

Logiciels pour l'analyse qualitative : innovations techniques et sociales

Recherches qualitatives

Collection hors série
« **Les actes** »

Sous la direction
de Véronique Dumont, Christophe Lejeune
et François Guillemette

Actes du colloque organisé par L'Unité Économie
de la connaissance et management
de l'innovation (IMaGinE) et l'Institut
des sciences humaines et sociales
de l'Université de Liège (Belgique),
avec le soutien du Fonds national
de la recherche du Luxembourg et de
l'Association pour la recherche qualitative

21-22 octobre 2008
Centre de Recherche Public
Henri Tudor, Luxembourg

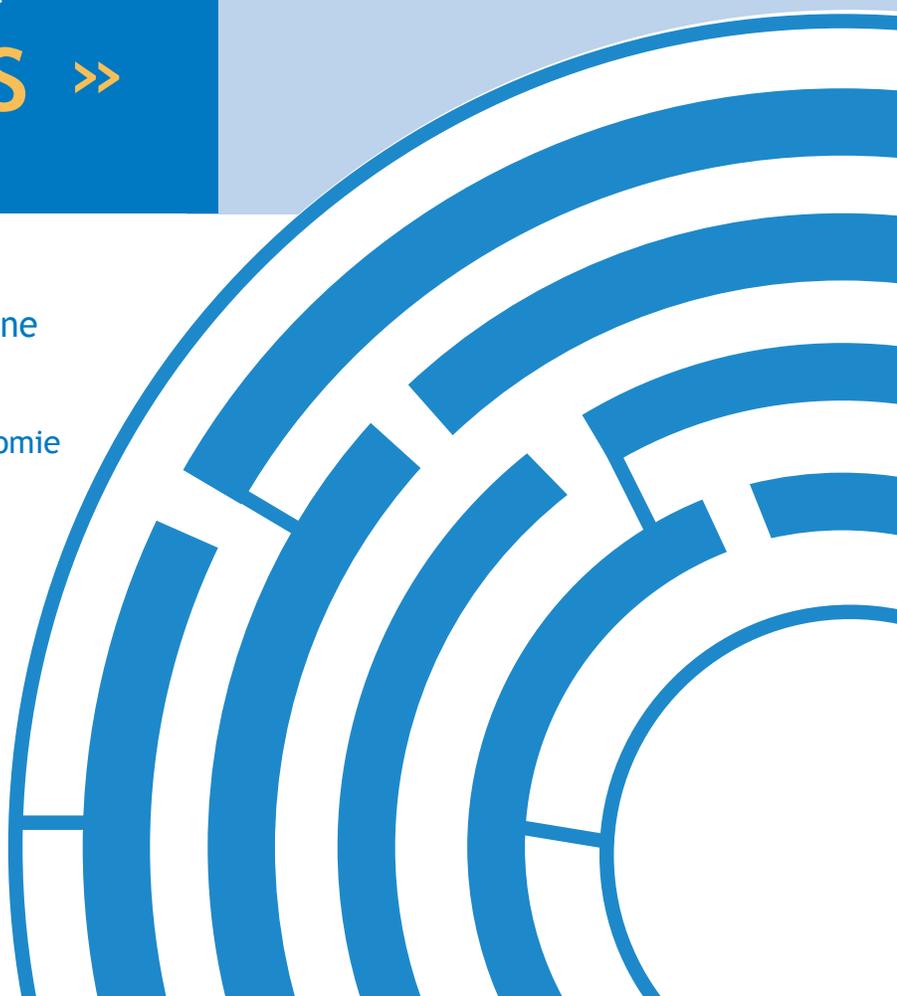


Table des matières

Hommage à la mémoire de Jean-Marie Jacques

Introduction

Du débat sur la place des logiciels dans l'analyse de données qualitatives

Véronique Dumont.....1

Texte de la conférence d'ouverture

Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative

Christophe Lejeune.....15

Section 1 : Développements logiciels : approche conceptuelle, fonctionnalités, visualisation

Langage de la crise ou crise du langage?

Jean-Marie Jacques, Nathanaël Laurent, Anne Wallemacq.....33

Analyse structurale de contenu et soutiens logiciels : une introduction au projet Anaconda

Jean-Pierre Hiernaux.....56

Logiciel NotaBene pour l'annotation linguistique. Annotations et conceptualisations multiples

Nicolas Mazziotta83

Contribution d'un outil de transcription dans la mise en mots de l'analyse d'un processus de conception collaborative observé dans sa profondeur phénoménale
Béatrice Arend.....95

Innovations dans un logiciel d'analyse qualitative de données : l'intégration d'outils de visualisation
Udo Kuckartz.....109

Section 2 : Usages de logiciels : comparaison, justification, confrontation

Stratégie d'utilisation différenciée de logiciels tout au long d'une enquête qualitative
Kerralie Oeuvery.....120

Traitement informatique de données orales : quels outils pour quelles analyses ?
Virginie André, Christophe Benzitoun, Emmanuelle Canut, Jeanne-Marie Debaisieux, Bertrand Gaiffe, Evelyne Jacquey.....131

Méthodes d'analyse des communications fonctionnelles en situation de travail collectif
Guillaume Gronier.....151

Section 3 : Méthodologie en sciences sociales. Place et rôle des logiciels

L'objet d'analyse comme pivot de l'analyse qualitative assistée par ordinateur
Jacques Hamel.....170

(Ne pas) Enseigner les méthodes qualitatives et (ni) former aux logiciels d'analyse des données qualitatives : une approche constructiviste de l'enseignement de la recherche qualitative
Diógenes Carvajal.....**181**

Ce numéro est dédié à la mémoire de Jean-Marie Jacques, décédé le 26 octobre 2009. Professeur à l'Université de Louvain, aux Facultés universitaires Notre-Dame de la Paix de Namur et membre de l'Académie royale des sciences de Belgique, Jean-Marie Jacques était aussi un des auteurs, enthousiaste, du logiciel Evoq. L'article qu'il avait soumis avec ses collègues Nathanaël Laurent et Anne Wallemacq du Research Center for Crisis and Conflict Management (RECCCOM, FUNDP) est publié dans ce numéro.

*Véronique Dumont, Christophe Lejeune et François Guillemette
Éditeurs de ce numéro*

Introduction

Du débat sur la place des logiciels dans l'analyse de données qualitatives

Véronique Dumont, Doctorante¹

Centre de recherche public Henri Tudor

Résumé

Les débats autour de l'utilisation de logiciels dans un processus d'analyse en sciences sociales nous renvoient dans différentes directions. Il y a, d'une part, la difficulté à catégoriser les outils logiciels disponibles pour l'analyse de données qualitatives et la multiplicité de justifications utilisées par les chercheurs. Au-delà d'une catégorisation de ce que « font » les logiciels et de ce qui les distingue entre eux, émerge la question des usages partiels, combinés, détournés de ceux-ci par les chercheurs. Si certaines injonctions de financeurs ou attentes de certains chercheurs tournent autour de l'utilisation d'un logiciel comme « garant » de la scientificité des résultats, il nous paraît utile d'insister sur la place du chercheur dans le processus d'interprétation et sur la validité globale de la démarche de recherche, avec ou sans outils logiciels.

Mots clés

LOGICIELS, DONNÉES QUALITATIVES, INTERPRÉTATION, MÉTHODOLOGIE, VALIDITÉ

Introduction

Dans la pratique des chercheurs, l'ordinateur fait partie des outils courants. Pourtant, il est intéressant de se poser la question de sa place dans un processus de recherche. Loin d'être une brique isolée du parcours, l'ordinateur est mobilisé, utilisé et agit en retour sur le processus. Il en devient un acteur plus ou moins actif en ce qu'il intervient dans la transformation, la traduction, l'inscription du phénomène en résultats au sens du travail scientifique tel que présenté par Bruno Latour (1989). Si l'incertitude sur la source de l'action reste ouverte, l'ordinateur ou les logiciels sont présentés dans des comptes-rendus de chercheurs, de commanditaires, d'évaluateurs, de collègues comme « faisant ou faisant faire quelque chose » (Latour, 2007).

RECHERCHES QUALITATIVES – Hors Série – numéro 9 – pp. 1-14.

LOGICIELS POUR L'ANALYSE QUALITATIVE: INNOVATIONS TECHNIQUES ET SOCIALES

ISSN 1715-8702 - <http://www.recherche-qualitative.qc.ca/Revue.html>

© 2010 Association pour la recherche qualitative

Est-on capable d'identifier les différents moments où l'ordinateur intervient? Certaines démarches ne sont-elles pas intégrées au point de le rendre invisible? Par ailleurs, certains usages sont aujourd'hui érigés en quasi-norme et en garantie de la qualité scientifique du travail de recherche effectué. Est-ce que l'utilisation d'un logiciel pour le traitement des données « rend » les résultats valides et scientifiques?

Parler de ressources logicielles permet de recadrer un peu la question tout en maintenant une ouverture très large, car la distinction « *hardware* – *software* » est très peu cohérente avec l'utilisation des chercheurs en sciences humaines. Peu de chercheurs sont capables d'identifier le type de processeur, de carte-mère, etc. qui composent les machines qu'ils utilisent. Contrairement à des ingénieurs ou des mathématiciens qui vont expliquer les choix matériels posés pour traiter leurs données, comme par exemple les réseaux d'ordinateurs constitués pour composer la puissance de calcul, les chercheurs sont relativement incapables de justifier leurs choix en ce sens. Comment s'attaquer à cette question dans nos disciplines?

Plusieurs alternatives s'offrent à nous. Nous pouvons, par exemple, identifier les fonctionnalités des ressources logicielles mobilisées ou créées par les chercheurs et les justifications données à l'utilisation de ces outils. Nous pouvons, par ailleurs, observer les pratiques des chercheurs. Nous pouvons enfin clarifier les exigences des disciplines et communautés scientifiques par rapport à l'usage de certains logiciels dans la recherche.

Pour introduire ce numéro spécial, il nous apparaît pertinent de raccrocher les aspects informatiques à la méthodologie de recherche pour en saisir les articulations, exprimées ou non, cohérentes ou non, et leurs implications. Si nous considérons que le logiciel agit dans le processus de connaissance scientifique, « il faut rendre compte de son action et, pour ce faire, il faut spécifier plus ou moins la nature des épreuves et celle des traces observables qu'elles ont laissées – ce qui ne veut pas dire qu'il faut se limiter aux actes de langage, la parole n'étant que l'un des nombreux comportements capables de produire un compte-rendu, et l'un des moins fréquents. » (Latour, 2007, pp. 76-77). En quelque sorte, il s'agit d'ouvrir la boîte noire que peut constituer l'ordinateur pour identifier les choix, les décisions et les positionnements d'acteurs divers et variés dans le processus d'analyse qualitative.

Identifier les ressources logicielles

Sans vouloir retracer ici un historique des développements logiciels pour les sciences humaines, d'autres l'ont fait avant moi (par exemple Lejeune, 2007),

je me propose de poser quelques points de repères qui permettent de situer ces ressources logicielles.

Entre les outils basés sur les mathématiques et les statistiques, disponibles très tôt pour la communauté de chercheurs étant donné leur proximité étroite avec le développement des ordinateurs, et l'éventail d'outils disponibles aujourd'hui pour les chercheurs en sciences humaines, il existe un vaste champ difficilement descriptible en quelques lignes.

Les chercheurs en sciences humaines et sociales collectent des données très variées sans toujours les traduire directement sous forme numérique. Ils « traitent » dès lors des données dites « qualitatives », que ce soit du texte, du son, des images voire tout cela à la fois. Ce numéro et les articles qui le composent n'ont donc pas abordé le traitement de chiffres, en tant que données de départ pour l'analyse.

Comment dresser le panorama de ce qui compose ce paysage de ressources logicielles? Quelques auteurs (Klein, 2001; Miles & Weitzman, 2003) se sont déjà risqués à cet exercice délicat. Citons notamment Christophe Lejeune qui est l'auteur du premier article dans lequel il propose une classification basée sur les fonctionnalités disponibles pour l'analyse de données textuelles. Il se distingue d'autres auteurs qui proposent des typologies des outils. La difficulté d'établir une typologie des outils réside dans le fait que le nombre et le spectre des outils s'élargit de façon croissante, certains devenant des couteaux suisses de plus en plus sophistiqués au fur et à mesure des versions (multi modules) dont la cohérence d'ensemble reste à démontrer. Au-delà des ressources logicielles dédiées directement à l'analyse de données qualitatives, d'autres ressources logicielles (traitement de texte, tableur, carte cognitive, etc.) sont mobilisées par des chercheurs avec ou sans programmation complémentaire de leur part.

À côté d'une catégorisation des fonctionnalités, on peut également entrevoir toute une série de distinctions variées (certaines sont inspirées de Miles & Weitzman, 2003) qui sont évoquées dans les choix posés par les chercheurs (ou leur laboratoire de rattachement) :

- l'accessibilité (coût, disponibilité dans le laboratoire) et/ou la philosophie de développement (*open source*, propriétaire);
- les compétences informatiques requises par l'utilisateur (compétences en programmation, temps de formation, ...);
- le type et la masse de données à traiter (son, vidéo, texte en format numérique ou scanné, un mixte);
- le terrain unique ou une comparaison multi-sites (classement ou regroupement des données);

- le traitement individuel ou coopératif des données (travail collectif sur l'encodage ou l'analyse, croisement des analyses);
- le format des résultats (export (ré)utilisable ou lisible par d'autres, modélisation);
- l'information / la formation à certains outils lors du cursus.

De la catégorisation et de la définition de la place des logiciels dans l'analyse

Si les chercheurs distinguent les logiciels entre eux par rapport à ce qu'ils « font ou font faire », créant ainsi certaines catégories, ils les distinguent également d'une méthode d'analyse « manuelle », « artisanale », de type « fluo, ciseaux et boîtes à chaussures ». Il ne s'agit donc pas d'une confrontation simple entre partisans et détracteurs des logiciels, il y a des arènes multiples de confrontation. Par exemple, Gaël de Peretti (2005) présente une partie de la controverse entre partisans de l'analyse statistique de texte sur les méthodes de transformation du texte en données permettant un traitement statistique.

À cet égard, l'article de Jean-Pierre Hiernaux présente les justifications qui lient la méthode d'analyse de contenu privilégiée par l'auteur, l'analyse structurale, et le choix du développement d'un outil spécifique permettant de rendre compte des phénomènes sémantiques, distinct d'instruments fondés sur les fréquences appropriés selon lui au traitement de phénomènes quantitatifs. Il définit ce qu'il entend par phénomène sémantique et en quoi cela constitue un fait sociologique. Cette définition lui permet de dégager la logique de l'analyse structurale et d'évaluer de la sorte l'adéquation des fonctionnalités de l'outil avec cette logique. À travers quelques illustrations, il démontre cette adéquation ainsi que la valeur ajoutée de l'outil par rapport à un traitement manuel équivalent. L'outil permet ainsi d'« articuler » les différents types d'opérations liées à l'analyse structurale « en un ensemble synoptique et interactif ». Une modification dans l'espace analytique, à quel que niveau que ce soit, est répercuté automatiquement dans les autres niveaux constitués. C'est la capacité de sélection, d'indexation, d'annotation et surtout de coordination de l'ensemble qui constitue ici l'atout de l'ordinateur. De plus, l'outil proposé a cette capacité de conserver les données de départ intactes, à l'inverse de la méthode manuelle, toutes modifications, sélections, annotations étant enregistrées dans un fichier distinct. La présentation de Jean-Pierre Hiernaux explicite également les différentes étapes, fichiers, processus qui accompagnent l'utilisation de l'outil. Le fichier de départ est ainsi transformé par le chercheur en catégorisation organisée et en modélisation de synthèse à partir des sélections et indexations qui se matérialisent par un surlignage jaune du texte sélectionné et par une marque d'indexation qui le précède et qui renseigne le

lecteur sur la 'collection d'extraits' (la boîte à chaussure qui se matérialise ici sous forme d'un fichier) dans laquelle ce segment a été déposé et à partir d'annotations et d'inscriptions qui condensent et décrivent les catégories qui se dégagent des collections.

Dans une optique similaire, Jean-Marie Jacques, Nathanaël Laurent et Anne Wallemacq présentent le logiciel qu'ils ont développé afin de suivre les processus de définition et de qualification du réel par des acteurs en situation (de crise, dans le cas présenté). Le processus d'analyse se concentre également sur le caractère plurivoque du langage et donc sur l'importance pour l'analyste de pouvoir étudier le rapport aux mots. La différence avec la démarche précédente se situe ici dans une vision processuelle en ce qu'ils souhaitent suivre l'émergence et la construction du sens qui intervient au fur et à mesure entre des acteurs en situation de coordination. L'outil est construit de façon différente afin de visualiser les relations entre les mots à travers des cartes sémantiques qui identifient les disjonctions et les conjonctions dans des textes, ce qui est appelé un paysage sémantique. La méthode se veut un aller-retour entre la cartographie proposée par l'outil et l'extrait à partir duquel elle a été construite, et ce, afin d'identifier comment les acteurs se positionnent dans cet espace sémantique. Différents extraits successifs sont analysés afin de reconstruire le processus symbolique à l'œuvre dans la situation étudiée. La différence fondamentale avec l'outil précédent (présenté par Hiernaux) réside dans le fait que l'identification des unités signifiantes et des relations est réalisée par l'outil, en fonction des critères intégrés au programme par les développeurs.

L'analyse structurale des données est une méthode d'analyse parmi d'autres. Certains logiciels ont été conçus en se rapportant à ce qui est appelé la théorisation ancrée ou *Grounded Theory*. C'est le cas d'ATLAS.ti et de NVivo. Jacques Hamel présente ce qu'il entend par analyse dans le cadre d'une telle approche et présente l'apport du logiciel comme soutien aux opérations méthodologiques identifiées en ce qu'il nécessite leur définition et leur explicitation. Ce sont les capacités d'indexation et de mise en liens (entre segments de texte, entre catégories) de l'ordinateur qui sont à nouveau mises en avant, en tant qu'instrument facilitateur pour l'analyste dans l'élaboration des relations entre les données collectées et la théorisation. Elles permettent de constituer le montage analytique de théorisation qui part des intuitions du chercheur, segmente les données, construit des catégories, y ajoute des définitions.

D'autres outils permettent de réaliser un travail d'exploration et d'analyse sur des données orales, des enregistrements vidéos tout en intégrant

une coordination avec une retranscription textuelle éventuelle. Un exemple d'outil permettant de visualiser différentes données issues d'une situation est présenté par Béatrice Arend. Elle souhaite étudier une situation de conception collective en suivant à la fois les traces pas à pas du résultat (sur l'écran de l'ordinateur) de cette activité, l'enregistrement audiovisuel (tour de paroles, gestuelles, mimiques, utilisation des objets, silence, etc.) et une retranscription réalisée par le chercheur. Les différentes modalités de retranscription et de sa visualisation ont constitué l'objet d'un questionnement dans son projet. Les échanges et discussions avec le concepteur ont finalement abouti au choix d'une présentation sous forme de partition de la retranscription. Celle-ci répond à la volonté de visualiser les dimensions verbales et non verbales des interactions et leur enchaînement. L'apport de la visualisation grâce à un outil logiciel dans un processus d'interprétation est également abordé par Udo Kuckartz, concepteur du logiciel MAXQDA. Cette dimension visuelle (graphes, modèles, cartes cognitives, etc.) est abordée dans sa dimension évaluative pour transmettre et discuter des résultats et du processus de recherche.

Tout comme pour les logiciels traitant de corpus textuels uniquement, il y a une série de « conditions » posées par les logiciels quant au format des données, qui doivent dès lors faire l'objet d'une numérisation et d'une traduction de certains formats vers d'autres. Les retranscriptions de données orales et visuelles sous forme de texte sont parfois accompagnées d'une symbolique conventionnelle qui permet d'intégrer tantôt des données contextuelles (bruit, pause, hésitations, etc.) tantôt des données propres au chercheur (passage qu'il ne comprend pas, commentaires, etc.). Certains outils proposent de créer des liens entre le fichier son ou vidéo et la retranscription. L'article des membres du laboratoire ATILF (André, Benzitoun, Canut, Debaisieux, Gaiffe & Jacquy) et celui de Nicolas Mazziotta explicitent bien ce travail réalisé sur les données collectées, leur mise en forme et en format, les liens entre les différents fichiers (alignement texte-son, fiche documentaire associée, annotations) ainsi constitués et les visualisations que permettent les outils. Il s'agit donc bien ici, en premier lieu, de faciliter la prise en main du corpus par le chercheur et plus largement par un ensemble de chercheurs (d'où les conventions et standards). À cet égard, d'autres outils sont également disponibles afin de structurer et coordonner un corpus contenant différents types de données. Dans le cadre de nos recherches, nous avons utilisé un wiki pour indexer et créer des liens entre différents fichiers textes, sons, vidéos de formats variés et également des sites internet. L'objectif est de pouvoir retracer l'historique des traces collectées tout en accédant facilement à celles-ci.

Dans tous ces débats, il s'agit d'identifier ce que le logiciel (ou la fonctionnalité) apporte à l'analyse (en positif et en négatif) et comment il le fait. Il y a, d'une part, les justifications apportées a priori ou a posteriori par les chercheurs et les développeurs (motivations épistémologiques, méthodologiques, théoriques, contingentes) et, d'autre part, les traces laissées pendant et par l'utilisation du logiciel. Il n'existe pas de consensus sur le positionnement de ces logiciels dans la sphère de l'analyse qualitative, la question se posant a priori pour les logiciels fonctionnant sur base de statistiques textuelles, mais également plus largement sur les pratiques d'utilisation des logiciels. Par contre, il existe un débat sur la définition de l'analyse qualitative, avec des distinctions importantes selon les disciplines des chercheurs. Il y a également la question de savoir qui mène l'analyse... et la réaffirmation de la place du chercheur dans le processus.

Dans l'approche présentée par Jean-Pierre Hiernaux, l'outil ne peut être « automatisable », il vient en « soutien », en « support », en complément aux instruments de base de l'analyste que sont « ses propres yeux et cerveau ». C'est l'analyste qui travaille, teste les structures en mobilisant certaines « capacités » ou « disponibilités » des ordinateurs qui peuvent *contribuer* « à la restitution des relations et topologies sous-jacentes aux données » en remplaçant « avantageusement les crayons, les surligneurs, les ciseaux, la colle et les boîtes à chaussures ». L'analyste est le principal acteur du résultat, et non ses instruments, en ce que « la rigueur des instruments conceptuels ou matériels n'ôte en rien à l'art que demeure leur mise en œuvre optimale ni à ce que doit à cet art le résultat obtenu ». Jacques Hamel rappelle également à quel point l'analyse est dirigée par le chercheur, en obéissant « à des opérations et à des règles précises ». L'ordinateur vient en soutien de certaines de ces opérations. Christophe Lejeune introduit la notion de division du travail scientifique pour souligner la place de l'ordinateur et du chercheur dans le processus de transformations successives du corpus. En ce sens, les dictionnaires permettent à l'ordinateur de remplacer les « en »codeurs, dont l'approche de l'analyse de contenu a tenté de limiter au maximum la subjectivité dans le travail de codage. À l'inverse, codeur et analyste ne font qu'un dans la création de registres telle que conçue par exemple à travers le logiciel *Cassandra*. La place de l'ordinateur souhaitée par les uns et les autres dans le processus scientifique dépend donc largement des attributs positifs et négatifs associés, d'une part, à la dimension humaine (et donc à la « subjectivité ») et, d'autre part, à la dimension informatique, donc à la position épistémologique et ontologique de ces chercheurs.

Des pratiques diverses et variées

Au-delà des justifications liées à l'utilisation d'un logiciel « dédié » à l'analyse, les pratiques des chercheurs impliquent des détournements et des combinaisons d'outils très variés.

Kerralie Oeuvery nous propose une analyse des différentes *logiques d'utilisation* des logiciels qu'elle a dégagées à partir de sa propre pratique. Dans le processus d'interprétation, elle distingue différentes phases (non linéaires) où les logiciels peuvent être présents. Il y a la phase d'examen minutieux des données qu'elle intitule *logique de proximité décomplexée* où les données sont analysées et déconstruites/reconstruites sous forme de catégories et de codes. C'est à ce stade que l'on retrouve l'utilisation des logiciels de type CAQDAS. Le travail du chercheur, avec ou sans logiciel, consiste à définir les ensembles cohérents et à les nommer sous un intitulé (code ou catégorie). L'utilisation du logiciel permet, selon elle, de multiplier les interprétations possibles et de revenir facilement aux extraits de texte pour modifier / améliorer le codage. À l'inverse, cette facilité devient parfois un piège pour certains chercheurs qui multiplient à outrance le nombre de catégories (au point de coder tous les segments de texte). Diógenes Carvajal précise que cette pratique repose sur une mauvaise compréhension de la logique de cette démarche et d'un amalgame avec une démarche quantitative où la masse et le traitement de l'ensemble du corpus est considéré comme gage de scientificité. La *logique de proximité créative* repose sur la construction de relations entre les codes et catégories et d'interprétations locales, prémises d'une construction théorique. Certains logiciels CAQDAS (NVivo par exemple) proposent de créer des mémos qui permettent notamment aux chercheurs de définir ces premiers cadres. D'autres (comme Weft QDA) n'intègrent pas la fonctionnalité mémo en tant que telle. Il est toutefois possible d'utiliser un traitement de texte traditionnel pour les créer et les intégrer ensuite dans le corpus du logiciel. Le suivi du questionnement et des réflexions du chercheur constitue le point essentiel de cette phase, qui permet de relier les schèmes explicatifs que constitue la *logique de distanciation* aux données empiriques. Un ensemble de mémos, articles scientifiques, communications, modélisations composent le paysage du chercheur à la recherche des inspirations et déclics qui lui apporteront une touche particulière par rapport aux théorisations existantes. À nouveau, certains logiciels intègrent ces outils, bien que l'usage d'un traitement de texte, d'un outil de présentation (powerpoint) ou de modélisation (freemind, inspiration, etc.) soit une alternative logicielle accessible et régulièrement utilisée. La structuration du rapport final peut se construire ici, soit directement fondée sur la catégorisation finale obtenue, soit à partir des réflexions multiples

que celle-ci a suscitées. Pour ce qui est de la *logique de vérification*, elle prend place au cours du processus jusqu'à la rédaction, soit pour extraire des illustrations utiles lors de la rédaction du rapport, soit pour provoquer une émulation suite à des interprétations divergentes ou des cas discordants avec la théorisation en cours. Les logiciels dédiés et d'autres ressources informatiques générales permettent de rechercher un extrait dans un ensemble de documents, soit en sélectionnant les extraits regroupés sous une catégorie, soit à partir de mots-clés.

Il est intéressant de noter également la différence d'usage d'une même fonctionnalité. Par exemple, le concordancier est présenté par Christophe Lejeune comme un réagencement de segments de texte autour d'un mot ou d'une expression qu'il rattache à une épistémologie exploratoire et qualitative, à la différence des outils lexicométriques qui produisent des statistiques textuelles. Cela ressemble aux outils de recherche intégrés dans certains CAQDAS et qui permettent d'extraire un ensemble de segments de texte d'un corpus à partir d'un mot ou d'une expression. Il y a ici un élément complémentaire dans la visualisation qui met au centre le mot recherché et de part et d'autre le contexte (avec une largeur variable, parfois définissable par le chercheur). Pourtant, l'usage proposé par le laboratoire ATILF (André et al.) montre que cet outil sert également, en plus de cette visée exploratoire, à sélectionner une partie du corpus (autour d'un phénomène) pour ensuite la soumettre à un traitement quantitatif (fréquence) au travers d'un autre outil (tableur). L'article insiste aussi sur le panel d'outils mobilisés dans le cadre de l'exploitation de leur corpus, outils qui restent parfois le crayon et le papier, voire une grille de codage imprimée sans qu'il leur paraisse souhaitable de vouloir tout informatiser. Ces différences démontrent que le fait pour un chercheur de citer un logiciel mobilisé sans explicitation du sens de la démarche et de son utilisation est insuffisant. Demazière (2005) souligne ainsi que :

si ces méthodes logicielles incorporent des procédures cadrées et spécifiques de traitement des matériaux, elles servent aussi des objectifs de connaissance variés et définis en amont par le sociologue. Si elles font des choses différentes inscrites dans leurs algorithmes, elles s'articulent à des approches sociologiques contrastées et élaborées avec d'autres outils, qu'on appelle habituellement théoriques, conceptuels, analytiques, etc.

Ces différences d'usages se retrouvent également pour les outils présentés par Christophe Lejeune comme des outils réflexifs, en ce qu'ils permettent la construction progressive d'une catégorisation vers une théorisation « enracinée », fondement du développement de ces outils. Ce type

d'usage est présenté par différents contributeurs de ce numéro. Toutefois, comme souligné plus haut en ce qui concerne les outils lexicométriques qui, de part leur construction, reposent sur des analyses statistiques de texte, certains outils réflexifs sont utilisés par les chercheurs dans une approche quantitative (pourcentage de segments codés, déclinés sur base de variables) et déductive (grille de codage prédéfinie) propre à l'analyse de contenu et qui repose sur des présupposés très différents. Tout comme l'habit ne fait pas le moine, l'outil ne fait pas l'analyse qualitative. L'enjeu ici semble donc moins être de pouvoir identifier l'outil utilisé (ou les outils utilisés), mais plutôt si le chercheur en précise l'usage et la finalité dans sa démarche particulière. L'utilisation de plusieurs méthodes d'analyse, assistée par ordinateur ou non, ou du croisement de plusieurs cadres d'analyse (construits par différents membres d'une équipe) peut apporter un plus à la démarche, non dans le sens d'une convergence obligatoire des résultats pour validation (ce qui est avancé dans certaines démarches plus proches de l'analyse de contenu), mais plutôt dans le sens d'une révélation de différences qui permettent d'enrichir l'interprétation.

Quels que soient les outils mobilisés, le chercheur reste au centre du processus d'interprétation. C'est ainsi que même l'usage d'outils logiciels de type « automate », comme le rappelle Christophe Lejeune, nécessite un travail important de la part du chercheur, d'une part, pour comprendre ce que le logiciel fait aux données (et éventuellement modifier les paramètres par défaut) et, d'autre part, pour donner un sens aux catégories qu'il propose. C'est la démarche présentée par Guillaume Gronier dans le cadre de l'utilisation du logiciel Alceste. Il souligne que cet outil ne fait que proposer « une structuration et une représentation des informations présentes au sein d'un corpus textuel » et que l'interprétation reste du ressort de l'analyste. Cette remarque est valable également pour les utilisateurs de logiciels réflexifs qui omettent de donner du sens à une modélisation de nœuds et de relations ou qui se limitent à un petit paragraphe à la suite de la figure; voir, à ce propos, l'article de Diógenes Carvajal.

Les attentes de la communauté scientifique ou des financeurs

Si les recherches basées sur des analyses statistiques se reposent systématiquement sur l'un ou l'autre outil logiciel explicitement identifié, il n'en est pas de même pour les analyses qualitatives encore régulièrement réalisées « à la main » ou pour lesquelles l'utilisation de logiciels n'est pas systématiquement mentionnée. Pourtant, de plus en plus, l'utilisation d'un logiciel devient un critère de validité scientifique en soi, sans lequel il est difficile d'obtenir un financement de recherche. Diógenes Carvajal souligne que les jeunes chercheurs et les étudiants, plus familiarisés avec les outils

informatiques, viennent solliciter les professeurs et les universités pour être formés et accéder à de tels outils dans le cadre de leur cursus de formation. La demande vise des outils qui sont cités largement dans la littérature et dont ils espèrent que la légitimité gagnée au sein de la communauté scientifique rejaillira sur leurs résultats. Toutefois, tant en Colombie que dans de nombreux autres pays, le coût de certains logiciels et des formations associées est un frein pour les universités et pour les étudiants. Il reste ensuite à résoudre l'épineuse question de savoir comment enseigner l'utilisation des logiciels aux étudiants. La proposition de Diógenes Carvajal consiste à ancrer cet apprentissage à travers la pratique en favorisant le développement d'un regard réflexif et critique des étudiants sur l'ensemble de la démarche de recherche.

Plutôt qu'une mode, une croyance en une boîte magique ou une exigence des financeurs qui poussent les chercheurs à utiliser ces outils logiciels, les auteurs de ce numéro présentent les arguments 'scientifiques' qui justifient leur utilisation par les chercheurs au sein d'une démarche de recherche. Ainsi, Jean-Pierre Hiernaux et Jacques Hamel soulignent chacun l'aide du logiciel pour restituer l'exigence du lien entre les données empiriques et les structures ou cadres d'analyse dégagés par le chercheur. Jacques Hamel insiste particulièrement sur l'importance de l'objet d'analyse comme élément fondamental, au-delà de l'utilisation d'un logiciel. Le logiciel vient principalement éclairer et illustrer les règles et les opérations menées par le chercheur pour obtenir les résultats présentés. Il s'agit, d'après lui, d'un complément utile pour donner « un éclat de rigueur » à une méthodologie qui souffre d'une image d'opacité et de manque de systématisation de l'analyse. Loin de remplacer l'explicitation et la présentation de la méthodologie, les logiciels sont ici placés en lien avec des épreuves scientifiques classiques où le chercheur va démontrer la cohérence et la pertinence de son dispositif d'enquête. L'outil est présenté comme un apport en termes de lisibilité et d'explicitation de la démarche. Il permet peut-être enfin, selon Jacques Hamel, « d'élaborer l'analyse qualitative tout en étant capable de dire ce que l'on fait et de faire ce que l'on dit. ».

Kerralie Ouevray souligne la méfiance existante dans la communauté scientifique par rapport à la place du chercheur et du logiciel dans le processus d'interprétation. En plus des débats portant sur la « technique », elle souligne l'injonction de plus en plus présente de présenter les preuves, les traces de l'ancrage des résultats du chercheur dans les données empiriques. Cette injonction tend à définir l'outil logiciel comme validateur et créateur unique de ces traces. Or, entre cette ligne théorique qui veut que le chercheur utilise un logiciel et celle qui veut qu'il n'utilise rien, il se dégage des usages bien plus

variés et complexes de différents outils. De plus, l'usage d'un outil ne procure aucun gage de scientificité de facto à la démarche et aux résultats.

Si certains auteurs parlent d'augmenter ainsi la rigueur et la systématisation du processus d'analyse là où d'autres envisagent plutôt l'émulation de la créativité du chercheur (Demazière, 2005), ces aspects ne sont pas antinomiques et se combinent pour aboutir au résultat final d'un projet de recherche. C'est ce qui permet de dire qu'il s'agit à la fois d'une science et d'un art.

Qu'en est-il de la cohérence globale?

Au-delà des outils, il y a une démarche de recherche que la majorité des auteurs de ce numéro ont soulignée, avec ses fondements épistémologiques, méthodologiques, conceptuels, ses contingences qui orientent les choix. Christophe Lejeune introduit ce lien fort en proposant d'identifier « ce que l'informatique fait (faire) à l'analyse qualitative » en l'inscrivant sur les axes description – dénombrement et exploration – analyse. Il s'agit en effet d'identifier et de justifier les choix de méthodes et d'outils mobilisés au sein du parcours de recherche. À ce niveau-là, il n'y a pas de changement majeur dans les exigences scientifiques.

Les différents auteurs présentent le processus d'interprétation comme un parcours aux multiples facettes dans lequel différents outils, logiciels ou non, peuvent trouver leur place. Toutefois la présence de ces outils augmente les choix à justifier et pose donc un défi supplémentaire pour le chercheur dans l'explicitation du sens de sa démarche.

À cet égard, la réflexion proposée par Pourtois, Desmet et Lahaye (voir Figure 1) qui vise à atteindre ce qu'ils ont appelé une validité de reliance entre les différents points charnières d'une démarche scientifique permet de rendre compte de la cohérence et également de l'originalité de tout parcours, quels que soient les choix et orientations pris par le chercheur. Cette proposition permet d'élargir la conception de la validité d'une démarche de recherche pour y intégrer les approches phénoménologiques et herméneutiques, exclues des épreuves classiques fondées sur l'effacement maximal de la subjectivité du chercheur et des sujets. Cette modélisation peut être utile pour situer et identifier la place des outils, logiciels ou non, mobilisés aux différents points charnières et à différents moments du parcours de recherche.

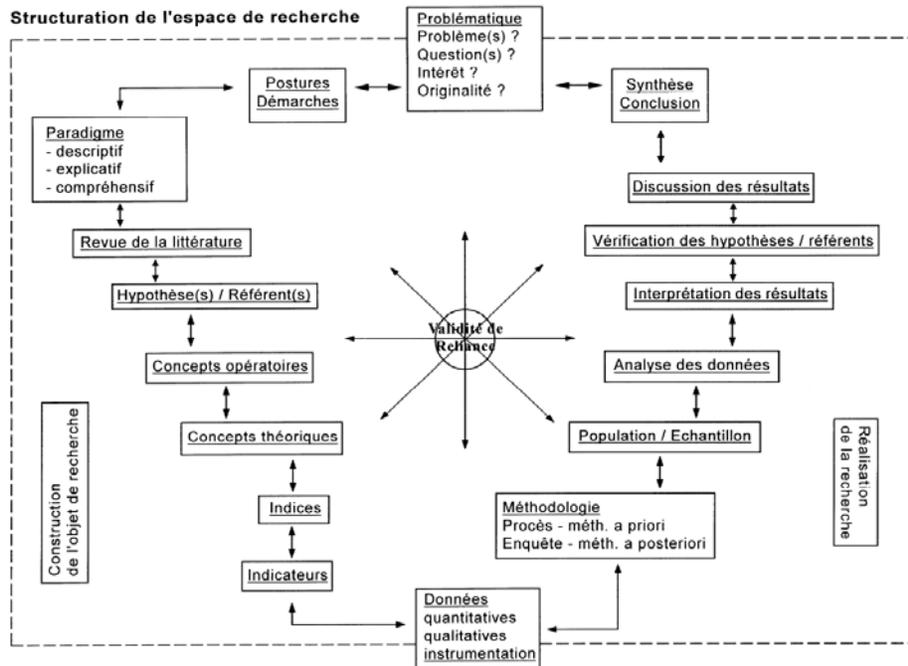


Figure 1. Les points charnières de la recherche (extrait de Pourtois et al., 2006)

Note

¹ Thèse financée dans le cadre du programme AFR du Fonds National de la Recherche luxembourgeois (www.fnr.lu)

Références

- Demazière, D. (2005). Des logiciels d'analyse textuelle au service de l'imagination sociologique. *Bulletin de méthodologie sociologique*, 85, 5-9.
- De Peretti, G. (2005). La « mise en variables » des textes : mythe ou réalité? *Bulletin de méthodologie sociologique*. 88, 5-30.
- Klein, H. (2001). Overview of text analysis software. *Bulletin de méthodologie sociologique*. 70, 53-66.
- Latour, B. (1989). *La science en action*. Paris : La Découverte.
- Latour, B. (2007). *Changer de société, refaire de la sociologie*. Paris : La Découverte.

- Lejeune, C. (2007). Petite histoire des ressources logicielles au service de la sociologie qualitative. Dans C. Brossaud, & B. Reber (Éds), *Humanités numériques, Tome 1. Nouvelles technologies cognitives et épistémologie* (pp. 197-214). Paris : Hermès.
- Miles, M.B., & Weitzman, E.A. (2003). Choisir des logiciels informatiques pour l'analyse des données qualitatives. Dans M.B. Miles, & A.M. Huberman (Éds), *Analyse des données qualitatives* (2^e éd.) (pp. 561-573). Bruxelles : De Boeck.
- Pourtois, J.-P., Desmet, H., & Lahaye, W. (2006). Postures et démarches épistémiques en recherche. Dans P. Paillé (Éd.), *La méthodologie qualitative. Postures de recherche et travail de terrain* (pp. 169-200). Paris : Armand Colin.

Véronique Dumont est doctorante en sciences politiques et sociales, inscrite aux Facultés universitaires Notre-Dame de la Paix de Namur (FUNDP), Belgique, et financée par le Fonds national de la recherche du Luxembourg et par le Centre de recherche publique Henri Tudor. Sociologue de formation, elle s'intéresse aux controverses entourant l'échange électronique de données de santé, en particulier aux débats relatifs à la vie privée. Elle est membre du comité d'animation du Comité de recherche 29 de l'AISLF « sciences, innovations technologiques et société » et a été chercheuse au sein de la Cellule interdisciplinaire de Technology Assessment des FUNDP pendant près de 5 ans, ainsi qu'au Centre de recherche publique Henri Tudor.

Texte de la conférence d'ouverture

Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative

Christophe Lejeune, Ph.D.

Université de Liège

Résumé

Au-delà de ses implications techniques, l'informatisation de l'analyse de textes soulève des questions de méthode. Après un recensement des motivations à recourir à des logiciels pour assister l'analyse de données textuelles, nous proposons une typologie distinguant cinq fonctionnalités proposées par ces outils : la lexicométrie, les concordances, les automates, les outils réflexifs et les dictionnaires. Chacune de ces fonctionnalités éclaire les textes d'une manière originale (et partielle). Rapportée à ses usages attestés, chacune de ces potentialités articule des présupposés épistémologiques spécifiques. En définitive, le choix d'un logiciel caractérise autant une démarche de recherche que l'inscription dans un paradigme.

