



# Le corpus en analyse de discours : perspective historique

Jacques Guilhaumou

## ► To cite this version:

Jacques Guilhaumou. Le corpus en analyse de discours : perspective historique. Corpus, Bases, Corpus, Langage - UMR 7320, 2002, pp.21- 50. halshs-00420270

**HAL Id: halshs-00420270**

**<https://halshs.archives-ouvertes.fr/halshs-00420270>**

Submitted on 11 Apr 2020

**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.

**Jacques Guilhaumou**

**Le corpus en analyse du discours: perspective historique.**

Jacques Guilhaumou CNRS/ENS Lettres et Sciences Humaines de Lyon, « Le corpus en analyse du discours. Perspective historique », *Corpus*, revue électronique, N°1, 2002, p. 21-50.

**Introduction**

Le problème de la constitution du corpus occupe une place centrale dans l'analyse de discours en ses débuts, c'est-à-dire à la fin des années 1970, et tout particulièrement dans le champ de l'analyse de discours comme objet de l'histoire mis en place par les travaux pionniers de Régine Robin (1973) et du Centre de lexicologie politique de l'ENS de Saint-Cloud.

Au départ, nous présentons donc les principes communs, en matière de corpus, aux premiers chercheurs dans le but cerner les différentes configurations qui délimitent l'apport et les limites de la démarche inaugurale en analyse de discours au cours des années 1970.

Puis nous restreignons notre réflexion à l'analyse de discours du côté de l'histoire dans la mesure où elle introduit, au début des années 1980 et en rupture avec le corpus construit, la notion de corpus « naturel », c'est-à-dire d'ensemble des énoncés attestés dans l'archive selon une logique propre de type configurationnel. Ainsi « le tournant interprétatif » de l'analyse de discours révoque le corpus clos, remet en cause la confrontation entre le corpus et le hors-corpus, refuse enfin la présentation référentielle des conditions de production, au profit d'une description de la réflexivité du discours à partir d'un travail configurationnel sur les énoncés d'archive, situé à la croisée des travaux de Michel Foucault et de ceux des ethnométhodologues. Il n'est plus question que de l'étude de « moments de corpus » caractérisés à l'intérieur de vastes trajets thématiques où s'entremêlent des évolutions conceptuelles, des émergences de sujets d'énonciation et des événements discursifs.

Le début des années 1990 est alors marqué par l'émergence d'une jeune génération d'historiens du discours qui s'interdisent toute instrumentalisation du corpus, et s'engagent résolument dans une approche des énoncés d'archive qui relève d'une histoire langagière des concepts situé désormais au centre de débats entre

chercheurs de nombreux pays (Guilhaumou, 2000). Le renouvellement de l'abord du corpus devient alors un défi pour les jeunes historiens lexicométriques qui tentent un nouveau croisement, tant du côté de la linguistique que de l'histoire, entre l'analyse de fonctionnements linguistiques et la description de la réflexivité des énoncés. Précisons d'emblée que la propriété de réflexivité du discours relève du fait que la langue a cette capacité spécifique de s'analyser elle-même, c'est-à-dire en utilisant ses propres termes. Le discours est ainsi porteur de ses propres ressources interprétatives, tout en n'étant pas déconnecté de la réalité (Guilhaumou, 2001b). La constitution de « très grand corpus » s'attache à la description de la réflexivité des énoncés au sein de ce qu'on appelle aujourd'hui la linguistique de corpus.

## **I - La première configuration de l'analyse de discours: les années 1970.**

Le corpus, dans sa définition classique, désigne « un ensemble déterminé de textes sur lesquels on applique une méthode définie » (Jean Dubois, 1969). Aspect d'emblée déterminant de la pratique de l'analyste de discours, son ancrage initial dans le discours politique induit une configuration spécifique.

### *1- La configuration de départ, une configuration méthodologique.*

La configuration de départ a pour origine la traduction dans *Langages* 13 (1969), à l'initiative de Jean Dubois, d'un texte du linguiste américain Harris de 1952, « Discours analysis » sous le titre « Analyse de discours ». Elle s'est immédiatement illustrée sous la désignation d'analyse d'énoncés. Pierre Kuentz (1977) montre alors que l'analyse de discours « préside à la constitution du corpus des phrases sur lesquelles se construit la théorie grammaticale ». Il désigne ainsi les naïvetés du linguiste de la langue qui croit pouvoir produire des exemples en extrayant des phrases du discours. En refusant de s'interroger sur cette opération d'extraction, le linguiste de la langue suppose une identité langue/discours: il neutralise l'effet discursif. On voit comment l'analyse de discours s'interroge, dès le début, sur la matérialité de la langue au sein même de la discursivité des énoncés.

Cependant une question symétrique se pose: dans quelle mesure la façon dont le linguiste de la langue constitue son corpus de phrases oriente-t-elle la constitution du corpus en analyse de d'énoncés ?

Evoquons rapidement les étapes de l'opération initiale de l'analyse de discours. On puise dans ce que Jean Dubois appelait « l'universel du discours », donc dans la totalité des énoncés d'une époque, d'un locuteur, d'un groupe social. Découpage

arbitraire à partir d'intérêts, de thèmes, de jugements de savoir. Dans un second temps, au sein du genre « discours politique » alors promu par les événements de mai 1968, on ne retenait finalement que l'ensemble des phrases contenant, en quelque position syntaxique que ce soit, tel ou tel mot pivot. C'est la dernière phase qui produit réellement le corpus: l'application des règles d'équivalence grammaticale proposées par Harris permet d'obtenir un ensemble paradigmatique de phrases transformées, la série des prédicats des mots-pivots.

Ainsi dans le travail conjoint de la linguiste Denise Maldidier et de l'historienne Régine Robin (1974), sur les Remontrances parlementaires face aux Edits de Turgot de 1776, le corpus des phrases régularisées autour des mots-pivots liberté et règlement est reproduit intégralement. La sélection des termes repose ici explicitement sur un savoir historique préalable: il est supposé d'évidence que c'est autour des notions de liberté et de règlement que se joue alors l'opposition entre noblesse et bourgeoisie dans la conjoncture de la tentative réformatrice, vouée à l'échec, de Turgot.

Cependant, ce type d'analyse s'est avérée très utile dans la quête d'un éventuel « sens caché ». Souhaitant appréhender la notion de féodalité à la fin du 18<sup>ème</sup> siècle, Régine Robin (1975) s'est intéressée à son usage supposé dans les Cahiers de la noblesse de 1789. Or, il apparaît d'emblée que les expressions de « droits féodaux » et « féodalité », présentes dans les cahiers de la bourgeoisie, sont absentes des cahiers de la noblesse, alors qu'y sont présentés avec insistance « les droits inhérents aux propriétés de la noblesse ». C'est donc à partir de l'étude des énoncés autour du mot-pivot propriété que s'explicite la volonté des nobles de maintenir « les droits seigneuriaux ». Nous obtenons en effet, après régularisation des énoncés, les trois « phrases de base » suivantes:

- 1- *Toute propriété doit être sacrée*
- 2- *Des (ces) propriétés doivent être respectées*
- 3- *Les X (de la noblesse) doivent être respectés*

La classe des X comprend certes « les droits attachés aux terres », mais aussi « les droits attachés aux fiefs » et « les droits attachés aux hautes, moyennes et basses justice », c'est-à-dire la quasi-totalité des « droits féodaux » qualifiés de « droits appartenant aux propriétés de la noblesse ».

