

QUEL "OBJET" POUR UNE "ANALYSE STATISTIQUE DU DISCOURS" ? QUELQUES RÉFLEXIONS À PROPOS DE LA RÉPONSE "ALCESTE"

Max Reinert
CNRS - ERSS & SLADE
Université de Toulouse-le Mirail
reinert@cict.fr

« Aussi objectifs que nous nous désirons, nous sommes donc engagés par le seul fait que nous travaillons sur le sens, que nous soyons sociologues, linguistes, historiens, économistes ou tout autre chose. » P. Achard (1997)

Abstract

Since Harris, several formal approaches to discourse analysis have been attempted (Pêcheux). The *Alceste* method has several points in common with Benzécri's research on the Analysis of Textual Data and with the discourse analysis approach developed by SLADE (Achard). The distribution of vocabulary within a particular discourse is studied in order to display the formal traces of language uses, or language games (Wittgenstein). This is in line with Harris' original aims.

Studying several *corpi* has enabled demonstration of the distinct ways in which vocabulary is distributed, and led to postulation of the idea of *lexical worlds*. These *worlds* indicate the most usual points of enunciation (*topoï*), i.e., sorts of spaces where the speaker takes up a position to coordinate the points of interest at a given moment of enunciation. We hypothesize that *all discourse automatically brings into play a system of topoï* and that a system of *topoï*, as a system of common places, is statistically definable.

The *Alceste* method offers a purely formal and entirely computerized way to isolate these *topoï*.

La méthodologie *Alceste*, au niveau de ses techniques, est une méthodologie de *statistique textuelle* selon la terminologie de Lebart et Salem. Elle se situe en continuité avec l'approche Benzécriste. Il s'agit surtout de rendre compte de l'organisation interne d'un discours plutôt que de rendre compte de différences statistiques entre les divers textes d'un corpus. Le problème posé est de savoir, en définitive, quel est l'objet visé par l'analyse. Notre propos, en cherchant à répondre à cette question, est de bien situer cette méthodologie, au niveau théorique, comme une méthodologie d'analyse de discours.

Cette étude comprend deux parties :

Dans la première, on évoquera rapidement les travaux de J.P. Benzécri qui sont à la base de notre méthodologie et le modèle de tableau qu'il propose de type "sujets x prédicats"

Dans la seconde, on présentera nos hypothèses en montrant à la fois la continuité de notre modèle avec le modèle précédent et le changement de point de vue impliqué par la modélisation du corpus "énoncés x lexèmes". On essaiera de montrer, d'une part la continuité de l'orientation générale qui est celle d'une analyse de discours et, d'autre part, en quoi ce changement de point de vue implique un changement d'objet de l'analyse.

1. L'analyse statistique du discours

1.1 Origines et objectifs

Comme on l'a mentionné, nos travaux se situent dans la continuité du travail de J. P. Benzécri dont les techniques statistiques ont été utilisées dès l'origine pour l'analyse de discours. Cet auteur évoque les travaux de Harris et son approche distributionnelle comme une source de son inspiration.

Comme on sait, Harris montra que l'approche distributionnelle appliquée à un discours unique pouvait mettre en évidence des traces de lois externes impliquées dans la production du discours [Harris, 1952] : *"L'analyse distributionnelle à l'intérieur d'un seul discours, considéré individuellement, fournit des renseignements sur certaines corrélations entre la langue et d'autres formes de comportement."* [Langage n° 13 mars 1969 traduit par Fr. Dubois-Chalier, p11].

Cela dit, la mise en oeuvre de la notion de distribution chez Harris se fonde essentiellement sur des aspects syntaxiques et transformationnels et non pas statistique. Cependant, l'objectif de l'approche Benzécriste, comme de la nôtre, reste globalement le même que celui de Harris dans la mesure où l'on pense que l'organisation interne des éléments d'un discours "mémorise" par sa forme même des processus externes qui ont conduit à sa production.

