



HAL
open science

La constitution d'un corpus, parcours initiatique en linguistique

Mariette Meunier

► **To cite this version:**

Mariette Meunier. La constitution d'un corpus, parcours initiatique en linguistique. 2008. halshs-00359903

HAL Id: halshs-00359903

<https://shs.hal.science/halshs-00359903>

Preprint submitted on 9 Feb 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mariette Meunier-Crespo
CEL, Université Jean-Moulin Lyon III

LA CONSTITUTION D'UN CORPUS ORAL, PARCOURS INITIATIQUE EN LINGUISTIQUE

Une longue tradition pour les dialectologues

Les dialectologues n'ont jamais prisé les longues considérations théoriques :

« La pratique de l'enquête ne s'apprend pas dans les livres. Les bons conseils et tous les discours des linguistes en chambre ne remplacent pas la relation unique, qui s'établit entre enquêteur et informateur. (...) A la lecture des pages introductives des atlas, pour ceux qui en comportent, il est très vite évident que la méthode suivie par les enquêteurs est souvent personnelle, spécifique, sans doute rendue nécessaire par les conditions de l'époque et des lieux. »¹.

Aux origines de la dialectologie, aller sur le terrain pour collecter des patois, c'était se démarquer de l'empire de la théorie, en l'occurrence des néogrammairiens et des lois phonétiques. Cette tradition s'est transmise, affinée, a élaboré ses propres méthodes au fil des réalisations des Atlas linguistiques et ethnographiques de la France. Sever Popp déclare en 1950 que « les futurs chercheurs [doivent] s'apercevoir par eux-mêmes des nombreux problèmes que soulève un examen soigné des parlans vivants. »² Cette longue tradition d'autonomie des chercheurs les a éloignés de la scène structuraliste des années soixante. En revanche, le champ de réflexion actuel que constitue la linguistique de corpus leur est familier.

La linguistique de corpus est actuellement un domaine de recherche en plein essor, avec des méthodologies très variées de constitution,

¹ Brasseur Patrice, 2005, « Enquêteur et enquêtés, l'enquête dialectologique en Normandie », in : *Mélanges offerts au Professeur Lothar Wolf*, « Je parle, donc je suis de quelque part ». Textes rassemblés et édités par Brigitte Horiot, Elmar Schafroth, Marie-Rose Simoni-Aurembou, Lyon 2005, CEL, hors série n°2, p.67 et 69.

² Popp Sever, 1950, *la dialectologie, première partie, dialectologie romane*, Université de Louvain.

d'exploitation et de conservation des corpus. Nombreux sont les travaux en cours pour transformer des archives, des enregistrements ou des bases de données en corpus le plus souvent informatisés en vue de leur exploitation linguistique avec des logiciels et selon des objectifs variés. Les dialectologues n'ont pas tardé à se mettre aux technologies de l'information pour la constitution et l'utilisation de leurs corpus³. En ce qui concerne les corpus oraux, les bases de données linguistiques répertorient en général toutes sortes de matériaux oraux : conversations, entretiens sociologiques, réponses de locuteurs à des enquêtes d'opinion, enquêtes dialectales, ethnotextes, etc. Ces matériaux se présentent sous forme d'enregistrements de données sonores, parfois accompagnées de données visuelles. L'ensemble de ces travaux a en commun la réflexion méthodologique initiée par la constatation toute simple pour l'étude de l'oral, que l'objet de la linguistique est empirique et consiste en *textes* oraux et non pas en *mots* ou *phrases*, qui ne s'observent pas à l'état brut⁴.

Le CNRS a publié un « *guide des bonnes pratiques* » qui dresse l'état des lieux en matière de corpus oraux, et aborde les questions techniques, méthodologiques, juridiques et éthiques liées à leur constitution⁵. L'objectif est d'offrir à tout chercheur les outils, les « bonnes pratiques » qui lui permettront de « [transformer] des productions verbales en un corpus oral, susceptible d'être étudié et conservé, et par conséquent de prendre place dans le patrimoine culturel de la nation »⁶.

Cette visée patrimoniale, également présente dans la *Charte européenne des langues régionales ou minoritaires* de 1992⁷, nous incite à

³ cf sur : <http://www.unice.fr/bcl/presente.php> le « Thesaurus occitan » (THESOC), élaboré par le laboratoire *Bases, Corpus, Langages* du CNRS à l'Université de Nice, qui propose une base de données de textes dialectaux oraux multimedia. Dans le cadre de cette base de données, des travaux sont en cours pour l'établissement d'un corpus de textes occitans destiné à l'étude syntaxique.

⁴ C'est la revue *Langages* qui initie en 1969 avec le n°13 sur « l'analyse du discours » la réflexion moderne sur les corpus, leur constitution, leur usage. Cf. également François Rastier, 2002, « enjeux épistémologiques de la linguistique de corpus », conférence aux deuxièmes journées de Linguistique de Corpus, Lorient, sur : http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html

⁵ *Corpus oraux, guide des bonnes pratiques*, 2006, CNRS.

