

(1995) - LAHLOU, Saadi. Vers une théorie de l'interprétation en analyse des données textuelles. JADT 1995. *3rd International Conference on Statistical Analysis of Textual Data*. S. Bolasco, L. Lebart, A. Salem (eds). CISU, Roma, 1995, Vol I, pp. 221-228.

VERS UNE THEORIE DE L'INTERPRETATION EN ANALYSE STATISTIQUE DES DONNEES TEXTUELLES

Saadi LAHLOU, EDF - Direction des Etudes et Recherches.
1, ave. du Gl. de Gaulle. 92141 Clamart France. saadi.lahlou@edf.fr

Summary : A theory of the interpretation process by the analyst in statistical analysis of textual data is proposed. Interpretation is presented as an abductive process through which the analyst gradually adjusts the results of analysis and elements of an ideal model using his previous knowledge of the real world (out of the text), so that the final result consists in two corresponding patterns (results & model). The abductive process for the interpretation of a class is based on the search of the « smallest common paradigm » of which most characteristic traits of this class can be considered as acceptable representants. This approach insists on the biases coming from limited corpus relevance and the analyst's knowledge of the world.

1. Introduction à une tentative téméraire de dévoilement du subreptice

Le champ de l'analyse statistique des données textuelles s'étend rapidement, par la disponibilité des logiciels et des corpus numérisés. Pourtant processus d'interprétation sur lequel reposent ces techniques, presque jamais explicité, reste subreptice, voire inconscient, contrairement aux hypothèses formelles de la théorie des tests, par exemple. Nous hasardons ici hardiment une formalisation du processus d'interprétation en statistique textuelle. En l'état, notre théorie est à la fois prétentieuse et insuffisante ; espérons qu'elle aura au moins l'intérêt de susciter des vocations et d'attirer l'attention sur le problème.

Nous supposons ici connu le fonctionnement d'ALCESTE (Reinert, 1990, 1993 a et b), logiciel qui a servi à analyser notre corpus. Pour dire vite, ALCESTE découpe le corpus analysé en « uce » (unités de contexte élémentaires : successions de segments ponctués d'une longueur -de quelques mots à quelques dizaines de mots- arbitrairement fixée par l'analyste) puis regroupe ces uce, sur la base des cooccurrences lexicales, en classes d'uce. La question que nous posons ici est celle de *l'interprétation* des classes, c'est-à-dire du processus par lequel l'analyste attribue un « sens » aux classes obtenues à partir de l'analyse des traits lexicaux typiques des unités regroupées dans chaque classes. Le problème est présenté à partir d'une analyse avec ALCESTE, mais a une portée générale, *mutatis mutandis*.

2. Un exemple : l'analyse du terme information

Interrogeant le dictionnaire Robert par « association libre » en utilisant un protocole d'analyse conceptuelle qui a fait ses preuves (Lahlou, 1995b), nous avons constitué un corpus de la façon suivante. Les définitions, en texte intégral, du terme *information* ainsi que de ses termes « associés » (synonymes, analogues et dérivés, tels que fournis par le Robert) et des associés de ces associés ont été extraites manuellement du Grand Robert (édition 1993). Nous

obtenons ainsi un corpus de 235 définitions (115 000 occurrences). Ce corpus est découpé en 7256 uce, qui vont être regroupées par ALCESTE.

L'analyse fournit 7 classes qui correspondent à des (présomptions) de noyaux sémantiques constitutifs du concept d'*information*. J'ai interprété ces classes, et leur ai donné les noms suivants : CONNAISSANCE, CODAGE, MEDIA, ACQUERIR, AVISER, INSTITUER, APPRENDRE (voir Lahlou, 1995a, pour l'analyse détaillée).

Trois classes fournissent le cœur technique du concept, et correspondent assez bien aux aspects pris en compte par la théorie de l'information de Shannon (1948). Il s'agit des classes CONNAISSANCE, CODAGE et MEDIA. Ces trois classes décrivent la forme de l'information : celle-ci est un contenu idéal, une représentation, la CONNAISSANCE (pensée, sentiment, etc.), présenté sur un support (courrier, publication, message...), le MEDIA, en utilisant un certain système de CODAGE (code, informatique, signal...). Pour le sujet qui nous intéresse ici, examinons simplement la classe 3, que j'ai interprétée comme renvoyant au noyau sémantique « MEDIA ».

