phonologiques et phonétiques faisant une large place à la variation. En automatisant le corpus, nous espérons favoriser les échanges entre les connaissances phonologiques et

les outils de traitement de la parole. Enfin, le corpus sera mis à la disposition de tous les intéressés car, comme le remarque Bergounioux (1992 : 52), "à quoi sert un corpus savamment constitué s'il n'est pas accessible ?".

3.1. L'enquête

PFC est un projet de *type socio-phonologique* qui vise à l'élaboration d'un grand corpus oral de plus de 350 heures d'enregistrements comprenant quatre registres différents et une ample couverture géographique. Dans le choix des points d'enquête, nous avons aussi inclus une dimension sociologique en choisissant, à côté de certaines villes, leur banlieue ouvrière et en tâchant d'équilibrer zones urbaines et zones rurales. Les points d'enquête en (1) sont soit achevés, soit en cours, alors que ceux en (2) sont programmés :

- (1) Lille, Béthune, Lyon, Villeurbanne, Paris-centre, Paris-banlieue est, Paris-banlieue ouest, Orléans, l'Orléanais, Rennes, Ile d'Yeu, Cherbourg, le Cotentin, le Perche, Caen, le Périgord, Bordeaux, la Corrèze rurale, Toulouse-ville, Toulouse-banlieue, Montpellier, Pézenas, Grenoble, Saint Etienne.
- (2) Strasbourg, l'Alsace rurale, Metz, la Lorraine rurale, Rouen, le Pays basque, Brives-la-Gaillarde, Nice, la Provence rurale, Marseille, Dijon.

A ce jour les points manquants sont l'Auvergne, le Poitou-Charente, la Champagne-Ardenne, la Picardie et la Corse. Puisque nous visons à une couverture aussi complète que possible des variétés de français, nous avons projeté de couvrir quelques points hors de France comme la Belgique, la Suisse, le Canada avec Québec Ville et Montréal, la Louisiane et le Maroc. Pour tous ces points d'enquête, le même protocole est utilisé par les enquêteurs qui sont choisis pour leur connaissance du terrain. Les enquêteurs travaillent dans la mesure du possible en équipe de deux, dont un chercheur connaît le/s témoin/s 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 (1972) dans certains de ses travaux sur le Lower East Side de New York et par les Milroy à Belfast (Milroy 1980). Cette méthode a l'avantage d'atténuer le "paradoxe de l'observateur" et de fournir relativement facilement l'accès à deux registres. L'enquêteur familier du témoin peut mener la conversation dite informelle pour accéder au vernaculaire de ce témoin, alors que l'autre enquêteur dirigera la conversation semi-formelle. 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 connaissance préalable des interviewés et sa participation au réseau dense constituent une condition *sine qua non* qui à elle seule a des effets d'atténuation du paradoxe de l'observateur.

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 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 du groupe est telle que si un sujet se laisse intimider par la vue du micro et adopte une langue plus standard mais éloignée de son vernaculaire, les autres membres du groupe 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 un texte et une liste de mots sans que cela ne devienne vite une tâche insurmontable. Cette exigence de scolarisation minimale exclut d'emblée un ensemble de milieux défavorisés, ce qui affaiblit évidemment la portée de la couverture sociale. Néanmoins, il est bien entendu que l'enquêteur s'efforcera d'obtenir à l'intérieur du groupe étudié une certaine différenciation sociale. En revanche, il est crucial que le groupe comporte un nombre égal d'hommes et de femmes, et idéalement le groupe sera composé de six hommes et six femmes répartis en quatre tranches d'âge : 20-30 ans ; 30-50 ans ; 50-60 ans ; 60 ans +. Pour certains points d'enquête, il peut être difficile de satisfaire à ces exigences, mais un minimum de deux tranches d'âge est requis.

Comme nous l'avons déjà souligné plus haut, 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 sont pas représentatifs, 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 néanmoins d'entreprendre des études sociolinguistiques plus fines et plus précises sur la variation phonologique.

3.2. Procédure

Les enregistrements sont effectués soit chez le témoin, soit dans un lieu neutre, comme à l'université, et comprennent dans tous les cas quatre volets sans qu'il y ait d'ordre prédéterminé pour les diverses tâches. Le témoin devra lire une liste de mots, un texte, participer à une conversation semi-formelle et à

une conversation informelle. L'enquêteur établit pour chaque sujet une fiche signalétique en dehors des entretiens et lui demande de **signer** un consentement de participation qui nous permet de disposer des **enregistrements** effectués à des fins de recherche ou à des fins pédagogiques.

