

Les facteurs de choix et d'utilisation de logiciels d'Analyse de Données Textuelles¹

Mathieu BRUGIDOU*, Caroline ESCOFFIER*, Helka FOLCH*,
Saadi LAHLOU*, Dominique LE ROUX*, Patricia MORIN-ANDREANI*,
Gérald PIAT*

**EDF- Pôle Industrie - Division Recherche et Développement, 1 avenue du Général de Gaulle, 92141 CLAMART Cédex, Fax : 01.47.65.57.00, Contact : dominique.Le-roux@edf.fr*

Summary

Several programs are at present available on the French market for assisting the qualitative tasks of the analyst. The purpose of this paper is not to compare their different theoretical characteristics and facilities nor to identify their strengths and limitations. This has been already been recently well done [JEN97]. Here, we intend to present the point of view of the user making his choice between all these programs according to his own ability, the characteristics of the study he is undertaking, the research hypothesis, the type of corpus, and so on. First, we propose a schedule representing the user's choice factors from our own experience of computerized textual analysis. Secondly, we will illustrate this schedule by rapidly presenting studies carried out on different types of textual material with the help of various programs such as ALCESTE, TROPES, DISCURSUS, ATLAS-Ti... Finally, we will present the developments which are planned for a better management of methods and programs.

Résumé

Plusieurs logiciels d'analyse textuelle sont actuellement disponibles sur le marché français. Le but de cet article n'est pas de comparer leurs différentes caractéristiques et équipements théoriques ni d'identifier leurs forces et leurs limites. Nous présenterons plutôt le point de vue de l'utilisateur essayant de choisir entre les différents logiciels, en fonction de ses connaissances, des caractéristiques de l'étude, de ses hypothèses de recherches, etc. Après avoir caractérisé les différents types d'études menées au GRETS, nous essaierons de mettre en évidence les différentes stratégies en matière de choix instrumental et nous présenterons rapidement les perspectives.

Mots-clés : *Analyse automatisée des données textuelles, Analyse du discours, Atelier logiciel*

1. Introduction

Quels sont les facteurs de choix de logiciels d'analyse textuelle ? L'analyste se contente-t-il de prendre l'outil qu'il a sous la main, celui dont il a une pratique à peu près correcte ? Nous avons fait le pari de dépasser cette hypothèse minimaliste, et pourtant parfois vérifiée, pour poser la question suivante : existe-t-il, au-delà d'un empirisme apparent, des « stratégies » implicites mises en oeuvre par les utilisateurs pour faire coïncider, à propos d'une étude, les quatre dimensions complémentaires constitutives d'une étude que sont : la représentation de la langue, la représentation du discours, la dimension statistique, la dimension instrumentale et expérimentale ? Sans vouloir donner, bien entendu, des réponses définitives, on se propose ici d'exposer le témoignage d'une équipe pluridisciplinaire de recherche en sciences sociales dans une entreprise. Le contexte de l'analyse textuelle en entreprise est tel que de nombreuses contraintes pèsent sur sa pratique comme les délais qui peuvent rendre difficiles le recul

¹ Les auteurs tiennent à associer à ce travail Isabelle Blot, Jean Vidal et Gro Waeraas de St Martin qui ont activement participé à la diffusion de l'analyse textuelle au GRETS.

théorique et le test de différents outils et, en ce qui nous concerne, la nécessité de s'adapter à des problématiques très variées. Conscients des limites de notre pratique, nous ne prétendons donc pas ni rendre compte d'une utilisation exhaustive des outils existants sur le marché, ni recenser les différents points de vue théoriques ayant présidé à l'élaboration des logiciels d'assistance à l'analyse de contenu et de discours disponibles pour le chercheur². En revanche, il nous a paru utile parce que peu fréquent, d'effectuer une présentation du point de vue d'un collectif de praticiens travaillant dans un domaine de connaissance hétérogène et dont les expériences diversifiées permettent d'appréhender la complexité des paramètres qui interviennent dans le choix d'un logiciel.

