

LANGUE LINGUISTIQUE COMMUNICATION
Collection dirigée par Bernard Quemada

initiation aux méthodes de
L'ANALYSE
DU DISCOURS
Problèmes et perspectives

par

Dominique MAINGUENEAU
Assistant à l'Université d'Amiens

Ami

D.

C
B
B
L

CLASSIQUES HACHETTE
79, boulevard Saint-Germain, Paris 6^e

TABLE DES MATIÈRES

Introduction	
Situation de l'analyse du discours	5
Polysémie du terme «discours»	11
Problèmes méthodologiques	16

LES APPROCHES DU DISCOURS

I. L'approche lexicologique	22
1. La lexicométrie	22
A. Les recherches lexicométriques du laboratoire de lexicologie politique de l'E.N.S. de Saint-Cloud	22
B. A propos de l'analyse factorielle des correspondances	36
2. Lexicologie non quantitative	46
II. L'approche «syntaxique»	65
1. L'analyse du discours de Z. S. Harris	65
2. L'analyse harrisienne au sens large	78
3. L'analyse automatique du discours	83
III. L'énonciation	99
1. L'aspect indiciel: problèmes de typologie	102
2. Les modalités	110
3. Les actes de langage	128
4. Les «places»	139
IV. A propos de la grammaire de texte	151
1. Une lacune grave pour l'analyse du discours	151
2. L'argumentation	163
3. Structures narratives	171
4. Sur la «grammaire de texte»: quelques remarques critiques	177
Conclusion	183
Bibliographie	186
Index	191

La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les «copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective», ci, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayant cause, est illicite (alinéa 1^{er} de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon, sanctionnée par les articles 425 et suivants du Code pénal.

INTRODUCTION

Écrire une *initiation* aux problèmes et aux méthodes de l'analyse du discours ne constitue nullement une entreprise dont il soit facile de tracer le contour: on peut même prétendre que cette discipline récemment annexée à la problématique linguistique n'a pas encore dépassé le stade des prolégomènes et cherche encore à constituer sa méthodologie et son objet. Il y a peut-être même plus grave: ce dont souffre l'analyse du discours, ce n'est pas seulement d'une difficulté à s'articuler sur le champ des sciences humaines, mais c'est aussi d'une difficulté à constituer son unité à l'intérieur de la théorie linguistique. En effet, elle apparaît souvent comme une sorte de parasite de cette théorie, lui empruntant concepts et méthodes sans s'astreindre à une rigueur suffisante. En outre, elle emprunte pour une grande part aux domaines de l'énonciation et de la sémantique qui constituent précisément ce qu'il y a de plus instable dans la réflexion linguistique contemporaine. Sur tous les plans, l'analyse du discours apparaît donc comme une sorte de parent pauvre, et bien des linguistes seraient tentés de renvoyer aux calendes sa constitution. Certes, la linguistique est fort loin de penser de manière satisfaisante son articulation sur les autres pratiques scientifiques, fort loin de présenter partout rigueur et unité, mais l'analyse du discours semble n'en être qu'un double dégradé et en aggraver les carences.

Dans ces conditions on est en droit de se demander quel intérêt peut présenter une introduction à une discipline encore si tâtonnante. On peut avancer plusieurs arguments pour justifier une telle entreprise: ce champ de recherche, pour être un des plus problématiques de la linguistique, n'en est pas moins un des plus actifs; paradoxe qui n'est qu'apparent, car l'analyse du discours possède le privilège de se situer au point de contact entre la réflexion linguistique et les autres sciences humaines, si bien que c'est le plus souvent par l'analyse du discours que beaucoup d'étudiants et de chercheurs sont confrontés à la théorie linguistique. Cela a une conséquence immédiate: il existe actuellement une sorte de «demande» théorique très forte de la part des sciences humaines, demande par rapport à laquelle l'«offre» de la linguistique apparaît ridiculement insuffisante. Cette situation doit amener le linguiste à ne pas se réfugier dans un quelconque mythe d'une «linguistique pure». Plutôt que de considérer avec amusement ou consternation la perversion des concepts linguistiques, mieux vaut essayer de contrôler dans la mesure du possible ce

phénomène inéluctable et de donner à ceux qui en ont besoin non tant les outils nécessaires pour disposer de procédures plus rigoureuses, que la possibilité de saisir les problèmes théoriques que pose leur pratique.

Beaucoup de non-linguistes souhaiteraient que l'analyse du discours leur fournisse une technique « scientifique » qui leur permette d'obtenir des résultats formalisés, directement utilisables pour leurs interprétations extra-linguistiques. A supposer même que la linguistique atteigne dans ce domaine un niveau de rigueur plus satisfaisant, elle exigera plus que jamais une réflexion critique de la part de l'utilisateur, si ce dernier ne veut pas être la victime des plus grandes naïvetés épistémologiques.

Notre projet est en fin de compte modeste: il ne s'agit ici nullement d'un ouvrage de recherche, mais, comme nous l'avons dit, d'un « état de la question », d'un panorama assez ouvert des principales orientations de la recherche linguistique dans ce domaine.

L'essentiel des matériaux se trouve dispersé dans diverses revues et quelques livres, mais il n'est pas toujours facile d'en dégager une vision d'ensemble, étant donné le caractère très tâtonnant de ces travaux ou les objets d'analyse très précis qu'ils se donnent. En raison du projet qui l'anime, une telle « introduction » devra bien souvent se contenter de schématiser, de simplifier des réalités que leur complexité, comme l'insuffisance des recherches linguistiques, laissent encore à l'état de nébuleuse; ainsi que nous le verrons, la rigueur des approches est ici souvent, malheureusement, corrélative de leur caractère réducteur. Nous espérons surtout provoquer une prise de conscience de la nécessité d'une élaboration théorique avant toute pratique; d'ailleurs, répétons-le encore, les choses ne sont pas suffisamment avancées dans ce domaine pour dispenser d'une telle élaboration, et chaque type de corpus exige un travail d'adaptation considérable des méthodes d'approche. Il se peut que certains lecteurs éprouvent une déception devant le caractère trivial et rudimentaire des procédures qui seront présentées: il est certain qu'en regard de la tâche à accomplir, la méthodologie est parfois très fruste, mais il faut bien comprendre aussi que l'on est la plupart du temps conduit à séparer ces travaux du contexte épistémologique dans lequel ils prennent leur pleine signification.

Quelques mots sur le plan de ce livre: le souci pédagogique qui a guidé sa rédaction se reflète dans son organisation. Nous commencerons par essayer de situer rapidement la problématique de l'analyse du discours dans l'histoire de la linguistique structurale; puis nous tâcherons d'éclairer la polysémie de la notion de *discours* que son usage significativement pléthorique rend difficilement maîtrisable.

La deuxième partie vise à apporter au lecteur une information sur les appareils méthodologiques que fournit la linguistique. Nous avons

regroupé ces appareils autour de trois approches principales: lexicologique, syntaxique, « énonciative ». Nous y ajouterons une quatrième partie sur des orientations qui ne sont ~~encore~~ guère développées, mais qui peuvent offrir dans l'avenir de grandes possibilités à l'analyse du discours: ce dernier volet visera essentiellement à apporter une information.

SITUATION DE L'ANALYSE DU DISCOURS *langue/parole/individu/social*

On définit souvent le *discours* comme un terme qui remplacerait celui de *parole* (Saussure) et s'opposerait donc à *langue*; s'il en était ainsi, l'analyse du discours n'aurait aucun fondement. En effet, s'il est nécessaire de remonter au *Cours de linguistique générale* de Saussure, c'est précisément pour construire le concept de *discours* sur une remise en cause de celui de *parole* et non pour reconduire ce dernier. Si la célèbre dichotomie fondatrice *langue/parole* a décidé du sort de la linguistique structurale pour plusieurs dizaines d'années, demandons-nous comment le *Cours* pense une telle opposition.

La langue est une réalité sociale et la parole une réalité individuelle.

• En séparant la langue de la parole, on sépare du même coup: 1°) ce qui est social de ce qui est individuel; 2°) ce qui est essentiel de ce qui est accessoire et plus ou moins accidentel. (1).

En outre, « la langue n'est pas une fonction du sujet parlant, elle est un produit que l'individu enregistre passivement, elle ne suppose jamais de pré-méditation, [...]. La parole est au contraire un acte individuel de volonté et d'intelligence. (2).

Autrement dit, la langue relève de la mémoire, et l'image du *dictionnaire*, du *trésor des signes* s'ensuit logiquement:

• La langue existe dans la collectivité sous la forme d'une somme d'empreintes déposées dans chaque cerveau, à peu près comme un *dictionnaire* dont tous les exemplaires, identiques, seraient répartis entre les individus. (3).

Étant donné les caractères qui permettent d'opposer *langue* et *parole*, on comprend que la phrase ne relève pas de la *langue*, mais de la *parole*, lieu de l'activité et de l'intelligence.

On pourrait, à partir des sources manuscrites du *Cours*, montrer les hésitations de Saussure sur ce sujet, mais l'important, c'est que l'opposition *langue/parole* a fini par devenir l'opposition entre un code homogène, un système de règles universelles et une liberté, une spontanéité qui échappent à toute règle. Cette opposition était évidemment rendue

1. *Cours* (Payot), 1969, p. 30.

2. *Ibidem*.

3. P. 38.

la voix de l'artiste

encore plus nette dès lors que la *langue* n'incluait pas la phrase. L'opposition langue/parole risque en effet d'opposer le systématique (la langue) à l'individuel, au contingent (la parole), à l'autonomie d'un sujet parlant, sujet psychologique soumis à des passions, des besoins, etc. Dans cette perspective, il n'y a pas place pour le «discours», concept qui vise à déposséder le sujet parlant de son rôle central pour l'intégrer au fonctionnement d'énoncés, de textes dont les conditions de possibilité sont systématiquement articulées sur des formations idéologiques.

Ainsi, le couple langue/parole peut amener à considérer qu'il y a d'une part ce qui est systématique et rationnel, un objet homogène et autarcique, la langue, et, de l'autre, ce qui relève de l'usage contingent du système, du rhétorique, du politique, etc. D'un côté, il y aurait un ensemble de mots doués d'un sens fixe et transparent, de l'autre, leur usage. En réalité il s'agit de savoir si le lien entre le sens des phrases d'un texte et ses conditions socio-historiques est quelque chose de secondaire ou est constitutif de ce sens même, indépendamment de l'illusion que peut avoir le locuteur que la signification de son discours coïncide avec ce qu'il «veut dire». Il n'est évidemment pas question d'adresser d'inutiles reproches à Saussure: l'occultation de ce qu'on définira par la suite comme le *discours* correspond à une sorte de nécessité, eu égard au contexte dans lequel a été définie la linguistique structurale.

On ne s'étonnera pas qu'un dépassement indirect de l'opposition langue/parole se soit accompli dans ce domaine privilégié que constituent les œuvres «littéraires». En mettant, dans une perspective qualifiée plus tard de «structuraliste», le principe d'immanence au centre de leur étude des textes, c'est-à-dire en essayant d'étudier la structure du texte en lui-même et par lui-même, en rejetant toute considération extérieure à celui-ci, les travaux des formalistes russes (1) ont permis de dégager une logique des enchaînements transphrastiques, dans le domaine du conte folklorique en particulier. Les recherches de Propp (2) sont maintenant bien connues en France et elles ont permis de construire les premières esquisses de syntaxe narrative, de logique du récit... (3). Certes, ce type de corpus semble un domaine trop spécifique pour avoir une

1. C'est le nom que l'on donne au groupe de jeunes linguistes soviétiques qui, dans les années 1910-1920, ont jeté les fondements et entrepris les premières analyses concrètes dans le domaine de l'analyse structurale des formes littéraires. Citons les noms de V. Propp, R. Jakobson, J. Tynianov, B. V. Tomachevski. Consulter l'anthologie publiée par T. Todorov au Seuil (1965) sous le titre de *Théorie de la Littérature*.

2. Spécialiste du conte folklorique; son ouvrage essentiel, *Morphologie du conte* (trad. frans., Gallimard, 1970), a montré qu'il était possible de construire un modèle élémentaire duquel on dérivait une infinité de contes folkloriques superficiellement très différents.

3. Cf. en France par exemple, les travaux de T. Todorov (*Grammaire du Décaméron*, Mouton, 1969); Cf. BRÉMOND (*Logique du récit*, Seuil, 1973); A. J. GREIMAS (*Sémantique structurale*, Larousse, 1966), (*Du sens*, Seuil, 1970).

incidence directe sur l'étude des langues naturelles ou d'énoncés apparemment moins structurés, mais on comprend facilement qu'en mettant en œuvre les principes fondamentaux de la linguistique structurale sur des contes populaires, des nouvelles..., l'école des formalistes russes préparait la prise en considération par la linguistique de ce qu'on allait appeler le *discours*. L'essentiel était d'appliquer la notion de «système» à des énoncés très vastes, car l'extension de ce type d'approches à des objets de moins en moins privilégiés n'était qu'une question de conjoncture épistémologique. Il va sans dire que l'analyse du discours ne pouvait, par définition, demeurer dans une perspective d'analyse immanente des textes. Les formalistes russes ont réussi à réaliser une rupture par rapport à l'approche impressionniste et philologique des œuvres littéraires, mais n'ont pas réussi à penser la relation de celles-ci à leurs conditions socio-historiques.

C'est dans les années 50 que s'exercent des actions beaucoup plus décisives sur la constitution de l'analyse du discours. Nous nous trouvons devant deux apports en un sens symétriques: l'extension des procédures de la linguistique distributionnelle américaine à des énoncés qui dépassent le cadre de la phrase (nommés *discours*) par Zellig S. Harris en 1952, et les travaux de Roman Jakobson et Emile Benveniste sur l'*énonciation*; d'un côté, comme on le verra, une problématique linguistique très américaine, et de l'autre très européenne.

Nous considérerons plus loin la conception de Harris, aussi suffit-il de dire qu'il est le premier linguiste à étendre directement les procédures utilisées pour l'analyse des unités de la *langue* à des énoncés dépassant le cadre de la phrase. Ce n'est pas un des moindres paradoxes de cette analyse du discours que son initiateur direct ait travaillé en dehors de toute référence à la signification. Ce paradoxe n'en est peut-être un que parce que nous assimilons hâtivement le structuralisme européen post-saussurien et le distributionnalisme américain, dont le contexte idéologique, les postulats, les méthodes, sont différents; chez Harris, «il n'y a pas le choix entre deux objets, ni deux linguistiques: celle de la langue et celle de la parole. La description formelle des données—grammaire ou si l'on veut structure—est un concept opératoire qui permet l'étude du phénomène linguistique» (1).

A l'inverse, E. Benveniste ou R. Jakobson cherchent à dégager comment le sujet parlant s'inscrit dans les énoncés qu'il émet: autrement dit, à une langue conçue comme un répertoire de signes combinés systématiquement, on tend à substituer l'idée que «le locuteur s'approprie l'appareil formel de la langue et énonce sa position de locuteur par des indices spécifiques» (2), selon les termes d'E. Benveniste; le locuteur pose

1. D. LEEMAN, *Langages* n° 29, p. 26.
2. *Langages* n° 17, p. 14.

de même un certain type de rapport à son propre énoncé et au monde. Cette perspective neuve a une très grande incidence sur la constitution d'une théorie du discours. (Il ne faudrait cependant pas se méprendre sur ce point: par le biais de l'énonciation, ce n'est pas la *parole* saussurienne qui est ainsi reconduite, mais il s'agit d'un ensemble de règles fondées sur un appareil formel; sinon, la théorie de l'énonciation seraient en contradiction avec les présupposés de l'analyse du discours, qui se fonde sur l'irréductibilité des intentions du sujet parlant aux mécanismes spécifiques des énoncés qu'il produit dans une situation déterminée.)