⁶ Op. cit. p. 11

⁷ Préambule de la *Charte*, alinéa 3, « la protection des langues régionales ou minoritaires historiques de l'Europe, dont certaines risquent, au fil du temps, de disparaître, contribue à maintenir et à développer les traditions et la richesse culturelles de l'Europe. » Il s'agit de « conforter un patrimoine, national et européen, dans sa

penser, au CEL, qu'il existe encore un champ de connaissance auquel la méthode dialectologique puisse s'appliquer, ce qui signifie qu'il faut encore « revenir sur le terrain quand c'est encore possible pour éclairer les points que les données disponibles laissent dans l'ombre »⁸. Plus d'un siècle après les propos de l'abbé Rousselot : « Les patois disparaissent et il est déjà tard pour les recueillir. Chaque année qui s'écoule emporte avec elle des sons, des constructions, des mots dont la perte est irréparable »⁹, les dialectologues ont donc encore et toujours à sauver ce qui reste des dialectes romans¹⁰ et ils doivent encore et toujours collecter des données pour constituer des corpus oraux. Selon la formule plaisante de Jean Le Dû « La dialectologie et la géolinguistique bougent encore¹¹ » !

Elles bougent d'autant plus à Lyon que, comme chacun sait, « c'est de Lyon qu'est venu en France, après plus d'un quart de siècle de silence, de stagnation, le renouveau de la dialectologie : renouveau du questionnaire,

diversité et sa richesse. Ce qui invite la République française à reconnaître les langues de la France, comme éléments du patrimoine culturel national ».

Cf. sur http://www.culture.gouv.fr/culture/dglf/lang-reg/rapport_cerquiglini/langues-france.html Bernard Cerquiglini, Avril 1999, « Les langues de la France », *Rapport au Ministre de l'Education Nationale, de la Recherche et de la Technologie, et à la Ministre de la Culture et de la Communication*.

Cf. également dans la perspective ouverte par le rapport Cerquiglini : <http://languesdefrance.nuxit.net/Colloque%20Paris/table-ronde.html> le Projet Histoire Sociale des Langues de France.

⁸ Le Du Jean Jean Le Dû et Yves Le Berre, 1991, « Où va la dialectologie... Quand on la voit passer à Brest ? », *Quaderni di Semantica*, a.XII, n° 2, 262-271.

http://www.webzinemaker.com/admi/m7/page.php3?num_web=7572&rubr=3&id=42594

⁹ Abbé Rousselot, 1887, « L'étude des patois », in : *Revue des patois gallo-romans* Tome I, Paris, Neuchâtel, p.3, et sur : <http://gallica.bnf.fr/ark:/12148/bpt6k929839>. (*Gallica*, bibliothèque numérique de la bibliothèque nationale de France).

¹⁰ Par exemple, « Les parlers normands de Jersey, Guernesey et Sercq se maintiennent, mais le nombre de locuteurs diminue chaque année et les autochtones de moins de quarante ans n'ont plus, au mieux, qu'une connaissance passive du dialecte. Ils le comprennent mais sont incapables de le parler. Ceci est d'autant plus à regretter que ces parlers restent magnifiquement conservés dans leur état ancien. Et ceux qui les emploient en connaissent encore souvent bien des finesses. Mais si l'illusion demeure encore pour quelques décennies, leur mort sera brutale. » Brasseur Patrice, 1990, *Le parler normand, mots et expressions du terroir*, Paris, Rivages, p.9.

¹¹ Le Dû Jean 2006, *La Bretagne Linguistique n°13*, "Dialectologie et Géolinguistique", Sous la direction de Jean Le Dû et Nelly Blanchard, Centre de Recherche Bretonne et Celtique. sur : <http://perso.orange.fr/fanch.broudic/articles/Diatopique.html>

de la méthode d'enquête, des interprétations des matériaux relevés¹² ». Les dialectologues de Lyon, comme ceux de Nice, Toulouse ou Brest, ne pensent pas que la dialectologie aurait « définitivement épuisé sa fécondité cognitive »¹³.

Il est rappelé dans la préface de l'ouvrage du CNRS qu'« un corpus oral n'est pas une simple collection d'enregistrements de la parole humaine, c'est un objet construit ». En effet, les travaux en linguistique de corpus ont totalement dissipé l'illusion qui consiste à croire à un partage en deux temps de l'activité du linguiste : le temps de la pratique sur le terrain, c'est-à-dire de la collecte de matériaux purement empiriques, et celui de la mise en cohérence théorique. Le tort de cette vision naïve est de conduire d'une part à une survalorisation de la théorie et à une surinterprétation des faits, et d'autre part à une occultation totale du travail sur le terrain perçu comme une boîte noire. Il suffit de confronter un étudiant à l'établissement d'un corpus dialectal, si modeste soit-il, pour constater son désarroi, ses interrogations sur la façon de procéder. On s'avise ainsi très concrètement que la dichotomie entre linguistique de terrain et linguistique théorique n'est qu'opératoire, comme toutes les dichotomies. Les interrogations sur la méthode et la théorie sous-jacente sont nombreuses dès le départ de la recherche et la paralysent tant qu'une réponse claire n'est pas apportée. « l'empirie et la théorie sont consubstantielles »¹⁴.