Voici une liste de ses traits typiques, c'est-à-dire ceux qui sont statistiquement plus fréquents dans les uce de cette classe que dans l'ensemble du corpus (suivant un simple test du chi²), ici classées par chi² décroissant (les + et < renvoient à des formes de lemmatisation, ex *journa+l* correspond à *journal* et *journaux*) :

journa+l, lettre+, courrier+, renseignement+, presse+, reclame+, annonce+, feuille+, depeche+, ecrit+, affiche+, prospectus, adress+er, imprime+, envo+yer, pli+, publicite+, expres+, missive, pli+er, bulletin+, livre+, papier+, correspondance+, depli+er, page+, spectacle+, note+, cahier+, service+, televise+, reportage+, communiqu+er, billet+, carnet+, reception+, vente+, periodique+, publication+, agent+, publicitaire, revue+, message<, agence+, radio+, edition+, expedition+, faire-part, actualite+, distribution+, magazine, nouvel+, titre+, destinataire+, voiture+, chronique+, recueil+, commerce+, recommand+er, achet+er, article+, somme+, depos+er, imprim+er, piece+, texte+, communicat<, poste+, diplomati+que, fiche+, traite+, mise+, television+, four+nir, pa+yer, reponse+, carte+, preface+, marque+, mur+, fete+, illustr+er, annonc+er, copie+, voyage+, signataire+, avis, dossier+, impression+, maison+, gros+, film+, remettre., bre+f, exemplaire+, large+, accuse+, campagne+, colonne+, marchandise+, tissu+, etranger+, emission+, commercia+l, verba+l, explication+, lire+, roman+, publi+er, demand+er, ambassad+eur, reference+,

La notion de support/véhicule apparaît sous de nombreux avatars classiques du média : les supports papier (*journa+l, lettre+, courrier+...*), oraux (*renseignement, annonce, avis...*), mais aussi la *radio*, la *télévision*, le *film*, et même des supports auxquels on penserait moins en première intention : le *mur*, l'*ambassadeur*, l'*agence*, le *spectacle*... Au vu de ces traits typiques, on nous concédera que notre interprétation paraît acceptable, sinon triviale. L'exemple ainsi planté, passons à la théorie.

3. *Le sens n'est pas dans le texte*

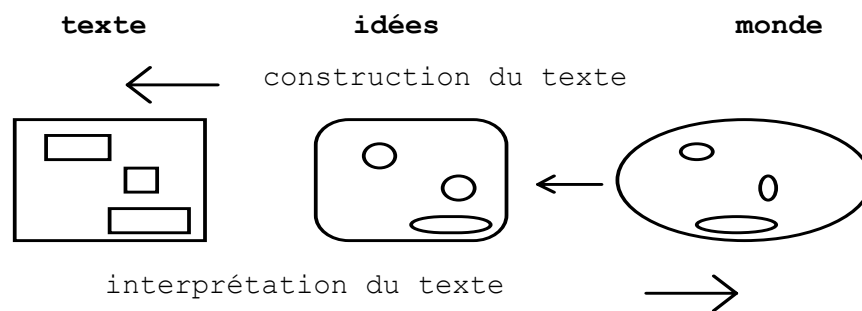
Le sens ne saurait se trouver dans la langue elle-même ; il est dans le rapport au monde des sujets : pas de sémantique sans pragmatique. C'est là une évidence : il faut connaître le monde pour interpréter des textes, et l'opération d'interprétation ne peut se reporter, en dernière analyse, qu'aux objets du monde auxquels se réfèrent les mots.

L'interprétation est possible car les associations des objets du Monde vécu se retrouvent dans le monde des idées, et celui-ci s'exprime par la langue. Dès 1660, exprime clairement la chaîne qui, fixant dans la mémoire les associations sensorielles, *propage* dans le monde idéal une image du monde vécu, faisant que la structure du monde des idées est analogue à celle du Monde des choses. Parce que l'âme enregistre, sous forme d'associations d'idées, les

associations entre choses rencontrées par le corps-sentant dans le Monde des choses, le réseau des idées sera analogue au réseau des choses perçues.