1.2. Le modèle "sujets-prédicats"

Benzécriste présente un exemple d'analyse, exemple simplifié et très didactique, dans son livre « Pratique de l'analyse des données : linguistique & lexicologie » qui se réfère directement à l'approche Harissienne. L'analyse de cet exemple permettra de mettre en place concrètement une première conception de la modélisation d'un corpus par un tableau de données et de la signification de cette modélisation. Le corpus considéré est un ensemble de propositions construit artificiellement, chaque proposition n'étant formée que de la conjonction de deux termes : le sujet et le prédicat :

{avion voler; chacal aboyer; avion ronfler; chat dormir ; chat miauler ; etc.}.

Ce corpus peut être représenté par un tableau de données à double entrée, avec en lignes, les sujets et, en colonnes, les prédicats, chaque constituant ou proposition dans ce cas étant repéré par l'association d'un sujet avec un prédicat :

	aboyer	dormir	manger	miauler	ronfler	voler
avion	0	0	0	0	3	7
chacal	1	1	1	0	1	0
chat	0	5	3	4	1	0
chien	3	3	3	0	3	0
coyote	2	1	1	0	1	0
moteur	1	0	3	0	8	0
oiseau	0	2	3	0	0	6

A l'intersection d'une ligne et d'une colonne figure le nombre de propositions du corpus qui associent tel sujet à tel prédicat. Ainsi, les fréquences observées dans le tableau de données permettent d'avoir un indice de la probabilité de rencontrer dans le "monde_" concerné tel ou tel comportement en supposant que le corpus ait été consigné par un ensemble d'enquêteurs rigoureux qui décriraient chaque "état de choses" par une proposition dans un ordre aléatoire.

Quoiqu'il en soit, le tableau de données modélise complètement le corpus dans la mesure où l'ordre des propositions est jugé sans signification (lié à l'aléa du parcours de chaque enquêteur).

En effet, on peut entièrement le reconstruire à partir du tableau de données, à l'ordre des propositions près. Autrement dit, analyser la forme du corpus dans son aspect distributionnel - au sens de Harris - revient à analyser la forme de ce tableau. L'analyse factorielle des correspondances permet une représentation iconique de cette "forme".

Mais qu'appelle-t-on la "forme" du discours ? Quels sont les présupposés d'une telle modélisation ? C'est par la jonction d'un sujet et d'un prédicat qu'un jugement - vrai ou faux - peut être énoncé

et donc qu'un lien entre un fait et une proposition est en quelque sorte constitué. Cette relation constituante entre propositions et faits apparaît au niveau de l'expérience scientifique usuelle par cette possibilité de considérer des objets et de leur attribuer des propriétés (éventuellement mesurables) à l'intérieur d'une modélisation particulière. On identifie ainsi les propositions du modèle à des faits du type "tel objet a telle propriété".

Autrement dit, la manière d'attribuer une propriété à un objet dans ce qu'on appelle un fait se reflète dans la forme logique de la proposition. Si les propositions reflètent les faits, un discours, comme ensemble de propositions est un modèle du monde où "existent" ces objets. On ne fait que suivre là le Wittgenstein du *Tractatus* pour qui : « 1.1 - *Le monde est l'ensemble des faits, non des choses. 1.13 - Les faits dans l'espace logique constituent le monde* »

Notons que nous ne sommes pas dans une attitude purement descriptiviste : le lien constitutionnel d'un fait et d'une proposition implique qu'un fait ne se définit pas en lui-même mais en relation à un espace logique, c'est à dire, en relation avec ce qu'on peut en dire à travers l'ensemble des propositions, à travers un certain discours. Même pour Claude Bernard : « Ce sont les faits qui jugent l'idée... Les faits sont la seule réalité qui puisse donner la formule à l'idée expérimentale, et lui servir en même temps de contrôle, *mais c'est à condition que la raison les accepte* ». C'est cette idée qui est reprise implicitement par Wittgenstein : un fait ne peut véritablement être identifié comme fait qu'en relation avec une manière générale de se représenter le monde (l'espace logique).

Notre tableau de données étant un équivalent du discours peut être considéré comme un modèle du monde. De ce point de vue, le tableau de données de Benzécri n'est pas si éloigné du tableau logique de Wittgenstein car il exprime par sa forme même une certaine "ressemblance" avec le "monde" et on pourrait intituler cette sorte de théorie comme une théorie du reflet.