Le chercheur doit – et ces contraintes sont l'autre visage de son autonomie – en construisant son corpus, déclarer ses choix théoriques, expliciter ses méthodes, ses partis pris, faire état de ses doutes, se situer par rapport au foisonnement des théories et des pratiques contemporaines. Il en résulte qu'en construisant son corpus, le linguiste se construit aussi en tant que tel. Par conséquent, le parcours initiatique dont nous parlons ici ne fait pas référence à une initiation superficielle, renvoyant à plus tard les occupations sérieuses de la théorie. C'est une prise de conscience que le travail de terrain, car travail il y a, est devenu de bout en bout théorique.

Les pages qui suivent ne sont pas une histoire des méthodes d'enquête de la dialectologie. Elles proposent seulement de retracer brièvement les étapes de ce parcours initiatique en linguistique à l'échelle du chercheur individuel et à l'usage de nos apprentis dialectologues au CEL. Nous

¹² Straka Georges, 1997, « Discours d'ouverture », in : *Les Cahiers du CEL*, n°1, p.14.

¹³ Le Dû Jean, 1991.

¹⁴ Texte de présentation du Laboratoire *Bases Corpus Langage*, université de Nice, cité note 2.

donnerons l'exemple concret de notre expérience en la matière : dans le cadre d'un travail de constitution et de transcription manuelle d'un corpus dialectologique oral, nous avons été amenée à réfléchir sur la dialectique entre le donné et le construit¹⁵. Cette étude a suivi non pas la lettre, mais l'esprit des travaux des dialectologues, en particulier par la conscience de la dialectalité de la langue et l'autonomie de conception et de réalisation des travaux.

Nous ferons ici un arrêt sur trois étapes de ce parcours, qui sont autant de moments de choix méthodologiques, et que nous nommons les trois paradoxes de l'observateur. Sans les avoir expressément nommés et sans les avoir théorisés, les dialectologues se sont confrontés à eux de longue date. Loin de tout esprit de système, les dialectologues s'interrogent depuis longtemps sur la méthodologie de leur discipline. Ils ont eu le souci dès la fin du XIX^e siècle, de délimiter soigneusement leur objet d'étude, de donner des instructions d'enquête, et d'établir des systèmes de transcription. Ils n'ont jamais prétendu à l'objectivité absolue, mais simplement à la cohérence des données, grâce à l'unité de conception, d'orientation et de dépouillement des enquêtes, ce qui est une attitude étonnamment moderne. Jamais un dialectologue n'est tombé dans le travers de survaloriser une théorie ou sur interpréter les faits. Tous ou presque ont souligné dans leurs introductions les limites du corpus recueilli et la nécessité de poursuivre ultérieurement les enquêtes. C'est ainsi qu'ils ont formulé diversement, dans des termes qui ont évolué avec le temps, et de plus en plus nettement, les conditions de la métamorphose des matériaux bruts en objets d'étude. Un siècle plus tard, Jean Philippe Dalbera rappelle ce qui vaut pour les dialectologues comme pour toute la linguistique de corpus, que ce dernier « ne saurait être qu'un construit et sa construction fait partie intégrante du prisme théorique à travers lequel le linguiste entend appréhender le réel. »¹⁶.

Le premier paradoxe de l'observateur : « rechercher et délimiter l'objet d'enquête ».

¹⁵ Mariette Meunier-Crespo, 1994, *Des données empiriques à l'objet d'étude, la constitution d'un corpus en démodialectologie, le français populaire d'Abidjan*, 2 volumes, Lyon, CEL.

¹⁶ Jean Philippe Dalbera, 2002, « le corpus entre données, analyse et théorie » *corpus* n°1, laboratoire *Bases Corpus Langage*, <http://corpus.revues.org/document.html?id=10>

Comme le dit Marcel Cohen au début de ses *Instructions d'enquête linguistique*, « la première chose à faire est de rechercher et délimiter l'objet d'enquête. »¹⁷ Nous avons qualifié de premier paradoxe de l'observateur le premier travail de définition de l'objet d'étude. Il s'agit d'isoler la variété de langue à étudier, alors que cette dernière ne se présente jamais à l'état pur. Ce paradoxe est pensé par Saussure. On connaît la formulation tant citée :