Jusqu'ici nous n'avons fait état que d'apports situés à l'intérieur même du champ de la linguistique, mais c'est un point de vue beaucoup trop restrictif, dans la mesure où les problématiques élaborées dans un domaine déterminé des sciences humaines interviennent dans d'autres disciplines (et cela est d'autant plus inéluctable dans le cas de l'analyse du discours). Le développement de l'analyse du discours nous semble ainsi pouvoir être abordé de deux points de vue différents: comme une tentative pour remédier aux insuffisances de la traditionnelle *analyse de contenu* (1), mais aussi comme symptôme et conséquence de ce contexte théorique souvent très diffus qu'est le «structuralisme».

Schématiquement, on peut dire que l'analyse de contenu s'est développée dans l'espace que le structuralisme linguistique avait laissé vide, mais dont les sciences humaines ne pouvaient se passer, la sociologie surtout.

L'analyse de contenu se veut une méthode de traitement contrôlé de l'information contenue dans des textes, au moyen d'une «grille» de lecture objective, dont les résultats seront interprétés: en gros, cela revient à disposer au départ d'un répertoire structuré de «catégories» servant à normaliser la diversité superficielle des textes pour les rendre comparables: une fois les «qualités» homogénéisées, il est alors souvent possible de quantifier. Tout le problème réside précisément dans la construction de ces catégories: même en recourant à une statistique lexicale ou à une syntaxe élémentaire, l'analyse de contenu ne résout pas la question de la structuration du texte. Il serait présomptueux de penser que la linguistique, par l'analyse du discours, sera à même, dans l'immédiat, de se substituer purement et simplement à l'analyse de contenu: cette dernière a des visées nettement plus «empiriques» que la linguistique, car elle ne vise qu'à «caractériser le corpus... au moyen d'un agencement particulier de concepts tirés de la métalangue, de manière à marquer ce que son contenu a de spécifique ou de singulier par rapport à d'autres corpus ou à d'autres parties du même corpus du

point de vue de la sociologie ou de la psychosociologie» (1); ce n'est pas vraiment le fonctionnement d'un discours qui est ainsi pris pour objectif.

Travailler sur des textes ne constitue pourtant pas une pratique évidente et innocente, quels que soient les résultats souvent intéressants auxquels on parvient par l'analyse de contenu. Peut-on étudier le contenu d'un discours sans une théorie du discours?

A vrai dire, ces insuffisances ne sont apparues comme telles qu'en raison de la pénétration progressive (pour ne pas dire l'envahissement) du «structuralisme», essentiellement fondé sur les apports de la linguistique structurale taxée de «science pilote». La linguistique elle-même a subi immédiatement le choc en retour de ce phénomène, se trouvant confrontée à des objets nouveaux et étant mise en demeure de produire de nouveaux concepts pour répondre à une demande considérable. Il est ainsi devenu impératif de considérer mythes, rêves, romans, films, tableaux, etc., comme des ensembles signifiants dont on cherche à établir les lois de la signification à l'aide de concepts empruntés à la linguistique, science de ce système de signification par excellence qu'est une langue naturelle. Dans ces conditions, la position d'une linguistique dont le domaine d'application s'arrêtait à la phrase devenait parfaitement intenable.

Plus généralement, l'apparition de l'analyse du discours n'est que le symptôme d'une mutation dans le statut accordé aux textes. La pratique des textes, jusqu'à la pénétration des analyses structurales, était dominée par le point de vue «philologique»; que ce soit en littérature, en histoire, en ethnologie, en philosophie, etc., il s'agissait de restituer les textes au vécu qui était censé les avoir produits. On cherchait sources, influences, allusions au contexte de l'époque, on voulait déchiffrer, reconstituer (au besoin) le texte original, repérer éventuellement les falsifications, évaluer leur intérêt... On aboutissait ainsi à entourer le texte d'un appareil de notes, de critiques qui devaient permettre sa lecture, c'est-à-dire, en fait, permettre de retrouver directement le passé dont ils émanent: la surface du langage était traversée, une fois éliminée l'opacité, et l'on atteignait son sens, celui de l'époque ou de l'auteur qu'il manifestait. Tout texte, par l'approche structurale, est maintenant devenu, de document, qu'il était, «monument», selon l'expression heureuse de M. Foucault. On ne traverse plus le langage pour se saisir de son sens, le dépouiller des accidents historiques qui l'ont rendu opaque, mais on cherche à dégager ses conditions de possibilité pour expliquer son fonctionnement, à l'aide de théories de la langue, de l'inconscient, des discours, de l'idéologie... systématiquement articulées. On déploie un système de corrélations qui échappent à un contact immédiat avec le vécu. L'analyse du discours est

1. L'analyse de contenu s'est développée aux Etats-Unis en particulier depuis la dernière guerre mondiale chez les sociologues et les historiens. — Citons deux classiques du genre: Bernard BERELSON, *Content Analysis in communication research* (Glencoe, The Free Press), 1952; LASSWELL (Harold J.) et alii, *Language of politics* (New York, G. Stewart), 1949. Pour une présentation et une critique, voir R. ROBIN, *Histoire et Linguistique* chap. 3; inversement, pour une défense de l'analyse de contenu et une critique de l'approche linguistique, consulter l'introduction de *Tracts en Mai 1968* (Armand Colin), 1974.

1. Des tracts en Mai 1968, Introduction.

le nom donné à un ensemble de réponses de la linguistique, très élémentaires pour le moment, à cette énorme pression. Du même coup, on le verra, se pose une grave question : qu'est-ce qui est du ressort de la linguistique dans un discours, et qu'est-ce qui n'en est pas ? Et ce qui n'en est pas, de quelle(s) discipline(s) relève-t-il ?

Entre-temps le distributionnalisme américain avait intégré la phrase dans la « langue » ; cet élargissement sera encore plus net par la suite, dans la théorie générative de Chomsky, qui pose comme symbole de départ P, c'est-à-dire la phrase. C'est ainsi que Chomsky affirme que, selon Saussure,

• la formation de la phrase n'est pas strictement un problème de *langue*, mais plutôt un problème de ce qu'il appelle la *parole*; elle se trouve donc hors du champ de la linguistique proprement dite. C'est un processus de création libre, sans contrainte de la part de règles linguistiques sauf celles qui gouvernent la forme des mots et l'agencement des sons. La syntaxe est de ce point de vue un problème secondaire et il y a effectivement peu d'ouvrages de syntaxe pendant la période de la linguistique structurale. (1).

De fait, Chomsky se démarquait rigoureusement de Saussure en affirmant le caractère dominant de la syntaxe et en remplaçant la conception d'une langue-stock de signes par celle d'une « créativité » du sujet parlant conçue comme système de règles. Contrairement à Harris qui envisageait nettement la distinction entre phrase et discours, Chomsky maintient une certaine ambiguïté sur ce point, semblant admettre implicitement que les sujets parlants produisent des phrases. A vrai dire, la théorie chomskienne ne se construit pas contre une problématique du discours, elle l'ignore presque complètement. La *compétence*, conçue comme le système abstrait de règles sous-tendant les phrases que produit le sujet parlant laisse dans l'ombre les variations dues aux conditions de production des énoncés comme aux locuteurs (2). On ne s'étonnera donc, pas qu'un théoricien de la linguistique du discours puisse écrire que « le concept de compétence... ne parvient pas à donner la connaissance de la réalité du discours, puisque, fondement d'une théorie de la phrase pour un locuteur-auditeur fictif, il est incapable de révéler l'existence du discours » (3).

Encore faudrait-il s'entendre sur le terme de «discours» que l'on trouve aussi bien au centre de la réflexion d'un philosophe comme Michel Foucault que dans des définitions linguistiques très restrictives; cette utilisation pléthorique doit être prise comme un symptôme de son statut théorique instable.

POLYSÉMIE DU TERME «DISCOURS»

Contrairement à ce qui se passe dans d'autres domaines de la linguistique, l'analyse du discours maîtrise très difficilement son objet; linguistes et non-linguistes font du concept de «discours» un usage souvent incontrôlé, et quand certains en ont une conception très restrictive, d'autres en font un synonyme très lâche de «texte» ou d'«énoncé».

En linguistique, moins qu'ailleurs peut-être, il n'y a pas d'évidences et il faut avoir clairement conscience de cette diversité d'emplois pour ne pas commettre de contresens. Nous nous contenterons de quelques remarques éclairantes, ne prétendant pas résoudre le moins du monde un problème terminologique aussi considérable.

Si nous considérons les divers emplois proprement linguistiques de *discours*, nous pouvons citer:

- 1°/ **discours 1:** synonyme de la *parole* saussurienne; c'est son sens courant dans la linguistique structurale.
 - 2°/ **discours 2:** le discours n'est plus tant rapporté à un sujet que considéré comme une unité linguistique de dimension supérieure à la phrase (*transphrastique*), un message pris globalement, un énoncé.
 - 3°/ **discours 3:** dans ce sens, le discours est proprement intégré à l'analyse linguistique puisqu'on considère l'ensemble des règles d'enchaînement des suites de phrases composant l'énoncé. Le linguiste américain Z. S. Harris est le premier à avoir proposé une procédure d'étude de ces enchaînements en 1952.
 - 4°/ **discours 4:** dans ce qu'on pourrait appeler l'« école française » d'analyse du discours, on oppose *énoncé* et *discours* dans une définition très réaliste, nous semble-t-il:

• L'énoncé, c'est la suite des phrases émises entre deux blancs sémantiques, deux arrêts de la communication; le discours, c'est l'énoncé considéré du point de vue du mécanisme discursif qui le conditionne. Ainsi un regard jeté sur un texte du point de vue de sa structuration «en langue» en fait un énoncé; une étude linguistique des conditions de production de ce texte en fera un discours. (1).

- 5/ **discours** 5: nous sommes obligés de signaler un usage de «discours» qui recoupe les précédents, mais avec quelque différence: ce serait sa reformulation dans le cadre des théories de l'énonciation. (Nous y reviendrons.) C'est en ce sens que pour E. Benveniste, «l'énonciation suppose la conversion individuelle de la langue en *discours*» (2); il donne par ailleurs la définition suivante: «Il faut entendre *discours* dans sa plus large extension:

1. *Le langage et la pensée* (Payot) 1970, p. 37.

1. *Le langage et la pensée* (1971), p. 10.
 2. C'est la fameuse conception du « locuteur-auditeur idéal » de Chomsky. Voir l'introduction de *Aspects de la théorie syntaxique* (Seuil).

³ D. SIECKA, *Linguas* n° 23, p. 109.

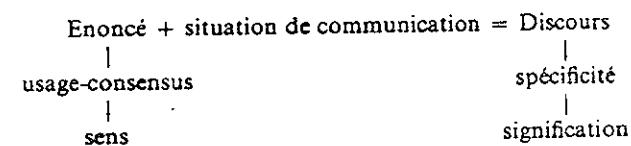
J. L. GUESPIN, *Langages* 23, p. 10.

2. *Languages* 17, p. 13.

toute énonciation supposant un locuteur et un auditeur, et chez le premier l'intention d'influencer l'autre en quelque manière » (1). 6°/ **discours** 6: à peine séparable de ses autres usages, la notion de « discours » entre fréquemment dans une opposition *langue*/*discours*. La *langue* s'oppose alors comme ensemble fini, relativement stable d'éléments au *discours*, entendu comme lieu où s'exerce la créativité, lieu de la contextualisation imprévisible qui confère de nouvelles valeurs aux unités de la *langue*. C'est ainsi qu'on dira que la polysémisation d'une unité lexicale est un fait de *discours* qui se convertit progressivement en fait de *langue*.

Si la définition 4 nous semble très recevable, elle n'en pose pas moins un problème, dans la mesure où de nombreux auteurs parlent d'« analyse de discours » au sens d'étude linguistique des unités transphrastiques, d'un strict point de vue grammatical et sans aucune référence aux conditions de production du discours (ainsi étudient-ils par exemple le jeu des substituts pronominaux dans un énoncé suivi). L'opposition énoncé/*discours* risque donc de masquer cette opposition, importante, *phrase*/*discours*. Peut-être serait-il souhaitable d'utiliser le terme d'énoncé pour les énoncés dépassant les limites de la phrase si on les considère dans leur structuration strictement linguistique, et d'utiliser celui de *discours* quand on considère les conditions de production d'un énoncé (qu'il comporte une seule phrase ou plusieurs: bien qu'il soit fort rare qu'on ait à considérer des discours d'une seule phrase). Il est bien évident que les deux approches sont inextricablement liées: beaucoup de linguistes pensent qu'une étude purement interne de la langue, en dehors de toute prise en considération du discours, est impossible, surtout dans le domaine de la sémantique.

Une source de confusions non moins importante réside dans l'opposition *énoncé/discours* qu'autorise la perspective de l'énonciation. O. Ducrot dans son ouvrage *Dire et ne pas dire* distingue ainsi deux composants dans l'interprétation sémantique d'un énoncé linguistique: un *composant linguistique* proprement dit, et un *composant rhétorique*; le composant linguistique assigne un sens «littéral» aux énoncés, en dehors de tout contexte énonciatif déterminé, alors que le composant rhétorique interprète cet énoncé en l'intégrant à une situation de communication précise. O. Ducrot ne s'inscrit pas explicitement dans le cadre d'une opposition énoncé/discours mais c'est dans la logique de son propos. Chez d'autres, les choses sont plus nettes: considérons par exemple ce schéma de P. Charaudeau (2): Ce dernier met *sens* là où O. Ducrot (p. 111) met *signification*, et réciproquement.



Autrement dit, le *sens* d'un énoncé est défini en dehors de tout cadre énonciatif, alors que sa *signification* est référée aux *circonstances de communication* qui en font un *discours*: « Si l'on considère l'énoncé dans son cadre énonciatif, alors cet énoncé devient *discours* avec, outre son sens-*consensus* [fondé sur le *consensus linguistique* des sujets parlants], une signification spécifique ».

Soit l'exemple suivant, emprunté à P. Charaudeau: « A la base des bonnes affaires en Irlande se trouve l'Allied Irish Investment Bank »; cet énoncé peut avoir, outre son « sens », des « significations » diverses, ainsi:

— Je peut être un anti-capitaliste qui veut convaincre un *tu* qui soutient que l'Irlande n'a pas une économie fondée sur le capitalisme; dans ce cas on dira que l'énoncé contient l'information: « vous voyez que l'Irlande est sous la domination des trusts financiers ».

—Il peut s'agir d'un slogan publicitaire de la forme « si vous achetez P, vous obtenez le résultat X ». On se doit alors de tirer de l'énoncé les informations suivantes, partie intégrante de son information: « si vous voulez que vos affaires marchent, IL FAUT mettre à leur base l'AIIB »; « OR vous voulez cela, DONC contactez l'AIIB ».

Ainsi, en changeant les relations *je-tu*, on change la totalité de la signification de cet énoncé.