En effet, ni Wittgenstein ni Benzécri n'identifie le "monde" à sa forme logique (ou distributionnelle), qui n'en est que le reflet. L'évocation du mythe de la caverne par Benzécri suggère à la fois une ressemblance du reflet avec son monde et une certaine dégénérescence de ce qui est représenté (à l'image du monde matériel relativement au «monde des idées platoniciennes»).

Un chemin parallèle est perceptible chez Wittgenstein lorsqu'il constate (*Tractatus*) «3.221 - *Je ne puis que nommer les objets. Les signes les représentent. Je ne puis que parler des objets, je ne saurais les prononcer. Une proposition ne peut que dire d'une chose comment elle est, non ce qu'elle est* »

Cela dit, dans cette manière de considérer les propositions comme reflets des "états de choses dans le monde", il y a là comme une sorte d'aporie, car, d'une part, les "états de chose" définissent le monde comme premier par rapport au langage et, d'autre part, ces "états de chose" ne peuvent eux-mêmes être définis que par la possibilité d'un jugement propositionnel qui les fonde comme "faits". La contradiction vient bien sûr de la présupposition des "états de chose" que les propositions ne feraient que décrire.

Cette conception a été remise en question par le Wittgenstein des *Investigations philosophiques* et par l'école de Palo-Alto .

2. Le modèle "énoncés-lexèmes"

La modélisation proposée par J.P. Benzécri n'est pas la seule à être utilisée dans la pratique de l'analyse des données en sciences humaines. Les enquêtes par questionnaires - ouverts ou fermés - conduisent plutôt à considérer, au moins dans une première approche, des tableaux à double entrée croisant, en lignes, des individus et, en colonnes, leurs différentes réponses. Autrement dit, par analogie au modèle précédent, on fait jouer dans cette modélisation le rôle de "sujet de la

proposition" à des "individus" et le rôle de "prédicat de la proposition" à des "réponses". Quel est le sens de cette pratique relativement au modèle précédent ? Il y a, en tout cas, un changement de perspective puisque l'on passe de la modélisation d'un monde conçu comme ensemble de faits à la modélisation d'un monde conçu comme un lieu d'interaction entre individus. Au niveau de la notion de discours, on passe d'un discours - reflet logique du monde - à un discours - trace des points de vue dans une communauté donnée.

Lorsque la réponse est verbale, ces notions d'individu et de réponse sont cependant faussement claires. Qui parle, de quelle place et pour qui ? que dit-il en définitive ?

Notre modèle est intermédiaire entre ce dernier et le modèle sujet-prédicat. Mais avant d'exposer cela, il est nécessaire d'aborder rapidement l'analyse de la proposition pour C. S. Peirce (qui est, rappelons-le, le philosophe-logicien américain de la fin du siècle dernier qui a créé la sémiotique et qui est à l'origine de la pragmatique).

2.1. L'analyse de la proposition pour Charles Sanders Peirce _

Reprenons rapidement la notion de proposition et plus précisément celle de vérité d'une proposition. En tant qu'une proposition parle de faits représentables dans un certain espace logique, sa "vérité" peut-être soit déduite de la "vérité" d'autres propositions, soit induite d'une expérience. La vérité, considérée d'un point de vue formel et déductif, sera plutôt appelée validité ; Ces deux aspects, vérité expérientielle et validité, qui semblent indépendants sont cependant intrinsèquement liés car seule la déduction permet de reconnaître dans une expérience une sorte de permanence du "fait" et vice versa, sans une expérience minimale, il est impossible de poser les premiers axiomes à partir desquels déduire.

La notion de vérité, pour être interprétée, implique une réflexion plus générale sur ce qu'est un signe, avec ses aspects perceptifs, dynamiques et logiques. Il s'agit donc d'un problème de sémiotique. Ce type de réflexion a été tenté par C. S. Peirce et c'est l'insuffisance d'une réponse dans un cadre purement logique qui a conduit cet auteur à concevoir une sorte de logique étendue à la genèse des signes. En effet, la logique étant l'expression d'un jeu entre signes déjà constitués, on ne peut attendre d'elle qu'elle en exprime le processus de formation. C'est dans cet essai de prolongement de l'analyse logique au processus même de formation des signes que Peirce a proposé d'appeler "sémiotique" cette nouvelle "science" des signes.