Bien loin que l'objet précède le point de vue, on dirait que c'est le point de vue qui crée l'objet, et d'ailleurs rien ne nous dit d'avance que l'une de ces manières de considérer le fait en question soit antérieure ou supérieure aux autres.¹⁸

Ce premier paradoxe concerne au premier chef les dialectologues car en inaugurant à la fin du XIX^e siècle les enquêtes sur le terrain pour l'étude des patois, ils ont été les premiers linguistes à être confrontés aux données réelles brutes, c'est-à-dire à ce qu'ils concevaient comme l'infinie diversité des parlers qu'ils voulaient collecter. Ce matériau oral contenait bien entendu, hier comme aujourd'hui et avant qu'elles ne soient répertoriées par la sociolinguistique variationniste et interactionnelle, les différentes formes que revêt la variation linguistique : la variation diachronique, diatopique, diastratique et diaphasique.

Il faut se représenter les difficultés méthodologiques d'une science à ses débuts, non encore rompue aux distinctions opératoires, et confrontée aux notions non encore définies de dialecte et patois, qui pour les linguistes (mais toujours pas pour le grand public) sont parfaitement claires : « Le patois est une variété linguistique de la plus petite communauté locale, c'est-à-dire en France de la commune » ; « le dialecte existe dans toute la région, mais il ne se réalise parfaitement nulle part ; le dialecte se réalise dans tous les patois de la région ».¹⁹

¹⁷.Cohen Marcel, 1950, *Instructions d'enquête linguistique*, Institut d'ethnologie de l'Université de Paris, p.15.

¹⁸ Ferdinand de Saussure, réédition 1972, *Cours de linguistique générale*, Paris, Payot, p.23.

¹⁹ Tuailon Gaston, 1982, « Qu'est-ce qu'un dialecte ? Qu'est-ce que le dialecte lyonnais ? » in : *Dialectologie lyonnaise*, Lyon, Institut Pierre Gardette.

Cependant, le point de vue de l'observateur est là, qui cherche à isoler les patois pour mieux les recueillir : L'abbé Rousselot signale d'emblée qu'il n'est pas simple d'isoler l'objet patois :

Il faut savoir dans quelle mesure ceux que l'on observe parlent patois. Les personnes qui parlent uniquement patois deviennent de plus en plus rares. Il n'y a guère qu'elles pourtant qui soient des témoins absolument sûrs. Une personne qui parle ordinairement français, bien que le patois ait été sa langue maternelle, pourra fournir des renseignements exacts sur les mots ; mais elle ne méritera qu'une confiance limitée pour les constructions et la syntaxe.²⁰

Marcel Cohen dans ses *Instructions d'enquête linguistique*, tient, plus de cinquante ans après l'abbé Rousselot, des propos semblables, assortis d'une formulation plus explicite de la variation linguistique :

Tout le monde parle-t-il de la même manière. Y a-t-il des différences entre les générations, parle-t-on dans une même famille des langues différentes ou tout le monde emploie-t-il le même langage avec quelques différences ? (...) Y a-t-il des langages de classes ou de castes ? En particulier un langage de hors castes ? Comment de réalise l'intercompréhension avec le reste de la population ?²¹

Nous avons bien entendu été confrontée à la même nécessité de ce travail de définitions pour l'établissement d'un point de vue lorsque nous avons entrepris d'étudier ce que nous pensons être une variété dialectale du français : le « français populaire d'Abidjan » (FPA). La dénomination de français populaire n'est qu'un pis aller, consacré par l'usage. Dans le cadre de leurs études et avec le point de vue qui est le leur, les créolistes proposent les dénominations « français zéro »²² ou « variété véhiculaire »²³. Pour notre part, nous avons vite constaté que ce parler ne relève pas de la catégorie expéditive de « petit nègre », ni de celles de patois, sabir, pidgin

²⁰ Abbé Rousselot 1887, p.20

²¹ Marcel Cohen, 1950 p. 17.

²² Pour une définition de la notion de français zéro et une comparaison avec des notions connexes, se reporter aux travaux de Robert Chaudenson et en particulier aux articles : « acrolecte », « basilecte », « français marginaux », « langue zéro » « mésolecte », « pidgin », « créole » in M.L. Moreau 1998, *Sociolinguistique. Les concepts de base*, Paris, Mardaga.

²³ Gabriel Manessy, 1995, *Créoles, pidgins, variétés véhiculaires. Procès et genèse*, Paris, CNRS.

ou créole. Le FPA nous est apparu comme le parler commun aux ivoiriens et à tous les migrants africains qui se retrouvent dans le creuset linguistique qu'est Abidjan. Nous avons choisi de le dénommer provisoirement « démodialecte ».