« L'ordre et l'enchaînement des idées est le même que l'ordre et l'enchaînement des choses » (Ethique, livre II, théorème VII)

La langue, à son tour, reproduit par des connotations lexicales les associations des idées, ce qui fait que le texte est, indirectement, une image des associations entre les choses du Monde vécu : en cela, le texte représente un tableau du Monde. En étudiant le texte, il est donc en principe possible, pourvu que l'analyste ait au préalable une idée des rapports entre les référents, de reconstruire le tableau du Monde qu'essaye de représenter le texte. Nous considérerons que ce tableau est représenté, dans le monde des idées, par une combinaison d'objets sémantiques plus simples, chacun correspondant à un noyau de sens représenté dans le texte par une classe d'énoncés. Ainsi, les traits peuvent renvoyer à des objets sémantiques comme les "taxèmes" ou les "champs" (Rastier, 1994, pp. 125-128). Il nous faut cependant abandonner l'espoir de déterminer des éléments "atomiques" de sens, en raison même de la nature holographique du sens : tout élément, vu de près, se scindera à son tour en éléments, renvoyant au reste de la langue (Saussure, 1915 ; Wittgenstein, 1965 p. 31).



Cette présentation simpliste, qui renvoie à la triade sémiotique (*vox*, *conceptus* et *res*, pour reprendre la terminologie de Rastier [1990]), pose le problème. L'analyste, dans sa pensée (*idées*) va essayer de faire coïncider ce qu'il observe dans le *texte* avec un tableau du *Monde* possible. L'opération d'interprétation est achevée quand l'analyste a pu attribuer un tableau du Monde au texte. L'opération est assez analogue à une traduction, avec tous les problèmes afférents (Mounin, 1963) : l'analyste donne une version idéelle du texte lexical. Cette version est faite sous la contrainte de vraisemblance, c'est-à-dire que les associations d'idées doivent être cohérente avec la vision du Monde de l'analyste. Celui-ci cherche donc à identifier dans le texte des « noyaux de sens » qui correspondent à des objets du Monde.

Pour cela, l'analyste va considérer le texte comme une succession de vues partielles du Monde, et supposer que l'ensemble de ces vues renvoie à un ensemble plus petit d'idées. Donc, que plusieurs vues partielles renvoient à une même idée. Un peu comme si, à partir de quelques centaines de clichés pris par un touriste en voyage à Paris, il essayait de retrouver les monuments de Paris, en classant les photos qui se ressemblent. Cette analogie est évidente quand le corpus est constitué de réponses d'un échantillon d'individus à une question ouverte.

Pour déterminer les noyaux de sens, dont la combinaison forme la représentation formée par le texte, on repère par des méthodes statistiques ce qu'il y a de commun dans différentes "vues" de la chose, énoncées au cours du texte.

L'espoir de faire émerger les éléments de base du sens (de la représentation de l'objet) par une telle méthode vient de trois hypothèses. D'une part, qu'il existe effectivement *quelque chose de commun* dans les différentes vues de chaque objet, quelque chose de stable qui viendrait du Monde et qui fonde l'existence de l'objet en tant qu'il est relativement

indépendant des vues qui peuvent en être obtenues. D'autre part que c'est par comparaison entre ces vues que l'on fera émerger, au niveau du texte, le même par contraste avec le différent ; c'est-à-dire, en termes distributionnels, ce qui se répète dans un contexte variable : les connotations, repérées par cooccurrence. Enfin, que les connotations systématiques entre termes permettent de repérer des « connexions » (Whorf, 1927) sémantiques, noyaux de sens culturellement partagés, qui correspondent dans la langue à des objet du Monde, construits par le processus évoqué par Spinoza. On fait donc, en somme, l'hypothèse que le texte est interprétable, c'est-à-dire qu'il parle, de manière répétée, d'une petite quantité d'objets du Monde, et de ceux-là principalement ; et que l'on pourra repérer ces objets à travers l'étude des cooccurrences de traits qui y renvoient.

4. *L'interprétation est un processus abductif*

Pratiquement, on a récupéré dans notre exemple, sous forme de texte, un grand nombre de vues locales énoncées à propos de l'objet « information ». Ces textes sont rassemblés dans un unique corpus, qui est ensuite scindé en énoncés, qui sont considérés comme des associations de traits (mots, locutions, ou leurs racines lexicales).