L'analyse de discours, dans sa configuration inaugurale, mettait donc ici en évidence une stratégie de masquage, par le fait de l'utilisation d'un mot du lexique « bourgeois » pour légitimer le maintien des « droits seigneuriaux » d'Ancien Régime.

Cette démarche repose alors sur un double pari:

- le pari sur la représentativité: le corpus « final » est censé représenter une réalité discursive. Dans les exemples évoqués ci-dessus, les corpus autour des mots pivots

liberté et règlement, propriété et droits féodaux résumés, pour les premiers, l'affrontement entre Turgot et les parlementaires en 1776, et, pour les seconds, l'opposition entre noblesse et bourgeoisie en 1789.

- le pari sur la systématique: on recherche une homogénéité dans le temps et dans l'espace, par exemple en étudiant, avec Jean-Baptiste Marcellesi (1971), les textes du Congrès de Tours de 1920, c'est-à-dire dans une même « situation de communication ». Cette homogénéisation a pour corollaire une démarche comparative, par exemple la prise en compte de textes de locuteurs différents au sein même du Congrès. Elle produit une systématique, un effet de cohérence.

La circularité de la démarche est alors évidente: si l'analyse de discours emprunte alors son modèle de systématique à la linguistique structurale, la linguistique elle-même, par ses jugements de grammaticalité, réitère des jugements idéologiques qui servent d'instance de jugement à l'analyse de discours. Ainsi en est-il de l'application du modèle historiographique de l'idéologie sans-culotte (Albert Soboul), d'abord au champ sémantique, puis aux phrases grammaticalement transformées autour des mots-pivots *peuple* et *sans-culotte* dans la presse pamphlétaire de l'été 1793 (Guilhaumou 1974, 1975).

L'analyse d'un corpus, certes issu d'un choix socio-historique explicite, neutralise donc son extérieur discursif, le hors-corpus, en le fixant dans un jugement de savoir, tout en constituant des entités discursives stables telles que le discours noble vs le discours bourgeois, le discours jacobin vs le discours sans-culotte. Le niveau interprétatif du corpus se situe alors exclusivement dans la recherche d'un positionnement adéquat du discours politique.

## *2- D'une configuration socio-linguistique au changement de terrain.*

De fait, la démarche inaugurale de l'analyse de discours s'avère très rapidement prendre la forme d'une approche sociolinguistique en ce sens qu'elle associe un modèle linguistique, essentiellement l'analyse d'énoncé, à un modèle sociologique, défini à travers la notion de conditions de production, autre désignation du contexte dans lequel on puise les éléments du corpus étudié. Tout est ici affaire de correspondances, de co-variance entre des structures linguistiques et des modèles sociaux. C'est ainsi que la recherche inaugurale de Régine Robin (1970) sur une ville à la fin de l'Ancien Régime, Semur-en-Auxois, comporte d'une part une analyse des structures sociales d'un bailliage bourguignon à la veille de la Révolution française, et d'autre part une analyse du contenu des *Cahiers de doléances* de la bourgeoisie et de la paysannerie à partir d'un certain nombre de mots-pivots, selon une approche linguistique combinant analyse d'énoncé et étude du vocabulaire socio-politique. Il en est de même de l'étude de J. B. Marcellesi

(1976) sur les Congrès socialistes et communistes de 1924 et 1925. Jean-Pierre Faye (1972) s'y associe en introduisant une approche discursive des effets de récit, et surtout des topologies narratives.

L'activité langagière apparaît alors immédiatement sociale. A l'isomorphisme langue/discours succède une unité dialectique entre le linguistique et l'extra-linguistique, un isomorphisme entre le discours et les pratiques sociales. Le choix des mots-pivots, donc du corpus d'énoncés, présuppose une définition du discours politique en terme d'acceptabilité sociale.

Ainsi, Jean-Baptiste Marcellesi, après avoir défini le discours politique comme un « discours tenu pour l'hégémonie par un intellectuel collectif », sous l'influence de Gramsci, fait porter son analyse des Congrès sur les énoncés contenant les termes Parti et Congrès. Il montre comment le P.C.F s'individualise par rapport aux socialistes, se constituant ainsi comme nouvel intellectuel collectif, en évitant l'équivalence Parti/Congrès. Le Parti est ce qui fait, à l'image de sa lutte contre l'hégémonie capitaliste.

De même Jean-Pierre Faye définit le discours politique comme un discours étatique visant à rendre crédible l'idéologie nazie au sein des idéologies nationalistes de l'entre-deux-guerres. Cependant il introduit une dimension pragmatique et stratégique dans l'approche du discours politique qui permet de désigner l'effet de récit à l'encontre de l'isomorphisme langue/discours. En effet, d'une part, Jean-Pierre Faye constitue le corpus des énoncés qui désignent le noyau régulateur du champ idéologique de l'extrême droite allemande de l'Allemagne de Weimar à partir de la formule de l'Etat total et de l'antithèse de la Révolution conservatrice. Alors s'ouvre la possibilité de jeux d'inversion complexes. autour de l'axe sémantique *Völkisch-Bündisch*: de « L'ouvrier c'est le conservateur » à l'énoncé d'Hitler en 1936, « Je suis le révolutionnaire le plus conservateur du monde », une conception raciste, antisémite et anti-internationaliste du monde est mise en acte au sein même du langage.

C'est alors la frappe narrative ou *l'effet de récit* qui donne consistance au totalitarisme nazi au sein de la circulation des langages dans le champ discursif de l'extrême droite allemande de l'entre-deux-guerres. Le fait discursif devient alors événement, action narrée. Le discours n'est pas uniquement ce qui énonce l'action rapportée, mais il est aussi ce qui produit l'action; il est fondamentalement acte de langage.

Le travail de Jean-Pierre Faye annonce ainsi un « changement de terrain », présentement une ouverture vers les récits idéologiques qui rompt avec toute définition implicite, donc non réfléchie, du discours politique.

Cependant le changement de terrain est alors essentiellement perceptible, dans l'analyse de discours des années 1970, à partir de la sortie interprétative du corpus

vers une réflexion théorique d'inspiration marxiste sur le discours politique comme pratique relativement autonome. Dans la lignée des réflexions de Michel Foucault dans *L'archéologie du savoir* (1969) et du renouveau de la pensée marxiste à l'initiative de Louis Althusser (1965), Régine Robin et Michel Pêcheux s'efforcent plus d'approcher les pratiques discursives dans leur historicité et leur spécificité que de constituer, de manière éphémère, une théorie du discours articulée sur une théorie des idéologies. Le discours politique n'est pas défini a priori, ses caractéristiques sont liées au moment actuel, au plus loin de toute démarche taxinomique.