Un démodialecte est défini par sa modalité d'acquisition, qui détermine sa communauté de locuteurs : il s'acquiert par imprégnation, dans le cadre de vie habituel, hors de l'influence de l'école et des médias. On voit que les notions de langue maternelle et langue populaire ne recouvrent pas la notion de démodialecte. Le FPA est un français véhiculaire d'Afrique, qui présente ailleurs qu'à Abidjan des traits semblables et d'identiques tendances évolutives. Nous avons tenté de définir la notion de démodialecte par rapport à celles qui existent déjà : les dialectes historiques, les sociolectes, les chronolectes et les argots. Notre souci était de ne pas laisser tomber le terme du haut de son ciel théorique, attitude incompatible avec l'intention de travailler sur corpus, c'est-à-dire de sauver avant tout les phénomènes. Puis nous avons émis l'hypothèse de départ que le FPA est un démodialecte de français.

Constituer un corpus de langue orale et « sauver les phénomènes » suppose que l'on se représente la langue réellement parlée comme étant d'essence dialectale. Recueillir les données du réel et établir des statistiques d'occurrences de formes linguistiques, c'est chercher à connaître des modalités d'un fonctionnement. C'est donc concevoir que la langue a du « jeu », qu'elle tolère des fonctionnements majoritaires et minoritaires et des dysfonctionnements qui ne rentrent dans aucun des cas précédents. Ces dysfonctionnements sont des propositions discursives atypiques qui peuvent être soit des sédimentations d'états antérieurs soit les signes avant-coureurs d'une évolution future. Ils peuvent être aussi des approximations dues à des stratégies d'acquisition du parler, de la part de migrants dont il n'est pas la langue maternelle.

L'image du système linguistique est englobée dans une autre image, celle du diasystème. Ce dernier, vu en coupe synchronique, serait un bricolage qui présente suffisamment de jeu pour permettre l'intercompréhension orale et écrite entre locuteurs d'une langue de grande communication. Il permet aux différents parlers et langages complexifiés de coexister. Cette représentation de la langue comme diasystème, intégrant tous les phénomènes de variation, est inhérente au travail des dialectologues.

Le gros inconvénient de proposer le néologisme démodialecte est de contribuer à l'inflation terminologique que nous redoutons par ailleurs.

« Ivan Ivanovitch », comme le proposait Jakobson, était une autre appellation à plus grande allure, mais peu évocatrice pour une hypothèse de travail et plus lourde à manier ! Le seul avantage que nous espérons retirer du terme démodialecte est d'ordre pédagogique : montrer concrètement, dans la lignée des dialectologues, qu'il n'y a pas des pans de vérité intrinsèque qu'un chercheur inspiré dévoilerait tout à coup. Il n'y a que la matérialité des corpus établis pour vérifier des hypothèses. Le chercheur propose son point de vue puis établit son dispositif d'observation pour approcher au plus près son objet dans le réel.

Le deuxième paradoxe de l'observateur : le dispositif d'observation :

Après avoir établi son point de vue et son hypothèse sur un parler, après donc avoir réduit conceptuellement la variation de la parole²⁴, on passe à l'étape suivante : faire exister son objet d'étude dans un corpus. C'est le deuxième paradoxe de l'observateur, qui installe son dispositif d'observation par lequel il filtre concrètement l'oral pour en réduire l'hétérogénéité. Il s'agit rien moins que d'établir les conditions de validité du futur corpus. Ce dernier n'existe pour l'instant que dans l'esprit du chercheur et uniquement parce qu'il lui servira à vérifier son hypothèse.

Revenons au texte de l'abbé Rousselot. On y voit déjà apparaître la préoccupation constante des dialectologues pour les conditions de validité de la collecte des données. Les différentes modalités de constitution des corpus sont déjà là : l'idéal de la conversation libre, le questionnaire, l'ethnotexte :

Il y a plusieurs manières de recueillir les mots d'un patois. Toutes n'ont pas la même valeur. Mais souvent on n'a pas le choix. La meilleure, c'est le tête à tête avec des parents ou des amis. Grâce au laisser-aller de la conversation, on peut faire les observations les plus profondes, recueillir les faits les plus curieux, pénétrer dans les secrets de la syntaxe (...) On peut demander des traits de chronique locale, des contes, des dictons, des proverbes, le nom des objets que l'on a sous les yeux.²⁵

La plus grande importance est accordée bien sûr au choix de l'informateur :

²⁴ Ce n'est pas le lieu de discuter la dichotomie langue parole. Nous employons ici le terme « parole » comme synonyme de production orale, « parler ».

²⁵ Abbé Rousselot, 1887, p. 18-19.

« Les renseignements que l'on fournit sur un patois n'obtiennent leur pleine valeur que si l'on fait connaître le lieu d'origine, l'âge, la condition, les habitudes de langage, les antécédents du sujet observé »²⁶

On trouve là en germe la méthodologie développée par les enquêtes linguistiques effectuées pour la confection des Atlas linguistiques et ethnographiques de la France, réalisés à partir des années 50²⁷.