A la question posée sur l'instance de la vérification du jugement sujet-prédicat, la réponse de Peirce fut de plonger cette vérification dans un processus sémiotique continu, plus exactement dans une série de "sémioses" passant par différents stades qu'on résumera de manière très schématique en trois mots : d'abord la perception, puis l'expérimentation, et enfin l'interprétation (si l'on entend par là, la construction d'un nouveau signe). Ce n'est qu'à ce dernier stade qu'existe une possibilité d'un jugement de vérité mais cela demande que soit suffisamment développée la cohérence symbolique du signe interprétant. Autrement dit, si la notion de validité est théoriquement séparable de la notion de vérité, cette dernière ne peut être concrètement séparée, la cohérence logique étant nécessaire à l'appréciation de l'expérience comme occurrence d'un même loi. On retrouve là cette inséparabilité du fait et de la forme logique.

Selon Peirce, la sémiose est un processus triadique. Ce par quoi un phénomène est reconnu et représenté pour quelqu'un, passe successivement par trois phases appelées par leur numéro d'ordre : priméité, secondéité et tiercéité. Pour imager ce processus, considérons le mot "sens". La construction du "sens" peircien de la sémiose, déroule successivement les trois acceptations de ce mot en français, allant du sens perçu en tant que sensation (priméité) au sens de la signification et de la représentation (tiercéité) en passant par le chemin qui le construit, c'est à dire, le sens comme mouvement dynamique, comme acte intentionnel concret (secondéité).

Certes, il est bien difficile d'introduire les catégories de Peirce en quelques mots. Illustrons ce processus triadique en prenant un exemple concret. On choisira le cas de cette institutrice prise en otage avec sa classe, voici quelques années. Lors d'une émission de "La marche du siècle", elle déclara que la première pensée absurde lui ayant traversé l'esprit, alors même que le preneur d'otage, cagoulé et armé, faisait irruption dans sa classe, fut cette proposition : «c'est le coursier». Cette formulation est une pure proposition sur un état de chose. Elle impliqua immédiatement son absurdité. C'est justement l'accès au niveau symbolique qui permet une distanciation de la première impression. Ce niveau du symbolique est celui de la tiercéité chez Peirce. Il est généralement le seul niveau qui occupe notre conscience.

Mais la première impression de l'institutrice, le *fondement* de son hypothèse selon la terminologie de Peirce, est tout entier contenu dans ses habitudes et, de cela, on ne peut rien dire. En effet, la seule personne inconnue qui est susceptible d'apparaître dans une classe pendant le cours est un coursier. Notons que si le *fondement* a motivé le choix du prédicat "coursier", ce prédicat n'était évocable qu'à travers une proposition et ne peut être confondu avec son *fondement*. Le *fondement* ne peut être jugé. Il se contente d'être (ou plutôt de s'imposer). Il est simple retour du "même" dans la diversité des apparences. Ce niveau relève de la pure priméité.

Quant au moment même de la prise d'otage en tant qu'expérience vécue par l'institutrice, il relève de la secondéité car c'est ce "ici et maintenant" qui met en contact, dans l'esprit de l'institutrice, le fondement de son impression première avec un existant (son corrélat). Cette mise en contact n'est pas du à une opération abstraite. Il s'agit d'un acte, d'une imposition. Ce qui est agi par ce contact est du reste double : il existe une pression des existants sur la situation du sujet pensant tel qu'il la perçoit. Mais le sujet n'est pas purement passif : son action consiste à porter son attention sur le lien entre le corrélat et le fondement. C'est du reste en appréciant ce lien dans la proposition "c'est le coursier" que l'institutrice arrive à juger de sa validité.

La proposition n'est pas seulement une pure forme logique pour se représenter quelque chose mais bien, en l'occurrence, un véritable acte pour se dégager d'une sorte d'envahissement angoissant de la réalité (qui aurait pu simplement pétrifier notre institutrice). Elle est donc tout autant l'expression d'un acte pour contenir à l'extérieur de soi ce qui apparaît comme une menace que l'expression d'un état de choses.

Ainsi, l'approche de Peirce introduit une analyse plus pragmatique de la proposition. Par exemple, dans la proposition : "c'est le coursier", le sujet est d'abord défini par ce qui est montré en situation et le prédicat est défini par l'hypothèse de la locutrice.