Mots clés

ANALYSE TEXTUELLE, LOGICIELS, OUTILS MÉTHODOLOGIQUES, ENJEUX ÉPISTÉMOLOGIQUES

Pourquoi recourir à un logiciel?

L'idée de recourir à l'informatique dans le cadre d'analyses qualitatives rebute encore beaucoup de chercheurs en sciences humaines et sociales (contrairement au travail quantitatif pour lequel l'ordinateur est devenu un point de passage obligé)¹. Cette réticence provient sans doute en partie de la méconnaissance de ces outils (Barry, 1998, § 2.4) et en partie de l'assimilation de l'ordinateur à un artefact technique produit par l'industrie et procédant de l'application systématique de la logique mathématique. Un tel outil serait par essence peu adéquat avec le travail de terrain, réclamant pour sa part une attitude compréhensive envers des modes de raisonnement et des logiques sociales qui échappent de très loin au logicisme. Il est pourtant possible de tirer parti des

RECHERCHES QUALITATIVES – Hors Série – numéro 9 – pp. 15-32.

LOGICIELS POUR L'ANALYSE QUALITATIVE: INNOVATIONS TECHNIQUES ET SOCIALES

ISSN 1715-8702 - <http://www.recherche-qualitative.qc.ca/Revue.html>

© 2010 Association pour la recherche qualitative

caractéristiques des outils informatiques sans pour autant faire violence à la démarche qualitative. Il s'agit dès lors d'outiller le travail d'analyse des documents rassemblés (entretiens, témoignages, rapports officiels, coupures de presse, monographies).

Pour se justifier, l'outil doit apporter une « valeur ajoutée » au travail du qualitatifiste. Cet apport consiste en une facilitation de l'exploration du matériau empirique, une explicitation (et, partant, une reproductibilité) des procédures analytiques, une ouverture sur l'analyse à plusieurs ou un archivage de l'activité scientifique.

La motivation première du passage par un logiciel est épistémologique. Celle-ci participe d'une option méthodologique forte : il s'agit, pour le chercheur, de se donner une contrainte de travail participant à garantir la rigueur de sa production. En passant par l'outil informatique, il s'oblige ainsi à expliciter chaque étape de la construction de son analyse (puisque, par définition, un logiciel ne peut appliquer une procédure implicite). Une telle explicitation ne passe pas (nécessairement) par une procéduralisation; il ne s'agit pas, pour le chercheur, de dégager les règles sous-jacentes à son analyse et de les incorporer à une machine. Plus simplement, il consigne chaque association qu'il opère entre un élément empirique et une composante thématique ou analytique. Il s'agit là d'une étape classique en analyse qualitative; l'introduction de l'ordinateur n'en change pas la nature mais lui offre un support (et, partant, en contraint l'inscription). Ce faisant, le chercheur explicite les médiations entre le terrain et l'interprétation, ce qui rend tangibles les éléments sur lesquels reposent les conclusions du chercheur. Non seulement, cette clarification rend observable et descriptible le cours de l'interprétation mais, bien plus, elle offre à l'analyse qualitative la possibilité que soit conduite une épreuve jusque-là réservée aux seules sciences expérimentales : la reproductibilité de l'analyse.

Cette explicitation draine également plusieurs apports de la médiation informatique. Clarifiées, les procédures analytiques en deviennent plus aisément transmissibles. Une fois inscrites dans le support informatique, les étapes successives du cheminement interprétatif peuvent être exposés à des tiers, comme les étudiants, les commanditaires ou les collègues (dans les publications ou les exposés). Ces derniers peuvent ainsi reconduire l'analyse (c'est dans ce sens que nous parlons de reproductibilité). Le dépôt dans l'outil informatique peut donc participer à la dynamique du champ scientifique. De Merton (1953) à Latour (1995), les travaux en sociologie des sciences ont insisté sur la dimension structurante et fédératrice de la discussion entre pairs. Celle-ci implique de se reposer sur les travaux antérieurs, de citer les travaux

portant sur des terrains homologues ainsi que d'être à même de comparer – et de critiquer – ces derniers. Les divergences apparaissant éventuellement à travers une telle confrontation des résultats renvoient les chercheurs aux spécificités de leurs recherches respectives et, par conséquent, à la singularité de leur matériau. La confrontation appelle donc la comparaison des analyses.

En outre, tout comme les notes couchées dans le journal de bord, le dépôt dans le système informatique permet d'assister et d'organiser la mémoire des recherches. En effet, au cours d'une même carrière, le chercheur est amené à conduire des recherches qui se font écho l'une à l'autre. Il arrive alors qu'il éprouve le besoin de mobiliser des éléments de cadres d'analyse construits antérieurement (Mayring, 2000, § 18). Disposer alors du journal de bord des analyses antérieures s'avère précieux.

Enfin, si elle facilite les comparaisons, la communication des médiations intermédiaires du travail analytique (et non seulement de ses conclusions) facilite également les collaborations. Cette fois, il ne s'agit plus tant de confronter des résultats que de collaborer dans l'analyse d'un même matériau. L'explicitation du cadre d'analyse restreint les risques de malentendu dans le travail à plusieurs. En outre, lorsque la plate-forme informatique tire parti des réseaux de communication (comme Internet), les collaborations deviennent plus aisées, même à distance. Avec la multiplication des projets internationaux impliquant sans cesse plus de partenaires, disposer d'une telle plate-forme dédiée à l'analyse qualitative en sciences humaines et sociales devient crucial.

Tout comme les logiciels de travail collaboratif (CSCW), une plate-forme informatique peut faciliter la comparaison, la mémoire et la collaboration. Les logiciels d'analyse qualitative peuvent ainsi assister la rigueur de la recherche scientifique. Les motivations – louables – à recourir à ces outils ne doivent cependant pas éclipser certains risques qu'engendre leur usage. Si certains chercheurs hésitent encore à y recourir, les commanditaires s'avèrent généralement rassurés par l'usage de techniques informatisées (Paillé, 2006, p. 152). Or, s'il peut effectivement être un adjuvant, le logiciel ne sera jamais un garant du sérieux de l'analyse conduite (ceci est d'ailleurs vrai tant pour l'approche quantitative que qualitative). Aucun outil ne doit donc éclipser la vigilance épistémologique et la rigueur méthodologique du chercheur (Bandeira-de-Mello & Garreau, à paraître). S'il l'on n'y prend garde, l'analyse se réduit à un relevé – plus ou moins systématique – des témoignages des acteurs ne s'articulant à aucune conclusion et dont la portée scientifique est, en conséquence, très limitée.

En tant qu'ils exigent l'explicitation des étapes successives de l'analyse, les logiciels d'analyse qualitative constituent des instruments, des laboratoires,

des espaces de calcul, auxquels le chercheur peut adosser son interprétation. Pour autant qu'ils s'intègrent à une démarche méthodologique réfléchie et rigoureuse, les logiciels contribuent donc à consolider les conclusions des recherches qualitatives.

Le jeu des cinq familles

La section précédente a présenté les raisons qui motivent le recours aux logiciels d'analyse qualitative. Dans la présente section, nous distinguons cinq fonctionnalités proposées par les logiciels d'analyse textuelle. Nous situons chacune d'entre elles par rapport aux autres en explicitant ses apports, ses limites et ses présupposés épistémologiques. Une telle explicitation nous amène bien entendu à rapprocher ces outils de l'une ou de l'autre technique de l'analyse textuelle (comme l'analyse de contenu, l'analyse de discours ou la méthode par théorisation ancrée).

Plusieurs recensions (Klein, 2001; Jenny, 1996; Popping, 1997; Weitzman & Miles, 1995) témoignent du grand nombre d'outils informatiques proposés pour assister l'analyse qualitative. Du point de vue de leur fonctionnement, nous proposons de distinguer ces logiciels selon deux axes : explorer-analyser et calculer-montrer (voir Figure 1).

La première dimension distingue les outils œuvrant à l'exploration du matériau empirique et ceux permettant de représenter les catégories d'analyse. Certains outils n'assistent que la phase d'exploration du matériau : les catégories d'analyse ne sont alors pas représentées dans le système informatique. C'est le cas des outils du haut de la Figure 1. À l'inverse, d'autres outils les représentent (c'est le cas des outils du bas de la Figure 1). Toute analyse repose évidemment sur une exploration préalable. Les deux pôles ne sont pas pour autant hiérarchisés. C'est plutôt l'existence d'outils du premier type qui nous a incité à établir cette distinction.

Le deuxième axe distingue le recours à un calcul ou à un réarrangement des textes. Même lorsqu'il est question de matériau textuel, iconographique ou vidéographique, l'usage du calcul est possible (et attesté). À l'opposé, l'analyse de ces matériaux peut également reposer sur des procédures d'organisation des matériaux permettant de les parcourir d'une manière non linéaire. Tant l'exploration que la construction de catégories d'analyse peuvent reposer soit sur des procédures quantifiées (à gauche de la Figure 1) soit sur des procédures non quantifiées (à droite de la Figure 1).

La combinaison des deux axes permet de répartir les outils d'analyse de textes en cinq familles : la lexicométrie, les concordanciers, les outils automatiques, les registres et les outils réflexifs. Ces cinq familles ne sont pas

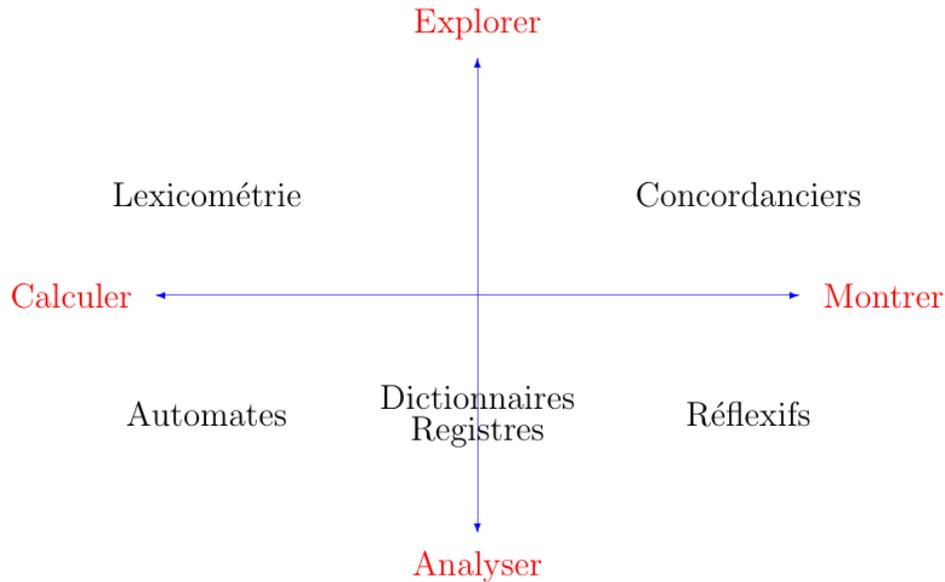


Figure 1. Familles de fonctionnalités

mutuellement exclusives : plusieurs outils (comme Unitex, Hyperbase ou T-LAB²) convoquent des fonctionnalités qui relèvent de différentes familles³. Dans la suite de cet article, nous examinerons alternativement chacune de ces cinq familles. Son positionnement épistémologique sera à chaque fois rapporté aux apports et aux limites des techniques informatiques.

Lexicométrie

L'analyse lexicale – connue en France en tant que *lexicométrie* – repose sur l'exploration statistique du lexique d'un corpus de textes. Les outils statistiques classiques, comme les tableaux de fréquences, les tris croisés voire l'analyse factorielle des correspondances, permettent (notamment) de déterminer la fréquences des mots⁴ d'un corpus, leur répartition dans différents textes ou leurs éventuelles sur- (ou sous-) représentation. Des procédures très rigoureuses ont été mises au point pour préparer le corpus ou pour exploiter ces résultats statistiques. Au niveau de la préparation, la lemmatisation ramène à une même forme les nombreuses variations présentes dans le langage (par exemple, les différentes formes conjuguées d'un verbe sont réunies sous sa forme infinitive). Au niveau du traitement, les cooccurrences identifient les termes qui

Terme	Frq Tot.	Frq P...	Spécif
nous	1270	66	5
faire	412	25	4
chaque	33	6	4
toi	88	10	4
marcher	41	6	3
année	14	3	3
comment	33	5	3
pendant	77	7	3
millions	43	5	3
présent	24	4	3
savons	12	3	3
mille	55	6	3
vous	1097	52	3
subsistances	47	6	3
département	14	3	3
queule	12	3	3

Figure 2. Liste de fréquences (copie d'écran de Lexico)

apparaissent souvent ensemble (dans la même phrase, le même paragraphe ou le même texte).

Seules les propriétés objectivement observables dans le texte sont prises en compte. La question de l'analyse, voire de l'interprétation, est confiée à l'analyste sans qu'un espace d'inscription soit prévu dans le logiciel. Des logiciels comme Lexico, Spad-T, Sato, Hyperbase et le projet d'unification Textométrie sont bien des outils mobilisant le calcul à des fins d'exploration (représentés dans le quadrant supérieur gauche de la Figure 1). Baser l'exploration sur des procédures de calcul témoigne d'une double inscription épistémologique quantitative et descriptive.

Tout d'abord, le recours à des procédures statistiques de dénombrement inscrit la lexicométrie dans une approche quantitative du texte. De tels calculs de la fréquence d'apparition de certaines formes textuelles peuvent être réalisés sur de très larges corpus de textes (éventuellement diachroniques). Associés aux dispositifs classiques de visualisation statistique (histogrammes, boîtes à moustaches, dendrogrammes), les listes de fréquences facilitent une vision d'ensemble, résumée, de ces larges corpus empiriques. En outre, elles en offrent un abord original, transversal, très différent de celui de la lecture linéaire (Gobin & Deroubais, 1994, pp. 69-70) et disposent, par conséquent, d'un vrai pouvoir de révélation. Ces vertus ont cependant leurs revers : d'une part, le découpage du texte en unités élémentaires – les mots – procède d'une

atomisation du texte, nuisible à sa compréhension. Par ailleurs, utilisée seule, la fréquence favorise des diagnostics à « grosse mouture », dont la généralité est parfois triviale.

Une telle approche quantitative repose sur l'hypothèse que les phénomènes de récurrence ont une pertinence. Une telle hypothèse constitue précisément le fondement des premières techniques d'analyse de contenu. Dès le début du vingtième siècle, ces techniques furent développées – notamment par Harold Lasswell – pour analyser la propagande dans la presse.

Outre ce postulat quantitatif, ces outils s'inscrivent dans une épistémologie essentiellement descriptive (c'est pour cette raison qu'ils se retrouvent dans le haut de la Figure 1). La lexicométrie se concentre en effet sur l'explicitation de propriétés « matérielles » du texte. L'interprétation de ces propriétés n'est pas prise en charge par le logiciel. Pour cette raison, la lexicométrie s'insère harmonieusement dans le courant dit de l'analyse de discours. Reposant sur les recherches allant de la linguistique de l'énonciation d'Émile Benveniste à la pragmatique du discours d'un Oswald Ducrot en passant par les travaux de Dominique Maingueneau et de Catherine Kerbrat-Orecchioni, l'analyse de discours postule que le discours comporte les traces des conditions (sociales) de son élaboration (en particulier, via l'engagement du locuteur dans son discours). Découlant en partie d'une critique de l'analyse de contenu, l'analyse de discours se veut essentiellement critique, tout en gardant un rapport de proximité avec les textes. L'effacement pur et simple de toute interprétation constitue la limite d'une telle démarche. Si elle constitue un apport à la stylistique, la simple compilation des propriétés matérielles d'un corpus de textes n'apporte par contre guère de valeur ajoutée à une analyse qualitative menée en sciences sociales.

Certains praticiens de l'analyse de discours tirent élégamment parti de l'approche quantitative de la lexicométrie sans tomber dans ses travers. En sciences politiques, Corinne Gobin en tire parti dans son impressionnante étude des acceptations du travail dans les institutions européennes. En sociologie, Jules Duchastel mobilise également de tels outils pour rendre compte des événements de crise.

Concordance

Les concordanciers sont des outils de recherche d'expressions dont les résultats sont présentés de manière singulière : tous les segments de textes sont affichés les uns en dessous des autres, verticalement centrés sur le mot ou l'expression recherché. Un tel alignement assiste bien le parcours du corpus. Par contre, l'articulation de cette lecture avec des éléments analytiques du chercheur n'est pas supportée par l'outil. Ce réagencement des textes ne fait appel à aucun

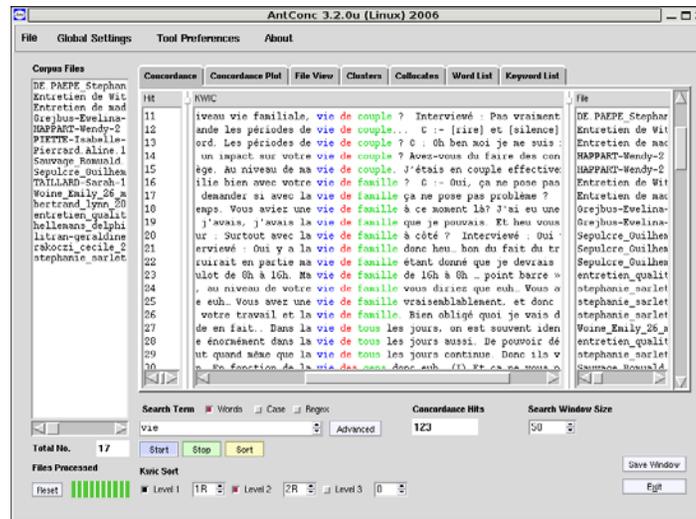


Figure 3. Concordancier (copie d'écran de AntConc)

calcul. Les concordanciers (comme AntConc ou Glossanet) ont donc leur place dans le quadrant supérieur droit de la Figure 1.

L'alignement central des éléments recherchés permet de parcourir transversalement la fenêtre des extraits, comme un index, et de saisir d'un coup d'œil des figures, des expressions récurrentes ou figées (Pincemin, Issac, Chanove & Mathieu-Colas, 2006). Cette lecture transversale constitue précisément une vertu des concordances. Contrairement aux statistiques lexicales, celles-ci donnent accès aux extraits dans leur environnement textuel. Pour cette raison, un grand nombre de logiciels lexicométriques (comme Lexico et Hyperbase, par exemple) complètent leurs analyses lexicales par un concordancier donnant accès aux textes. Comparés aux outils lexicométriques, les concordanciers offrent une plus grande finesse mais un piètre pouvoir de révélation : seuls les éléments explicitement (et délibérément) recherchés par l'utilisateur apparaissent dans les concordances. Là où la lexicométrie fait émerger des récurrences sans l'intervention du chercheur, les concordanciers apparaissent frappés d'un certain mutisme.

Ces caractéristiques trahissent une épistémologie essentiellement exploratoire et qualitative. Ce cadrage provient de l'histoire (relativement ancienne) des concordanciers. Ceux-ci ont en effet émergé, au treizième siècle, des débats entre différentes traditions d'exégèse des textes sacrés. Les Juifs et

les Catholiques rivalisent alors d'ingéniosité dans l'équipement de leur travail sur les textes sacrés et dotent respectivement leurs index d'environnements textuels et de renvois aux textes. C'est dans ce contexte qu'émergent les concordances, simples outils composés alors à la main par des moines afin d'assister le commentaire de textes. Avec la mécanographie puis l'informatique, le vingtième siècle se charge d'automatiser la création des concordances et de les appliquer aux textes littéraires. À l'opposé d'une épistémologie quantitative, l'outil repose essentiellement sur une modification de l'agencement initial du texte, offrant ainsi des options de visualisation et des parcours inédits. Aujourd'hui, les concordanciers sont essentiellement mobilisés dans des travaux littéraires, voire linguistiques. Dans ce numéro de *Recherches qualitatives*, des contributions émanant de ce champ les mobilisent effectivement, comme celle de Virginie André, Christophe Benzitoun, Emmanuelle Canut, Jeanne-Marie Debaisieux, Bertrand Gaiffe et Evelyne Jacquy ainsi que celle de Nicolas Mazziotta. Ces contributions montent bien l'intérêt d'une telle technique pour l'analyse qualitative des transcriptions d'entretiens. Cependant aucune recherche en sciences sociales n'a, à notre connaissance, fait appel à ce type d'instrument.

Automates

Outre leur utilisation descriptive en lexicométrie, les calculs statistiques de distributions des *mots* dans un corpus de textes sont également mobilisés dans des outils de catégorisation automatique. En reposant sur les cooccurrences ou sur l'analyse factorielle, il est en effet possible d'automatiser le calcul d'agrégats de portions du corpus et de leur assigner (automatiquement) une étiquette. C'est ce que font respectivement Leximappe, Candide et Réseau-Lu (à partir du calcul des cooccurrences⁵) et Spad-T ou Alceste (sur base de l'analyse des correspondances). Souvent utilisés en France, ces logiciels mobilisent bien des calculs (et se situent donc dans la partie gauche de la Figure 1) et ce, à des fins qui dépassent la description, puisque les agrégats sont proposés par le logiciel comme les premières catégories d'analyse du corpus (on se déplace donc dans le quadrant inférieur gauche de la Figure 1). Il incombe bien entendu ensuite au chercheur d'interpréter les propositions de l'outil.

Avec ces outils automatiques, on retrouve le potentiel de révélation et la vision d'ensemble des statistiques textuelles. Les outils automatiques renforcent cette possibilité de visualisation en représentant les agrégats calculés sous forme d'arborescences ou de réseaux (Figure 4). Ceux-ci fournissent des illustrations appréciées des chercheurs (que l'on retrouve dans leurs publications). Ces vertus ont néanmoins un revers : au mutisme des

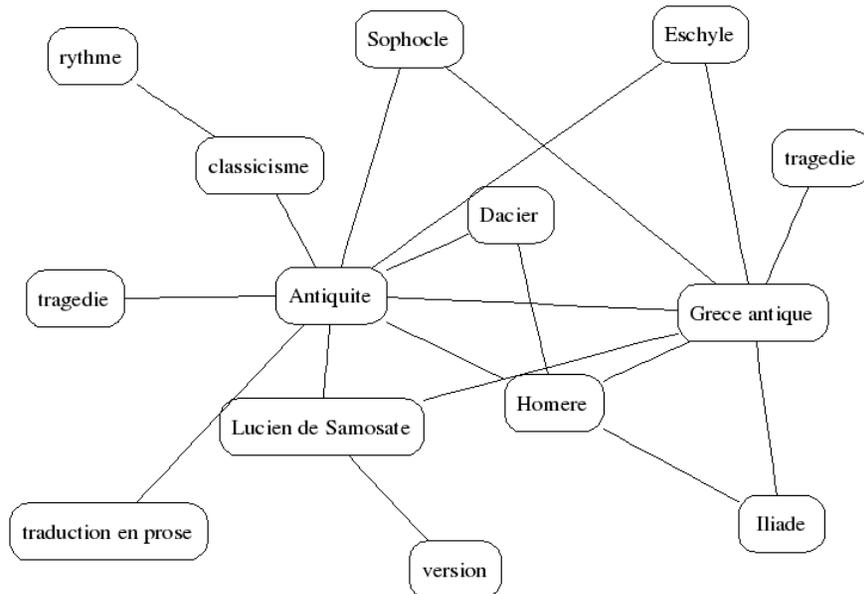


Figure 4. Automate (d'après Candide)

concordances, les automates opposent une volubilité graphique. Non articulée à une connaissance intime du corpus (fondement de toute analyse qualitative), une telle profusion d'illustrations comporte la tentation d'une interprétation non ancrée (et dès lors possiblement fausse). Cette limite est bien connue des concepteurs de tels outils, comme Max Reinert, qui leur assignent avant tout un rôle heuristique (aider la formulation d'hypothèses) et non probatoire.

D'autres limites tiennent à la nature très automatisée des traitements. Les cooccurrences doivent en effet être interprétées avec précaution : la cooccurrence de *maman* et de *amour* ne renseigne aucunement sur le fait que de l'amour était témoigné (ou non) ni sur le fait que la *maman* est le bénéficiaire ou le créancier du sentiment en question. Une telle limitation est connue depuis les premières automatisations informatiques, comme le souligne Krippendorff (2004, p. 326), qui attribue la précédente mise en garde à un texte de Charles Osgood (1959, pp. 73-77), publié dans les années cinquante.

Les opérations de préparation du corpus sont susceptibles d'introduire des artéfacts trompeurs : la lemmatisation – présentée plus haut – d'un corpus mentionnant les transports en commun peut ainsi erronément comptabiliser les occurrences de *bus* comme des variations du verbe *boire*⁶. Il importe également de tenir compte de la métrique particulière de chaque dispositif de

visualisation. Selon la façon dont ces graphes sont calculés, une proximité entre deux points ne traduit pas nécessairement une association pertinente. Dans les graphes factoriels, par exemple, l'angle par rapport au centre du graphe et l'appartenance à un même quadrant sont plus pertinents que la distance entre deux points. Par ailleurs, une thématique centrale se retrouve dans le quadrant supérieur droit d'outils comme Leximappe ou Calliope (on parle de *diagrammes stratégiques*) alors qu'elle se situe au centre du « réseau de mots associés » de Candide. Au niveau épistémologique, les outils automatiques combinent donc les postulats quantitatifs des outils lexicométriques à une valorisation des vertus synthétiques de la représentation graphique.

S'il est conscient des transformations opérées lors des traitements, le chercheur attentif n'a évidemment aucun mal à éviter les pièges évoqués ci-dessus. La contribution de Guillaume Gronier à ce numéro est particulièrement remarquable à cet égard.

Réflexif

À l'opposé de l'automatisation des outils précédemment présentés, les outils réflexifs proposent d'assister le chercheur dans sa lecture et son analyse des textes. Ces outils offrent une simple fenêtre d'annotation, comparable à un traitement de texte voire à une feuille de papier. La fonctionnalité informatique proposée permet au chercheur de surligner en couleur les passages qui l'intéressent et d'y attacher une annotation. Cette annotation apparaît dans la marge ou dans une collection d'annotations pouvant être organisées entre elles (Figure 5). De tels outils assistent bien la création de catégories d'analyse sans déléguer la moindre tâche à la machine. Pour cette raison, ils occupent le quadrant inférieur droit de la Figure 1.

La simplicité de tels outils n'entache pas leur succès puisqu'ils sont les plus utilisés au monde, en recherche qualitative, de la gestion aux sciences de l'éducation en passant par la sociologie (ce qui se traduit également dans ce numéro par le nombre de contributions relatives à ce type d'outils). Désignés par l'acronyme CAQDAS pour « Computer Assisted Qualitative Data Analysis Software », des outils comme NVivo, ATLAS.ti, Weft QDA ou TamsAnalyser reposent essentiellement sur le travail intellectuel du chercheur⁷. Pour reprendre le terme utilisé à propos des concordanciers, ces outils sont donc caractérisés par un certain mutisme. Ce trait, qui peut être vu comme une limite, constitue également un atout : c'est bien le chercheur qui procède à l'analyse. Pour autant, les chercheurs rapportent certains problèmes (Trivelin, 2003, pp. 135-138). La procédure de codage du matériau semble ainsi opérer une attraction pouvant distraire le chercheur de son but (la construction d'une interprétation sur base du matériau).

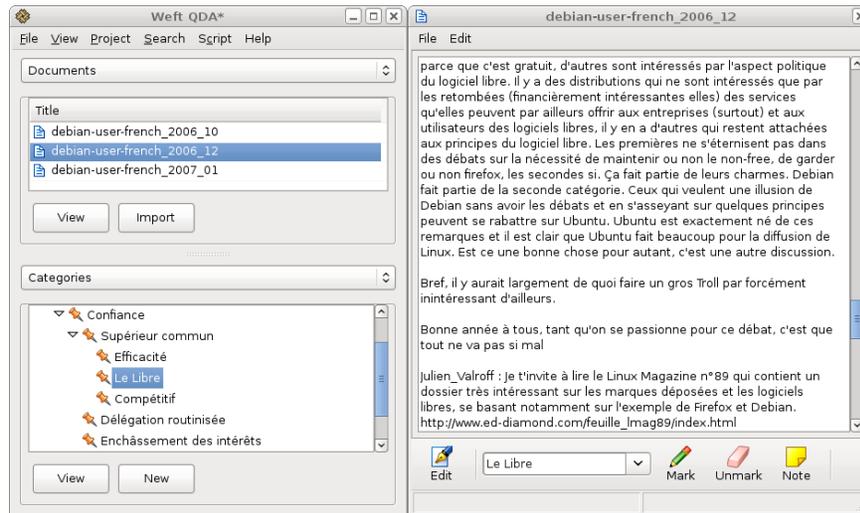


Figure 5. Outil réflexif (copie d'écran de Weft QDA)

Épistémologiquement, ces outils se réclament explicitement de la méthode par théorisation ancrée (ou enracinée), la « Grounded Theory Methodology ». Le nom des fonctionnalités – voire des outils eux-mêmes – empruntent ainsi au répertoire d'une analyse résolument qualitative. De fait, le support informatique assiste l'explicitation des codages. À la manière d'un journal de bord, les « mémos » permettent de conserver la trace des formulations successives de l'analyste. Ces fonctionnalités rencontrent donc les préoccupations résolument qualitatives de la méthode conçue par Barney Glaser et Anselm Strauss (2010).

C'est dans ce cadre que s'inscrivent les contributions de Béatrice Arend, de Udo Kuckartz, de Kerralie Œuvray et de Jacques Hamel. Ce dernier apporte une réflexion épistémologique sur la place de tels outils dans la recherche qualitative. Kerralie Œuvray présente comment le recours à un tel outil s'inscrit explicitement dans une démarche qualitative. En tant qu'auteur de MaxQDA, Udo Kuckartz présente l'apport d'une interface de visualisation des codages de segments de textes; Béatrice Arend apporte une contribution comparable appliquée aux sources vidéographiques.

Paradoxalement, selon certains utilisateurs, les logiciels réflexifs semblent particulièrement adaptés à l'application d'une grille d'analyse pré-existante, plus qu'à l'élaboration continue et progressive de catégories d'analyse. Or, à la différence de l'analyse de contenu, la méthode par théorisation enracinée repose précisément sur la condensation des catégories

d'analyse tout au long du processus. Cela contribue peut-être en partie à ce que la majorité des analyses conduites avec des logiciels réflexifs ne s'inscrivent pas dans ce paradigme. En particulier, de nombreux chercheurs les mobilisent pour quantifier les codages créés par leurs soins, voire pour opérer des croisements permettant par exemple de déterminer qui, des hommes ou des femmes, mobilise le plus tel ou tel argument. Un tel usage s'inscrit bien entendu dans la démarche quantitative typique de l'analyse de contenu. Il détourne donc les outils réflexifs de leur inscription épistémologique originale. Un tel détournement ne soulève bien évidemment aucune difficulté lorsqu'il est opéré consciemment. La situation devient par contre problématique lorsqu'une démarche reposant exclusivement sur la quantification se réclame d'une approche qualitative.

Dictionnaires et registres

Les dictionnaires et les registres constituent une voie intermédiaire entre les automates et les outils réflexifs. Ils se composent d'une série de marqueurs utilisés pour annoter le corpus de textes. L'attribution d'un registre ou d'un dictionnaire à un passage est semi-automatique. D'une part, les marqueurs sont des mots-clés ou des expressions choisis réflexivement par le chercheur pour leur pertinence par rapport à ses investigations. D'autre part, le logiciel se charge d'attribuer automatiquement le registre (ou le dictionnaire) à tous les passages comportant le marqueur en question. Le logiciel assiste donc bien la construction des catégories (partie inférieure de la Figure 1) dans une stratégie qui combine réflexivité et automatisme.

Le moment du choix des marqueurs détermine la différence entre un dictionnaire et un registre. Les dictionnaires préexistent à l'analyse. Ils peuvent être définis une fois pour toute, de manière totalement indépendante de l'analyse. Ils peuvent également être définis pour une recherche particulière, mais toujours avant que l'analyse ne débute : c'est le cas lorsque les catégories découlent de propositions théoriques qu'il s'agit de confronter au terrain (dans une optique hypothético-déductive). Les registres, eux, se construisent (et se révisent) tout au long du processus analytique. Le chercheur les élabore dans une démarche abductive, à partir des textes.

Comme l'analyse lexicale, dictionnaires et registres facilitent une lecture transversale. Comme les outils réflexifs, ils autorisent en outre une analyse tenant compte des objectifs d'une recherche et des spécificités d'un corpus déterminé. Procédant à une annotation par mots-clés, ils sont néanmoins sujets aux mêmes limitations que les procédures automatiques face aux phénomènes d'homographie et restent démunis face à des phénomènes non marqués, contrairement aux outils réflexifs.



Figure 6. Registre (copie d'écran de Porphyry / Cassandre)

Au niveau méthodologique, le recours aux dictionnaires rappelle la division du travail scientifique pratiqué en analyse de contenu. Le chercheur définit les catégories d'analyse, alors que l'attribution de ces catégories à des passages de textes est confiée à différents codeurs. Afin de limiter l'expression de la subjectivité des codeurs, le chercheur définit une liste d'indicateurs explicites pour chaque catégorie. Une fois effectué le codage des textes, la variabilité inter-individuelle est calculée au moyen d'un indice de fiabilité inter-codeur. Celui-ci permet d'écarter les catégories ambiguës et de ne conserver que les codages consensuels, entendus comme plus « objectifs ». L'informatisation de la procédure de codage offre une sous-traitance nécessairement dénuée de subjectivité. N'exécutant strictement que ce qui leur est explicitement demandé, les programmes informatiques correspondent en définitive mieux à la fonction de codeur tel que l'analyse de contenu la conçoit. C'est sans doute pour cette raison que le premier logiciel d'analyse textuelle – General Inquirer – a été développé pour assister cette technique, dès la fin des années cinquante. Cet outil a été mis à profit par une large communauté de chercheurs depuis son concepteur, le psychologue Philip Stone, jusqu'à l'ethnométhodologue George Psathas.

Le recours aux registres témoigne une conception diamétralement opposée de la subjectivité du codeur. Il interdit en effet la séparation du travail de conceptualisation et de codage. Résolument qualitative, la démarche

implique que les catégories d'analyse se définissent progressivement, par condensation conceptuelle des extraits rassemblés par le chercheur. La définition des catégories d'analyse – donc la conceptualisation – s'élabore parallèlement au codage. La subjectivité du chercheur – découlant de sa culture scientifique – s'entend cette fois comme source de richesse pour l'analyse, et non plus comme source de biais (Bénel & Lejeune, 2009, pp. m8-9).

Bien que les dictionnaires et les registres présentent des différences considérables dans la façon de conduire l'analyse, très peu de logiciels proposent une architecture informatique dédiée à la construction de registres. Si des outils comme Tropes et Prospéro se proposent d'assister la construction de registres, leur architecture logicielle les inscrit plutôt dans l'« application » de dictionnaires prédéfinis : ils sont d'ailleurs tous deux livrés avec des listes de dictionnaires prédéfinis, que le chercheur est invité à amender en fonction de ses recherches. Jean-Pierre Hiernaux montre, dans sa contribution, comment il a élaboré une technique par registres assistant une analyse structurale de contenu. Pour notre part, nous avons proposé un logiciel libre d'analyse de textes dédié à la construction de registres, du nom de Cassandra (Lejeune, 2008a).

Conclusion

Un grand nombre de logiciels sont aujourd'hui disponibles pour assister l'analyse de textes afin d'en rendre explicites, comparables et discutables les différentes étapes. Ce faisant, ces logiciels peuvent contribuer à la scientificité d'une démarche qualitative. Pour autant, les procédures proposées sont très variées : tantôt elles assistent uniquement l'exploration du matériau, tantôt elles s'étendent à l'élaboration des catégories d'analyse. Certaines reposent sur un haut degré d'automatisme là où d'autres délèguent au chercheur le choix des agencements susceptibles de faire avancer sa recherche. Chaque stratégie apporte un éclairage original, mais recèle également ses limites. Il importe par conséquent de multiplier les points d'entrée et les voies d'exploration.

Certains outils logiciels offrent dès lors aux chercheurs un véritable laboratoire, tout équipé. Ces logiciels, comme MaxQDA, la suite Provalis ou T-LAB, combinent des outils permettant de mettre en œuvre plusieurs stratégies à la fois ou parallèlement. Un concordancier permet, par exemple, d'accéder au contexte des récurrences mises en évidence de manière statistique. Les annotations réflexives ou par dictionnaires sont intégrées à des calculs automatisés. En outre, certains outils offrent des fonctionnalités spécifiques et originales.

Si elle ouvre des voies originales (et parfois productives), la multiplication des fonctionnalités ne garantit en aucune manière la qualité des

analyses conduites. Au contraire, de nombreux praticiens n'utilisent qu'un nombre limité d'outils, notamment parce qu'une trop grande complexité entraverait la communication des résultats. On l'aura compris, il n'existe pas d'outil ultime, loin s'en faut. Mais cette conclusion ne constitue pas un aveu d'échec des outils informatiques. Elle rappelle simplement qu'aucun dispositif méthodologique ne peut s'affranchir d'une réflexion (et d'une vigilance) épistémologique... ce qui, en définitive, est heureux.

Notes

¹ L'auteur tient à remercier les participants au séminaire sur les ressources informatiques pour l'analyse qualitative et au colloque *AnaLogiQual*, en particulier, Françoise Mulkay et Jean-Pierre Hiernaux dont les commentaires ont contribué à la maturation de ce texte.

² Lejeune (2004a) évalue cet outil dans le cadre d'une analyse sociologique de forums de discussion.

³ Lejeune (2007) propose une histoire de ces fonctionnalités.

⁴ La notion de *mot* est l'objet de nombreux débats en sciences du langage. Nous conservons néanmoins ce terme compréhensible par un large lectorat.

⁵ L'anthropologie des sciences et des techniques a vu dans un tel recours aux cooccurrences une métaphore du concept de réseau qui lui est cher (Lejeune, 2004b).

⁶ Cette confusion entre le substantif *bus* et le participe passé masculin pluriel du verbe *boire* est un exemple du problème soulevé par les *homographes*.

⁷ Aurélien Bénél a attiré notre attention sur le fait que cette opération mérite d'être qualifiée d'*intellectuelle* – plutôt que de *manuelle*. C'est pour cette raison que nous qualifions ces outils de réflexifs.

Références

- Bandeira-de-Mello, R., & Garreau, L. (à paraître). L'utilisation d'ATLAS.ti pour améliorer les recherches dans le cadre de la théorie enracinée : panacée ou mirage? *Recherches qualitatives*.
- Barry, C.A. (1998). Choosing qualitative data analysis software : ATLAS.ti and Nudist compared. *Sociological Research*, 3(3). En ligne : <http://www.socresonline.org.uk/3/3/4.html>
- Bénél, A., & Lejeune, C. (2009). Partager des corpus et leurs analyses à l'heure du Web 2.0. *Degrés : revue de synthèse à orientation sémiologique*, 36-37(136-137), m1-20.
- Glaser, B.G., & Strauss, A.L. (2010). *La découverte de la théorie ancrée : stratégies pour la recherche qualitative*. Paris : Armand Colin.

- Gobin, C., & Deroubaix, J.-C. (1994). *Quand la commission se présente devant le parlement* [rapport de recherche]. Bruxelles : Recherche européenne en Sciences Humaines.
- Jenny, J. (1996). Analyse de contenu et de discours dans la recherche sociologique française : pratiques micro-informatiques actuelles et potentielles. *Current Sociology*, 44(3), 279-290.
- Klein, H. (2001). Overview of text analysis software. *Bulletin de méthodologie sociologique*, 70, 53-66.
- Krippendorff, K. (2004). *Content analysis. An introduction to its methodology*. Thousand Oaks : Sage.
- Latour, B. (1995). *La science en action*. Paris : Gallimard.
- Lejeune, C. (2004a). L'analisi sociologica di un corpus derivato da fori di discussione. Arricchimento reciproco delle analisi della conversazione, dei testi e di corpus. Dans F. Lancia (Éd.), *Strumenti per l'analisi dei testi. Introduzione all'uso di T-LAB* (pp. 104-106). Milano : FrancoAngeli.
- Lejeune, C. (2004b). Représentations des réseaux de mots associés. Dans G. Purnelle, C. Fairon, & A. Dister (Éds), *Le pouvoir des mots. Actes des 7^{es} Journées internationales d'analyse statistique des données textuelles* (pp. 726-736). Louvain : PUL.
- Lejeune, C. (2007). Petite histoire des ressources logicielles au service de la sociologie qualitative. Dans C. Brossaud, & B. Reber (Éds), *Humanités numériques. Nouvelles technologies cognitives et concepts des sciences sociales* (pp. 197-214). Paris : Hermès.
- Lejeune, C. (2008a). Au fil de l'interprétation. L'apport des registres aux logiciels d'analyse qualitative. *Revue suisse de sociologie*, 34(3), 593-603.
- Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum : Qualitative Social Research*, 1(2). En ligne : <http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204>
- Merton, R.K. (1953). *Éléments de méthode sociologique*. Paris : Plon.
- Osgood, C.E. (1959). The representational model and relevant research methods. Dans I. De Sola Pool (Éd.), *Trends in content analysis* (pp. 33-88). Urbana : University of Illinois Press.
- Paillé, P. (2006). Lumières et flammes autour de ma petite histoire de la recherche qualitative. *Recherches qualitatives*, 26(1), 139-153.