Toute la difficulté de cette nouvelle opposition énoncé/discours tient dans la question de savoir quelles sont les relations entre les *conditions de production* (L. Guespin) et les *circonstances de communication* (P. Chaïraudeau), question d'autant plus délicate que les deux perspectives se réclament explicitement d'une analyse du discours. Il semble toutefois possible d'avancer que l'analyse du discours au sens de L. Guespin pré-suppose l'autre perspective. Le type de signification qu'il permet de dégager l'insertion de l'énoncé dans son cadre énonciatif reste finalement relativement immédiat, même s'il s'agit d'une signification implicite: il existe un ensemble de règles, encore mal connues, dans la pratique habituelle de la langue, qui permettent à tout sujet parlant le français de décoder correctement l'interprétation de tels énoncés. En revanche, l'autre perspective vise essentiellement des significations construites à partir d'hypothèses et de méthodes fondées sur une théorie de l'articulation du discours sur les conditions socio-historiques. Alors que les premiers cherchent plutôt à élucider les mécanismes de la pratique usuelle du langage, les seconds s'intéressent surtout à des textes et à des types de

1. *Problèmes de linguistique générale*, p. 242.

1. *Problèmes de linguistique générale*, p. 242.
2. *Etudes de linguistique appliquée* n° 11, sept. 1973, p. 28: « Réflexion pour une typologie des discours ».

textes choisis en fonction de préoccupations dépassant le cadre d'une éventuelle théorie de la communication ordinaire. En outre, les problèmes d'interprétation d'énoncés hors contexte ne se posent guère dans les discours étudiés en tant qu'objets socio-historiques puisque les contraintes contextuelles éliminent des ambiguïtés de ce type (mais sont-ce des ambiguïtés réellement « linguistiques » ?).

En conséquence, ces deux oppositions énoncé/discours se recoupent très nettement sans qu'il soit possible de définir ce qui en droit revient à chacune. Dans ces conditions, on ne peut évacuer ce flou conceptuel et il faut y voir le corrélat du statut très instable de l'« énonciation » (voir *infra*, III^e partie) et de la sémantique, du moins à l'heure actuelle. Pour notre part, dans cet ouvrage, nous nous attacherons essentiellement aux discours-objets socio-historiques et non à ce que O. Ducrot appelait « composant rhétorique ».

Il existe aussi des usages de la notion de *discours* qu'on pourrait qualifier de « paralinguistiques », plutôt que de non linguistiques. C'est le fait en particulier de Jacques Derrida et de Michel Foucault dont la réflexion s'articule très souvent sur la linguistique, mais sans jamais s'y fixer.

Dans un article intitulé « La structure, le signe et le jeu dans le discours des sciences humaines » (1), J. Derrida cherche à penser l'événement qui a fait qu' « on a dû sans doute commencer à penser qu'il n'y avait pas de centre, que le centre ne pouvait être pensé dans la forme d'un étant-présent, que le centre n'avait pas de lieu naturel, qu'il n'était pas un lieu fixe mais une fonction, une sorte de non-lieu dans lequel se jouaient à l'infini des substitutions de signes. C'est alors le moment où le langage envahit le champ problématique universel; c'est alors le moment où, en l'absence de centre ou d'origine, tout devient discours [c'est nous qui soulignons], à condition de s'entendre sur ce mot—c'est-à-dire système dans lequel le signifié central, originaire ou transcendantal, n'est jamais absolument présent hors d'un système de différences. L'absence de signifié transcendantal étend à l'infini le champ et le jeu de la signification » (2).

La réflexion du philosophe déborde largement le cadre de la linguistique et son concept de « discours » vaut pour l'ensemble des systèmes de signes auxquels sont confrontées les sciences humaines. Cette extension d'un concept linguistique n'en est pas moins intéressante: à un niveau très général, on peut dire que la définition du discours comme rejet d'une instance centrale est assez satisfaisante; elle rend compte du double processus par lequel s'est constituée une linguistique du discours: rejet de la *parole saussurienne*, considérée comme « liberté du locuteur, signe de son unicité et de son incomparabilité » (3), refus corrélatif de s'en tenir aux limites de la phrase, qui libère la théorie du discours de la notion de

« jugement » rapportée à la subjectivité. Ce faisant, la théorie du discours se fonde comme une « analyse non subjective des effets de sens » contre l'illusion qu'a le sujet « d'être à la source du sens » (1).

Le « discours » semble aussi au centre de la réflexion de Michel Foucault, du moins si l'on en croit l'usage abondant qu'il en fait: « régularités discursives », « unités du discours », « formations discursives », « ordre du discours »... sont au nombre de ses concepts fondamentaux (on relèverait de même un emploi très particulier du terme « énoncé »). Nous nous contenterons d'une seule remarque: il semble que M. Foucault n'ait pu distinguer si nettement son projet de l'entreprise linguistique que parce que la linguistique à laquelle il renvoie est la linguistique proprement saussurienne, fondée sur le couple langue/parole, ignorant la problématique de l'analyse du discours comme les théories de l'énonciation. Ainsi trouve-t-on dans l'*Archéologie du savoir*:

« L'analyse des énoncés ne prétend pas être une description totale, exhaustive du « langage » ou de « ce qui a été dit »... Elle ne prend pas la place d'une analyse logique des propositions, d'une analyse grammaticale des phrases, d'une analyse psychologique ou contextuelle des formulations: elle constitue une autre manière d'attaquer les performances verbales, d'en dissocier la complexité, d'isoler les termes qui s'y entrecroisent et de respecter les diverses régularités auxquelles elles obéissent. En mettant en jeu l'*énoncé en face de la phrase ou de la proposition*, on n'essaie pas de retrouver une totalité perdue »... (2) (c'est nous qui soulignons).

Les travaux de M. Foucault entretiennent donc des relations quelque peu fausses avec la linguistique du discours.

Nous ne pouvons malheureusement, dans le cadre de cet ouvrage, développer les perspectives de J. Kristeva, qui cherche à réarticuler théorie des idéologies, psychanalyse, sémiologie dans une nouvelle formalisation des systèmes signifiants. Considérons seulement le concept de *texte*: s'il fonctionne souvent, dans la pratique linguistique, comme un simple doublet de « discours », au sens 3, il convient cependant de signaler que J. Kristeva a considérablement élaboré cette notion, dans des directions très enrichissantes pour l'analyse du discours. La sémiologie qu'elle envisage « se donne actuellement pour objet plusieurs pratiques sémiotiques qu'elle considère comme translinguistiques c'est-à-dire faites à travers la langue et irréductibles à ses catégories » (3). Autrement dit, le discours se construit aussi à travers des structurations qui lui sont propres, d'où cette définition du texte comme « un appareil translinguistique qui redistribue l'ordre de la langue [...] », une productivité, ce qui veut dire: 1. son rapport à la langue dans laquelle il se situe est

1. « Mises au point et perspectives à propos de l'analyse automatique du discours », M. PÈCHEUX et C. FUCHS, *Langages* 37, p. 8.

2. *Archéologie du savoir*, p. 142.

3. *Le texte clos*, 1968.

1. 1966 — repris dans *L'écriture et la différence*, 1967, p. 409 sq.

2. P. 411.

3. M. PÈCHEUX et C. HAROCHE, *in t.o. informations* 1972, 1, p. 13.

redistributif (destructivo-constructif), par conséquent il est abordable à travers des catégories logiques plutôt que purement linguistiques. 2. Il est une permutation de textes, une intertextualité: dans l'espace d'un texte plusieurs énoncés, pris à d'autres textes, se croisent et se neutralisent. (1).

PROBLÈMES MÉTHODOLOGIQUES

Un discours n'est donc pas une réalité évidente, un objet concret offert à l'intuition, mais le résultat d'une construction. La conception d'un discours donné à l'intuition et qu'il suffirait de découper s'accompagne souvent du postulat implicite qu'il existe une structure unique qu'il suffirait de dégager pour connaître l'« essence » de ce discours. On aurait de même facilement tendance à opposer une langue dite « naturelle » qui serait sans règle, celle de la conversation courante par exemple, à des types de discours nettement définis (discours polémique, argumentation, etc.), livrant sans ambiguïté les signes évidents de leur existence en tant que discours; ce serait réintroduire subrepticement l'opposition langue-contrainte/parole-liberté. Il faut plutôt considérer que tous les énoncés, en droit, relèvent de typologies, de mécanismes transphrastiques d'un certain degré de généralité, mais de manière plus ou moins diffuse. Par exemple, la conversation courante obéit à des règles d'enchaînement, des contraintes qui, pour ne pas, à l'évidence, relever du même type de « rigueur » qu'un discours électoral n'en obéissent pas moins à un ordre propre. De même, au-delà de ces contraintes générales portant sur toute conversation, il existe des types de contrainte en fonction des types de condition de production (selon le statut social des locuteurs, l'environnement, les rôles joués...).

On n'opposera donc pas un hypothétique langage libre qui serait « naturel », hors de toute contrainte, et des énoncés soumis à différentes contraintes qui seraient des « discours »: on considérera plutôt le discours comme le résultat de l'articulation d'une pluralité plus ou moins grande de structurations transphrastiques, en fonction des conditions de production. Cela suppose néanmoins qu'on maintienne l'existence d'une « langue », d'une base linguistique commune; il ne peut être en effet question de prétendre qu'en raison de la diversité des idéologies des locuteurs, de la variation de l'environnement, la langue éclate en une pluralité hétérogène de « micro-langues ». Cette « base » à partir de laquelle se construisent les discours ne doit pas être conçue comme s'arrêtant à la phrase; un certain nombre de linguistes (2) tentent à l'heure actuelle de montrer qu'un texte quelconque constitue une totalité linguistique spécifique

1. *Ibidem*.

2. Citons en particulier un ensemble de linguistes allemands (Thümmel, Heidolph, Stempel, etc.); consulter à ce sujet le n° 26 de la revue *Langages* et sa bibliographie, p. 122.

au-delà de la simple somme des phrases qui le constituent. Il s'agit donc d'un niveau d'analyse supérieur qui fait partie de la « langue », c'est-à-dire des conditions nécessaires à l'existence d'un discours quelconque. Il convient de distinguer soigneusement cette problématique en « langue » des règles de structuration textuelles liées à des types discursifs particuliers. Ainsi ne faudrait-il pas mêler les considérations portant par exemple sur les substituts pronominaux, ou la désambiguisation des phrases par le contexte et celles qui portent sur l'organisation des parties d'une dissertation philosophique!

Entre des structures discursives très fines et les structures de la langue, on peut construire plusieurs échelles de complexité: ainsi distinguerait-on des régulations transphrastiques d'un grand niveau de généralité (narratif, argumentatif, dialogue, etc.). A un niveau inférieur, on repérerait des types de structuration plus spécifiés, pour parvenir finalement à des régulations fines très particulières.

Ces quelques remarques ne doivent pas donner à penser que les régulations s'emboîtent harmonieusement les unes dans les autres; il s'agit au contraire d'une interaction: ainsi, au niveau le plus élémentaire, un dialogue comprend très souvent des composantes narratives, et inversement, ou des composantes argumentatives; un discours didactique peut se donner la forme d'un dialogue, etc. Dès qu'on aborde des structures discursives complexes, ce phénomène est encore plus nettement marqué. En outre, dans les discours complexes auxquels nous faisons allusion, le poids de l'intertexte est considérable; l'« intertexte » désigne l'ensemble des relations avec d'autres textes se manifestant à l'intérieur d'un texte (citation, parodie, paraphrase, négation, etc.); dans ce cas la mise en relation avec cet intertexte s'avère souvent plus pertinente qu'une mise en relation immédiate et primaire avec les conditions de production.

L'analyse du discours a pour caractéristique d'opérer le plus souvent sur plusieurs discours mis en relation par la prise en considération de leurs conditions de production; c'est d'ailleurs dans ce sens que vont la plupart des recherches dans ce domaine.

L'analyse du discours suppose la mise ensemble de plusieurs textes, étant donné que l'organisation interne du texte pris isolément ne peut renvoyer qu'à lui-même (structure close) ou à la langue (structure infinie, réitération des mêmes processus) [...]. Lorsque l'analyse est conduite sur un seul texte, on se réfère à une typologie existante. (1).

Il faut nécessairement poser un invariant pour mener à bien une analyse de discours. On retrouve ici le problème de la constitution du corpus; la comparabilité des discours suppose des invariants: le même

locuteur dans des environnements différents, avec des rôles différents, différents locuteurs dans les mêmes conditions de production, etc. N'oublions surtout pas la diversité des relations que peuvent entretenir des discours entre eux: alliance, inclusion, antagonisme, etc., et la nécessité d'une réflexion théorique extérieure à la linguistique *stricto sensu*, pour déterminer si l'état des conditions de production présente suffisamment de stabilité et d'homogénéité.

Si l'on se tourne maintenant vers les corpus possibles, ils sont évidemment en nombre infini, mais les types de discours analysables ne le sont pas. A la réflexion, on s'aperçoit vite que bien des difficultés surgissent si l'on prétend considérer comme équivalents tous les corpus théoriquement possibles. L'analyse du discours a pour spécificité de chercher à construire des modèles de discours en articulant ces modèles sur des conditions de production; il s'agit donc de mettre en relation ces deux instances grâce à une « théorie du discours »; cela étant, on risque de se heurter à deux obstacles symétriques l'un de l'autre: être confronté à des discours très complexes dont la mise en relation avec des conditions de production est délicate, ou, inversement, se trouver en face de discours à structuration très diffuse dont la mise en relation avec les conditions de production est si immédiate que le déséquilibre se fait cette fois au profit des conditions de production. Cela ne signifie pas que ces deux types de discours n'aient pas d'intérêt pour la recherche, bien au contraire, mais seulement qu'à essayer de commencer par eux, l'analyse du discours risquerait fort d'adopter une mauvaise stratégie heuristique.

Soit par exemple un corpus de discours théologique: ces textes se définiraient essentiellement par les relations qu'ils entretiendraient avec une foule d'autres textes théologiques eux-mêmes soumis à la même règle, ce renvoi de texte à texte étant une des caractéristiques essentielles de ce type de discours. Etant donné le statut lui-même très complexe des ecclésiastiques et de l'ensemble des institutions religieuses dans une formation sociale, la mise en rapport directe et sans médiation des conditions de production et de tels types de discours constituerait une tâche très difficile à mener. A l'inverse, un corpus composé de discours « spontanés » et peu structurés serait, dans l'état des connaissances linguistiques, c'est-à-dire de nos ignorances, une mauvaise stratégie: les conditions de production prendraient beaucoup trop d'importance. Autrement dit, pour travailler avec le maximum d'efficacité, mieux vaut chercher à équilibrer discours et conditions de production pour que leur articulation soit la plus riche possible; il nous semble que cela explique en partie le succès actuel de l'analyse de corpus de discours politiques.

Il convient de préciser ce point: ce choix n'est pas une exclusion des autres types de discours, mais un choix « stratégique » qui doit permettre un élargissement progressif des types de corpus pris en considération. L. Guespin parvient d'ailleurs à la même conclusion quant aux raisons

profondes de cette restriction du champ d'investigation: « si l'on ne veut pas introduire le laxisme dans ce domaine encore neuf qu'est l'analyse de discours, il s'impose de se spécialiser dans des textes dont les règles discursives soient le moins capricieuses possible. A ce titre l'énoncé politique est particulièrement satisfaisant » (1). Pour cet auteur, dans le discours politique, « il s'agit généralement d'énoncés sur énoncés; ainsi le problème de la situation est-il réduit au minimum, la situation étant un précédent discours, elle est déjà linguistique » (2).

Il semble cependant que le discours politique entretienne avec les conditions de production une relation moins médiate que beaucoup d'autres types de discours, et que si la situation est souvent déjà linguistique, toutefois, de par le statut de ses protagonistes (chefs de parti, députés etc.) et de par sa fonction très remarquable dans la formation sociale, le discours politique constitue un domaine privilégié: discours très structuré, mais moins complexe (3) que bien d'autres, discours produit dans le cadre d'un ensemble d'institutions fixant et délimitant nettement ses conditions de production, discours le plus propre à une lecture en termes d'idéologie.