Le deuxième paradoxe de l'observateur apparaît en filigrane :

Quelque moyen que l'on emploie, il ne faut pas oublier qu'il y a une différence à établir entre les faits que l'on a provoqués et ceux qui se sont produits spontanément.²⁸

C'est Sever Popp²⁹ qui formule le plus nettement, ce deuxième paradoxe : « Les données dialectales sont toujours conditionnées – je ne cesse de le répéter, par le procédé d'enquête. »

En 1960 ; Emmanuel Company³⁰ résume le dispositif d'observation adopté par les enquêteurs des Atlas. Les différentes façons d'interroger les informateurs sont :

« 1 – provoquer des conversations ou des textes libres qu'on enregistre, le dialectologue extrayant ensuite de ses enregistrements les réponses de son questionnaire.

2 – la traduction pure et simple

3 – la question indirecte

4 – la désignation des objets et des actions, ou bien de leurs images photographiques ou dessinées ».

Enfin, Jean Philippe Dalbera a fait le tour de la question du corpus en dialectologie. A propos du protocole d'enquête dialectologique et des

²⁶ *Ibidem* p. 20. On pense au profil des informateurs que les anglo-saxons appellent NORM : *Non educated, Old, Rural, Male*.

²⁷ Certains e'entre eux comportent un tome entier consacré à la méthodologie : Nauton Pierre, *Atlas Linguistique et Ethnographique du Massif Central, vol. IV, Exposé général, table, questionnaire, index alphabétique*, CNRS 1963.

Gardette Pierre, *Atlas Linguistique et Ethnographique du Lyonnais, vol V, Commentaires et index*. CNRS.1976.

²⁸ Rousselot 1887, p. 19.

²⁹ Popp, Sever, 1950.

³⁰ Company Emmanuel, 1960, « A propos des questionnaires dialectologiques », *Premier Congrès international de dialectologie générale*, Louvain 1960, publ. Par A.J. wan Windikens, 1964-1965.

questionnaires, il présente une vision organique à laquelle nous souscrivons totalement :

De l'exécution d'un questionnaire qui se voulait standard au départ, le travail d'enquête, au fur et à mesure que les faits se dévoilent, s'adapte et s'approfondit là où cela en vaut la peine et le produit résultant, ce que nous nommons le « responsable », peut n'avoir plus qu'une relation lointaine avec le questionnaire initial qui a servi de base. Il en va d'un protocole d'enquête dialectologique comme du choix et de l'appropriation progressive d'une tenue vestimentaire. On se voit dans certains cas affublé d'un vêtement anonyme, passe-partout (uniforme, unisexe et taille unique) comme l'enquêteur se voit confier un questionnaire standard ; puis celui-ci devient à mesure de la progression de l'enquête et de l'analyse, l'équivalent d'un « prêt à porter » pour finir comme un « sur mesure » bien ajusté. Le questionnement mécanique conçu a priori se corrige, se complète et se peaufine tout au long de l'enquête et, par suite le corpus de réponses auxquelles il donne lieu et, en définitive, l'échantillon même qui en est issu.³¹

Nous nous sommes démarquée de la méthodologie des dialectologues et avons pour notre part choisi d'enregistrer onze heures de conversations « libres » avec des locuteurs de FPA. Notre corpus est un cas extrême, aux franges de la francophonie, menacé par les parlers africains. Nous avons enregistré pendant dix heures des boys, des cuisiniers, des vendeurs de cigarettes dans la rue, des gardiens de nuit, et nous avons ajouté une heure d'émissions de radio où étaient interviewés des cordonniers et des chauffeurs de taxi. Notre procédé avait le mérite d'être uniforme et sans équivoque. Refusant le procédé du micro caché, que nous trouvons contestable, nous avons toujours orienté comme nous l'avons pu³² nos conversations avec nos locuteurs, chacun pris séparément, vers le récit de vie, avec le magnétophone bien en vue. Nos locuteurs ne pouvaient donc utiliser que le FPA, n'ayant que nous-même comme interlocutrice.

Patrice Brasseur, évoquant les conditions concrètes des enquêtes dialectologiques, dit avec nostalgie ce que confirment tous les dialectologues : « Ces années passées sur le terrain constituent une expérience inoubliable pour nous tous, mais datée et sans doute devenue intransmissible »³³.

³¹ Dalbera Jean Philippe, 2002.

³² « Il faut un art bien subtil, une diplomatie bien exercée, pour amener sans violence les formes que l'on recherche. » Abbé Rousselot, 1887.