- Pincemin, B., Issac, F., Chanove, M., & Mathieu-Colas, M. (2006). Concordanciers : thème et variations. Dans J.-M. Viprey, A. Lelu, C. Condé, & M. Silberstein (Éds), *Actes des 8èmes Journées internationales d'analyse statistique des données textuelles* (pp. 773-784). Besançon : Presses Universitaires de Franche-Comté.
- Popping, R. (1997). Computer programs for the analysis of texts and transcripts. Dans C.W. Roberts (Éd.), *Text analysis for the social sciences. Methods for drawing statistical inferences from texts and transcripts* (pp. 209-221). New Jersey : Lawrence Erlbaum.
- Trivelin, B. (2003). Une aide à l'analyse de contenu : le tableau Excel. *Recherches sociologiques, 1*, 135-147.
- Weitzman, E., & Miles, M. (1995). *A software source book. Computer programs for qualitative data analysis*. London : Sage.

Christophe Lejeune est expert scientifique à l'Université de Liège. Docteur en sociologie, il s'intéresse particulièrement aux collectifs médiatisés. Il a notamment développé une analyse ethnométhodologique de la confiance sur Internet ainsi qu'une comparaison de la régulation interne de plusieurs collectifs actifs au sein du mouvement pour le logiciel libre. Il est par ailleurs l'auteur du logiciel libre d'analyse qualitative Cassandra.

Langage de la crise ou crise du langage?

Jean-Marie Jacques, Docteur en biologie

Université Notre-Dame de la Paix

Nathanaël Laurent, Docteur en sciences biomédicales

Université Notre-Dame de la Paix

Anne Wallemacq, Docteur en sociologie

Université Notre-Dame de la Paix

Résumé

Le papier porte sur l'étude au moyen du logiciel Evoq du langage de crises lors d'un accident nucléaire (simulé). La communication de crise joue un rôle central dans le processus d'évolution de la crise (aggravation ou résolution). Si une communication claire est souhaitée dans ces conditions, il est cependant évident que les gestionnaires de crise contribuent à rendre confuse la communication de par leur appartenance à des disciplines et à de organisations différente (chacune d'elles ayant une culture distincte). En utilisant une méthodologie qualitative basée sur l'analyse structurale du langage et qui permet de pointer la polysémie, nous avons isolé quelques phénomènes communicationnels qui permettent de mieux comprendre pourquoi le langage peut en même temps être source de conflit et de créativité durant la gestion de la crise.

Mots clés

RECHERCHE QUALITATIVE, CRISES ET CONFLIT, COMMUNICATION, LANGAGE NATUREL, CRÉATIVITÉ, ANALYSE STRUCTURALE

... il suffit que nous nous prêtons à sa vie,
(...)à sa gesticulation éloquente. (Merleau-
Ponty, 1960, p. 66)

Introduction

« La crise est une rançon de la gloire. Plus on communique, plus on s'expose. »
(Candellier, 2004).

« Plus on communique, plus on s'expose », cette phrase retiendra toute
notre attention dans ce texte qui tentera d'éclairer d'une manière nouvelle les

liens entre crise et communication. Pour ce faire, il faut nous tout d'abord définir ou redéfinir le concept de crise. La définition de la crise contrairement (à la catastrophe) a toujours été hésitante et ambiguë. Il n'existe pas de définition ni de théorie unifiée de ce concept. Lalonde (2004) dans « In search of archetypes in crisis management » illustre parfaitement cette ambiguïté :

Crisis is simultaneously characterized by negative effects (perturbation, deregulation, conflict, confusion of action, excessive stress leading to action which is hasty or poorly thought out) to positive effects (mobilisation, solidarity, co-operation, improved adaptation to the environment, experimental learning) (p. 7).

De même, on peut rappeler le sens double et ambivalent de l'idéogramme chinois du mot crise : opportunité et danger. Cette ambiguïté se marque aussi dans le fait que la crise, selon Jacques, Peris Mora et Gatot (1997) peut être considérée soit comme un événement, soit comme un processus suivant le cas. Comme processus, la crise peut être vue comme relevant du champ de la perception (Jacques & Gatot, 1998).

S'attendre à l'inattendu, prévoir l'imprévisible

La gestion de crise s'est surtout développée en catégorisant la crise en fonction de sa nature (industrielle, naturelle, alimentaire, politique, etc.) et, encore davantage, en fonction de l'événement déclencheur. Cette focalisation sur l'élément déclencheur a conduit à un paradoxe, assez paralysant pour l'action : comment s'attendre à l'inattendu? Ainsi par exemple, paradoxale est la recommandation de Candellier (2004) :

La crise, qui arrive toujours à un moment inattendu, se gagne avant son émergence. Il faut l'anticiper par le recensement de toutes les crises potentielles, l'évaluation de leurs probabilités et de leurs effets, la mise en place d'une cellule de crise, la désignation d'un porte-parole, la réalisation d'outils de communication spécifiques, de simulations...

En réduisant ainsi la gestion des crises à un problème de planification (on peut prévoir de prévoir), plusieurs erreurs peuvent apparaître :

- on oublie d'envisager la possibilité d'une forme plus radicale de crise qui est non seulement inattendue mais aussi imprévisible;
- on joue seulement sur les procédures qui peuvent préparer les acteurs à faire face à la crise, avec comme conséquence qu'une telle approche procédurale de la crise risque de normaliser et ainsi concentrer les perceptions sur des attributs de crises connues.

Mais caractériser la crise par l'élément déclencheur a une autre conséquence : elle peut conduire à une certaine méconnaissance du processus même de la crise. C'est ce processus que nous nous attachons, quant à nous, à saisir. Dès lors, nous voyons la crise comme une forme, peut-être brutale, d'un processus beaucoup plus général de définition et redéfinition de soi et du monde. Au lieu de définir la crise à partir de la nature de l'élément déclencheur ou de facteurs environnants nous avons essayé de caractériser le processus dynamique à partir de sa propre nature : la crise est pour nous changement et énonciation du paradoxe (double-bind) de définir notre propre identité (une personne ou une organisation) en perdant notre représentation antérieure de nous-mêmes.

Crise et langage

Nous pensons que l'analyse du discours peut nous aider à mieux comprendre les processus de crise et qu'inversement, cette étude du processus de la crise peut nous apporter de nouvelles connaissances sur les propriétés du langage naturel. Crise et langage naturel sont, au minimum, deux processus homologues : tous deux sont des processus complexes qui peuvent être des sources de créativité permettant de vivre et d'évoluer dans un monde imprévisible. Mais on peut aller plus loin dans la relation entre crise et langage : la crise est inhérente à la manière propre qu'a l'homme de communiquer avec ses semblables, et cela quelle que soit la forme de communication. De notre point de vue, même la forme de communication la plus simple, routinière et ancienne qu'est le langage courant (« langage in use ») présente une complexité telle, qu'elle suscite sans arrêt la crise. Inversement, une forme de communication très formalisée, même si elle permet de réduire l'anxiété, ne réduit jamais à rien la marge d'incertitude liée à la définition même de la réalité.

Notre hypothèse est que c'est de la nature même du langage que dépend ce que nous appelons la crise : l'indexicalité du langage naturel conduit à la permanente redéfinition de ce que les choses sont, et plus fondamentalement encore, à la permanente définition et redéfinition du réel.

Analyser la communication en situation de crise nous paraît dès lors particulièrement intéressant pour étudier le fonctionnement du discours. D'une part, parce que l'échange d'informations prend ici une valeur de survie : de lui peut dépendre la vie d'une, de plusieurs, voire de nombreuses personnes. D'autre part, parce que l'on peut analyser dans le détail le processus même qui conduit à *nommer* : ici à dire « c'est une crise », et comprendre la vertu salvatrice de ce recouvrement de la capacité de nommer vis à vis d'une situation où l'évidence du monde tel que défini habituellement se dissout. Dire

« c'est une crise » permet de sortir de l'anomie (Wallemacq, 2001) d'une manière similaire à celle d'Alice au pays des merveilles lorsque celle-ci, perdue dans le bois des choses sans nom découvre que ce bois a un nom : « le bois des choses sans nom ». Un tel processus a déjà été étudié dans le modèle de facteurs aggravants (Jacques & Gatot, 1998).

Analyse du discours : les axes de notre positionnement

La relation entre le discours et l'organisation a été abordée par différents champs théoriques. La première question consiste à déterminer le point de départ de l'analyse : l'organisation est-elle une structure préexistante au langage dans laquelle le discours est produit. Ou, à l'inverse, l'organisation n'est-elle finalement formée que dans le discours qui la définit comme objet? Le discours définit-il ou produit-il l'organisation? La seconde question est celle du type de discours visé. De quel discours s'agit-il? Est-ce, par exemple, le discours correspondant à la langue en usage formé de conversations ou plutôt, à la Foucault, le discours archétypique d'un système de pensée situé historiquement (Alvesson & Kärreman, 2000)? Est-ce le discours dominé, dominant, officiel, etc.? Fairhurst et Putnam (2004) ont proposé une synthèse intéressante des principaux courants contemporains qui organisent les champs scientifiques « organisational discourse ». Les trois principaux courants considérés sont :

- centrés sur la structure : du haut vers le bas, l'organisation comme une entité ou un objet pré - existant au discours produit en son sein;
- centrés sur l'énonciateur : de bas en haut, le discours comme processus vécu d'où l'organisation émerge;
- centrés sur l'action : à partir de l'intérieur de l'organisation, ancré dans les pratiques sociales, le discours forme l'organisation qui le forme.

Le discours, on le voit, peut avoir des statuts très différents dans l'analyse : la réification de l'organisation, l'acculturation des membres, et l'ancrage de ces deux éléments dans les pratiques sociales. De surcroît, la nature même du langage et des processus de communication se trouve à la croisée de traditions différentes dans lesquelles on peut distinguer, pour faire bref :

- la tradition (logico - mathématique et positiviste) qui considère le langage comme un véhicule qui transmet de manière efficace des informations quand un maximum de codification et désambiguïsation a eu lieu;
- la tradition phénoménologique, et plus particulièrement ethnométhodologique, qui traite la langue naturelle comme telle, considérant d'emblée sa complexité naturelle et l'ambiguïté comme une condition préalable à la signification.

La perspective que nous adoptons ici est immédiatement issue de l'ethnométhodologie pour qui le réel est moins un fait qu'un processus (*accomplishment*). Dès lors, cette redéfinition permanente fait l'objet de jeux de pouvoir. Ceux-ci portent sur la définition de « ce qui est », sur la définition même de la situation bien plus que, comme on les envisage habituellement, sur ce qui se passe à l'intérieur d'une situation bien définie (Heritage, 1984) :

(...) instead of beginning from the assumption that the terms of a language invoked a fixed domain of substantive content and that their intelligibility and meaning rest upon a shared agreement between speakers as to what this content consists of, Garfinkel proposed an alternative procedural version of how descriptions works. In this alternative version, he argues that the intelligibility of what is said rests upon the hearer's ability to make out what is meant from what is said according to methods which are tacitly relied on both speaker and hearer. These methods involve the continual invocation of common-sense knowledge and of context as resources with which to make definite sense of indefinite descriptive terms. (p. 144).

Dès lors « la crise » est autant un processus de définition (« il y a crise ») qu'une situation concrète (ce qui nous situe dans le troisième courant de Fairhurst et Putnam (2004)) et on sait combien la définition même de la situation de crise comme crise est cruciale pour la gestion future de la situation.

Le caractère ouvert du sens : indexicalité, réflexivité, code

Dans le domaine de l'analyse du discours, la nature polysémique, voire ambiguë des mots et des expressions est rarement considérée comme condition de possibilité de la communication langagière. Le fait que les mots puissent revêtir des significations différentes selon leur contexte d'usage (grammatical, syntaxique, sémantique, performatif) explique sans doute qu'ils permettent de s'adapter à l'immense variété des situations communicatives dans lesquelles ils sont amenés à faire sens et, dans certains cas, conduire à l'action. Or, si le mot signifie avec le contexte linguistique particulier dans lequel il prend place, et dans la situation originale qui l'accueille, comment expliquer que des locuteurs différents puissent comprendre son sens? Si une telle indétermination au niveau du contenu informationnel langagier est une condition de possibilité de son utilisation dans des situations toujours nouvelles, et de son pouvoir créateur de sens, comment expliquer néanmoins que les locuteurs choisissent tels ou tels mots et expressions, et qu'ils puissent être compris correctement par leurs interlocuteurs? C'est pourtant la perspective qu'adoptent d'emblée les ethnométhodologues (Wallemacq, 2001) :

Une des caractéristiques principales du langage, selon les ethnométhodologues, est son indexicalité : un mot n'a pas une signification définie. Il prend sa signification dans le contexte même de son énonciation. Mais (...) le contexte lui-même ne peut être défini intrinsèquement. La conséquence est que la signification n'est pas un état défini. C'est un processus perpétuel, de nature intersubjective. Loin d'être une caractéristique dont le langage devrait être purifié, l'indexicalité est, pour les ethnométhodologues, une ressource utilisée par l'individu en situation pour définir un sens parmi une infinité d'interprétations possibles. L'interprétation est ouverte et c'est le système des présuppositions réciproques – la confiance que nous avons dans le fait que l'autre va interpréter de la même manière que moi – qui crée la signification (p. 65).

Ces présuppositions réciproques forment ce qu'ils vont appeler le code. Le code n'est jamais défini une fois pour toutes, il est serti, enchâssé dans l'interprétation réciproque même. « C'est cette réciprocité (...) qui garantit la stabilité et l'indubitable facticité de la définition des choses » (Wallemacq, 2001, p. 65). C'est sans doute sur base de considérations de ce genre que des chercheurs ont été amenés à remettre en question l'usage de certaines catégories linguistiques devenues inopérantes dans l'analyse de la langue, telle que l'entité que représentent le mot et sa signification. Ainsi, pour Ploux et ses collaborateurs qui s'intéressent de près au développement d'outils de traduction automatique, l'unité de signification (« unit of meaning ») devient un ensemble de termes reliés entre eux par la synonymie ou la co-occurrence (Ploux & Ji, 2003). Nous nous intéressons pour notre part à des unités de sens davantage inscrites dans la singularité de la relation intersubjective vécue au moment de la communication. La situation avec laquelle les termes font sens ne se limite pas aux relations grammaticales structurant le discours, ni aux relations sémantiques passées dont les termes gardent certaines traces (synonymie, antinomie, « contonymie », etc.), mais englobe plutôt le processus d'émergence de sens qui survient dans la conversation présente entre des locuteurs différents. Une telle création de sens est dès lors autant dépendante des termes convoqués par les acteurs de la communication, que de la structure nouvelle qu'ils co-instituent et qui devient soumise à l'épreuve de la compréhension mutuelle. Dans un tel schéma, les unités de sens sont non seulement formées par des formes langagières polysémiques composées de mots et d'expressions porteurs des traces de leurs usages passés (sens concret, objectivable), mais elles sont également forgées dans la situation présente originale dépendante des états mentaux des différents locuteurs et de leur

environnement (sens abstrait, subjectif). L'unité de sens prend donc ici deux aspects (Varela, 1998) : l'un concret et performatif (porteur des promesses d'action) faisant référence à l'usage passé (« conventionnel ») des termes d'une langue et permettant de prévoir et d'anticiper significations et actions (cf. notamment les modèles informatiques de Ploux et son équipe); l'autre abstrait et métaphorique étant inscrit dans la nouveauté et l'imprévisibilité de toute situation présente, dépendant à la fois des capacités cognitives originales des locuteurs et des propriétés polysémiques du langage naturel.

Espaces sémantiques et analyse de conversation

Le processus de création d'une réalité intersubjective dans le discours en temps réel a été analysé en détail par « l'analyse de conversation » développée par une certaine branche de l'ethnométhodologie. Pour notre part, nous avons privilégié une autre approche, directement inspirée du post-structuralisme et d'une relecture phénoménologique du structuralisme. Nous avons détaillé ailleurs l'articulation précise de ces différents courants (Wallemacq & Jacques, 2003). Sans rentrer dans les détails, nous avons utilisé ici un logiciel : Evoq¹ (Wallemacq & Jacques, 2003). Celui-ci est basé sur les principes suivants :

- Nous cherchons à retracer les champs sémantiques, c'est-à-dire les systèmes de différence (association/opposition) qui organisent le discours d'un individu.
- Ces systèmes de différences impliquent qu'un terme est toujours considéré dans ses rapports avec les autres termes et, au minimum avec son opposé possible, éventuellement absent, différé. L'unité de base de l'analyse n'est donc pas un terme mais, au minimum un couple de termes, ce que nous appelons une disjonction.
- La différence n'est pas statique mais dynamique et au lieu de parler de différence il y aurait lieu, de parler de différenciation (ou de *différance* pour suivre Derrida). Il y a toujours une relations de pouvoir entre les termes opposés, le terme choisi « différant », rejetant son ombre, mais l'ombre venant toujours défier le terme dominant. On n'assiste donc pas à une coexistence paisible des paires d'opposés mais au contraire à une conflit violent entre les termes ainsi séparés (Cooper, 1989, p. 483).
- Ces systèmes de « *différance* » ne sont pas seulement dans la main de l'individu, ils constituent son environnement, ce avec quoi il écrit son texte et ce sur quoi il l'écrit. Les champs sémantiques ne sont pas devant l'individu, ils sont autour de lui et l'individu y est immergé.
- Le sens de ce qu'il dit n'est pas entièrement contenu dans son intention mais peut venir au rebond, des termes différés mais qui constamment viennent menacer les termes dominants. Ces termes différés

appartiennent autant au champ sémantique dans lequel il s'inscrit qu'à son discours propre.

- Ces systèmes de différence invitent à un raisonnement « horizontal », à surfer à la surface des mots, à voir comment ils se renvoient l'un à l'autre plus qu'à un raisonnement vertical où le mot est dans un simple rapport de désignation avec les choses.
- Ces champs sémantiques organisent le monde tel qu'il est perçu par les individus qui s'y meuvent. Ils organisent l'expérience (et non l'inverse).

Le logiciel produit des cartes sémantiques dessinant un système de relations sémantiques disjonctives et conjonctives. Grâce à cet outil d'aide à la représentation, les mots révèlent leur sens à l'intérieur d'un réseau sémantique relationnel, qui trace les contours d'un espace intersubjectif. Les relations sont toujours inédites et pourtant toujours ancrées dans un univers sémantique, qui structure un discours donné.

Nous allons à présent passer à l'analyse d'un discours en situation de crise. L'analyse par cartographie sémantique va nous permettre de mettre en lumière les racines d'un conflit dans la structure sémantique ambiguë du langage naturel, et de comprendre comment ce dernier est surmonté. Nous verrons comment le processus même de crise implique une crise du code.

Contexte, méthodologie, données

Les données empiriques (enregistrements audio et vidéo) ont été collectées durant un exercice de simulation d'accident nucléaire à Tours en France. Cette collecte de données s'inscrit dans un projet de recherche CNRS regroupant trois centres de recherche : l'École des mines de Paris, l'IRIT-UMR 5505 (Université Paul Sabattier – Toulouse), et le ReCCCoM – Crecis), (Louvain School of Management, Université de Namur). L'exercice porte sur un transport (fictif) de déchets radioactifs par voie ferrée. Le jeudi 30 septembre, un convoi composé d'une motrice et d'un wagon transportant un colis radioactif circule sur une voie unique menant vers le site de la gare de triage de Saint Pierre des Corps (situé à environ 5 Km de Tours en France). A 7h, à l'intersection de la voie SNCF et de la rue de la Vicairerie, un camion citerne transportant des hydrocarbures roulant à très vive allure percute le wagon. Les barrières du passage à niveau avaient été normalement abaissées. Le chauffeur du camion venait d'être victime d'une crise cardiaque. La cabine du camion et son chargement prennent feu immédiatement. Le passager du camion est blessé mais il parvient à s'extraire de la cabine avec le chauffeur inconscient; ils se mettent à l'abri mais ils ont inhalé des fumées. Sous l'impact du choc, le wagon a déraillé mais n'a pas basculé. L'attelage est rompu. Sous l'effet de l'incendie du camion, le wagon et le colis sont pris progressivement dans

l'incendie. Le camion a par ailleurs enfoncé un côté du colis (le canopis, les ailettes et le support du colis sont déformés). Le conducteur de la motrice dételle sa motrice, évacue son engin de la zone de l'accident, et donne l'alerte auprès du cadre opérationnel du triage de Saint-Pierre des Corps. Dix minutes après l'accident, un témoin parvenu sur les lieux de l'accident alerte les sapeurs pompiers en composant le 18 à partir de son portable. Suite à une coupure de batterie, son appel est bref. Il indique :

- le lieu de l'accident;
- un important incendie impliquant un wagon.

Les conditions météorologiques sont clémentes : temps clair, température de 18 à 23°C, vent de 3 à 5 km / h.

L'analyse porte sur certains échantillons issus des données d'enregistrement vidéo et audio et que nous avons réalisé avec l'autorisation des acteurs. Les échantillons qui ont été retenus sont ceux qui montrent la présence d'un conflit entre différents acteurs impliqués. Nous avons analysé en profondeur les échanges entre les acteurs en présence, ces acteurs appartiennent au poste de commande opérationnel (PCO) localisé dans la préfecture. Nous avons choisi deux extraits (mentionnés extrait 1 et 2). Dans l'extrait 1, cinq personnes sont présentes, le coordinateur de crise et quatre experts techniques (Figure 1).

Les échanges entre ces différents acteurs (durée totale 8 minutes 26 secondes) ont été retranscrits intégralement. Le conflit qui s'y produit est déclenché par le coordinateur de crise qui manifeste son désaccord suite à des propos tenus par les experts. La Tableau 1 ci-dessous reprend la retranscription du passage en question, nous y avons identifié trois phases reprises :

- [phase 1] la discussion sur l'état des lieux de la situation suite à l'accident ferroviaire (le discours de l'expert 1 prédomine);
- [phase 2] l'émergence d'un conflit suite au désaccord exprimé par le coordinateur;
- [phase 3] le traitement du conflit essentiellement suite à l'intervention de l'expert 3.

Langage de crise ou crise du langage

Un premier espace

Nous avons dans un premier temps analysé à l'aide du logiciel de cartographie sémantique Evoq© la première phase de l'extrait (1) considéré. Nous nous sommes penchés sur le discours de l'expert 1 qui y est prédominant. La Figure 2 présente la carte sémantique obtenue, avec le système composé des disjonctions et des conjonctions qui structurent le discours. En lisant

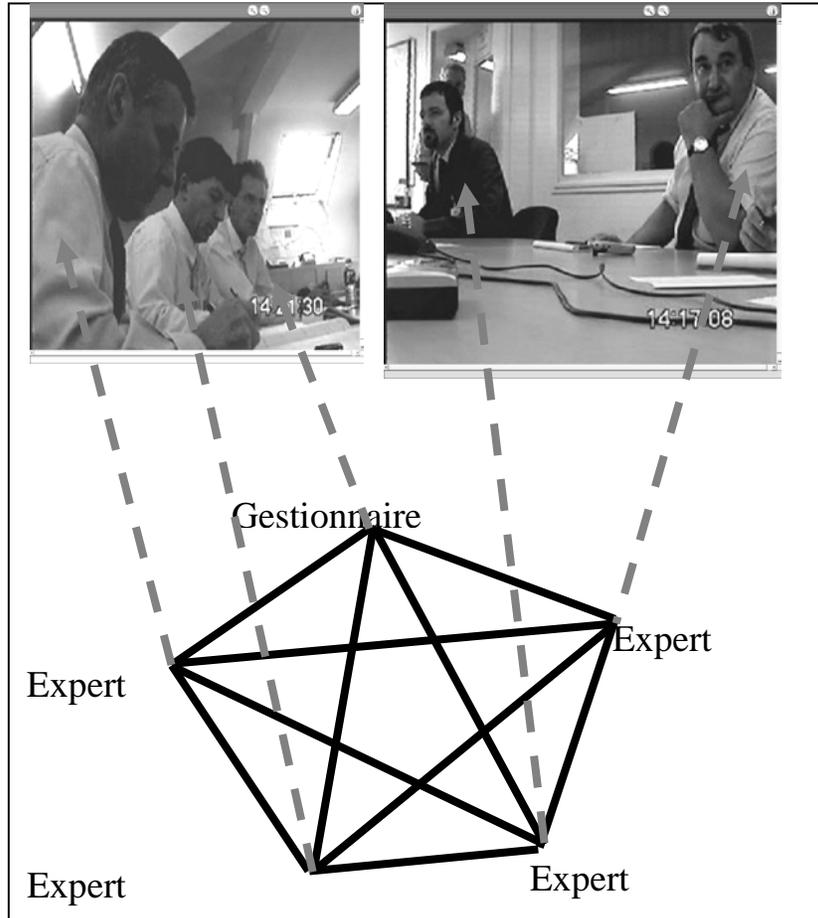


Figure 1. Structure de l'espace communicationnel occupé par les cinq acteurs présents lors de l'enregistrement de l'extrait n°1

attentivement le schéma obtenu, nous voyons apparaître des contradictions (flèches noires). Ces dernières nous donnent une première indication sur ce qui a pu déclencher le conflit (phase 2). Il est important de noter que ces contradictions n'apparaissent pas d'emblée dans le discours de l'expert 1. Elles apparaissent plutôt suite à l'analyse structurale au moyen d'Evoq©. En effet, elles concernent des éléments du discours pouvant être éloignés temporellement (cf. la première contradiction ci-dessous). Ceci est une première illustration montrant que notre analyse des relations sémantiques

Tableau 1
 Extrait n°1 : Transcription de l'extrait
 et découpage de ce dernier en trois phases

Phase 1: Discussion concernant l'état des lieux de la situation

- Expert 1** [...] les mesures de contamination des rails autour du wagon ne sont pas contaminées, l'environnement en dehors du wagon ne sont pas contaminé, donc tout ça sont quand même des éléments qui sont très concordants pour dire qu'il n'y a pas de contamination à l'extérieur du convoi, ce qui est aussi cohérent avec ce qu'on sait, ce qui a été expliqué plus tôt sur la robustesse de ces emballages. Deuxième point, c'est que au-delà de 20 mètres il y a confirmation qu'il n'y a pas de radioactivité ajoutée par rapport à la radioactivité naturelle, donc il n'y a pas de danger pour les populations voisines et même les installations industrielles voisines du lieu de l'accident. Eu hm, le euh l'élément qui est quand même anormal c'est la mesure des neutrons autour de l'emballage, dont on commence à avoir confirmation; il y a plusieurs types de radioactivités, pour faire simple, il y a des combustibles qui ont été irradiés euh génèrent aussi des neutrons, ces neutrons là ils sont arrêtés dans le euh dans l'emballage par de la résine qui pourrait avoir été partiellement dégradée par l'incendie, et effectivement on mesure plus, ce sont des mesures EDF, réalisées par l'équipe EDF qui est sur site, et on attend des confirmations par les mesures réalisées par l'IRSN, euh ils en mesurent plus que ce qu'on mesurait quand le convoi est parti hier de Chine. Par contre au niveau des autres types de radioactivité, on mesure exactement la même chose que sur ce qui est mesuré dans un convoi euh dans le convoi qui est parti hier et qui était intègre; donc il n'y a pas de euh de dommage autre sur la protection métallique apparemment du colis.
- Coordinateur de crise** Et quand vous dites qu'on en mesure plus que ce qu'on mesurait hier, on en mesure euh que dans l'absolu on atteint des doses qui euh...
- Expert 1** Non, donc ça reste faible, quelques dizaines de micro [?] / heure à quelques mètres du convoi, au-delà du de euh, il faut regarder les mesures au-delà de 20 mètres n'ont été réalisées que pour un type de radioactivité, qui est la radioactivité gamma; mais euh pour les neutrons on va être sur la même chose et je pense que, il n'y a pas de neutrons très fréquents dans la radioactivité naturelle, euh au-delà de quelques dizaines de mètres on est dans une situation où on ne mesurera rien du tout, on sera en dessous de euh, en dessous des seuils de détection.

Tableau 1 (suite)

Coordinateur	Ça en terme de communication tout à l'heure, est-ce qu'on a intérêt à évoquer ce genre de point ou est-ce qu'on en reste à l'essentiel?
Expert 2	Je vous conseillerais volontiers de dire que comme prévu dans la construction des emballages, la résine qui a été soumise au feu a joué son rôle de protection notamment en se carbonisant, ou en se détériorant un peu sous l'effet du feu; qu'il y a un effet de protection, il faut le comprendre. Et donc, l'augmentation de rayonnement, au moment global, c'est pas la peine de distinguer le neutron et le [?], reste très faible puisqu'on retrouve le bruit de fond à 20 mètres.
Expert 3	Moi je resterais [?], en disant que c'est un accident entre un poids lourd et un train qui contenait un colis. Le colis n'a pas été très endommagé puisqu'il n'y a pas eu de dissémination des produits radioactifs... que vous avez à déplorer les victimes et que euh votre émotion pour les familles des victimes. En ce qui concerne le colis nucléaire, les spécialistes sont sur place, on ne voit aucun danger pour la population, et euh vous maintenez le dispositif de protection, sans entrer dans le détail.
Expert 1	Je crois qu'il faut utiliser le mot « intègre », il n'y a pas ...
Expert 3	Il n'y a pas de dissémination.
Expert 1	Le colis est intègre et il n'y a pas de dissémination ...
Expert 3	Il n'y a pas de dissémination, ce n'est pas la peine de rentrer dans les détails.
Phase 2 : Émergence d'un conflit ayant pour sujet : « le colis est-il intègre? »	
Coordinateur	Comment est-ce que je peux dire que le colis est intègre alors que nous avons une équipe de euh, supposée de spécialistes qui est depuis 1 heure et demi sur place et qui est toujours pas fichue de dire ce qu'il en est?

Tableau 1 (suite)

Phase 3: Traitement du conflit par la métaphore de l'étanchéité

Expert 3	Alors n'utilisez pas le mot « intègre » si, s'il vous inquiète, mais il n'y a pas eu de DISSEMINATION de produits radioactifs... ils sont restés dans une enveloppe qui est restée étanche sous l'aspect dissémination Elle n'est pas étanche, totalement étanche sous l'aspect rayonnement, soit de neutrons, soit de de de gammas, mais sous l'aspect alpha, tout ce qui peut être euh dissémination, elle est intègre.
Expert 2	Bon, je je pense que là on ...
Expert 3	Il ne faut pas rentrer dans ce détail.
Expert 2	... le détail est trop fort.
Expert 3	Ah oui, oui, il ne faut pas rentrer dans ce détail.

superficielles structurant le discours permet d'atteindre un niveau d'organisation plus profond de ce dernier. Les principales contradictions que nous pouvons relever sont les suivantes :

- « [...] ces neutrons là ils sont arrêtés dans le euh l'emballage par de la résine qui pourrait avoir été partiellement dégradée par l'incendie [...] » contredit par « Le colis est intègre et il n'y a pas de dissémination » ;
- « [...] ces neutrons là ils sont arrêtés dans le euh l'emballage par de la résine qui pourrait avoir été partiellement dégradée par l'incendie [...] » contredit par « [...] ce qui est aussi cohérent avec ce qu'on sait, ce qui a été expliqué plus tôt sur la robustesse de ces emballages » ;
- « [...] il n'y a pas de contamination à l'extérieur du convoi [...] » contredit par « l'élément qui est quand même anormal c'est la mesure des neutrons autour de l'emballage [...] et effectivement on en mesure plus [...] ».

Les disjonctions sont représentées par une distance importante entre les termes et des flèches grisées. Les conjonctions sont indiquées par une grande proximité entre les termes. Des contradictions sont repérables que nous avons identifiées au moyen de flèches noires.

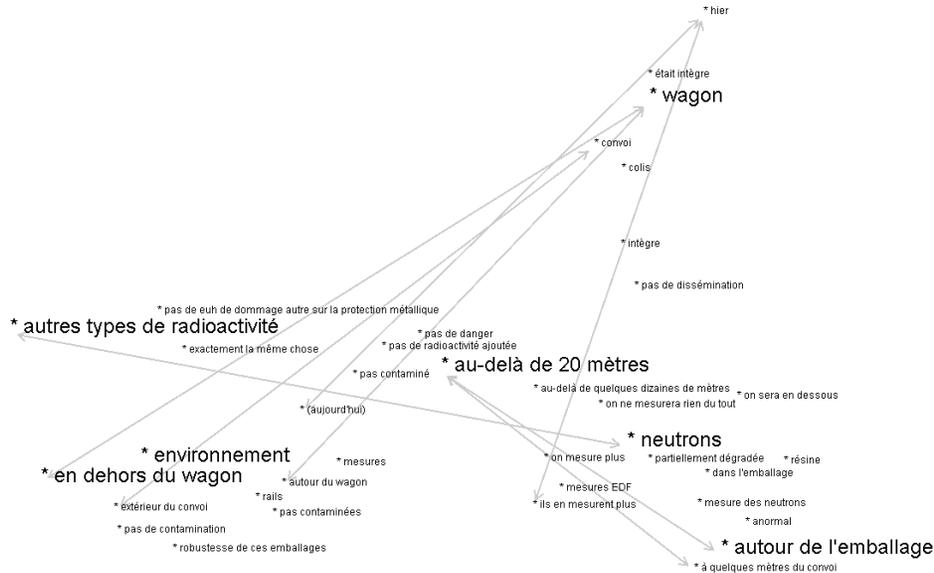


Figure 2. Carte sémantique correspondant à la première phase du discours analysé

Si on retourne à présent dans l'extrait (étape réflexive de la méthode), on constate que les experts 2 et 3 s'accordent avec les propos de l'expert 1. L'expert 3 affirme par exemple que « le colis n'a pas été très endommagé puisqu'il n'y a pas eu de dissémination des produits radioactifs »; de même l'expert 2 déclare que « comme prévu dans la construction des emballages, la résine qui a été soumise au feu a joué son rôle de protection ».

Si, pour ces spécialistes, un tel discours semble tout à fait justifié, il importe de se demander s'il en va de même pour le non spécialiste qu'est le coordinateur de crise : comment le colis peut-il encore être intègre (aujourd'hui, après l'accident) alors qu'on mesure une émission de neutrons? Dit autrement : comment le colis peut-il être intègre (pas de dissémination) et non intègre (on mesure plus de neutrons)?

Nous allons à présent analyser la deuxième phase de l'extrait au cours de laquelle le coordinateur de crise exprime son désaccord (manifestation du conflit).

Un second espace

La réaction du coordinateur de crise, manifestant son désaccord suite au discours des trois experts est particulièrement intéressante. Alors que le discours des experts présente manifestement des éléments contradictoires, on observe que le coordinateur de crise ressent la contradiction, mais situe son discours en amont du discours des experts. Le fond de l'affaire, pour lui, c'est qu'on ne sait pas qualifier la situation : « et qui n'est toujours pas fichue de dire ce qu'il en est ». S'agit-il oui ou non d'une crise? La situation est d'autant plus inconfortable que deux ethnométhodes interviennent ici qui devraient ordonner la prévalence des discours :

- prévalence du concret sur l'abstrait : « une équipe (...) qui est depuis 1 heure et demi sur place ». Ils sont sur place vs les experts sont dans les bureaux ;
- ce sont eux aussi des spécialistes et en cette qualité c'est à eux que sont confiés le rôle et la charge de définir la situation : « une équipe de, euh, supposé de spécialistes ».

La contradiction entre intègre et non intègre se trouve ramenée à un niveau beaucoup plus inquiétant : peut-on dire quelque chose (qu'il est intègre) ou ne peut on rien dire car il n'y a rien à dire puisqu'on ne sait pas...

Ce faisant le coordinateur de crise met le doigt sur le problème d'interprétation devant lequel se trouvent les experts. Ceux-ci ont des mesures et des indications concrètes mais ils ne parviennent pas à déterminer une fois pour toutes « ce que cela veut dire » c'est-à-dire, en termes ethnométhodologiques à trouver le pattern (la forme, le motif, la gestalt) qui leur permettrait de dire ce que les choses sont. On baigne en pleine crise de la définition, mais pas seulement de la définition de ce que les choses sont mais de la situation elle-même : sommes-nous oui ou non devant une crise?

Cette joute sur les ethnométhodes (les méthodes qui permettent à un groupe en situation de définir « ce qui est »), est encore très bien illustrée dans un autre extrait au cours duquel le coordinateur de crise se trouve à nouveau en désaccord avec l'expert 1 (Tableau 2). Cet extrait survient plus tôt dans le déroulement de l'exercice, mais il ne nous avait pas paru aussi complexe que le premier sur lequel nous avons décidé de diriger notre analyse. Ici, le coordinateur réagit à l'attitude particulière de l'expert consistant à s'appuyer sur des éléments théoriques a priori pour évaluer la situation concrète actuelle.

Tableau 2
Extrait n°2 : échange entre le coordinateur de crise et l'expert 1

Coordinateur	bon ce produit là qui brûle [?] qu'est-ce qu'il donne ...
Expert1	Ce n'est, ce n'est pas un produit qui brûle. Pour l'instant on a un château de plomb qui, autour duquel il y a eu des matières inflammables qui se sont enflammées... mais là, il n'y a aucune matière qui est en train de brûler. Euh, on a un emballage qui résiste au feu, qui a été agréé pour, et euh, sur lequel on été déversées des matières inflammables qui ont brûlé. Pour l'instant c'est tout ce qu'on sait. Il n'y a aucune raison que euh le colis soit a priori dans une autre situation puisqu'il n'y a pas eu de choc euh particulier, il n'y a pas eu de dégradation particulière de son état, de sa qualité de confinement des matières.
Coordinateur	Et quel est le euh, alors ça c'est l'hypothèse, ou c'est plutôt le constat actuel c'est-à-dire que l'enveloppe qui protège le produit n'est pas atteinte ... n'est pas altérée.
Expert1	Voilà ... n'a pas de raison de l'être si il n'y a pas eu d'autres types d'agression, dont on n'a pas l'information pour l'instant.
Coordinateur	Qu'est-ce qui permet de, de savoir, de confirmer que l'enveloppe n'est pas fragilisée, c'est un examen visuel, c'est un examen technique?
Expert1	Alors, c'est, c'est dans la façon dont on a autorisé cette enveloppe à transporter des euh, des matières. C'est-à-dire qu'avant qu'on autorise l'utilisation de ces enveloppes, il y a des (...) de tests pour obtenir un agrément
Coordinateur	D'accord, donc ce que vous me dites pour le moment c'est une réponse théorique. C'est-à-dire dans les conditions dans lesquelles l'accident a lieu, a priori l'enveloppe doit résister.
Expert1	Voilà.
Coordinateur	Bon, maintenant ça c'est de la théorie, maintenant en pratique comment est-ce que l'on vérifie que effectivement ça a résisté?

Tableau 2 (suite)

Expert1	On, on peut le savoir déjà visuellement si il n'y a pas eu de choc c'est mieux, et puis, pour vraiment avoir la ceinture de sécurité c'est la mesure de radioactivité autour ...
Coordinateur	Alors visuellement qui est-ce qui est capable de faire un examen visuel? Est-ce que c'est tout un chacun ou est ce que ce sont des spécialistes?
Expert1	Non, c'est tout un, enfin tout un chacun sera capable de dire si euh, si euh, bon j'ai déjà demandé aux pompiers tout à l'heure, ils m'ont dit que le wagon n'est pas renversé.

La question très pratique que pose le coordinateur de crise est la suivante :

- « Qu'est-ce qui permet de, de savoir, de confirmer que l'enveloppe n'est pas fragilisée, c'est un examen visuel, c'est un examen technique? ».

Et la réponse de l'expert s'appuie uniquement sur sa connaissance théorique générale (non liée à la situation présente) :

- « Alors, c'est, c'est dans la façon dont on a autorisé cette enveloppe à transporter des euh, des matières. C'est-à-dire qu'avant qu'on autorise l'utilisation de ces enveloppes là, il y a un ensemble de tests pour obtenir un agrément ... ».

La réaction du coordinateur manifestant son désaccord est cette fois-ci directe, en opposant la réponse théorique de l'expert à son besoin de réponse pratique, en rapport avec l'évènement concret actuel :

- « Bon, maintenant ça c'est de la théorie, maintenant en pratique comment est-ce que l'on vérifie que effectivement ça a résisté? ».

Ce passage peut donc nous aider à comprendre la réaction du coordinateur de crise dans le premier extrait analysé. En effet, dans son discours, l'expert 1 utilisait également des éléments théoriques pour évaluer la situation liée à l'accident ferroviaire :

- « [...] mais euh pour les neutrons on va être sur la même chose et je pense que, il n'y a pas de neutrons très fréquents dans la radioactivité naturelle, euh au-delà de quelques dizaines de mètres on est dans une situation où on ne mesurera rien du tout, on sera en dessous de euh, en dessous des seuils de détection ».