Il y a, pour la linguistique, deux manières de penser le discours; la première consiste à n'y voir qu'une simple extension de la linguistique quant à la deuxième, elle fait du discours le symptôme d'une crise interne à la linguistique, dans le domaine de la sémantique en particulier. La situation n'est pas stabilisée pour le moment, et cette phase de transition laisse coexister des acceptations multiples, enchevêtrées, voire contradictoires.

Une question qu'il est impossible de ne pas poser et que nous avons déjà rencontrée, c'est donc celle qui consiste à se demander ce qui est proprement du domaine de la linguistique dans l'analyse du discours. A une telle question il est absolument impossible d'apporter une réponse satisfaisante, pour le moment du moins. Une chose est sûre: ce n'est pas à la linguistique seule de tracer une telle ligne de partage, mais plutôt à une théorie du discours. On peut avoir une idée de la complexité du problème en considérant la solution qu'a essayé d'y apporter E. Benveniste. En 1962, il proposait l'opposition suivante: « Il y a d'un côté la *langue*, ensemble de signes formels, dégagés par des procédures rigoureuses, étagés en classe, combinés en structures et en systèmes, de l'autre, la manifestation de la langue dans la communication vivante. » La phrase n'est pas unité de la langue, mais du discours: « Avec la phrase on quitte

1. *Langages* 23, p. 23.

2. *Ibidem*.

3. L'utilisation du qualificatif de « complexe » est extrêmement approximative; nous ne prétendons pas qu'il existe des discours « simples » et d'autres « complexes », mais nous renvoyons ainsi au type de régulations transphrastiques qui sont articulées dans l'unité du discours.

Avec

le domaine de la langue comme système de signes et l'on entre dans un autre univers, celui de la langue comme instrument de communication, dont l'expression est le *discours* (1). Mais un flou conceptuel subsiste: l'étude du discours relève-t-elle de la linguistique seulement? Quel statut donner aux discours dépassant le cadre de la phrase? Prend-on en considération les conditions de production du discours? E. Benveniste reprend cette question sept ans plus tard: « Il faut dépasser la notion saussurienne du signe comme principe unique, dont dépendraient à la fois la structure et le fonctionnement de la langue. Ce dépassement se fera par deux voies: dans l'analyse *intralinguistique*, par l'ouverture d'une nouvelle dimension de signification, celle du *discours*, que nous appelons *sémantique*, désormais distincte de celle qui est liée au signe, et qui sera *sémiotique*; dans l'analyse *translinguistique* des textes œuvres, par l'élaboration d'une métasémantique qui se construira sur la *sémantique* de l'énonciation » (2).

L'équivoque n'est que déplacée: comment séparer ce qui relève de l'*intralinguistique* et du *translinguistique*, en dehors de cas extrêmes? C'est précisément le statut très obscur de l'énonciation qui se trouve au cœur de la difficulté. De même, une des questions essentielles que se pose à l'heure actuelle la linguistique a trait à la *sémantique*: peut-on affirmer qu'elle relève uniquement de la linguistique? Les linguistes préfèrent éviter les positions tranchées, que ce soit pour exclure l'étude du sens de la linguistique ou pour l'y inclure totalement et préfèrent parler d'interaction entre langue et discours ou, simplement, comme E. Benveniste, en faire deux univers différents. Si l'énonciation est à la source de bien des équivoques, c'est qu'en s'en servant comme d'une caution, beaucoup de travaux prétendent rester dans un cadre *intralinguistique* tout en développant des perspectives que d'autres jugent « *translinguistiques* » (relever de l'analyse du discours), au nom d'une conception moins laxiste, ou différente, de l'énonciation.

Après une phase d'expansion « *structuraliste* », au sens strict, dominée par l'analyse distributionnelle, après une phase d'expansion de la grammaire générative et transformationnelle qui a amené un remarquable développement de la syntaxe, la linguistique se heurte de plein front aux problèmes d'une *sémantique* non lexicologique. La notion de *discours* vient alors combler un manque, mais sans qu'il soit encore possible de lui assigner un statut théorique rigoureux. D'un point de vue tout à fait pragmatique, dans ce livre nous entendrons par « *discours* » essentiellement des organisations transphrastiques relevant d'une typologie articulée sur des conditions de production socio-historiques.

LES APPROCHES DU DISCOURS

AVERTISSEMENT

Les « méthodes » que nous allons présenter ont été rangées sous différentes rubriques (l'approche lexicologique, l'approche syntaxique, l'énonciation); cette division n'est nullement fondée sur une théorie, mais n'est que le symptôme des difficultés auxquelles est confrontée l'analyse du discours: elle ne constitue donc qu'un classement pragmatique en vue d'une présentation claire.

En effet, ces tendances (plutôt que ces *approches*) n'ont pas d'autre fondement qu'un certain morcellement de fait du champ de l'analyse du discours, morcellement dont les chercheurs souhaitent voir la fin de façon à préparer la construction d'une théorie capable d'intégrer ces diverses voies.

Nous ne dissimulerons pas que cette présentation est trop sélective et que bien d'autres travaux auraient pu être commentés dans ce même cadre. L'analyse du discours, on le sait, a le dangereux privilège de se trouver à l'intersection de disciplines elles-mêmes en voie de constitution et dont l'articulation sur la linguistique est malheureusement aussi nécessaire qu'encore balbutiante.

En conséquence, on ne s'étonnera pas que des « méthodes » (au sens rigoureux du terme) comme celles de M. Pêcheux, de Harris, ou les approches statistiques, coexistent avec des hypothèses souvent rudimentaires dont les éléments ne présentent pas encore d'unité, comme l'énonciation; sans sombrer dans un pessimisme trop grand, il n'en faut pas moins faire les plus grandes réserves quant à la stabilité des composants de ce qui a pris le nom d'« analyse du discours ». Pour reprendre la métaphore trop célèbre de Claude Levi-Strauss, on dira que l'analyse du discours se constitue par un travail de « bricolage » et qu'elle n'en est encore qu'au stade où elle essaie un peu tous les outils méthodologiques disponibles.

1. *Problèmes de linguistique générale* (Gallimard), 1966, p. 130.

2. *Problèmes de linguistique générale* II, 1974, p. 66. C'est nous qui soulignons.

PROBLÈMES DE L'ANALYSE DE CONTENU

1. Les conditions de production des textes et leur analyse.

1.0. L'analyse de contenu est un ensemble disparate de techniques utilisées pour traiter des matériaux linguistiques. Ces matériaux peuvent avoir été recueillis au moyen d'enquêtes ou d'interviews ou bien ce sont des matériaux « naturels », regroupés en vue d'une recherche, tels que des articles de journaux, des récits, des témoignages, des discours politiques, des œuvres littéraires, etc. Tout ce qui est dit ou écrit¹ est susceptible d'être soumis à une analyse de contenu.

Considérée globalement, bien qu'elle puisse s'appliquer en principe au traitement de tous les matériaux linguistiques, l'analyse de contenu ne conduit pas à une étude du langage (en tant que faculté individuelle par exemple) ou de sa manifestation comme telle. Elle ne débouche pas non plus sur une étude de la langue (en tant qu'institution ou produit socioculturel par exemple). C'est un instrument analytique sans ambitions, sinon sans fondements, théoriques.

1.1. L'analyse de contenu sert à analyser des *textes*, c'est-à-dire des écrits ou des paroles enregistrées et transcrites. Toutefois une analyse de contenu ne s'applique pas au traitement de n'importe quels textes, mais de textes délibérément choisis. Lorsque les textes sont obtenus au moyen d'enquêtes ou d'interviews, ce choix se manifeste dans la formulation des questions et dans l'échantillonnage de la population soumise à l'enquête, ou dans le choix des thèmes d'interviews et la conduite de ces derniers. Lorsque l'analyse est appliquée à des textes « naturels », ceux-ci sont soigneusement choisis (V. un exemple dans Holsti et al., 1963, pp. 39-41).

1. Nous excluons du champ d'application de l'analyse de contenu tout ce qui n'est pas proprement linguistique tel que films, représentations picturales, comportements (considérés comme « symboliques »), etc., bien que, à certains égards, le traitement de ces matériaux pose des problèmes apparentés à ceux de l'analyse de contenu.

Nous excluons du champ de notre étude des pratiques analytiques, parfois appelées « analyses de contenu », utilisées en linguistique ou en recherche documentaire. Enfin, nous nous concentrerons principalement sur les méthodes utilisées en psychologie sociale.

Ce choix est nécessaire car toute analyse de contenu vise, non l'étude de la langue et du langage, mais la détermination, plus ou moins partielle, de ce que nous appellerons les *conditions de production* des textes qui en sont l'objet. Ce qu'on cherche à caractériser ce sont ces conditions de production et non les textes eux-mêmes. L'ensemble des conditions de production constitue le champ des déterminations des textes. Nous préférerons l'expression « condition de production » à celles qui sont communément utilisées et qui font référence soit à la situation de communication, soit à l'environnement socioculturel, soit à la psychologie individuelle. Ces dernières expressions ont toutes l'inconvénient de restreindre arbitrairement le champ des déterminations possibles des textes en mettant l'accent sur tel ou tel aspect particulier.

Les conditions de production sont multiples et dans la pratique on ne s'intéresse qu'à quelques-unes d'entre elles. Par exemple, la situation d'interview est une condition de production (nous devrions dire un ensemble de conditions de production) lorsque les textes sont obtenus par ce procédé. Cette situation influe sur certaines particularités des textes, mais ne les détermine que partiellement. Ce n'est qu'une condition de production parmi d'autres, on ne s'y intéresse qu'indirectement (d'un point de vue méthodologique) et il est toujours possible d'envisager d'autres modes de détermination. Ainsi, lorsqu'on cherche, par l'analyse des textes, à définir l'attitude d'une ou plusieurs personnes à l'égard d'un objet (marque de savon, « coopération technique », ou « démocratie » peu importe), il est toujours supposé que cette attitude existe indépendamment des textes et qu'elle en a déterminé certains aspects (ceux sur lesquels devra s'appuyer l'analyse). On peut donc dire que cette attitude figure parmi les conditions de production des textes. Même lorsque l'analyse de contenu se réduit à une analyse stylistique, on suppose que le style existe indépendamment du texte, ce qui permettra, par comparaison, de dire s'il est légitime ou non d'attribuer un texte anonyme à un auteur connu (Hart, 1934). Le style, ou tout au moins certains de ses aspects, est considéré comme une condition de production des textes en ce sens qu'il affecte tous les textes produits (ou productibles) par un auteur à des fins précises (par exemple littéraires).

1.2. Ces exemples, proposés à titre illustratif, ne suffisent pas à donner au concept de « condition de production » une signification précise : il n'a pour l'instant qu'une valeur opérationnelle. Il nous permet d'examiner les procédures d'analyse de contenu en référence à deux plans fondamentaux. Le premier plan de référence est celui de l'*analyse des conditions de production*. Le psychologue, le psychologue social, le sociologue disposent d'un vocabulaire théorique ou plus exactement d'un modèle d'analyse plus ou moins précis de ces conditions de production qui constituent leurs objets d'étude. Ce modèle sert à structurer le premier plan de référence que nous appellerons par commodité de langage, le *plan vertical*. Le second plan de référence est celui de l'*analyse des* le *plan horizontal*.

textes dont on ne retient que certains éléments ou certains traits; nous l'appellerons le *plan horizontal*.

1.3. Nous sommes maintenant en mesure d'énoncer le problème fondamental de l'analyse de contenu. Il est clair que les deux plans de référence que nous venons de définir ne sont pas indépendants. C'est ce qui simultanément rend l'analyse possible et fait problème. Selon le plan horizontal, la procédure analytique est déterminée par le plan vertical; le choix des éléments du texte que l'on retiendra dépend naturellement des objectifs de l'analyse, c'est-à-dire de celles parmi les conditions de production que l'on veut étudier ainsi que de la manière de les définir. Inversement, le plan vertical est déterminé par le plan horizontal. En effet, les conditions de production sont caractérisées à partir des éléments distingués dans le texte et les possibilités de cette caractérisation dépend de leur définition. En l'absence de définition autonome de la structuration de ces deux plans et de leurs rapports, on aboutit à une impasse méthodologique et théorique.

1.4. Tous les problèmes de l'analyse de contenu dérivent de celui-ci. Les pratiques analytiques ne sont que des moyens d'échapper à la circularité de cette problématique sans pour autant la résoudre. L'ensemble de cet article sera consacré à l'examen de ce problème sous ses diverses formes. Afin de l'énoncer d'une manière moins abstraite et plus précise, nous commencerons par présenter succinctement les méthodes existantes. Dans la même perspective, nous examinerons plus en détail une étude entreprise par W. Ackermann et R. Zygouris dans le cadre du Centre d'Études et de Recherches psychotechniques. Puis nous esquisserons des solutions théoriques.

2. Trois types d'analyse de contenu.

2.0. Pour présenter les méthodes utilisées en analyse de contenu, nous allons naturellement faire usage des concepts opérationnels que nous avons introduits. Les procédures analytiques seront classées en fonction de la structuration de leur plan vertical (c'est-à-dire en fonction de la caractérisation des conditions de production qu'elles servent à étudier) et de leur plan horizontal (c'est-à-dire en fonction de leur analyse des textes) ainsi que des relations qu'elles postulent entre ces deux plans. Nous obtiendrons ainsi trois grandes familles de procédures que nous aborderons dans l'ordre de la complexité croissante (laquelle se trouve coïncider grossièrement avec l'ordre chronologique de leur apparition).

2.1. Les procédures les plus simples ont été minutieusement décrites par Berelson (1952). Leur caractéristique fondamentale est qu'elles privilégièrent le plan vertical au détriment du plan horizontal en n'envisageant que des rapports simples entre ces deux plans. La description des textes est commandée par la définition des conditions de production considé-

rées et par la manière de les caractériser. Ces conditions de production, ce sont principalement les attitudes à l'égard d'objets déterminés, les opinions et croyances sur un sujet précis et délimité et des variables plus contingentes telles que les intentions ou les motivations. Les attitudes sont caractérisées par leurs composantes, leurs directions et leurs intensités. Classiquement on distingue dans les attitudes des composantes affectives (ou évaluatives) et cognitives. Les opinions ont un caractère plus exclusivement cognitif et n'ont pas, en principe, de composante affective. Ces composantes globales peuvent être subdivisées en sous-composantes dont la définition est plus directement liée aux attitudes particulières considérées. Par direction, on entend par exemple favorable opposé à défavorable ou bien, dans un cas particulier, autoritarisme opposé à égalitarisme. Par intensité, on entend le caractère plus ou moins saillant d'une composante dans une direction donnée. L'attitude, comme l'opinion, est toujours une attitude ou une opinion à l'égard de quelque chose de précis mais qui peut être abstrait — comme la démocratie — ou concret — comme un produit de consommation. Si l'on cherche à définir les attitudes à l'égard d'un objet par l'analyse de textes, cet objet doit faire partie (explicitement ou implicitement) du référent des textes. C'est le critère de choix des textes.

L'analyse de contenu est une méthode très utilisée dans l'étude et la catégorisation des attitudes, opinions et croyances. Envisagée comme condition de production des textes, l'attitude ou l'opinion n'est qu'une condition relativement peu déterminante. De ce point de vue l'analyse est donc très partielle pour ne pas dire ponctuelle. En contrepartie seuls certains éléments isolés des textes sont retenus. On repère des mots ou des fragments de textes qui sont regroupés en classes thématiques ou classés par items. Les mots, les fragments de textes sont retenus dans la mesure où ils dénotent quelque chose de pertinent du point de vue de l'attitude considérée, c'est-à-dire en fonction de la possibilité de les rattacher à l'une des composantes ou sous-composantes de cette attitude. Composantes et directions définissent les catégories thématiques. Le dénombrement des éléments de chaque catégorie doit en principe caractériser l'intensité de l'attitude selon telle ou telle composante et dans telle ou telle direction.