2. Contexte d'une pratique de l'analyse textuelle

2.1. Une cellule de sciences humaines au sein de l'entreprise

Les auteurs de ce papier appartiennent tous à une cellule sciences humaines de la Division Recherches et Développement (le GRETS), qui traite principalement à la demande de clients internes, des sujets aussi variés que la veille sociale au sein de l'entreprise, les évolutions organisationnelles et l'évolution des métiers, l'impact d'EDF sur l'environnement, les produits et services liés à l'électricité... Cette pluralité des thèmes traités induit une pluralité des approches disciplinaires : sociologues, psychosociologues, politologues, ethno-anthropologues, linguistes, sémiologues, statisticiens... forment l'équipe de travail.

L'introduction d'outils logiciels s'est effectuée, d'abord dans le domaine des études quantitatives, puis dans le champ des études qualitatives en commençant par le traitement des réponses à des questions ouvertes. Nous ne bénéficions pas à proprement parler d'un « atelier logiciel » intégré et disponible à tous les chercheurs et analystes du GRETS. La réalité de notre travail est beaucoup plus artisanale et nous oblige à mettre « bout à bout » des outils dans un chaîne de traitement dont certains maillons ne sont pas reliés. Dans une logique essentiellement opératoire, le GRETS a procédé, dans un passé plus ou moins récent, à des tests évaluatifs de logiciels de nature différente parmi lesquels : SPAD-T, Moda Lisa, Termino, Prospero. Actuellement, les principaux outils utilisés dans le cadre de nos études sont, pour des usages diversifiés, Alceste, Tropes, Atlas-ti dans la famille des CAQDAS³ et deux outils *ad hoc*, développés dans le cadre d'un partenariat avec IBM, Discursus et ATAL).

2.2 Caractérisation des études effectuées (cf. tableau en annexe)

La grille qui suit est le fruit d'un effort de mise à plat des différents critères intervenant dans la caractérisation d'une étude et qui vont orienter un choix instrumental :

- **Définition de la problématique** : Type d'étude : prestation pour un « client interne » ou travail de prospective ou de recherche appliquée ; Objectifs de l'analyse : définition des objectifs de l'étude, des présupposés, des hypothèses, des moyens à mettre en oeuvre, de la méthodologie de travail .
- **Profil de l'analyste** : un / multiple (en fonction des modes de partage du travail : définition du questionnaire, recueil des données, analyse des données...) ; discipline d'origine et référentiel théorique ; limites de l'analyste : facteurs de contraintes quant au choix et à la pratique d'un outil ; manque d'information sur les outils existants, enfermement dans un type d'approche unique.
- **Mode de constitution d'un corpus** : Type d'analyse : analyse « primaire » à partir de matériaux *ad hoc*; analyse « secondaire » à partir de matériaux produits indépendamment du cadre de l'étude ; Type de corpus :

² Une excellente présentation a déjà été effectuée, de manière récente [JEN97], des points de vue théoriques ayant présidé à l'élaboration des logiciels d'assistance à l'analyse de contenu et de discours disponibles pour le chercheur.

³ Computer Aided Qualitative Data Analysis Systems.

taille (nombre de textes), degré d'homogénéité du corpus ; Type de texte : entretiens, questions ouvertes, articles de presse, publicités, courrier...

- **Type de traitement en fonction des résultats attendus** : Classer des textes ou de fragments de textes ; Extraire des informations ; Effectuer une synthèse ; Effectuer l'inventaire des thèmes traités ; Enrichir un corpus de commentaires et de codes.

L'expérience montre que ces différents facteurs ne se situent pas au même niveau et sont en perpétuelle interaction. Nous avons remarqué, de manière empirique, l'importance de la formation initiale et de la « famille théorique » de l'analyste sur le choix d'un outil. Nous avons recensé de nombreuses pratiques de détournement par rapport aux objectifs initiaux de l'outil (définis en terme de spécifications à la fois techniques et théoriques). Plus classiquement, on notera des niveaux d'usage la plupart du temps en deçà de l'usage « idéal » prévu par le concepteur.