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

Tableau 1. Liste de mots utilisée dans PFC

La liste de mots 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 littérature phonétique et des grandes enquêtes déjà réalisées. Ceci garantit une comparabilité de PFC avec les grandes enquêtes qui l'ont précédée et permet de donner aux analyses qui en découleront une profondeur historique certaine. Ceci permettra 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 aurait exigé une liste bien trop longue, cependant certaines tendances peuvent probablement être dégagées. En ce qui concerne les consonnes, nous avons été assez restrictifs en nous fixant comme objectif le statut de <gn>, la réalisation du <r> dans divers contextes, le groupe consonantique /ks/ et quelques assimilations de voisement. En fin de liste, nous reprenons quelques paires minimales afin de tester la conscience linguistique du locuteur, mais comme le montre la liste ci-dessus, qui est celle que nous présentons à tous nos témoins, l'ordre des mots dans le corps de la liste est aléatoire.

On demande au témoin de faire précéder la lecture de chaque mot du chiffre correspondant afin d'éloigner l'attention de la voyelle testée, mais aussi pour obtenir des indications supplémentaires sur la lecture des chiffres. Les résultats phonologiques obtenus sur la seule base de la liste de mots ne peuvent être considérés comme fiables. L'informateur lira sans doute comme on lui a appris à lire. La liste de mots, comparée à la grille des phonèmes établie indépendamment, donnera cependant des indications sur l'attitude de l'informateur par rapport à l'orthographe et à la norme. Les réactions verbales des enquêtés à cet exercice sont aussi une source précieuse quant à leur cantonnement sociolinguistique. D'autre part, même si le locuteur fait des distinctions artificielles, un ensemble de phénomènes échappe à sa vigilance. Dans nos enregistrements, nous avons ainsi remarqué pour de nombreux locuteurs le dévoisement de /r/ en position finale, une semi-vocalisation proche de la parole spontanée, la réduction de groupes consonantiques, des assimilations de voisements, y compris *slip* prononcé [zlip] chez certains locuteurs du Midi !

Le texte que nous soumettons à nos témoins a été établi avec soin en termes de composition lexicale et grammaticale. Il a la forme d'un petit article dans un journal régional et ne pose aucune difficulté de compréhension. Tous nos informateurs ont d'ailleurs souligné la simplicité de la tâche. Si le style est qualifié de 'mauvais' ou de très peu littéraire, nous y voyons non pas une critique mais plutôt un compliment. Nous avons voulu éviter un style littéraire ou même le style d'un bon journalisme. Ce texte a été construit afin d'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 nous ne pensons pas que ce reste culturel constitue un frein à la lecture de locuteurs belges ou canadiens, par exemple.

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 donc a 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, nous avons tenté d'inclure non seulement la majorité des segments vocaliques et consonantiques du français, mais aussi tout un ensemble de phénomènes phonologiques typiques du français. Tout comme la liste de mots dont il reprend certains mots et paires minimales, le texte permet d'approcher l'inventaire des phonèmes du locuteur. Il permet aussi de tester dans un premier temps deux phénomènes en détail : la prononciation du schwa et la liaison (y compris les phénomènes connexes de nasalisation et de 'h' aspiré). La lecture du texte ne fournit évidemment aucune donnée pour une étude de la parole spontanée, mais vu la diversité des phénomènes en jeu, l'informateur est beaucoup moins conscient de sa performance linguistique dans cet exercice. On obtient alors un style soutenu certes, mais dans certains cas un style moins surveillé, voire naturel.

Si dans un point d'enquête particulier il apparaît nécessaire de tester des phénomènes laissés pour compte ici, l'enquêteur a tout loisir de proposer des mots supplémentaires ou un autre texte. Dans les deux cas, ces mots et ce texte s'ajoutent à ceux proposés dans notre protocole. Au-delà de la lecture, chaque témoin participe à deux types d'interviews, une interview semi-formelle et une interview informelle.

Dans l'entretien semi-formel d'une durée de vingt à trente minutes, le locuteur est interrogé par exemple sur ses activités, son travail, ses projets,

son enfance, 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 n'est pas d'obtenir un ensemble de renseignements sur le locuteur, ceux-ci seront rassemblés à la fin de la séance dans le questionnaire, mais de 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 informelle 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. Le but de cette discussion 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 à l'analyse quatre registres bien définis.