La description empirique des attitudes et des opinions fournit dans ce cas un cadre analytique dans lequel sont classés certains éléments des textes. Il n'y a aucune définition autonome de la structuration du plan horizontal, mais les problèmes posés par ces méthodes intéressent aussi bien le plan vertical que le plan horizontal ou les rapports entre ces plans. En ce qui concerne le plan vertical, il est clair que la validité de l'analyse dépend essentiellement de la valeur de la description empirique des conditions de production considérées. Pour le plan horizontal, se pose le problème de la dimension des unités de contenu. Les méthodes qui reposent sur le repérage de mots isolés ne tiennent pas compte du

contexte dans lequel ces mots apparaissent. Or ce contexte peut modifier considérablement leur signification. Les méthodes qui utilisent des unités de contenu plus larges négligent les nuances de l'expression et même, puisqu'il est difficile de parler de nuances lorsqu'on substitue à tout un paragraphe ou à plusieurs pages de texte un seul item pour les représenter, altèrent profondément la signification. Le rôle de l'interprétation est considérable. Enfin et surtout, il n'est jamais tenu compte des relations entre les différentes unités de contenu, de leurs rapports mutuels.

En ce qui concerne les rapports entre les deux plans, la validité de l'analyse dépend entièrement de la possibilité de décider ce que l'emploi de tel ou tel mot ou expression dénote quand à l'attitude considérée. A la limite, il faudrait pouvoir faire a priori l'inventaire de tous les mots ou de toutes les expressions qui révèlent les différents aspects des attitudes ou opinions. Or, si on s'intéresse par exemple à l'attitude à l'égard de la coopération technique telle qu'on peut la définir à partir d'articles de presse, on ne peut même pas décider de ne retenir que les jugements de valeur par exemple (on y serait enclin du fait que, dans toute attitude, il y a une composante affective ou évaluative). Le fait que certains aspects de la coopération sont ou ne sont pas abordés dans le texte est en lui-même significatif. On n'est pas en mesure de faire l'inventaire de tout ce qu'on pourrait dire sur un tel sujet, donc de déterminer les manques même à l'intérieur d'un système de catégories pré-établi.

On a cherché à tourner ces obstacles de plusieurs manières. La première consiste à avoir recours à une sorte de pré-analyse qu'on pourrait qualifier de « sauvage » ou d'intuitive. Cette analyse aurait pour but d'explorer le contenu global des textes et d'en déterminer empiriquement les dimensions saillantes par rapport aux conditions de production considérées. On peut également procéder par approximations successives en alternant analyse systématique et retour aux textes. La seconde manière de contourner les difficultés revient à considérer que l'analyse a un but principalement comparatif. Il ne s'agit plus de caractériser, ou de « mesurer » des attitudes ou des opinions mais de comparer, d'un texte à l'autre ou d'un ensemble de textes à l'autre, ces attitudes et ces opinions. Pour tester des hypothèses ou des prédictions théoriques, on s'intéresse plus souvent aux changements ou aux différences d'attitudes ou d'opinions qu'à leurs caractéristiques intrinsèques. Utiliser l'analyse de contenu à des fins comparatives répond donc à un besoin pratique de la recherche. On évite du même coup la question des manques mais on n'élimine pas les autres difficultés car les critères de comparaison sont définis à partir de descriptions psychologiques empiriques et ce choix laisse peu de place à la comparaison directe des textes. Il faut mentionner enfin que, pour résoudre la question de la structuration du plan vertical, on a tenté de constituer des systèmes de catégories universelles (en particulier Laswell, 1938, et White, 1951). La construction de ces grilles repose entièrement sur des considérations empiriques étrangères au problème de l'inférence

à partir de textes; leur utilisation pose avec une acuité accrue le problème des rapports entre les deux plans d'analyse comme l'ont fait remarquer divers auteurs (Pool, 1959).

Comme nous l'avions annoncé en abordant la présentation de cette première famille de méthodes, leur caractéristique fondamentale est l'importance qu'elle accorde à la structuration du plan vertical, laquelle ne s'accompagne pas d'un effort parallèle d'analyse du texte. Cette structuration n'a été possible que parce qu'elle concerne un ensemble très délimité et particulier de conditions de production en faisant abstraction de toutes les autres. Elles postulent que les conditions de production prises en considération sont indépendantes des autres déterminations des textes. C'est, pour une part, en réaction contre ces différentes limitations et pour échapper à leurs conséquences que les méthodes que nous allons maintenant examiner ont été conçues.

2.2. Ces méthodes n'ont pas été construites dans le seul but de surmonter des difficultés méthodologiques : elles devaient en premier lieu répondre à des exigences nouvelles issues des développements de la psychologie et de la psychologie sociale. La validation des hypothèses théoriques ne reposait plus sur la mesure de certains écarts isolés mais nécessitait l'étude des systèmes de variables interdépendantes. Ainsi, dans le prolongement du paragraphe précédent, la théorie des attitudes s'intéressait désormais aux relations entre attitudes à l'égard d'objets différents. Cette évolution, qui n'intéresse pas que la théorie des attitudes, impliquait une nouvelle structuration du plan vertical et surtout du plan horizontal. Les techniques d'analyse devaient prendre en considération les relations entre les éléments des textes, c'est-à-dire tenir compte des *contextes*. Elles sont donc caractérisées par une analyse plus approfondie des textes, analyse qui fait appel parfois simultanément à des techniques inspirées par la théorie de l'information et par la linguistique structurale et aux possibilités de traitement automatique sur calculateur électronique.

Il est possible de définir de très nombreux types de relations entre éléments d'un texte. Les techniques varient selon le type de relation considéré et corrélativement selon la manière de définir les contextes, c'est-à-dire les unités de contenu. La théorie de l'information et la linguistique ont fourni des modèles opérationnels pour définir ces relations. Dans d'autres cas (celui du General Inquirer en particulier), on ne s'intéresse qu'aux relations marquées dans le texte entre certains aspects définis des conditions de production considérées : c'est alors le plan vertical qui commande la délimitation des unités de contenu.

La variété des modèles utilisés pour structurer le plan horizontal a pour corollaire une grande diversité dans les méthodes; nous devrons nous limiter à l'examen de quelques types représentatifs.

2.2. a Un premier type de relation entre mots d'un texte est défini à partir de la fréquence des co-occurrences dans un contexte de dimension fixée arbitrairement. Imaginons par exemple qu'on ait recueilli une série

d'interviews non directifs dans le but de déterminer les attitudes à l'égard de plusieurs objets et les relations entre ces attitudes. On structure le plan vertical comme précédemment, c'est-à-dire qu'on définit des catégories thématiques correspondant à divers aspects des attitudes. On repère non seulement les mots ou les séquences de chaque catégorie, mais les co-occurrences de mots ou de séquences de catégories différentes dans des contextes de dimension fixée arbitrairement (par exemple la phrase : on estime que, si les occurrences de termes sont trop espacées, il ne peut y avoir d'interférences entre elles). On peut ainsi dresser des matrices de fréquence des co-occurrences des éléments de chaque catégorie pour chaque texte et la matrice des fréquences moyennes des co-occurrences pour l'ensemble des textes. Cette dernière matrice fournit une sorte de point de référence; on considère que les différences interindividuelles s'annulent les unes les autres. Il est alors possible de répartir les co-occurrences dans un texte particulier en deux catégories, d'une part celles dont la fréquence est significativement supérieure à la moyenne pour l'ensemble des textes et d'autre part celles dont la fréquence est significativement inférieure à cette moyenne.

A partir de là, on peut conclure à l'existence d'associations ou d'exclusions entre catégories. Une telle procédure permet de répondre à des questions telles que :

a) Pour une personne déterminée, l'objet A évoque-t-il l'objet B ou au contraire le fait de penser ou de mentionner un objet A empêche-t-il de penser simultanément à un objet B ou de le mentionner?

b) Telle direction dans une composante de l'attitude à l'égard d'un objet est-elle liée à telle direction dans une autre composante de cette attitude, etc.?

On ne peut mettre en évidence par cette méthode que des propriétés relatives, définies par comparaison à la moyenne d'un échantillon, d'une population ou d'un groupe de « juges ». La validité de cette comparaison est basée sur des considérations empiriques qui ont été soutenues par le développement de la théorie de l'information. Pour justifier de l'emploi d'une méthode apparentée à celle que nous venons de décrire, Osgood (1952) s'appuie sur Shannon et Weaver pour affirmer que les mots n'apparaissent pas au hasard mais par paquets, qu'on peut reconstituer les mots manquant dans un texte mutilé en se basant, non sur le *sens*, mais sur un modèle de la langue qui fixe la fréquence d'association des mots. Tout écart par rapport à cette norme génératrice est considéré comme porteur d'une signification contingente, c'est-à-dire qu'il n'est pas déterminé par la langue ou par la culture mais par d'autres conditions de production, propres au producteur particulier du texte.

Dans cette direction peut être envisagée une généralisation de ces méthodes qui éliminerait le caractère relatif des résultats. Des dictionnaires de fréquence d'apparition des mots ont été constitués pour la plupart des langues usuelles à partir de très vastes corpus. A partir de là,

on peut calculer des fréquences de co-occurrences statistiquement normales par rapport à cette distribution de fréquence pour certains corpus de mots sélectionnés, puis comparer les fréquences de co-occurrences de ces termes dans un texte déterminé à cette fréquence théorique. La conception d'un tel instrument suscite deux remarques complémentaires : (a) Quelle est la signification des fréquences « normales » de co-occurrence; autrement dit, parvient-on par ce procédé à représenter opérationnellement ce qui, dans le modèle hypothétique légitimant la procédure, intéresse les éléments retenus? (b) L'instrument est trop grossier car le corpus de référence est trop général : la structure qu'il détermine par projection sur l'échantillon est trop lâche pour cerner les variables. Tout au plus, selon nous, on peut étalonner l'échantillon, la population ou le groupe de « juges », délimiter un « thème » dans un complexe culturel.

Tout en étant instructives car elles posent les questions du corpus et des relations entre les textes rassemblés en vue de l'analyse et un ensemble culturel plus vaste (questions sur lesquelles nous reviendrons), ces objections ont peu de poids comparées à celles qui portent sur la structuration du plan vertical et l'inventaire des catégories thématiques. Sur ce point, les difficultés sont identiques à celles que nous avons rencontrées à propos des méthodes plus simples. Il en est de nouvelles qui tiennent (a) à la nécessité d'une préparation du texte (« édition ») — par exemple il est nécessaire de mettre sur le même plan les mots et leurs modalités (pronom, etc.) pour que le dénombrement des co-occurrences ait une quelconque signification — (b) au découpage arbitraire du point de vue linguistique des unités de contenu — la délimitation des contextes soulève de grandes difficultés lorsqu'on a affaire à des phrases de structures complexes.

2.2. b Une méthode proposée par Osgood sous le titre « Evaluative Association Analysis » (EAA, Osgood et al., 1956) se heurte aux mêmes difficultés « linguistiques » et leur apporte une solution normative au niveau de l'édition des textes. Là encore il s'agit d'une technique d'analyse des attitudes, l'accent étant mis sur la composante affective. L'édition, c'est-à-dire la préparation des textes, a pour but de donner aux séquences retenues dans les textes une structure canonique. Celle-ci peut avoir deux formes :

- (a) $A_0_1 - C - P$
- (b) $A_0_1 - C - A_0_2$

(ces formes sont dites « affirmatives »; en anglais « assertives »).

(A_0_1) est un objet sur lequel porte une attitude de même que (A_0_2) (« Attitude Object »), C est un connecteur verbal, P un prédicat d'un type particulier (*common meaning term*). Les objets d'attitude sont des symboles dont l'évaluation (*evaluative meaning*) varie d'une personne à l'autre (par exemple : la démocratie, le capitalisme, une marque d'essence,

le cinéma, un film, etc.). Les prédictats considérés (*common meaning terms*) sont des symboles dont l'évaluation est stable (par exemple, bon, traître, efficace, etc.). Objets d'attitude et prédictats sont plus des catégories thématiques que des unités linguistiques. Des équivalences entre noms sont posées par le codeur au moment de l'édition en fonction du contexte.

De par la définition même des unités de contenu, on ne retient des textes que des jugements de valeur. Le choix de ces formes canoniques est étroitement lié à la structuration du plan vertical. Cette structuration est fondée sur une technique multidimensionnelle de mesure des attitudes qu'Osgood a proposée. C'est une généralisation des techniques connues sous le nom « d'échelles d'attitudes ». Les échelles utilisées par Osgood sont des échelles en sept points aux extrémités desquelles figurent deux prédictats antinomiques (exemple bon-mauvais). Lorsqu'on veut déterminer l'attitude d'une personne à l'égard d'un objet, il lui est demandé d'évaluer cet objet sur une série d'échelles en indiquant le point de chaque échelle, entre les extrêmes, qui lui semblent le caractériser. Osgood a étudié la distribution des réponses à des séries d'échelles en les soumettant à des analyses factorielles et a dégagé trois dimensions indépendantes (évaluation : bon-mauvais ou beau-laid; activité : rapide-lent ou tendu-relâché; puissance : fort-faible ou lourd-léger), ce qui ouvrait la voie à une caractérisation multidimensionnelle des attitudes. De plus, il s'est intéressé aux rapports entre attitudes : schématiquement le principe de congruité énonce que, lorsqu'une relation est perçue entre deux objets, leur évaluation dépend de cette relation et de son sens (relation associative ou disjonctive).

Le rapport entre cette technique de mesure des attitudes et la EAA est défini par une pondération des connecteurs verbaux (indice du poids) et du sens de la liaison entre deux objets ou entre un objet et un prédictat) et des prédictats (évaluation des objets) sur des échelles en sept points (pour déterminer ces pondérations on a encore recours à des groupes de juges). Connaissant les pondérations, on calcule indirectement l'évaluation dans chaque dimension des objets des attitudes, évaluation qui tient compte des relations entre les différents objets telles qu'elles sont mesurées par la pondération des connecteurs.

Les problèmes que soulève la mise en œuvre de cette technique sont d'une part la détermination des équivalences entre termes à l'intérieur d'une catégorie thématique et d'autre part la réduction des séquences utiles en formes canoniques. Cette réduction normative est plus fondée sur une interprétation que sur une analyse des textes bien que la définition de la forme des unités de contenu se rapproche d'une définition linguistique. Les autres objections qu'on pourrait formuler portent sur les rapports entre les deux plans d'analyse. Elles concernent en fait les techniques de mesure des attitudes au moyen d'échelles et nous nous contenterons ici de les mentionner.

Les deux types de méthode que nous venons d'esquisser — mesure des fréquences des co-occurrences et EAA — ont de nombreux points communs. Entre autres, elles nécessitent l'une et l'autre une préparation ou édition des textes qui précède l'analyse proprement dite. Cette caractéristique, nous la retrouverons dans un troisième type de méthode conçue en vue du traitement automatique sur ordinateur.