Deux concepts, tels qu'ils sont formulés par Michel Pêcheux (1975), deviennent centraux, ceux de formation discursive et d'interdiscours. L'étude des formations discursives permet alors de déterminer ce qui peut et doit être dit dans une conjoncture donnée. Le risque était là, nous l'avons déjà souligné, de classer les diverses formations discursives d'une formation sociale, à l'exemple de l'opposition noblesse/bourgeoisie sous l'Ancien Régime. Le concept d'interdiscours introduit alors une approche plus dialectique, dans la mesure où il permet de dire que toute formation discursive dissimule, dans la transparence du sens propre à la linéarité du texte, une dépendance à l'égard d'un « tout complexe à dominante » selon la formule du philosophe marxiste Louis Althusser (1965), ensemble qui n'est autre que l'interdiscours, cet espace discursif et idéologique où se déploient les formations discursives en fonction de rapports de domination, de subordination et de contradiction. Cette conceptualisation « forte » rencontre alors le souci de l'historien du discours d'inscrire durablement son interrogation du côté de la tradition marxiste<sup>1</sup>. Là où dominait au départ une approche taxinomique, isolant des éléments simples (discours bourgeois/discours féodal; discours jacobin/discours sans-culotte) dans le corps complexe des discours, il était question en fin de compte, dans le jeu de l'interdiscours et de l'intradiscours, d'intrication de stratégies discursives, d'affrontements et d'alliances langagiers.

En fin de compte, le « changement de terrain », sévèrement jugé après-coup, ne visait pas l'installation du discursif dans un système conceptuel figé et externe, mais au contraire poussait le chercheur à un perpétuel travail de remaniement des concepts sur une base discursive. Dans cette ultime métamorphose de la configuration de départ, la question de la constitution du corpus rendait obligatoire l'étude des conditions de production du discours, en liaison avec l'histoire des formations sociales. Quant à la notion de conditions de production, elle désignait alors moins la réalité stable de la situation de communication qu'un travail sur les

---

<sup>1</sup> Nous n'avons pas la place de développer cette relation fondamentale au marxisme, dont nous résumons les étapes, au sein de notre propre itinéraire de recherche, dans Guilhaumou, 1996.

effets de conjoncture, de l'effet-sujet à l'effet-appareil en passant par l'effet majeur de l'événement.

Cependant *l'analyse du discours comme objet de l'histoire* n'arrivait pas à sortir, dans cette configuration, d'un triple écueil. En premier lieu, elle introduisait une coupure nette entre le corpus choisi, à vrai dire fort restreint au terme de la procédure d'analyse, et le hors-corpus défini de façon référentielle et idéologique par la notion de conditions de production. En second lieu, le choix des mots-pivots reposait sur le jugement de savoir de l'historien, pris lui-même dans le champ des débats historiographiques, et non dans les termes du moment actuel. Enfin, elle constituait, sur des bases idéologiques et historiographiques, des entités discursives séparées telles que le discours noble, le discours bourgeois, le discours jacobin, le discours sans-culotte, etc.

## **II - La démarche configurationnelle au premier plan.**

La lexicométrie est mise en place au cours des années 1970 avec l'étude pionnière des tracts de mai 1968 (Tournier et alii, 1975): ses avancées sont perceptibles à travers la première formule de la revue *Mots* jusqu'à la fin des années 1980. En quantifiant les faits langagiers, elle ouvre la voie à une linguistique de corpus qui définit le corpus comme une collection de données langagières sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage (Habert, Nazarenko, Salem, 1997, 11), résolvant ainsi en partie les problèmes des linguistes face au corpus. Le linguiste finit en effet par concentrer son attention sur l'enrichissement des corpus, via les banques de données, l'accroissement de leur taille et l'amélioration des accès aux corpus.

Mais l'historien du discours a d'abord fait appel à la lexicométrie pour démêler, en première approche, l'intrication des phénomènes énonciatifs et rhétoriques qui constituent la surface discursive d'un texte, par contraste avec les énoncés qui le structurent sémantiquement autour de mots-pivots étudiés en analyse harissienne. Là encore, la procédure d'analyse portait au départ sur un corpus réduit, non plus un corpus d'énoncés, mais le tableau lexical à double entrée des formes recensées automatiquement du corpus qui sont ventilées sur la base de leur fréquence absolue et relative dans les diverses parties du discours.

Le passage par l'archive, si spécifique de la démarche de l'historien du discours pendant les années 1980 mais sur lequel nous serons fort succinct dans la mesure où elle a fait l'objet d'un ouvrage-bilan (Guilhaumou, Maldidier, Robin, 1994), devait permettre d'enrichir l'approche lexicométrique de l'historien du discours. Certes l'analyse de corpus, à l'aide d'outils linguistiques adéquats, demeure plus que jamais d'actualité. Mais d'une part il ne s'agit plus de construire d'emblée un



corpus sur la base d'un jugement de savoir, au titre de la désignation préalable de conditions de production, mais il convient de décrire d'abord des configurations d'archives significatives à la fois d'un thème, d'un sujet, d'un concept, et en fin de compte d'en événement. Il est alors toujours temps d'isoler en leur sein un « moment de corpus », c'est-à-dire un ensemble d'énoncés sur des critères lexicaux, syntaxiques ou énonciatifs et de constituer ainsi un moment de corpus susceptible d'une approche linguistique fine.

L'apport de la démarche lexicométrique s'en trouve singulièrement modifié. Nous prendrons l'exemple d'un corpus chronologique d'environ 150.000 occurrences, le *Père Duchesne* d'Hébert en 1793-1794, que nous avons étudié (1986a) en collaboration étroite avec le linguiste lexicomètre André Salem (1986). Il ne s'agit plus seulement, à partir de l'analyse du tableau lexical de répartition des formes, de décrire une évolution du vocabulaire sur la base automatisée de regroupements quasi-sémantiques de termes. La démarche à la fois descriptive et interprétative consiste plutôt à articuler des indices quantitatifs de fonctionnement discursif, introduisant des éléments contrastifs dans le corpus, à des *descriptions locales* en prise sur des énoncés hors-corpus.

Prenons l'exemple de la mise à l'ordre du jour de la Terreur en 1793, mot d'ordre dont nous avons retracé par ailleurs le cheminement (1981), en nous posant la question suivante: quelle place discursive occupe cet événement discursif dans la narration duchênienne ?

Le réseau d'énoncés de la thématique de la terreur s'organise alors en deux temps au sein du *Père Duchesne* d'Hébert. Au moment de la mise en place du mot d'ordre de terreur à l'ordre du jour, pendant le mois d'août 1793, l'usage répété de la figure de l'aristocrate à travers les multiformes « tous les aristocrates », « aux aristocrates », « les aristocrates » légitime de l'extérieur l'existence même de la terreur. Puis l'inversion de l'usage de l'expression populaire « perdre le goût du pain » qui, d'abord appliquée aux amis de la République, concerne ses ennemis, réitère le processus de retournement de la terreur, exercé initialement par les ennemis, contre ces mêmes ennemis. Dans un second temps, une fois la mise à l'ordre du jour de la terreur légitimée par la Convention Nationale, le *Père Duchesne* use systématiquement de la multiforme « tous les ennemis », indice de l'émergence d'une systématique de terreur dans le discours duchesnien.

En fin de compte, une telle description locale d'origine lexicométrique permet de situer les enjeux stratégiques à l'intérieur même de la description textuelle, et non dans l'articulation d'une surface discursive à une formation discursive déterminée par ses conditions de production, comme cela était le cas au début de l'analyse de discours, lorsqu'il s'agissait d'établir un lien entre les effets populaires dans le *Père Duchesne* et son contenu jacobin (Guilhaumou, 1974).