En se basant sur des prévisions tout à fait théoriques, l'expert 1 donne son point de vue sur la situation et manifeste encore cette attitude qui lui est propre et qui consiste à faire une totale confiance en la qualité du matériel employé pour transporter les produits radioactifs (identifiant ainsi sa qualité d'expert).

Nous assistons donc au cours de ces réunions d'urgence à négociation sur les ethnométhodes qui vont permettre de déterminer « ce qui est ». La joute se joue :

- entre la qualité de celui qui a la charge de savoir, l'expert, mais qui ne sait pas définir la situation. Comme il ne voit pas, il déduit;
- et celle de l'homme ordinaire qui utilise ses cinq sens.

D'une part, l'expert théoricien se référant au modèle théorique de la situation auquel il fait une totale confiance (ici au sujet de la résistance des matériaux employés pour protéger les produits radioactifs). Il s'agit de sa vérité à lui, de ses propres repères :

- « [...] on a un emballage qui résiste au feu, qui a été agréé pour, et euh, sur lequel ont été déversées des matières inflammables qui ont brûlé ».

D'autre part, le coordinateur de crise qui cherche à évaluer les risques réels inhérents à la situation de l'accident qui vient de se produire. Ses repères à lui sont les conditions de l'accident et les mesures/observations de la situation actuelle.

Et voici le troisième

La « résolution » du conflit émergeant dans l'extrait 1 (phase 3) sera apportée par l'expert 3 qui, employant une métaphore, distinguera les concepts de dissémination et de rayonnement en montrant que seul le premier peut-être associé à la notion d'intégrité. Comme le montre la Figure 3, la métaphore de l'étanchéité permet de distinguer deux types de radioactivité, à savoir la radioactivité gamma des neutrons qui est capable de rayonner à travers l'enveloppe du colis, et la radioactivité alpha qui reste cloisonnée à l'intérieur du colis si ce dernier est intègre :

- « Alors n'utilisez pas le mot « intègre » s'il vous inquiète, [...] les produits radioactifs] sont restés dans une enveloppe qui est restée étanche [...] ».

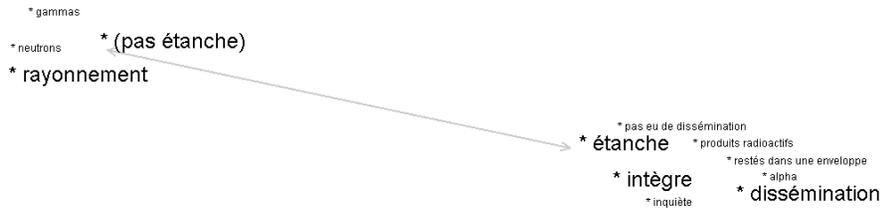


Figure 3. Carte sémantique correspondant à la troisième phase du discours analysé

Ainsi, dans la situation de l'accident nucléaire qui les concerne et dont ils doivent faire l'état des lieux, on peut dire que le colis est « étanche sous l'aspect dissémination », mais « pas totalement étanche sous l'aspect rayonnement ». La notion d'intégrité qui posait précédemment un problème de contradiction se voit de cette manière éclairée différemment :

- Le mot intègre ne souffre pas que l'on distingue des niveaux : le colis est ou n'est pas intègre.
« Intègre » = « intacte », « non altéré », « dans l'état initial », évoquant le non changement d'état, l'invariance.
- Par contre, on peut être étanche sous certains rapports et non sous d'autres.
« Étanche » = « ne laisse pas passer », « est imperméable à », la métaphore de l'étanchéité permet de sortir de la contradiction.

Au cours de leurs échanges, un problème surgit de l'usage de la notion d'intégrité par le coordinateur de crise : comment le colis peut encore être intègre (aujourd'hui, après l'accident) alors qu'on mesure plus de neutrons (radioactivité gamma)? La résolution du conflit viendra de l'introduction par un tiers (l'expert 3) d'un nouveau terme qui permet de distinguer les deux types de radioactivité mesurés sur le terrain et la signification de ces mesures quant à l'état d'intégrité du colis nucléaire. Plus précisément, la solution consiste dans le choix d'un terme polysémique et métaphorique : « étanche ». Ce dernier peut être à la fois employé pour décrire le rayonnement et la dissémination alors que ces deux termes s'opposent. L'étanchéité à la dissémination ne signifie pas forcément une étanchéité au rayonnement, cette dernière n'impliquant pas le même degré d'altérité physique du container. Alors que le terme « intégrité » ne permettait pas cette structure polysémique bivalente, le terme « étanche » permet à tous les experts (et plus tard au public) de comprendre pourquoi le container reste intègre (étanche à la dissémination) alors qu'on mesure la

présence de neutrons (radioactivité gamma) indiquant une perte d'étanchéité au niveau du rayonnement.

La contradiction résolue, le coordinateur et les experts débouchent sur un consensus sur la définition de la situation d'incertitude : on sait à présent ce que l'on sait et ce que l'on ne sait pas.

Discussion et conclusion

1. La focalisation sur les éléments déclencheurs de la crise, oblitèrent le processus même de la crise et, en particulier le processus symbolique qui est à l'œuvre dans une crise à savoir : la crise du code. Les choses, dont la définition s'opère habituellement spontanément ne se laissent plus définir aussi aisément. Sont soulignés alors les processus ordinairement infraconscients de définition intersubjective du réel. Les intervenants s'affrontent dans une joute fébrile qui met en jeu (et clairement en lumière) les ethnométhodes utilisées pour voir « émerger le sens ». Le problème est que le sens n'émerge pas, non seulement le sens de ce qui est mais également, en deçà de la qualification de ce qui est, l'existence même de la crise. Pouvoir dire : « c'est une crise » (ou « ce n'est pas une crise ») permet de sortir de la crise du code.

Cette situation s'apparente à celle de Alice au Pays des Merveilles qui, perdue dans le Bois des choses sans nom, voit le réel se dissoudre et, à la suite, sa propre identité. Alice n'est sauvée qu'en réalisant que le bois des choses sans nom (la crise du code) possède un nom : « Bois des choses sans nom » (crise) et aussitôt le caractère profondément anémique de la situation disparaît.

Un enjeu de poids pèse sur cette crise du code, sur la crise de l'activité de nommer, de dire « ce qui est » puisque de la définition en temps utile de l'existence d'une crise va dépendre des vies humaines.

2. Il peut paraître dès lors moins étonnant que la résolution du conflit analysé ici repose moins sur les qualités d'expertise technique des acteurs que sur la propension à utiliser le langage dans sa nature polysémique et métaphorique. La solution qui surgit de la rencontre entre le coordinateur et les experts nucléaires nous enseigne les leçons suivantes :

- La diversité des points de vue est responsable à la fois de l'émergence des conflits et de leurs solutions!
- L'importance de la métaphore. C'est elle en effet qui permet au tiers de résoudre la contradiction et par conséquent le conflit. Elle permet aux acteurs d'accéder à un niveau plus abstrait mais aussi plus créatif du langage, que celui concret et informatif au sein duquel les contradictions sont apparues.

3. La situation d'urgence s'analyse ici au travers des éléments langagiers. Ces derniers s'avèrent alors autant «°crisogènes°» (contradictions) que « thérapeutiques » (métaphore). D'une manière générale, c'est la nature ambiguë du langage naturel qui est ici soulignée. Dès lors, il est moins crucial de gérer le risque à force de procédures que d'accompagner celles-ci de véritables compétences langagières.

La controverse serait : certains voudraient épurer le langage pour qu'il soit univoque et justement peu susceptible d'interprétations dans une situation où l'urgence est prédominante. Nous nous demandons si, à l'inverse, le caractère plurivoque du langage naturel n'est pas une condition nécessaire pour sortir de situations critiques. Favoriser l'utilisation du langage naturel et des métaphores peut devenir une recommandation cruciale pour la formation des acteurs impliqués dans la gestion des crises et conflits. La capacité des acteurs de la gestion de crise à pouvoir user du langage comme d'un outil créatif est selon nous un élément clé leur permettant de construire un cadre de compréhension intersubjectif solide. Il faut pouvoir tirer parti des propriétés de robustesse et de plasticité du langage naturel qui est tout à fait capable de tenir compte des expériences passées tout en se tenant prêt à créer de nouvelles solutions dans un environnement nécessairement changeant et imprévisible.

Note

¹ Evoq est un logiciel de cartographie cognitive qui vise à produire des paysages sémantiques au départ des principes de l'analyse structurale. Ce logiciel permet de saisir le ou les champs sémantiques qui organisent la perception d'un locuteur dans une relation dialogique (avec lui-même ou avec autrui).

La spécificité de ce logiciel est qu'il tente moins de saisir une articulation de concepts au sein d'un raisonnement, que d'isoler les champs sémantiques dans lesquels se situe et se débat le locuteur. Pour ce faire, Evoq[®] s'appuie sur une conception du langage qui ne s'accommode pas des modes de représentation hiérarchiques classiques (en « branches et nœuds ») reposant sur la logique causale.

À partir de sa dimension pragmatique, la question de savoir *comment représenter le langage* a pris une importance éminemment théorique. Il s'agit en effet de trouver un espace de représentation dont les propriétés renvoient bien à la conception du langage portée par les courants de pensée dont nous avons précédemment extrait nos ingrédients. Evoq[®] permet de mettre à jour et de visualiser les termes de référence (présupposés) à l'intérieur desquels communiquent, raisonnent, et décident les acteurs et groupes d'acteurs. Il trace la carte des relations sémantiques et des choix intentionnels sur base desquels un individu ou un groupe désigne et structure la réalité, en l'organisant en fonction d'unités signifiantes. Ce logiciel saisit donc *le niveau de la construction symbolique de la réalité qui va organiser la*

perception. Evoq[®] est donc un outil de cartographie cognitive en ce sens qu'il vise ce niveau très fondamental de la constitution du « ce dont on parle », ce que nous appellerons la *perception*.

Références

- Alvesson, M., & Kärreman, D. (2000). Varieties of discourse : On the study of organizations through discourse analysis. *Human Relations*, 53(9), 1125-1149.
- Candellier, D. (2004). Ni recette miracle ni entreprise modèle. *Le bulletin de l'Ilec*, 352. En ligne : www.ilec.asso.fr.
- Derrida, J. (1978). *Writing and difference*. Chicago : University of Chicago Press.
- Cooper, R. (1989). Modernism, posmodernism and organisational analysis 3 : the contribution of Jacques Derrida. *Organisation Studies*, 10(4), 479-502.
- Fairhurst, G.T., & Putnam, L.L. (2004). Organizations as discursive constructions. *Communication Theory*, 14, 1-22.
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge : Polity Press.
- Jacques, J.M., Peris Mora, E., & Gatot, L. (1997). Aproximacion cognitiva de crisis y conflictos : el caso del Parque natural de la Albuferra (Valencia, Espana). *Humedales Mediterraneos*, 1, 11-18.
- Jacques, J.M., & Gatot, L., (1998). Environmental and industrial risk and crises assessment : a cognitive approach. Dans Ph. Hubert, & Cl. Mays (Éds) *Risk analysis : opening the process* (pp. 39-48). Fontenay aux roses : Institut de protection et de sécurité nucléaire CEA Saclay.
- Lalonde, C. (2004). In search of archetypes in crisis management. *Journal of Contingencies and Crises Management*, 12(2), 76-88.
- Varela, F.J. (1998). Construction du réel et affect : expérience du sujet, performances et narrations. *Cahiers critiques de thérapie familiale et de pratique de réseaux*, 19, 281.
- Merleau-Ponty, M. (1960). *Signes*. Paris : Gallimard.
- Ploux, S., & Ji, H. (2003). A model for matching semantic maps between languages (French/English, English/French). *Computational Linguistics*, 29(2), 155-178.

Wallemacq, A. (2001). Logic, language and objects, the « natural logic » of organizations. Dans B. Hellgren, & J. Löwstedt (Éds), *Management in the thought-full enterprise* (pp. 54-92). Bergen : Fagbokforlaget.

Wallemacq, A., & Jacques, J.-M. (2003). Spaces of thought, spaces for thought. Dans S. Linstead, & A. Linstead (Éds), *Thinking organization*. (pp. 39-61) London : Routledge.

Jean-Marie Jacques est docteur en biologie et professeur en stratégie et sciences de l'organisation, aux Facultés universitaires ND de la Paix (Namur-Belgique). Il a développé une approche originale de la gestion des crises basées sur les facteurs de perception et de cognition. Il a fondé et dirige le ReCCCoM, Centre de recherche des Facultés universitaires de Namur spécialisé dans la question. Sa contribution scientifique s'enrichit de nombreuses expériences de terrain. Un accent tout particulier est mis sur la gestion préventive des crises.

Nathanaël Laurent est assistant au département sciences philosophie et société, il est docteur en sciences biomédicales. Il est également licencié en philosophie. Avec Jean-Marie Jacques, il mène une réflexion en profondeur sur les processus de cognition et leur positionnement sur la scène philosophique. Le concept de conscience est au centre de leurs travaux.

Anne Wallemacq est docteur en sociologie et professeur dans le domaine des sciences de l'organisation. Ses intérêts de recherche sont le langage et la perception. Avec Jean-Marie Jacques elle a mis sur pied un logiciel de cartographie cognitive, Evoq, qui loin d'épurer le langage ordinaire, tente d'en restituer le caractère polysémique.

Analyse structurale de contenu et soutiens logiciels : une introduction au projet Anaconda

Jean-Pierre Hiernaux, Ph.D.

Université Catholique de Louvain

Résumé

Seront d'abord clarifiés l'objet spécifique, les fondements et les modes opératoires propres à l'analyse structurale de contenu. Au regard des conditions pratiques que celle-ci demande de gérer face à des données concrètes, se raisonne ensuite l'adéquation des types de soutiens logiciels possibles. Enfin seront présentés quelques dispositifs plus particulièrement appropriés, tels que nous les développons et testons actuellement à l'échelle du projet ANACONDA - pour ANALYse de CONtenu Disjonctive et Associative¹ - exploitant les ressources du traitement de texte Microsoft Word sous programmation VBA.

Mots clés

ANALYSE STRUCTURALE, ANALYSE DE CONTENU, LOGICIELS, CAQDAS

L'analyse structurale de contenu - objet, fondements, pratique

L'objet

Devant des manifestations de sens, notamment verbales et discursives, l'objet de l'analyse structurale de contenu consiste à saisir non pas le mode de manifestation du sens, non pas la forme ou la pratique discursive, mais bien le mode d'existence du sens comme tel, c'est-à-dire ce par quoi ce qui est évoqué se constitue en « réalité » perceptible ou concevable tout en se dotant de propriétés spécifiques. Par delà l'échelle discursive, cet objet de connaissance désigne donc l'ordre même du sens, un phénomène proprement sémantique.

Pour le praticien des sciences sociales cependant, pour l'analyste des comportements humains, le phénomène sémantique dont il s'agit ne se réduit pas à l'échelle de la langue comme telle où il renverrait aux « dénotations » des mots et expressions, c'est-à-dire à leurs investissements de sens standard dont rendent compte les définitions des dictionnaires et qui permettent de communiquer en sachant que si je lis ou entends « masculin », cela veut dire

RECHERCHES QUALITATIVES – Hors Série – numéro 9 – pp. 56-82.

LOGICIELS POUR L'ANALYSE QUALITATIVE: INNOVATIONS TECHNIQUES ET SOCIALES

ISSN 1715-8702 - <http://www.recherche-qualitative.qc.ca/Revue.html>

© 2010 Association pour la recherche qualitative

ceci. Alternativement, en effet, c'est sur les « connotations » que l'analyste se centrera en tentant de saisir – par delà ce qui en permet la communication – ce qui compose comme telles les représentations, les « réalités » ou les évidences particulières qu'assument des locuteurs ou des fractions de locuteurs spécifiques, ce qui – en même temps qu'un phénomène sémantique – constitue un fait sociologique, psychologique et/ou historique.

Ainsi, d'une page de « catéchisme » qu'un prêtre catholique, professeur de religion, composa pour ses élèves – il y a plus de 25 ans déjà, mais cependant bien après mai '68 – , une analyse structurale de contenu retiendra non le texte comme tel, ni son style, ni sa rhétorique, ni non plus ce qui fait que quiconque comprenant la langue utilisée comprendra aussi ce que désignent les mots « masculin », « matière », « exploiter » ou tout autre élément évoqué, mais bien le réseau constitutif de sens spécifique et particulier dont rend compte le graphe de la Figure 1.

Dans cette figure, on voit notamment comment – dans les lignes d'abord – les destinées se construisent au nombre de deux seulement et ce en se contre-définissant comme mâles ou femelles, comment aussi – au sein des objets de chacune de ces destinées – la matière se construit et tant qu'inverse de l'humain et réciproquement tandis que les actions des mêmes destinées se spécifient exclusivement et typiquement soit comme exploitation, soit comme valorisation, chacun de ces termes prenant sens comme inverse de l'autre. Par ailleurs, – en colonnes cette fois – l'exploitation se donne comme une action propre aux espaces extérieurs (extra-domestiques, publics), compatible seulement avec les objets d'ordre matériel tandis que la chaîne EXPLOITER – EXTÉRIEUR – MATIÈRE caractériserait la destinée mâle en la disjoignant de l'assortiment inverse où l'humain se valorise dans l'espace intérieur et constituerait ainsi la destinée femelle... Voilà exactement ce qui nous intéresse.

Les fondements

Au fondement de l'analyse structurale de contenu, la sémantique structurale prétend à des concepts et outils descriptifs congruents avec la nature intrinsèque que l'élaboration scientifique reconnaît aux phénomènes à décrire. À cet égard, si intellectuellement les phénomènes quantitatifs s'élaborent parfaitement dans l'ordre de la fréquence et tirent de ce qui régit cet ordre les instruments appropriés à leur traitement, les phénomènes sémantiques de leur côté, donc les phénomènes de sens, tiennent à l'ordre de la différence et de la ressemblance dont, à leur manière, ils peuvent recevoir des modalités de description adéquate. En particulier, l'existence même du sens peut se saisir comme résultant de la mise en oeuvre de différences : sans différencier, en effet, rien n'émerge à la conscience. C'est ce que, sans pour autant être

destinées	
MASCULIN	FÉMININ
objets	
MATIÈRE	HUMAIN
espace	
EXTÉRIEUR	INTÉRIEUR
actions	
EXPLOITER	VALORISER

Figure 1. Exemple d'un réseau constitutif de sens

linguiste mais de manière néanmoins fort séduisante, illustre le romancier Blaise Cendrars (né Frédéric-Louis Sauser, 1887-1961) en annexe de son roman *Moravagine* où il apparaît que la langue martienne ne connaît qu'un seul mot - *areukeukokex* - lequel veut tout dire... et qui donc ne veut rien dire, faute de faire la différence avec quoi que ce soit. A cet égard justement, la sémantique structurale pose comme forme minimale constitutive de sens, la relation de disjonction qui au sein d'une totalité initiale (Ti) fait émerger comme réalités spécifiques un couple d'unités de sens (Ai, Bi) qui se constituent comme telles en se contre-définissant : Ai / Bi (Greimas, 1966). Ainsi, par exemple, les « réalités » que désignent les mots « gauche » et « droite » naissent de la contre-définition qui se donne non seulement à penser mais même et d'abord à expérimenter lorsque l'on étend les bras à l'horizontale au sein de la totalité initiale qu'est à ce moment-là la « latéralité ». Ainsi aussi naissent par disjonction au sein de leurs totalités initiales les couples contre-définis que nous avons observés dans le graphe présenté plus haut. Et il en va de même des milliers d'autres couples du même type qui peuplent notre esprit d'autant de « réalités » et de notions qui orientent notre existence.

Certes, si la forme minimale constitutive de sens est du type que nous venons de situer, toute structure de sens plus large peut se saisir comme l'articulation d'une pluralité de telles formes selon un mode déterminé (Hiernaux, 1978). Ainsi, une articulation du genre évoqué et le mode qui la caractérise apparaissent-ils dans le graphe que nous présentions plus haut : celui-ci en effet doit sa structure typique au fait que lorsqu'un terme d'une disjonction donnée (mettons Ai de Ai / Bi) s'y associe à un terme d'une autre disjonction (mettons à Aj de Aj / Bj), ce qui engendre donc l'association Ai-Aj), les deux termes restants s'associent entre eux (soit Bi-Bj). C'est ce qu'un jargon commode, basé sur l'apparence générale du graphe obtenu, a fini par nommer une « structure en parallèle » (voir Figure 2).

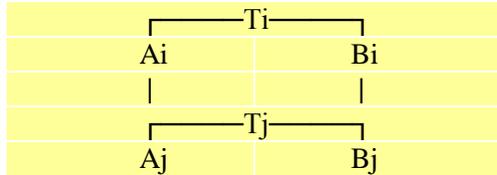


Figure 2. Principe d'articulation des structures en parallèle

Ces structures en parallèle, dont le mode d'articulation est le plus simple, sont aussi les plus fréquentes au sein des manifestations de sens. En même temps, elles ne sont pas les moins décisives quant à l'orientation des comportements.

D'autres modes basiques d'articulation, moins fréquents et que nous ne ferons ici qu'évoquer, peuvent évidemment prendre corps :

- ainsi, lorsqu'au moins deux disjonctions élémentaires interviennent dans un ordre contraint (rang "i" avant "j") et qu'un terme de la première (mettons Bi) forme à son tour la totalité de la seconde (Bi=Tj), apparaît la « structure en éventail » comme ci-dessous :

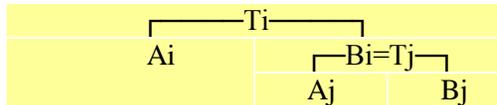


Figure 3. Principe d'articulation des structures en éventail

- ainsi encore, lorsqu'au moins deux disjonctions élémentaires interviennent simultanément comme dans la structure en parallèle mais, cette fois, en réalisant toutes les combinaisons possibles entre leurs termes, apparaît la « structure croisée »² :

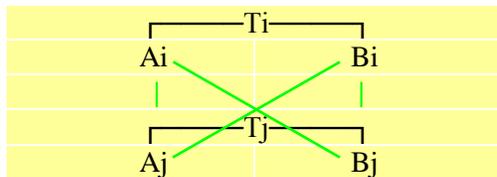


Figure 4. Principe d'articulation des structures croisées

Les graphes et modes d'articulation ci-dessus, faut-il le dire, s'en tiennent aux types de structure basiques c'est-à-dire formellement purs, la

possibilité existant par ailleurs que des manifestations de sens déterminées articulent entre-elles des séquences relevant tantôt de l'un ou de l'autre de ces types. Quoiqu'il en soit, là où il s'agit de rendre compte du mode d'existence du sens, il apparaît combien aux yeux de l'approche proposée le sens est une question de différences et de ressemblances, autrement dit de relations et de rapports qui séparent ou rapprochent, donc de positionnements respectifs qui font les unités en les distribuent les unes par rapport aux autres au sein de ce qui apparaît en substance comme un phénomène topologique³. Ainsi, notamment, l'investissement sémantique des unités de sens tel qu'il se donne à observer, leur sens tel qu'il peut être décrit le plus objectivement possible, se saisit finalement comme résultant de – et équivalant exactement à – la place que ces unités occupent au sein des structures dont elles émergent et telles que permettent de les saisir les limites d'une manifestation déterminée. Par exemple, en posant la question de savoir ce qu'est – dans le cadre de la Figure 1 considérée plus haut – la « matière », la réponse – dans les limites mêmes de sa validité empirique – s'affirme simple et précise : la « matière » est ce qui s'associe à l'« extérieur », à l'« exploitation », à la « destinée masculine » tout en se disjoignant de l'« humain », de l'« intérieur », de la « valorisation » et de la « destinée féminine ». Autrement dit, la définition qu'assumerait un dictionnaire issu de l'analyse des données retrace, au sein de l'ensemble topologique qui la génère, la position du terme qu'elle définit.

La pratique

La pratique de l'analyse structurale de contenu consiste ainsi et en substance à dégager des données concrètes les disjonctions élémentaires, les unités de sens qu'elles constituent et les structures plus larges dans lesquelles les unes et les autres prennent place et corps. Les aboutissements classiques d'une telle pratique sont – au terme d'autant d'esquisses provisoires et de (re)mises au point successives qu'il est nécessaire – les graphes des structures sémantiques qui sont à l'œuvre, soit la représentation topologique des relations qui déploient et articulent les unités de sens d'un même espace logique, d'une même « isotopie », même si les données à leur propos sont dispersées en différents lieux ou moments physiques de manifestations de sens plus ou moins amples.

Ainsi, pour compléter l'illustration, considérons la structure parallèle présentée ci-dessous en forme de tableau et qui résulte d'éléments dispersés sur plusieurs pages du récit de vie d'un immigré maghrébin (Figure 5).

<p>'ULT'A ACCOMPLISSEMENT+ VIE MEILLEURE+</p> <p>'ESP'A VILLE+</p> <p>'TPS'A PRÉSENT EUROPÉEN+</p> <p>'APP'A CONNAISSANCE+ MULTIDIMENSION+ OUVERTURE+ AUTONOMIE+</p>	<p>'ULT'B INACCOMPLISSEMENT- VIE FRUSTE-</p> <p>'ESP'B CAMPAGNE-</p> <p>'TPS'B PASSÉ COLONIAL-</p> <p>'APP'B IGNORANCE- UNILATÉRALITÉ- CLÔTURE- ALIÉNATION-</p>	<p>'ULT'T devenir de soi qualité existentielle</p> <p>'ESP'T socio-morphologie</p> <p>'TPS'T socio-chronologie</p> <p>'APP'T °cognitivité °supplétivité °perméabilité °liberté</p>
--	---	--

Figure 5. Exemple de structure en parallèle sous forme de tableau

On remarquera, du point de vue technique, d'une part que les totalités ('T) des disjonctions ont reçu leur description dans la colonne de droite du tableau, sur la ligne même des couples des contre-définitions correspondantes et, d'autre part, que les disjonctions ont par ailleurs été classées selon les plans de perception qu'elles contribuent à structurer et auxquels correspondent empiriquement une ou plusieurs disjonctions : deux pour le plan des ultimités ('ULT), une respectivement pour les plans de l'espace et du temps ('ESP; 'TPS), quatre pour le plan des appuis ('APP).

D'autre part, du point de vue du contenu, il apparaît notamment comment le locuteur met en jeu une quête dont les ultimités ('ULT'A / 'ULT'B) se raisonnent en termes de devenir de soi et de qualité existentielle tout en liant à des alternatives spatio-temporelles où le milieu urbain ('ESP'A) et le présent européenisé ('TPS'A) s'arrachent comme polarités positives à la campagne ('ESP'B) et au passé colonial ('TPS'B), chacun des binômes ainsi formés s'articulant du reste avec les appuis positifs ('APP'A) ou négatifs ('APP'B) à la quête poursuivie, lesquels se contrastent en termes de développement cognitif, de richesse des apports, de perméabilité sociale et culturelle et enfin d'autonomie des individus... soit toute une perception du monde articulée à l'histoire et au devenir de soi au sein de ce monde.

Ainsi encore les structures parallèles comparées ci-dessous, issues cette fois de centaines de pages de données reprenant d'une part les retranscriptions de débats de groupe entre locuteurs d'âge mûr dans des villages traditionnels d'il y a bientôt un demi-siècle et, d'autre part, des manifestations de sens diverses (livres, articles, tracts, publicités, manifestes, déclarations...) collectées au sein de milieux urbains « dans le vent » avant la fin de l'expansion économique des années '70 (Figure 6).

SOI		SOI	
RÉGULATION+	IRRÉGULATION-	COMPLÉTUDE+	INCOMPLÉTUDE-
ACTIONS		ACTIONS	
S'EFFORCER+	SE LAISSER ALLER-	S'AFFIRMER+	SE SOUMETTRE-
MOYENS		MOYENS	
CONTRAINTE+	LIBERTÉ-FACILITÉ-	LIBERTÉ-FACILITÉ+	CONTRAINTE-
SOURCES		SOURCES	
CONSTRICTEURS+	DÉCONSTRICTEURS-	DÉCONSTRICTEURS+	CONSTRICTEURS-
ULTIMITÉS		ULTIMITÉS	
ORDRE SOCIAL+	CHAOS-	BIEN-ÊTRE+ BONHEUR+	MISÈRE- SOUFFRANCE-
ESPACE		ESPACE	
INTÉRIEUR SOCIAL+	EXTÉRIEUR SOCIAL-	EXTÉRIEUR SOCIAL+	INTÉRIEUR SOCIAL-
TEMPS		TEMPS	
PASSÉ- CONTINUÉ+	FUTUR DE CHANGEMENT-	PRÉSENT et/ou FUTUR+	PASSÉ-

Figure 6. Exemple de comparaison de structures en parallèle

Du point de vue technique, l'on remarquera l'usage d'une grille de plans commune permettant des comparaisons sur base standard. Du point de vue du contenu, apparaît alors le contrepoint entre d'une part, une vision de soi et du monde de type traditionnel-ascétique et d'autre part, une vision de promotion individuelle et sociale de type hédoniste. Selon l'une, doit se conquérir la régulation de soi par l'effort que permet la contrainte qu'offrent divers « constricteurs » (les autorités et tout ce qui nous résiste) de manière à garantir l'ordre social tandis que l'intérieur socioculturel et la tradition sont les espaces-temps élus. Selon l'autre, motivera la complétion de soi à tous égards en s'affirmant grâce à la liberté et à la facilité qu'offrent autant de « dé-constricteurs » (politiques, psycho-socio-pédagogiques, industriels et commerciaux...), l'enjeu étant une ultimité de bien-être et de bonheur tandis que l'extérieur socioculturel et le changement actuel et futur forment, au regard de désirs insatiables, les réservoirs inépuisables de tous les possibles. On remarquera également comment, en passant d'un scénario à l'autre, des éléments objectivement identiques changent de position, donc de sens et de rôle : ainsi s'inversent les statuts de la contrainte et de la liberté, des constricteurs et des dé-constricteurs, de l'intérieur et de l'extérieur, du passé et du futur; ainsi aussi, changent non seulement de statut mais même de nom des éléments tels « s'efforcer » que la négativisation traduit en « se soumettre » (s'efforcer, c'est évidemment se soumettre à une exigence...) ou « se laisser aller » que la positivisation cette fois traduit en « s'affirmer » (qui, en effet,

consiste pour le moins à ne pas renoncer à notre cours spontané...). Voilà aussi que la parure des mots s'assortit aux perspectives qui s'en drapent!

Face aux données et aux soutiens logiciels

De masses de données emmêlées l'analyste fait ainsi surgir des graphes ou tableaux ordonnés, condensés et structurés. Mais entre la tête et la queue du processus, quelle est la chair et quelle est l'alchimie? quelle est la systémativité de la démarche? quelle est la complétude de l'analyse? quelles sont la vérifiabilité et la transférabilité des descriptions? peut-on en retracer la genèse autrement que par l'énoncé formel de ce qu'elle aurait dû être? et la pratique analytique pourrait-elle se partager avec ceux qui, comme Saint Thomas, aimeraient mettre les mains dans les plaies du Seigneur? ou encore pourrait-elle être transférée à d'autres qui la poursuivraient à partir du stade déjà atteint? Ou bien serions-nous plutôt dans le règne de la divine inspiration, l'artiste, sans rien laisser transparaître d'un quelconque procédé, tirant lestement du chapeau le blanc lapin qui ébahit les foules? voire serions-nous devant de plus obscurs tours de passe-passe dont l'un ou l'autre s'illustrerait marginalement d'extraits succulents sélectionnés pour le besoin de la cause et pour l'effet escompté?

Des alternatives positives, sans doute, pourraient tirer profit de quelque utile soutien logiciel. Mais, s'agissant de décrire des relations et des positions, les outils à base fréquentielle – et par là volontiers automatisables – relèveraient en même temps du non-sens : ni le sens ni la topologie, justement, ne tiennent à la fréquence. Quant aux phénomènes pertinents, l'automatisation de leur description demeure largement hors portée tant elle rencontre des obstacles difficilement surmontables. Ainsi en va-t-il de la complexité voire de la subtilité des modes de manifestation des relations entre unités de sens : si un discours peut énoncer directement et clairement que quelque-chose est à l'opposé d'une autre ou, au contraire, s'associe à celle-ci, une infinité d'autres modalités manifestent des relations assumées comme évidentes... même et y compris lorsque les locuteurs – pas plus qu'ils ne sont conscients de la grammaire qu'ils pratiquent – ignorent les rapports qu'ils assument tout en les manifestant néanmoins. Ainsi en va-t-il aussi des manifestations polymorphes des unités de sens, lesquelles peuvent, selon le cas, correspondre à un mot ou à autant de synonymes, à une locution ou à plusieurs et parfois à beaucoup, voire à des séquences plus amples ou à d'autres modalités encore. Ainsi encore le fait que la description des unités de sens ne va jamais sans requérir quelque abstraction pour transcender les variations non-pertinentes au regard du propos de l'analyse tandis qu'en même temps des recherches différentes impliquent des niveaux de pertinence et donc d'abstraction variables servis par la production de vocabulaires descriptifs assortis. Le dernier tableau présenté plus

haut en livre un exemple avec les termes « CONSTRICTEURS » et « DÉCONSTRICTEURS », où le premier vaut pour toute source de contrainte ou de limitations et le second pour toute source de licence ou de facilités. Sans cette condensation-réduction descriptive, qui abstrait à un degré relativement élevé, il n'eut jamais été possible de rendre compte du rapport d'inversion que les deux perspectives décrites présentent sur le plan des « SOURCES », rapport qui n'apparaît évidemment que lorsque se saisit sa base commune. Or, les termes concrets qui apparaissent à l'échelle des manifestations respectives sont, d'un côté, la religion comme police des mœurs, Dieu comme source de la règle, le prêtre, les enseignants et les parents comme surveillants de la conformité, les contraintes mêmes de l'existence, celles de la nature, etc. et, de l'autre côté, le progrès, le socialisme, le Concile du Vatican, Che Guevara, les psychothérapeutes et les gourous, les pédagogues libertaires, les producteurs de biens et de services, etc. Ce n'est donc que lorsque se transcendent toutes les différences non pertinentes du point de vue considéré qu'apparaissent dans les collections ci-dessus les convergences pertinentes que tente d'exprimer le vocabulaire descriptif créé ad hoc à l'aboutissement d'un processus d'observation et en même temps d'élaboration intellectuelle auquel – à l'heure actuelle à tout le moins – l'on voit mal un automate s'atteler et, surtout, aboutir avec quelque pertinence.

Les soutiens logiciels appropriés se réduisent donc modestement à ceux où :

- les instruments essentiels de l'analyste demeurent ses propres yeux et cerveau;
- cet analyste ébauche, en autant d'esquisses et de (re)mises au point successives – tels des escarpins essayés aux pieds des données – les graphes des structures qu'il voit émerger, ce qui confirme qu'en matière d'analyse structurale de contenu comme en bien d'autres domaines la rigueur des instruments, conceptuels ou matériels, n'ôte rien à l'art que demeure leur mise en œuvre optimale ni à ce que doit à cet art le résultat obtenu;
- s'offre le soutien d'un certain nombre de commodités marginales qui, tirant profit des disponibilités informatiques contemporaines, remplacent avantageusement les crayons, les surligneurs, les ciseaux, la colle et les boîtes à chaussures qui depuis belle lurette déjà aidaient les analystes à identifier, collecter et classer les traces de leurs objets de connaissance et les pièces à conviction de leurs démonstrations.

Au travers des énoncés qui précèdent, on devrait avoir reconnu les types de soutiens logiciels affublés de l'inélegant acronyme CAQDAS, qui – en une

langue étrangère mais dominante – vaut pour *Computer-Assisted Qualitative Data Analysis Software*, longue locution dénotant comme telle peu de choses spécifiques⁴ mais qui vaut là où, alternativement à la capacité qu’ont les ordinateurs de compter, s’exploitent plutôt celle qu’ils ont aussi à agir sur du texte : annoter, sélectionner, copier, coller, collecter, comparer, classer, etc. Des types de soutien donc auxquels devrait intéresser pour notre cas moins la raison – ou la mode? – des nœuds et arborescences⁵ que la contribution à la restitution des relations et topologies sous-jacentes aux données.

Le projet Anaconda – dispositifs de base et détails, compléments et technique

Les dispositifs de base

Les soutiens logiciels aux analyses structurales de contenu tels que testés sous le projet Anaconda adaptent des procédures relativement simples proposées déjà antérieurement pour le traitement manuel de corpus de données volumineux (Hiernaux, 1995). Là où l’élaboration des données consiste essentiellement à dépasser la dysmorphie entre d’une part la manifestation de sens observée et d’autre part l’organisation structurale sous-jacente, ces procédures et modalités contribuent à trois types d’opérations distribuées à leur tour sur trois espaces de travail :

- d’une part, sur l’espace de travail initial que forment les données à traiter (texte, retranscription...), les opérations de classement qui, par delà l’entremêlement des données à l’échelle de la manifestation, vont en rassembler les segments jugés pertinents dans des collections davantage homogènes;
- d’autre part, sur l’espace de travail que forment à leur tour les collections obtenues, les opérations de description qui, par delà la variété des modes de manifestation des contenus collectés, vont en condenser et énoncer au niveau d’abstraction jugé opportun les unités de sens et leurs disjonctions constitutives;
- enfin, sur un dernier espace de travail, l’esquisse de la structure d’ensemble qui, jusqu’à sa mise au point finale, va se dégager progressivement des relations observées entre les différents couples disjonctifs qui auront été décrits.

Le dispositif de base d’Anaconda vise à articuler les trois espaces de travail concernés en un ensemble synoptique et interactif où notamment :

- ce qui dérive des opérations réalisées dans l’un des espaces et qui importe à l’un ou au deux des autres y est automatiquement répercuté et visualisé;

- depuis les éléments présents dans un espace donné, (i) il est possible de retracer la voie vers les espaces en amont dont ces éléments procèdent et notamment d'y reprendre, poursuivre ou amender les opérations correspondantes, (ii) à terme, il est possible également, de re-visionner la progression des élaborations vers l'aval.

Techniquement, ce qui vient d'être évoqué peut être réalisé de différentes façons. Anaconda, pour sa part, a opté pour offrir comme support aux espaces de travail et à leur articulation de simples fichiers de traitement de texte : un par espace de travail. Gérés par le logiciel, mais lisibles, imprimables, échangeables... même hors celui-ci, ces fichiers – sauf évidemment celui qui contient les données en l'état initial – se créent et se dérivent automatiquement les uns des autres, s'affichent de même au moment opportun – en plein écran pour la commodité – et sont tout aussi automatiquement coordonnés quant à leurs contenus, noms, enregistrements, ouvertures, suppressions, protections, etc.⁶

Chacun des trois espaces de travail illustrés dans l'articulation de la Figure 7 donne ainsi exemple de l'affichage sur écran d'une partie de l'un des fichiers correspondants... même s'il a fallu ici, vu la surface réduite, raccourcir singulièrement les éléments illustrant les contenus.

Sur base de l'illustration ci-dessous, le fonctionnement du dispositif peut se détailler davantage dans ce qui suit.

(1) Au sein de l'espace/fichier 1, celui des données à traiter, l'analyste peut :

- par des procédures familières aux traitements de texte, sélectionner tout segment désiré depuis un seul mot jusqu'à un paragraphe entier;
- insérer en tête de chaque sélection au moins une marque d'indexation, une étiquette, qui détermine à quelle collection se joindra, dans l'espace/fichier des collections d'extraits et comme davantage explicité plus bas, un extrait sous forme de copie de la sélection concernée.

Dès qu'une sélection est réalisée et indexée, Anaconda le confirme par un surligné jaune. En même temps, dès la première sélection et indexation, les données initiales sont automatiquement réenregistrées dans un nouveau fichier – le « fichier indexé » – qui se substitue dans l'espace de travail en cours au fichier contenant les données vierges, lequel aura été fermé intact et sera conservé comme tel à toutes fins ultérieures.

