Revue *Corpus* n°8, Faculté des Lettres de Nice, 2009.

Dis-moi quel est ton corpus, je te dirai quelle est ta problématique

Patrick CHARAUDEAU
Université de Paris 13 - CAD

Les problèmes que pose la notion de corpus sont relativement bien connus pour avoir été longuement discutés dans le champs des sciences du langage[1], mais ils n'ont toujours pas donné lieu à un consensus qui aurait permis de s'en remettre à une définition faisant autorité et à laquelle on se référerait chaque fois que l'on aurait à justifier un corpus d'analyse. Cela est peut-être le symptôme de ce que le corpus n'existe pas en soi, mais dépend —comme en témoigne l'article «corpus» du *Dictionnaire d'analyse du discours*[2]— du positionnement théorique à partir duquel on l'envisage.

S'agissant du langage, on connaît les problèmes dont je ne cite que les principaux :

- le problème qui concerne le recueil des données, recueil qui dépend du choix de la matérialité langagière (paroles orales, paroles écrites), du choix du support qui véhicule ces paroles en relation avec une situation de communication (pour l'écrit : lettres, rapports, journaux, tracts, circulaires, affiches, etc. ; pour l'oral : radio, télévision, réunions diverses, meetings, conversations du quotidien, etc.). Ce sont autant d'aspects qui ont des incidences sur la manière de recueillir des données : exploration du terrain, procédés d'enregistrement libres ou contraints, au su ou à l'insu des acteurs de parole, etc. ;

- le problème qui concerne l'importance du matériel recueilli et de sa valeur de représentativité ; le corpus peut-il être considéré exhaustif et clos, ou partiel et ouvert, et, conséquemment, peut-il être considéré comme un objet en soi ou un simple outil (Mayaffre 2005). On sait que l'hypothèse de l'exhaustivité —vieux rêve de l'attitude positiviste— n'est plus tenue, et ce malgré le développement récent de la dénommée *linguistique de corpus* initiée dans le monde anglo-britannique[3], et prolongée en France par quelques auteurs avec une certaine prudence[4], car le langage est en perpétuelle production. Il peut se faire cependant que la clôture du corpus soit revendiquée, à titre expérimental, par certains analystes[5]. Si le corpus est considéré comme partiel se pose alors le problème de sa valeur comme échantillon, et de la possibilité de le faire varier en sous-corpus. Quant à considérer si le corpus est un objet en soi ou un instrument, cela renvoie à la question du contexte que nous analyserons plus loin ;

- le problème qui concerne, à l'intérieur du matériau langagier, les catégories qui vont faire l'objet de l'analyse : grammaticales (connecteurs, pronoms, verbes, etc.), lexicales (par champs ou de façon aléatoire), syntaxiques (selon divers types de construction) ; mais aussi les variables externes à la production des actes langagiers, telles que les types de locuteurs, les dispositifs de communication, de même que les variables concernant le temps (l'historicité) et l'espace (les cultures) ;

- le problème enfin qui concerne l'outil de traitement des données : dépouillements manuels, traitement informatique à l'aide de logiciels ad hoc, constitution d'échantillons à partir de bases de données (voir ci-dessous)

---

[1] Voir la Revue *Corpus* éditée par l'université de Nice-Sophia Antipolis, particulièrement les numéros 1 (2000) et 4 (2005).

[2] Charaudeau P. et Maingueneau D., *Dictionnaire d'analyse du discours*, Seuil, Paris, 2002.

[3] Voir particulièrement Williams G. (2005).

[4] Voir Habert B. et *alii* (1997) et Mayaffre (2005).

[5] Dans l'article «Lexicométrie» du *Dictionnaire d'analyse du discours*, op.cit., M. Tournier écrit :«la lexicométrie n'est pas une théorie mais une *méthodologie* d'étude du discours, qui se veut exhaustive, systématique et automatisée» ; et plus loin : «[Le corpus d'étude] Il est fermé, au moins le temps d'une expérience, car on ne peut compter que sur des ensembles stabilisés».

Dans tous ces cas, se pose la question de la pertinence de ces choix en relation avec les présupposés théoriques, et de l'éventuelle circularité qui peut s'instaurer entre ceux-ci et le corpus.

Je voudrais montrer dans cet article de quelle façon la construction d'un corpus, en analyse de discours, dépend d'un positionnement théorique lié à un objectif d'analyse, ce que j'appelle une *problématique*. On envisagera trois types de corpus selon qu'ils se réfèrent à une problématique dite *cognitive*, *communicative* ou *représentationnelle*. Puis on montrera comment un corpus de textes se construit dans un jeu de déconstruction et de reconstruction selon des critères d'*ouverture* ou de *clôture* (intertexte et hypertexte), de *contraste* (interne et externe) et de *pertinence quantitative* et/ou *qualitative*. On terminera par le renvoi à un exemple d'analyse pratiqué par le Centre d'Analyse du Discours à propos du discours d'information télévisé, afin de montrer comment on peut tirer parti de certaines de ces options, et comment elles peuvent se combiner.

Mais auparavant, il est nécessaire de circonscrire la spécificité de l'analyse du discours au sein des sciences du langage, car on ne peut traiter la question des corpus de discours sans s'interroger sur certaines notions souvent employées dans ce domaine.

De quelques distinctions indispensables pour former des corpus de discours

Dans les sciences humaines et sociales, certaines disciplines privilégient des études de terrain avec des procédures suivant une démarche descriptivo-empirique (sociologie, anthropologie), d'autres privilégient des procédures d'expérimentation (psychologie sociale), d'autres des analyses à partir d'archives ou en construisant des corpus avec des procédures plus ou moins systématiques de recueil et traitement du matériel sémiologique réuni (histoire, sciences du langage).

Les sciences du langage font donc partie des disciplines de *corpus* : rassemblement de données linguistiques (sous forme de textes écrits ou oraux, de documents divers, d'observations empiriques raisonnées ou d'enquêtes provoquées) que l'on constitue en objet d'analyse. Dès lors se pose la question de savoir quelle est la nature de ces données. On ne peut donc traiter de la question du corpus si l'on n'est pas au clair (c'est-à-dire si on ne prend pas position) sur certains notions qui sont au cœur des sciences du langage, dont j'envisagerai ici les distinctions qui me paraissent les plus fondamentales : *langue/discours, texte/discours, texte/contexte*.

Langue/discours

Vieille lune diront certains. Et pourtant, malgré Saussure lui-même déclarant que la langue n'est créée qu'en vue du discours, de nombreuses définitions de la notion de *discours* ont eu cours depuis les années soixante-dix qui ne facilitent pas la clarté de son opposition à *langue*. Renvoyons à Maingueneau qui, en 1976[6] puis 1996[7], en répertorie les définitions, et rappelons que certaines de ces définitions font du discours un prolongement de la langue, opposant l'unité de phrase à une unité «transphrastique» (Harris 1969), d'autres intègrent des catégories de discours dans la langue[8], d'autres encore opposent *discours* à *énoncé* en rapportant cette notion aux «conditions de production» d'un texte[9] ; d'autres enfin qui font de ce terme de discours un usage que Maingueneau qualifie de «paralinguistique» dans la mesure où il s'intègre dans une réflexion qui participe de plusieurs disciplines telles que l'histoire, la philosophie, la sémiologie, voire la psychanalyse[10]

6 Initiation aux méthodes de l'analyse du discours, Hachette-Université, Paris, 1976.

7 Les termes clés de l'analyse du discours, Seuil, coll. «Mémo», Paris, 1996.

8 Comme O. Ducrot qui relie cette notion à son «composant rhétorique» mais à l'intérieur de la langue, jusqu'à dire que«l'argumentation est dans la langue».

9 Voir sa mise en application dans le numéro 81 de la revue *Langages*, mars 1986, *Analyse de discours, nouveaux parcours*, Larousse, Paris, 1986.