La proposition n'est plus la simple expression logique du lien unissant un objet à une propriété, elle recouvre un acte, une assertion engageant quelqu'un dans une situation donnée.

Reprenons un exemple banal de proposition comme "Le ciel est bleu", en situation, où "le ciel" est sujet et "bleu", prédicat. Pour Peirce, la décomposition sujet/prédicat est toujours opérante dans ce cas mais sa signification est changée : "Le ciel" recouvre ce dont on veut parler, ici et maintenant, au moment où on parle et que l'on peut éventuellement montrer ; Quant au prédicat "bleu", il est une manière du locuteur de poser une hypothèse en rapport avec son usage du monde. Cela dit, cette proposition simple est déjà complexe car si le propos principal semble bien affirmer un lien entre ciel et bleu, le syntagme "le ciel" peut aussi être considéré comme une mini-proposition dans laquelle "ciel" est le prédicat (ce que l'usage du sujet lui permet de reconnaître en situation) et le déterminant "le" est sujet car c'est par ce déictique qu'un existant est montré en situation (comme le démonstratif "ce" dans "c'est le coursier").

Par rapport à ce point de vue, la proposition peut donc être vue comme la proposition de "quelqu'un" sur "quelque chose" et on passe insensiblement d'un modèle "sujet-prédicat" à un modèle "individu-réponse". Mais si la matérialité de ce "quelqu'un" est figurée en quelque sorte

par l'occurrence même de la proposition (qui implique bien un acte), ce sur quoi porte la proposition (son objet) reste encore des plus mystérieux.

2.2. Le fondement topique des énoncés

Cette analyse conduit à plusieurs remarques :

Ce processus triadique s'applique à toute sémiologie qui à partir d'un contenu quelconque de l'esprit (saisi comme *fondement* d'un signe) se développe dans un nouveau signe (son *interprétant*). Cet interprétant - en tant que signe - peut être perçu à nouveau par quelqu'un et donc sera également développé en un nouveau signe, etc.. Il s'en suit une certaine récursivité du processus de sémiologie qui conduit à une fractalisation des contenus. Par exemple, cette fractalisation de la représentation expliquerait bien, au niveau syntaxique, l'enchâssement des niveaux de propositions comme on l'a vu, avec l'énoncé "le ciel est bleu", la place vide de l'objet dans la proposition pouvant être occupée par une autre proposition et ainsi de suite.

Cette réflexion à peine ébauchée suffira ici pour justifier notre passage de la notion de proposition à la notion d'énoncé et à notre mode d'opérationnalisation de cette notion.

D'une part, un énoncé n'est pas une entité que l'on peut définir avec précision du fait même de cette fractalisation de l'énonciation, de l'aspect continu du processus sémiotique, processus qui se poursuit bien au delà de la phrase. Qu'on pense à la notion de paragraphe, de chapitre, de livre, des oeuvres complètes d'un auteur, d'une époque, etc..

D'autre part, elle relativise la distinction entre sujet et prédicat à un moment de l'analyse. On peut par contre distinguer dans un énoncé au moins deux types de mots : ceux qui sont associés par l'usage à une notion ou un concept (les catégorèmes_ou mots pleins) et donc, qui peuvent être utilisés à un moment ou à un autre en position de prédicat pour "dire à propos de quelque chose" et les autres mots qui servent plutôt à relier ces notions entre elles ou à des existants en situation dans un espace pragmatico-logique particulier (les syncatégorèmes ou mots vides), le problème étant qu'on ne peut que montrer ce dont on veut parler mais qu'on ne peut pas le dire (on ne donne que des hypothèses à propos de ce dont on veut parler).

Par exemple dans la proposition «c'est le coursier», le seul mot plein est «coursier». Les autres mots ne servent qu'à indiquer qu'il y a bien quelque chose qui se passe, ici et maintenant.

Aussi, en reprenant l'analyse de Peirce, on distinguera dans un énoncé trois aspects :

- 1) son *fondement* qui est relatif à l'impression première laissée par l'énoncé en lui-même.
- 2) sa *dynamique* : le fait que l'énoncé est l'expression d'un acte intentionnel en situation, d'un engagement, et répond à une situation (le plus souvent composée d'autres actes intentionnels).
- 3) sa *représentation* : le fait que l'énoncé renvoie à une certaine schématisation de ce dont on veut parler, à un moment du discours ou du dialogue, ce que Grice appelle un micro-univers (amorce d'un espace logique entre colocuteurs qui permet un certain consensus).