De la démarche configurationnelle de l'historien du discours qui se met en place dans les années 1980, nous avons décrit les principales entrées, *configuration/archive*, *trajet thématique* et *événement discursif*, sous forme d'un second bilan, dans le récent *Dictionnaire d'analyse de discours* (Chareaudeau, Maingueneau, 2002). Dans la mesure où cette démarche tend à remplacer la démarche usuelle de construction du corpus par une approche en corpus « naturel » des énoncés d'archive, il est difficile d'y trouver une quelconque centralité de la question du corpus.

L'archive est ici définie comme « le jeu de règles qui déterminent dans une culture l'apparition et la disparition des énoncés, leur rémanence et leur effacement, leur existence paradoxale d'événements et de choses » (Foucault, 1994, I, 708). Elle est alors associée à un *geste de lecture* de facture herméneutique, donc foncièrement anti-constructiviste, c'est-à-dire basé sur les capacités réflexives, donc interprétatives, des acteurs « naturels » de l'histoire. Elle s'inscrit ainsi dans l'activité ordinaire des gens ordinaires, à l'encontre de toute préséance des acteurs jugés légitimes. Désormais, l'énoncé « rare » ne prend sens qu'au terme d'un travail configurationnel sur un ensemble d'énoncés attestés, dispersés et réguliers, donc nous renvoie à l'articulation de la description et de la réflexion au sein d'un acte configurant centré sur une intrigue, bien souvent concentrée dans un événement discursif. Il se situe au plus loin du découpage par le linguiste, de la linéarité du discours, en expressions, propositions et phrases.

Paradoxe apparent, la démarche révoquant la primauté de la construction du corpus permet à l'historien et au linguiste de se rencontrer au mieux sur le terrain de la matérialité de la langue au sein de la discursivité de l'archive. Ainsi peut-on évaluer l'exemplarité de notre étude, avec Denise Maldidier, des modalités discursives de la récurrence de l'expression « Du pain et X » (Guilhaumou, Maldidier, 1986) au sein du trajet thématique des subsistances, décrit par ailleurs dans toute son amplitude (Guilhaumou, 2001). Il en est de même du « moment de corpus » constitué des énoncés autour de « Marat est mort/ Marat n'est pas mort » dans la description configurationnelle de l'événement « Mort de Marat » (Guilhaumou, 1986b, 1987).

Les ouvertures problématiques des historiens du discours dans les années 1980 ont été rendues possible par la neutralisation progressive de toute démarche analogique de type sociolinguistique et la constitution de l'analyse de discours comme discipline interprétative à part entière (Guilhaumou, 1993). Mais il a fallu d'abord récuser la notion de conditions de production, et son corollaire, la situation de communication, en situant les ressources interprétatives des textes en leur sein. Puis l'histoire du discours s'est libérée en partie des modélisations sémantiques *a priori* dans la manière d'agencer les « phrases de base » de l'analyse d'énoncés

et/ou d'organiser les listes de termes différenciés issus des analyses lexicométriques de tableaux lexicaux. Enfin il est devenu possible de ne plus instrumentaliser la linguistique en s'en servant comme d'un simple outil, pour ensuite l'exclure du moment interprétatif. Il s'agit donc bien là d'un *tournant langagier* de facture fortement herméneutique, qui s'intègre dans ce qu'il est convenu d'appeler « le tournant interprétatif »: l'acte d'interpréter est mis au centre de « la constitution linguistique du monde » et se concrétise dans l'historicité des discours par un vaste gamme pragmatique d'actes de langage. Il s'agit bien alors d'attester de l'existence de l'événement discursif, d'en déclarer la signification attestée par rapport à ce qui est dit dans un moment donné au sein d'une configuration d'énoncés. A partir de la lecture d'archives, l'intérêt justifié des analystes de discours pour l'événement de communication (Chareaudeau, 1997), échelle commune de notre société médiatique, est déplacé vers des événements où la mise en évidence de l'attitude réflexive des acteurs, c'est-à-dire de leur capacité à l'autodétermination, permet d'insister plus sur leur liberté de sujet s'émancipant que sur leur réalité d'être social déterminé.

Que devient alors la question initiale de la constitution du corpus hors du champ restreint du questionnement linguistique sur l'enrichissement des corpus, en particulier par le biais des corpus annotés (Habert, Nazarenko, Salem, 1997), auquel l'historien peut être sensible ? Sans aucun doute elle paraît quelque peu marginalisée à la lecture des travaux récents de jeunes historiens du discours engagés du côté de l'histoire langagière des concepts.

De fait, l'historien du discours tend de plus en plus, au cours des années 1980, à considérer l'analyse de discours comme une *ethno-méthode*. Il s'agit d'abord d'adopter la posture d'indifférence ethnométhodologique qui refuse d'induire de la position sociale les pratiques discursives d'un individu, ou d'un groupe. Puis la frontière entre texte et contexte s'estompe dans la mesure où les individus utilisent le langage naturel comme contexte, ressource, thème, donnant ainsi un sens à leur action par la seule réflexivité du discours. « Ethno » se réfère ici conjointement à la culture, aux croyances locales d'un membre de la société au-delà de son appartenance à une groupe déterminé et à la connaissance dont dispose ce membre de cette culture dont il est à la fois l'acteur, le protagoniste, l'auteur et le spectateur (Widmer, 1986; Coulon, 1987). La perspective ethnométhodologique permet également d'aborder les « ordres de grandeur » qui hiérarchisent l'idéologie sociale, donc les relations d'ordre entre les individus au sein des institutions, d'après « ce dont les gens sont capables » (Boltanski, 1990; Thévenot, Boltanski, 1991).

L'analyse de discours tend ainsi à prendre un « tournant anthropologique » où les problèmes de l'intersubjectivité et de l'altérité sont au centre des

préoccupations des jeunes historiens du discours. L'étude de Sophie Wahnich (1997) sur la thématique de l'étranger pendant la Révolution française s'avère tout à fait significative de ce tournant.

Sans entrer dans les résultats de l'étude de trajets thématiques très intriqués, constatons d'abord que ce travail discursif sur l'espace interlocutif du discours d'assemblée pendant la Révolution française montre la capacité à produire des résultats historiques d'un travail configurationnel sur les énoncés d'archive. Son apport central au « tournant anthropologique » tient à une recherche en cours (2002) sur la thématization des émotions au sein même des discours argumentatifs politiques de la Révolution française. La réflexivité langagière du thème des émotions devient ainsi un objet de recherche à part entière au sein d'une économie émotive de l'événement qui peut prendre, dans certains cas, la forme d'une économie linguistique. Ainsi lorsque les Montagnards en l'an II dénoncent l'usage fortement émotif par les anglais de syntagmes figés (« les soldats de Robespierre », « les hordes conventionnels »), de structures syntaxiques (« Robespierre a fait ordonner.. », « les hordes conventionnels ont fait.. ».) à l'encontre de l'appropriation historique et mémorielle des événements révolutionnaires (Wahnich, 1999). Nous pouvons terminer ce point par un autre exemple particulièrement significatif de la mise à l'écart de la question de la construction du corpus dans l'analyse configurationnelle des énoncés d'archive. Lorsque Sophie Wahnich montre que l'acte de demande constitutif de la radicalité révolutionnaire dès 1789 devient lui-même un acte émotif, qui plus est que la demande émotive de terreur est présente dès 1792 et précède donc la mise à l'ordre du jour de la terreur pendant l'été 1793, elle contourne le corpus des énoncés d'archive, aussi vaste soit-il, de l'événementialité particulièrement dense de l'été 1793, pour aller au plus profond de la rationalité pratique des émotions.