Les énoncés sont classés, par analogie et contraste, sur la base de leur contenu lexical. On obtient des classes qui contiennent des énoncés analogues (au sens où ils ont des traits en commun). Les énoncés classés sont similaires entre eux dans une même classe, et aussi différents que possible des énoncés des autres classes. Chaque classe possède des traits typiques lexicaux, ceux qui reviennent fréquemment dans les énoncés de cette classe. Si ces traits renvoient tous à un même noyau de sens, celui-ci peut être considéré comme un des noyaux du sens du mot *information* : cf. notre exemple de la classe « MEDIA ».

On a alors trouvé, par abduction, les éléments que nous cherchions, à travers l'interprétation des classes. Pourquoi par abduction ? Examinons cette phase obscure et rarement décrite qu'est l'opération d'interprétation, où l'opérateur introduit subrepticement sa connaissance du monde.

La méthode de classification descendante utilisée ici construit d'un même mouvement les traits caractéristiques et les classes, car la probabilité que le trait "a" (en l'occurrence, une forme graphique, le lexème) puisse être un trait pertinent pour caractériser un énoncé provient précisément de ce que "a" est observé sur un certain nombre d'énoncés, mais pas sur tous. En d'autres termes, émergent comme traits classificatoires les traits *discriminants*, ceux qui sont effectivement susceptibles d'être productifs de classifications intéressantes. Ceux-ci permettent de caractériser un nombre suffisant d'expressions à classer, ils sont donc efficaces. Les traits trop répandus (non discriminants) ou les traits trop rares (hapax, traits aberrants) ne sont pas intéressants. C'est là une résolution empirique du problème du choix des traits pertinents. Les traits non discriminants sont rendus inopérants par leur répartition homogène dans les énoncés, ils constituent le "fond". Certains traits qui pourraient être discriminants mais sont présents en trop petit nombre n'ont pas une fréquence assez élevée pour influencer l'analyse : ils sont également rendus inopérants par la classification descendante. C'est ce qui a fait que, dans notre exemple, un certain nombre d'énoncés se sont retrouvés classés ensemble (dans la classe 3), et cela par les traits typiques qui ont été énumérés plus haut.

Mais nous n'avons pour le moment que procédé à un classement, pas encore à une interprétation.

L'opération d'interprétation d'une classe C réside dans la décision de considérer ses traits typiques comme renvoyant tous une "idée" unique A, à laquelle et par laquelle cette classe est identifiée. Autrement dit, d'abord l'analyste remarque que la classe C est caractérisée par un ensemble (a_i) de traits typiques. Il *suppose* que la classe C correspond bien à une idée unique, c'est-à-dire que ces (a_i) sont des avatars d'un unique paradigme A, à déterminer, qui serait un

noyau de sens, l'idée qui sous-tend la classe C. En examinant les traits (a_j) il cherche à inférer la nature de ce paradigme générateur. Par exemple, "journal" lui semble renvoyer à MEDIA. Il s'agit là d'une induction.

L'analyste pose alors que la classe C réfère bien au paradigme "A" (ex : MEDIA). Ce faisant, l'analyste décide que "A" est effectivement un noyau de sens acceptable comme signifié d'une classe paradigmatique. L'analyste vérifie alors que tous les traits (a_j) sont effectivement acceptables comme avatars lexicaux de A, aux artefacts de méthode près. Si c'est le cas, L'analyste reconnaît que la classe C signifie bien "A", et il l'interprète comme telle : C(A). Ceci se fait progressivement, en essayant éventuellement plusieurs "A" candidats.

Par exemple, l'induction que "journal" et "lettre" renverraient à une classe « SUPPORT PAPIER » (signifié potentiel pour l'interprétation de la classe) se voit invalidée par la présence de "télévisé" et "radio". « SUPPORT PAPIER » est trop étroit. Inversement, il semblerait peu intéressant d'interpréter la classe en question par le paradigme ARTEFACTS HUMAINS, qui est trop large. L'analyste retient finalement un signifié tel que son paradigme soit quelque chose comme le plus petit commun paradigme (PPCP) rendant compte des traits observés.

Il s'agit donc d'une opération d'abduction, comme nous allons le voir en passant l'opération "au ralenti".