Les marques d'indexation, pour leur part, répondent au type de structure

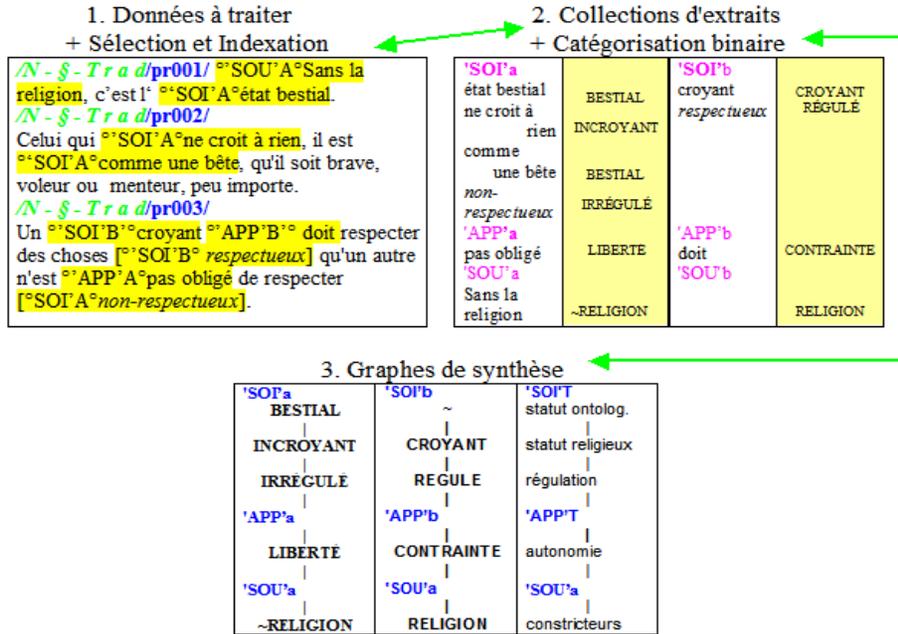


Figure 7. Les trois espaces de travail articulés d'Anaconda

auquel l'analyste pressent que participent les données qu'il sélectionne. Ainsi, en Figure 7 plus haut, espace/fichier 1, on verra l'exemple d'un fichier indexé à l'échelle de sélections qui préfigurent une structure en parallèle, type de structure auquel nous nous tiendrons dans ce qui suit⁷. Les marques d'indexation se composent dès lors des éléments suivants :

- un radical formé du sigle que l'analyste crée pour désigner le plan de perception du réel dont il voit manifester un contenu dans la sélection indexée;
- un suffixe 'A ou 'B notant la polarité du contenu concerné au sein d'une disjonction binaire du plan considéré.

Repris de la Figure 7 plus haut, espace/fichier 1, les exemples du Tableau 1 illustrent et explicitent davantage l'usage de telles marques d'indexation⁸.

Il importe de noter que les radicaux des marques d'indexation du type ici concerné n'équivalent pas aux totalités T des disjonctions binaires élémentaires, comme le serait par exemple la « latéralité » pour la disjonction gauche/droite. Ils réfèrent au contraire à des plans de perception du réel qui peuvent contenir plus d'une disjonction binaire : ainsi, en Tableau 1 ci-dessous,

Tableau 1
 Marques et pratiques d'indexation - exemples

°°SOI'A°état bestial	L'analyste voit dans la séquence /état bestial/ une manifestation du plan des « états de soi »; il crée donc le sigle 'SOI et l'insère en tête de la sélection, suivi de la mention de sa polarité : 'A ⁹ .
°°SOI'A°ne croit à rien	L'analyste voit /ne croit à rien/, i.e. « incroyant », manifester également le plan des « états de soi » avec la même polarité que ci-dessus : 'A.
°°SOU'A°Sans la religion	L'analyste voit /Sans la religion/ manifester le plan des « sources des états de soi »; il crée le sigle 'SOU et - /Sans la religion/ s'associant à /état bestial/ ('A) - il l'insère avec la polarité 'A.
°°SOI'B°croyant	L'analyste voit /croyant/ manifester la polarité inverse à /ne croit à rien/ ('A) plus haut ; il insère donc le sigle 'SOI suivi de la polarité 'B.
°°APP'B°doit	L'analyste voit /doit/, i.e. la contrainte, manifester le plan des « appuis aux états de soi »; il crée le sigle 'APP et l'insère avec la polarité 'B : /doit/ s'associe en effet à /croyant/ ('B).

on pourrait – en anticipant sur des étapes d'élaboration ultérieures – voir le plan 'SOI annoncer déjà deux disjonctions binaires BESTIAL / NON-BESTIAL et INCROYANT / CROYANT. Ce caractère relativement englobant des marques d'indexation est essentiel. En effet, confronté aux données sous la forme où elles se donnent, l'analyste n'est dans bien des cas pas à même ni de saisir ni d'énoncer d'emblée les unités de sens impliquées et encore moins leurs disjonctions constitutives. Souvent, il ne pourra le faire qu'après qu'aient été collectées un nombre suffisant de manifestations qui, sous des formes diverses, éclairent ces unités et disjonctions. Dans l'intervalle, et pour ouvrir à des progressions ultérieures, l'analyste doit donc rassembler les données en collections compatibles avec les possibilités d'observation qu'il peut actuellement pratiquer, soit celles en termes de regroupements encore relativement larges. De tels regroupements ne se ramènent cependant pas à des pis-aller : le nombre et la variété relative des manifestations qu'ils peuvent contenir offriront aussi les bases les plus favorables pour des descriptions ultérieures dont les niveaux d'abstraction alternatifs ne pourront apparaître qu'à ces conditions et à ce moment-là¹⁰. Enfin, lorsque des descriptions plus fines deviendront possibles – telle celle des disjonctions élémentaires –, leurs

résultats, plutôt que d'apparaître en vrac, prendront place au sein de tels regroupements et bénéficieront ainsi du classement de rang supérieur que commencent à esquisser, sous forme de plans, les radicaux des marques d'indexation. – Voilà qui non seulement aura situé au niveau des procédures d'indexation une condition de production des collections d'extraits, mais encore qui aura suggéré l'importance de telles collections à différents égards et pour différents moments du traitement des données.

(2) Dès que l'analyste réalise une sélection et indexation dans l'espace/fichier des données à traiter, l'espace/fichier des collections d'extraits (i) est automatiquement créé s'il n'existe pas encore et (ii) en tous cas – et tant que l'analyste ne décide pas de revenir à l'espace/fichier antérieur ou d'en rejoindre un autre – est affiché en visualisant l'extrait qui vient d'être inséré au sein de la collection à laquelle ses marques d'indexation l'ont destiné, collection qui, si elle n'existait pas encore, aura elle aussi été simultanément et automatiquement créée.

Au détail :

- l'espace/fichier des collections d'extraits anticipe sous forme de tableau le modèle du type de structure qu'il contribue à traiter : ainsi, en Figure 7 plus haut, le modèle de l'espace/fichier 2 est celui des structures en parallèle;
- horizontalement, le tableau pour structures en parallèle s'organise en sections transversales selon les plans de perception dont l'analyste a, lors des opérations de sélection et d'indexation dans l'espace/fichier des données à traiter, constaté la manifestation et créé les sigles qu'il a ensuite insérés aux marques d'indexation de ses sélections : dès qu'au moins une de ces sélections reçoit une marque d'indexation contenant le sigle d'un plan donné, ce sigle est automatiquement reporté dans l'espace/fichier des collections d'extraits et y définit une section horizontale qui pourra recevoir les extraits correspondants; – ainsi, en Figure 7 plus haut, espace/fichier 2, les sections/plans apparaissent en rose et traversent toutes les colonnes du tableau respectivement pour 'SOI (« états du soi »), 'APP (« appuis aux états du soi ») ou 'SOU (« sources des appuis »);
- verticalement, le même tableau s'organise en colonnes qui divisent les sections ci-dessus en parties gauche ou droite correspondant aux polarités 'A ou 'B des marques d'indexation que l'analyste attribue aux sélections réalisées dans l'espace/fichier des données à traiter et dont dérivent les extraits; – en fait, comme l'illustre la Figure 7 plus haut, espace/fichier 2, les espaces-colonne gauche et droite sont encore une

fois divisés en deux : les extraits s'insèrent plus précisément en colonne 1 ou 3 tandis que les colonnes 2 et 4 restent libres pour des traitements que nous préciserons bientôt;

- sur le canevas résultant des paramètres ci-dessus, et en fonction des marques d'indexation attribuées à la sélection dont il procède, chaque extrait rejoint donc automatiquement l'espace d'une collection qui correspond, en ligne, à un plan de perception donné et, en colonne, à une polarité donnée; – à ce stade, notons-le, il n'est pas visé à ce que, au sein de la collection des extraits d'un plan de perception donné, plan auquel peuvent appartenir plusieurs totalités et leurs disjonctions respectives¹¹, les extraits 'A relevant de la disjonction d'une totalité donnée se trouvent nécessairement en face d'extraits 'B relevant de la même totalité : comme déjà évoqué, il est en effet et fréquemment impossible de garantir de telles identifications lors des sélections et indexations dont dérivent tant les extraits que leur classement tandis que, comme il sera montré à l'instant, dès que la possibilité de ces identifications survient, elle ouvre aux traitements ultérieurs que l'étape actuelle a justement comme but de préparer.

Automatiquement et progressivement composé, développé et tenu à jour sous l'effet des sélections et indexations que l'analyste réalise dans l'espace/fichier des données à traiter, l'espace/fichier des collections d'extraits est cependant et également le lieu d'opérations subséquentes dès que l'analyste, au terme de – ou même pendant – la production des collections, voit émerger la possibilité de condenser en un énoncé descriptif adéquat les unités de sens, voire même les couples disjonctifs, que des extraits déterminés manifestent¹². L'analyste peut alors en effet, dans l'espace/fichier des collections d'extraits :

- en face d'un ou de plusieurs extraits, insérer une catégorisation, un vocable qui – à cette étape de progression de l'analyse – semble convenir pour condenser l'unité de sens sous-jacente aux contenus concernés : ainsi en Figure 7 plus haut, espace/fichier 2, les catégorisations créées et insérées en colonne 2 pour les extraits de la colonne 1, dont notamment BESTIAL pour les extraits « état bestial » et « comme une bête », ou les catégorisations créées et insérées en colonne 4 pour les extraits de la colonne 3¹³;
- en face d'un ou de plusieurs extraits relevant de polarités opposées au sein d'une même collection, condenser l'énoncé complet de la disjonction sous-jacente sous forme d'une catégorisation binaire enregistrée comme telle par le logiciel, et en insérer les termes respectifs au regard des extraits correspondants : ainsi, en Figure 7 plus haut, espace/fichier 2, au plan 'APP, l'enregistrement comme tel du couple

LIBERTÉ / CONTRAINTE pour les extraits « pas obligé » et « doit » ou, au plan 'SOI, l'enregistrement comme tel du couple INCROYANT / CROYANT même si les extraits correspondants n'étaient pas – et donc si ces termes eux-mêmes ne sont pas – face à face dans la collection, situation que géreront notamment les fonctionnements présentés dans ce qui suit.

(3) Dès que l'analyste insère une catégorisation en colonne 2 ou 4 de l'espace/fichier des collections d'extraits, l'espace/fichier des graphes de synthèse (i) est automatiquement créé s'il n'existe pas encore et (ii) en tous cas – et tant que l'analyste ne décide pas de revenir à l'un ou l'autre des espace/fichiers antérieurs – est affiché en visualisant le plus récent état de l'esquisse de structure d'ensemble qui résulte des traitements jusque là appliqués aux données.

Au détail :

- l'espace/fichier du graphe de synthèse reprend automatiquement le modèle du type de structure que contribue à traiter l'espace/fichier des collections d'extraits et que préfiguraient déjà les marques d'indexation insérées aux sélections dans l'espace/fichier des données à traiter;
- s'agissant de structures parallèles, les deux premières colonnes de l'espace/fichier du graphe de synthèse reprennent ainsi la grille d'intitulés par plans et polarités que tient constamment à jour l'espace/fichier des collections d'extraits qui répercute lui-même l'état des sélections et indexations réalisées dans l'espace/fichier des données à traiter; – une troisième colonne s'y ajoute où l'analyste pourra, comme précisé plus bas, insérer des suppléments d'élaboration; – en Figure 7 plus haut, espace/fichier 3, on verra les intitulés concernés figurer en bleu et l'on y reconnaîtra notamment ceux qui figuraient déjà en rose dans l'espace/fichier 2;
- sur le canevas de plans et polarités qu'offrent ainsi les deux premières colonnes de l'espace/fichier du graphe de synthèse, et selon les coordonnées des extraits concernés dans l'espace/fichier des collections d'extraits, Anaconda va cette fois automatiquement et à chaque attribution d'une catégorisation à un extrait (i) reporter uniquement les catégorisations telles qu'elles condensent et expriment les contenus des extraits sous forme d'unités de sens, et ce sans doublets¹⁴ et (ii) lorsque de telles catégorisations ont été mémorisées comme constituant des couples disjonctifs, en aligner les termes en face à face dans le graphe¹⁵; – l'analyste pourra, s'il le souhaite et comme illustré en Figure 7,

espace/fichier 3, compléter dans la troisième colonne sa description des totalités qui correspondent aux couples disjonctifs;

- comme les positions que vont occuper les catégorisations dans l'espace/fichier du graphe de synthèse correspondent à celles de l'espace/fichier des collections d'extraits et comme à leur tour celles-ci préfiguraient – déjà depuis les sélections et indexations dans l'espace/fichier initial – une forme de structure déterminée, c'est bien le contenu de la structure correspondant à cette forme en même temps que sous-jacente aux données retenues qui émerge au fur et à mesure dans le graphe de synthèse, avec un degré de dépouillement, de netteté et de confirmation empirique accru à chaque fois que, par delà la multiplicité des extraits collectés et des formes de manifestation qu'ils recèlent, davantage d'éléments se traduisent dans les vocables condensés des catégorisations.

(4) Au travers de l'ensemble du dispositif, et passant par deux échelles articulées qui confrontent à chaque fois l'analyste au résultat de chacune de ses opérations – d'une part les sélections et indexations répercutées dans les collections d'extraits et, d'autre part, les catégorisations répercutées dans le graphe de synthèse – s'accomplit ainsi l'objectif même de la description structurale : transcender la masse entremêlée et polymorphe des données initiales pour dégager les structures de sens sous-jacentes et les restituer en graphes synthétiques, condensés et ordonnés en même temps qu'empiriquement justifiés c'est-à-dire – le logiciel aidant ici encore à tenir l'exigence – des graphes finaux dont n'importe quelle catégorisation donne, par la commande adéquate, accès aux collections d'extraits qui la justifient tandis que, de la même façon, chacun des extraits de ces collections livre à son tour l'accès au contexte dont il est issu au sein des données traitées, données qui de la sorte, pour avoir conduit au graphe final, se rejoignent également à partir de lui et sans solution de continuité.

Notons que l'articulation des deux échelles rappelées plus haut ne constitue en rien une séquence aux temporalités contraintes. Tout au contraire, l'analyste peut à son gré et/ou selon les opportunités qu'il discerne lors du traitement d'un lot de données :

- tantôt privilégier, dans une première étape, la réalisation des sélections et indexations, en réservant pour une étape ultérieure la tâche de condensation et de catégorisation descriptive;
- tantôt, chaque fois que l'opportunité en apparaît – et même avant que ne s'atteigne le terme d'une collection –, privilégier la progression jusqu'à

l'attribution de catégorisations aux extraits collectés et donc jusqu'à la contribution à l'esquisse du graphe de synthèse.

Soit, à l'échelle des trois espaces/fichiers articulés, les usages soit séquentiel soit directement transversal schématisés comme le montre la Figure 8.

En même temps, la navigabilité en marche-avant et en marche-arrière au sein du dispositif, jointe à l'articulation de ses différents niveaux, contribue à sa plasticité permanente aux corrections et ajustements : ainsi, partant de n'importe quel élément en aval du processus d'élaboration, n'importe quel élément en amont peut être rejoint et amendé, supprimé ou remplacé, avec répercussion automatique des implications dans n'importe quel espace qui en dérive¹⁶. Concevant du reste que l'analyse – l'essayage des escarpins aux pieds des données – est dans son principe une pratique d'affinement par esquisses provisoires et mises au point successives, Anaconda offre également – d'une part dans l'espace/fichier des données à traiter, d'autre part dans celui des collections d'extraits – des dispositifs de recherche/remplacement susceptibles de gérer notamment des adaptations/modifications en série portant respectivement sur des indexations ou des catégorisations, dispositifs qui eux aussi gèrent, aux différentes échelles concernées, toutes les répercussions impliquées aux niveaux qui en dérivent.

Enfin, lors de l'enregistrement des modifications progressives aux trois fichiers qui servent de support à chacun des espaces de travail décrits jusqu'à présent, Anaconda offre systématiquement l'option de conserver et d'archiver ou non les enregistrements des états antérieurs. Il est ainsi possible, au titre de contribution directe à la pratique analytique ou encore à des fins didactiques – par exemple la révision du mode de travail suivi par un analyste en formation –, de retracer les étapes du traitement des données tout autant que les éventuelles modifications, reprises et réorientations appliquées à ce traitement lui-même.

Quelques détails et compléments

À l'échelle des détails opératoires, les dispositifs de base tels que décrits jusqu'à présent, bénéficient d'un certain nombre de modalités ou de fonctionnalités spécifiquement assorties.

Ainsi, pour les processus de sélection et d'indexation au sein des données à traiter, Anaconda offre notamment les latitudes et appuis suivants :

- la possibilité de multi-indexation, permettant qu'une sélection insécable où l'analyste voit se manifester des contenus relevant de plus d'un plan de perception puisse donner lieu à autant d'extraits-doubles que de plans concernés, chaque collection correspondante se complétant automatique-

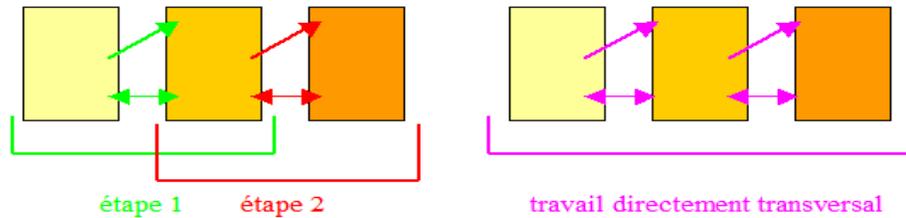


Figure 8. Usage séquentiel ou directement transversal des espaces/fichiers

ment par l'un de ces extraits; – soit, référant aux données traitées en Figure 7 plus haut, espace/fichier 1, paragraphe « pr003 », la modalité alternative d'indexation présentée au Tableau 2 ;

- la possibilité d'éditer les données initiales, permettant par exemple à l'analyste d'insérer au regard de l'original des « traductions » ou des clarifications qui pourront également entrer dans les sélections indexées, ou encore permettant – alternativement à la pratique illustrée en Tableau 2 ci-dessous – de dédoubler des sélections insécables par des copiés-collés de celles-ci ou de parties de celles-ci dans le fichier-même des données; – on trouvera ainsi, en Figure 7 plus haut, espace/fichier 1, paragraphe « pr003 », entre crochets, des exemples de « traduction »;
- la tenue à jour¹⁷ et la mémorisation automatiques des marques d'indexation dans des listes stockées sous forme de fichiers en même temps qu'affichées dans l'espace de travail correspondant d'Anaconda, favorisant ainsi, si l'analyste le souhaite, l'usage de listes-standard pour le traitement de divers lots de données et permettant en tous cas – lors des sélections – l'insertion des marques d'indexation par un simple clic.

À l'échelle des collections d'extraits et des opérations de catégorisation de leurs contenus, par ailleurs, Anaconda offre notamment les modalités et appuis suivants :

- l'adjonction à tout extrait de la référence à la localisation de la sélection dont il procède au sein des données traitées et, s'il y a lieu, de l'information sur la présence simultanée de l'extrait dans plus d'une collection (cas des multi-indexations)¹⁸;
- une tutelle automatique des pratiques de l'analyste lors de la formulation et de l'usage des termes des couples de catégorisations disjonctives, de manière à garantir et/ou gérer la cohérence du vocabulaire descriptif¹⁹;

Tableau 2
Multi-indexation - exemple

<p>°'SOI'A°°'APP'A°pas obligé</p>	<p>L'analyste voit /pas obligé/ manifester un « état de soi » : celui où on est sans contraintes; il insère donc la marque d'indexation 'SOI.</p> <p>L'analyste voit aussi /pas obligé/, i.e. l'absence de contrainte, manifester un appui à un « état de soi » : il insère donc également la marque d'indexation 'APP.</p> <p>L'extrait correspondant ira à la collection 'SOI et aussi à la collection 'APP.</p>
-----------------------------------	--

- de manière analogue à ce qui se passe pour les marques d'indexation, la tenue à jour et la mémorisation automatiques des marques de catégorisation et de leurs couples disjonctifs dans des listes stockées sous forme de fichiers en même temps qu'affichées dans l'espace de travail correspondant d'Anaconda, favorisant ainsi, si l'analyste le souhaite, l'usage de listes-standard pour le traitement de diverses collections d'extraits et permettant en tous cas – lors des opérations de catégorisation – l'insertion des marques correspondantes par un simple clic;
- selon les besoins de l'analyste, le reclassement automatique des extraits au sein des collections d'extraits de manière à favoriser leur consultation soit dans l'ordre historique de la réalisation des sélections et indexations, soit dans l'ordre physique de leur apparition dans les données traitées, soit en fonction de leur traitement en termes de catégorisations; – la dernière modalité citée correspond au fait que, entrés dans une collection dans l'ordre même de leur obtention, les extraits s'y entremêlent volontiers du point de vue des unités de sens que l'analyste pourra y reconnaître lors des opérations de catégorisation; or, pour soutenir ces opérations, il est avantageux que d'une part l'analyste puisse consulter synoptiquement les extraits qui ont déjà reçu une catégorisation déterminée tandis que, d'autre part, ceux qui restent à catégoriser seront eux-mêmes regroupés; c'est l'effet produit comme illustré à la Figure 9 où – passant du tableau de gauche à celui de droite –, les collections d'extraits respectivement 'A (colonnes 1 et 2) ou 'B (colonnes 3 et 4), au départ non autrement classées, sont réorganisées en fonction de l'état de

SOF'a état bestial	BÊTE	SOF'b respectueux	RÉGULÉ
non-respectueux	IRRÉGULÉ	croyant	
ne croit à rien			
comme une bête	BÊTE		

SOF'a ne croit à rien		SOF'b croyant	
état bestial	BÊTE		
comme une bête	BÊTE		
non-respectueux	IRRÉGULÉ	respectueux	RÉGULÉ

Figure 9. Reclassement automatique des extraits - exemple

catégorisation de leurs extraits pour présenter d'abord ceux de ces extraits qui restent à catégoriser (cf. cellules vacantes respectivement en colonne 2 ou 4), puis ceux qui le sont déjà et qui apparaissent à leur tour regroupés selon les marques de catégorisation qui leur correspondent, marques elles-mêmes classées en ordre alphabétique²⁰.

Outre les modalités et fonctionnalités ponctuelles du type qui précède, Anaconda peut ou pourra mobiliser également des dispositifs complémentaires à son fonctionnement de base. Ainsi :

- est disponible un dispositif de normalisation du texte des données et de référencement automatique de ses paragraphes, susceptible de traiter soit un seul fichier de données initiales à la fois, soit une série complète de tels fichiers – par exemple une collection de retranscriptions d'entretiens –, et qui réalise dès le départ les deux opérations automatiques que sont (i) d'une part la standardisation de la présentation des données et la vérification/adaptation de certaines de leurs formes aux modalités des traitements ultérieurs²¹ et, (ii) d'autre part, la numérotation automatique des paragraphes des données, insérant également à chacun de ceux-ci la référence au fichier concerné²²;
- est disponible également un dispositif extracteur automatique de segments de texte, susceptible lui aussi de traiter soit un seul fichier de données initiales à la fois soit une série complète de tels fichiers, et qui réunit, à partir de ces sources et sous forme de copies, des collections de segments-standard – par exemple des paragraphes – où, au gré de l'analyste, Anaconda reconnaît (i) soit la présence d'une chaîne de

caractères déterminée (un mot, une chaîne de mots, un lemme...), (ii) soit la présence d'au moins une chaîne de caractères déterminée parmi un ensemble de telles chaînes défini par l'analyste (par exemple en tant qu'indicateurs d'un champ lexical²³), (iii) soit la cooccurrence de deux ou plusieurs chaînes de caractères; – la constitution de telles collections peut être pratiquée notamment comme une étape d'exploration en même temps que de réduction de données initiales volumineuses; les résultats sont produits sous forme de fichiers qui peuvent à leur tour être traités selon le dispositif de base d'Anaconda tel qu'antérieurement exposé;

- est actuellement en développement un dispositif de modélisation/typologisation par comparaison et intégration automatique de graphes de synthèse tels que présentés plus haut et issus de sources multiples; – le propos correspond grosso modo à ceci : une série d'unités de données initiales, par exemple un certain nombre d'entretiens, livrent au terme de leur analyse autant de graphes de synthèse; un certain nombre de ces graphes peuvent présenter des homologues qui suggèrent leur participation à un « modèle » ou type commun qu'ils contribuent ensemble à manifester avec des degrés de consolidation variables notamment selon les échelles d'abstraction de la comparaison et le type de langage descriptif initialement produit par l'analyste; il s'agirait d'automatiser la production des résultats en la matière – donc de produire des « graphes de synthèse des graphes de synthèse » – à différents niveaux d'abstraction tout en rendant compte des degrés respectifs de consolidation, de manière notamment à créer des conditions où, stimulé à tenter des révisions de ses descriptions initiales, l'analyste pourrait aussi expérimenter les incidences de celles-ci sur les résultats de la modélisation; – on conçoit certes et par ailleurs que la modélisation à partir de graphes de synthèse multiples ouvre aussi à la possibilité de vérifier automatiquement la correspondance entre les modèles dégagés et les caractéristiques des sources et/ou acteurs qui supportent ces modèles, ceci notamment à l'échelle – et dans les limites – d'une sorte de micro-statistique²⁴... comme quoi toute chose, et donc aussi la fréquence, demeure là où elle a quelque pertinence.

Aspects techniques

Les procédures testées sous Anaconda sont en principe susceptibles de multiples réalisations aux détails variables et peuvent bénéficier d'outils de programmation divers. En pratique cependant, les développements actuellement disponibles ont été produits sous VBA – Visual Basic for Applications – le langage de programmation des applications de Microsoft Office : Word, Excel, PowerPoint... Puissant, souple et facile à utiliser²⁵, VBA

est accessible au très grand nombre : même si beaucoup l'ignorent, il fait normalement partie de l'équipement standard de tout détenteur de la suite bureautique citée. Si VBA permet d'automatiser ponctuellement des tâches en produisant des « macros »²⁶, il offre aussi le moyen de créer des logiciels de grande ampleur, tel notamment Anaconda.

Conçu pour programmer des applications préexistantes, VBA sert évidemment à développer des dispositifs qui usent des ressources de ces applications et qui, de ce fait, auront comme particularité d'être dépendants de celles-ci. Ainsi, Anaconda mobilise et exploite les potentialités du traitement de texte Word; aussi ne peut-il fonctionner sans celui-ci. Des points de vue militants pourraient le regretter, mais sans doute le feront-ils moins alors pour la dépendance elle-même que pour qui en bénéficie²⁷.

En contre-partie s'offrent notamment les avantages pratiques et techniques suivants :

- Anaconda se lance et s'affiche dans l'environnement même d'un traitement de texte largement répandu, comme s'il en était une fonctionnalité : il opère donc dans un cadre familier à un grand nombre d'utilisateurs;
- dans ce cadre, ces utilisateurs – hormis les fonctionnements contrôlés par Anaconda et qui obéissent à sa logique spécifique – gardent à leur disposition toutes les fonctionnalités du traitement de texte habituel ou qui sont accessibles à partir de celui-ci;
- pour la programmation, d'autre part, les fonctionnalités préexistantes ne sont évidemment d'aucune façon à produire à nouveaux frais ni distinctement : il suffit d'asservir de manière adéquate celles requises pour les propos spécifiques que les développements visent à servir;
- enfin, même si les praticiens des sciences sociales sont loin d'être tous des bricoleurs de soutiens logiciels, il n'est – pour la poignée de ceux qui le seraient tout de même, pour ceux aussi qui pourraient leur apporter un appui ou des contributions plus qualifiées, et finalement pour la production et le perfectionnement des soutiens logiciels eux-mêmes – il n'est sans doute pas sans intérêt de disposer en même temps d'outils de création, d'adaptation et de perfectionnement aussi accessibles et largement diffusés que ceux évoqués dans ce que nous clôturons ici.

Notes

¹ En zoologie, l'anaconda – *eunectes murinus*, ordre des Ophidiens, famille des Boïdés – est un boa aquatique du bassin amazonien, long de sept à onze mètres, non venimeux

et qui vit des fruits de son étreinte. Les deux derniers caractères flattent évidemment notre choix homonymique.

² A la différence du graphe qui suit, dès l'implication de trois disjonctions élémentaires, la représentation des structures croisées n'est techniquement plus réalisable qu'au moyen de matrices appropriées. Ces matrices peuvent également recevoir les données correspondant aux structures en éventail et parallèles; elles révèlent alors comment ces autres formes structurales, loin de former des natures sui generis, ne sont – à des échelles variables – que des réalisations réduites par rapport au maximum de combinaisons possibles qu'offre, pour un nombre donné de disjonctions élémentaires, l'articulation croisée de leurs termes respectifs; alternativement aux graphes, les matrices pourraient donc offrir une base commune aux descriptions et analyses comparées (Hiernaux, 1994).

³ Au simple sens étymologique : lié à l'organisation des lieux, sans compromettre ici ni une branche particulière des mathématiques, ni une dérive lacanienne.

⁴ Pour situer plus précisément les propositions qui suivent au sein des ressources logicielles destinées à l'analyse de corpus discursifs, voir Lejeune, 2007.

⁵ Allusion aux procédures typiques – et du reste tout-à-fait légitimes dans leur segment particulier – que soutient spécifiquement l'un ou l'autre logiciel dont la pénétration sur le marché se fait telle que la totalité des démarches non fréquentielles pourrait sembler lui appartenir.

⁶ Par exemple, lorsque après fermeture de fichiers articulés comme décrit, l'un d'entre eux est ré-ouvert pour reprendre le travail, ceux dont il dérive et/ou ceux qui dérivent de lui peuvent être automatiquement et simultanément ré-ouverts. Sous la tutelle du logiciel, il n'est par ailleurs pas possible de supprimer des fichiers participant à une série articulée sans être appelé à gérer les incidences sur celle-ci, etc.

⁷ Mutatis mutandis, des traitements analogues à ceux conçus pour les structures en parallèle pourraient s'appliquer aux structures en éventail ou croisées; considérablement moins fréquentes et moins amples ces derniers types de structures rendent cependant aussi moins urgents les soutiens logiciels à leur mesure.

⁸ Le propos, notons-le, n'est pas ici de valider l'analyse des données comme telle mais bien et uniquement d'illustrer la pratique d'indexation, à charge de l'analyste de justifier – dans et par les données – ce qu'il fait émerger de celles-ci.

⁹ Lors d'une toute première insertion, le sigle d'une polarité disjonctive – 'A ou 'B – est en principe indifférent. Par la suite, l'analyste tiendra compte des associations observées d'une sélection à l'autre.

¹⁰ Ainsi, en examinant anticipativement la Figure 7, espace/fichier 2, on comprendra que c'est seulement parce que l'on a réuni dans le plan des « états de soi » – 'SOI – tous les extraits repris en colonnes 1 et 3 que peut surgir la question de savoir si, à l'échelle des disjonctions structurant ce plan, il est plus optimal pour une finalité de recherche donnée de mettre en avant trois disjonctions – BESTIAL / NON-BESTIAL, INCROYANT / CROYANT et IRRÉGULÉ / RÉGULÉ – ou, à une échelle d'abstraction supérieure et valant pour l'ensemble, une seule disjonction que pourraient

reprendre les vocables IRRÉGULÉ / RÉGULÉ en admettant que la bête, l'incroyant et celui qui ne respecte rien ne seraient tous que des manifestations de l'être irrégulé.

¹¹ Pour l'illustrer, anticipons en examinant la Figure 7, espace 3, où le plan 'SOI est structuré par trois disjonctions, chacune correspondant à une des totalités décrites en colonne 3.

¹² Le moment où une collection atteint l'échelle requise pour que puisse en émerger l'esquisse de son principe, le moment de la « saturation », dépend tout autant de ce qui s'y donne à voir que du regard qu'y porte un analyste donné : certains verront vite, d'autres auront besoin de plus de temps et/ou de données, d'où – pour le soutien logiciel – la nécessité de laisser place à des conjonctures et rythmes variables.

¹³ Si l'illustration contient souvent la correspondance entre une catégorisation et un seul extrait, on se souviendra que l'espace réduit de cette illustration n'aura pas permis ici de montrer le cas le plus fréquent : la catégorisation unique pour une multiplicité d'extraits au contenu convergeant mais aux formes variables.

¹⁴ Ainsi la catégorisation BESTIAL qui figure deux fois en Figure 7, espace/fichier 2, colonne 2, n'est reportée qu'une seule fois dans le graphe de l'espace/fichier 3 et il en irait de même de toutes autres redondances, y compris nombreuses : le graphe de synthèse ne se complète donc que lorsqu'une catégorisation nouvelle est attribuée à au moins un extrait; dans les collections d'extraits, par ailleurs, l'association d'une même catégorisation à plusieurs extraits les identifie comme autant d'occurrences de l'unité de sens correspondante, occurrences qu'il est possible de rejoindre automatiquement en marche-arrière à partir de la catégorisation insérée au graphe de synthèse.

¹⁵ C'est le cas, en Figure 7, des catégorisations CROYANT et INCROYANT qui, dans l'espace/fichier 2 et bien que formant un couple, occupaient des lignes différentes du fait de la localisation initiale des extraits dont ils étaient issus. – Voilà pourquoi dans les collections d'extraits – et sans même considérer les difficultés qu'il y aurait à le faire, il n'importe guère de placer en face à face ceux qui se rapporteraient respectivement aux polarités 'A ou 'B d'une même totalité : cette organisation doit apparaître dans les résultats de l'analyse tandis que les étapes qui y conduisent peuvent et doivent disposer de la labilité qui leur convient.

¹⁶ Pour donner un seul exemple : la suppression dans l'espace/fichier des données à traiter d'une sélection qui soutiendrait, dans l'espace/fichier des collections d'extraits, l'unique extrait soutenant lui-même une catégorisation donnée dans l'espace/fichier du graphe de synthèse... modifierait directement ce dernier graphe lui-même en y supprimant la catégorisation concernée.

¹⁷ Incluant notamment que lorsque l'analyste insère pour la première fois la marque 'NNN'A, celle-ci est intégrée à la liste de même que 'NNN'B et que, lorsque la dernière sélection portant une marque d'un radical déterminé est supprimée ou ré-indexée, cette marque – sauf objection de l'analyste – est ôtée de la liste.

¹⁸ L'espace réduit des illustrations en Figure 7 plus haut, espace/fichier 2, n'aura pas permis d'y visualiser ces éléments. En grandeur nature, cependant, chaque extrait au sein d'une collection est – sur le modèle du référencement des paragraphes tel que présenté en Figure 7 plus haut, espace/fichier 1 – muni du nom du fichier et du numéro

du paragraphe de la sélection dont il procède ainsi que de la reproduction des marques d'indexation de cette sélection, lesquelles, lorsqu'elles sont multiples, livrent immédiatement l'information sur la localisation des « doublets ». – Ici comme ailleurs, Anaconda veille ainsi à ce que ses résultats demeurent lisibles même hors son cadre, par exemple sous forme de simple affichage de fichiers ou de consultation de copies imprimées.

¹⁹ Lorsqu'il a fait d'un terme la manière d'exprimer l'inverse d'un autre, l'analyste ne peut en effet modifier l'un quelconque des membres du couple sans notamment (i) modifier de la même façon toutes les occurrences de ce couple, (ii) avoir vérifié si les termes de remplacement n'engagent pas eux-mêmes d'autres couples et, le cas échéant, avoir géré la situation et ses incidences.

²⁰ On se souviendra de ce que les tableaux d'extraits, à la différence des graphes de synthèse, ne visent pas à placer en face à face sur une même ligne les éléments respectivement 'A et 'B correspondant à une même totalité. Lorsqu'un tel alignement survient néanmoins, comme en première et en dernière ligne du tableau de droite de notre illustration, il s'agit d'un effet purement aléatoire. En fait et comme exposé, le reclassement des extraits s'opère distinctement à l'échelle de la collection soit 'A soit 'B d'une section de tableau. Le tableau de droite de notre illustration présente un aboutissement où chacune de ces deux collections a été reclassée selon l'état des catégorisations attribuées à ses extraits.

²¹ Un seul exemple pour donner à comprendre : Anaconda reconnaît ses marques d'indexation au fait qu'elles commencent et finissent par le caractère °. Ce caractère ne peut donc figurer spontanément dans les données sans risquer de dérouter certaines procédures automatiques. Si c'est le cas, il doit donc être supprimé ou remplacé... ce qui, dans le cas de ° (code 176), se présente très favorablement : le caractère alternatif ° (code 186; oui, il est différent du précédent – regardez bien!) existe en effet et peut servir de remplacement à un grain près.

²² Ces informations serviront ultérieurement de base au référencement automatique des extraits dérivant des sélections opérées dans les données. Leur insertion aux paragraphes des données initiales tient cependant et une fois de plus au souci d'Anaconda de soutenir y compris les pratiques les plus « primitives » d'un quelconque analyste, en l'occurrence la copie et/ou le découpage des paragraphes avec des ciseaux ou encore la pratique du copier-coller avec l'ordinateur, deux variantes pour lesquelles il est bon de munir des extraits devenus autonomes de l'adresse de leur origine, adresse que l'on pourra également retrouver dans le fichier initial.

²³ On peut penser, par exemple, au champ familial avec les chaînes « père », « mère », « frère », « sœur », « oncle », etc.

²⁴ Pour illustrer la perspective de manière ne fut-ce que primitive, on conçoit comment, devant un modèle auquel participeraient des locuteurs des deux sexes, on pourrait vérifier le taux de contribution de chacun de ceux-ci, soit en soi, soit même relativement à la probabilité qu'en offrent les effectifs au départ.

²⁵ S'il n'était que la documentation qui l'accompagne d'office n'est pas des plus immédiatement pénétrables aux novices... et le dire ainsi est sans doute une litote.

²⁶ Pour « macro-instruction » : séquence d'instructions préenregistrées pouvant être répétées par l'ordinateur (www.le-dictionnaire.com).

²⁷ Comme le rappellent les noms de famille ci-dessous, il n'y aurait de ce point de vue aucune différence à dépendre de Microsoft Visual Basic plutôt que de Microsoft Word, même si le premier – VB, sans le « A » de « for Applications », un « grand frère » de VBA obéissant au même langage – produit des applications au fonctionnement totalement autonome.

Références

Greimas, A.J. (1966). *Sémantique structurale*. Paris : Larousse.

Hiernaux, J.P. (1978). *L'institution culturelle. Méthode de description structurale*. Louvain-la-Neuve : Presses Universitaires de Louvain.

Hiernaux, J.P. (1994). « *Et hic tres unum sunt* ». *Structures croisées et théorie des réductions*. [Note de travail inédite disponible auprès de l'auteur], 940430.

Hiernaux, J.P. (1995). Analyse structurale de contenus et modèles culturels. Application à des matériaux volumineux. Dans L. Albarello, F. Digneffe, J.P. Hiernaux, C. Maroy, D. Ruquoy, & P. de Saint-Georges (Éds), *Pratiques et méthodes de recherche en sciences sociales* (pp. 111-144). Paris : Armand Colin.

Lejeune, Ch. (2007). Petite histoire des ressources logicielles au service de la sociologie qualitative. Dans C. Brossaud, & B. Reber (Éds), *Humanités numériques, Tome 1. Nouvelles technologies cognitives et épistémologie* (pp. 197-214). Paris : Lavoisier.

Hiernaux, Jean-Pierre est docteur en sociologie, professeur au département des sciences politiques et sociales de l'Université Catholique de Louvain et membre du Laboratoire d'anthropologie prospective de la même université. Depuis la réalisation de sa thèse doctorale - L'institution culturelle : systématisation théorique et méthodologique (1977) - il n'a cessé de se consacrer aux développements propres à saisir dans différents contextes la composition et la dynamique des systèmes de sens socialement produits et socialement opérants. Il a contribué dans ce cadre à la mise au point de procédures d'analyse structurale de contenu dont il assure depuis de nombreuses années les formations en second et troisième cycles en même temps que des applications pratiques en divers domaines. Plus récemment, il s'est intéressé aux possibilités d'offrir à de telles analyses des supports logiciels appropriés, quitte à les produire soi-même.

Logiciel NotaBene pour l'annotation linguistique. Annotations et conceptualisations multiples

Nicolas Mazziotta, Ph.D.

Université de Stuttgart

Résumé

Nous présentons le logiciel libre NotaBene. Il s'agit d'un logiciel destiné principalement à l'annotation linguistique de ressources textuelles écrites en ancien français, mais dont les principes se veulent suffisamment génériques pour servir à l'exploitation d'autres types de ressources et éventuellement dans d'autres disciplines. NotaBene est basé sur des standards (XML, RDF et OWL) et permet l'affichage simultané d'analyses de nature différentes (p. ex. : lemmatisations vs arbres syntaxiques ou de progression thématique) ou simplement concurrentes (effectuées par deux chercheurs différents). Le logiciel permet également à chaque utilisateur de définir son propre modèle d'annotation et son jeu d'étiquettes personnalisé. Le modèle est enregistré de manière formalisée pour permettre les échanges et les traitements automatiques.