### **III - Les nouvelles conditions du retour au corpus.**

A vrai dire, tant chez l'historien que chez le linguiste, la construction du corpus, selon une procédure explicite et sur la base d'une validation empirique, était perçue comme une donnée scientifique permettant d'élargir la continuité du social vers les phénomènes langagiers. Désormais, le choix d'une échelle d'observation, et plus spécifiquement la détermination d'un observatoire des discours selon le point de vue adopté prime dans l'opération de construction du corpus.

Nous l'avons constaté au moment où nous avons pu confronté, de concert avec d'autres linguistes, diverses démarches en analyse de discours qui ont pour trait commun de s'intéresser à des *corpus historiques* d'une part, et de prendre en

compte la matérialité du langage et des formes de la langue d'autre part (Branca-Rosoff, Collinot, Guilhaumou, Mazière, 1995).

Au terme de cette confrontation, il est apparu au moins trois modalités différentes d'observatoire linguistique, dont dépend l'opération même de construction du corpus.

Il est tout d'abord question d'un *observatoire des discours* où se confrontent en permanence la matérialité des formes linguistiques et les valeurs de l'institution sociale fortement mobilisées dans l'opération de construction du corpus. Ici l'observateur-linguiste porte toute son attention sur les ressources interprétatives des discours en les interprétant pas la médiation des formes linguistiques, et plus largement des formes de l'écriture. C'est ainsi que Sonia Branca et Nathalie Schneider (1994) ont constitué un corpus des manières de dire - formes socialement réglées mais soumises à des déplacements, en particulier sur l'axe oral/écrit - des peu-lettrés provençaux pendant la Révolution française à partir de discours publics manuscrits. Cet observatoire des discours accorde une place importante à la lexicologie sociale sur la base d'une approche du mot comme notion hétérogène (Branca éd., 1998). De ce fait, la lexicologie peut tout aussi bien s'intéresser à la valeur d'usage d'un mot dans un vaste corpus ouvert comme Frantex, enrichi de lectures complémentaires, à l'exemple de l'étude du mot « classe » par Marie-France Piguet (1996) que s'en tenir à une première analyse exhaustive d'un corpus bien délimité, pour l'enrichir ensuite sur la base de nouvelles sources. Ainsi la saisie à vif de l'usage des mots résulte d'une clôture en profondeur du corpus, par la rencontre avec l'archive, et à propos de laquelle le mieux est...de ne pas clore (Eluerd, 2000, 101).

En second lieu, l'arrivée récente de l'analyste du discours dans le champ de l'histoire de la linguistique (Auroux, 1989-1999) fait apparaître un *observateur-historien* soucieux de décrire, dans la connexion empirique entre la réalité et le discours à un moment historique donné, les capacités réflexives de sujets de la langue explicitement dotés d'outils et de conscience linguistiques, par exemple les grammairiens-philosophes du 18ème siècle (Guilhaumou 2001b, 2002). Le travail d'archives permet non seulement de fournir un matériau empirique inédit, mais aussi de cerner la langue empirique en tant qu'elle est composée d'états et de sujets cognitifs fixant le possible en langue.

Il ne s'agit donc pas de décrire de simples manifestations de conscience linguistique au sein de l'événement discursif, mais plus avant de repérer dans l'espace-temps de la communication humaine les sujets et les objets, par exemple le *quelqu'un qui parle* et le *quelque chose qui existe* comme sujet politique de la langue française au 18ème siècle, au sein d'une événementialité originaire vide d'une signification première mais qui juge de l'appartenance de chacun à une

communauté de langage. Cet observatoire constitué dans la connexion empirique entre un matériau archivistique et une langue archétypique, ne peut s'autoriser, à vrai dire, d'aucune autre opération de construction de corpus que de l'existence externe d'objets linguistiques. Sur cette base cognitive, toutes sortes de connexions sont possibles, y compris dans l'espace fortement subjectif et réflexif de l'événement discursif, processus qui continue à s'expliquer pour une part par la capacité réflexive des acteurs agissant en son sein.

La réflexivité généralisée des discours, propre à la démarche herméneutique en histoire du discours des années en 1980, est actuellement fortement nuancée par une interrogation sur un référent toujours intersubjectivement négociable, étant alors entendu que la forme langagière de l'accord humain, de l'intentionnalité collective importe beaucoup plus que la signification du contenu des représentations. Dans cette perspective, le corpus est soit une modalité de structuration du savoir en analyse de discours éminemment transitoire donc qui peut disparaître, soit un nouvel objet technique analogue au mode d'existence matériel de la langue. Le débat sur le corpus peut alors reprendre, nous le verrons, si l'on part de la remarque suivante, à vrai dire très matérielle, « La langue a l'air assez différente quand on examine un grand morceau d'un coup » (cité par Habert, Nazarenko, Salem, 1997, 217).

Cependant, il convient d'abord de délimiter un troisième observatoire possible, *l'observatoire de la langue* constitué à partir de l'hétérogénéité même du discours. Cet observatoire est toujours pris dans la tension d'une instrumentation confrontée en permanence avec les avancées de la linguistique au sein de moments de corpus où se manifeste la matérialité de la langue dans la discursivité de l'archive. Cependant, le processus de construction du corpus est alors motivé par l'association de l'exigence formaliste à un geste de lecture effectué au sein d'univers discursifs non stabilisés logiquement. Il relève de la constitution de « machines paradoxales » qui interdisent toute construction du corpus par étapes et à ordre fixe. Une fois acquis, sans connotation intentionaliste, que c'est de l'intérieur du champ discursif, dans l'immense circulation des énoncés, que la question du corpus est désormais posée, l'interaction de moments de l'analyse linguistique et de l'analyse discursive relève de la production en spirales de reconfigurations de corpus (Pêcheux, 1990). Ce troisième observatoire peut s'étendre aux objets linguistiques eux-mêmes, par exemple les Dictionnaires (Collinot, Mazière, 1997), en vue de mettre en évidence des fonctionnements aveugles des discours dans la mise en collocation des mots entre eux et dans les perturbations apportées par le recours aux citations.

Dans tous les cas évoqués ci-dessus, l'exigence initiale d'un corpus stable, homogène et clos a été abandonnée. Pour le jeune historien du discours Marc

Deleplace, cet abandon, justifié par la nécessaire ouverture au large éventail des publications imprimées et des sources manuscrites permet de passer du discours comme simple objet de l'histoire au *discours constitué comme objet social en lui-même* (Deleplace, 1996). Ainsi rejoint-il la réflexion de l'historien allemand Reinhart Koselleck (1997) sur la connexion empirique entre réalité et discours dans la nécessaire distinction entre les faits sociaux et les faits de langage. Mais il lui importe tout autant d'assurer la cohérence du corpus utilisé de manière à garantir la permutation des énonciateurs. Et de citer Antoine Prost, un de ses aînés historiens les plus intéressés par l'analyse de discours: « Nous avons vu que ce corpus devait présenter trois caractères; être contrastif, pour permettre des comparaisons; être diachronique, c'est-à-dire s'échelonner dans le temps, pour permettre de repérer continuités et tournants; enfin être constitué, sinon de textes d'organisation, du moins de textes significatifs, assignables à des situations de communication déterminées » (1988, 280). Caractères que nous retrouvons donc dans le vaste travail de Marc Deleplace sur le concept d'anarchie de Mably à Proudhon (2001).