Remarquons que ces trois aspects reprennent, à un autre niveau plus général de l'analyse, la forme même de la proposition de départ : le *fondement* est ce qui joue, en quelque sorte, le rôle de prédicat, le rôle des hypothèses dans une expérimentation ; l'*énoncé* en tant que trace d'un acte joue le rôle de cette place de l'objet qui ne peut être que montré en situation (par quelqu'un). L'énoncé-acte amalgame justement, en une même trace, l'objet dont on parle et l'individu qui le montre. Quant à la représentation, elle permet justement de séparer l'objet logique du sujet de l'énonciation, d'objectiver l'acte en le référant à son fondement.

Ainsi nous proposons de passer d'une modélisation du discours de type sujet-prédicat à une modélisation du discours de type "énoncés-occurrences" - "fondements".

Reste à définir de manière plus opérationnelle cette notion de *fondement*. Le *fondement* d'un énoncé, par définition, ne peut être décrit avec précision (il serait alors représentation et non

fondement). Pour notre part, nous nous sommes intéressés à un aspect qui nous paraît être une part essentielle de l'impression qu'il laisse mais qu'on ne peut véritablement identifier au fondement lui-même sans un abus de langage. C'est la raison pour laquelle nous avons choisi de l'appeler le *fondement topique* de l'énoncé.

Pour concevoir ce qu'on peut entendre par là, considérons ces vers de Reverdy « *L'heure pleine est passée sur une autre qui sonne. Les pas des voyageurs courent déjà loin* » la simple présence des mots pleins comme *heure, passée, pas, voyageurs, courent, loin* dessinent dans ce cas une isotopie du passage indépendamment de la manière dont sont reliés ces mots au niveau syntaxique. Cette première impression isotopique de l'énoncé nous l'appelons son *fondement topique*. Mais contrairement à l'isotopie, on ne cherchera pas à la thématiser ou à l'identifier dans un "sème".

Notre hypothèse est que ce *fondement topique*, dans ce qu'il a de plus archaïque, c'est à dire de plus inconscient, n'est pas attaché à un mot singulier ni à aucune forme en particulier et qu'il se diffuse globalement dans tous les éléments de l'énoncé, et principalement dans le choix des mots pleins indépendamment de leur position syntaxique. Autrement dit, notre hypothèse consiste à dire : *la trace lexicale des mots pleins d'un énoncé est un indice de son fondement topique*.

Donnons un autre exemple : P. Achard (1993) définit le discours comme « *l'usage du langage en situation pratique, envisagé comme acte effectif, et en relation avec l'ensemble des actes (langagiers ou non) dont il fait partie* ». (p 10). Si l'on compare cette définition avec celle du "Quillet", pour qui le discours est « *Toute espèce de manières d'exprimer verbalement sa pensée* », on constate qu'aucun mot n'est commun entre ces deux définitions. Le fondement topique de l'énoncé du "Quillet" prend évidemment racine dans une vision plus cognitive du discours, que la marque des seuls mots pleins suffit à appréhender : *espèce, manière, exprime, verbe, pensée*.

Dans la définition de P. Achard, l'accumulation de termes comme *usage, situation, pratique, acte, effectif*, oriente vers une vision plus dynamique, plus pragmatique du discours. Cette classe de mots, indépendamment des règles syntaxiques, définit bien une sorte d'isotopie activant un certain lieu de "l'arrière plan" (le *Background* de Searle) du colocuteur. Notons à ce propos que la force perlocutoire de l'énoncé ne dépend pas uniquement de la forme de l'énoncé ni même de l'intentionnalité consciente du locuteur (ordre, promesse, prière, objection, etc.) mais aussi de son *fondement topique* même, de cette possibilité d'activer chez le colocuteur un "lieu" particulier à partir duquel il pourra comprendre ce qui est dit. Il est clair que l'enveloppement du colocuteur par un ensemble de mots relatifs à ce lieu de "l'arrière plan" est en rapport direct avec l'intensité de la force perlocutoire, intensité d'autant démultipliée que ce lieu aura été souvent activé dans le passé.