3.3. Etudes phonologiques : quelques directions

Notre objectif est de remettre en question le mythe de l'homogénéité de la prononciation du français. La grammaire générative traditionnelle s'est attachée exclusivement à analyser le français dit "standard" et s'est le plus souvent basée sur des descriptions normatives comme Fouché (1959). Si notre compréhension des principaux phénomènes de la phonologie du français doit progresser, il est indispensable d'élargir la base empirique qui sert de support à la formulation de généralisations théoriques et, de ce fait, nous souscrivons à la recommandation de Séguy dans son article sur les *Atlas* :

les dialectologues doivent s'efforcer de s'intégrer dans la linguistique générale, tandis que les généralistes se feront un devoir d'enrichir et de consolider leurs spéculations en puisant dans le trésor de faits concrets que les chercheurs de terrain mettent à leur disposition. (Séguy 1973 : 88)

A court terme, nous nous proposons d'établir pour chaque point d'enquête l'inventaire phonologique de chaque locuteur et de survoler la question de la liaison et celle du schwa. Dans un premier temps, ces questions sont considérées en relation avec la liste de mots et le texte. Nous avons établi pour le texte une grille de liaison à remplir par l'enquêteur. Cette grille apparaît sous forme de tableau où sur l'axe vertical sont donnés tous les cas de liaisons possibles ou 'impossibles' avec leur catégorie (obligatoire, facultative, interdite),

alors que sur l'axe horizontal, on note l'apparition ou non de la liaison, s'il s'agit d'une liaison non enchaînée ou s'il y a hésitation. Les catégories sont reprises de la littérature phonologique classique sur le sujet (Delattre 1966, Encrevé 1988) et font office de point de référence. Sur la base des enquêtes menées à Grenoble, Caen et Québec, trois points géographiquement très distants, on constate un assez grand consensus dans l'usage de la liaison.

Si l'on reprend les trois catégories traditionnelles, nous observons que dans la catégorie dite 'facultative', les 'règles d'usage' ne sont pas adéquates, alors qu'il n'y a pratiquement pas de déviation par rapport à ces mêmes règles dans les deux autres catégories. Pour ce qui est des liaisons dites 'interdites', l'usage est conforme aux attentes et la liaison est absente dans :

comment / en plus
 dans le coin / on est jaloux
 quelques fanatiques / auraient
 vraiment / une étape nécessaire

Parallèlement, la catégorie dite 'obligatoire' semble relativement stable. Les écarts portent sur deux cas précis : le groupe *grand émoi* et l'expression *jeux olympiques*. Dans le premier groupe, tous les locuteurs canadiens et la plupart des locuteurs français font la liaison. On observe cependant que trois locuteurs omettent la liaison et un locuteur fait la liaison non pas avec la consonne dentale, mais avec la nasale [gränemwa]. Pour ce qui est de *jeux olympiques*, la lexicalisation avec liaison est généralisée chez les locuteurs de Grenoble et de Caen, mais pas chez les Canadiens interviewés.

La liaison dite 'facultative' ne correspond pas à ce que nous sommes en mesure d'attendre sur la base des descriptions dites classiques. Cette divergence par rapport à une certaine norme n'est pas géographiquement marquée et les observations concordent en tous points dans les trois villes examinées. Lors de la lecture du texte, exercice auquel nous avons attribué l'étiquette de registre formel et où, de ce fait, on prédit une recrudescence de liaisons, celle-ci n'apparaît que sporadiquement et exclusivement à l'intérieur d'un syntagme nominal, dans le contexte 'substantif pluriel-adjectif'. On peut entendre par exemple *les circuits-habituels, des visites-officielles, des pâtes-italiennes*, mais pas *toujours-autour*. Ces résultats ne sont pas surprenants si, comme le suggérait déjà Morin (1986), la liaison fonctionne comme un marqueur du pluriel. Cette généralisation ne rend pas compte de la liaison après la troisième personne de l'auxiliaire *être* au présent (*est*) : *Le village de Beaulieu est-en grand émoi, est-en revanche*. Cette liaison est qualifiée de facultative, mais ne devrait-on pas plutôt parler de liaison obligatoire, au moins à certains niveaux stylistiques ? Ces résultats confirment les observations de De Jong (1994) et suggèrent que les catégories traditionnelles demandent impérativement à être réexaminées.