En fin de compte, l'analyse de discours du côté de l'histoire en France s'ouvre sur une histoire linguistique des usages conceptuels en tant que partie intégrante de l'histoire langagière des concepts (Guilhaumou, 2000). Elle garde ainsi sa spécificité tout en s'intégrant au sein d'un cadre européen plus large dans la mesure où elle introduit, dans la désignation commune d'histoire des concepts, deux notions centrales dans les travaux français: l'une, *l'usage*, associée à une longue tradition d'analyse lexicologique de l'usage des mots en discours (Eluerd, 2000), l'autre, *la linguistique*, issue d'un champ de la linguistique en France où chaque évolution majeure nécessite un repositionnement par rapport à la référence relativement stable au métier d'historien. La récente publication collective du laboratoire de lexicologie politique de Saint-Cloud, sous la responsabilité de Pierre Fiala (1999), amorce une telle évolution. En son sein, nous notons en particulier la présence de l'historienne de la Révolution française, Raymonde Monnier, l'une des responsables par ailleurs du réseau européen « History of concepts » dont les récents travaux (1999a et b) marquent une évolution sensible vers l'analyse de discours.

Nous avons vu que l'analyse de discours a considérablement progressé par la prise en compte de la réflexivité du discours. Plutôt que de préciser les enjeux d'une telle catégorisation sur une base réflexive, dans la mesure où nous l'avons déjà fait dans la postface relative aux historiens et le « tournant herméneutique » à notre ouvrage de synthèse sur les porte-parole pendant la Révolution française (1998), nous préférons montrer en quoi l'impact de cette évolution sur la lexicométrie ouvre une perspective novatrice en matière de construction de corpus.

Au-delà de l'apport de la lexicométrie au travail du linguiste sur des matériaux historiques, relatifs en particulier à la Révolution française (Moreau-Steuckardt, 2000), nous retiendrons comme particulièrement significatif de cette évolution le récent travail de Damon Mayaffre (2000), qui nous entraîne sur les rives du discours politique français de l'entre-deux-guerres. Dans la mesure où ce jeune chercheur précise ici même l'originalité de sa démarche en matière de construction de corpus, nous nous contentons de souligner les apports qui nous semblent essentiels de sa première recherche, lui laissant ainsi le soin d'en systématiser les ouvertures nouvelles.

La première originalité de ce travail réside dans la constitution d'un vaste corpus d'environ 1.500.000 occurrences réparties de manière à peu près équivalentes entre les discours de quatre dirigeants politiques des années 30, Thorez, Blum, Flandin et Tardieu, sur une dizaine d'années, justifiant ainsi une étude à la fois contrastive, synchronique et diachronique. La lexicométrie, avec sa boîte à outils (les listes de fréquences et de répartition des formes par sous-corpus, l'analyse factorielle des correspondances à valeur globalisante, les énoncés répétés, etc.), est mise à contribution tout au long de ce travail. Mais ce qu'il importe avant tout de souligner, c'est qu'il s'agit d'une expérimentation tout à fait novatrice en matière de « très grand corpus »: il est avant tout question ici du *corpus discursif lui-même comme objet d'histoire* à l'horizon d'un genre discursif spécifique.

Ainsi les critères usuels de constitution de corpus (homogénéité, contrastivité et diachronicité) sont respectés, mais ils n'aboutissent pas à l'échantillonnage habituel. Le choix se veut quasi-exhaustif dans le champ discursif étudié, le discours parlementaire de quatre dirigeants politiques, sur la base d'un travail d'archivage et de numérisation systématique. C'est donc le corpus lui-même, plus exactement sa description à partir de critères quantitatifs qui constitue sa propre norme interprétative.

Qui plus est, l'analyse de discours sur des corpus de faible envergure devait faire appel en permanence à des données contextuelles pour valider l'interprétation historique. Mais ici la taille même du corpus rend possible, dans de nombreux cas, la présence du contexte dans le texte lui-même. Cette question de *la réflexivité du corpus*, donc de la disponibilité, dans le corpus lui-même, d'une grande partie des ressources nécessaires pour interpréter les discours politiques étudiés marque un bond qualitatif par rapport aux études lexicométriques antérieures. Nous en percevons tout particulièrement l'intérêt de cette démarche au moment de l'interprétation historique des résultats, là où les textes à l'appui du débat sur des questions historiographiques précises sont le plus souvent des textes du corpus qui font en quelque sorte réseaux. Là s'établit un lien fructueux avec la manière dont les historiens des concepts reposent actuellement la question du contexte à partir



de la prise en compte de du lien entre les croyances au sein de réseaux interconnectés (Guilhaumou, 2001c).

Au système citationnel usuel constitué sur la base d'une lecture ordinaire d'un hors-corpus définissant la situation de communication se substitue donc une mise en configuration construite à l'intérieur même de l'espace du corpus. Le temps du corpus limité, échantillonné, clos est désormais bien révolu chez les historiens du discours. Et dans la même temps, la spécificité de la démarche française en analyse de discours acquise dès son origine, faire appel à des outils linguistiques d'analyse au sens large, demeure avec les modifications inhérentes à l'enrichissement de l'outillage méthodologique de l'analyse de discours.

### **Conclusion: les nouveaux corpus.**

L'analyse de discours n'accordait en ses débuts aucune place particulière à la question des genres discursifs, d'autant plus que la notion de genre était généralement rapportée à une catégorie idéologique pré-construite. Cependant il est aisé de constater que les nouveaux historiens du discours évoluent majoritairement, en matière de corpus historiques, dans un genre politique fortement légitimé, le discours parlementaire.

A ce titre, il importe désormais d'explicitier, dans la lignée des travaux de Pierre Achard (1995) et de Sonia Branca (1999), la dynamique des genres dans la constitution de nouveaux corpus.

Ainsi Pierre Achard s'intéresse à la manière dont tout chercheur peut homogénéiser des discours, donc les constituer en corpus, sur la base de l'organisation interne de leur énonciation. Il s'agit alors, pour le sociologue du langage, de prendre en compte conjointement « l'énonciation telle que le développement interne du discours la déploie » et « les actions sociales dont le discours est le support » (Achard, 1995, 87). Le registre discursif désigne alors la modalité ainsi mise en oeuvre de construction sociale du sens à laquelle s'associe nécessairement un genre discursif repérable par des éléments linguistiques internes. La langue comme objet social, dans la lignée des travaux de Renée Balibar sur le colinguisme (Branca-Rosoff, Wionet, 1998), est un exemple particulièrement intéressant de registre discursif.

Dans cette voie, il est également possible d'envisager, sur le versant de l'événement linguistique, la formation de nouveaux corpus par une attention particulière « à la restructuration des formes langagières plus ou moins stabilisées que constituent les genres » (Branca-Rosoff, 1999, 125). Il ne s'agit pas de repérer la formation d'un nouveau genre par une simple accumulation de faits linguistiques, mais de mettre en valeur la montée progressive de formes, leur

visibilité, par un déplacement de normes autorisées dans un contexte de rupture. Ainsi en est-il de l'écriture des peu-lettrés pendant la Révolution française (Branca-Rosoff, Schneider, 1994) qui peut faire corpus par la manière dont les chercheurs valorisent sa dimension réflexive, sa créativité au sein même de l'événement révolutionnaire

Nous constatons donc que l'aspect réflexif du discours prend une importance toujours plus grande dans la procédure de formation de nouveaux corpus. Il s'agit en effet d'associer dans un corpus issu de l'activité langagière de locuteurs précis non seulement des acteurs stratégiques et des acteurs émergents, des experts et des militants ordinaires, des protagonistes et des spectateurs, mais aussi les chercheurs en science sociales eux-mêmes. Nous entrons ainsi dans un processus de *co-construction des corpus* où la question de l'articulation du discours et du métadiscours constitue une dynamique interne à la formation de nouveaux corpus.