Dans cette manière de voir, la relation des propositions aux états de choses se complexifie singulièrement en ce sens que la proposition ne correspond plus à un état de choses en soi dans une identité fermée mais doit être intégrée à un processus discursif ouvert, l'acte discursif se complexifiant au fur et à mesure des ajustements à des niveaux logiques divers, avec des compromis divers en fonction d'une pression de la situation, de "ce qui se passe ici et maintenant" perçu d'abord inconsciemment, de manière globale et synthétique, en relation avec nos usages. Aussi sous la complexité de la forme, parfois même ses contradictions au niveau logique, on peut espérer trouver la simplicité de son fondement topique (le fait que le sujet parlant se situe dans un certain lieu).

D'un point de vue plus opératoire, cette analyse conduit à distinguer très grossièrement dans un énoncé deux types d'unités : les "lexèmes" et les "morphèmes grammaticaux". Les lexèmes sont associables à des fondements topiques orientant l'interprète vers des "lieux" privilégiés par les

habitudes sociales d'un groupe et de son idéologie. Les autres morphèmes sont davantage impliqués dans la mise en forme logico-sémantique de l'énoncé, en situation. Pour le contexte situationnel, on retrouve les déictiques, les modalisations, les temps ; pour le contexte logique, les connecteurs, les anaphores, les quantificateurs.

2.3. La modélisation proposée

Si l'on reprend l'exemple proposé par Benzécri, la modélisation serait donc plutôt :

	aboyer	dormir	manger	oiseau	chacal	chat
propos 1	1	0	0	0	1	0
propos 2	0	1	0	0	1	0
propos 3	0	1	0	0	0	1

Ce tableau, comme le précédent, permet, dans le cas particulier de l'exemple, de reconstruire le corpus d'origine. L'analyse des deux tableaux est, du reste (dans ce cas) équivalente. Il met en relation trois types d'entités : a) les termes (en colonnes) ; b) les propositions (en lignes) ; c) Le corpus (le tableau).

Il est remarquable que ces trois niveaux d'analyse sont aussi évoqués par Peirce dans sa division triadique des "symboles" (c'est à dire les "signes" conventionnels, dont les productions langagières font partie).

Peirce distingue en effet trois sortes de symboles : les "termes", les "propositions" et les "arguments" :

1) les termes ne jouent que le rôle de marques de quelque chose par habitude (nos mots pleins ci-dessus); (ils relèvent de la priméité sous cet aspect).

2) les propositions sont susceptibles d'être jugées, vraies ou fausses, du fait que, dans la proposition, une marque est mise en rapport avec un objet particulier (en situation : nos propositions) ; seul un acte peut permettre de saisir l'adéquation entre une proposition et ce qu'elle montre (secondéité).

3) enfin les arguments qui déterminent une manière d'interpréter l'ensemble de ces propositions dans une démarche déductive. C'est par là que l'activité de l'interprète s'harmonise avec les contraintes exprimées dans la mesure où chacune des propositions de base est valide. Il y a alors comme un déploiement des contraintes réelles dans des contraintes logiques_. C'est le lieu même des représentations. Ce niveau relève de la tiercéité.

Nous voulons montrer par là qu'il n'y a pas rupture au niveau de la forme entre notre modélisation et celle proposée par Benzécri dans le cas d'un corpus de propositions. Cela dit, dans le cas d'énoncés naturels, cette modélisation se prolonge en conservant une correspondance triadique : il suffit de mettre en colonnes, les "lexèmes" et en lignes, les "énoncés" en tant que prise de position des énonciateurs. Ces énoncés ont toutefois des frontières incertaines_, ce qui nous a amené à effectuer plusieurs essais de classement en faisant varier leur longueur pour ne retenir que les classements stables (pour la définition opérationnelle des énoncés, se reporter à Reinert, 1993).

	mot 1	mot 2	mot 3	mot 4	mot 5	mot 6
énoncé 1	1	0	0	0	1	0
énoncé 2	0	1	1	0	1	0
énoncé 3	1	1	0	1	0	1

En résumé, ce modèle ne cherche plus à représenter la forme logique d'un ensemble de propositions mais à "cartographier" les principaux *fondements topiques* ou "*lieux*" sur lesquels ce monde du discours est construit par des énonciateurs. Cela dit, il est nécessaire de distinguer la notion de *fondement topique* de celle de *topos* comme on distingue l'occurrence de son type. Les *topoi*, en tant que pré-système de lieux se définissent par leur opposition, alors que le fondement topique ne se définit pas : il est ce qui fonde la croyance d'être quelqu'un - quelque part.