2.2. c Cette méthode comporte en fait un très grand nombre de variantes. Toutes ces variantes sont cependant dérivées d'un système d'analyse documentaire baptisé le « General Inquirer », lequel est un sous-système du langage artificiel « COMIT ». De très nombreux ouvrages ont été consacrés aux différentes versions du « General Inquirer » et à ses applications (Stone et al., 1966); nous devrons nous contenter ici d'énoncer les principes. A bien des égards, les systèmes « General Inquirer » apparaissent comme une synthèse et une systématisation des méthodes examinées jusqu'ici. Cette systématisation est liée à l'automatisation, laquelle constraint de rendre explicite des règles qui restaient jusque-là dans le domaine des préoccupations méthodologiques. Les systèmes « General Inquirer » se présentent comme des outils permettant de traduire les données empiriques, extraites des textes, en données théoriques. Le cadre théorique, c'est-à-dire la structuration du plan vertical, dans lequel ces données vont être organisées, est là encore défini d'une manière autonome. Le rapport entre le plan vertical d'analyse des conditions de production considérées et le plan horizontal d'analyse du texte est défini en partie par un dictionnaire qui est enregistré, en mémoire. Ce dictionnaire fixe des correspondances entre le vocabulaire théorique servant à décrire les conditions de production et le vocabulaire des textes par l'intermédiaire d'un système de « mots-clefs » (*tag concepts*). Ces mots-clefs désignent les catégories de mots qui sont la trace de tel ou tel aspect des conditions de production considérées. Si la recherche a pour but de caractériser des attitudes à l'égard d'objets définis, le mot « objet » figurera dans le vocabulaire théorique et les mots-clefs correspondants seront des classes d'objets. Le dictionnaire est construit à partir de l'inventaire des aspects des conditions de production dont on doit tenir compte pour les besoins de la recherche, c'est-à-dire qu'on commence par faire l'inventaire du vocabulaire théorique lequel traduit la structuration du plan vertical. On établit ensuite les listes de mots-clefs correspondant à chaque terme du vocabulaire théorique. Le degré de généralité de ces termes — qui doit être homogène — fixe la finesse de l'analyse. Enfin, par l'exploration systématique d'un thésaurus, on fait l'inventaire de tous les mots qui peuvent apparaître dans les textes et qui peuvent être considérés comme autant d'indices de la présence des variables représentées par les mots-clefs. Ces listes d'entrée peuvent être enrichies après une première exploitation car les programmes prévoient l'établissement de la liste de tous les mots rejetés. Il est donc possible de réajuster le dictionnaire. Ce réajustement reste cependant partiel car il n'affecte ni le vocabulaire théorique

ni le système de mots-clefs qui en dépend. En d'autres termes, on ne peut pas modifier la définition du rapport entre les deux plans d'analyse. Ce fait est évidemment très important car il a pour conséquence qu'on ne peut pas modifier, à partir de l'analyse des textes, la structuration du plan vertical, introduire de nouveaux aspects de conditions de production plan vertical, introduire de nouveaux aspects de conditions de production autrement qu'en reconstruisant un nouveau dictionnaire, c'est-à-dire en reprenant l'analyse théorique des conditions de production. »

Le dictionnaire à lui seul ne définit pas complètement les rapports entre les deux plans d'analyse sauf dans les cas où on ne s'intéresse qu'à la présence et à l'absence de certaines variables des conditions de production parmi les déterminations des textes (et éventuellement à leur « intensité » mesurée par des fréquences d'apparition des mots d'entrée pour chaque mot-clef). Les systèmes « General Inquirer » peuvent utiliser des unités de contenu plus larges que le mot. Celles-ci sont parfois appelées « thèmes ». Leur définition est là encore commandée par le plan vertical et non basée sur des critères linguistiques (sémantiques ou syntaxiques) : la dimension de l'unité de contenu est fixée par le nombre maximum de relations qui peuvent intervenir entre les différents aspects des conditions de production apparaissant dans une même unité de contenu. En pratique, on fixe arbitrairement le nombre d'aspects des conditions de production qui peuvent figurer simultanément dans une unité, ce qui revient au même. Dans une étude sur les conflits internationaux, R. Holsti (Holsti et al., 1963) a défini le thème comme étant une séquence à l'intérieur de laquelle ne figurait pas simultanément plus d'un des quatre éléments suivants : (a) celui qui perçoit l'action ou l'état, (b) l'agent ou celui qui est perçu, (c) l'action, (d) le but de l'action ou la cible. La définition de ces unités de contenu de dimension fixée rend nécessaire une préparation des textes dont le but est de séparer à l'intérieur des phrases les unités de contenu. Par ailleurs, il est possible, connaissant le dictionnaire, de faire un inventaire théorique a priori de tous les thèmes possibles. Il ne reste plus alors, à l'aide d'une indexation appropriée, qu'à repérer ces thèmes manuellement dans les textes. (Cette procédure d'indexation manuelle permet de faire l'économie de la construction d'un dictionnaire des thèmes analogues au dictionnaire des mots. La liste des mots d'entrée sous chaque mot-clef pouvant comporter plusieurs dizaines d'unités, la simple combinatoire donnerait un dictionnaire des thèmes de dimensions considérables entraînant des temps de recherches disproportionnés.)

Les systèmes « General Inquirer » résultent d'une tentative dans la construction d'un modèle universel d'analyse de contenu. Leur caractéristique fondamentale est encore la prédominance du plan vertical sur le plan horizontal et l'impossibilité de modifier la structuration du plan vertical par inférence à partir de l'analyse des textes. On exprime parfois ce fait en disant que ce sont des systèmes à interprétation close. Par ailleurs, le nombre des relations qui peuvent être prises en considération entre les éléments d'interprétation est fixé a priori, ceci tient au fait que

ces systèmes ne sont pas dotés d'une syntaxe autonome permettant de construire, par exemple par imbrication, des unités de contenu de dimensions variables.

**

Les deux premières familles de procédure que nous avons examinées ont un point commun. Pour l'une comme pour l'autre, la structuration du plan vertical commande l'analyse du texte. Dans le cas des procédures du premier type, on ne cherche pas à atteindre autre chose qu'une description de conditions de production particulières. Le cadre théorique de cette description est fourni par une analyse autonome de certaines caractéristiques psychologiques ou d'autres facteurs (description des attitudes, catégories de Bales, 1950, par exemple) mais ceux-ci ne sont considérés implicitement comme facteurs de production qu'à l'occasion de la procédure analytique. On ne repère donc dans les textes que des indices de ces états.

Avec les méthodes de la seconde famille, on quitte le stade purement descriptif, leur but est de mettre à l'épreuve des hypothèses concernant par exemple les changements d'opinion, d'attitudes, les phénomènes d'influence, etc. Cette ambition théorique implique qu'on a défini préalablement un certain nombre de *variables* des phénomènes correspondants. La définition de ces variables est donnée par exemple par la théorie des attitudes (divers principes de consistance ou de congruité qui fixent certaines liaisons théoriques entre divers aspects des attitudes) ou par certaines hypothèses concernant les processus d'interaction. Ces variables ne sont pas définies en tant que facteurs des processus de production, mais en tant que variables des phénomènes qui, même lorsqu'ils font intervenir le langage, sont supposés avoir leur lois propres. Là encore, ce n'est qu'à l'occasion de la mise en œuvre de la procédure analytique que ces variables et leur relation sont considérées en tant que facteur de production. Dans ces conditions, les textes sont utilisés comme fournit un ensemble d'indices. La signification des éléments apparaissant (ou susceptibles d'apparaître) dans les textes détermine leur qualité d'indices de telle ou telle variable, la fréquence de ces éléments est l'indice d'autres variables — intensité ou degré de liaison.

Dans un cas comme dans l'autre, on est prisonnier du mode théorique de description ou de modèle théorique définissant les variables précisément parce qu'il n'y a d'ordre ou de variable que dans le plan vertical. C'est la structuration du plan vertical qui fixe quels sont les éléments pertinents des textes, quelles sont les équivalences entre éléments à l'intérieur des catégories thématiques, quelles sont les relations à retenir entre catégories d'éléments, quelle est la dimension des unités de contenu. La mise en œuvre des procédures ne permet pas d'aboutir à la formulation d'hypothèses nouvelles, à la définition de variables qui

sortiraient du cadre théorique fixé. Du point de vue de leur rentabilité théorique, ces procédures débouchent donc sur un système clos d'interprétation.

D'autre part, et c'est là leur seconde caractéristique fondamentale, leur validité sur le plan strictement méthodologique est suspendue à la légitimité de l'utilisation des traits des textes comme indice de telle ou telle variable particulière. Le mode de liaison entre les deux plans d'analyse n'est pas systématique; il est établi par des interprétations ou doit être assujetti à des épreuves empiriques.

Ces deux faits suffisent à expliquer qu'on ait cherché à construire d'autres systèmes d'analyse qui, premièrement, laissent ouvert le système d'interprétation et, deuxièmement et corrélativement, ne définissent pas a priori les traits des textes susceptibles d'être utilisés comme indices. Les méthodes correspondantes, que l'on pourrait qualifier de procédures de découvertes, doivent s'appuyer sur un mode d'analyse des textes, offrant, de plus larges possibilités. En effet (du fait que jusqu'ici c'était la structuration du plan vertical qui commandait les caractéristiques des méthodes), à la clôture du système d'interprétation correspondaient certaines particularités du mode d'analyse des textes. Ceux-ci consistent en la répartition des éléments retenus en un nombre fixe de catégories pré-établies, ou bien on utilise un système de représentation dérivé d'un langage artificiel à un nombre fini d'états, comme le COMIT. Or, la définition d'indices dans un système à interprétation ouverte repose essentiellement sur la mise en évidence de propriétés des textes, de relations entre leurs éléments définissant des unités fonctionnelles. C'est une méthode de ce type que nous allons maintenant examiner en détail.

2.3. Cette procédure a été utilisée dans le cadre d'une enquête sur l'assimilation des connaissances scientifiques. Cette enquête elle-même faisait suite à une série d'études traitant de la diffusion des connaissances scientifiques, des transformations des modes de penser, des métamorphoses des notions scientifiques consécutives à la diffusion de ces connaissances (Barbichon et Moscovici, 1965). Le vocabulaire scientifique comporte de nombreux termes qui ont une signification « vulgaire », hors du champ de chaque spécialité (lumière, vitesses, poids, etc.). Les objets de la connaissance scientifique font partie de l'environnement naturel. Les techniques sont bouleversées par les progrès de la science. En conséquence, le domaine scientifique n'est pas un domaine clos que l'on pourrait apprêhender de l'intérieur sans qu'il y ait interférence avec une connaissance antérieure empirique ou déjà théorique. L'assimilation des connaissances doit être envisagée du point de vue de la réorganisation des connaissances antérieures et de l'altération des connaissances nouvelles (altération du point de vue de la source) nécessaires à leur intégration dans un contexte comportant des débris du cadre préexistant. Le but de l'enquête était donc de clarifier, voire de réformer, les conceptions classiques du mécanisme de l'assimilation en l'expliquant. Pourquoi

l'accueil est-il sélectif? Quels sont les facteurs qui entraînent le refus de notions nouvelles? En quoi l'organisation nouvelle des connaissances dépend de l'organisation ancienne? Quel est le processus de la construction de cette nouvelle organisation? Enfin quel est le rapport existant entre l'organisation des connaissances à la source et chez celui qui a été soumis à l'influence de cette source? Autant de questions auxquelles on ne sait donner que des réponses fragmentaires.

L'enquête (Ackermann et Zigoiris, 1966-1967) a été effectuée auprès d'un groupe de stagiaires techniciens d'un centre de formation professionnelle de l'AFPA. Une série de questions relatives à des phénomènes qui leur étaient familiers et dont l'explication faisait appel à leurs connaissances scientifiques et techniques a été soumise à ces stagiaires à deux reprises, en début de stage et, quatre mois plus tard, en fin de stage. Entre ces deux passations du questionnaire, les stagiaires avaient suivi une formation théorique. L'analyse des réponses obtenues lors de la première passation devait permettre de définir l'organisation initiale des connaissances et l'analyse du second groupe de réponses devait révéler les transformations de cette organisation.

Dans une étude de ce genre, on ne dispose pas d'une structuration précise du plan vertical; les hypothèses concernant les mécanismes d'assimilation restent vagues. Tout au plus sait-on dans quelles directions on doit les formuler. La procédure analytique doit donc être exploratoire. Alors que jusqu'ici c'était la structuration du plan vertical qui avait dominé la définition de la procédure, dans le cas présent, ce sera le choix du système de représentation et d'analyse des textes.

2.3. a L'instrument d'analyse des textes est dérivé d'un système général d'analyse documentaire, le SYNTOL (Cros, Gardin, Lévy, 1964).

Ce système permet de construire à l'aide d'un *lexique organisé* et d'un système de *relations syntaxiques* des représentations condensées, normalisées et manipulables d'énoncés scientifiques. Ces règles fixent la forme des représentations ainsi que les opérations et les transformations auxquelles elles peuvent être soumises. Il comporte deux axes de références. L'un, *syntagmatique*, dans lequel les relations entre les termes sont explicitement marquées à l'aide des relations syntaxiques. L'autre, *paradigmatique*, est lié à l'organisation hiérarchique du lexique, laquelle permet d'exprimer certaines relations implicites de « voisinage » ou « d'équivalence » sémantique. Les relations exprimées selon l'un et l'autre axe sont de même nature et jouent des rôles complémentaires dans l'expression de l'information. L'organisation paradigmique est fixée une fois pour toutes. Il est seulement possible, lors d'une recherche rétrospective, et grâce à des instructions précises, d'en tenir ou de ne pas en tenir compte. Elle permet d'exprimer des relations stables, universellement acceptées donc le plus souvent implicites. L'organisation syntagmatique offre au contraire de grandes possibilités combinatoires propres à exprimer des rapports contingents, vérifiés seulement dans un contexte particulier.

Un tel système a l'avantage de ne pas fixer arbitrairement la dimension des unités de contenu mais laisse en suspens la question du découpage de ces unités. Par ailleurs, le système des relations SYNTOL ne permet pas d'exprimer la totalité des règles rhétoriques du langage naturel, une réduction est nécessaire. Celle-ci a pu au moins partiellement être rendue automatique par la construction d'algorithmes d'analyse automatique opérant au niveau syntaxique (et applicable au français). Dans ces conditions, en théorie sinon en pratique, il n'est pas nécessaire d'avoir recours à des interprétations non systématiques pour opérer ces réductions. Le gros problème reste bien entendu la construction du lexique et le choix de son organisation. Comme pour la délimitation des unités de contenu, la nature et les buts de la recherche sont sur ce point déterminants, et les deux questions ne sont pas indépendantes. En analyse documentaire, cette construction ne soulève pas de problèmes majeurs tant que n'est pas posée la question de l'enrichissement du lexique. Les besoins spécifiques de la recherche rétrospective des informations, d'une part, et la connaissance du domaine scientifique sur lequel portent les documents traités, d'autre part, fournissent les principes de sélection des termes et de leur organisation. Cette référence à l'organisation du savoir scientifique auquel renvoient les textes joue le rôle d'une norme qui permet la réduction des énoncés et assure l'efficacité de l'exploitation en recherche documentaire.

2.3. b Cet instrument ne peut pas être utilisé en analyse de contenu sans un minimum d'adaptation. Il n'est pas possible d'introduire brutalement la référence au savoir scientifique, à la grille culturelle, sans risquer de plaquer sur les textes une organisation des connaissances qui n'existe que chez celui qui les analyse. Comme nous venons de la rappeler, cette grille culturelle constitue en recherche documentaire, le paradigme de la méthode d'analyse et le principe de normalisation. Le problème est donc celui de la définition des critères de comparaison permettant de considérer comme équivalentes des expressions distinctes quant à leur forme mais non quant à ce qu'elles révèlent de l'organisation des connaissances ou des mécanismes de leur assimilation. Or, c'est là précisément l'inconnue et on ne dispose pas d'un modèle théorique de cette organisation et de ces mécanismes.