Lorsque l'analyste compare plusieurs expressions, il constate que :

(1) - de nombreuses expressions (X_i) de la classe C possèdent le trait typique " a_j "

Il en induit :

(2) - il existe *probablement* un "bonne" classe C(A), dont " a_j " est un trait caractéristique

soit : il est probable que (Il existe C(A) tel que, si X_i contient a_j , alors X_i appartient à C(A))

L'analyste examine alors la plausibilité de C(A) en tant que classe, c'est-à-dire qu'il cherche dans sa connaissance du monde le A en question. S'il trouve un tel A acceptable et pertinent (PPCP), alors :

Il abduit de (1) et (2) :

(3) - alors X_i fait partie de la classe C(A).

Dans cette opération d'abduction, ce qui caractérise une "bonne" classe, c'est qu'elle renvoie à "quelque chose" identifié par l'analyste, dans notre cas un signifié A, le PPCP, dont les (a_i) seraient des expressions, ou signes (est-ce que tous les a_j renvoient au même A?). Ensuite, on vérifie que tous les X_i de C contiennent bien des traits " a_j " qui renvoient au même A. En d'autres termes : les X_i sont interprétables en A, et tous les X_i de C renvoient à la même interprétation A.

5. Les analystes ont des pratiques discutables

Quelques remarques. D'abord, dans la pratique, le PPCP n'est pas aisé à trouver. Par exemple on ne sait pas très bien ce qui doit être considéré comme artefact, ou comme élément contradictoire de l'hypothèse interprétative. Ainsi, nous avons délibérément choisi de « négliger » d'interpréter les traits *maison*, *gros*, *étranger*, sous prétexte qu'ils étaient « peu caractéristiques ». Tandis que nous utilisons sans vergogne les traits *émission* ou *ambassadeur*, certes moins significatifs statistiquement, mais qui rentrent bien dans notre PPCP.

Ensuite, le logiciel, sourd au sens, constitue des classes sur la base de tests statistiques (partition optimale). Ces classes peuvent être "bonnes" (interprétables : on reconnaît "A") ou "pas bonnes" (non interprétables : on ne reconnaît rien). L'analyste "vérifie" donc que ces

classes correspondent effectivement à "quelque chose" d'interprétable, c'est-à-dire un objet du monde ou un concept connu. Or cette opération de vérification est subreptice. Et la manière dont sont présentés les "résultats de l'analyse" la passe sous silence, ou plutôt, fait croire qu'elle est postérieure à la constitution des classes.

Car, une fois l'analyse terminée, l'analyste présente toujours les résultats de telle manière que, effectivement, les (a_j) sont "naturellement" interprétables comme "signes" de "A". Donc, il ne présente que des "bonnes" classes dans le contexte culturel de son public. Ceci occulte la façon dont le travail réel s'est fait. Puisque, en fait, souvent l'analyste a obtenu une, ou plusieurs classifications préalables, dans lesquelles certaines classes "ne faisaient pas sens" : certains des (a_j) n'étaient pas admissibles comme représentants de "A". C'est-à-dire, il n'a pas trouvé de bon "A" candidat à être le signifié dont tous les " a_j " caractérisant C seraient des représentants lexicaux. Il a alors, utilisant les facilités techniques de son dispositif de classification (ici, les paramètres du logiciel) réalisé une autre classification, modifiant par essais successifs la partition pour obtenir le résultat final, qui, lui, est acceptable. Plus encore, il peut, une fois les $C(A)$ trouvées à son goût, modifier marginalement les partitions pour obtenir une meilleure ventilation des X_i , ajustant ainsi les résultats à l'interprétation pour avoir les classes les plus « pures » possibles (et cela, en utilisant, bien sûr, des méthodes licites statistiquement, par exemple d'échanges de X_i sur la base des distances aux barycentres de classes etc.).

Il nous arrive souvent, personnellement, de réaliser plusieurs essais pour obtenir une "bonne" analyse. Un analyste prudent se doit d'ailleurs de réaliser systématiquement plusieurs essais en variant les paramètres, pour vérifier que les résultats obtenus sont solides (stables, c'est-à-dire rebelles à ses manipulations). C'est au cours de ces analyses successives que les contours de l'objet se précisent pour lui. Et, naturellement, il choisit de présenter "le meilleur", c'est-à-dire celui qui correspond le mieux à sa conception du Monde.

La boîte noire qu'est l'analyste utilise donc sa connaissance du Monde pour reconnaître l'objet "A" à travers ses traits manifestes. L'interprétation suppose donc, implicitement, que l'analyste connaisse déjà plus ou moins les objets qu'il cherche, et soit capable de les identifier à partir des combinaisons que son matériau lui présente comme pertinentes.