De plus en plus, le chercheur est confronté, dans le matériau d'enquête et/ou le matériau d'archive, à des configurations de sens inédites qui contribuent à valoriser des sujets émergents, c'est à dire pris, selon la formule de Michel Pécheux (1990), dans « la délocalisation tendancielle du sujet d'énonciation ». Il est alors interpellé par des formes nouvelles de subjectivation et par des objets notionnels inédits. Nous l'avons constaté dans notre recherche conjointe avec des sociologues et des ethnologues (Mesini, Pelen, Guilhaumou, 2002) sur les acteurs émergents, dits « exclus » dans le champ de la lutte contre l'exclusion. L'historien linguiste tend ici à singulariser, dans les « récits de vie » des dits exclus, un trajet narratif où l'autonomie interprétative des ressources du sujet d'énonciation, l'enquêté, est très forte au point que les arguments avancés sont co-construits à part égale entre l'enquêteur et l'enquêté. Le corpus analysé consiste alors dans un trajet énonciatif où l'étude de fonctionnements linguistiques met en évidence une reconnaissance réciproque de l'enquêteur et de l'enquêté, un copartage des ressources entre coauteurs, traduisant ainsi la violence exercée par la société sur les dits « exclus » en une certaine forme de réconciliation discursive, sans pour autant invisibiliser les formes de résistance et de conflit. En effet, le chercheur est confronté lui-même à sa propre responsabilité en matière de quête d'autonomie. De fait l'irruption de la parole du dit « exclu » est si forte qu'elle fait éclater conjointement la notion de sujet social dépendant et celle de chercheur-expert maîtrisant son métadiscours. Ainsi, interpellé par une recherche d'autonomie, l'enquêteur ne se contente pas de se laisser guider par les convictions de l'enquêté. Prenant sa part de responsabilité dans l'émergence d'une dimension universalisante de l'expression authentique de l'individu empirique, il participe aux moments forts producteurs d'arguments dans le cours du trajet narratif. Son

point de vue fait donc partie intégrante du corpus analysé, et ne nécessite donc pas la construction d'un observatoire séparé.

La réflexivité du discours, thème désormais majeur de l'analyse de discours, a ouvert la voie au corpus réflexif non seulement sous la forme du « très grand corpus » où texte et contexte se retrouvent dans un même espace construit, mais aussi sous la forme de corpus co-construits à l'intérieur desquels le point de vue du chercheur est l'une des données majeures de la construction de ces corpus. Double défi que la lexicométrie semble actuellement apte à relever. Nous l'avons montré à propos des travaux récents de Damon Mayaffre sur les « très grands corpus ». Il en est de même, en matière d'encodage du corpus, lorsque Serge Heiden (1999) prévoit une solution d'extensibilité du format d'encodage mis en oeuvre dans l'annotation du corpus, non seulement sur la base de l'évolution future des logiciels d'analyse, mais aussi en fonction de la variation du point de vue du chercheur.

### Références bibliographiques

- Achard, Pierre, *La sociologie du langage*. Paris: PUF, 1993.
- Althusser, Louis: *Pour Marx*. Paris: Maspero, 1965.
- Auroux, Sylvain (éd.): *Histoire des idées linguistiques*. Bruxelles: Mardaga, trois volumes, 1989-1999.
- Boltanski, Luc: *L'amour et la justice comme compétences*. Paris: Métailié, 1990.
- Boltanski, Luc/Thévenot, Laurent: *De la justification. Les économies de la grandeur*. Paris: Gallimard, 1991.
- Branca-Rosoff, Sonia (éd.): *Le mot, analyse de discours et sciences sociales*. Aix: Publications de l'Université de Provence, Langues et langage N°7, 1998.
- Branca-Rosoff, Sonia (éd.): Types, modes et genre en discours. In: *Langage & Société*, 87, mars 1999.
- Branca-Rosoff, Sonia/ Schneider, Nathalie: *L'écriture des citoyens. Une analyse linguistique des écritures des peu-lettrés pendant la Révolution française*. Paris: Klincksieck, collection « Saint-Cloud », 1994.
- Branca-Rosoff, Sonia/Wionet, Chantal (éd.): colinguisme et lexicographie. In: *Langage & Société*, 83-84, mars-juin 1998.
- Branca-Rosoff Sonia, Collinot André, Guilhaumou Jacques, Mazière Francine: Questions d'histoire et de sens. In: *Langages*, 117, mars 1995.
- Chareaudeau, Patrick: *Le discours d'information médiatique. La construction du miroir social*. Paris: Nathan, 1997.
- Chareaudeau, Patrick/Maingueneau, Dominique: *Dictionnaire de l'analyse de discours*. Paris: Seuil, 2001.

- Collinot, André/Mazière, Francine: *Un prêt à parler: le dictionnaire*. Paris: PUF, 1997.
- Coulon, Alain: *L'ethnométhodologie*. Paris: PUF, 1987.
- Deleplace, Marc: La notion d'anarchie pendant la Révolution française. In: *Revue d'histoire moderne et contemporaine*, 43-2 (1996).
- Deleplace, Marc: *L'Anarchie de Mably à Proudhon (1750-1850)*. Lyon: ENSéditions, 2001.
- Dubois, Jean: Lexicologie et analyse d'énoncé. In: *Cahiers de lexicologie*, 1969-II.
- Eluerd, Roland: *La lexicologie*. Paris: PUF, Que sais-je ?, 2000.
- Faye, Jean-Pierre: *Langages totalitaires*. Paris: Hermann, 1972.
- Fiala, Pierre (éd.): *In/égalité/s. Usages lexicaux et variations discursives (18ème-20ème siècle)*. Paris: L'Harmattan, 1999.
- Foucault, Michel : *L'archéologie du savoir*. Paris : Gallimard, 1979.
- Foucault, Michel : *Dits et écrits*. Paris, Gallimard, trois volumes, 1994.
- Guilhaumou, Jacques: L'idéologie du Père Duchesne. In: Guilhaumou J. et alii, *Langage et idéologies. Le discours comme objet de l'histoire*. Paris: Editions ouvrières, 1974.
- Guilhaumou, Jacques: 'Moment actuel' et processus discursifs. *Le Père Duchesne d'Hébert et le Publiciste de la république française* de J. Roux (14 juillet - 6 septembre 1793). In: Robin Régine/ Guilhaumou Jacques (éds), *Sur la Révolution française. Bulletin du Centre d'Analyse du Discours de l'Université de Lille III*, 2, 1975.
- Guilhaumou, Jacques: La formation d'un mot d'ordre : 'Plaçons la terreur à l'ordre du jour' (13 juillet - 5 septembre 1793). In: Guilhaumou Jacques (éd.), *La rhétorique du discours, objet d'histoire (XVIIIe-XXe siècles)*. *Bulletin du Centre d'analyse du discours de l'Université de Lille III*, N° 5, 1981.
- Guilhaumou Jacques : L'historien du discours et la lexicométrie. Etude d'une série chronologique : le "Père Duchesne" d'Hébert (juillet 1793 - mars 1794). In : *Histoire & Mesure*, Editions du C.N.R.S., vol. I, 1986a.
- Guilhaumou, Jacques : La mort de Marat à Paris (13 juillet-16 juillet 1793. In : J. C. Bonnet Jean-Claude (éd.): *La Mort de Marat*. Paris : Flammarion, 1986b.
- Guilhaumou Jacques : Enoncés et récits sur la mort de Marat (juillet 1793). La matérialité de la langue dans la description de l'archive. In : *Lexique: Presses Universitaires de Lille. Lexique et faits sociaux*, N° 5, 1987.
- Guilhaumou, Jacques : A propos de l'analyse de discours: les historiens et le 'tournant linguistique' (l'exemple du porte-parole pendant la Révolution française). In : *Langage et Société*, 65, 1993.