3. Existe-t-il des stratégies implicites en matière de choix instrumental ?

Au cours de notre effort, parfois laborieux, en raison de la complexité des critères en présence, de mises à plat de « stratégies » des analystes en matière de choix instrumental, nous avons retenu deux points importants :

- de quelle manière l'analyste essaie-t-il de faire coïncider la représentation de la langue, celle du discours, la dimension statistique, avec la dimension expérimentale liée au choix d'un instrument ?
- à quels « garde-fous » peut-on faire appel pour parer aux dérapages interprétatifs en puissance quel que soit le type d'approche retenu ?

3.1. Quel niveau d'analyse appliquer à quel type de texte ?

Notre équipe a eu à traiter des types de corpus variés dont il serait naïf de chercher à établir la typologie tant est grande la complexité des paramètres en présence, profondément imbriqués. Etablir une typologie de textes relève toujours, comme le note Habert [HAB98] d'un «bricolage». Il nous paraît néanmoins important de lister quelques-uns de ces paramètres dans la mesure où ils ont nécessairement une forte influence sur le type d'analyse qu'on peut leur appliquer :

- le canal de communication : écrit ou retranscription d'un discours oral plus ou moins respectueuse de l'émission initiale. A ce propos, on peut faire trois types de remarques :
 - ⇒ Les corpus disponibles sont pour des raisons pratiques le plus souvent des corpus écrits. L'analyse des discours des acteurs internes privilégie ainsi des prises de positions plus ou moins institutionnelles (presse interne, tracts, communiqués etc.). Les paroles d'agents, par exemple, sont le plus souvent issus de dispositifs de recueils *ad hoc* (entretiens lors d'enquête sociologique, réunions internes etc.).
 - ⇒ On sait par ailleurs que la communication orale est en générale moins structurée (notamment d'un point de vue thématique) qu'une intervention écrite. L'analyse thématique des entretiens montre une très forte redondance et une structure « en boucle » qui est caractéristique des récits oraux. La répétition, sous la forme d'une répétition littérale ou bien anaphorique, agit comme un moyen de pallier les insuffisances de la mémoire, et constitue un outil privilégié de la production de la cohérence interne et séquentielle du récit. Par ailleurs, la structure syntaxique de l'oral n'est pas la même que celle de l'écrit.
 - ⇒ Enfin, le mode de retranscription de l'oral peut impliquer la réécriture de certains mots ou certaines phrases (style prise de notes, erreurs manifestes de retranscription...). Une de nos premières préoccupations est de décider à partir de quel moment et jusqu'où s'impose le

nettoyage d'un corpus. La solution la plus rigoureuse qui consiste à ne rien modifier est très rarement satisfaisante.