10 A travers les écrits de Michel Foucault, Julia Kristeva et Jacques Derrida.

Ici n'est pas le lieu de discuter ces définitions. Il s'agit seulement de bien marquer la différence entre *langue* et *discours* dans ce qu'elle implique du point de vue de l'analyse et de la constitution de son objet. Ces deux notions signalent deux lieux de structuration du langage :

-) la langue, comme lieu de conformation entre des *formes* et du *sens* s'organisant en systèmes, c'est-à-dire en réseaux de relations entre des unités minimales selon des règles de combinaisons syntagmatiques et paradigmatiques, systèmes dont on pourra dire qu'ils témoignent de catégories de pensée prenant position sur des visions du monde[11]. Les unités prises en compte sont des unités phonologiques, morphologiques, sémantiques, d'ordre grammatical ou lexical, et les corpus sont constitués d'un ensemble de co-occurrences par ressemblances de forme ou de sens.

-) le discours, comme lieu, à la fois, de structuration des usages en fonction des conditions de production dans lesquels ces usages se manifestent, témoignant des comportements langagiers des sujets parlant, et de catégorisation de sens qui témoigne des systèmes de connaissance et de croyance auxquels adhèrent les individus ou groupes sociaux. Ici, on n'est plus dans une combinaison d'unités, seraient-elles transphrastiques, car si l'on est dans un au-delà de la phrase, le sens de discours résulte d'une multiplicité de facteurs d'ordre contextuel (voir ci-dessous) qui font que le sens n'est repérable dans aucune des unités qui composent une production discursive quelconque, et ne peut être inféré que par le jeu des combinaisons de tous ces facteurs. Pour reprendre un mot de Roland Barthes, le sens s'attache aux formes de façon «diffractée»[12]. Un corpus de discours ne peut être constitué que par un ensemble de productions langagières en situation d'usage.

De là, la constitution de deux linguistiques, l'une que l'on appellera «linguistique de la langue» orientée vers la description des systèmes intrinsèques à chaque langue, à corpus plus ou moins clos, et qui présuppose l'existence d'un sujet opérateur de catégories «cognitivo-linguistiques» ; l'autre que l'on appellera «linguistique du discours» orientée vers la description des usages et des significations sociales, à corpus, par définition ouvert, et qui présuppose un sujet opérateur de catégories «sociodiscursives» et porteur d'imaginaire social[13].

Prenons un exemple pour illustrer cette différence. Un énoncé comme «J'ai trente ans» pourra faire l'objet d'une analyse linguistique de langue lorsque, inséré dans un corpus d'énoncés similaires conformes au schème :

< Pr. x Vb. x quant. x Sb. >, on en déterminera l'actant (*celui qui parle*) auquel est attribué (*avoir*) une certaine propriété (*ans*), laquelle est quantifiée (*trente*), le tout dans un acte d'énonciation qui dit que cette assertion doit être rapportée au sujet parlant lui-même (*Je*) dans une modalisation élocutive affirmative ; de plus, en observant un certain paradigme lexical, on dira que ce «ans» désigne un certain segment du temps qui renvoie à l'âge de la vie. Mais globalement, «J'ai trente ans» ne signifiera ici rien d'autre que /*j'ai trente ans*/.

Si l'on considère maintenant cet énoncé dans son contexte de production, à savoir un sportif répliquant à la réflexion d'un ami qui s'étonne de le voir quitter la compétition : «J'ai trente ans», alors cet énoncé signifie entre autres choses : /*Je suis trop vieux*/. Mais s'il s'agissait d'une personne qui, venant d'être licenciée d'une entreprise, répond à un ami lui suggérant que c'est peut-être parce qu'il a passé l'âge, alors le «J'ai trente ans» signifie : /*Mais je suis encore jeune !*/ Ces sens dépendent évidemment de la suite des mots contenus dans l'énoncé, mais ils ont été obtenus par opération d'inférence à partir de la connaissance que l'interlocuteur peut avoir de l'identité de celui qui lui parle

---

11 Voir la tradition de l'analyse linguistique française depuis A. Martinet en passant par G. Guillaume, B. Pottier, R. Martin, A. Culioli.

12 *Roland Barthes par roland barthes*, Le Seuil, Paris, 1975. Je me permets de renvoyer également le lecteur aux idées de «texte étoilé» et «texte brisé» que R. Barthes développe dans *S/Z*, Le Seuil, 1970.

13 Je ne peux discuter ici la notion d'imaginaire pour laquelle je renvoie provisoirement le lecteur à ce que j'en dis dans mon Le discours politique. Les masques du pouvoir, Vuibert, Paris, 2005, et dans «Tiers où es-tu ?», in *La voix cachée du Tiers. Des non-dits du discours*, L'Harmattan, Paris, 2004.

(*sportif/employé*) et de l'univers de sens dont il est à chaque fois question (*sport/entreprise*), autant d'éléments inclus dans un corpus mémoriel de discours.

Reste que le sens de discours, s'il est inféré, l'est à partir de formes dont la combinaison constitue un texte. Il s'agit donc de savoir si *texte* et *discours* sont une seule et même chose.

Texte/discours

Étant donné le nombre d'écrits qui existent sur ces notions, voire sur cette distinction qui n'est pas toujours opérée, on ne pourra s'étendre longuement. Seulement faire un constat et prendre position, car cela est nécessaire pour déterminer un corpus de discours.

Le constat : rarement sont opposés de façon nette *texte* et *discours*. Au point que des ouvrages se réclamant de l'«analyse textuelle» emploient souvent le terme «analyse de discours» et inversement. Conscient de la chose, Jean-Michel Adam propose de «distinguer *texte* et *discours* comme les deux faces complémentaires d'un objet commun pris en charge par la *linguistique textuelle*»[14], précisant que celle-ci privilégie l'organisation du *cotexte* selon des critères de cohésion, alors que l'*analyse du discours* s'occupe davantage du *contexte* selon des critères de cohérence. Cette distinction pointe bien que l'objet *texte* se définit par l'organisation de sa configuration en rapport avec ce qui l'entoure et que le discours concerne davantage l'organisation sémantique, même si le premier est aussi porteur de sens et le second se repère à travers des formes. Mais cette distinction ne me semble pas assez nette pour distinguer ensuite corpus de textes et corpus de discours. D'autant que si l'on fait intervenir la série des notions proposées par G. Genette : *paratexte, métatexte, épitexte, intertexte* et *hypertexte*, à quoi on peut ajouter la notion d'*interdiscours* introduite par l'école française d'analyse du discours des années soixante-dix/quatre-vingt[15], on ne voit pas toujours bien où se trouve le frontière entre ces deux notions.

La prise de position. Je ne pense pas avoir un point de vue très différent de celui de Jean-Michel Adam, sauf à considérer que le rapport de complémentarité entre analyse de texte et analyse de discours n'est pas de symétrie. Ce ne sont pas les deux faces d'une même pièce, parce qu'il y a *des* discours dans un même texte, et un même discours dans plusieurs textes. Le discours n'est pas le texte mais il est porté par des textes. Le discours est un parcours de signifiance qui se trouve inscrit dans un texte, et qui dépend de ses conditions de production et des locuteurs qui le produisent et l'interprète. Un même texte est donc porteur de divers discours et un même discours peut irriguer des textes différents. Du discours traverse des textes différents, et un même texte peut être porteur de discours différents. Pour comprendre le slogan publicitaire : «Obernai, la première grand bière avec un tiers de calories en moins», je suis amené à mobiliser une série de discours qui ne sont pas dits explicitement dans ce slogan, mais qui s'y trouvent implicitement, à savoir : /*je dois avoir des connaissances en diététique*/>/*cette bière ne fait pas grossir*/>/*la minceur est valorisée dans ma société*/>/*les femmes peuvent boire cette bière, traditionnellement réservée aux hommes*/. Cela n'empêche pas que par ailleurs les discours sur la diététique, la minceur et sa valeur culturelle puissent se trouver configurés dans d'autres types de texte des plus savants aux plus vulgarisés.