Mots clés

LINGUISTIQUE, VISUALISATION MULTIPLE, ONTOLOGIES, RDF, LOGICIEL LIBRE

Introduction

Nous voudrions présenter ici le logiciel libre NotaBene (disponible à l'adresse <https://sourceforge.net/projects/notabene/>), dont nous avons commencé la conception et le développement d'un prototype en Python entre janvier et juin 2008 dans le cadre d'un post-doctorat au laboratoire ICAR (Université Lyon 2, UMR 5191, <http://icar.univ-lyon2.fr/>). Il s'agit d'un logiciel destiné principalement à l'annotation linguistique de ressources textuelles écrites en ancien français, mais dont les principes se veulent suffisamment génériques pour servir à l'exploitation d'autres types de ressources et éventuellement dans d'autres disciplines. Le logiciel est actuellement développé dans le cadre du projet nommé *Syntactical Reference Corpus of Medieval French* – projet subventionné par l'Agence nationale de la recherche (ANR, France), sous la responsabilité de Sophie Prévost (LaTTiCe, UMR 8094, Paris), et la Deutsche

Forschungsgemeinschaft (DFG, Allemagne), sous la responsabilité d'Achim Stein (Université de Stuttgart), de janvier 2009 à janvier 2012.

NotaBene est basé sur des standards (XML, RDF et OWL, voir la section 3 ci-dessous) et implémente un modèle de visualisation multiple des ressources.

Abordons l'interface du logiciel au travers d'un exemple très simple pour commencer (nous montrerons ci-dessous d'autres emplois correspondant à une utilisation réelle). Nous voyons, dans la Figure 1, l'analyse de la phrase forgée *Demain, Jean ira à Paris* (les principales zones de l'interface sont délimitées par des rectangles noirs).

Comme le montre la Figure 1, l'environnement de travail est subdivisé en deux parties. Le panneau de gauche contient de multiples annotations d'un même document, visualisées simultanément sous des représentations diverses (texte courant, concordance, arbre). Ces annotations multiples feront l'objet de la section 2. Le panneau de droite contient un éditeur de terminologie, utilisé pour sélectionner ou modifier les jeux d'étiquettes mobilisés lors de l'annotation. Les questions de conceptualisation et de construction des jeux d'étiquettes seront abordées dans la section 3. Au préalable, nous dirons quelques mots concernant la nécessité de développer un tel logiciel.

Section 1 : Pratiques d'annotations et besoin d'un outil

Nous présenterons tout d'abord, sans nous étendre outre mesure, l'intérêt des annotations linguistiques, ainsi que la nécessité de développer un nouvel outil permettant l'annotation manuelle.

Corpus annotés en linguistique

À la différence des corpus annotés en sciences sociales, les corpus annotés dans le cadre de recherches linguistiques sont souvent enrichis de manière relativement exhaustive, car les questions posées nécessitent une vue d'ensemble; l'étude quantitative des propriétés morphosyntaxiques, description des structures syntaxiques ou des courbes intonatives n'ont de pertinence que dans la mesure où toutes les unités sont traitées.

L'exploitation des *métadonnées* associées par l'analyste au corpus nu nécessite que ce surplus d'information soit enregistré d'une façon ou d'une autre; on parle d'*annotation*. L'annotation ajoute ainsi une couche d'abstraction aux données, regroupant les différentes occurrences de mots ou de structures dans des ensembles plus ou moins homogènes. L'objectif de pareilles annotations est de permettre l'extraction de données en fonction d'une question spécifique. Ainsi, pour ne prendre qu'un exemple, un étiquetage complet en « parties du discours » permet de sélectionner, par exemple, toutes les phrases

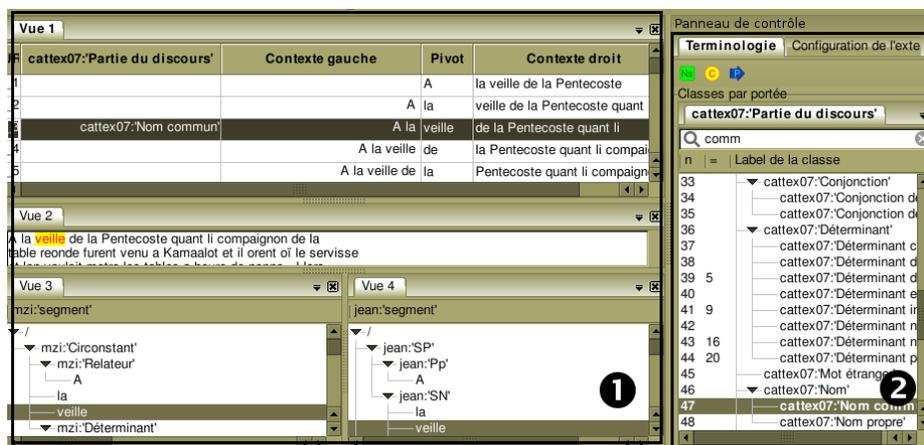


Figure 1. Environnement de travail

qui comportent un nom directement suivi d'un adverbe, ou un verbe directement précédé d'un pronom, etc. Ce type de sélection peut servir à porter au regard du chercheur des faits singuliers, qui méritent un commentaire ou une étude approfondie. D'autre part, une fois couplées à des critères extralinguistiques, comme la date de composition, le nom de l'auteur ou le type de texte, les annotations linguistiques permettent des études lexicométriques (relevant des statistiques textuelles) utiles pour étudier, par exemple, l'attribution des œuvres ou le style d'un auteur.

Besoin d'un outil

Les principales entreprises de construction de corpus de français médiéval, rassemblées en un groupe nommé *Consortium* pour les corpus de Français médiéval (<http://ccfm.ens-lsh.fr/>), ont choisi de partir des recommandations de la *Text Encoding Initiative* (Sperberg-McQueen, Burnard, Bauman, DeRose & Rahtz, 2007), c'est-à-dire, entre autres, d'encoder les documents en XML.

Les membres de cette communauté font relativement peu usage de logiciels pour mener à bien leurs études, qu'elles soient qualitatives ou quantitatives. Les relevés sont généralement faits manuellement et les annotations sont, le cas échéant, encodées directement dans le fichier sans interface spécifique, si bien que les corpus d'ancien français sont généralement peu étiquetés. La situation n'a que peu évolué par rapport à celle que décrit Habert, Nazarenko et Salem (1997, pp. 125-134).

Certains logiciels sont parfois employés pour aider à l'annotation des documents; par exemple, le *Système d'annotation de texte par ordinateur* SATO (<http://www.ling.uqam.ca/sato/satoman-fr.html>), qui permet l'annotation au niveau du mot. Malheureusement, le modèle de données de SATO ainsi que son implémentation sous la forme de service web rendent parfois ce travail extrêmement laborieux, en particulier quand il est question d'étiqueter *tous* les mots d'un texte. Il n'est par ailleurs pas adapté à l'annotation des relations syntaxiques. D'autres logiciels pourraient être utilisés dans ce cadre – par exemple, *WordFreak* (<http://wordfreak.sourceforge.net>) ou encore le *Linguistic Tree Constructor* (<http://ltc.sourceforge.net/index.html>) –, mais cela impliquerait de nombreuses difficultés d'interfaçage entre les différents formats de fichier annotés.

La situation peut se résumer comme suit :

- il n'existe aucun environnement d'*annotation manuelle* intégrant les fonctionnalités généralement requises;
- les formats employés pour définir les jeux d'étiquettes et la manière dont les documents sont annotés sont à chaque fois idiosyncratiques.

Partant, nous avons décidé de mettre au point NotaBene, qui se définit, déjà au stade du prototype que nous présentons ici, comme un environnement d'annotation manuelle de documents textuels XML, fondé sur les standards RDF et OWL. Enfin, bien que nous ne traiterons pas cet aspect, NotaBene a été conçu pour être extensible.

Section 2 : Annotations multiples

Une des premières conséquences du champ d'étude dans lequel NotaBene a été conçu est la multiplicité des manières d'analyser un même matériau. Si les modèles linguistiques foisonnent, les sous-branches interconnectées sont également pléthoriques. Ces dernières, bien que relevant de domaines distincts, sont reliées dans une dialectique complexe.

Ainsi, la syntaxe de la phrase, la morphologie verbale et la manière dont les verbes déterminent la forme de leurs compléments doivent pouvoir être mises en relation. Par exemple, dans la phrase présentée ci-dessus, l'emploi du circonstant *demain* est en relation avec le futur du verbe; par ailleurs, la forme et la présence du complément *Paris* est déterminée par le choix du verbe *aller*. Dans le cadre d'une analyse fine, les différents niveaux s'interconnectent « naturellement » pour le linguiste.

Principe général

Ces informations sont souvent encodées séparément avant d'être exploitées. L'annotateur n'a généralement pas besoin de connaître la fonction syntaxique

de *demain* ou d'*ira* pour indiquer qu'il s'agit respectivement d'un adverbe et d'un verbe; avoir accès au reste de la phrase n'a pas forcément d'intérêt. Néanmoins, le recours à d'autres niveaux d'analyse est sporadiquement requis, notamment pour lever les ambiguïtés syntaxiques. Une vue du texte sous forme de liste ordonnée de mots permet d'en faire une lemmatisation ou une catégorisation grammaticale rapide, mais il faut que la structure syntaxique soit également visualisable dans certains cas, sachant que ce dernier type de vue est trop complexe pour permettre une annotation rapide des lemmes.

On peut voir un exemple d'affichage simultané d'analyses différentes dans NotaBene dans la Figure 2.

À gauche se trouve une analyse syntaxique arborescente en cours d'élaboration. À droite se trouve la même phrase, affichée sous la forme de concordance (où chaque mot est présenté avec les mots qui le précèdent ou le suivent directement, c'est-à-dire ses contextes de gauche et de droite), et dont les mots ont été classés suivant leur classe morphosyntaxique.

Exemples de calibrage de visualisation

Prenons quelques exemples simples. Imaginons que le texte nommé *La quête del saint Graal* (texte daté de ca1230; édition en cours par Christiane Marchello-Nizia), soit en train d'être annoté par quatre équipes différentes :

- une première équipe en ferait l'annotation morphosyntaxique;
- une seconde équipe annoterait les mots désignant les personnages;
- une troisième équipe en ferait l'annotation syntaxique;
- une quatrième équipe en ferait également l'annotation syntaxique, mais suivant un modèle théorique différent.

Soit quatre ensembles d'annotation différents, qui peuvent ou non être mobilisés simultanément par les utilisateurs. Voyons ce qu'il en est au travers des trois exemples suivants.

Exemple 1. Pour commencer, l'annotation morphosyntaxique baptisée *CATTEX*, appliquée aux textes de la Base de Français Médiéval (<http://bfm.ens-lsh.fr>), considère chaque mot selon la classe morphosyntaxique à laquelle il appartient (on parle de « parties du discours » en grammaire traditionnelle). Pour mener à bien sa tâche, l'annotateur n'a besoin que de visualiser chacun des mots et son contexte immédiat sous la forme d'une concordance. Il n'est pas nécessaire qu'il ait accès aux trois autres couches d'annotation. Ainsi, une visualisation simple comme la suivante suffit amplement (Figure 3).

IR	partie du discours	Contexte gauche	Pivot	Contexte droit
1	Adverbe		Demain	, Jean ira à Paris
3	Nom	Demain ,	Jean	ira à Paris .
6	Nom	Demain , Jean ira à	Paris	.
2	Ponctuation	Demain ,	,	Jean ira à Paris .
7	Ponctuation	, Jean ira à Paris	.	
5	Préposition	Demain , Jean ira à	à	Paris .
4	Verbe	Demain , Jean	ira	à Paris .

Figure 2. Deux approches différentes d'une même donnée

URI	mzi:'partie du discours'	Contexte gauche	Pivot	Contexte droit
#_1	cattex:'Préposition'		A	la veille de la Pentecoste
#_2	cattex:'Déterminant article défini'		A	veille de la Pentecoste quant
#_3	cattex:'Nom'		A la	veille de la Pentecoste quant li
#_4	cattex:'Préposition'		A la veille	de la Pentecoste quant li compaignon
#_5	cattex:'Déterminant article défini'		A la veille de	Pentecoste quant li compaignon d
#_6	cattex:'Nom'		A la veille de la	Pentecoste
#_7	cattex:'Conjonction'	la veille de la Pentecoste	quant	li compaignon de la
#_8	cattex:'Déterminant article défini'	veille de la Pentecoste quant	li	compaignon de la table reonde
#_9	cattex:'Nom'	de la Pentecoste quant li	compaignon	de la table reonde furent
_10	cattex:'Préposition'	Pentecoste quant li compaignon	de	la table reonde furent venu
_11	cattex:'Déterminant article défini'	entecoste quant li compaignon de	la	table reonde furent venu a
_12	cattex:'Nom'	quant li compaignon de la	table	reonde furent venu a Kamaalot
_13	cattex:'Adjectif'	li compaignon de la table	reonde	furent venu a Kamaalot et
_14	cattex:'Verbe'	compaignon de la table reonde	furent	venu a Kamaalot et il
_15	cattex:'Verbe'	de la table reonde furent	venu	a Kamaalot et il orent
_16	cattex:'Préposition'	la table reonde furent venu	a	Kamaalot et il orent oï
_17	cattex:'Nom de lieu'	table reonde furent venu a	Kamaalot	et il orent oï le
_18	cattex:'Conjonction'	reonde furent venu a Kamaalot	et	il orent oï le servisse
_19	cattex:'Pronom'	furent venu a Kamaalot et	il	orent oï le servisse et
_20	cattex:'Nom'	venu a Kamaalot et il	orent	oï le servisse et len

Figure 3. Concordance (annotation morphosyntaxique)

On voit que la deuxième colonne de la concordance (la première colonne contient la référence correspondant au mot) contient un ensemble d'annotations sous la forme *cattex* : 'Préposition', *cattex* : 'Nom', etc. Ces « parties du discours » sont associées aux mots de la colonne *Pivot* de la ligne

correspondante. Comme on le voit, un contexte très limité suffit à sélectionner la classe adéquate. Par ailleurs, ce type de vue permet des tris aisés sur les différentes colonnes, en vue d'une annotation plus rapide ou de corrections.

Exemple 2. En guise de deuxième exemple, voyons ce que nécessite l'annotation des mots désignant les personnages du texte – projet de thèse de Stéphanie Obry, ENS-LSH Lyon. Les noms et surnoms des personnages, mais également les unités linguistiques telles que les pronoms ou les adjectifs employés comme vocatifs sont autant d'unités susceptibles d'être annotées. Il est évident que les mots pronominaux, du fait de leur potentiel anaphorique, ne portent pas les marques qui permettent de repérer à coup sûr à quel personnage ils font référence. Par ailleurs, l'homonymie existant entre les pronoms de la troisième personne et les déterminants articles définis (*le, la*) oblige à prendre en compte le contexte immédiat. Ainsi, pour annoter les occurrences du mot *le*, l'utilisateur aura besoin de deux vues : une première vue permettant de discerner rapidement les pronoms des déterminants (par exemple, une concordance) et une seconde vue permettant d'associer les pronoms à leur antécédent (par exemple, le texte sous une forme courante).

Dans cette capture d'écran (voir Figure 4), on peut voir que la troisième ligne de la concordance (au-dessus) a comme pivot un pronom, ce qui n'est pas le cas de la première et de la quatrième ligne. En tant que pronom, le mot fait potentiellement référence à un personnage, en l'occurrence, Lancelot, ce qui est aisément dégagé d'une vue d'ensemble du passage où figure le pronom à annoter (en-dessous).

Exemple 3. Imaginons à présent que les équipes travaillant sur la base de modèles syntaxiques différents aient besoin de comparer les résultats de leurs analyses pour évaluer l'adéquation de leur modèle aux données analysées. Dans ce cas, il est nécessaire d'afficher simultanément les deux analyses (voir Figure 5).

Les deux analyses concurrentes suivent, à gauche, un modèle inspiré de Lucien Tesnière (1965), et, à droite, l'analyse en constituants immédiats (voir p. ex. Gleason 1969). Comme on le voit, le découpage des unités et la manière dont elles sont groupées sont fondamentalement différents, de même que la valeur des étiquettes qui sont placées sur les structures.

On dispose ainsi d'un moyen visuel de comparer les analyses. En recherchant et en alignant les différentes analyses gravitant autour d'un mot en particulier, on est en mesure de déterminer laquelle des deux modélisations convient le mieux aux données.

IF	vo:personnage	Contexte gauche	Pivot	Contexte droit
72		ele descent et vient devant	le	roi si le salue ,
75		vient devant le roi si	le	salue , et il dist
10	vo:Lancelot	fet li rois , veez	le	la . " Si li
37		je vos di de par	le	roi Pellés que vos avec

Project 3

table reonde furent venu a Kamaalot et il orent oï le servisse
 et len vouloit metre les tables a heure de nonne , ! lors
 en sale une mout bele damoisele , et fu venue si grant oirre
 que bien le pooit len veoir , car ses chevaux en fu encore
 toz suanz , et ele descent et vient devant le roi si le
 salue , et il dist que Diex la beneïe . " Sire , fet ele ,
 por Dieu dites moi se **Lancelot** est ceenz . - Oïl voir , fet
 li rois , veez **la** . " Si li mostre , et ele va maintenant
 la ou il est et li dit : " **Lancelot** je vos di de par le
 roi Pellés que vos avec moi venez jusqu' en cele forest . "
 Et il li demande a qui ele est . " Je sui , fait ele , a
 celui donc je vos paroil . - Et quel besoin , fet il , avez
 vos de moi ? - Ce verniz vos bien " . fet ele . " De par

Figure 4. Concordance et texte courant (référence aux personnages)

Eichier Vues Vue active Aide		Visualisation Nicolas		Visualisation Jean	
segment	partie du discours	segment	partie du discours	segment	partie du discours
▼ /		▼ /		▼ /	
▼ Circonstant		▼ SP		▼ SP	
Demain	Adverbe	Demain	Adverbe	Demain	Adverbe
▼ Sujet		▼ SN		▼ SN	
Jean	Nom	Jean	Nom	Jean	Nom
▼ Prédicat		▼ SV		▼ SV	
ira	Verbe	▼ SP		▼ SP	
▼ Adjet		▼ Pp		▼ Pp	
▼ Relateur		à	Préposition	à	Préposition
à	Préposition	▼ N		▼ N	
Paris	Nom	Paris	Nom	Paris	Nom
.		▼ V		▼ V	
		ira	Verbe	ira	Verbe

Figure 5. Analyses syntaxiques concurrentes (heuristique épistémologique)

Section 3 : Conceptualisations multiples

En sciences humaines les annotations sont le reflet d'une conceptualisation subjective de l'objet observé. Très souvent, comme on vient de le voir dans l'exemple associé à la Figure 5 ci-dessus, deux annotateurs de formation proche construisent deux ensembles différents de concepts. Un « inévitable éparpillement des étiquetages » (Habert, Nazarenko & Salem, 1997, pp. 23-24) se forme ainsi, faute d'un formalisme unificateur. Cette difficulté déteint fatalement sur les possibilités de comparaison entre les modèles. Pour pallier ce manque, nous avons décidé de choisir un langage d'expression unique de ces modèles (métamodèle), qui nous permet de représenter de manière similaire des conceptualisations différentes. Ainsi, les jeux d'étiquettes correspondant aux analyses syntaxiques concurrentes dont nous venons de parler peuvent être exprimés comme le montre la Figure 6.

En effet, dans la perspective de la mise en relation des différentes analyses, la visualisation simultanée de celles-ci est un premier pas, mais ne nous paraît pas suffisante. Il y a une nécessité méthodologique d'ordre collectif de documenter les schémas d'analyse employés, la structure des classes de concepts mobilisés et la manière dont ceux-ci interagissent avec les données. En d'autres termes, il faut que le chercheur puisse décrire son vocabulaire dans un langage qui puisse être compris par d'autres. Ce vocabulaire est désigné sous le nom d'*ontologie*.

La représentation présentée dans la Figure 6 est une manière de visualiser les vocabulaires formalisés à l'aide d'un langage spécifique d'une manière adaptée à la l'interaction entre l'humain et la machine. Le choix de NotaBene au niveau du langage formel est d'employer le *Web Ontology Language* (OWL, voir Bechhofer, Van Harmelen, Hendler, Horrocks, McGuinness, Patel-Schneider & Stein, 2004) pour assurer la communication entre les entreprises d'analyse. Ce langage a de multiples qualités, dont les plus intéressantes dans le cadre ici défini sont :

- plusieurs degrés de formalisation de l'ontologie sont possibles et varient du plus lâche au plus strict (Bechhofer et al., 2004, §8);
- OWL exploite le potentiel des espaces de nommage (Bray, Hollander, Layman & Tobin, 2006), qui délimitent précisément chaque ontologie, de façon à ce que plusieurs d'entre elles puissent interagir ou être comparées sans introduire d'équivoque;

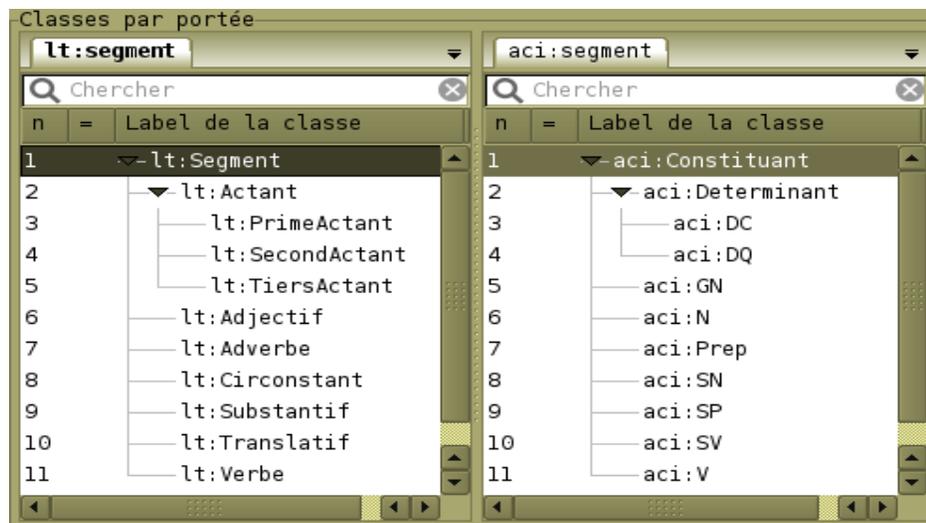


Figure 6. Jeux d'étiquettes correspondant aux analyses concurrentes

- OWL, conçu pour être exprimé en XML, peut être traduit dans n'importe quel langage de représentation de triplets (association entre trois éléments ordonnés) RDF (Klyne & Carroll, 2004, en partic. §3).

Cela dit, l'adoption du langage OWL ne signifie pas que nous adhérons à l'idée couramment véhiculée avec lui que les ontologies sont formulées *a priori* et sont nécessairement consensuelles (Laublet, Charlet & Reynaud, 2007, p. 105). Dans le cadre de la linguistique historique et, plus généralement, des sciences humaines, le travail scientifique implique un constant effort de conceptualisation nouvelle et de révision des concepts anciens. Techniquement, les ontologies OWL sont évolutives et mutualisables. Dans cette optique, il est important que l'utilisateur puisse modifier lui-même la terminologie d'une ontologie simple, comme le montre la Figure 7.

En refusant le figement, nous adoptons le point de vue selon lequel les ontologies sont dynamiques et conçues suivant une dialectique permanente (Iacovella, Bénel, Pétard & Helly, 2007, pp. 118-119). Toutefois, les ontologies ainsi élaborées peuvent être momentanément capturées pour permettre la comparaison de points de vues ou la mutualisation de la terminologie.

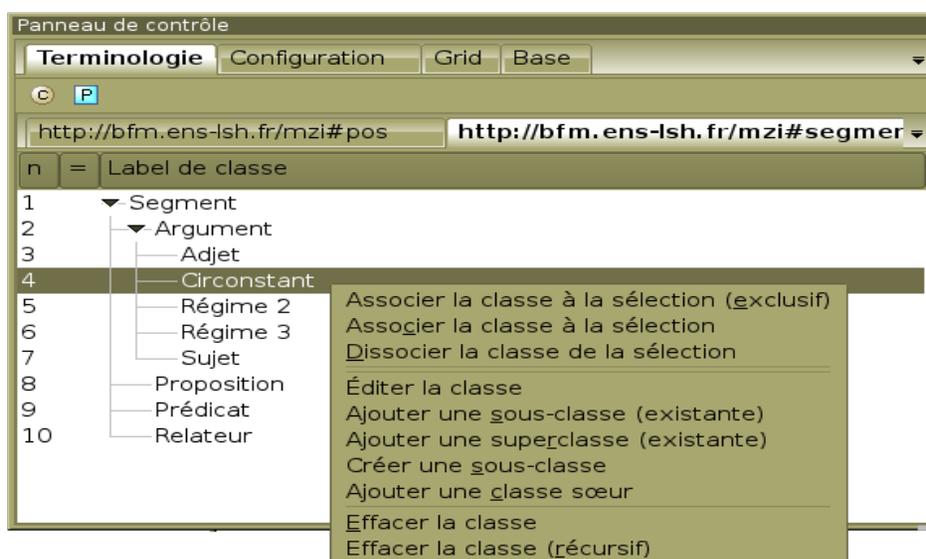


Figure 7. Éditeur d'ontologie intégré

La dernière version de travail de la spécification de NotaBene, qui détaille les formalismes choisis pour le logiciel, est accessible en ligne à l'adresse suivante : <http://notabene.svn.sourceforge.net/viewvc/notabene/trunk/doc/specification.pdf>.

Références

- Bechhofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., & Stein, L.A. (Éds). (2004). *OWL Web ontology language reference. Reference. W3C recommendation 10 february 2004*. Document consulté le 26 mai 2010 de <http://www.w3c.org/TR/2004/REC-owl-ref-20040210>.
- Bray, T., Hollander, D., Layman, A., & Tobin, R. (Éds). (2006). *Namespaces in XML 1.0 (2^e éd.)*. *W3C Recommendation 16 August 2006*. Document consulté le 26 mai 2010 de <http://www.w3c.org/TR/2006/REC-xml-names-20060816>.
- Gleason, H.A. (1969). *Introduction à la linguistique*. Paris : Larousse.
- Habert, B., Nazarenko, A., & Salem, A. (1997). *Les linguistiques de corpus*. Paris : Armand Colin.

- Iacovella, A., Bénel, A., Pétard, X., & Helly, B. (2007). Corpus scientifiques numérisés : savoirs de référence et point de vue expert. Dans R.T. Pédaque (Éd.), *La redocumentarisation du monde* (pp. 117-130). Toulouse : Cépaduès.
- Klyne, G., & Carroll, J. (Éds). (2004). *Resource description framework (RDF) : Concepts and abstract syntax W3C Recommendation 10 February 2004*. Document consulté le 26 mai 2010 de <http://www.w3c.org/TR/2004/REC-rdf-concepts-20040210/>.
- Laublet, P., Charlet, J., & Reynaud, C. (2007). Sur des aspects primordiaux du Web sémantique. Dans R.T. Pédaque (Éd.), *La redocumentarisation du monde* (pp. 99-116). Toulouse : Cépaduès.
- Sperberg-McQueen, C.M., Burnard, L., Bauman, S., DeRose, S., & Rahtz, S. (2007). *TEI P5. Guidelines for electronic text encoding and interchange. Version 1.0.1*. Document consulté le 26 mai 2010 de <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>.
- Tesnière, L. (1965). *Éléments de syntaxe structurale*. Paris : Klincksieck.

Nicolas Mazziotta est actuellement akademischer Mitarbeiter pour le projet Syntactic Reference Corpus of Medieval French à l'Université de Stuttgart. Il est titulaire d'un doctorat en Langues et littératures romanes (Université de Liège). Ses centres d'intérêt sont essentiellement les systèmes graphiques médiévaux, l'édition numérique et la syntaxe du français (en partic. médiéval). Parallèlement à ses recherches, il développe les logiciels nécessaires à l'annotation de structures syntaxiques et à l'exploitation de ces annotations. Ses contrats de recherche l'ont amené à collaborer avec les principales initiatives de constitution de corpus de français médiéval (Base de français médiéval à l'ENS-Lyon et Nouveau corpus d'Amsterdam à l'Université de Stuttgart).

Contribution d'un outil de transcription dans la mise en mots de l'analyse d'un processus de conception collaborative observé dans sa profondeur phénoménale

Béatrice Arend, Doctorante

Université du Luxembourg

Résumé

L'article relate comment une question de méthodologie d'analyse « comment retracer et visualiser un processus de conception collaborative dans sa profondeur phénoménale lors de la mise en mots? » a initié le développement d'un outil de transcription, support de traçabilité et de visualisation du processus analysé. Nous analysons la dynamique de modelage des formes langagières, gestuelles et artefactuelles dans une activité de conception collaborative instrumentée. En cours de recherche, en rapport avec notre posture épistémologique interactionniste et socio-constructiviste, a émergé un objectif préoccupant concernant le traitement et l'analyse des données, celui de visualiser l'épaisseur phénoménale des interactions verbales et non-verbales observées en situation. Nous montrerons au travers de l'analyse d'un extrait de corpus que la représentation visuelle du phénomène interactionnel analysé permise par l'outil de transcription développé contribue à la lisibilité de la mise en mots de l'analyse.

Mots clés

PROCESSUS DE CONCEPTION COLLABORATIVE, MULTIMODALITÉ, ANALYSE DES INTERACTIONS, ANALYSE CONVERSATIONNELLE, OUTIL DE TRANSCRIPTION¹, VISUALISATION DU PHÉNOMÈNE

Cadre de recherche

L'analyse d'un processus de conception collaborative dans son épaisseur phénoménale est le sujet de recherche. Le phénomène observé et analysé dans le cadre de ma thèse² consiste en une situation d'écriture conjointe médiée par ordinateur. Nous avons demandé à quatre enseignants d'expertises et d'approches différenciées, par rapport à leur métier d'enseignant et par rapport à l'artefact texte, d'écrire conjointement une fiction destinée à leurs élèves. Les quatre acteurs ont été placés en situation de rédaction conversationnelle. La séance de travail était instrumentée, la consigne prévoyant que le texte soit tapé

RECHERCHES QUALITATIVES – Hors Série – numéro 9– pp. 95-108.

LOGICIELS POUR L'ANALYSE QUALITATIVE: INNOVATIONS TECHNIQUES ET SOCIALES

ISSN 1715-8702 - <http://www.recherche-qualitative.qc.ca/Revue.html>

© 2010 Association pour la recherche qualitative

sur ordinateur dont l'écran était projeté sur un tableau blanc. Les enseignants rédigent la narration de manière collaborative. Ils sont installés à une table sur laquelle est placé l'ordinateur portable utilisé. La distribution de la tâche de scribe a été négociée au préalable.

Notre analyse de la séance collaborative s'inscrit dans une approche épistémologique interactionniste et socio-constructiviste de la cognition humaine, en accord avec une approche dialogique de construction de sens se matérialisant dans l'artefact texte considéré comme la concrétisation de la cognition dite située et distribuée des quatre enseignants. Nous considérons l'interaction de cette élaboration conjointe comme une dynamique de modelage de formes langagières, gestuelles et artefactuelles. « Cet ensemble de formes constitue la matière sémiotique conjointe travaillée par les interactants » (Brassac & Arend, 2007a, p. 5).

Nous abordons le processus d'écriture conjointe dans sa multimodalité en soulignant les dire et les faire des acteurs et nous relevons les caractéristiques propres à une séance de rédaction conversationnelle en tenant compte à la fois des discours oraux, des gestes et de la mobilisation des objets. « [Les enseignants] agissent en parlant et en mobilisant des objets [et] accomplissent un ensemble d'activités qui prennent des significations en égard à leur tâche commune » (Gregori & Brassac, 2001, p. 22).

Nous posons qu'il y a des articulations dialogiques entre l'activité d'écriture et l'activité cognitive conjointe. Nous nous intéressons non seulement aux interactions entre acteurs humains, mais nous centrons notre analyse aussi sur les modalités d'interaction avec l'objet. Nous étudions le rapport que les acteurs entretiennent avec l'ordinateur via clavier et écran et nous soutenons que l'ordinateur est une entité qui prend sens dans l'interaction. En cours d'analyse des données, nous avons en effet relevé que l'ordinateur via l'écran et via le clavier mobilisé par les acteurs est constitutif du processus conversationnel. « Autrement dit, l'interaction telle que nous la concevons n'est pas une seule intersubjectivité, mais également une interobjectivité (...) » (Brassac, 2007b, p. 27).

L'ordinateur s'inscrit dans la dynamique interactionnelle. Si nous convenons que l'ordinateur agit comme médiateur dans la co-construction de sens et dans le processus d'élaboration du texte, nous lui voyons aussi le statut d'acteur. Nous nous référons ici à la perspective de Latour pour qui l'objet est acteur social qui interpelle l'acteur humain. « À chaque fois qu'une interaction dure dans le temps et s'allonge dans l'espace, c'est qu'on l'a partagée avec des non-humains (...) il faut traiter les objets comme des faits sociaux » (Latour, 2006, p. 96).

Nous postulons, au sens de Latour, que l'ordinateur, dans le cas de notre recherche, vient modifier une situation donnée en introduisant « ... une différence dans le déroulement de l'action ... (et qu'il) existe une preuve qui permette à un observateur de détecter cette différence » (Latour, 2006, p. 103).

Analyser un processus de conception collaborative dans une posture interactionniste en adoptant une approche dialogique de construction de sens demande de recourir à une méthodologie qui permette d'accéder à et de visualiser ce qui « se réalise, se matérialise, sous forme d'interaction (... et qui) ne se situe pas quelque part à l'intérieur ... » (Bakhtine, 1977, p. 38).

Méthode

Les conduites verbales et non-verbales des enseignants élaborant une fiction destinée à être lue par leurs élèves respectifs ont été enregistrées en situation et en temps réel par captation audiovisuelle, le traçage vidéographique de cette séance de travail instrumentée servant à établir l'histoire rédactionnelle du processus dynamique de conception du texte. L'enregistrement total compte 1 heure 52 minutes et 14 secondes. Le traçage vidéographique in situ nous permet de voir dans son épaisseur phénoménale l'interaction entre les échanges verbaux oralisés des acteurs humains (nommés A, B, C, D) et la verbalisation graphique via l'activation du clavier par C et, synchroniquement, l'affichage sur l'écran.

Sur le plan méthodologique nous nous sommes interrogés sur le *comment* analyser le phénomène en rendant compte de son épaisseur capturée en images et en sons réels :

- Comment, -pour relater son analyse, le chercheur aura recours au mode narratif et sera tenu à la linéarité-, pointer dans le narratif³ la médiation instrumentalo-sémiotique en relevant à la fois la synchronisation voire la simultanéité dans les formes de modelage et la modélisation dans sa chronologie?
- Comment conférer à la mise en mots linéaire la profondeur phénoménale des images et du son et visualiser, dans la description du phénomène analysé, que l'ordinateur, instrument et support de traçage, (inter)agit en cours de processus d'écriture en acteur non-humain?

Pour le traitement des données vidéographiques, nous avons recours à l'analyse conversationnelle, démarche méthodologique qui requiert un travail de transcription précis et laborieux, la transcription faisant partie de l'analyse dont elle est en même temps support.

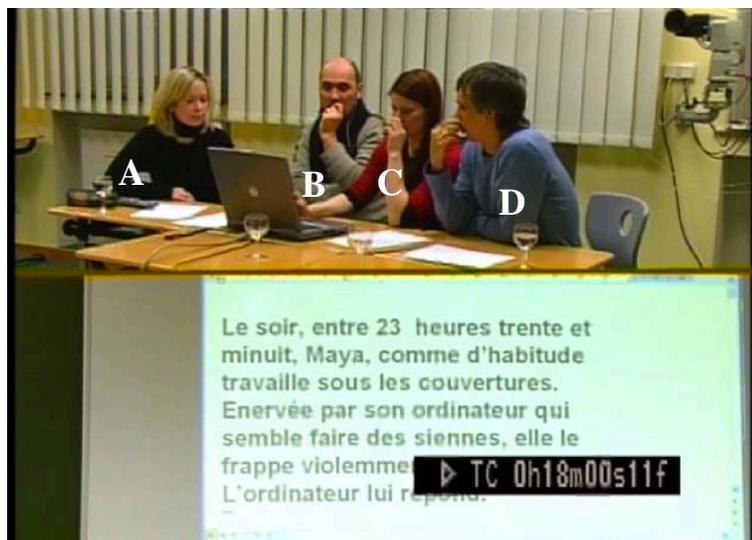


Figure 1. Traçage vidéographique d'une situation d'écriture conjointe

Comme nous ne traitons pas seulement les verbalisations, mais aussi les formes gestuelles et artefactuelles à la fois comme des constituants de l'interaction et constituant l'interaction, et comme nous postulons que l'ordinateur est inscrit dans cette dynamique interactionnelle en y participant, d'autres questions se sont posées :

- Comment tenir compte dans la transcription de ce que nous appelons la convocation récurrente de l'ordinateur par les acteurs humains et vice versa?
- Comment afficher dans la transcription le statut d'acteur non-humain à la fois constitutif du processus conversationnel et constituant la dynamique cognitive?

Souvent, la transcription retraçant prioritairement l'échange verbal entre les acteurs par l'attribution des tours de parole, annotée et pourvue de commentaires entre parenthèses pour relater les fautes ou la mimique, nous paraît surchargée voire difficilement lisible ou peu précise. Comme les gestes, la mimique, la posture des acteurs, les manipulations d'objets sont pour notre analyse tout aussi éloquentes que les occurrences langagières, nous étions à la recherche d'un outil de transcription permettant, à partir de la transcription, une visualisation graphique support de lisibilité de l'histoire conversationnelle

analysée, visualisation représentant les occurrences des formes de modelage dans leur simultanéité et dans leur chronologie.

C'est au travers de l'analyse d'un extrait choisi que nous relèverons les phénomènes dialogiques d'une situation de conception collaborative s'actualisant et actualisés dans les dire et les faire des acteurs humains et non-humains et que nous montrerons aussi comment l'outil de transcription confère plus de lisibilité à l'analyse.

Analyse empirique d'une séquence d'écriture conjointe et apport de l'outil de transcription pour cette analyse

A) Nous décrivons dans l'analyse de cet extrait comment des acteurs co-construisent du sens émergeant du et dans le modelage interactionnel des formes verbales et non-verbales et nous relèverons comment l'ordinateur participe au processus conversationnel.

Transcription de l'extrait choisi qui se situe entre 1h 42' 00'' et 1h 42' 42'' :

- 1 C 1h42'00'' et: May[a:: ((C écrit *ya*))
- 2 B 1h42'0.5'' [Maya: à la ferme ((C se tourne lentement vers B, étonnement dans le regard))
- 3 B 1h42'5'' () ((rire))
- 4 C 1h42'7'' je_ ((rire)) j'étais en train de [:_ [Maya Maya ou::
- 5 B 1h42'8'' [bon
- 6 D 1h42'9'' [Maya
informaticienne [((rire))
- 7 B 1h42'11''
[ou Maya:: la ferme c'est pour dire_ [pour sortir un peu des: &
- 8 A 1h42'12'' [oui/
- 9 D 1h42'15'' ou[i::
- 10 B 1h42'15'' & [des images qu'on a quoi tu vois/
- 11 C 1h42'16'' () ((expire)) ((C entame l'inscription de *la ferme*, les sourcils levés))
- 12 D 1h42'19'' oui moi je veux bien
- 13 B 1h42'21'' A la ferme pas non pas Maya la ferme ((rire))
- 14 C 1h42'22'' ((rire)) c'est pour ça/ à la [ferme ((C efface *la ferme*))
- 15 A 1h42'25'' ((rire)) [et moi aussi j'ai regardé
comme ça ((C se met en arrière et croise les bars))

- 16 C 1h42'28'' bah on va () maintenant ((rire)) oh pourquoi pas/
[c'est une expression . une expression qui eh:: qu'il faut [connaître ((C écrit à))
- 17 B 1h42'32'' [Maya à la ferme oui
- 18 A 1h42'36''
[()
- 19 A 1h42'37'' [((rire)) ((C écrit *la ferme*))
- 20 B 1h42'37'' [((rire))
- 21 B 1h42'39'' Maya à la [ferme . et puis elle c'est vrai
- 22 C 1h42'40'' ((inspire)) [oui voilà\

Les quatre enseignants construisent une fiction destinée à être lue par leurs élèves respectifs. Ils ont inventé un personnage principal Maya, une petite fille, qui est en communication avec un astronaute via ordinateur.

Dans la genèse du texte, à ce stade du travail d'écriture, Maya invitée par l'astronaute à s'imaginer sa vie future, voit dans les morceaux d'un miroir brisé les possibilités que lui offre l'avenir.

Les quatre acteurs A, B, C et D sont en train de négocier l'avenir du personnage principal de la fiction. Il y a consensus sur trois propositions affichées et fixées à l'écran dans le champ de vision des quatre enseignants.