Cette opération reste cependant une opération logique en ce sens que si le discours est un discours à propos de quelque chose, en séparant les lieux, on sépare du même coup les différents points de vue sous lesquels on peut considérer cette "chose", chose qui restera toujours dans un au-delà du discours. De plus l'iconification des points de vue permet de séparer ce qui vient des sujets et de leur interaction. En se coordonnant dans une même représentation, les points de vues peuvent finir par apparaître comme les différents aspects d'un même objet, objet qui, en quelque sorte, émerge "objectivement" de sa forme logique. C'est par ces interactions qu'une construction de la réalité peu à peu se codifie dans des lois "objectives". Évidemment, le modèle d'analyse proposé est une première et grossière approche. Elle ne permet aucunement de différencier avec précision ces lois mais elle donne une première base à l'analyste pour articuler ses conceptions.

Conclusion

De même que, selon Bourdieu, les activités humaines sont structurées par des "habitus", les discours, qui ne sont que les traces langagières de ces activités (Achard), sont structurés par des systèmes de "lieux" (ou *topoi*) agissant comme des attracteurs pour le locuteur. Un "lieu" ne peut véritablement être défini en soi. Un lieu se définit par l'existence même d'autres lieux auquel il s'oppose et sans lesquels il perd son identité. Dans le cadre de son discours, un locuteur doit gérer au fil du processus énonciatif ce passage d'un lieu à l'autre qui est aussi le passage d'une identité à l'autre. Cela est le propre même de l'argumentation rhétorique et dialectique. En conséquence, la mise en place d'un système de lieux n'a pas qu'une visée référentielle, elle vise également à gérer l'activité d'un « *locuteur* » dans la manière de définir ses positions d'énonciation. Dans un corpus particulier, les "lieux habituels" peuvent être indexés par leurs « *mondes lexicaux* ». Cela, une analyse statistique peut le montrer de manière relativement automatique. C'est l'objectif même de la méthodologie "Alceste" (pour « Analyse des Lexèmes Cooccurrents dans les Énoncés Simples d'un Texte »).

REFERENCES

- Achard, P. (1993) *La sociologie du Langage*, PUF
- Achard, P. (1997) L'engagement de l'analyste à l'épreuve d'un événement, *Langage et Société*, 79, 5-38.
- Anscombe, J. C., & coll. (1995) *Théorie des Topoi*, Kimé.
- Bakhtine, M. (Tr. fr. 1977), *Le marxisme et la philosophie du langage*, Éditions de Minuit.
- Benzécri, J.P., & coll. (1981) *Pratique de l'Analyse des Données : linguistique et lexicologie*, Dunod.
- Bourdieu, P. (1982) *Ce que parler veut dire*, Fayard.
- Chauviré, Ch. (1979) Peirce, le langage et l'action, *Les Études philosophiques*, n° 1.
- Grize, J.B. (1982) *De la logique à l'argumentation*, DROZ, Genève.
- Harris, Z.S. (1952) Discourse analysis, *Language*, 28, 1-30 [tr. fr. Dubois-Chalier, *Langage* n° 13 (1969)]
- Lalande, A. (1988, 1ère ed 1926) *Vocabulaire technique et critique de la philosophie*, PUF.
- Lebart, L., Salem, A. (1994) *statistique textuelle*, Dunod.
- Pêcheux, M. (1969) *L'analyse automatique du discours*, Dunod.
- Peirce, Ch. S. (tr. fr. 1987) « *Textes fondamentaux de sémiotique* », Klincksieck.
- Reinert, M. (1993) « Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars », *Langage et Société*, 66, 5-39.

Reinert, M. (1997) Les "mondes lexicaux" des six numéros de la revue "le surréalisme au service de la révolution", *Cahiers du centre de recherche sur le surréalisme (Mélusine)*, L'age d'Homme, XVI, 270-302.