Comme le discours a besoin de configuration textuelle pour signifier, cela veut dire que cette signifiance, à un moment donné, a été texte. Il se produit alors un phénomène étrange de va-et-vient entre différents textes se faisant écho, au terme duquel se construit une signifiance abstraite qui se trouve dans différents textes sans être uniquement l'un d'entre eux. Pour prendre un exemple simple, le discours du péché originel et l'imaginaire de culpabilité et de rédemption dont il est porteur s'est construit en Occident du fait de l'«échoïsation» entre divers textes (Bible, Évangiles, écrits des

14 Voir l'article «Texte» du *Dictionnaire d'analyse du discours*, op.cit.

15 Voir Pêcheux M. *Les Vérités de La Palice. Linguistique, sémantique, philosophie*, Maspero, Paris, 1975 ; et Courtine J.-J. «Quelques problèmes théoriques et méthodologiques en analyse du discours. A propos du discours communiste adressé aux chrétiens», revue *Langages* n°62, Larousse, Paris, 1981.

théologiens, commentaires de philosophes, prédications de l'Église, jusqu'à des textes publicitaires, des magazines people, des films, etc.). Ainsi se déposent des discours dans la mémoire collective.

Texte/contexte

A considérer le discours comme un lieu de signifiance abstrait se configurant diversement dans des textes, on retrouve les notions de d'*intertextualité* de Genette et de *dialogisme* de Bakhtine. Partant de cette dernière notion qui a donné lieu à diverses interprétations[16], je dirai que ce jeu d'échos peut s'opérer entre des textes mais aussi entre des discours. Dans le premier cas, on aura affaire, au sens strict du terme, à une *intertextualité* qui se manifeste dans divers types de citation — ce que Jacqueline Authier-Revuz nomme «l'hétérogénéité montrée»[17] — : il y a du texte dans du texte, même partiel, même déformé. Dans le second cas, on aura affaire à une interdiscursivité, «un jeu de renvois entre des discours qui ont eu un support textuel mais dont on n'a pas mémorisé la configuration»[18]. Cela nous explique qu'un corpus ne peut être que de textes et non point de discours, à moins de décider que les textes réunis le sont au nom d'un certain discours. Mais n'anticipons pas, je reviendrai là-dessus. Nous voilà, en tout cas, arrivé à la notion de *contexte*.

Depuis, les premiers travaux en linguistique de la langue, la notion de contexte s'est considérablement étendue avec des dénominations variables. En résumé, on parlera : de contexte strictement *linguistique*, lorsqu'il s'agit de se référer à l'environnement verbal qui se trouve avant et/ou après une unité déterminée ; de contexte *textuel*, lorsqu'il est constitué de textes produits par une même source (par exemple, les divers écrits d'un même auteur) ; de contexte *paratextuel* (Genette), lorsqu'il est constitué de textes ou fragments de textes se trouvant en coprésence dans un même espace scriptural (les titres, sous-titres, chapeaux, légendes d'une page de journal) ; de contexte *métatextuel* (Genette), lorsqu'il est constitué de textes qui se commentent les uns les autres ; de contexte *hypertextuel* (Genette), lorsqu'il est constitué de textes qui se citent, se renvoient les uns aux autres, se reprennent et se transforment, comme sont les pastiches et parodies ; enfin, de contexte *intertextuel* et *interdiscursif* (au sens où je viens de définir ces notions), lorsque sont convoqués des textes et discours pour justifier les interprétations du sujet interprétant (ou analysant). Autrement dit, dans cet élargissement progressif de la notion de contexte, apparaît une prise de conscience progressive, non seulement du rapport entre texte et tout l'environnement textuel qui peut s'y rapporter, mais aussi entre le texte et un «hors-texte» (parfois appelé *cotexte*), c'est-à-dire des données présentes dans les *conditions de production* de l'acte de langage.

Ce que je voudrais souligner ici, c'est un aspect, qui est rarement pris en compte par les différentes courants d'analyse du discours, et qui me permet d'introduire la notion de contexte *situationnel*. L'aspect qui concerne, dans les conditions de production, non seulement les présupposés de positionnement interdiscursif (présupposés idéologiques pour certains), mais aussi le conditionnement de la situation de communication elle-même : la nature de l'*identité* des partenaires de l'acte de langage, la *finalité* de la situation, les *dispositifs* et les *circonstances matérielles* de celle-ci. Cet ensemble constitue ce que j'appelle un *contrat de communication*, lequel contraint le sujet parlant en lui donnant les instructions discursives qu'il devra suivre pour procéder à son acte d'énonciation[19]. Certes les ethnographes et ethnométhodologues de la communication interactionnelle nous ont habitué à prendre ces variables en considération pour analyser les conversations, mais elles ne sont pas suffisamment prises en compte dans l'analyse du discours[20].

16 Voir l'article «Dialogisme» du *Dictionnaire d'analyse du discours*, op.cit.

17 Authier-Revuz J., «Hétérogénéité montrée et hétérogénéité constitutive : éléments pour une approche de l'autre dans le discours», *DRLAV*, 26, 91-151, 1982.

18 Parce que, pour reprendre ce que dit D. Maingueneau, «Tout discours (…) a pour propriété constitutive d'être en relation multiforme avec d'autres discours», voir l'article «Interdiscours» dans le *Dictionnaire d'analyse du discours*, op.cit.

19 Charaudeau P. (2004).

20 Et pourtant, de mon point de vue, l'analyse conversationnelle fait partie de l'analyse du discours.

C'est que l'analyse du discours a surtout travaillé sur l'hypothèse que la force du langage se trouve davantage dans ce que disent les mots que dans ce qu'ils transmettent. Et donc on s'est longtemps penché sur ce que dit le langage sans s'interroger sur le pouvoir du langage comme effet d'influence. Autrement dit, pour reprendre une observation de Roland Barthes, on pense toujours que le signe «signifie» et on oublie qu'il «signifie à»[21]. Or le phénomène de signifiance résulte des deux : le langage signifie en même temps qu'il transmet ; c'est dans l'acte même de transmission qu'il signifie avec l'effet qu'il produit sur l'autre, c'est-à-dire son interprétation. Ne pas se poser la question des effets, c'est amputer l'acte de langage de sa signification phénoménale, à savoir qu'il est le résultat d'une co-construction. On ne peut saisir le sens d'une expression comme «préférence nationale» que si on la considère dans son contexte situationnel, lequel nous dira *qui* l'emploie (Jean-Marie Le Pen), à l'adresse de *qui* (citoyens-électeurs), dans quel *dispositif* communicationnel (campagne électorale), avec quelle *finalité* (séduire), et, évidemment, en le rapportant à une certaine tradition historique française (*interdiscours*).

Le problème qui se pose alors est celui de l'articulation entre, d'une part ce à quoi renvoient les discours du point de vue des valeurs, et d'autre part les effets qu'ils sont susceptibles de produire. Par exemple, le pouvoir d'influence des mots du discours politique n'est pas seulement dans la force du projet d'idéalité sociale dont ils sont porteurs, mais aussi, et de plus en plus, dans leur écho aux systèmes d'attente des citoyens (effet de miroir), lesquels sont davantage sensibles au charisme des personnalités politiques, à l'émotion que celles-ci suscitent, qu'aux idées.

Le contexte situationnel doit donc être pris en compte à l'égal des autres pour former des corpus. Évidemment, aucun corpus ne peut prendre en compte la totalité des ces contextes, mais c'est de la prise en considération de cette multiplicité contextuelle que l'on pourra construire des corpus de façon raisonnée.

Spécificités du corpus en analyse du discours

A considérer ces différentes distinctions : le discours comme signifiance/le texte comme configuration ; le contexte sous ses trois aspects (*paratextuel*, *interdiscursif* et *situationnel*)[22], on est conduit à justifier le fait qu'un corpus de discours soit toujours une construction résultant de divers types de regroupement : corpus selon le *paratexte* (de mots, d'énoncés, de modes d'énonciation), corpus selon l'*interdiscours* (savoirs de connaissance, savoirs de croyance), corpus selon la *situation* (locuteurs, finalité et dispositif). Ces regroupements se font en fonction de la *problématique d'analyse* et de la *mise en contraste* choisies.