Pour le traitement de schwa dans nos données, nous nous préoccupons de mettre au point un système de codage afin de relever de façon précise tous les environnements où un schwa peut être réalisé. Un tel système de codage a été

développé pour le texte, ce qui nous permettra de disposer de données d'une comparabilité fiable. L'ensemble des contextes est déterminé à partir de l'ouvrage de référence sur la question (Dell 1973/85) : schwa initial d'énoncé, schwa initial de mot syllabique, schwa interne, schwa final devant voyelle ou devant consonne. Le système de codage proposé permet uniquement un balayage général des données, il restera ensuite à préciser les différents contextes. Ce codage comprend quatre chiffres : le premier dénote la présence ou l'absence de la voyelle, le deuxième la position du schwa dans le mot (monosyllabe ou première, dernière ou autre de polysyllabe), le troisième l'environnement à gauche de la syllabe qui comprend le schwa (voyelle, consonne ou début de groupe intonatif) et le quatrième l'environnement qui succède à la syllabe avec schwa (voyelle, consonne, frontière intonative faible ou forte).

Au-delà de ces trois objectifs (inventaire phonologique de base, traitement de la liaison et du schwa), communs à l'ensemble des équipes impliquées dans le projet, chaque équipe s'investit dans des analyses plus fines : à Orléans, des études de changements linguistiques en temps réel sont entreprises qui comprennent une comparaison du système vocalique décrit par PFC au système vocalique établi avec l'ESLO, même si les mêmes témoins ne sont pas repris. Dans une analyse de ce type, le choix des locuteurs est déterminant puisque, comme nous l'avons vu, les locuteurs de l'ESLO n'étaient pas toujours géographiquement représentatifs. Durand et Lyche (1999) proposent dans le cadre de la théorie de l'optimalité une analyse des glissantes dans le français du Midi où, par exemple, une frontière morphologique bloque la synérèse dans une forme verbale alors que celle-ci est insensible au groupe consonantique composé d'une obstruante et d'une liquide. Dans les parlars étudiés, *scier* est toujours réalisé [sije] et *tablier* est prononcé [tablje] par certains locuteurs. Dans le domaine des glissantes, les premiers résultats obtenus semblent confirmer que dans le français du Nord, la synérèse augmente en accord avec une échelle de sonorité interne des voyelles hautes (Klein 1991 ; 1993). Sur la base des enregistrements faits à Grenoble et à Caen, on remarque que la tendance se dégage nettement aussi bien pour la liste de mots que pour la lecture, et la synérèse est plus fréquente avec *nier* qu'avec *louer* et qu'avec *tuer*.

4. Conclusion : un corpus informatisé

Les interviews une fois terminées sont transcrites orthographiquement. Comme dans le cas de VALIBEL et d'autres corpus oraux, nous avons en effet opté pour une transcription orthographique (et non pas phonétique / phonémique) des données. Ceci permet d'alléger la charge de formatage du rendu qui, s'agissant d'une enquête aussi importante que PFC, est déjà considérable. D'autre part, on sait que la transcription orthographique constitue l'un des

accès les plus aisés aux corpus oraux et à leur manipulation. Elle permet également un balisage et une circulation simple dans ces corpus, de même qu'elle assure une comparabilité maximale entre les diverses interviews du corpus comme avec d'autres corpus oraux. Les transcriptions phonétiques ou phonémiques, utiles pour des études fines et ponctuelles, n'offrent pas du tout ces avantages et, s'agissant d'une enquête mettant en jeu un grand nombre d'enquêteurs et de transcripteurs différents, favorisent les erreurs et les biais multiples.

Ces transcriptions orthographiques seront alignées sur l'enregistrement sonore à l'aide de logiciels d'alignement multilocuteurs et multilinéaires (PRAAT). Ceci nous permettra de constituer une base sonore totalement numérisée de la phonologie du français contemporain avec un alignement fin texte / parole. Des transcriptions phonétiques ponctuelles correspondant aux grands locus de la phonologie du français (*e* muet, liaison, nasales, glissantes) seront également proposées. En même temps, des outils d'indexation précise et de recherche rapide seront fournis. Le corpus sonore sera entièrement numérisé et découpé en fichiers de taille réduite, donc manipulables, offrant ainsi un ensemble d'échantillons de parole spontanée. L'ensemble du corpus, des transcriptions, des alignements et des outils d'indexation, de circulation et de recherche rapide multicritères sera édité sous forme de CD-ROM. En dehors des analyses phonologiques proposées par les phonologues partie prenante au projet PFC, le corpus sera ainsi mis à disposition de l'ensemble de la communauté scientifique, offrant, nous l'espérons, une base empirique et descriptive utile aux études portant sur la phonologie du français contemporain dans la réalité de ses usages.