Il ne nous est pas possible d'entrer dans les détails de l'analyse (Ackermann et Zygouris, 1966). Nous devrons nous limiter à ce qui en constitue les caractères saillants. Pour contourner la difficulté, on a eu recours à une procédure d'approximation en se basant simultanément sur la comparaison des réponses de l'ensemble des sujets (repérage des invariants ou des séquences recurrentes et des commutations) et sur une référence au savoir scientifique guidée par cette comparaison. C'est par la lecture des énoncés qu'on a fixé ce qu'on pourrait appeler « le niveau d'abstraction » auquel il fallait se situer dans la formulation des théories pour obtenir une norme de comparaison ni trop grossière ni trop fine.

Cette norme a été traduite par le choix des termes du lexique et par l'organisation en arbre de ce lexique (les termes généraux dominant les termes plus particuliers selon le modèle du SYNTOL) ainsi que par la détermination des unités de contenu (séquences constituant des réponses minimales à la question posée).

L'application de cette technique a permis de construire des représentations des textes à l'aide d'un ensemble de « schémas-types » lesquels sont des constellations de concepts unis par des relations SYNTOL. Partant de là, on a pu définir une typologie des réponses fondée sur l'inventaire des schémas-types servant chacune à les représenter. Cette typologie a été exploitée lors d'une comparaison systématique des réponses aux deux passations, comparaison qui a permis de mettre en évidence certains faits intéressants. On peut représenter la démarche à l'aide d'une

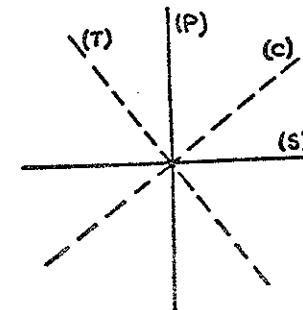


image graphique (laquelle vaut ce que vaut toute image). (P) et (S) représentent respectivement les axes paradigmatisques (lexique hiérarchisé) et syntagmatique (schéma de représentation des textes), (C) est l'axe de la comparaison des textes, (T) l'axe de référence aux théories ou en d'autres termes le paradigme de l'interprétation par l'analyste de la comparaison brute des textes². Le décalage entre ces deux systèmes d'axes traduit la distance entre l'énoncé brut et sa représentation. Plus la réduction est poussée, plus le décalage est grand. Une édition ou préparation manuelle des textes est nécessaire pour passer d'un système à l'autre: On construit un état intermédiaire des textes entre leur état brut et leur représentation. Cet état intermédiaire est une sorte de résumé qui élimine les variations stylistiques, les éléments inanalysables et prépare le découpage en unité de contenu.

2.3. c Le passage des résumés aux schémas-types peut être systématisé mais non le passage des textes bruts à leurs résumés. Ce biais révèle ce qui est contestable du *point de vue méthodologique* dans la procédure d'autant plus qu'un problème important, celui des erreurs, n'a pu être résolu de manière satisfaisante. Ce problème ne se pose pas en analyse documentaire ou ce qui n'est pas conforme à la norme relève de l'inven-

2. On doit considérer que ces systèmes d'axes sont situés dans le plan horizontal de référence puisqu'ils définissent le mode d'analyse des textes.

tion et implique un enrichissement ou une modification de cette norme. Ce n'est pas le cas ici et les textes énoncés échappent à la norme et ne sont représentables que pour autant qu'ils sont corrigés. On n'a en effet qu'une connaissance empirique des *causes d'erreur* (très imparfaite du reste) et non des systèmes d'erreurs, ce qui fait que ces données ne peuvent être intégrées à la procédure analytique. La correction a pour effet de morceler les textes, de briser les relations entre les éléments puisque c'est au niveau de ces relations (et non des termes qui sont toujours pris dans leur sens usuel, comme en recherche documentaire) que se manifestent ces erreurs. Les auteurs ont dû représenter les énoncés par les schémas-types correspondant à l'énoncé corrigé (en leur adjoignant un indice pour signaler l'erreur), quitte à revenir par la suite à une analyse autonome des erreurs.

En fin de compte, l'impossibilité de rendre compte des erreurs dans le système de représentation utilisé ne remet pas en cause toute la procédure. La question des erreurs est importante, ce n'est cependant pas la seule question à laquelle on peut s'intéresser dans une étude sur les mécanismes d'assimilation. Les questions du niveau et du degré d'organisation sont non moins centrales. Ce sont ces problèmes principalement qui ont pu être étudiés, ce qui signifie que les conséquences de la non-résolution du problème des erreurs ont pu être circonscrites. Le vrai problème reste celui de la coupure dans le passage des textes à leur représentation au niveau du résumé. Dans une démarche exploratoire, ce fait revêt une importance particulière. En effet, on est en droit de s'inquiéter de savoir si les propriétés des représentations résultent ou non de l'introduction de la norme qui rend nécessaire cette coupure. A fortiori, les inférences, concernant les conditions de production, faites à partir de ces propriétés risquent de manquer de fondement. On ne peut pas décider *a priori* que les opérations effectuées sont illégitimes mais la procédure d'approximation, bien qu'elle constitue un effort de systématisation laisse une marge d'incertitude.

2.3.4 Les problèmes que nous venons d'évoquer ne sont cependant pas les plus importants du point de vue de l'analyse de contenu.

Le but d'une analyse de contenu n'est pas d'obtenir une représentation systématique des textes analysés. Tous les problèmes de passage des textes à leur représentation sont mineurs car le véritable problème est de savoir quel type de représentation fournira des indications utiles aux progrès de la recherche entreprise (ici, l'étude des processus d'assimilation des connaissances).

Nous avons dit que la procédure avait permis de construire une typologie des réponses. On doit remarquer que, selon le point de vue où l'on se serait placé, on aurait pu en obtenir d'autres sans qu'en puisse décider que l'une soit meilleure. Bien au contraire, on devrait convenir qu'elles constituent autant de descriptions analytiques complémentaires d'une même réalité car les unités de représentation, schémas-types ou

autres, ne constituent pas des variables mais tout au plus des indices. Même dans l'hypothèse où l'on serait parvenu à définir avec précision des transformations à l'intérieur de l'ensemble des schémas-types qui servent à représenter les textes (en faisant apparaître certains d'entre eux comme des développements ou des synthèses *logiques* de schémas plus simples), rien ne permettrait d'affirmer que ces transformations reflètent celles qui permettent effectivement de passer d'un type de réponse à un autre. La question de l'adéquation entre ces processus théoriques et les processus réels resterait entière. En effet, si dans une telle perspective les unités de représentation apparaissent bien comme des variables des systèmes de transformation correspondants, ces transformations comme ces variables ne sont définies qu'au niveau du plan horizontal : elles ne concernent que les propriétés propres des textes par rapport à la langue ou les unes par rapport aux autres. On fixe ainsi un mode d'approche pour l'*observateur* des textes à analyser qui ne tient aucun compte de l'unité phénoménale des processus étudiés.

Ce type de problème n'est pas circonscrit à un domaine de recherche particulier. Il se pose dans les mêmes termes, à quelques nuances près, pour toutes les méthodes qui tentent d'atteindre une caractérisation théorique des conditions de production à partir d'une analyse autonome des textes et de cela seulement. C'est en particulier le cas pour la *Discourse analysis* de Harris (1952) qui, fondée sur la seule analyse linguistique des textes en termes distributionnels et transformationnels, ne peut fournir autre chose que des indices linguistiques devant être interprétés en termes de condition de production, ce que Harris lui-même a fait remarquer en opposant analyse du discours et analyse de contenu.

Dans un autre ordre de faits, puisqu'il ne s'agit que de la génération linguistique des énoncés, des remarques analogues peuvent être faites à propos de certaines inférences faites par Chomsky (1967) à partir de la construction des modèles théoriques de la génération que constituent les grammaires génératives. Ces inférences concernent une rationalité intrinsèque du sujet parlant qui permettrait de rendre compte de distinctions telles que celles qui ont été introduites entre modèle fort et modèle faible. Cette rationalité intrinsèque constituerait un ensemble de conditions de production de tout énoncé. Le fait important reste que des variables concernant les rapports entre énoncés sont définis par les systèmes de transformations introduits par Chomsky. Toutefois, les tentatives faites pour tester l'adéquation entre les processus théoriques de génération et les processus réels par Miller (1962), Miller et Mc Kean (1964) se sont soldées par des échecs (Fodor et Garrett-Sutherland et Cohen, 1966). Ces résultats négatifs n'excluent cependant pas, comme nous le verrons, une utilisation rationnelle des systèmes de transformations définis entre énoncés.

3. Perspectives théoriques : Analyse de contenu taxinomique et analyse de contenu systématique.

3.0. En examinant les principaux types de méthodes, notre objectif était de rassembler les éléments nécessaires pour formuler avec précision les problèmes de l'analyse de contenu.

La critique interne des méthodes peut être poursuivie dans le sens d'une remise en question des rapports entre l'instrument et les phénomènes observés bien au-delà de ce que nous en avons dit. On peut chercher à tenir compte plus encore de la lettre des textes en raffinant les indices linguistiques ou statistiques et simultanément on peut repenser les conceptions théoriques qui structurent l'analyse des conditions de production. Nous n'avons nullement été exhaustifs car ce n'était pas notre propos. Par la critique interne, on ne peut poser que des questions méthodologiques, or les problèmes de l'analyse de contenu ne sont pas seulement des problèmes de méthode. Tant qu'on tentera sur deux plans séparés d'améliorer l'analyse des conditions de production et l'analyse des textes, on ne pourra juger ni de la valeur opérationnelle ni de la légitimité des opérations auxquelles conduit la mise en œuvre des techniques. Les problèmes sont simultanément théoriques et pratiques. Théorique en ce qui concerne la liaison entre les deux plans d'analyse, pratique en ce que les méthodes n'offrent que des possibilités limitées (a) dans la mise à l'épreuve d'hypothèses (b) dans la construction d'hypothèses nouvelles. Les méthodes correspondent en effet à deux types de démarches analytiques que nous avons appelées respectivement procédures closes et procédures exploratoires. Mettre en œuvre une procédure close consiste à partir d'un cadre empirique ou théorique d'analyse de certains états psychologiques, psychosociologiques ou autres, qu'on cherche à particulariser ou bien à propos desquels on a formulé des hypothèses ou des questions. On rassemble des textes produits dans des conditions telles que ces états fassent partie de leurs conditions de production. Puis, on observe ces textes à travers le cadre théorique fixé pour aboutir à une particularisation des conditions de production considérées ou à une mise à l'épreuve des hypothèses les concernant. Ces procédures sont dites closes car le cadre pré-établi ne peut être modifié.

En ce qui concerne les procédures exploratoires, on commence également par rassembler des textes produits dans des situations particulières correspondant aux états qu'on veut observer. Le cadre d'analyse de ces dernières n'est pas fixé et on part de la mise en évidence de propriétés des textes. On définit des différences, des similitudes et éventuellement des transformations faisant se correspondre les textes ou leurs représentations. Les différences, les similitudes, les transformations doivent ensuite être interprétées pour aboutir à une caractérisation des états observés.

En résumé, ce qui sépare procédure close et procédure d'exploration c'est le fait que le principe d'organisation qui fixe le mode d'observation

est dans un cas fourni par ce que nous avons appelé le plan vertical et dans l'autre, le plan horizontal. Ces procédures se rejoignent en ce qu'elles constituent les unes et les autres des *techniques d'observation* d'états qu'on ne cherche pas à apprêhender en tant que condition de production des textes (bien qu'implicitement on les considère comme tels), mais en quelque sorte en eux-mêmes, indépendamment du fait qu'ils peuvent intervenir dans le processus de production. Bien au contraire, ces états interviennent dans des phénomènes d'un tout autre ordre et, sion postule l'existence d'une systématique de ces états, les procédures sont nécessairement en marge de cette systématique. Les techniques opèrent en quelque sorte par coups de sonde, la seule unité phénoménale à l'échelle des textes étant celle des processus de leur production et non celle des phénomènes étudiés. Dans ces conditions, il n'y a rien d'étonnant à ce que les textes ne puissent fournir que des indices des états permettant de les classer en établissant entre eux des différences et des similitudes. La logique des transformations entre états est établie ou postulée sur un autre plan. Pour cette raison, nous dirons que ces procédures sont des méthodes d'*analyse de contenu taxinomiques*. Nous entendons ainsi insister sur ce qui constitue la limite de leur rentabilité théorique.

3.1. Ce qui précède nous permet du même coup d'énoncer à quelle condition on peut espérer s'affranchir de ces limitations. Le problème est en fin de compte celui du rapport entre les processus de production des textes et les phénomènes étudiés. Nous allons commencer par montrer que ce n'est là qu'une manière nouvelle d'énoncer la problématique que nous avions posée au début de cet article.

D'une part, c'est bien la dimension des phénomènes étudiés, celle du champ dans lequel ils opèrent, qui rend nécessaire la séparation opérationnelle des deux plans d'analyse. Elle fait que ces phénomènes apparaissent à des moments de leur déroulement dans le temps comme conditions de production des textes. Le mode d'action de ces conditions de production particulières sur les caractéristiques des textes est fixé par les processus de production qui ont leur spécificité propre. Ce sont les lois de ces processus qui définissent le rapport entre les deux plans d'analyse.

Les méthodes d'analyse de contenu taxinomique considèrent les processus de production comme un invariant dans la démarche analytique. En procédant de la sorte, on fait une double approximation. Premièrement, on postule que, quels que soient les phénomènes étudiés, c'est-à-dire la nature des conditions de production, les lois des processus de production qui interviennent sont les mêmes. Deuxièmement, on pose qu'il y a indépendance entre les phénomènes étudiés et les processus de production que ces derniers n'en modifient pas le cours ou encore que le fait de la verbalisation ne constitue pas un événement en soi³.

3. Il n'est pas étonnant que l'on considère alors le langage comme un mode « symbolique » de comportement, c'est-à-dire dont l'ordre serait entièrement réductible à celui d'autres phénomènes.

Dire que les phénomènes étudiés et les processus de production ont leurs lois propres ne revient pas à poser ces deux ordres de phénomènes comme indépendants. Simultanément c'est la condition de toute approche scientifique des problèmes posés.

3.2. Sur le plan théorique, que signifie le fait de considérer que les processus de production des textes et les phénomènes qu'on étudie par leur analyse ont leurs lois propres? Cela revient à dire qu'il y a un ordre des phénomènes étudiés et un ordre du langage même si, dans une perspective théorique plus large, on doit poser que l'ordre du langage comme tel résulte de son intervention dans des phénomènes d'un autre ordre; tels que le fonctionnement de la pensée, la communication, les échanges, la culture, etc., et des conditions biologiques. C'est à ce niveau que se situe l'articulation réelle entre les deux ordres.

Pratiquement, en ce qui concerne la démarche analytique elle-même, on doit poser qu'il existe un système de transformations (dont la logique est celle des processus) qui permet de passer des conditions *globales* de production aux textes. En fait, on ne s'intéresse pas aux conditions globales de production parmi lesquelles il faudrait ranger les caractéristiques de la langue (institution sociale), mais seulement à celles d'entre elles qui sont en rapport avec les phénomènes étudiés. Corrélativement, on ne s'intéressera pas à toutes les particularités des textes mais seulement à celles d'entre elles qui sont liées par l'intermédiaire des processus de production à ces conditions particulières.