Corpus et problématique d'analyse

On n'utilisera pas le terme *problématique* avec le sens qu'il a dans l'usage courant, comme dans l'expression «C'est problématique» pour signifier qu'une question est difficile à expliquer, voire sans solution, ou qu'une affirmation est incertaine, douteuse ou parfois suspecte. Une problématique (emploi substantif) est un ensemble cohérent de propositions hypothétiques (ou de postulats) qui, à l'intérieur d'un champ d'étude, déterminent à la fois un objet, un point de vue d'analyse et un questionnement par opposition à d'autres questionnements possibles. On ne confondra donc pas

21 *Roland Barthes par roland barthes*, op.cit..

22 C'est que, au bout du compte, on peut attribuer à *contexte* une fonction englobante se composant du contexte *paratextuel*, qui prend en compte tout ce qui est coprésent dans un même espace textuel, du contexte *interdiscursif*, qui prend en compte tout texte (y compris les *métatextuels*) et discours se référant à un même univers de discours, et du contexte *situationnel* qui prend en compte les conditions énonciatives, quitte à opérer, pour autant que de besoin, des distinctions au sein de chacun d'eux.

problématique et *théorie* qui se différencient en ce que cette dernière a un corps de concepts et de catégories plus ou moins modélisé s'accompagnant d'un cadre méthodologique.

En observant les études qui se réclament de près ou de loin du discours, et en les interrogeant selon le type d'*objet* qu'elles analysent et la conception qu'elles se font du *sujet* qui en est l'ordonnateur, on peut repérer trois grandes problématiques qu'on appellera *cognitive*, *communicative* et *représentative*, chacune déterminant un type de corpus.

Une problématique dite cognitive et catégorisante :

L'*objet* d'étude est considéré comme un ensemble de *mécanismes discursifs* dont il s'agit soit de repérer l'existence et le mode d'agencement à l'intérieur d'une production discursive quelconque (texte ou énoncés aléatoires), soit de décrire le comment de leur fonctionnement, et particulièrement leur mode de production. Dans le premier cas, on étudie, par exemple, les jeux de cohérence et cohésion dans un texte, ou certains mots du discours (connecteurs, modalités). On trouve le second cas dans les travaux de certains psycho-cognitivistes du langage qui cherchent à décrire des parcours d'organisation discursive «down» ou «up» de réalisation des schèmes, scripts et scénarios. Dans ces deux cas, il en sort des catégories instrumentales d'analyse qui servent à repérer ou à produire des configurations discursives.

Le *sujet* concerné par la détermination d'un tel objet sera appelé *cognitif* dans la mesure où n'est pris en compte chez celui-ci que son aptitude à produire ou repérer des opérations d'articulation discursive : relations anaphoriques ou cataphoriques, connexions coordonnées ou subordonnées, relations de cohérence et de cohésion, de progression et de non contradiction, etc., sans aucune prise en compte de sa nature sociale.

Dans une telle problématique, le *corpus* n'a pas besoin d'être finalisé en fonction d'une situation de communication particulière. De ce point de vue, il peut être dit *aléatoire*, et seulement contraint par les contextes linguistiques dans lesquels apparaissent ces mécanismes. On a affaire à un corpus de co-occurrents par catégories d'opération linguistico-discursive.

Appartiennent à cette problématique, des travaux dans la ligne des théories de la présupposition[23] et du sens des mots du discours[24], des topoï et de l'orientation argumentative[25] et de la pragmatique (les connecteurs, les embrayeurs et les actes de parole)[26], ainsi que les travaux en grammaire de texte qui traite des problèmes de cohérence et cohésion textuelles[27].

Une problématique dite communicative et descriptive

L'*objet* d'étude est ici empirique, ou plus exactement il est déterminé à partir de l'observation (plus ou moins naïve) des manifestations du monde phénoménal. Par exemple, nous pouvons observer que les comportements des individus dans leurs actes d'échange social ont une certaine récurrence ce qui permet de faire l'hypothèse qu'ils sont soumis à des conditions de réalisation et qu'ils obéissent à des règles ; c'est cela qui autorise à les constituer en objet d'analyse. Dès lors, l'objet d'étude issu de cette empirie peut être structuré en «types idéaux» (au sens durkhemien) de communication à partir, et à l'intérieur, desquels sont décrites les caractéristiques discursives qui s'attachent à chacun d'eux (*identité* des locuteurs, *finalité*, *dispositif*). Une telle démarche aboutit, entre autres choses, à l'établissement de diverses typologies dites de discours, de

---

[23] Ducrot O., *Dire et ne pas dire*, Hermann, Paris, 1991.

[24] Ducrot O. et alii, *Les mots du discours*, Minuit, Paris, 1980.

[25] AnscombreDucrot O., *L'argumentation dans la langue*, Mardaga, Bruxelles, 1983.

[26] On pourra, sur ces questions, consulter le *Dictionnaire encyclopédique de Pragmatique* de J. Moeschler et A. Reboul, Seuil, 1994.

27 Charolles M., «Cohérence, cohésion et pertinence du discours», revue *Travaux de linguistique*, n°29 (125-151), 1995 ; et l'on pourra se référer aux travaux de Van Dijk T.A. et Kintsch W., particulièrement : *Strategies of discourse compréhension*, New York, Academic Press, 1983.

textes, de genres, ou de situations communicatives, mais qui supposent toutes une théorisation de ces types idéaux de communication. Cette problématique peut être appelée descriptive du fait qu'elle vise à décrire des types de situation communicative.

Le *sujet*, ici, est lié à cette empirie des échanges communicatifs, mais il est lui aussi construit et théorisé en fonction de la façon dont on construit et théorise ces échanges. Le sujet est donc un sujet de communication qui se définit par son identité psychologique et sociale, par un comportement qui est finalisé à la fois par les contraintes qu'il subit s'il veut s'insérer dans l'échange (de ce point de vue il est un «on», un sujet de la norme contractuelle), et par ses propres intentions vis à vis de l'autre (de ce point de vue il est un «je», un sujet s'individuant).

Le *corpus* se ressent de cette vison empirique de l'objet d'étude. Il est généralement constitué de textes (productions langagières en situation) qui sont regroupés en fonction de leur appartenance à tel ou tel type de situation. Par exemple, textes publicitaires, textes journalistiques, textes de manuels scolaires, textes administratifs, textes de programmes politiques, et divers textes conversationnels (échanges téléphoniques, demandes de renseignements, interviews, entretiens, débats, etc.). Ce type de regroupement permet ensuite de procéder à des comparaisons autour d'un même contexte *paratextuel* et *situationnel*.

Appartiennent à cette problématique, des travaux autour de la description des genres du discours et des objets empiriques catégorisés en situation de communication[28]

Une problématique dite représentationnelle et interprétative

L'*objet* d'étude de cette problématique est défini à travers des hypothèses de représentations socio-discursives dont on suppose qu'elles sont dominantes à un moment donné de l'histoire d'une société (elles sont donc socio-historiques), et qu'elles caractérisent tel ou tel groupe social. En cela elles sont interprétatives, puisqu'il faut bien avoir, au départ, une hypothèse sur ce que sont les «positionnements sociaux» en relation avec les «pratiques discursives» et les «types de sujets» qui s'y attachent. La difficulté de cette problématique réside justement dans le fait qu'il faut faire ces hypothèses et que pour ce faire on doit s'appuyer sur des manifestations discursives extrêmement diverses, non toujours explicites, voire diffuses et même floues, ce qui est un problème pour la constitution du corpus.

Le *sujet*, dans cette perspective, fait lui aussi problème et a été souvent discuté. Du fait qu'il est à l'origine des pratiques discursives qui construisent les représentations, il peut être considéré comme un sujet *passif* ou *actif*. De là, deux positions :

- l'une, radicale, qui ne confère à ce sujet d'autre existence que celle d'une «illusion»[29], car il serait complètement surdéterminé par ce que Pêcheux appelle le «pré-construit» des «formes discursives». Le sujet ne s'appartient pas, il est parlé par un ailleurs[30] qui sera appelé tantôt idéologie (Althusser), tantôt inconscient (Authier-Revuz) en se diluant dans la conscience du groupe social. Dans un cas comme dans l'autre, le sujet n'est pas une «je» mais un «ça» (idéologique ou inconscient) qui parle à travers le sujet parlant, et ce n'est pas un hasard si les tenants de cette position ont toujours opéré un parallélisme entre idéologie et inconscient.