- la sphère de communication : communication «privée» (entretiens, lettres de réclamation), interne à l'entreprise (discours syndicaux et directoriaux), sociale, culturelle et politique (coupures de presse). Par exemple, dans le cas de prises de paroles retranscrites dans le cadre de compte-rendu de réunions internes plusieurs facteurs font que le corpus ne peut en aucun cas être confondu avec des paroles d'agents pris individuellement. Il fait état dans le meilleur des cas d'une parole collective, organisée.
- la longueur du texte : du niveau du syntagme parfois, (réponses à des questions ouvertes) jusqu'au texte de plusieurs pages.
- le niveau de complexité de l'encodage linguistique en ce qui concerne l'organisation syntaxique, l'inventaire lexical, la présence d'éventuels procédés rhétoriques ;
- le niveau de complexité des thématiques abordées et de leur articulation. La plupart des outils ont notamment pour objet d'identifier les thématiques. Par exemple, l'analyse avec Alceste d'un corpus de presse ou d'entretiens fait apparaître des thèmes ; il est évidemment intéressant d'identifier les articulations, et les schèmes (notamment idéologiques) qui sont au principe de ces articulations (par exemple, dans les discours syndicaux, l'opposition entre les riches et les pauvres etc.). Or ni l'unité retenue (forme graphique, lemme etc.), ni la qualification de la relation (basée sur la distance : cooccurrence, opposition etc.) ne sont tout à fait adaptées à ce type d'analyse. Par ailleurs, certains logiciels de type « CAQDAS » (ou des logiciels comme Prospero) permettent de manipuler des unités non-linguistiques ou du moins des unités qui sont hétérogènes de ce point de vue : ce ne sont en effet ni des lemmes, ni phrases mais des paragraphes, des propositions ou des mots définis sur des critères sémantiques (bref des notions). Ces mêmes logiciels permettent de qualifier la relation identifiée autrement que d'un point de vue statistique (corrélation) ou même grammatical. Toutefois, on entre ici dans un processus d'interprétation qui est assez peu formalisé : les indicateurs linguistiques notamment sont peu sollicités.
- le mode de structuration du texte : dialogue orienté par un meneur de jeu (le sociologue, muni d'une grille d'entretien), débat de groupe, texte argumentatif... Le statut des acteurs est déterminant en ce qui regarde l'interprétation et la constitution du corpus : par exemple, pour l'étude d'un corpus de retranscriptions de réunions internes, les remarques du rapporteur sur l'ambiance de la réunion n'ont évidemment pas le même sens que les propos des participants qu'il est censé rapporter : même si le clivage n'est pas toujours si net.
- la fonction de communication dominante : fonction émotive dans les récits de vie des agents ou des utilisateurs interviewés, fonction référentielle dans les textes de presse interne ou externe, fonction conative dans les tracts et discours syndicaux...
- le statut des acteurs de la communication déterminant quant au choix des actes de parole : locuteur parlant au nom d'une structure ou s'exprimant en son nom propre, émetteur et récepteur clairement identifiés ou dissimulés derrière une structure anonymisante. On sait par exemple que les réponses sont conditionnées par la nature des commanditaires d'une enquête (interne / externe) et la relation de subordination entre commanditaire et enquêtés : absence supposée d'anonymat, questionnaire utilisé à des fins revendicatives ou modulations des réponses en fonction des conséquences directes envisageables sur la situation du répondant (salaires, charge de travail, etc.).

Face à ces matériaux, constitués en corpus en fonction de la définition d'un objectif de recherche (dans la plupart des cas élaborée avec le commanditaire de l'étude), il est du ressort de l'analyste de déterminer le niveau d'analyse auquel il prétend opérer et à quel type de questionnement il entend soumettre les textes :

- niveau lexico-sémantique : inventaire lexical et terminologique du corpus traité, élaboration d'un réseau de notions ;
- niveau de construction syntaxique et textuelle : étude des connecteurs et des marques de progression thématique, des marqueurs énonciatifs, notamment les déictiques comme référence aux interlocuteurs, au temps et à l'espace du discours, des marqueurs d'actes de discours (forme assertive, interrogative, impérative, reprise dans un dialogue..), processus à visée rhétorique de focalisation et de thématisation...
- niveau extra-linguistique : prise en compte du sémantisme propre à la structuration «physique» du texte, couplage avec des données sur le contexte de communication (notes du sociologue sur le déroulement d'un entretien), ou des données socio-démographiques sur l'émetteur du discours.

Les différents outils à notre disposition n'opèrent pas tous sur les mêmes types d'objets et n'offrent pas tous, quelle que soit la nature des outils utilisés (approche statistique, linguistique ou mixte) et dans l'état actuel de leur développement, les mêmes possibilités d'analyse, laissant à l'utilisateur le soin de dépasser les résultats fournis en approfondissant «à la main» certains niveaux d'analyse :