A 1h 42' 00'', nous lisons sur l'écran *Maya star; Maya directrice de cirque; Maya joueuse de foot; Maya*

Avant que les enseignants n'approuvent une quatrième identité virtuelle de Maya et ne la fixent, il y a dans cette séquence un malentendu ou, en termes bakhtiniens, une dissonance à peine perceptible des voix se manifestant « pas tant dans les mots que dans les silences (...) dans un rire ... » (Bakhtine, 1970, p. 354).

Voilà qu'à 1h 42' 00'' B émet une nouvelle proposition (2)⁴. C se retourne vers B et son regard marque de l'étonnement. Elle n'écrit pas et B réagit à ce regard par un rire (3). C se met à rire à son tour et répète le nom *Maya* deux fois suivi d'un *ou* allongé comme si elle cherchait une proposition à son tour (4). Le *bon* de B (5) marque une légère impatience. D avance *Maya informaticienne*, et rit lui aussi (6).

Le clavier est toujours en repos et nous avançons que la dissonance se manifeste dans ce silence, actualisation du « combat dialogique aux frontières fluctuantes entre les mots d'autrui et les mots personnels » (Bakhtine, 1984, p. 364).

La scribe n'active pas le clavier, il n'y a pas encore consensus. Le pourquoi de la discorde n'est cependant pas visible. Pour Bakhtine, en tout acte de parole, l'activité mentale subjective se dissout dans le fait objectif de l'énonciation ayant pris forme, tandis que la parole énoncée se subjectivise dans l'acte de décodage qui doit tôt ou tard provoquer l'encodage d'une réplique (1977, p. 67).

Jusqu'alors C a « encodé sa réplique » en regard étonné, en rires, en ébauche de nouvelle proposition (4), mais elle n'a pas encore répondu en termes d'activation du clavier c'est-à-dire en termes de signes graphiques. N'approuvant pas la contribution de B en matière sémiotique, C n'exécute pas sa tâche de scribe et tant que la proposition n'est pas inscrite il y a fluctuation de sens.

B justifie l'idée émise en disant que c'est pour sortir des images qu'ils avaient (7, 10). A et D marquent leur accord (8, 9) et D confirme par un *moi je veux bien* (12). La forme forte du pronom personnel et le verbe *vouloir* marquent ici non seulement l'accord avec B, mais la volonté que cette proposition soit fixée.

Ce durant, C entame l'inscription (11) en expirant audiblement, les sourcils levés, ce que nous avons lu comme un autre signe de réticence.

Jusqu'ici, (1h 42' 16''), C n'a pas encore explicité verbalement cette réticence face à la proposition de B, elle affiche son désaccord par sa posture, sa mimique, en ne répondant pas de suite à sa tâche de scribe.

En 1h 42' 20'', C « encode sa réplique » en tapant *la ferme*, expression comportant « [...] deux facettes : le contenu (intérieur) et son objectivation extérieure pour autrui » (Bakhtine, 1977, p. 121).

En 1h 42' 21'' l'écran affiche *Maya la ferme* et B agit immédiatement en reformulant *à la ferme* et en invalidant par un *non* ce que C avait inscrit (13). Il y a des rires et C efface de suite *la ferme*. B voit dans l'inscription du signe sur l'écran la représentation conceptuelle que C avait construite autour de l'expression qu'elle croyait avoir entendue⁵. Dans le signe extériorisé sur l'écran, ici dans *la ferme*, s'actualise l'acte de parole dans sa dimension dialogique; c'est la matérialisation des deux faces du mot qui est « [...] déterminé tout autant par le fait qu'il procède de quelqu'un que par le fait qu'il est dirigé quelqu'un. Il constitue justement le produit de l'interaction du locuteur et de l'auditeur. » (Bakhtine, 1977, p. 123).

L'ordinateur agit comme médiateur en affichant la contribution de C et conjointement le malentendu, le pourquoi de la dissonance. L'ordinateur rend visibles les deux extrémités du mot se croisant et émergeant dans le signe

affiché et parle ainsi aux autres acteurs. Il fait voir que « si le mot prend appui sur moi à une extrémité, à l'autre extrémité il prend appui sur mon interlocuteur » (Bakhtine, 1977, p. 123).

La pensée de C se réalise dans le mot extériorisé sur l'écran et devient ainsi accessible aux autres acteurs :

la pensée (...) se réalise dans le mot. C'est pourquoi on pourrait parler d'un devenir de la pensée dans le mot. (...) Ce déroulement de la pensée s'opère sous la forme (...) d'un passage de la pensée dans le mot et du mot dans la pensée (Vygotsky, 1985, p. 329).

Si le passage de la pensée dans le mot et du mot dans la pensée est médié par l'ordinateur, ce dernier initie aussi une transformation et participe activement à la co-rédaction : B est interpellé et procède à une invalidation et à une reformulation de l'expression *la ferme*. Au sens de Latour, l'ordinateur ici figure parmi les acteurs qui⁶ « ... font quelque chose, [qui] ne sont pas seulement les écrans ou les rétroprojecteurs de notre vie sociale » (Latour, 2006, p. 96).

Dans cette dynamique interactionnelle, l'ordinateur via le clavier et via l'écran est un constituant de l'interaction faisant émerger et fixant le sens co-construit. Il participe ici à l'échange dialogique et est un acteur parmi les autres dans ce lieu de l'intercommunicabilité. L'ordinateur participe à la stabilisation de sens; il porte et visualise le flux du sens et le sens en fluctuation.

C réagit à l'invalidation de l'expression en l'effaçant. Ensuite, en riant (14), C se met néanmoins en arrière sur sa chaise et croise les bras; une scribe ayant les bras croisés ne peut répondre à sa tâche. C s'est visiblement distanciee de l'ordinateur et par ce biais de sa tâche. Le silence du clavier est révélateur de la posture de C par rapport à sa tâche et par rapport aux autres acteurs. Elle explique et justifie encore (16) ce qui est déjà effacé avant de déplacer sa chaise légèrement en direction de l'ordinateur, de reposer les doigts sur le clavier et de taper dans une première phase *à*, ce petit mot qui manquait et dont l'absence était révélatrice. 9 secondes plus tard seulement elle tape *la ferme*, elle n'a pas écrit l'expression entière en un seul flux. A et B rient toujours. C clôture cette séquence en émettant *oui voilà*. C'est à cet instant qu'est stabilisé le consensus négocié et approuvé par tous les acteurs.

Impliqué dans cette activité d'écriture, l'ordinateur « transporte les encodages » et conjointement les « décodages » des acteurs et véhicule la matérialisation du processus de sémiotisation dialogique. Toutefois, l'objet mobilisé n'est pas que médiateur. En fonction de ses possibilités d'engagement, l'ordinateur contribue à construire le processus interactionnel étant un des acteurs de ce processus qui interpelle les acteurs humains en ce

sens que le rapport dialogique entre l'énonciation en cours et l'énoncé en transformation émerge dans l'inscription.

B) Toute transcription relève du traitement de données et se révèle être une interprétation. Selon l'importance que le chercheur décide d'accorder aux formes non verbales (selon sa posture épistémologique), il organise la transcription de l'extrait à analyser. Dans une première transcription, j'avais décidé de relater les dires des acteurs et de mettre en commentaire entre parenthèses les gestes, la mimique et l'activation du clavier par C. Un choix nullement arbitraire comme j'analyse le processus interactionnel dans son modelage de formes verbales, gestuelles et artefactuelles.

Lors de la mise en mots de l'analyse de l'extrait, nous avons cependant trouvé que la transcription manquait de « profondeur » et nous avons souhaité visualiser davantage le phénomène relaté.

Ainsi nous avons pensé dresser un tableau qui permettrait de voir en un coup d'œil les formes de modelage (langagières, gestuelles et artefactuelles), avec l'indication de leur actualisation dans le temps. Nous avons opté de relever dans ce tableau les formes par lesquelles mettre en évidence la réticence de C d'inscrire la proposition de B de même que l'émergence du malentendu et l'importance de l'activité respectivement du silence du clavier dans cette interaction.

Nous n'étions cependant pas satisfaits du résultat : nous aurions dû conformément à notre épistémologie tenir compte de tous les acteurs et prévoir en conséquence l'organisation du tableau. Un deuxième essai dans cette direction a donné un tableau difficilement lisible qui n'a guère apporté d'élément nouveau par rapport à la transcription produite. Nous avons mis dans les lignes horizontales les indications de temps et en colonnes verticales les acteurs; présentation abordable si les occurrences langagières sont minimales et là encore en format A3.

En fait notre problème relevait de l'exigence suivante : concilier voire faire fusionner deux présentations à format différent, une transcription en format texte et une transcription en partition. La lecture en partition permet une lecture à deux entrées des interactions; une lecture de la synchronisation des interactions des différents acteurs sur l'axe vertical et en même temps une lecture de l'échange interactionnel dialogique dans son déroulement chronologique sur l'axe du temps horizontal. Une discussion avec l'expert développant le programme de transcription a permis que l'outil offre, à partir de notre transcription, sur simple click dans le menu une représentation graphique en partition. Nous avons la possibilité de souligner visuellement le geste, le regard, la mobilisation de tel objet selon notre perspective de

Tableau 1
Résultat de la démarche méthodologique

B	C scribe	Ordinateur (écran, clavier)
1h 42' 0,5'' Maya à la ferme	((regard étonné, autre proposition, rire))	1h 42' 20'' la ferme
1h 42' 21'' A la ferme pas non pas Maya la ferme	((efface)) ((rire)) ((bras croisés ensuite))	1h 42' 23''
	((remet les mains sur le clavier))	1h 42' 28'' à
	((rire, explication, scription))	1h 42' 37'' à la ferme
	1h 42' 40'' ((inspire)) oui voilà	

recherche. Dans la séquence analysée ci-dessus, ce sont plus particulièrement les interactions entre les acteurs humains et l'acteur l'ordinateur que nous voulons mettre en évidence.

Les formes de modelage figurent en « boîtes » bleues et rouges, les « boîtes » rouges représentant les contributions verbales de C et l'activité du clavier respectivement l'affichage en mouvement sur l'écran de l'ordinateur. Les « caisses » bleues signalent les occurrences langagières de A, B et D; à chaque fois la numérotation du tour de parole renvoie à la transcription.

Cette représentation graphique en partition sert de support de visualisation et de lecture à la transcription format texte. Sur le graphique apparaît notamment (pour ne relever qu'un passage clé de cette séquence) que B (13) agit de suite dès que l'écran affiche *la ferme* et que suite à l'intervention de B qui incite à la reformulation, il y a effaçage. Pendant un moment il n'y a plus trace de la représentation de C. L'expression du malentendu est effacée bien que C y revienne (16) en demandant *pourquoi pas*. Elle justifie encore une fois sa lecture de la proposition de B tout en tapant la préposition *à*. En même temps que verbalement elle reprend la première expression affichée, elle enchaîne sur un nouvel affichage, cette fois-ci celui de l'expression du consensus de tous les acteurs.

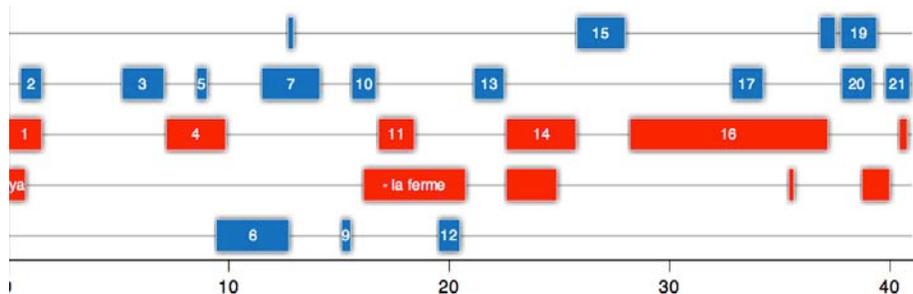


Figure 2. Voici une représentation possible de la séquence analysée^{7 8}.

Sur le graphique, nous pouvons suivre la migration de l'expression en mouvement et ainsi le flux de la construction de sens. Evidemment, le graphique à lui seul ne fait que soutenir notre analyse et ne peut pas être généré par le programme sans transcription préalable minutieuse.

Conclusion

Analyser, dans une posture épistémologique interactionniste en accord avec une approche dialogique de construction de sens, une activité de conception collaborative lors de laquelle les acteurs parlent, font des gestes et manipulent des objets c'est relever les formes verbales et non-verbales dans et par lesquelles s'actualise le processus interactionnel dans sa profondeur phénoménale. Le vidéotraçage permet de capter et de fixer les interactions dans leur synchronisation et dans leur chronologie.

Mais, à un moment de son analyse, le chercheur est amené à relater par écrit l'analyse du processus et se trouve alors « enfermé » dans la linéarité du mode narratif. Il a recours à la mise en mots sous forme de transcription et de description dense de la séquence analysée pour retracer et faire voir le phénomène. La méthodologie est censée à la fois faire accéder à la matérialisation langagière, gestuelle et artefactuelle de l'interaction et contribuer à visualiser cette extériorité dans toutes ses dimensions.

La troisième dimension n'y est toujours pas, mais c'est suite à ses réflexions que nous avons eu recours à un logiciel de transcription permettant, à partir de la transcription et la complétant ainsi, une visualisation graphique des interactions verbales et non-verbales des acteurs. C'est pour nous un support de retraçabilité et de lecture du processus interactionnel relaté.

Nous avancerons ainsi que ce programme⁹ nous a fait avancer dans notre objectif de conférer plus de lisibilité au retraçage de l'activité de conception collaborative analysée en ce sens que nous pouvons y faire figurer l'ordinateur comme acteur de la co-rédaction au même titre que les quatre enseignants.

Notes

¹ Henri Hansen, TranScripter : <http://www.transcripiter.lu>

² *Co-construction de sens : Élaboration conjointe d'une production écrite. Analyse du processus d'écriture.*

³ Nous recourons au format texte et au mode narratif pour relater notre analyse et pour en faire part. Mon objet de chercheur étant d'analyser le processus d'écriture conjointe des 4 enseignants, une partie importante de ma tâche consiste à relater les résultats d'analyse, et je me trouve alors à mon tour acteur dans une construction de sens médiée, acteur constituant du rapport dialogique entre le processus d'analyse et le processus de traçage scriptural de cette analyse.

⁴ Dans la transcription, nous avons écrit *à la ferme*, donc adapté l'orthographe à ce que nous avons entendu, ce qui s'est confirmé dans la suite. Toute transcription relevant d'une interprétation (nous soulignons tel geste, tel regard, nous marquons les pauses et les allongements tels que nous les avons entendus ...) nous avons choisi de ne pas adopter le regard de C ce que nous aurions pu faire en allongeant (avec une pause) la dernière syllabe du nom *Maya*. Nous convenons que c'est en connaissance de cause du sens qui s'est stabilisé à la fin de la séquence que nous avons fait la transcription de ce tour de parole.

⁵ B a fait l'élision en prononçant *Maya à la ferme* et C a compris *Maya la ferme*; elle s'étonnait de ce choix sémiotique : pourquoi demander au personnage de se taire et encore sur ce ton? C restait immobile en tant que scribe face à ce qu'elle croyait être un changement de registre de langue inexplicable. D'ailleurs, nous pensons que c'est probablement la rupture syntaxique qui a contribué au malentendu. Dans les 3 premières propositions, un groupe nominal (*star, directrice de cirque, joueuse de foot*) placé en apposition du nom propre *Maya* désigne la profession future virtuelle du personnage alors que dans la dernière proposition, la petite fille est vue dans une localité, *à la ferme*, groupe prépositionnel faisant fonction de complément circonstanciel renvoyant à un lieu.

⁶ Sans être un acteur humain doté d'une intentionnalité. « Un tel renversement de la direction dans laquelle s'exerce l'influence ne serait qu'une façon de transformer les objets en cause dont les effets seraient véhiculés par une action humaine. » (Latour, 2006, p. 103).

⁷ Si le programme, dans un rôle d'opérateur peut modifier la tâche du chercheur en ce sens qu'il modifie la façon de réaliser la tâche de transcription, une de nos préoccupations lors du processus de développement du programme a été toujours que le design n'impose pas de contraintes; ce n'est pas le programme qui oriente la démarche méthodologique, mais l'outil est un support construit dans et pour la réalisation de la tâche de transcription s'inscrivant dans notre cadre conceptuel.

⁸ Le programme permet de définir la dissolution et la taille de l'image; les bandes horizontales peuvent être déplacées : ici, nous avons choisi de suivre la dénomination A, B, C et D (acteurs de gauche à droite) et d'intercaler 'la bande clavier' entre B et C pour bien mettre en évidence les interactions mentionnées ci-dessus.

⁹ Sans qu'il (ré)oriente notre posture épistémologique (nous y insistons une fois de plus)

Références

- Bakhtine, M. (1970). *La poétique de Dostoïevski*. Paris : Editions du Seuil.
- Bakhtine, M. (1977). *Le marxisme et la philosophie du langage. Essai d'application de la méthode sociologique en linguistique*. Paris : Éditions de Minuit.
- Bakhtine, M. (1984). *Esthétique de la création verbale*. Paris : Gallimard.
- Brassac, C. (2007b). Co-responsabilité cognitive et dissolution des frontières. Dans P. Hert, & M. Paul-Cavallier (Éds), *Sciences et frontières. Délimitations du savoir, objets et passages* (pp. 159-176). Ferrelmont, BE : Éditions Modulaires Européennes & Intercommunications .
- Brassac, C., & Arend, B. (2007a). *Production située et distribuée d'une forme discursive à visée pédagogique : le cas d'une rédaction conjointe*. Symposium La dynamique des rapports entre individuel et collectif en formation : quels apports de l'approche de l'activité. Réseau éducation formation (REF), Sherbrooke, 23 au 27 septembre.
- Gregori, N., & Brassac, C. (2001). *La conception collaborative d'artefacts : activités cognitives en situation dialogique* (pp. 21-31). Actes des premières journées d'études en psychologie et ergonomie. Nantes, France, 29-30 octobre.
- Latour, B. (2006). *Changer de société, refaire de la sociologie*. Paris : Édition La Découverte.
- Vygotsky, L. (1985). *Pensée et Langage*. Paris : Messidor/ Éditions Sociales.

Béatrice Arend est enseignant-chercheur à l'Université du Luxembourg. Elle assure des cours sur l'apprentissage des langues dans la formation du Bachelor en sciences de l'éducation. Son travail de recherche porte sur les processus interactionnels et plus particulièrement sur les activités médiées instrumentées. Dans sa thèse qui est en voie de finalisation, elle analyse un processus d'écriture conjointe; le présent article relève essentiellement de la méthodologie et de l'épistémologie d'analyse.

Appendice 1
Conventions de transcription

:, ::	allongement d'une sonorité
., ..	pause
(3)	pause de plus de 2 secondes
&	suite d'un tour de parole
[chevauchement
()	inaudible
(())	description de gestes, de mimique et autre entre parenthèses
–	troncation
/	intonation montante
\	intonation descendante
	la majuscule marque une voix élevée

Innovations dans un logiciel d'analyse qualitative de données : l'intégration d'outils de visualisation

Udo Kuckartz, Ph.D.

Philipps-Universität Marburg

Résumé

La présentation et la visualisation de données sont des domaines de la recherche qui deviennent de plus en plus importants, en particulier dans les approches qualitatives et mixtes. Cet article couvre ces thèmes à partir de deux perspectives. Premièrement, l'attention est portée sur les questions méthodologiques et théoriques afin de démontrer que la visualisation est créative et utile pour l'exploration et la découverte de logiques, de relations et de structures dans les données. Des scientifiques ont montré de façon concluante que les informations complexes et les corrélations peuvent être traduites en des représentations graphiques facilement lisibles. Deuxièmement, la partie principale de l'article s'intéresse aux pratiques de recherche, et en particulier à l'utilisation de nouveaux outils de visualisation intégrés dans des logiciels d'analyse de données qualitatives, tel que MAXQDA. Une distinction majeure est opérée entre deux types de visualisation : les visualisations internes à un cas et les visualisations multi-cas. La présentation visuelle des données a plusieurs fonctions dans le processus d'analyse : l'agrégation, la réduction et l'exploration des données, la gestion des idées ainsi que la présentation des résultats.

Mots clés

VISUALISATION, DONNÉES QUALITATIVES, LOGICIELS, MAXQDA

Introduction

Cet article prend une perspective qui est, dans une certaine mesure, la contrepartie du point de vue dominant de la sociologie du visuel. Mon attention ne se porte pas sur le visuel, ou l'image, en tant qu'objet de la recherche et de la méthodologie en sciences sociales. Le visuel est approché plutôt de l'autre côté : je vais de la recherche vers le visuel. Plus spécifiquement, je me demande comment visualiser les connexions rencontrées à travers l'analyse qualitative de données et le gain que cela peut apporter au processus d'analyse qualitative.

Le sujet de ce papier concerne la recherche en tant qu'activité productrice d'images, par exemple la transformation de résultats de recherche et des processus de recherche dans le monde des images, tels que les formes, graphes, modèles, carte conceptuelle, etc. Ce sujet est présent dans mon esprit depuis plusieurs années et j'aimerais présenter une série de visualisations récemment développées qui ont été entretemps mises en œuvre à travers le logiciel d'analyse qualitative MAXQDA.

La visualisation a gagné en importance dans de nombreuses disciplines scientifiques tant comme outil d'évaluation que comme mode de présentation des résultats d'analyse. Les sciences sociales sont, par contre, une branche des sciences dans laquelle le travail et les arguments sont basés principalement sur des textes.

Contrairement aux sciences naturelles ou à la médecine, où il est de pratique courante de présenter les résultats de façon visuelle et, de plus en plus pour les diagnostics médicaux, en utilisant des techniques d'imagerie, la méthodologie en sciences sociales utilise ces possibilités de façon plus que modérée. Il existe très peu de littérature en méthodologie des sciences sociales qui traitent explicitement de possibilités de modélisation des résultats d'analyse (Banks, 2001; Blasius & Greenacre, 1998). Au mieux, nous le rencontrons plus régulièrement dans le cadre d'analyse de réseaux (par exemple : Krempel, 2005). La raison, telle que Feldmann (2003) la présente, est une aversion générale des sciences sociales envers les images. La sociologie est un enfant de la rationalité et des lumières, et est très sceptique quant aux images de par son héritage.

The holy Sociology writings of Marx, Durkheim, Weber, Parsons, Luhmann, Habermas, Giddens etc. are mostly image-less, seldom defiled with statistics. You should not make any images! Yet, not only the classics, but also the newer sociological writings that use older and newer media as well as the internet are largely image-free (Feldmann, 2003).

Visualisation et perception humaine

Dans son film concernant le changement climatique, Al Gore exprime ainsi que « le non-expert peut « voir » les relations et est impressionné par sa modélisation ». En effet, les représentations graphiques nous permettent de présenter des faits complexes de façon à les rendre facilement et rapidement compréhensible par tous. Seuls quelques scientifiques en sciences sociales se sont intéressés à ce sujet et à la question de comment configurer au mieux ces représentations.

Lothar Krempel représente une exception dans sa dissertation à propos de la visualisation en réseau, en ce qu'il a également abordé l'histoire de la visualisation de l'information (Krempel, 2005) : les cartes géographiques, dit Krempel, existent déjà en 2000 A.J.-C. et il est de notoriété que la cartographie a joué un rôle central dans le transport maritime. Tout qui voudrait savoir comment se rendre d'un endroit à un autre peut trouver facilement en regardant une représentation graphique. Imaginons un instant que nous devons transcrire en mots toute l'information contenue dans la carte du métro londonien (*The Tube*). Ou encore si, en tant qu'utilisateur du métro, nous avons à lire un texte qui décrit les connexions et les stations durant le trajet afin de trouver comment se rendre d'un point A à un point B. De telles cartes de métro ne représentent plus une topographie métrique; elles sont plutôt conçues spécialement pour les passagers. Par exemple, la distance entre les stations est toujours la même et, dès lors, pas d'échelle. Les lignes de connexion entre les stations sont toujours dessinées en ligne droite. Parfois, le caractère construit de telles représentations laisse la place à des conclusions erronées pour l'observateur. En l'occurrence, beaucoup d'utilisateurs du métro newyorkais supposent que Manhattan est beaucoup plus grand que Brooklyn, alors qu'en réalité Brooklyn fait trois fois la taille de Manhattan.

Ces exemples des cartes de métro montrent à quel point les visualisations intégrant de nombreuses et complexes informations peuvent être plus efficaces que l'écrit en terme de séquence et d'ensemble de caractères. Toutefois, un pré-requis pour une bonne représentation visuelle est la sélection des moyens adéquats pour celle-ci, tels que les couleurs, les symboles, etc. Des résultats en psychologie cognitive montrent que les humains emmagasinent des savoirs complexes sous forme de modèles mentaux. C'est pour cela que depuis le début, nous comprenons des modèles tels qu'une carte de métro mieux qu'un texte.

Visualisation et analyse qualitative des données assistée par ordinateur (CAQDAS)

Formes de visualisation

Dans la recherche quantitative, il est habituel de présenter les résultats de simples procédures statistiques tels que des graphes par exemple, sous forme de tartes ou de diagrammes ou d'histogrammes. Les formes de visualisation qui représentent des résultats complexes et des connexions sont utilisées plus rarement. Les exemples incluent les diagrammes de trajectoire ou des modèles de causalité linéaire, dans lesquels les connexions entre un grand nombre de variables – latentes ou manifestes – peuvent être représentées.

Au sein de l'analyse qualitative de données traditionnelle, nous trouvons encore moins de visualisation que dans l'approche quantitative. C'est dans la forme d'analyse assistée par ordinateur (souvent connue sous le terme CAQDAS) que nous commençons à percevoir un intérêt pour la question de la visualisation. Nous y trouvons quatre types différents de visualisation :

- Visualisations avec le logiciel QDA lui-même. Leur principale fonction est d'améliorer la prise en charge de données durant le processus d'analyse et de rendre les données et les résultats de l'analyse plus accessibles.
- Les cartes conceptuelles. Tout comme les diagrammes de structure linéaire dans la recherche quantitative, elles permettent d'illustrer les connexions entre des concepts, des catégories, des mémos, ainsi que les interconnexions et les relations causales.
- Les visualisations interne à un cas (*within-case display*). De telles représentations présentent des séquences chronologiques (comme par exemple dans une recherche biographique) ou illustrent l'ordre et l'interconnexion des codes et des concepts. Le terme « cas » peut se référer à une variété de choses selon le matériau empirique disponible, comme par exemple une personne, une institution ou une organisation.
- Les visualisations multi-cas (*cross-case display*). L'ordre des codes, leurs chevauchements ainsi que la fréquence peuvent être représentés afin de comparer les différents cas.

Dans le cadre de cet article, les deux premières formes de visualisation ne seront pas plus explicitées. Il est possible de s'en passer car d'une part le type a nécessite des explications très spécifiques au logiciel utilisé et le type b est déjà plus connu dans la littérature sur les cartes cognitives. J'ai adopté la différence entre visualisation propre à un cas ou multi-cas d'Huberman et Miles dans leur ouvrage du milieu des années 90 « *Qualitative Data Analysis: An Expanded Sourcebook* » C'était un travail novateur dans le champ de la visualisation des données. Tout comme pour n'importe quelle visualisation, celle intervenant dans le contexte de la recherche qualitative dépend de l'acquisition de moyens efficaces. Deux questions doivent être posées :

Premièrement, quelles sont les informations pertinentes pour une analyse qualitative de données?

Dans les approches de recherche qui suivent la méthode comparative, telle que la théorisation enracinée (*Grounded Theory*), les informations pertinentes incluent les concepts théoriques, les codes et sous-codes et leurs dimensions, autant que l'ordre et la progression des codes, leur apparition simultanée ou leur proximité.

Deuxièmement, quels sont les moyens visuels disponibles pour représenter ces informations pertinentes?

Quelques exemples comprennent les symboles, les couleurs, les lignes directionnelles ou non, etc.

Les visualisations internes à un cas (Within-Case Displays)

Les visualisations internes à un cas présentent des éléments d'un cas intéressant sous forme de tableau, de graphe ou de diagramme. Un cas peut, tel qu'indiqué ci-dessus, impliquer un individu, une organisation, une institution ou une configuration. L'analyse qualitative de données comprend habituellement un texte, tel qu'une interview. Deux exemples de visualisations sont présentées dans les sections suivantes :

- *Codeliner* produit une sorte de partition du texte sélectionné, dans laquelle les codes apparaissent en tant que lignes et les parties codées sont illustrées comme des notes, apparaissant dans la chronologie de l'interview.
- *Text Portrait* crée une image qui est contruite comme une photographie. Les parties codées du texte sont présentées ligne par ligne.

Codeliner

Codeliner est une fonctionnalité utilisée pour la visualisation qui présente le texte sous forme d'une image séquentielle de son codage (Figure 1). L'axe des x reprend les paragraphes du texte. Il commence à gauche avec le numéro 1 et termine avec le dernier paragraphe du texte. L'axe des y est construit sur base des codes. La cellule du graphe indique si le code ou sous-code est attribué à un paragraphe donné ou à une section du texte. Une couleur est assignée au code spécifique et reprise dans le tableau.

Codeliner peut être mis en œuvre afin d'obtenir un aperçu rapide des sujets principaux du texte. Grâce au caractère interactif du graphe, chaque passage est accessible immédiatement par un simple click. Une telle représentation est particulièrement utile pour des analyses exploratoires, car les chevauchements et proximités peuvent être identifiées et examinées de façon approfondie.

Codeliner est particulièrement utile pour l'évaluation de *focus group*. Dans ce cas, chaque interlocuteur est représenté comme un code pour leur attribuer leur prise de parole respective. La progression des orateurs est facile à reconnaître sur le graphe. De plus, il est possible d'explorer les données pour identifier quel orateur s'engage (ou reste silencieux) sur un certain sujet. Enfin,

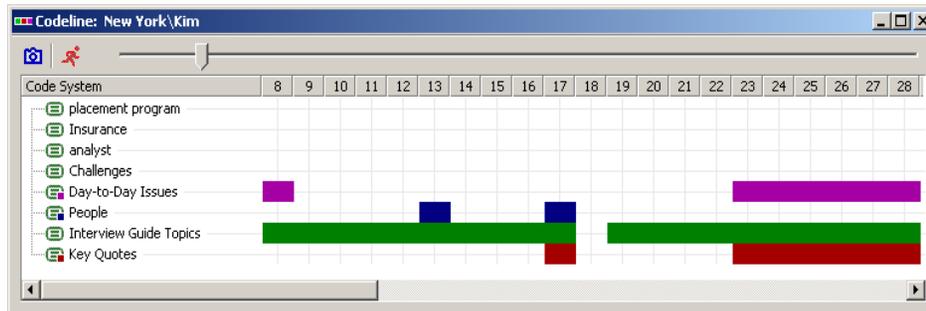


Figure 1. Visualisation des codages d'un texte

il est facile de reconnaître les effets de tels engagements sur la modération si, par exemple, le sujet ou l'orateur se modifie après l'intervention.

Text Portrait

Le second type de visualisation interne à un cas, le *Text portrait*, est une nouvelle méthode de visualisation du contenu des textes et est conçu pour attribuer des catégories (codes) aux passages du texte. Chaque catégorie est associée à une couleur. Le contenu du document est présenté de telle manière que l'attribut coloré peut être observé dans le diagramme dans l'ordre séquentiel. Le diagramme contient un nombre spécifique de carreaux colorés qui sont répartis en lignes et en colonnes.

En commençant par la première ligne et le carreau à l'extrême gauche, les carreaux sont décrits par couleurs et se suivent de gauche à droite. À la fin d'une ligne, il continue à la ligne suivante de la même façon – à la façon d'une machine à écrire. Dans une présentation normale, le nombre total de carreaux est réparti selon les passages codés du texte afin que le nombre de carreaux symbolisant un segment ou une couleur corresponde au pourcentage du segment, déterminé par la taille totale de l'ensemble des segments codés. Par exemple, si il y a seulement un segment codé, auquel deux codes sont assignés, rouge et vert, alors, dans un diagramme qui consiste en 30x40, c'est à dire 1200 carreaux, il y aura 600 rouges et 600 verts. S'il y a 3 codes (rouge, vert et magenta) pour le même segment, le diagramme contiendra 400 carreaux de chaque couleur.

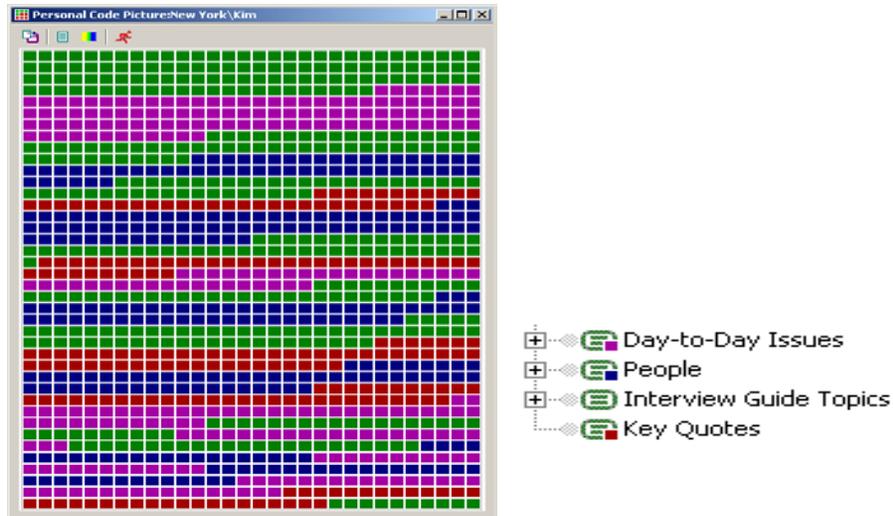


Figure 2. Exemple d'un *Text Portrait*

Text Portrait peut rendre la structure thématique d'un texte accessible sous forme d'image. Par exemple, si des couleurs sont attribuées à des émotions spécifiques dans le cadre d'une interview psychologique, l'humeur dominante ainsi que les sections particulièrement illustratives du texte peuvent être repérées d'un regard.

Les visualisations multi-cas (Cross-Case Displays)

La seconde sorte de visualisation considérée dans cet article est la visualisation multi-cas. Cette nouvelle possibilité permet de présenter des échantillons et des structures contenus dans le texte d'une façon compréhensible et de comparer les différents cas entre eux. Nous présentons deux exemples : le *Code Matrix Browser* et le *Code Relation Browser* de MAXQDA. Les deux représentations visuelles permettent de présenter les codages d'une façon graphique. Avec un coup d'œil à la matrice, on peut extraire quel texte appartient à quelle catégorie et le nombre (élevé ou non) de segments qu'elle contient. Cette visualisation est plus vérifiable et interprétable qu'une matrice numérique.

L'organigramme de la Figure 3, basé sur les travaux de Nadin (1998), illustre la progression des données à la reconnaissance des régularités et des structures :

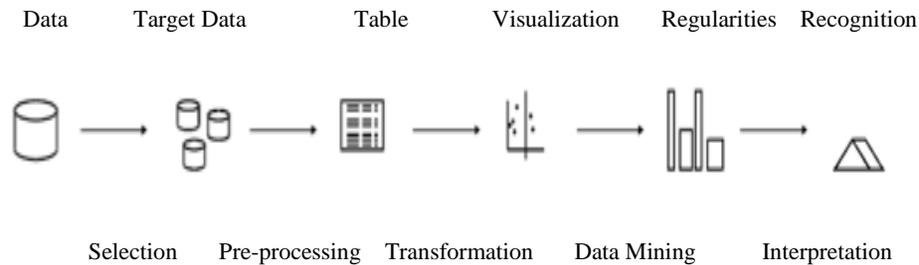


Figure 3. La progression des données à la reconnaissance des structures

Cela signifie qu'une présentation multi-cas des codes commence par la sélection des textes et des codes et sous-codes à visualiser. L'étape décisive est la transformation d'un tableau en graphe, en ce que les régularités sont plus facilement reconnaissables dans un graphe que dans un tableau qui contient une grande quantité d'information. Par exemple, afin de pouvoir identifier le plus grand nombre dans un tableau, il faut lire et comparer tous les nombres entre eux. Mais avec une visualisation où le nombre le plus élevé est présenté sous une forme ou une couleur spécifique, une reconnaissance immédiate est possible.

Afin de pouvoir comparer des textes à partir de leur codage, une visualisation croisée peut être activée. Les textes sont présentés les uns à côté des autres dans la présentation pour faciliter leur comparaison directe au regard des catégories. Le *Code Matrix Browser* nous donne un diagramme tel que le montre la Figure 4.

Les nœuds individuels dans la matrice symbolisent – à travers leur taille et couleur – combien de codages du code ou sous-code spécifique sont issus du texte. Les colonnes de la matrice sont construites à partir du texte : au plus le nœud est grand dans la colonne, au plus on peut trouver de segments codés dans le texte qui entrent dans cette catégorie ou sous-catégorie. Avec un double « clic » sur le nœud, tous les segments du texte correspondants sont listés. La matrice ne présente donc pas seulement une vision globale du codage mais contient également – même de façon invisible – les différents morceaux de texte qui se retrouvent derrière les nœuds.

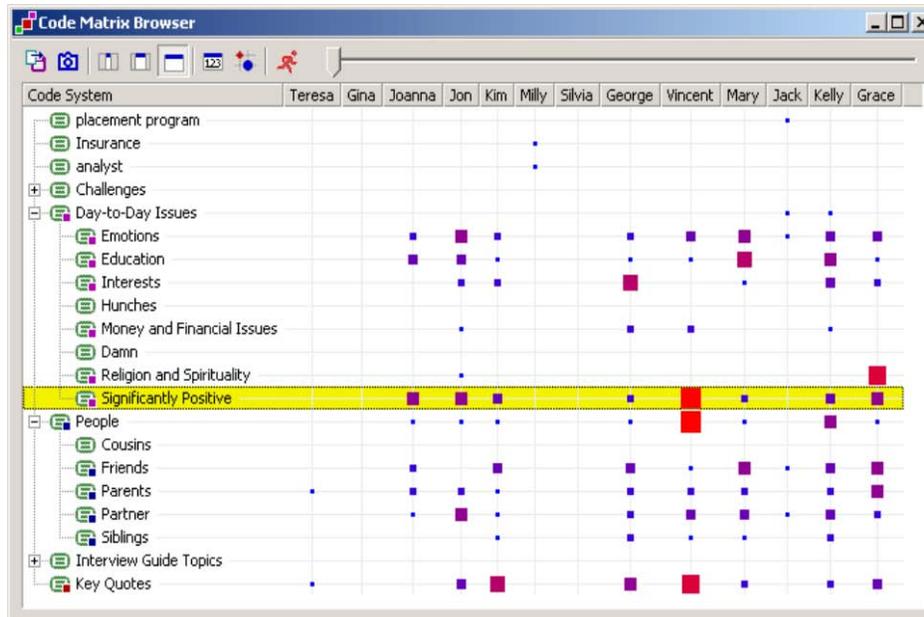


Figure 4. Diagramme des codages dans les textes

La seconde forme de visualisation multi-cas présentée ici affiche les connexions entre les codes, par exemple, combien de fois les codes se chevauchent ou sont attribués simultanément à un même passage de texte. Localiser les chevauchements entre catégories manuellement est une tâche trop énergivore, alors qu'elle peut être réalisée rapidement et de façon fiable à l'aide de l'ordinateur. La Figure 5 montre une présentation du *Code-Relations Browser* dans MAXQDA. Les chevauchements des différents codes sont analysés grâce à cette matrice symétrique des codes. À la place du nombre de chevauchements trouvés, les nœuds de tailles et de couleurs différentes apparaissent dans les cellules de la matrice : au plus le nœud est large, au plus nombreux sont les chevauchements trouvés dans le matériau. L'information numérique apparaît dès que le curseur de la souris est positionné sur le nœud. Un double-clic permet d'extraire du texte l'ensemble des passages correspondant.



Figure 5. Diagramme de chevauchements des codages

Conclusion

Dans la recherche sociale empirique, nous devrions parvenir à travailler avec des visualisations, tout comme d'autres disciplines scientifiques reconnues ont pu les pratiquer avec succès. Toutefois, nous devons apporter une attention particulière à l'accroissement de notre propre compréhension et à l'amélioration de la présentation extérieure de nos résultats. Le proverbe approprié de Confucius « Une image vaut mille mots » est particulièrement valable ici tant la publicité donnée par les images dévoile bien plus que les milliers de mots de scientifiques en sciences sociales pour expliquer le contenu. Les images ont acquis un rôle prédominant dans notre société; elles nous entourent et laissent des impressions durables dans notre mémoire – souvenons-nous du 11 septembre.

Les visualisations deviendront plus complexes et compliquées dans le futur. Si, par exemple, un médecin doit évaluer une radiographie ou une raisonance magnétique, il doit apprendre à le faire. Nous, en tant que profane, nous ne reconnaissons pratiquement rien lorsque nous regardons de telles images. Mais si quelqu'un nous les explique, par exemple une échographie lors d'un diagnostic prénatal de grossesse, alors nous apprenons rapidement et sommes capables de reconnaître la tête, les bras, les jambes et même les ventricules du cœur.