- l'autre position est moins globalisante. Elle ne nie pas le jeu de la surdétermination du sujet, mais au lieu d'en faire une illusion, elle en fait quelque chose de positif. Il existe dans toute société des discours de doxa (lieux commun, stéréotypes, savoirs partagés). Ceux-ci circulent et se répartissent de façon non aléatoire dans les différents groupes sociaux. Tout sujet est donc porteur,

28 Branca-Rossof S., «Les lettres de réclamation adressées au service de la redevance», revue *Langage et société* n°81, 1997 ; et Charaudeau P., «Visées discursives, genres situationnels et construction textuelle», in *Analyse des discours. Types et genres*, Éd. Universitaires du Sud, Toulouse, 2001.

29 Pêcheux dit se situer dans un champ socio-historique «où se construisent les sens, dans l'obscurité inévitable des illusions du sujet», in «Mise au point et perspectives à propos de l'Analyse Automatique du Discours», revue Langages n°37, Larousse, Paris, 1975.

30 Voir «l'effet Munchausen» décrit par Michel Pêcheux dans *Les Vérités de La Palice*, op.cit.

pour une part, de certains de ces discours, et du même coup cet appariement révèle son positionnement social. On peut donc dire que, d'une certaine manière, ce sujet est «responsable» de ses représentations, ce qui ne veut pas dire qu'il en soit conscient. Font écho à cette position que nous qualifierons de sociologique, les travaux de certains sociologues français qui s'inscrivent dans une sociologie constructiviste[31] , ainsi que ceux qui travaillent à analyser les représentations de l'espace public dans les médias[32].

Le *corpus* varie selon l'une ou l'autre position. Dans la première, le corpus est constitué d'un ensemble de *textes-archives* qui sont sélectionnés pour leur valeur emblématique de discours dominant, le plus souvent liés à une institution sociale : dictionnaires, encyclopédies, productions discursives de groupes constitués (cahiers de doléances, partis politiques, syndicats) et toutes productions ayant valeur de discours fondateur[33].

Dans la deuxième position, le corpus est constitué d'un ensemble de ce que l'on appellera des *signes-symptômes* qui représentent de façon emblématique des systèmes de valeurs. Ces signes peuvent être des mots («racisme», «immigration», «solidarité» etc.)[34] ou des formules diverses («purification ethnique»)[35] révélatrices de manières de dire ; ce peut être également des signes iconiques (des mises en scène d'images) qui permettent d'étudier par exemple les «représentations de la femme»[36] dans les publicités ou de la «violence» à la télévision[37].

Le corpus est ici d'ordre *interdiscursif*, mais il pourrait inclure d'autres composantes dans la mesure où l'on peut faire l'hypothèse que les divers éléments d'une situation de communication (identités, finalité, dispositifs) témoignent également des valeurs et des idéologies. Par exemple, si l'on compare la façon dont la presse concevait son rubriquage et ses genres d'écriture il y a 50 ans et maintenant, on peut voir que la façon d'envisager son rôle dans la symbolique démocratique a changé. Ce qui nous renvoie à la problématique précédente.

La mise en contraste : un jeu de déconstruction/reconstruction

Suivant l'hypothèse saussurienne qui dit que le *sens naît de la différence*, le sens de discours ne peut être saisi que dans ce qui fait contraste, du fait de la mise en regard de diverses séquences discursives. Ainsi les corpus doivent-ils être construits selon certaines variables permettant de les comparer, des variables *externes* ou *internes*.

Les variables *externes* permettent de mettre en regard des ensembles textuels ou discursifs appartenant à des époques différentes (variable temporelle) : par exemple, la presse du XIX° siècle comparée à celle d'aujourd'hui. Ou bien, ces ensembles peuvent appartenir à des espaces différents (variable spatiale et culturelle) : par exemple, la publicité française comparée à la publicité brésilienne. Ou encore ils peuvent appartenir à des dispositifs situationnels différents (variable de genre) : par exemple, le discours politique et le discours publicitaire afin de savoir si les stratégies de persuasion sont les mêmes. Ces corpus contrastés permettent de trancher dans les discussions qui opposent divers points de vue : celui qui défend la pérennité des discours («c'est toujours la même chose à travers le temps et l'espace»)[38], opposé à celui qui défend à tout prix le changement des

---

[31] Voir à ce propos *Choses dites* de Pierre Bourdieu.

[32] Quéré L. (1995).

33 Voir, ce que Dominique Maingueneau et Frédéric Cossutta appellent des «textes constituants» (religieux, scientifiques, philosophiques, littéraires, juridiques, etc.) dans «L'analyse des discours constituants», revue *Langages* n°117, Larousse, Paris, mars 1995.

[34] Boyer H. (1983) ; et voir *Sans distinction… de race*, le numéro 33 de la revue *Mots*, Presse de la Fondation nationale des Sciences politiques, Paris, décembre 1992.

35 Krieg-Planke A. (2009).

[36] Soulages J.-C. (2003).

37 Boyer H. et Lochard G. (1998).

38 Par exemple, comme on l'a entendu à propos de la dernière campagne électorale : «les campagnes se ressemblent toutes, les discours électoraux ne changent pas quelle que soit l'époque».

discours («ce n'est jamais pareil selon les époques)[39] ; ou encore le point de vue qui globalise divers genres discursifs : «tout ça (discours politique, discours publicitaire), c'est de la propagande», opposé à celui qui les distingue. Ressemblances et différences entre les caractéristiques de divers corpus permettent de prendre acte des changements ou de la pérennité des discours, de leur généralité ou de leur spécificité quant au contexte de production ou quant au genre.

Mais il y a aussi des variables dites *internes* dans la mesure où elles se situent à l'intérieur d'un même champ de discours. Elles correspondent aux composantes situationnelles qui structurent un domaine de pratiques sociales. Le champ politique, par exemple, se compose de divers domaines d'activité : ceux du gouvernement, des parlements, des partis, des Conseils (constitutionnel, d'État, économique et social), etc. La parole n'y circule pas de la même façon, les normes contractuelles ne sont pas les mêmes, les stratégies discursives ne s'y déploient pas de la même manière, partant, le sens des mots varient selon ces situations. Sans compter le fait que le discours politique n'est pas le même non plus selon que les acteurs se trouvent en situation de conquête du pouvoir (campagnes électorales) ou d'exercice du pouvoir, la première conduisant à produire des discours de *promesse* qui sont de l'ordre du souhaitable, la seconde des discours de *justification* qui sont de l'ordre du possible. Il conviendra alors de construire à l'intérieur d'un vaste corpus politique des sous-corpus à contextes situationnels divers.

Il est également une autre façon d'opérer des contrastes à variables internes. C'est celle qui consiste à procéder à des *rapprochements* entre textes traitant de mêmes notions. C'est le travail que l'on est conduit à faire lorsque, analysant des déclarations d'acteurs politiques, on cherche à comprendre le sens attribué à des mots comme «autorité», «égalité», «devoir» et «droit». Il est alors nécessaire d'avoir recours à des écrits philosophiques traitant de la république ou de la démocratie[40]

Cette variété de corpus selon les diverses problématiques auxquelles ils se rattachent et selon un jeu de déconstruction/construction aboutissant à des sous-corpus, nous renvoie à deux questions qui sont largement discutées en analyse du discours : la question de l'*ouverture/clôture* du corpus ; la question de son traitement *quantitatif/qualitatif*.

Une question heuristique : l'ouverture/clôture du corpus

En résumant les considération précédentes, on dira qu'un corpus dépend du traitement que l'on a décidé de lui appliquer (sous influence d'une problématique et d'un certain jeu de construction contrastée). On pourra dire qu'il s'agit là d'une circularité peu démonstrative, mais comment faire autrement ? On n'ira pas jusqu'à dire que «C'est finalement le corpus qui fait la théorie»[41], car ce serait nier le double mouvement déductivo-inductif de la recherche en sciences humaines et sociales, mais on dira : «Dis-moi quel est ton corpus, je te dirai quelle est ta problématique / dis-moi quelle est ta problématique, je te dirai quel est ton corpus».