- certains outils, relevant d'une approche de type lexicométrique, proposent à l'analyste le calcul des fréquences d'occurrence des mots d'un corpus et l'élaboration de classes fondées sur ces dernières comme support de l'interprétation. Dans ce cas, ce sont les traitements de données qui guident l'interprétation. Certains logiciels appartenant à cette famille opèrent au niveau de la forme graphique, d'autres tentent d'approcher la notion de lemme au travers d'une simplification de la réduction morphologique.
- un autre type d'approche est celui offert par les outils proposant un inventaire terminologique du texte analysé ; dans ce cas, il y a une prise en compte de ce qui est approché dans l'analyse lexicométrique au travers des segments répétés mais seulement à partir d'une fréquence d'apparition suffisante. Des outils tels que ATAL (prototype IBM) ou Termino (actuellement Nomino) permettent, à partir de constructions syntagmatiques figées ou en voie de figement terminologique de décrire le paysage notionnel ou ses évolutions dans le temps d'un domaine d'étude.
- d'autres outils, comme Tropes, effectuent une analyse sémantique (de classification, de visualisation et de surlignage des composants sémantiques de diverses natures).

3.2 A quels garde-fous faire appel pour parer aux dérapages interprétatifs ?

Le risque de dérapages se pose en analyse textuelle comme dans d'autres types d'études mais les données textuelles se prêtent sans doute davantage aux « projections » interprétatives. Dans tous les cas de figure, l'analyse produite est inséparable des conditions de production des données, c'est-à-dire l'ensemble des choix opérés par l'analyste à chaque étape de l'étude (constitution du corpus, choix du logiciel et de ses paramètres). A ce titre, les points évoqués dans le paragraphe précédent nous semblent constituer autant de garde-fous à considérer tout au long du travail. De façon plus précise, les possibilités de contrôle qu'offrent les différents logiciels que nous utilisons sont diverses. Alceste laisse la possibilité de procéder à des

analyses multiples pour lesquelles ne varient que peu de paramètres : la robustesse des résultats est donc vérifiable. En revanche, si elles rendent ce logiciel particulièrement transparent, la multiplicité des résultats qu'il produit et l'accessibilité de ses dictionnaires ne constituent pas pour autant des garde-fous automatiques. Tropes, bien que plus opaque, offre par exemple la possibilité de contrôler et de modifier partiellement les univers de référence qu'il propose. Atlas-Ti, selon le principe des CAQDAS n'offre pas le type de mise à distance et d'objectivation proposé par exemple par Alceste : le codage, laissé aux soins de l'analyste et faiblement assisté par quelques outils simples de recherche de chaînes de caractère, est d'emblée forcément interprétatif, ce qui confère au processus d'analyse un caractère « hyperprojectif ». On peut craindre qu'une telle analyse « top-down » ne permette aucune résistance des matériaux et ne conduise qu'à retrouver la confirmation naturelle des hypothèses posées par le chercheur. Or, l'expérience montre que ces outils offrent bien, une possibilité de découverte et de surgissement (« *serendipity* ») de notions nouvelles capables d'orienter la construction progressive de la théorie. La possibilité de tenir un journal de bord de l'analyse ainsi que de partager le codage des données à plusieurs sont également autant de garde-fous à des interprétations trop projectives. Par ailleurs, s'il n'existe pas de règles standardisant la définition des catégories dans un outil tel qu'Atlas-ti, ce qui fait la puissance d'Alceste, en revanche, la liberté laissée à l'utilisateur et la souplesse de l'outil, lui permet d'augmenter la finesse de l'analyse.

4. Les perspectives

L'expérience du GRETS montre que deux moments importants de l'analyse des données textuelles ne peuvent être dissociés : d'une part, nous considérons que les textes constituent bien des « données » et qu'ils sont donc de ce fait justiciables de processus d'objectivation et de formalisation, d'autre part, que ces mêmes textes sont le fruit d'intention signifiante de la part des acteurs et l'objet de « parcours interprétatif » de la part de l'analyste. La double attention portée au texte comme « donnée » (i.e des unités textuelles fruit d'un processus de réduction et de formalisation qui peuvent être comptées) et comme « interprétation » répond à deux traditions de la sociologie mais aussi de la linguistique représentée au GRETS. Cette position rend compte des choix du GRETS en terme de logiciels et de capitalisation.