Guilhaumou, Jacques: Révolution française et tradition marxiste: une volonté de refondation ». In: *Actuel Marx*, 20, 1996.

Guilhaumou, Jacques :*L'avènement des porte-parole de la République (1789-1792). Essai de synthèse sur les langages de la Révolution française*. Lille : Presses Universitaires du Septentrion, 1998a.

Guilhaumou, Jacques : De l'histoire des concepts à l'histoire linguistique des usages conceptuels. In : *Genèses*, 38, 2000.

Guilhaumou, Jacques :Substances (pain, bled(s), grains). In : Lüsebrink Hans /Reichardt Rolf (Hrsg.), *Handbuch politisch-sozialer Grundbegriffe in Frankreich, 1680-1820.*, Munich : Oldenbourg, Heft 19-20 2001a.

Guilhaumou, Jacques : La connexion empirique entre la réalité et le discours. Sieyès et l'ordre de la langue. In : *marges-linguistiques.com*, revue électronique, N°1, juin 2001b.

Guilhaumou, Jacques: L'histoire des concepts: le contexte en débat. In: *Annales*, N°3, mai-juin 2001c.

Guilhaumou, Jacques: L'histoire des événements linguistiques de « la Langue française » au 18ème siècle. In: Moreau-Steuckart, Agnès, *La norme lexicale*. Montpellier: Presses Universitaires de Montpellier, à paraître en 2002.

Guilhaumou Jacques, Maldidier Denise, Prost Antoine, Robin Régine: *Langage et idéologies. Le discours comme objet de l'histoire*. Paris: Editions ouvrières, 1974, p. 80-116.

Guilhaumou Jacques,/Maldidier Denise : Effets de l'archive. L'analyse de discours du côté de l'histoire. In : *Langages*, 81, mars 1986b.

Guilhaumou Jacques/ Maldidier Denise/Robin Régine : *Discours et archive. Expérimentations en analyse du discours*. Liège : Mardaga, 1994.

Habert, Benoît/Nazarenko, Adeline/Salem, André: *Les linguistiques de corpus*. Paris: Armand Colin, 1997.

Harris, Zellig: Discourse analysis. In: *Language*, 28, 1952. Traduit dans *Langages* N°13, 1969.

Heiden, Serge: Encodage uniforme et normalisé de corpus. Application à l'étude d'un débat parlementaire. In: *Mots*, N°60, 1999.

Kuentz, Pierre:le linguiste et le discours. In: *Langages*, 45, 1977.

Koselleck, Reinhart: *Vergangene Zukunft: zur Semantik geschichtlicher Zeiten*. Frankfurt/M., Suhrkamp, 1979.

Koselleck, Reinhart: L'expérience de l'histoire. Paris: EHESS/Gallimard, 1997

Lebart, Louis/Salem André: *Statistique textuelle*. Paris: Dunod, 1994.

Le Trocquer Olivier/Wahnich Sophie: La trame du discours et le fil du temps. In: *EspaceTemps, les cahiers*, N°59-60-61, 1995.

- Malidier, Denise/Robin, Régine: Polémique idéologique et affrontement discursif en 1776. In: Robin Régine (éd.): *Langage et idéologies. Le discours comme objet de l'histoire*. Paris: Editions ouvrières, 1974.
- Maingueneau, Dominique: *L'analyse de discours. Introduction aux lectures d'archive*. Paris: Hachette, 1991.
- Marcellesi, Jean Baptiste: *Le Congrès de Tours (décembre 1920). Etudes sociolinguistiques*. Paris: Roger Maria, 1971.
- Marcellesi, Jean-Baptiste: Analyse de discours à entrée lexicale. Application à un corpus de 1924-1924. In: *Langages*, 41, 1976.
- Mayaffre, Damon: *Le poids des mots. Le discours de gauche et de droite dans l'entre-deux-guerres*. Paris: Champion, 2001.
- Mesini, Béatrice/Pelen Jean-Noël/ Guilhaumou, Jacques: *La résistance à l'exclusion. Récits de vie et du Monde*. Paris: L'Harmattan, collection « Logiques sociales », 2001.
- Monnier, Raymonde: *Démocratie et révolution française*. In: *Mots*, 59, juin 1999.
- Monnier, Raymonde: Liberté/égalité dans l'événement révolutionnaire: le discours thermidorien. In: Pierre Fiala (éd.), *In/égalité/s. Usages lexicaux et variations discursives (18ème-20ème siècle)*. Paris: L'Harmattan, 1999.
- Moreau-Steuckardt Agnès: *L'eau-forte des mots. Analyse lexicale de la violence dans l'Ami du peuple*. Thèse de doctorat, Université de Paris III, dir. A. Salem, 2000.
- Pêcheux, Michel: *Les vérités de la Palice*. Paris: Maspero, 1975.
- Pêcheux, Michel: *L'inquiétude du discours*. Paris: Editions des Cendres, textes choisis et présentés par Denise Malidier, 1990.
- Piguet, Marie-France: *Classe. Histoire du mot et genèse d'un concept des Physiocrates aux Historiens de la Renaissance*. Lyon: Presses Universitaires de Lyon, 1996.
- Prost, Alain: Les mots. In: Rémond René (éd.), *Pour une histoire politique*. Paris: Seuil, 1988.
- Robin; Régine : *La Société française en 1789*. Paris : Plon, 1970.
- Robin, Régine: *Histoire et linguistique*. Paris: Armand Colin, 1973.
- Robin, Régine: Le champ sémantique de « féodalité » dans les cahiers de doléances généraux de 1789. In: *Bulletin du Centre d'analyse de discours de Lille III*, N°2, 1975.
- Salem, André: Segments répétés et analyse statistique des données textuelles. In: *Histoire & Mesure*, 1-2, 1986.
- Tournier, Maurice et alii: *Des tracts en mai 1968*. Paris: Armand Colin, 1975.
- Wahnich, Sophie : *L'étranger dans le discours de la Révolution française*. Paris : Albin Michel, 1997

Wahnich, Sophie : Puissance des concepts et pouvoir des discours. Quelques débats révolutionnaires sur la souveraineté. In : *Ethnologie française*, 4, 1999.

Wahnich, Sophie: Voix, parole, discours: dire et entendre les émotions du peuple en révolution. A paraître en 2002.

Widmer, Jean: *Langage et action sociale*. Fribourg: Editions Universitaires, 1986.