³³ Brasseur Patrice, 2005, p. 75

En hommage au travail des dialectologues, à leur magnifique moisson de données, à leur expérience tout simplement humaine, nous leur adressons ce premier message : le terrain continue à être une expérience inoubliable, parfois difficile, et immanquablement parsemée d'épisodes saugrenus. Par exemple, nous avons dû participer aux psalmodies de Karim, sur fond de coassements de crapauds-buffles et de fumées de braseros, car pour notre locuteur, veilleur de nuit, l'heure de la prière était arrivée. Contrairement à nombre d'enquêteurs des Atlas, nous n'avons jamais eu affaire à la maréchaussée, ni à la sourde hostilité de l'entourage des informateurs, ni à des comportements étranges de la part de ces derniers. Cependant, le terrain n'est pas une préface romancée au chapitre de la linguistique. Les dialectologues sont restés trop discrets pour ce qui est de la valeur de leur travail, en disant que ce dernier ne saurait être formalisé, qu'il ne se transmet pas, qu'il s'éprouve.

Nous adressons donc aux dialectologues un deuxième message : nous nous permettons de les inciter à écrire un livre à plusieurs voix sur leur expérience du terrain, leurs « dispositifs d'observation ». Ce qui en premier lieu nous procurerait le récit d'aventures souvent désopilantes, toujours passionnantes, nous ferait voir du pays et entendre des mots, nous distrairait un moment (ou pour toujours ?) de l'étude des concepts, progiciels et autres modélisations. On y constaterait que la particularité de chaque terrain actualise un travail de questions et de décisions, ainsi qu'une stratégie d'enquête. Loin des déclarations des « linguistes en chambre », on verrait comment se construit une unité d'analyse. On verrait surtout la ruine de toute forme de méthodologisme qui poserait *a priori* l'angle d'approche de la démarche. La dialectologie est riche de toute cette réflexion moderne engagée en ethnologie et en sociologie³⁴.

Pour conclure sur le deuxième paradoxe de l'observateur, les méthodes d'enquête qui réduisent l'hétérogénéité de la langue parlée, nous emploierons à nouveau une image : on lance des filets et on ramène les prises que les mailles du filet ont pu capturer. Avec un autre maillage, les prises sont différentes. Il apparaît à la modeste échelle de la collecte des données par un enquêteur individuel, que le travail scientifique est fragmentaire et se définit d'abord par la conscience de ses limites. Le chercheur ne peut que mettre bon ordre dans son objet d'étude. Ce n'est que par sa médiation, c'est-à-dire ses explications et ses manies de

³⁴ Nous nous contenterons de citer Beaud, Stéphane et Weber, Florence, 1997 *Guide de l'enquête de terrain* Paris, La Découverte.

rangement que l'on jugera de l'ancrage de cet objet d'étude dans le réel. Ces considérations prennent un relief particulier avec le troisième paradoxe de l'observateur.

Le troisième paradoxe de l'observateur, la transcription réduit la variabilité de la langue parlée

Dès le premier numéro de la *Revue des patois gallo-romans*, Gilliéron et Rousselot mettent au point l'alphabet qui porte leur nom et qui va servir à Gilliéron pour l'Atlas Linguistique de la France. On peut lire les remarques fondamentales suivantes :

Sans une éducation spéciale de l'oreille, on n'entend bien que les sons auxquels on est habitué. (...) Dans l'intérêt de la science, le transcripateur devrait signaler toutes les circonstances qui peuvent montrer jusqu'à quel point il est capable de bien entendre les sons et de les bien noter.³⁵

Marcel Cohen reprend ceci de façon très claire :

On n'entend bien que ce que l'on comprend ; Mais, par une difficulté inverse, celui qui comprend ou croit comprendre risque d'entendre, de bonne foi, des sons qui n'ont pas été émis. La perception se fait aussi bien que possible si l'observateur sait le sens de ce qu'il va entendre, pressent à peu près ce qui va être prononcé, mais n'a aucune idée préconçue sur le détail de ce qu'il va enregistrer. On doit donc réunir les meilleures conditions possibles d'attention aisée et de liberté d'esprit.³⁶

On voit que l'esprit de système est totalement étranger aux dialectologues. Leurs instructions de transcription des données orales est, quel qu'en soit l'auteur, une leçon de responsabilité et de liberté. L'envers de cette autonomie des chercheurs est une certaine disparité entre les atlas en ce qui concerne l'usage des signes phonétiques, disparité que les dialectologues sont les premiers à reconnaître :

Conséquence de la trop grande autonomie de chaque chantier, chaque zone a son originalité dans ce domaine. Tel chercheur a jugé indispensable d'introduire de nouveaux signes phonétiques, tel autre préfère une transcription plus proche de la phonologie, tel autre enfin

³⁵ Abbé Rousselot, 1887, p.17.

³⁶ Cohen Marcel, 1950 p. 21.

multiplie à l'envi les signes superposés, marquant ainsi une grande hésitation dans la transcription du timbre des voyelles. Il résulte de tout cela un foisonnement de signes et de combinaison qui a rendu nécessaire la création pour chaque atlas d'un alphabet phonétique particulier³⁷.