L'expérience du GRETS montre en effet la nécessité pratique de recourir à différents logiciels en fonction des problématiques très diversifiées (types de données, objectifs etc.) traités par le département et du « profil » des utilisateurs (et de leurs horizons théoriques en particulier) et surtout d'avoir recours à des logiciels de la famille de l'Analyse de Données textuelles (standardisation de la construction de la données, formalisation des « traitements » pour assurer notamment le caractère cumulatif des résultats) comme aux CAQDAS (prise en compte des processus interprétatifs dans la construction de la donnée).

La politique de capitalisation qui a été mise en œuvre s'appuie sur deux projets :

- Le projet *Verbatim* est spécifiquement consacré à la capitalisation des entretiens qualitatifs en vue de pratiquer des analyses secondaires. Ici, on s'intéresse particulièrement à la standardisation de l'archivage des entretiens [LER00]. Cette base de **données** nous permet de réunir le descriptif signalétique de l'étude, les entretiens effectués au cours des études qualitatives, les résultats obtenus à l'issue de leur dépouillement.
- le projet *Scriptorium*, dédié à des données textuelles hétérogènes, vise à construire des collections de corpus textuels en accroissement permanent. Ces documents sont décrits par une DTD⁴ permettant d'associer à chaque fragment de texte une « étiquette » contenant le

⁴ Description of textual data, dérivée de la text Encoding Initiative.

signalétique du texte. L'objectif de Scriptorium étant d'effectuer du profilage de corpus, il s'agit de procéder à des extractions homogènes en fonction de certains critères qui sont également modélisés par une DTD permettant de renseigner sur le contexte de l'extraction. Enfin, un troisième niveau de codage correspond aux résultats de l'analyse, permettant ainsi d'enrichir le corpus. C'est en effet parce que les points de vue (et donc les interprétations) sur le texte sont multiples mais toujours susceptibles d'une objectivation, qu'il est nécessaire d'assurer la traçabilité et la rétro-projection des résultats des analyses vers les documents d'origine.

Références

1. Références externes

- [BLO94] BLOT, I.; HAMMER, B.; LE ROUX, D. (1994) Traitement des questions d'opinion "ouvertes" : utilisation d'Alceste, outil d'assistance à l'analyse. *Revue ICO Québec* ; 6. (1 & 2)
- [BRU99] BRUGIDOU, M. ; LE QUEAU P. Les « rafales », une méthode pour identifier les différents épisodes d'un récit - *BMS - Bulletin de Méthodologie Sociologique*, n°64, octobre 1999, p 49 à 81.
- [FIE91] FIELDING, N.G. & LEE, R.M. (éds); (1991) *Using Computers in Qualitative Research* ; Londres : Sage Publ.
- [GUI98] GUIGLIONE, R. ; LANDRE, A. ; BROMBERG, M. ; MOLETTE, P. ; L'analyse automatique des contenus, 1998, Paris Dunod, 156 p.
- [HAB98]; HABERT, B., FABRE, C., ISSAC, F. (1998) De l'écrit au numérique : constituer, normaliser et exploiter les corpus électroniques; Paris Masson ; 320 p..
- [JEN97] JENNY, J. (1997) Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine : état des lieux et essai de classification ; *Bulletin de méthodologie sociologique (BMS)* ; 54; pp.64-112.
- [LEG95] LEGRIS-DESORTES, C. (1995) Semiotic analysis concerning the communication about the electric vehicle, in *12th international electric vehicle symposium* ; Los Angeles
- [LER00] LE ROUX, D. (2000) *Verbatim* : une expérience de capitalisation d'entretiens qualitatifs - *BMS - Bulletin de Méthodologie Sociologique*, n°65, à paraître.
- [LAH95] LAHLOU, S. (1995) Vers une théorie de l'interprétation en analyse statistique des données textuelles, in *JADT 1995* ; vol II pp. 221-227
- [REI95] REINERT, M. (1995) Quelques aspects de choix des unités d'analyse et de leur contrôle dans la méthode Alceste, in *JADT 1995* ; vol I p.27-34