Le corpus participe ainsi d'une démarche heuristique, en ce que celle-ci propose un certain cheminement intellectuel à partir d'hypothèses de base pour découvrir «des faits et des idées»[42], c'est-à-dire pour *interpréter*. Or, interpréter est une opération délicate qui ne peut se faire qu'après coup, même lorsque dans une démarche plus expérimentale on fait des hypothèses préalables. Interpréter consiste en une opération de mise en relation des faits observés avec d'autres faits pour en tirer, par inférence, des hypothèses de sens. Mais il est une autre activité interprétative, celle qui consiste à

39 Par exemple, comme on l'a également entendu au cours de la même campagne électorale : «Cette campagne n'est pas comme les autres»; «On n'a jamais aussi peu parlé des questions de fond».

40 C'est ce à quoi je me suis employé dans l'analyse de la dernière campagne électorale : *Entre populisme et peopolisme. Comment Sarkozy a gagné*, Vuibert, Paris, 2008.

41 Dalbera J.-Ph. (2002), «Le corpus entre données, analyse et théorie», Corpus 1 : 99.

42 Voir le *Dictionnaire de philosophie* (entré «Heuristique»), de Christian Godin, Fayard, Paris, 2004.

mettre en relation les résultats d'une analyse descriptive avec ceux d'autres analyses : ceux d'autres corpus connexes (confrontation des articles de différents journaux pour en interpréter les ressemblances et différences) ; ceux de corpus de textes d'un même domaine mais de situations différentes (confrontation des écrits journalistiques de différentes époques) ; ceux, enfin, des analyses proposées par d'autres disciplines sur le même domaine discursif (philosophie, histoire, sociologie, psychologie sociale), sur, par exemple, le domaine politique.

Dans les sciences humaines et sociales la démarche est double : «empirico-descriptive» et «hypothético-déductive». L'une ne va pas sans l'autre : la première est davantage dépendante d'outils méthodologiques, la seconde de concepts fondateurs et de catégories explicatives. La première se déroule selon un mouvement centripète, la seconde selon un mouvement centrifuge, ce qui explique que les corpus qui s'inscrivent dans ces mouvements tendent, tantôt à se fermer sur eux-mêmes, tantôt à s'ouvrir.

C'est pourquoi un corpus n'est jamais qu'un *prétexte*, au sens de ce qu'il est un point de départ indispensable pour une analyse du discours, les textes devant être confrontés en permanence à d'autres textes et d'autres corpus. Car les significations sociales se répartissent dans un large combinatoire des mots en rapport avec leurs différents contextes. C'est également pourquoi aucun corpus n'est jamais définitivement fermé, sa clôture ne pouvant être que le fait d'une décision provisoire à des fins opératoires. Ce sont les paramètres de «contrastivité» (externe et interne) du corpus qui par le jeu des ressemblances/différences font sens.

Comment faire, dès lors, entre la conception d'un corpus complètement fermé sur lui-même, mais dont on sait qu'il a besoin de se confronter à d'autres corpus pour signifier, et la conception d'un corpus déclaré ouvert, mais dont il faut chercher à maîtriser les horizons sous peine de ne pouvoir le rendre opératoire ?

Je procède pour ma part selon la méthode de l'escargot : partir d'un premier corpus noyau déterminé selon des paramètres de temps, d'espace, de genres, de dispositifs, de locuteurs, de thèmes, etc., et ce en fonction des objectifs d'analyse que l'on se propose ; puis étendre progressivement ce corpus en le confrontant à d'autres, autant que de besoin, en fonction des questions qui surgissent au fur et à mesure des analyses.

J'en donnerai pour exemple l'étude des discours d'une campagne électorale. D'abord, on construit le corpus des déclarations du candidat X dont on fait l'analyse selon différentes catégories. Puis on peut confronter ces résultats aux corpus des déclarations des candidats concurrents pour en repérer les différences et les jeux de renvois des uns aux autres, une sorte de jeu de miroir, comme le dit D. Mayaffre qu'il désigne par l'expression «corpus réflexifs»[43]. On peut aussi comparer ces résultats aux corpus des candidats du même parti des élections antérieures pour repérer en quoi ce candidat suit la même doctrine que ses prédécesseurs ou s'en différencie, et l'on peut même élargir encore la comparaison en y intégrant le corpus des déclarations et écrits qui concernent l'idéologie d'appartenance du candidat afin de mesurer les degrés de filiation politique ; et l'on peut faire ce même travail de confrontation pour chacun des autres candidats. On peut aussi comparer les prestations d'un même candidat selon les situations de communication dans lesquelles il se trouve (meeting, interviews journalistiques, déclarations télévisée, débats, etc.), afin de vérifier si ses stratégies de persuasion sont les mêmes ou différentes dans chacun de ces cas.

Tout corpus renvoyant à d'autres corpus dans un enchaînement de contextes *situationnels*, *paratextuels* et *interdiscursifs*, se construit progressivement sous forme de ce que G. Genette appelle une *architextualité*, notion reprise par les tenants d'un traitement informatisé des textes[44]. Un jeu de *réflexivité* entre les corpus dont on voit qu'il ne peut être qu'ouvert, car c'est la découverte progressive des questions qui se posent lors des analyses issues de ces confrontations qui renvoie à d'autres textes et donc à d'autres corpus.

---

43 Voir : «Les corpus réflexifs : entre architextualité et hypertextualité», Corpus 1 : 51-69.

44 *Ibid*.

La question du traitement quantitatif/qualitatif du corpus

Je partage donc le point de vue de ceux qui disent que le corpus n'est pas l'*outil* de la recherche mais l'*objet* de la recherche[45], du moins l'objet que l'on a construit en corpus en fonction de présupposés théoriques et méthodologiques. Se pose alors la question du mode de traitement du corpus, corrélativement à la question de la pertinence d'une analyse quantitative au regard de l'analyse qualitative.

De ma pratique de sujet analysant et de la méthode d'analyse que nous avons éprouvée dans les travaux du Centre d'analyse du discours[46] sur des corpus médiatiques, j'en tire qu'il faut distinguer ce qui est de l'ordre du *recueil des données*, du *relevé des récurrences*, du *repérage des conditions de communication*.

Le *recueil des données* relève d'une logique documentaire. L'objectif est de recueillir, décrire et classer les objets du monde et les phénomènes sociaux, selon des critères de forme, sans hypothèses particulières orientant la recherche dans une direction ou une autre, et ce avec un idéal d'exhaustivité. Cela produits classements archivistiques divers, catalogues et bases de données.

La démarche est ici «empirico-inductive», s'appuyant sur une méthodologie techniciste : les marqueurs de description et les catégories de classement sont pensés simultanément avec les caractéristiques de l'outil technologique (la performance de l'outil peut influer en retour sur les marqueurs et les catégories), ainsi que les modes de représentation de ces classements (arbres, nuages statistiques, tableaux à corrélations diverses) et d'extractions des données pour fournir à l'utilisateur les moyens de les consulter[47]. C'est une méthode «descriptiviste». Un tel catalogage est indispensable à toute recherche sur corpus, mais il ne faudrait pas faire passer ce stade de recueil d'un matériel d'analyse pour l'analyse elle-même, car il se situe en amont de l'analyse.

C'est en utilisant la base de données de l'Inathèque sur la télévision, la radio et le Web (en constitution) que l'équipe du CAD a procédé à ses analyses[48].

Le *relevé des récurrences* implique que la recherche soit orientée. Ici, contrairement au cas précédent, les données sont corrélées à des variables externes qui sont en rapport de pertinence avec un objectif d'analyse, et les récurrences sont relevées et classées dans un certain ordre en fonction de ces variables. Par exemple, un ordre de récurrence de mots et expressions corrélé à un type de locuteur dans une situation donnée (un candidat dans une campagne électorale), ou un ordre de récurrence thématique corrélé à un certain support, comme par exemple le relevé des thèmes traités par un journal à propos d'un certain événement et comparé à ceux relevés dans d'autres journaux. On peut évidemment multiplier les variables, et l'on obtiendra un panorama des résultats statistiques d'occurrences corrélées à diverses variables externes.