En conséquence, on ne peut se contenter de raffiner les indices statistiques ou linguistiques puisque rien ne permet de savoir a priori si ces indices ont une signification par rapport aux phénomènes étudiés. Le problème est bien celui du rapport entre les deux plans d'analyse et renvoie à une étude des processus de production.

3.3. L'étude de ces processus peut être envisagée dans deux perspectives différentes. L'une est à long terme et théorique. Elle vise la constitution d'une théorie des processus de production. Cette entreprise dépasse le cadre de chacune des disciplines qui s'intéressent au langage ou utilisent l'analyse de contenu, c'est-à-dire entre autres, la linguistique, la psychologie, la sociologie, la psychologie sociale, l'ethnologie, etc., car il faudrait tenir compte des conditions globales de production et de l'ensemble des propriétés linguistiques des textes. En fin de compte, la validation d'une telle théorie (qui fournirait du même coup une théorie de l'instrument d'analyse) repose sur la pratique analytique ce qui nous renvoie à la seconde voie possible. Celle-ci est opérationnelle et à court terme bien qu'elle puisse fournir des éléments de l'approche théorique. Elle consiste à isoler expérimentalement les variables des processus de production qui sont liées aux phénomènes étudiés. Autrement dit, il s'agit de dégager le système partiel de transformations permettant de passer des conditions de production particulières liées aux phénomènes étudiés aux traits correspondants des textes. Dans cette voie, on doit avoir recours à des pro-

cédures expérimentales qui permettent soit une manipulation des conditions de production, non plus en tant que variables des phénomènes étudiés, mais des processus de production, soit un échantillonnage des situations de production qui par recouplement a le même effet. Dans le champ des phénomènes étudiés, les conditions de production considérées sont des variables de ces phénomènes organisés en fonction de la connaissance partielle qu'on en a ou d'hypothèses de recherches. Cette organisation se traduit par exemple par le fait que certaines variables sont considérées comme indépendantes et d'autres comme dépendantes. De la même manière, en ne considérant que les textes dans leurs rapports avec la langue, particulièrement dans l'optique de la linguistique transformationnelle, on fait apparaître certains traits des textes comme variables. Le développement de ce qu'on pourrait appeler la linguistique « traductrice » ou transformationnelle et la théorie des grammaires génératives (Chomsky et Miller, 1963; Quine, 1959) a permis de donner un sens précis au concept d'« événement linguistique » à l'intérieur du plan de la langue. Par événement, il faut entendre la manifestation de liaisons entre énoncés en relation d'expansion par exemple. Jusqu'à une période récente, on ne pouvait concevoir d'événement linguistique que dans le cadre de la linguistique diachronique et les phénomènes correspondants ne pouvaient se manifester en synchronie que sous la forme des pressions à l'intérieur des structures synchroniques de la langue par le jeu des principes fonctionnels ou d'économie. Ces événements ne pouvaient intéresser des procédures d'analyse de contenu puisque leur dimension dépassait largement celle des phénomènes étudiés. Pour l'analyse de contenu, la linguistique restait descriptive et ne pouvait fournir que des indices (même à l'intérieur du plan linguistique). Il n'en est plus de même dès qu'on envisage la construction de modèles opérationnels du langage définissant les transformations entre formes de sens donné, c'est-à-dire en relation sémantique d'expansion. Les formes deviennent alors des variables à l'intérieur d'un système. Toutefois, de la même manière que l'organisation des variables dans le plan vertical est celle de l'ordre des phénomènes étudiés, le système des variables du plan horizontal concerne l'ordre des énoncés par rapport à la langue.

Du point de vue de l'étude des processus de production, le statut de ces variables change. Globalement⁴, les variables du plan vertical deviennent des variables indépendantes, celles du plan horizontal, dépendantes. En d'autres termes, on construit le plan horizontal comme plan de variables dépendantes. Il n'en reste pas moins que, du point de vue des phénomènes étudiés, les variables du plan vertical gardent leur statut et leur organisation. A cause de la séparation entre variables dépendantes et indépendantes dans ce plan, la mise en œuvre de procédures expé-

4. A quelques nuances près, puisqu'il faut laisser la possibilité d'un retour tenant compte de l'autodétermination des processus de production dans leur déroulement séquentiel.

mentales de ce type peut conduire à la construction de véritables méthodes d'*analyse de contenu* que nous qualifierons de *systématiques*.

En effet, bien qu'on ne puisse pas déterminer simultanément les conditions de production et les manipuler, le fait qu'en tant que variables des phénomènes étudiés elles aient des statuts différentiels, qu'elles soient hiérarchisées, permet de les déterminer dans leur relation avec les phénomènes par une construction convenable du plan expérimental.

Les méthodes d'*analyse de contenu systématiques* s'intéressent à l'étude du langage et à son fonctionnement du point de vue des conditions de production exactement de la même manière que la linguistique par exemple vise l'ordre du discours et son fonctionnement du point de vue de la langue⁵. L'avantage de ces méthodes est qu'elles se situent dans le cadre de la liaison fonctionnelle entre les deux plans d'*analyse* et qu'elles permettent de faire se correspondre des variables sur ces deux plans. Par ailleurs, il est possible de délimiter avec précision le champ d'*investigation* alors que, dans le cas des méthodes taxinomiques, cette délimitation est toujours arbitraire.

Le passage des méthodes d'*analyse de contenu taxinomiques* à celles que nous venons de définir constitue un renversement. Les techniques taxinomiques sont des méthodes d'*observation*, leur rentabilité théorique est limitée comme celle de toute méthode d'*observation*. L'introduction de l'*expérimentation* dans les procédures correspond non seulement à un progrès méthodologique mais également à un stade plus avancé de la recherche. Les techniques d'*observation* ont longtemps suffi au besoin de celle-ci, elles s'avèrent maintenant insuffisantes. En particulier l'*analyse de contenu systématique* devrait jouer un rôle important dans l'étude des processus cognitifs sur lesquels on sait peu de choses et qui pourtant posent des problèmes centraux en psychologie sociale.

De ce qui précède, on peut dégager quatre types de problèmes intéressant l'*analyse de contenu*.

En premier lieu, il faut choisir entre méthodes taxinomiques et méthodes systématiques. Les premières conduisent à un classement d'éléments des textes en fonction de critères internes ou externes. Il reste à interpréter la signification de ce classement du point de vue de la recherche entreprise : l'étude de phénomènes particuliers. C'est choisir la voie de l'*induction* laquelle ne peut conduire à la construction de modèles théoriques parce que la liaison fonctionnelle entre les variables des deux plans d'*analyse* lui échappe. L'*analyse de contenu taxinomique* est une technique d'*observation*, elle ne peut être utile qu'à ce titre. Les méthodes systématiques ouvrent la voie à une démarche déductive en distinguant clairement les phénomènes étudiés et les processus de production des

5. Il est remarquable de constater que la linguistique s'est tournée vers l'étude du fonctionnement du discours du point de vue de la langue au moment même où les techniques de traduction automatique et de fabrication de résumés automatiques introduisaient en quelque sorte l'*expérimentation* en linguistique.

textes par l'*analyse* desquels est menée cette étude et en se pliant aux déterminations des uns et des autres. Elles permettent en cernant les liaisons entre les différentes variables de mettre à l'épreuve des modèles dynamiques des phénomènes étudiés. Le choix entre méthodes taxinomiques et systématiques coïncide donc avec celui qui est posé à l'occasion de toute investigation entre observation et expérimentation.

Si on opte pour des méthodes systématiques, c'est donc en fonction des objectifs et de la nature de la recherche entreprise. Il reste qu'il ne saurait y avoir de méthode générale d'*analyse de contenu systématique*. De par sa définition même, l'*instrument* doit être adapté au type de phénomènes étudiés. Ceci nous conduit à distinguer deux autres catégories de problèmes, lesquels ont précisément trait à la mise en œuvre des méthodes systématiques.

Le choix des modèles définissant les variables sur les deux plans d'*analyse* est fondamental. Il dépend pour les unes du développement antérieur de la théorie des phénomènes étudiés et pour les autres de l'*analyse linguistique*. Toutefois, dès ce stade, il est nécessaire de faire ces hypothèses sur les processus de production de manière à pouvoir considérer les variables *dans le cadre de ces processus et non plus dans leurs plans respectifs*. Ce sont ces hypothèses qui, en fin de compte, déterminent le choix des modèles particuliers sur l'un et l'autre plan d'*analyse*. Elles devront être formulées à partir de ce que l'on sait de la théorie des processus de production et c'est parce qu'il s'agit d'hypothèses dérivées de cette théorie générale qu'on devra avoir recours à l'*expérimentation*. Il est donc clair qu'on ne doit pas confondre les manipulations que l'on pourrait faire sur les variables en tant que variables de phénomènes étudiés et en tant que variables des processus de production.

En conséquence, la mise en œuvre des méthodes systématiques suppose qu'on ait clairement défini les plans expérimentaux et construit des situations expérimentales. En d'autres termes, la méthodologie expérimentale doit nécessairement intervenir et être adaptée.

Il reste une dernière catégorie de problèmes qui découlent du fait que ces méthodes de recherche se situent au carrefour de la linguistique et de la psychologie sociale. Ces deux disciplines se sont en quelque sorte tournées le dos, les linguistes dénonçant à juste titre le psychologisme et les psychologues assimilant le langage à un ensemble quelconque d'indices de phénomènes non linguistiques. On a donc confondu variables des textes et variables des phénomènes étudiés en niant, entre autres, l'*existence d'effets idéologiques*. On admet habituellement sans difficultés que la linguistique puisse apporter beaucoup à la psychologie sociale, mais inversement l'*analyse de contenu systématique* devrait jouer vis-à-vis de la linguistique un rôle analogue à celui de la traduction automatique. Elle conduit à une mise à l'épreuve des concepts et des méthodes et simultanément elle permet d'introduire dans la recherche linguistique elle-même une dimension nouvelle, celle des phénomènes dans lesquels

intervient le langage. Or, la prise en considération des déterminations que ces phénomènes imposent à la production du discours doit permettre de résoudre certains problèmes strictement linguistiques qui, autrement, resteraient obscurs. Nous croyons que les méthodes d'analyse systématique devraient fournir des indications utiles à la théorie linguistique de la génération des énoncés et à l'étude de la logique du langage.

- BALES, R. F. (1950), « A set of categories for the analysis of small group research », *Amer. Sociol. Rev.*, 1950, 15, 146-159.
- BARBICHON, G. et MOSCOVICI, S. (1965), « Diffusion des connaissances scientifiques », *Inf. Sciences Soc.*, 1965, 4-1, 7-22.
- BERELSON, B. (1952), *Content analysis in communication research*, The Free Press, Glencoe.
- CHOMSKY, N. et MILLER, G. (1963), *Introduction to the formal analysis of natural language*, in LUCE, R. D., BUSH, R. R., GALANTER, E. (eds), *Handbook of mathematical psychology*, Wiley, New York, Londres, vol. 2, pp. 269-322.
- CHOMSKY, N. (1967), *Cartesian linguistics*, Harper and Row, New York.
- CROS, R. C., GARDIN, J. C., LÉVY, F. (1964), *L'automation des recherches documentaires, un modèle général*, le « SYNTOL », Gauthier Villars, Paris.
- FODOR, J. et GARRETT, M. (1966), *Some reflections on competence and performance*. Commenté par SUTHERLAND, N. S. et COHEN, L. J., in *Psycholinguistics papers*. Proceedings of the 1966 Edinburgh conf. LYONS, J. et WALES, R. J. (eds), Edinburgh University Press.
- HARRIS, Z. S. (1952), « Discourse analysis », *Language*, 28, 1-30.
- HART, A. (1934), *Shakespeare and the homelies*, Melbourne University Press.
- HOLSTI, O. R., NORTH, R. C., ZANINOVICH, H. G., DINNES, D. A. (1963), *Content Analysis*, Northwestern University Press.
- LASWELL, H. D. (1938), « A provisional classification of symbol data », *Psychiatry*, 1938, 1, 197-204.
- MILLER, G. A. (1962), « Some psychological studies of grammar », *Amer. Psychologist*, 17, 748-762.
- MILLER, G. A., MC KEAN, K. (1964), « A chronometric study of some relations between sentences », *Quant. Journ. of Exp. Psychology*.
- MOSCOVICI, S., *Communication processes and the properties of language*, in BERKOWITZ L. (ed), *Advances in Experimental Social Psychology*, 1967, 3, Academic Press, New York et Londres.
- ZYGOURIS, R. et ACKERMANN, W. (1966), *Représentation et assimilation des connaissances scientifiques*. Doc. n° 3, ronéotypé CERP/AFPA.
- OSGOOD, C. E. (1962), *Symbols of Democracy*, Stanford University Press.
- OSGOOD, C. E., SAPORTA, S., NUNNALLY, J. C. (1956), « Evaluative assertion Analysis », *Litera*, 3, 1956, 47-102.
- POOL, I. de S. (1959), *Trends in Content analysis*, The University of Illinois Press, Urbana, Ill., 189-233.
- QUINE, N. V. (1959), *Meaning and translation*, in BROWER, R. A. (ed), *On Translation*, Harvard University Press, Cambridge Mass.
- STONE, P. J., DUNPHY, D. C., SMITH, M. S., OGILIRE, D. M. (1966), *The general inquirer a computer approach to content analysis in the behavioral sciences*, Mit Press, Cambridge. Mass.
- WHITE, R. K. (1951), *The study of behavior: Q-technique and its methodology*, Glen Gardner, New Jersey. The society for the psychological study of social issues.
- ZYGOURIS, et ACKERMANN, W. R. (1967), *Code d'analyse et domaine de référence*. Bulletin du CERP.

MAURICE MOUILAUD

Faculté des Lettres et Sciences humaines de Poitiers
Institut français de Presse

LE SYSTÈME DES JOURNAUX
(Théorie et méthodes pour l'analyse de presse.)

Cette étude est consacrée aux moyens d'information, plus précisément à la presse, et au sein de celle-ci, au journal quotidien, elle utilise des concepts qui présentent une analogie avec ceux de la linguistique : on ne prétendra pas que ces concepts soient les seuls possibles, qu'ils couvrent tous les besoins de l'analyse de presse ni qu'ils effacent les voies de recherches qui existent et qui sont certainement fructueuses. Mais il s'agit de dégager des problèmes qui ont été masqués et qui constituent un domaine d'investigation inexploré; ce domaine, nous l'appelons l'étude du journal comme champ d'information, expression que nous chercherons à justifier; en effet, la situation des études de presse ressemble à celle que Saussure décrivait en linguistique : caractère hétéroclite des points de vue et absence de théorie ; c'est qu'on ne s'est pas assez soucié de définir l'objet auquel on avait affaire; pourtant, la plupart des tentatives de définition du journal avouent leur embarras; les définitions qui existent sont faites de concepts hétérogènes qui juxtaposent les caractères les plus visibles, mais aussi les plus extérieurs de leur objet; bref, il se pose un problème de méthode et de théorie, l'une et l'autre étant, comme ailleurs, deux aspects d'une même tâche.

* *

Nous évoquerons d'abord quelques grands types de recherches qui sont pratiqués; le premier pourrait être appelé : les recherches historiques; les études des historiens consacrées à la presse sont relativement récentes et rares; cela se comprend : pour l'historien, le journal est d'abord une source où il puise l'événement; on pourrait dire que dans un premier temps cette source ne l'intéresse pas pour elle-même : elle n'est qu'un tremplin vers l'actualité; l'être de la presse n'apparaît que pour disparaître; le journal est considéré comme un moyen et son contenu est essentiellement l'objet d'une critique. Cette lecture armée n'est pas radicalement