La seule différence entre une analyse appuyée sur du matériel empirique et la réflexion en chambre consiste donc en ce que la technique (ici, le logiciel ALCESTE) sélectionne certaines combinaisons particulières de traits candidates à être les "bons" noyaux de sens ("signifiés"), suivant un certain principe légitime (ici : la redondance dans les différents énoncés, mesurée statistiquement). Ce point crucial est rarement explicité par les praticiens. Or c'est la connaissance préalable de l'objet, et celle des limites de la technique utilisée, qui fait la différence entre les bonnes et les mauvaises analyses.

6. *Une théorie de la démarche d'interprétation qui a des prétentions générales*

Résumons les opérations, dans ce que nous appelons pompeusement une théorie de la démarche d'interprétation :

- 1) Rassembler un corpus d'énoncés "à propos de" l'objet qui nous intéresse (corpus pertinent).
- 2) Classer ensemble les énoncés qui ont des traits linguistiques analogues
- 3) Interpréter les classes en termes sémantiques. Soit :
 - a) Repérer les combinaisons de traits lexicaux typiques de chaque classe ;
 - b) En induire, dans l'espace sémantique, des noyaux de sens, constitués de combinaisons de traits sémantiques correspondants ;

c) Formuler des lois d'expression de cette combinaison sémantique dans la langue. (Pratiquement, on considérera toutes les expressions linguistiquement "bien formées" du noyau sémantique sont des productions acceptables.)

4) Vérifier que les expressions effectivement classées dans chaque classe sont interprétables comme expression de ces noyaux sémantiques sous-jacents, (règles de langue, relations sémantiques sur le monde issues non seulement des associations attestées du corpus, mais de la connaissance du sens commun sur le monde, par exemple celle de « catégories » d'objets).

Nous pensons donc que cette « théorie » correspond à une démarche heuristique plus générale, appliquée dans les sciences expérimentales, et que l'on pourrait décrire de la façon suivante :

1. On obtient, par une série d'expériences (observations...), des vues subjectives et partielles de l'objet étudié, exprimées sous la forme d'arrangements de signes conventionnels (descriptions en langue, mesures...).

2. On compare ces expressions pour en tirer des régularités, c'est-à-dire des arrangements analogues (c'est-à-dire : ayant des traits communs). Lors de cette opération, on ne retient que des régularités considérées comme "pertinentes", et cela par rapport à une connaissance du monde dont on dispose préalablement

3. On construit des règles (ou lois) de production de ces expressions (règles d'assemblages des signes conventionnels).

4. On vérifie que les productions de ces règles correspondent effectivement à des observables, par des expériences de validation.

Ceci n'apparaîtra peut-être pas encore assez clair au lecteur. Pour imaginer, disons que le processus d'interprétation consiste à essayer de mettre en correspondance deux univers. Le premier est celui des observables (les résultats statistiques obtenus sur le texte, qui fournissent des classes et une liste de traits caractéristiques). Le second est celui des modèles (dans le monde des idées). L'opération d'interprétation consiste en un va-et-vient abductif entre ces deux univers, par lequel l'analyste, jouant littéralement sur les deux tableaux, va essayer d'obtenir deux représentations en correspondance : d'une part les « résultats » de l'analyse, d'autre part « l'interprétation ». Nous avons vu que l'analyste peut pour cela procéder par ajustements successifs, en modifiant un peu les contours des différentes catégories obtenues dans les résultats (variation des paramètres) et dans l'interprétation (en choisissant des paradigmes explicatifs) jusqu'à obtenir un état final présentable. Dans ce dernier, la correspondance entre résultats et interprétation est considérée comme bonne si elle correspond bien au sens commun, c'est-à-dire si elle semble compatible avec le Monde (du moins dans sa partie censée être représentée par le texte), tel que le connaissent ceux qui lisent l'analyse. C'est ce qui fait que notre interprétation « MEDIA » pour la classe 3 peut sembler acceptable.