Ici, on peut dire que, contrairement au cas précédent, on a affaire à une analyse quantitative fournissant de précieux renseignements pour d'autres analyses postérieures, d'autant que le développement d'outils techniques de plus en plus performants (logiciels et numérisation) permettent à l'heure actuelle de procéder rapidement à de tels relevés.

C'est dans ce genre d'analyse quantitative que s'inscrivent nombre de chercheurs en se réclamant, chacun à sa façon, d'une *linguistique de corpus* —de *grands* corpus faudrait-il dire, car il n'y a aucune linguistique qui ne soit de corpus. Ces travaux livrent des résultats précieux, parfois

---

45 Voir l'introduction de Damon Mayaffre au numéro 4 de *Corpus*, Nice, 2005 : 10.

46 Avec Claude Chabrol, Guy Lochard et Jean-Claude Soulages.

47 Voir, par exemple, l'«Hyperbase» de l'Inathèque de France.

48 Voir : *La télévision et la guerre. Déformation ou construction de la réalité ? Le conflit en Bosnie (1990-1994),* in Charaudeau P. (dir.) 2001 ; et *La médiatisation de la science. Clonage, OGM, manipulations génétiques*, in Charaudeau P. (dir.), 2008.

indispensables pour fonder plus objectivement certaines analyses sémantiques. Mais pour qui s'intéresse, non seulement au sens des mots mais à leur effet social en situation de communication (les mots comme facteur d'influence sociale), on ne peut s'empêcher de faire les remarques suivantes :

- l'importance et l'impact d'un mot (c'est-à-dire l'effet de sens) ne sont pas nécessairement liés à sa récurrence : l'expression «fracture sociale» employée par Jacques Chirac lors de la campagne présidentielle de 1995, les mots «karcher» et «racaille» employés par Nicolas Sarkozy, ne furent employés qu'une seule fois par leurs auteurs, et l'effet qu'ils ont produit tient évidemment au sens qu'ils avaient dans leur contexte situationnel et interdiscursif, mais aussi au fait qu'ils ont été relayés par les médias, donc par d'autres corpus. L'impact des mots n'obéit pas nécessairement à un critère quantitatif. L'important est qu'ils s'inscrivent dans un certain réseau discursif et qu'ils fassent écho à un système d'attente plus ou moins conscient.

- le sens des mots, on le sait, est pluriel. Non seulement du fait de leur polysémie, mais aussi parce qu'ils acquièrent selon les contextes linguistique, paratextuel et interdiscursif ce que l'on pourrait appeler une «polydiscursivité». Et cela est bien difficile à mesurer quantitativement. J'en donnerai un exemple : récemment, à l'occasion de la dernière campagne électorale, des relevés lexicométriques[49] ont mis en évidence la propension des deux principaux candidats, Nicolas Sarkozy et Ségolène Royal à employer le pronom de première personne, *Je*. Cela a conforté les commentateurs qui ont conclu à la grande *personnalisation* de cette campagne. On fera remarquer, tout d'abord, que pour pouvoir juger de cette personnalisation, il faudrait comparer ces relevés avec ceux des campagnes précédentes. Or, si l'on se reporte à la campagne de 1988, on peut observer également un emploi important du pronom *Je*, de la part de certains candidats[50]. Cela est bien naturel dans la mesure où un jeu d'alternance *Je/Nous* est inscrit dans le contrat communicationnelle du discours de campagne électorale : le candidat se présente à la fois comme le porteur de la souveraineté populaire (*Nous*), et comme un leader différent des autres (*Je*).

Mais surtout, si l'on étudie l'emploi de ces *je*, non seulement dans leur contexte linguistique, mais en rapport avec l'environnement propre à chaque déclaration politique (paratextuel) et avec l'ensemble des discours auxquels il est fait référence (interdiscours), on perçoit deux choses. D'une part, que les *Je* de Nicolas Sarkozy et Ségolène Royal obéissent à un même souci de se démarquer de l'appareil de leur parti ; ce sont là des *je* de *différenciation*. Mais d'autre part, on observe une différence entre les deux : les *Je* (plus exactement les *Je veux*) de Nicolas Sarkozy construisent la figure d'un chef puissant et fort, prenant en charge toutes les responsabilités et les décisions, s'instituant en homme d'action avec ardeur guerrière ; alors que les *Je* de Ségolène Royal construisent la figure d'une femme combattante animée d'une conviction profonde, témoignant d'une attitude de protection et d'écoute, de sincérité des sentiments, d'une ardeur faite davantage de courage, de ténacité que de désir de puissance[51] ; ce sont là des *je* d'*appropriation*.

De ce point de vue, les analyses quantitatives, sans être récusées, ne peuvent fournir les résultats que les analyses qualitatives, soucieuses de prendre en compte des réseaux d'interdiscursivité, pourront livrer. Il faudra même se garder (étant donné l'engouement actuel pour ces analyses du nombre) de réifier les premières, car cela aboutirait à la neutralisation des secondes, ce qui réduirait considérablement le pouvoir explicatif de l'analyse du discours.

Rappelons que la logique de la recherche repose sur une démarche «empirico-déductive» de va-et-vient entre, d'un côté, détermination des objectifs d'analyse et des hypothèses signifiantes, et de l'autre, les méthodes d'observation des objets et de recueils de données selon une instrumentation qui doit pouvoir rendre compte des objectifs et des hypothèses préalablement établis. C'est dans un va-

49 Voir entre autres relevés ceux de l'Equipe PSYchologie de la COMmunication , LERASS / IUT Information & Communication de l'université de Toulouse, dirigée par Pierre-Olivier Dupuy et Pascal Marchand (consultable en ligne).

50 Voir : Groupe Saint-Cloud, *Les présidentielles. Regards sur les discours télévisés*, Ina-Nathan, Paris, 1995.

51 Je résume ici, de façon quelque peu cavalière, l'étude que j'ai faite de cette campagne, et à laquelle je me permets de renvoyer le lecteur : : *Entre populisme et peopolisme. Comment Sarkozy a gagné*, op.cit.

et-vient entre ces deux pôles que s'élaborent concepts, instruments d'analyse et interprétation des résultats.

* *
*

Il me reste à renvoyer le lecteur aux études entreprises pas le Centre d'analyse du discours, pour illustrer cette démarche d'articulation entre analyse quantitative et analyse qualitative que je résume ici en quelques mots[52].

S'agissant de la façon dont la télévision rend compte d'événements sociaux (une guerre, une controverse sociale), doivent être relevés, dans un premier temps, les sujets diffusés sur la question traitée ; pour ce faire, on se plonge dans la base de données de l'Inathèque dont un extrait les sujets la question et dont on constitue une «sous-base» (cela correspond à la phase de relevé des données). Puis on procède à un décryptage systématique du matériel réuni, en fonction d'un certain nombre de paramètres externes renvoyant à la situation d'énonciation et à la thématique.

Ainsi, a-t-on pu établir pour chacune de nos études :

- pour le dispositif énonciatif :
  - le *type d'émission* (Journal télévisé, reportage, interviews, débats, etc.) qui permet de déterminer l'importance accordée par chaque chaîne au mode de traitement discursif (description, récit, commentaire, argumentation) ;
  - l'*identité des acteurs* impliqués (politique, scientifique, journaliste spécialisé, envoyé spécial, institution, citoyen, etc.) qui permet de repérer la légitimité et l'autorité des prises de parole ;
- pour la thématique :
  - le *capital thématique* (en nombre de sujets et en durée d'apparition) que chaque chaîne consacre à l'événement-thème. Cela fournit une donnée objective quant à l'importance donnée à l'événement par chaque chaîne comparativement à sa concurrente ;
  - la *hiérarchisation thématique* (en heure moyenne) relative au moment d'apparition de l'événement-thème dans chaque édition de JT (début, milieu, fin), donnée objective quant à la stratégie de traitement à l'intérieur de la programmation (le «conducteur») de chaque chaîne.
  - la *distribution des domaines scéniques*[53], comme variable de contenu, pour déterminer l'importance respective des univers de référence convoqués dans le traitement de l'événement (scènes civile, humanitaire, diplomatique, etc. pour la guerre en Bosnie ; scènes scientifique, politique, éthique, etc. pour la controverse sur le clonage), qui témoignent de l'orientation dramatisante ou argumentative que propose chaque chaîne.