Pour l'étude du système verbal du FPA, la variabilité des occurrences verbales était telle que nous avons douté du fonctionnement binaire du caractère discret du signe linguistique. Nous avons dû élaborer à notre propre usage la notion existante de diaphone³⁸ et celle de diamorphe afin de rendre compte de l'extrême labilité des structures verbales. Ces dernières fonctionnent plus ou moins bien, par intermittences, avec quantité de zones de flottements. Les diaphones et les diamorphes ne remplacent pas les phonèmes et les morphèmes. Ils sont d'une autre nature, plus proche du réel dont ils donnent une première interprétation. Ils contiennent en germe les phonèmes et les morphèmes, produits des nécessaires choix théoriques ultérieurs. Leur inconvénient est d'être trop lourds à manier. Nous n'avons pas imaginé d'autre façon de les noter que de signaler leur aire de dispersion, en juxtaposant les différents allophones dont ils se composent. Nous avons noté les diamorphes de la même façon, en juxtaposant les différents allomorphes dont ils se composent, pour une même fonction verbale. Diaphones et diamorphes ne se révèlent qu'après l'étude morphologique qui montre leurs aires de réalisation pour un fonctionnement donné. En toute rigueur, un corpus de langue orale ne peut être transcrit que par les allophones que l'on entend réellement. C'est ce que nous avons fait manuellement pour l'intégralité du corpus.

Puis, ayant choisi d'étudier la morphologie du système verbal, nous avons relevé toutes les occurrences verbales. Nous avons classé ces occurrences selon leurs fréquences et selon d'autres critères pragmatiques que nous avons sans cesse explicités. On obtient ainsi une photographie réaliste de la morphologie de ce système verbal, ou plutôt diasystème, avec des fréquences majoritaires, des fréquences minoritaires, quelques hypercorrections et bon nombre de réalisations tout à fait aberrantes. Pour montrer que ce parler appartient à l'aire d'intercompréhension du français parlé, il fallait montrer que ses dysfonctionnements formels (diamorphisme) et la labilité de ses réalisations phonétiques n'empêchent

³⁷ Le Du, Jean, 1997 « La disparition du groupe des atlas et l'avenir de la géographie linguistique » in *Le Français Moderne*, LXV, 1.

³⁸ Jones Daniel, 1950, "The phoneme: Its nature and use", in *Language* vol. 28 n°3, Cambridge, Heffer & Sons.

pas l'intercompréhension car ils sont statistiquement minoritaires. Nous avons donc élaboré deux instruments de mesure :

- l'indice de diamorphisme est une indication numérique qui exprime le rapport des occurrences de diamorphisme fonctionnel aux occurrences de formes « correctes »³⁹. L'indice 1 serait le point d'équilibre théorique au-delà duquel la communication passe de plus en plus difficilement pour aboutir à l'incompréhension.

- l'indice de dialectalisation est une indication numérique qui exprime le rapport des occurrences de flottements phonétiques et lexicaux aux occurrences de formes correctes.

L'établissement des indices de diamorphisme et des indices de dialectalisation sur différents points de la morphologie du système verbal réserve des surprises : ces indices ne sont pas très élevés et restent en deçà du seuil où l'intercompréhension est compromise. Autrement dit, ce corpus situé au bord extrême de la francophonie, établi sur une population de migrants dont on pouvait penser qu'ils s'expriment en sabir, ne s'éloigne que peu du français fondamental. Quant aux flottements, nommés écarts ou fautes selon un point de vue normatif, ils semblent se produire de préférence sur les points fragiles du diasystème, là où les locuteurs natifs de la langue butent également lors de leur second apprentissage du français, l'apprentissage scolaire. L'intercompréhension passant majoritairement bien, le FPA, parler d'une communauté stable de locuteurs, dont il est de plus en plus la langue maternelle, est donc un démodialecte du français.

Conclusion :

Collecter et transcrire un corpus oral sont des activités qui supposent des prises de position théoriques sur la langue sans lesquelles le corpus sera inutilisable : une délimitation de l'objet d'étude une hypothèse de travail et des conventions de transcription. Rien de tout cela n'existe à priori, aucun manuel ne donne de recette éprouvée, aucune théorie n'est disponible pour faciliter la tâche. Tout est à construire, à chaque fois. L'aventure intellectuelle qui consiste à ancrer des hypothèses dans le réel, c'est l'histoire même de la recherche pour chaque chercheur.

³⁹ « Correctes » non par rapport à la correction grammaticale, mais par rapport au seuil d'intercompréhension en deçà duquel une communauté de locuteurs continue à se comprendre et au-delà duquel on ne se comprend plus. Dans ce sens, parler correctement, c'est parler à la satisfaction de son entourage habituel.

Seul celui qui a constitué le corpus en connaît la qualité, c'est-à-dire l'ensemble de ses caractéristiques bonnes ou mauvaises. Lui seul connaît sa dette envers ses devanciers, ainsi que les contraintes assumées de son travail. Il connaît donc *in fine* sa marge de liberté ainsi que la validité de son aventure. A lui de la faire partager.

janvier 2007