2. Choix de notes internes EDF

- AMAR, M. - Assistance automatisée à l'exploration textuelle : présentation du progiciel Termino, système de dépouillement terminologique assisté par ordinateur - 1992 - Note interne EDF-DER HN-52/5/92/040
- BRUGIDOU M. - LABBE D. - Le discours syndical CGT - CFDT - FO en 1996 et 1998- 1999 Note interne EDF-DER HN-51/99/013
- BRUGIDOU, M.; ESCOFFIER, C. - L'avenir des métiers clientèle : réunions préparatoires à la vidéotransmission du 21 janvier 1999: analyse de discours - Note interne EDF-DER HN-51/99/015
- HAMMER, B.; LE ROUX, D. - Enquête interne "Vous et votre Entreprise" 1992 : analyse des réponses aux questions ouvertes - 1993 - Note interne EDF-DER HN-52/93/032
- LAHLOU, S. - La construction du sens dans l'analyse statistique de données textuelles : théorie et méthodologie illustrées par deux analyses - 1995 - Note interne EDF-DER HN-51/95/012
- LEMOINE, J-C. - VVE 1995 Analyse des réponses à la question ouverte Q45 : "Pour vous, qu'est-ce que la mission de service public ?" - 1997 - Note interne EDF-DER HN-51/97/012
- PIAT, G. - Spécifications et fonctionnement de l'application d'aide à l'analyse de discours DISCURSUS - 1996 - Note interne EDF-DER HN-52/96/025

Annexe : tableau illustratif des différents types d'études traités par le GRETS.

Etudes	Lettres de réclamation	Les éditoriaux de la presse syndicale	Questions ouvertes de l'enquête « Vous et Votre Entreprise »	Comptes-rendus de réunion	Climatisation et santé	Entretiens de fumeurs avec le CREDOC
commanditaire	interne	interne	interne	interne	interne	externe/ recherche interne
objectifs	thèmes récurrents	identification de thématiques et de rhétoriques syndicales	représentations caractérisation de sous-population dans l'entreprises	discours développés enjeux des débats caractérisation géographique	lien entre désaffection pour la climatisation et santé pré étude test des analyses secondaires	Méthodologie d'analyse des entretiens. Identification des différents épisodes d'un récit de vie et de leur articulation dans une intrigue
type d'analyse	primaire	primaire	primaire	secondaire	secondaire	primaire
type de corpus	de 10 à 10000 lettres	200 articles	7000 réponses	86 comptes-rendus	45 rapports et entretiens	Une quinzaine d'entretiens
type de texte	lettres	éditoriaux	réponses de 3 à 50 mots	oral et écrit	varié	entretiens
analyste	socio linguiste / sémologue	sociologue, politologue	statisticien	statisticien / politologue	linguiste	sociologue/politologue
limites de l'analyse	expression « affective » et éléments « objectifs » connaissance du domaine commercial	connaissances des enjeux sociaux et politiques, des stratégies des organisations syndicales	connaissance de l'entreprise	conditions de constitution du corpus inconnues connaissance des métiers	analyse secondaire qualité de la capitalisation connaissance du domaine	Difficulté de comparer différents entretiens.
classification de textes	oui	oui	oui	oui		non
inventaire thèmes	oui	oui	oui	oui	oui	oui
structures thématiques	oui					oui = relation entre les thèmes/ analyse synchronique (*)
croisement extralinguistique	oui	oui	oui	oui	oui	oui
logiciel	Discursus	Alceste, Tropes	Alceste	Alceste Tropes	Atlas Ti	Tropes

*Analyse synchronique : dans un texte homogène comme un entretien non-directif ou un discours l'identification d'une série de termes répétés en même temps de manière importante dans une courte séquence de texte (i.e des « rafales ») permet de mettre en évidence les « paquets de sens » qui constituent des épisodes. En suivant ces rafales de termes, on peut espérer repérer les ruptures thématiques (fin d'une série de rafales), les passages où les fils du récit se dénouent cependant qu'un nouveau thème (ou épisode ici) est développé (nouvelle série de rafales). L'algorithme des rafales a été intégré dans le logiciel d'analyse sémantique TROPES. [BRU99].