Pour finir, quelques évidences : la constitution du corpus, et la présomption qu'il parle d'un certain objet, est cruciale. Elle conditionne la validité de la démarche de l'analyse, et oriente l'interprétation. Comment aurions-nous interprété la classe 3 si elle était censée se rapporter par exemple à l'analyse du terme « administration » ? ou du terme « colloque » ? Faites l'essai de relire les traits typiques dans cette perspective : c'est instructif, mais angoissant quand on sait que, de fait, nos corpus ne sont jamais *rigoureusement* centrés sur l'objet que nous voulons étudier.

Ensuite, ceci met l'accent sur l'importance des connaissances que l'analyste a de l'objet étudié. Notre interprétation actuelle des résultats de certaines de nos analyses anciennes, après une meilleure connaissance du sujet, a un peu varié. Il y a un contre-transfert du chercheur sur

son matériau, ici comme dans les méthodes qualitatives (Devereux, 1967). Ceci devrait nous inciter à la prudence.

7. Conclusion : à suivre

Notre approche paraîtra sans doute naïve, et lacunaire, et en l'état elle l'est, et pas seulement parce que le court espace d'exposition accordé, limitant le recours aux références, accentue ces caractères déplorable. Notre "théorie" en l'état, est largement perfectible, mais nous pensons qu'il y a là quelque chose à creuser. Il y a, en tous cas, à l'heure actuelle un manque cruel dans ce domaine, comme le signalait Ludovic Lebart dans sa conclusion orale au 2èmes JADT. Les analystes, dans l'intérêt de la discipline, ont intérêt à réfléchir aux fondements de leurs pratiques, afin d'éviter qu'une fois encore les méthodes issues des travaux séminaux de Benzecri ne se dévoient dans une ivresse technique et logicielle, et donnent lieu à des interprétations hâtives ou incorrectes qui feraient noyer le bébé avec l'eau du bain.

REFERENCES

- DEVEREUX, G. (1967). *De l'angoisse à la méthode dans les sciences du comportement*. Paris : Flammarion, 1980.
- LAHLOU, S. (1995a). L'utilisation de l'information dans l'entreprise : quelques réflexions théoriques et une analyse lexicale. EDF-Direction des Etudes et Recherches. HN-5194055.
- LAHLOU, S. (1995b). *Penser Manger. Les représentations sociales de l'alimentation*. Thèse, Ecole des Hautes Etudes en Sciences Sociales.
- LE ROBERT ÉLECTRONIQUE (1993), outil d'aide à la rédaction sur la base du Grand Robert de la Langue Française. Paris : Dictionnaire le Robert.
- MOUNIN, G. (1963). *Les problèmes théoriques de la traduction*. Paris : Gallimard, 1963.
- RASTIER, F. (1990). La triade sémiotique, le trivium et la sémantique linguistique. *Nouveaux actes sémiotiques*, 9, PULIM (Limoges) 1990.
- RASTIER, F. (1994). Microsémantique, lexique et contexte. Pp. 109-147. IN Eveline Martin (éd.) *Traitements informatisés de corpus textuels*. Collection "Etudes de sémantique lexicale". INaLF. Paris : Didier Erudition. 1994. ISBN 2-86460-224-5.
- REINERT, M. (1990). ALCESTE, une méthode d'analyse des données textuelles. Application au texte "Aurélia" de Gérard de Nerval. *Bulletin de Méthodologie Sociologique*, 26. pp. 25-54.
- REINERT, M. (1993a). Quelques problèmes méthodologiques posés par l'analyse de tableaux "énoncés x vocabulaire". Actes des secondes journées internationales d'analyse statistique des données textuelles. Montpellier, 21-22 octobre 1993. Paris, ENST, 1994. pp. 539-549.
- REINERT, M. (1993b). Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars. *Langage et Société*, 66, Décembre 1993. pp. 5-39.
- SAUSSURE, F. de. (éd. posth., 1915). *Cours de linguistique générale*. Paris : Payot, 1972.
- SHANNON, C. E. , WEAVER, W. (1948). *The Mathematical Theory of Communication*. Illini Books edition, University of Illinois Press, 1963.
- SPINOZA, Baruch (1677). *Ethique*. Trad. Raoul Lantzenberg. Paris : Flammarion, 1947.
- WHORF, B. L. (1927). On the Connection of Ideas. In : John B. Carroll (ed.), *Language, Thought and Reality*. Cambridge (Massachusetts) : The MIT Press, 1956. pp. 35-39..
- WITTGENSTEIN, Ludwig (1958). *Le cahier bleu*. Paris : Gallimard, 1965.