Bibliographie

- Agren, John 1973 : *Enquête sur quelques liaisons facultatives dans le français de conversation radiophonique* (= Studia Romanica Upsaliensia ; 10). Uppsala : Université d'Uppsala.
- Bergounioux, Gabriel (éd.) 1992 : *Enquêtes, corpus et témoins en France hier et aujourd'hui* (= Langue française 93). Paris : Larousse.
- Bourdieu, Pierre 1993 : *La Misère du monde*. Paris : Seuil.
- Coquebert de Montbret 1807-1812 : *Recueil des Paraboles et de la correspondance* (Paris, Bibliothèque Nationale, NAF, 5910 à 5913 et Archives Nationales, F17A 1209).
- Dausès, August 1973 : *Etudes sur l'"e" instable dans le français familier* (= Beihefte zur ZRPh ; 135). Tübingen : Niemeyer.
- De Jong, Daan 1994 : Sociophonologie de la liaison orléanaise ; in : Lyche, Chantal (éd.) : *Generative Phonology: Retrospective and Perspectives*. Middlesex : AFLS / ESRI, 95-130.
- Delattre, Pierre 1966 : *Studies in French and Comparative Phonetics*. La Haye : Mouton.
- Dell, François 1985 (1973) : *Les Règles et les sons*. Paris : Hermann.

- Durand, Jacques / Lyche, Chantal 1999 : Regard sur les glissantes en français, français standard, français du Midi. *Cahiers de Grammaire* 24, 39-65.
- / Slater, Catherine / Wise, Hilary 1987 : Observations on schwa in southern French. *Linguistics* 25, 983-1004.
- Encrevé, Pierre 1988 : *La liaison avec ou sans enchaînement. Phonologie tridimensionnelle et usages du français*. Paris : Seuil.
- Fouché, Pierre 1959 : *Traité de prononciation française*. 2e éd. Paris : Klincksieck.
- Gillieron, Jules / Edmont, Edmond 1902-1910 : *Atlas linguistique de la France (ALF)*. 9 vol. Paris : Champion.
- Gougenheim, Georges *et al.* 1956 : *L'élaboration du Français Elémentaire*. Paris : Didier.
- Hansen, Anita Berit 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. Copenhague : Université de Copenhague (mémoire de maîtrise).
- Klein, Marc 1991 : Vers une approche substantielle et dynamique de la constituance syllabique. Le cas des semi-voyelles et des voyelles hautes dans les usages parisiens. Saint Denis : Université de Paris VIII (thèse de doctorat).
- 1993 : La syllabe comme interface de la production et de la réception phoniques ; in : Laks, Bernard / Plénat, Marc (éds.) : *De Natura Sonorum. Essais de phonologie*. Saint Denis : Presses Universitaires de Vincennes, 101-141.
- Labov, William 1972 : *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia : University of Philadelphia Press.
- Laks, Bernard 1977 : Contribution empirique à l'analyse socio-différentielle de la chute des /r/ dans les groupes consonantiques finals. *Langue française* 34, 109-125.
- 1980 : Différenciation linguistique et différenciation sociale : quelques problèmes de sociolinguistique française. Saint Denis / Vincennes : Université de Paris VIII-Vincennes (thèse de doctorat).
- 1983 : Langage et pratiques sociales. Etude sociolinguistique d'un groupe d'adolescents. *Actes de la recherche en sciences sociales* 46, 73-97.
- Martinet, André 1945 : *La prononciation du français contemporain*. Genève : Droz.
- / Walter, Henriette 1973 : *Dictionnaire de la prononciation du français dans son usage réel*. Paris : France-Expansion.
- Milroy, Lesley 1980 : *Language and Social Networks*. Oxford : Blackwell.
- Morin, Yves-Charles 1986 : On the morphologization of word-final consonant deletion in French ; in : Anderson, Henning (éd.) : *Sandhi Phenomena in the Languages of Europe*. Berlin : Mouton de Gruyter, 167-210.
- 2000 : Le français de référence et les normes de prononciation ; in : Francard, Michel *et al.* (éds.) : *Actes du colloque 'Les français de référence' de Louvain-la-Neuve, 3-5 novembre 1999* (= Cahiers de Linguistique de Louvain ; 26), 91-135.
- Séguy, Jean 1973 : Les Atlas linguistiques de la France par région. *Langue française* 18, 65-90.
- Walter, Henriette 1982 : *Enquête phonologique et variétés régionales du français*. Paris : PUF.