Cette phase quantitative permet d'une part de constituer des indices reposant sur des résultats statistiques, indices qui sont susceptibles de jouer le rôle de *symptôme*, et sur lesquels on fera porter les analyses qualitatives ultérieures[54] ; elle permet, d'autre part, de constituer un corpus-échantillon, c'est-à-dire un ensemble de fragments de texte qui peut être considéré comme représentatif au regard des catégories qui serviront à l'analyser de façon qualitative : la parole des acteurs, les caractéristiques du dispositif, le traitement de la thématique. Ainsi peut-on échapper à la critique qui est faite parfois à l'échantillonnage, à savoir que ce n'est qu'une partie qui prétend représenter un tout. Notre échantillon n'est pas la partie d'un tout (comme serait un échantillon de mots représentatifs

52 Voir : *La télévision et la guerre. Déformation ou construction de la réalité ? Le conflit en Bosnie (1990-1994)*, op.cit. ; *La médiatisation de la science. Clonage, OGM, manipulations génétiques*, op.cit.

53 Pour la définition de «domaine scénique» voir les études elles-mêmes (note 49).

54 Voir la justification de cette démarche in : Charaudeau P., Lochard G., Soulages J.-C., «La construction thématique du conflit en ex-Yougoslavie», revue *Mots* n°47, Fondation des sciences politiques, Paris, 1996.

d'un tout textuel) ; il est constitué de catégories dont chacune joue le rôle d'un point focal sur lequel faire porter l'analyse qualitative (les locuteurs, les genres, les thèmes).

Ainsi l'étude quantitative fait-elle sens en soi, mais un sens provisoire devant être confirmé, corrigé, voire contredit, et en tout cas étendu et approfondi par l'analyse qualitative.

Références bibliographiques

Anscombre J.-C. et Ducrot O. (1983), *L'argumentation dans la langue*, Mardaga, Bruxelles, 1983.

Authier-Revuz J. (1982), «Hétérogénéité montrée et hétérogénéité constitutive : éléments pour une approche de l'autre dans le discours», revue *DRLAV*, 26 : 91-151.

Barthes R. (1975), *Roland Barthes par roland barthes*, coll. «Écrivains de toujours», Seuil, Paris.

Barthes R. (1970), *S/Z*, Le Seuil, Paris.

Bourdieu P. (1987), *Choses dites*, Minuit, Paris.

Boyer H. (1983), ««Solidarité». Fortune politique, dérive lexicale», revue *Mots* 7, Fondation des sciences politiques, Paris, octobre.

Boyer H. (2002), «Sociolinguistique : faire «corpus» de toute(s) voix ?», *Mots* n°69, ENS édition, 97-101, Lyon, juillet.

Boyer H. et Lochard G. (1998), *Scènes de télévision en banlieue*, Ina-L'Harmattan, Paris.

Branca-Rossof S. (1997), «Les lettres de réclamation adressées au service de la redevance», revue *Langage et société* n°81, Larousse, Paris.

Charaudeau P. (dir.) (2001), *La télévision et la guerre. Déformation ou construction de la réalité ? Le conflit en Bosnie (1990-1994)*, Ina-De Boeck, Bruxelles.

Charaudeau P. (2001), «Visées discursives, genres situationnels et construction textuelle», in *Analyse des discours. Types et genres*, Éd. Universitaires du Sud, Toulouse, 45-73.

Charaudeau P. (2004), "Le contrat de communication dans une perspective langagière : contraintes psychosociales et contraintes discursives", in *Psychologie sociale et communication*, Bromberg M. et Trognon A (dir.), Dunod, Paris.

Charaudeau P. (2005), *Le discours politique. Les masques du pouvoir*, Vuibert, Paris.

Charaudeau P. (dir.) (2008), *La médiatisation de la science. Clonage, OGM, manipulations génétiques*, Ina-De Boeck, Bruxelles.

Charaudeau P. *(*2008)*, Entre populisme et peopolisme. Comment Sarkozy a gagné*, Vuibert, Paris.

Charaudeau P., Lochard G., Soulages J.-C. (1996), «La construction thématique du conflit en ex-Yougoslavie», revue *Mots* n°47, Fondation des sciences politiques, Paris, 89-107.

Charaudeau P. et Maingueneau D. (2002), *Dictionnaire d'analyse du discours,* Seuil, Paris.

Charolles M. (1995), «Cohérence, cohésion et pertinence du discours», revue *Travaux de linguistique*, n°29 : 125-151.

Courtine J.-J. (1981), «Quelques problèmes théoriques et méthodologiques en analyse du discours. A propos du discours communiste adressé aux chrétiens», revue *Langages* n°62, Larousse, Paris, 9-128.

Dalbera.J.-Ph. (2002), «Le corpus entre données, analyse et théorie», revue *Corpus* 1 : 89-101, Faculté des Lettres de Nice, .

Ducrot O. et alii (1980), *Les mots du discours*, Minuit, Paris.

Ducrot O. (1991), *Dire et ne pas dire*, Hermann, Paris.

Godin C.*î* (2004), *Dictionnaire de philosophie*, Fayard, Paris.

Groupe Saint-Cloud (1995), *Les présidentielles. Regards sur les discours télévisés*, Ina-Nathan, Paris.

Habert B. et alii (1997), *Les linguistiques de corpus*, Colin, Paris.

Harris Z.S. (1969), «Analyse du discours», trad. fr., revue *Langages*, 13 : 8-45 (1ère éd. 1952, «Discourse analysis», *Language*, vol. 28, 1-30), Larousse, Paris.

Krieg-Planke A. (2009), *La notion de «formule» en analyse du discours. Cadre théorique et méthodologique*, Presses Universitaires de Franche-Comté, Besançon.

Maingueneau D. (1976), *Initiation aux méthodes de l'analyse du discours*, Hachette-Université, Paris.

Maingueneau D. (1996), *Les termes clés de l'analyse du discours*, Seuil, coll. «Mémo», Paris.

Maingueneau D. et Cossutta F. (1995), «L'analyse des discours constituants», revue *Langages* n°117 : 112-125, Larousse, Paris.

Mayaffre D. (2002), «Les corpus réflexifs : entre architextualité et hypertextualité», revue *Corpus* 1 : 51-69, Faculté des Lettres de Nice.

Mayaffre D. (2005), «Les corpus politiques : objet, méthode et contenu», revue *Corpus* n° 4 : 5-19, Faculté des Lettres de Nice.

Moeschler J. et Reboul A. (1994), *Dictionnaire encyclopédique de Pragmatique*, Paris, Seuil.

Pêcheux M. (1975), *Les Vérités de La Palice. Linguistique, sémantique, philosophie*, Maspero, Paris.

Pêcheux M. (1975), «Mise au point et perspectives à propos de l'Analyse Automatique du Discours», revue *Langages* n°37 : 7-80, Larousse, Paris.

Quéré L. (1995), «L'espace public comme forme et comme événement», in Isaac J. (dir.), Recheches, Paris.

Soulages J.-C. (2003), «Identités discursives et imaginaires figuratifs», in Houdebine A.-M. (dir.), *L'imaginaire linguistique*, L'Harmattan, Paris, 103-109.

Tournier M. (2002), «article «Lexicométrie»», in *Dictionnaire d'analyse du discours,* Seuil, Paris.

van Dijk T.A. et Kintsch W. (1983), *Strategies of discourse compréhension*, New York, Academic Press.

Williams G., (éd.) (2005), *La linguistique de corpus*, Presses Universitaires de Rennes, Rennes.