C'est l'approche nécessaire pour rendre observable et compréhensible les analyses visuelles de données et de problèmes en sciences sociales.

Références

- Banks, M. (2001). *Visual methods in social research*. London : Sage.
- Blasius, J., & Greenacre, M.H. (1998). *Visualization of categorical data*. San Diego : Academic Press.
- Feldmann, K. (2003a). Du sollst dir kein Bild machen! (Nicht)Visualisierung in der Soziologie. *TRANS. Internet-Zeitschrift für Kulturwissenschaften*, 14. Document consulté le 15 juillet 2009 de <http://www.inst.at/trans/14Nr/feldmann14.htm>.
- Huberman, A.M., & Miles, M.B. (1994). Data management and analysis methods. Dans N. Denzin, & Y. Lincoln (Éds), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA. : Sage.
- Krempel, L. (2005). *Visualisierung komplexer strukturen. Grundlagen der darstellung mehrdimensionaler Netzwerke*. Frankfurt, M. : Campus.
- Nadin, M. (1998, 27 Novembre). *Zwischen daten und display. Communication présentée au Third Visualization Workshop, University of Paderborn*. Document consulté le 15 juillet 2009 de <http://www.nadin.name/pdf/between.pdf>.

Udo Kuckartz est professeur au département d'éducation de l'Université de Marburg (Allemagne). Il est titulaire d'un doctorat en sociologie et d'un master en sociologie et science politique. Ses domaines d'intérêt se situent en sociologie de l'environnement, en recherche qualitative et en méthodes mixtes. Il est l'auteur du logiciel MAXQDA qui est utilisé dans plus de 70 pays.

Stratégie d'utilisation différenciée de logiciels tout au long d'une enquête qualitative

Kerralie Oeuvray, Ph.D.

Université de Fribourg

Résumé

En réduisant un logiciel qualitatif à un outil parmi d'autres, on peut examiner, tout au long de l'analyse, la place que lui accordent les chercheurs. Différentes *logiques d'utilisation* se dégagent : une logique de proximité décomplexée (examen minutieux des données), une logique de proximité créative (processus consistant à monter vers des catégories de plus en plus abstraites), une logique de distanciation (émergence d'une théorie par la mise entre parenthèse du logiciel qualitatif et par un détour éventuel vers un logiciel de dessin conceptuel), une logique de vérification (reprise du logiciel qualitatif pour débusquer et traiter les cas négatifs). En différenciant ainsi les logiques d'utilisation, on peut considérer les pratiques innovatrices à partir de la capacité des chercheurs à sélectionner l'outil approprié, voire à s'en passer.

Mots clés

INTERPRÉTATION, CAQDAS, LOGIQUES D'UTILISATION, STRATÉGIES CHERCHEURS

Introduction

Comment un chercheur utilise-t-il *concrètement* un logiciel qualitatif lors d'une démarche qualitative? Quelle place accorde-t-il aux logiciels spécialisés et techniquement déjà formatés lorsqu'il entreprend une analyse en vue d'élaborer une théorie compréhensive? Autrement dit, comment fait-il pour « prendre le risque de l'interprétation » (Kaufmann, 2007, p. 92)?

De telles questions rappellent bien sûr tout un ensemble de réserves, documentées depuis plusieurs années déjà, à propos de l'intrusion des logiciels dans les démarches auparavant réservées aux capacités analytiques (quoique parfois opaques) des chercheurs. En effet, l'introduction des logiciels spécialisés capables de soutenir les activités consistant à élaborer une analyse qualitative, fait craindre qu'elle ne conduise à un morcellement de l'intégrité du corpus des données et/ou à une décontextualisation des données (Seale, 1999), à une émulation scientiste des méthodes quantitatives (Bryman, 2004) ou à une standardisation des pratiques (Lonkila, 1995). Sous-jacente à toutes ces

critiques, on rencontre bien sûr la crainte que, face aux exigences techniques dictées par le logiciel, le chercheur ou la chercheuse ne tende justement à disparaître avec sa capacité d'interpréter les données (Kelle, 1995). Même les réactions des utilisateurs CAQDAS semblent parfois donner raison à ces craintes. Ainsi, on distingue les activités consistant à *gérer les données* de celles touchant à *l'élaboration véritable* de la théorie (Fielding & Lee 1998). Certes, de la sorte, l'honneur du chercheur est sauf car l'activité essentiellement analytique reste de son ressort (Gibbs, 2002). Quant au logiciel « technique », il se contentera d'effectuer les tâches administratives. Une variante de l'argument soulignant la place du chercheur est celle qui déconstruit la dichotomie induction – déduction. Ainsi, on démontre que « même » les utilisateurs de SPSS font preuve de capacités inductives en parallèle à leurs capacités déductives (Bryman, 1996; 2004). En faisant alors le parallèle avec les logiciels qualitatifs, au pire, l'utilisation d'un logiciel CAQDAS n'exclut pas l'exercice de l'induction. Et, au mieux, elle réhabilite les activités de déduction grâce aux outils de recherche capables de fournir les preuves que les théories sont véritablement ancrées dans les données (Bazeley, 2007).

En dépit de leur pertinence au sein d'un débat qui continue, tous ces arguments partagent néanmoins le désavantage de renforcer l'opposition entre activité *interprétative* et activité *technique*. Quant à l'activité interprétative, elle se fonde elle-même sur une interprétation littérale de l'injonction d'ancrer à tout prix (Glaser & Strauss, 1967) car la priorité est donnée aux *traces* que, justement, un logiciel peut fournir (en formes de comptes-rendus datés, de catégorisations effectuées, de modifications de l'arborescence, etc.) sur les prises de distance du chercheur par rapport au *corpus* de données. Mais peut-être l'inconvénient majeur est celui qui laisse entendre une sorte de « tout ou rien », autrement dit, une démarche d'analyse se faisant soit *avec* (et exclusivement avec) un logiciel, soit *sans* logiciel (sollicitant évidemment des images folkloriques de crayons de couleurs, de ciseaux, de colle...). Ensemble, ces inconvénients tendent d'une part à évacuer la question de fond de savoir si et en quoi les possibilités techniques influencent les types ou les contenus des interprétations produites, et, d'autre part sur un niveau plus pragmatique, à appréhender l'*usage ordinaire* des logiciels par les chercheurs lorsqu'ils entreprennent une analyse qualitative.

Sans écarter l'importance de la question de fond, nous voulons contribuer à éclaircir le niveau pragmatique en voulant justement capter cet usage ordinaire des logiciels, c'est-à-dire à passer en revue les pratiques utilisées par un chercheur en train d'exercer le métier d'analyste qualitatif en ayant à disposition les divers outils aussi bien conceptuels que techniques. Il

s'agit donc d'une réflexion sur le déroulement de nos propres recherches, menées seule ou en équipe, pendant lesquelles nous avons utilisé les logiciels qualitatifs du type CAQDAS (*Nud.ist N4*, *NVivo 2*), un logiciel de dessin conceptuel (*Inspiration*), de même qu'un logiciel de traitement de texte (*Microsoft Word*). Quant aux recherches, elles se situaient dans le domaine de la toxicodépendance et privilégiaient l'élaboration de théorie à partir de pratiques particulières (Soulet, 2002) ou l'évaluation qualitative de traitements ordonnés par le système judiciaire (Stevens, 2005).

Nous avons ainsi dégagé quatre *logiques d'utilisation*. Il en ressort en outre que non seulement ces utilisations se différencient selon la phase de l'analyse mais aussi que l'activité d'*interprétation* est omniprésente tout au long de la démarche, une activité qui se fait avec, sans ou grâce aux logiciels. Pour illustrer chacune des logiques, nous utiliserons le fil d'une recherche visant à saisir les formes et les logiques des prises en charge médico-sociales « sans fin » allant donc à l'encontre des visées officielles de guérison ou d'autonomie des patients/usagers (Oeuvray, 2007). Pour capter cette tension entre discours officiels et pratiques officieuses voire inavouables, les données incluaient les récits des usagers et de professionnels portant sur les descriptions détaillées des contacts actuels avec le réseau médico-social, des trajectoires de prises en charge et des projections pour le temps à venir.

La logique de proximité décomplexée

Caractéristique du traitement initial des données, notamment discursives, cette logique accompagne l'activité intense des chercheurs consistant à déconstruire les récits ou les documents en codes et catégories ou, pour utiliser le terme CAQDAS, en nœuds/nodes (Deschenaux & Bourdon, 2005). C'est la phase de l'examen minutieux des données et elle est grandement facilitée par l'aisance avec laquelle les logiciels permettent de créer ces nœuds, voire de les multiplier pour un seul extrait de données. La facilité technique (offerte par exemple par *Nud.ist* ou *NVivo*) ne devrait pas toutefois cacher l'enjeu interprétatif. Car, en conformité avec les procédures constitutives des théories ancrées, la compétence analytique fondamentale consiste à *nommer* en tant que *code* ou *catégorie* chacun des nœuds ainsi créés.

Les exemples du Tableau 1 proviennent de deux patients d'un centre de traitement de substitution à propos de leurs contacts avec le centre.

Le thème plus spécifique des *interactions* ou contacts entre usagers et professionnels se complexifiait avec les usagers d'un centre social dont les offres incluaient des suivis personnalisés et un centre de jour fréquenté aussi bien par les usagers que par le personnel (voir Tableau 2).

Tableau 1
 Désignation de codes lors du premier traitement des discours

Discours	Codes
<i>Je viens ici tous les jours ; ça me fait sortir de chez moi.</i>	Quotidienneté : structuration Fréquence de contacts Liens sociaux
<i>J'suis déjà là avant les autres... je reste un moment encore.</i>	Routinisation Contacts facultatifs (allant au-delà de l'obligatoire)

Tableau 2
 Approfondissement d'un thème particulier

Discours	Codes
<i>Je croise mon assistant social dans les escaliers. Ça va ? Ça va.</i>	Interactions / interactions informelles Interactions spontanées Croisement (au lieu du rendez-vous)
<i>Si on se croise au centre de jour, on boit un café ensemble... c'est sympa et on fait le point [de ma situation].</i>	Proximité permettant les croisements Coprésence facilitée

Si le défi pour le chercheur, à ce stade de l'analyse, est donc de saisir en quoi un extrait est un exemple (Strauss & Corbin 2004), la facilité technique lui permet de le faire de façon *décomplexée*, c'est-à-dire sans qu'il ait à se soucier du maintien ou non de la voie ouverte par la création d'un nœud ou code particulier. Le codage initial de l'extrait suivant rappelle justement une sorte d'incertitude concernant l'utilisation par l'utilisateur du mot « nous ».

Le sens du « nous » en effet a été difficile à cerner pendant la première phase de l'analyse mais, couplé avec d'autres données, sa pertinence s'est imposée par la suite (Tableau 3). Et, il est aisé d'y revenir vu que les traces subsistent grâce à une facilité technique consistant à multiplier les interprétations possibles plutôt que de se limiter aux interprétations probables car gérables par des supports physiques moins « généreux » (les ciseaux et la

Tableau 3
Facilitation du questionnement par le codage

Discours	Code
<i>Nous, ici, on aide les jeunes</i>	Nous / identification avec centre ?
(Usager, femme de 35 ans participant à un programme d'occupation)	Nous / solidarité entre usagers ?
	Nous / travail bénévolat ?

colle mais aussi la rédaction continue de l'analyse en cours dont les bifurcations seraient progressivement de moins en moins bienvenues). Tout en restant dans la proximité étroite des données, le travail d'interprétation facilité, tout en restant exigeant peut de la sorte être repris et redirigé vers des pistes inédites ayant l'avantage de rompre avec les cadres explicatifs habituellement privilégiés (malgré lui) par le chercheur.

La logique de proximité créative

Le logiciel CAQDAS reste de mise avec l'utilisation accrue de mémos. Certes l'élaboration de mémos liés aux codes ou aux passages particuliers se développe déjà en rapport de proximité étroite avec les données et les interprétations locales. Mais l'activité interprétative exige de la part du chercheur un engagement créatif accru. En mettant en forme ses réflexions, grâce notamment à l'outil mémo, le chercheur commence à poser les cadres d'une théorie éventuelle. Cet engagement s'intensifie avec le travail centré sur les codes eux-mêmes. La proximité ici se rapporte donc moins aux données brutes qu'aux premières étiquettes nommées par le chercheur. Les regroupements de codes, la création de nouveaux nœuds et, avant tout, les constructions successives des rapports entre eux (représentés par les « arbres ») expriment à la fois l'attachement aux données ayant généré les étiquettes et à l'interprétation approfondie et de plus en plus abstraite du chercheur (Tableau 4). La proximité créative est donc aussi une proximité savante qui rebondit avec les nouveaux questionnements.

C'est cette logique – à la fois attachée aux nœuds et aux arbres mais qui plane aussi au-dessus d'eux – qui favorise la création de nouvelles catégories, via l'utilisation accrue des mémos. Quant aux réflexions continues, certes le chercheur n'est pas contraint d'employer exclusivement l'outil mémo fourni par les logiciels. Mais la possibilité de retrouver les discours originaux fondant ces réflexions constitue un avantage indéniable.

Tableau 4
Catégorisation des codes

Codes	Catégorisation des codes
Ensemble des codes portant sur les espaces-temps de la quotidienneté « inspirés » par les types d'interactions entre usagers et professionnels	Organisation spatio-temporelle dynamique <ul style="list-style-type: none"> - extension des lieux d'échange formel (bureaux, clinique) vers l'informel (centre de jour, bistros du coin) - extension des temps de contact hors du cadre officiel - émergence d'espaces-temps ambigus (salles d'attente conviviales)
Ensemble des codes portant sur les types de rapport entre clients et usages « inspirés » par leur caractère qui est indissociable des espaces-temps informels émergeant des cadres institutionnels formels	Interactions hors rôles conventionnels <ul style="list-style-type: none"> - rôles négociés ou glissements de rôle émergeant de la proximité ?
Ensemble des codes portant à la fois sur un sentiment de différence engendré par la problématique drogue et sur le sentiment d'aisance (usagers) et de soulagement (professionnels) concernant la protection fournies par les prises en charges en cours	Émergence du nous de proximité <ul style="list-style-type: none"> - connivence entre usagers et professionnels pour former un « nous » autour de la connaissance de la problématique <i>drogue</i>

La logique de distanciation

Progressivement, le chercheur dispose ainsi de schèmes explicatifs sous forme de cumul de mémos ou de textes ponctuels ainsi que de versions multiples ou successives d'arbres hiérarchiques et/ou relationnels; *bref*, de versions différentes d'un corpus reconstitué de données. Certes, la voie indiquée peut être celle de la poursuite progressive voulant l'abstraction des catégories, tout en précisant, pour prendre le langage de Glaser et Strauss (1967), les dimensions et les propriétés des catégories centrales. Mais parallèlement à une telle démarche, d'ailleurs parfaitement compatible avec les logiciels CAQDAS, un chercheur peut vouloir souhaiter une sorte de « temps d'arrêt ». En effet, devant la concurrence éventuelle d'explications, ou plus vraisemblablement, face à la multiplicité des directions qui se présentent, le chercheur peut être

amené à poser la question : « *finalement que veut dire tout cela ?* ». La logique de *distanciation* traduit alors la capacité du chercheur de délaisser aussi bien les données brutes que ses propres schèmes explicatifs.

Et voilà! Le logiciel spécialisé peut être mis entre parenthèses. C'est alors le moment pour les marches en forêt dans l'attente du célèbre *insight* ou alors pour les discussions à bâtons rompus dont connaissent les équipes de recherche. C'est ainsi que le cadre de la théorie commence à s'affirmer. Certes, la théorie sera toujours ancrée (la marche en forêt ne signifie pas pour autant une rupture avec la démarche en cours) mais elle reflète aussi un travail *d'interprétation globale et transcendante*, en contraste avec une démarche linéaire vers l'abstraction, aidée tout de même par des catégorisations transversales (voir Tableau 5).

N'empêche, le chercheur peut donner des coups de pouce informatiques pour favoriser l'émergence de l'*insight*. En plus d'une sorte de retour aux sources par la production de textes avec un logiciel de traitement de texte, nous avons recherché une forme ludique afin d'examiner les catégories dans une perspective moins scolaire. Un logiciel de dessin conceptuel, tel que « *Inspiration* » (voir Figure 1), peut justement fournir un support concret et facile à cette logique de distanciation. Avec ce genre de logiciel, il suffit en effet de manipuler le positionnement du curseur et des « cliques » de souris pour produire de nouveaux symboles ainsi que des liens entre eux. Le schéma suivant est un exemple du temps d'arrêt bien productif.

Relevons une dernière activité qui peut être menée lors de cette phase de distanciation avec le logiciel CAQDAS : la définition de l'organisation du texte final de recherche. Quoiqu'il est techniquement faisable de transposer simplement l'arbre des catégories progressivement remaniées, le résultat de cette période du temps d'arrêt peut plaider pour une autre organisation qui s'inspirera des schèmes explicatifs mais sans les reproduire fidèlement.

La logique de vérification

Cette dernière logique intervient tout au long de l'écriture du rapport final. Tout en étant plongé dans la production discursive de la théorie retenue, le chercheur reprend contact avec le logiciel qualitatif. À un premier niveau, il s'agit de retrouver les citations les plus pertinentes pour les insérer dans le rapport, opération facile compte tenu des outils de recherche. Mais un deuxième niveau plus exigeant consiste à confronter les interprétations (presque) abouties avec les données brutes ou encore avec un exemple discordant, voire un cas négatif (voir Tableau 6). C'est donc non seulement

Tableau 5
Émergence de la théorie

Catégories	<i>Temps d'arrêt</i>	Éléments de la théorie
Organisation spatio-temporelle dynamique		En formulant les attentes à poursuivre pendant longtemps encore l'activité d'utilisateur, ou de devoir encadrer encore longtemps ces mêmes usagers, les acteurs privilégient une sorte de convivialité contrôlée.
Interactions non conventionnelles		
Le « nous » de l'appartenance et de la différence		Ainsi, un sentiment de communauté informelle, ancrée néanmoins dans les espaces-temps formels, émerge et remplace les rapports visant auparavant l'affranchissement des prises en charge

Qu'est-ce que tout cela veut dire?

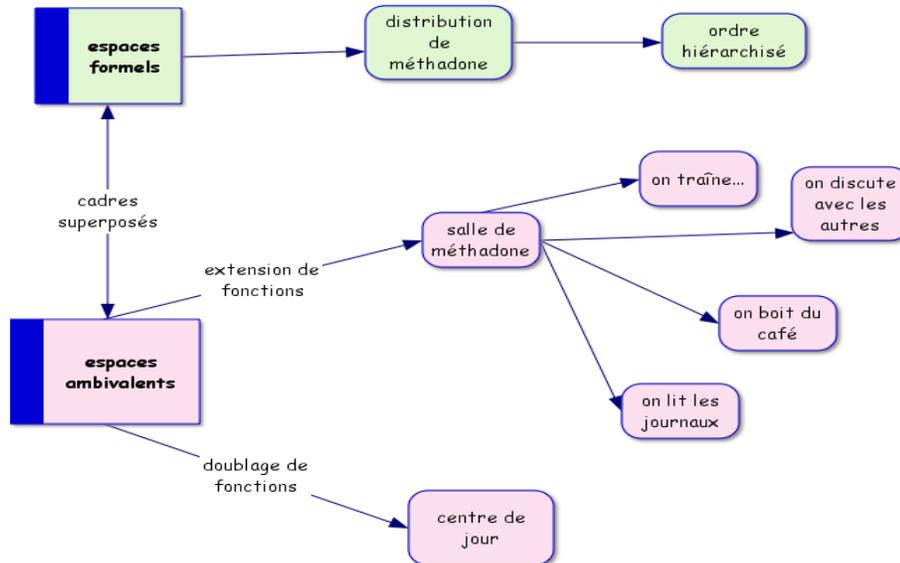


Figure 1. Exemple d'un dessin conceptuel visant à clarifier les éléments d'une théorie

Tableau 6
 Mise en évidence des éléments discordants

Éléments de théorie	Contre-exemple
<p>Les acteurs justifient leurs attentes de poursuivre pour longtemps encore leurs contacts par des interprétations catastrophiques des périodes vécues sans traitement et une anticipation d'une urgence certaine ainsi que la conviction de l'absence d'alternatives ailleurs pour le temps à venir...</p> <p>... en effet, une autre durée (la durée de vie) s'impose comme concurrente de celle des longues prises en charge.</p>	<p><i>Je n'ai pas envie de passer le reste de ma vie ici... Lorsqu'on a le sida... (homme, 39 ans)</i></p>

l'enjeu de l'ancrage de la théorie qui se joue ici mais aussi la capacité du chercheur à concilier sa théorie avec les éléments apparemment discordants.

Contrastes et complémentarités

Si on prend en compte l'ensemble de la démarche d'analyse qualitative, alors les différentes logiques d'utilisation de logiciels apparaissent (Tableau 7).

Certes, présentées ainsi, ces quatre *logiques d'utilisation* de logiciels de recherche qualitative semblent reproduire de manière trop linéaire la démarche d'analyse. En fait, selon les exigences du moment, le chercheur va privilégier l'une ou l'autre logique : une nouvelle récolte de données, l'élaboration d'un rapport intermédiaire, le besoin de changer de perspective ou simplement une réaction de fatigue suite à plusieurs séances de travail de micro-analyse! Cela dit, notre tentative de différencier les logiques d'utilisation réside moins dans l'établissement d'un ordre progressif ou itératif que dans l'explicitation des différentes manières entreprises par le chercheur pour interpréter le corpus des données. La contribution décisive des logiciels CAQDAS, est justement de faciliter cette ouverture, sans toutefois que le chercheur cède à la facilité interprétative – heureusement. Autrement dit, en combinant habilement les logiques de liberté intellectuelle et de proximité étroite des données avec celles de la distanciation et de la rigueur, le chercheur s'ouvre vers des interprétations nouvelles car sortant des cadres théoriques initiaux, tout en assumant les postures fondamentales des théories ancrées émergeant des données. De la sorte, le logiciel n'est qu'un outil parmi d'autres, mais un outil de valeur dont nous ne pourrions plus nous passer.

Tableau 7
Utilisation différenciée des logiciels

Logiques d'utilisation	Caractéristiques	Logiciels mobilisés
LA PROXIMITÉ DÉCOMPLEXÉE L'examen minutieux des données	Interprétation locale Codifications plurielles grâce à la facilité technique	CAQDAS (Nud.ist, NVivo)
LA PROXIMITÉ CRÉATIVE monter vers des catégories de plus en plus abstraites	Interprétation savante et conceptuellement abstraite Catégorisation des ensembles de codes via la manipulation des arborescences Formulation du questionnement continu via les mémos	CAQDAS (Nud.ist, NVivo)
LA DISTANCIATION <i>Finement, que veut dire tout cela ?</i>	Interprétation de l'ensemble Recherche et encouragement de l'insight Mise entre parenthèse du logiciel Organisation de la table des matières	----- Logiciel de dessin conceptuel Traitement de texte
LA VÉRIFICATION Débusquer et traiter les cas négatifs	Interprétation argumentée Recherche de citations exemplaires La chasse aux cas négatifs Rédaction finale	CAQDAS (Nud.ist, NVivo) Traitement de texte

Références

- Bazeley, P. (2007). *Qualitative data analysis with NVivo*. London : Sage.
- Bryman, A. (1996). *Quantity and quality in social research*. London : Routledge.
- Bryman, A. (2004). *Social research methods*. Oxford : Oxford University Press.
- Deschenaux, F., & Bourdon, S. (2005). *Introduction à l'analyse qualitative informatisée à l'aide du logiciel QSR NVivo 2.0*. Les cahiers pédagogiques de l'Association pour la recherche qualitative.

- Fielding, N.G., & Lee, R.M. (1998). *Computer analysis and qualitative research*. London : Sage.
- Gibbs, G.R. (2002). *Qualitative data analysis. Exploration with NVivo*. Buckingham : Open University Press.
- Glaser, B., & Strauss, A. (1967). *The discovery of Grounded Theory : Strategies for qualitative research*. New York : Aldine.
- Kaufmann, J.-C. (2007). *L'entretien compréhensif* (2^e éd.). Paris : Armand Colin.
- Kelle, U. (Éd.). (1995). *Computer-aided qualitative data analysis. Theory, methods and practice*. London : Sage.
- Lonkila, M. (1995). Grounded Theory as an emerging paradigm for computer-assisted qualitative data analysis. Dans U. Kelle (Éd.), *Computer-aided qualitative data analysis. Theory, methods and practice* (pp.41-51). London : Sage.
- Oeufray, K. (2007). *Rester face aux temps certains. Destins sociotemporales de personnes toxicodépendantes sans perspective de travail*. Thèse de doctorat inédite, Université de Fribourg, Suisse.
- Seale, C. (1999). *The quality of qualitative research*. London : Sage.
- Soulet, M.-H. (avec la collaboration de M. Caiata Zufferey et K. Oeufray) (2002). *Gérer sa consommation. Drogues dures et enjeu de conventionnalité*. Fribourg : Éditions Universitaires.
- Stevens, A., Berto, D., Hunt, N., Kerschl, A., McSweeney, T., Schaaf, S., Soulet, M.-H., Trinkl, B., Turnball, P., Waidner, G., Werdenich, W., & Uchtenhagen, A. (2005). *Quasi compulsory treatment for drug dependant offenders*. Commission européenne Quality of Life (5^{ème} programme), CT 2002-01446.
- Strauss, A., & Corbin, J. (2004). *Les fondements de la recherche qualitative. Techniques et procédures de développement de la théorie enracinée*. Fribourg : Academic Press.

Kerralie Oeufray est sociologue, rattachée au domaine sociologie, politiques sociales et travail social de l'Université de Fribourg. Privilégiant l'approche qualitative, ses activités de recherche s'insèrent dans une sociologie des problèmes sociaux et touche notamment au traitement social et pénal de la toxicodépendance ainsi qu'à la pauvreté. En outre, elle enseigne les cours de méthodologie qualitative y compris, au niveau du Masters, l'utilisation des logiciels pour l'analyse des données qualitatives.

Traitement informatique de données orales : quels outils pour quelles analyses ?

Virginie André, Maître de conférences

Université de Nancy 2

Christophe Benzitoun, Maître de conférences

Université de Nancy 2

Emmanuelle Canut, Maître de conférences

Université de Nancy 2

Jeanne-Marie Debaisieux, Maître de conférences

Université de Nancy 2

Bertrand Gaiffe, Chargé de recherche

Université de Nancy 2

Evelyne Jacquy, Chargée de recherche

Université de Nancy 2

Résumé

Dans cet article, nous présentons le projet TCOF (Traitement de corpus oraux en français) dont l'objectif est la pérennisation, la collecte et le traitement des données de français parlé. Nous mettons plus particulièrement l'accent sur la façon dont les outils informatiques, utilisés de manière conventionnelle ou détournés de leur usage habituel (logiciel d'alignement texte-son, concordanciers, etc.), font avancer le travail d'archivage, de codage et d'analyse qualitative de nos données en vue d'une diffusion internationale libre et gratuite.

Mots clés

CORPUS ORAUX, TRANSCRIPTION ASSISTÉE PAR ORDINATEUR, MÉTADONNÉES ET CODAGE, ANALYSE LINGUISTIQUE QUALITATIVE, MUTUALISATION ET CUMUL DES CONNAISSANCES.

Contexte scientifique de la recherche sur les données orales

Dans le contexte de la recherche française, les données recueillies par les chercheurs en sciences humaines ne sont pas toujours librement à la disposition de l'ensemble de la communauté scientifique. Il existe ainsi, en sciences du langage, une multitude de données orales et écrites sur le français mais leur accès est très inégalement assuré :

- en ce qui concerne l'écrit, des bases de données textuelles existent et sont consultables, tel que Frantext, corpus à dominante littéraire constitué de textes français qui s'échelonnent du XVI^e au XXI^e siècle. Le Centre national de ressources textuelles et lexicales (CNRTL) regroupe également des corpus informatisés comme les deux années des éditions intégrales du quotidien régional *l'Est républicain*. Internet constitue en soi une base de données;
- la situation est plus problématique en ce qui concerne les données orales. Ces dernières sont éparpillées et rarement consultables par des personnes extérieures à la recherche locale. « On peut poser qu'il y a sans doute entre quatre ou cinq millions de mots effectivement disponibles mais l'absence de coordination rend l'exploitation de l'ensemble impossible » (Debaisieux, 2005). Les données sont, en outre, fortement hétérogènes dans la mesure où le recueil et l'analyse dépendent de l'objectif d'étude, de l'orientation épistémologique et de la connaissance des outils à disposition de chaque chercheur. En conséquence, la mutualisation des connaissances issues de chacune des exploitations pour les corpus oraux reste relativement exceptionnelle, ce qui rend difficile le cumul des analyses (pour une tentative de synthèse voir Savelli (2005), pour un inventaire des corpus oraux en France voir Cappeau & Seijido (2005)).

Certains pays européens (Allemagne, Angleterre, Espagne, Portugal, Italie) ont dépassé ces difficultés et ont pu constituer un corpus de référence contenant aussi bien de l'écrit que de l'oral :

- le British national corpus (BNC) : cette banque de données compte 100 millions de mots enrichis par des annotations morphosyntaxiques ;
- le Corpus de référence de l'espagnol actuel (CREA) : cette banque de données compte actuellement 100 millions de mots et devrait encore s'enrichir de vingt-cinq millions. Elle présente une grande variété

d'extraits écrits et oraux, produits dans tous les pays hispanophones depuis 1975;

- le Corpus de référence allemand COSMAS II (Corpus search, management and analysis system : Institut *Für Deutsche Sprache* à Mannheim);
- le Corpus de référence du portugais contemporain (CRPC) : cette banque de données orales et écrites compte actuellement 80 millions de mots.

Prenant conscience du retard de la France dans le développement des ressources et la constitution d'une banque de données textuelles informatisée, notamment pour la langue parlée, la communauté universitaire a entamé depuis 2000 une réflexion de fond (Bilger, 2000; Habert, Nazarenko & Salem, 1997), en particulier pour ce qui concerne la constitution et l'hébergement de corpus :

- création d'un *Guide des bonnes pratiques* qui fait le point sur les aspects déontologiques, juridiques et techniques du recueil et de l'analyse de données orales (Baude, 2006);
- regroupement sur une même base internet des enregistrements concernant, entre autres, le français et ses variétés et les langues de France. Ce projet, sous l'égide de la Délégation générale à la langue française et aux langues de France (DGLFLF), via le conseil scientifique de l'Observatoire des pratiques linguistiques, s'inscrit dans le cadre d'un accord entre le CNRS et le ministère de la Culture pour prolonger et rendre cohérent sur le long terme le programme *Corpus de la parole* qui donne la priorité aux ressources orales;
- création par le CNRS d'un Centre de ressources pour la description de l'oral (CRDO) pour la conservation et la diffusion de corpus oraux;
- création de base de données informatisées : projet ESLO (Enquête sociolinguistique à Orléans, laboratoire CORAL, Université d'Orléans), projet international PFC (Phonologie du français contemporain), projet CLAPI (Corpus de langue parlée en interaction », Unité mixte de recherche CNRS : ICAR, Université Lyon 2).

Nous présenterons ici le projet TCOF (Traitement de corpus oraux en français), rattaché à l'Unité mixte de recherche CNRS - Nancy Université : Analyse et traitement informatique de la langue française (ATILF), qui est développé dans le contexte nancéien et qui illustre cette avancée de la réflexion sur la collecte et le traitement des données de français parlé (Canut, 2008). Nous mettrons l'accent sur la façon dont les outils informatiques, utilisés de manière conventionnelle ou détournés de leur usage habituel (logiciel d'alignement texte-son, concordanciers, etc.), font avancer le travail

d'archivage, de codage et d'analyse qualitative de nos données en vue d'une diffusion internationale libre et gratuite.

Origine et objectifs du projet Traitement de corpus oraux (TCOF)

À l'origine du projet se trouvent des linguistes qui ont accumulé depuis les années 1990 un grand nombre de données orales (des enregistrements audio) dans des domaines de recherche distincts : syntaxe, sociolinguistique, didactique et linguistique de l'acquisition. La convergence entre les différentes orientations s'est faite autour d'un projet portant sur la description et la comparaison des productions langagières, notamment du point de vue de l'usage des formes linguistiques chez les locuteurs adultes et enfants, en lien avec les interactions verbales. Le projet a démarré en septembre 2005 et a depuis bénéficié d'un soutien financier et logistique du laboratoire ATILF. Il a un double objectif de recherche :

- la description linguistique, aspects lexicaux, syntaxiques, pragmatiques et interactionnels, de pratiques langagières dans des situations d'interaction entre adultes et d'interaction entre adultes et enfants au cours de l'apprentissage du langage;
- la comparaison, du point de vue de leurs caractéristiques linguistiques, entre les verbalisations entre adultes, récits, conversation, explication, etc., et les verbalisations des adultes adressées à de jeunes enfants (moins de 7 ans) dans des situations d'apprentissage ou de vie quotidienne.

L'étude doit permettre de mener des analyses descriptives sur un grand nombre de corpus de français parlé pour affiner la description syntaxique de la langue orale (Benzitoun, Campione, Deulofeu, Henry, Teston, Valli & Véronis, à paraître; Véronis, 2000). Elle doit permettre de rendre compte de phénomènes d'acquisition comme l'évolution du langage de l'enfant (augmentation et diversité du répertoire linguistique) et les caractéristiques linguistiques du langage de l'adulte adressé à l'enfant.

Nous avons dès le départ envisagé la diffusion des données sans restriction (mise à disposition libre et gratuite pour la communauté scientifique). Ainsi, avant même de pouvoir envisager une analyse des données, la question de la conservation des données, de leur archivage et de leur pérennité s'est posée, d'autant que leur visibilité pouvait être accrue grâce au traitement informatique. À la fin de la chaîne, et c'est la principale réflexion que nous menons actuellement, se pose le problème du croisement possible des analyses (sur le plan méthodologique en particulier mais aussi au niveau de l'analyse linguistique des données) et donc de leur mise en commun.

Constitution des données primaires

Il s'agit de données orales authentiques (enregistrées dans des situations réelles) : conversations entre adultes (récits divers, réunions de travail en entreprise, entretiens...), entre adulte et enfant de moins de 7 ans (conversations libres, narrations à partir de livres illustrés).

Le matériel qui a été utilisé pour procéder aux enregistrements est classique : magnétophone à cassettes avant l'ère du numérique, enregistreurs numériques depuis deux ou trois ans. Il a donc fallu, dans un premier temps, procéder à l'harmonisation du son en numérisant toutes les cassettes (son Wav, avec le logiciel Wavelab) et, dans un deuxième temps, archiver le tout sur un même espace (machine dédiée).

La réflexion a porté sur les aspects juridiques liés au recueil des données – notamment en ce qui concerne les autorisations – et a pris en compte les positions publiées dans le *Guide des bonnes pratiques* (Baude, 2006). Sur ce point, même si nous étions en possession d'autorisations écrites des locuteurs, nous nous sommes mis d'accord sur le principe d'une anonymisation des données graphiques et sonores interdisant l'identification des locuteurs. Pour ce faire, nous avons instauré :

- un accès restreint (aux seuls chercheurs de l'équipe) des données susceptibles de renseigner l'identité des personnes comme le nom de famille ou la date de naissance;
- une anonymisation de la transcription et du son (masquage par du bruit) des noms de villes ou de sociétés, des noms de famille... renseignant sur l'identité des personnes. Par exemple, dans la transcription, nous avons systématiquement remplacé les noms de familles par «P» et dans le fichier son nous avons remplacé ces noms de famille par un « bip ».

Constitution des données secondaires

La transcription

Au fil de la réflexion et de l'usage généralisé des technologies informatiques, nous avons élaboré plusieurs versions des conventions de transcriptions et profondément modifié le format de nos données. Ainsi, nous sommes passés de conventions de transcriptions divergentes (adulte-adulte, adulte-enfant, réunions de travail) à des conventions communes, moyennant quelques informations supplémentaires spécifiques. Nous reproduisons ici deux exemples de corpus transcrits ainsi que les symboles utilisés dans les transcriptions. Les transcriptions sont faites en orthographe standard, sans trutage orthographique (voir Tableau 1).

Nous sommes également passés de transcriptions sous format Microsoft Word ou papier vers des transcriptions alignées texte-son (points de synchronisation entre la transcription et le fichier son) avec le logiciel Transcriber (voir Figure 1).

Le passage à la transcription assistée par ordinateur a eu un impact important sur la fiabilité et le temps de transcription, notamment grâce à l'intégration dans une interface simple et conviviale de tous les modules utiles. Transcriber étant un logiciel libre, il ne fait pas obstacle à la diffusion de nos corpus (Tableau 2).

Nous avons par ailleurs complété les fichiers générés par Transcriber à l'aide de traitements spécifiques, en fonction de nos nécessités d'analyse. Nous avons ainsi ajouté des modules supplémentaires comme les chevauchements de plus de deux locuteurs et la numérotation des tours de parole.

Nous avons également, grâce à une collaboration avec le Laboratoire Lorrain de recherche en informatique et ses applications (LORIA) (UMR CNRS, Nancy), développé un logiciel en Java comparable à Transcriber (logiciel JTRANS). Il permet un transfert semi-automatique de transcriptions sous format Microsoft Word vers des transcriptions alignées texte-son. Ce logiciel, en cours de test, permettra un alignement rapide, précis et fiable des transcriptions existantes et permettra de traiter de façon rapide les différents états des transcriptions rassemblées depuis le début du projet (voir Figures 2 et 3).

Le format de codage

Une fiche documentaire accompagne chaque fichier son et sa transcription. Sa mise au point a nécessité une phase de standardisation et d'uniformisation assez longue. La fiche décrit ce qu'on appelle classiquement les « méta-données » relatives au corpus considéré (participants à la conversation, titre du corpus, responsable(s) du projet, nom du transcripteur, matériel d'enregistrement utilisé, date d'enregistrement, pour n'en citer que quelques-unes) (voir Tableau 3).

D'un point de vue pratique, la séparation entre méta-données d'un côté et fichier transcrit de l'autre, peut poser problème :

- d'une part, pour chaque corpus, il faut s'assurer de garder le lien entre sa fiche documentaire, son fichier son et sa transcription (pour l'instant ce lien repose sur une convention de nommage des fichiers et leur présence dans un même répertoire);

Tableau 1
Exemples d'application des conventions de transcription

Corpus adulte-enfant	Corpus adulte
VERONIQUE	L1
<ul style="list-style-type: none"> là je vois la poupée avec un nounours là 	<ul style="list-style-type: none"> je sais pas elle a dit qu'on trouverait un moyen
ADULTE	L2
<ul style="list-style-type: none"> tu crois que c'est un nounours ça ? c'est bien plus gros qu'un nounours 	<ul style="list-style-type: none"> ah bon sûrement parce que moi je peux pas prendre de voiture
VERONIQUE	L1
<ul style="list-style-type: none"> non c'est un Babar [<i>pron=baba</i>] 	<ul style="list-style-type: none"> oui je sais mais euh peut-être que Jean aura une bagnole
ADULTE	L2
<ul style="list-style-type: none"> ah c'est un Babar 	<ul style="list-style-type: none"> c'est combien de temps jusqu'à Troyes
VERONIQUE	L1
<ul style="list-style-type: none"> non c'est un éléphant [<i>pron=efa~</i>] 	<ul style="list-style-type: none"> pff ça fait trois heures je crois c'est long hein
ADULTE	L2
<ul style="list-style-type: none"> c'est un oui c'est un éphant [<i>sic</i>] 	<ul style="list-style-type: none"> mh oh c- ça va

- d'autre part, certaines recherches sur le corpus sont difficiles à formuler parce qu'elles font intervenir des fichiers différents, ce qui est le cas, par exemple, si l'on souhaite trouver tous les corpus qui contiennent une expression donnée (à trouver dans le fichier de transcription) et pour lesquels un des locuteurs est une femme entre 30 et 40 ans (à trouver dans la fiche documentaire).

Ces deux points nous ont poussés vers un format de représentation accumulant dans un même fichier les données et les méta-données.

Un autre problème auquel nous devons faire face est celui de l'hétérogénéité dans les formats initiaux des corpus : fichiers alignés à l'aide de Transcriber vs fichiers transcrits sous Word et donc non alignés.

Tableau 2
Récapitulatif des symboles de transcription

{...}	Commentaires (balise via Transcriber)
[...]	Prononciations particulières notées avec l'alphabet phonétique SAMPA (balise via Transcriber)
(...)	Variante graphique indécidable
+	Pauses
///	Pauses très longues
=	Liaison non standard remarquable
/..., .../	Hésitations entre transcription
...-	Amorces
*	Syllabe incompréhensible
***	Suite de syllabes incompréhensibles
###	Passage enregistré non transcrit
\$\$\$	Coupure de l'enregistrement

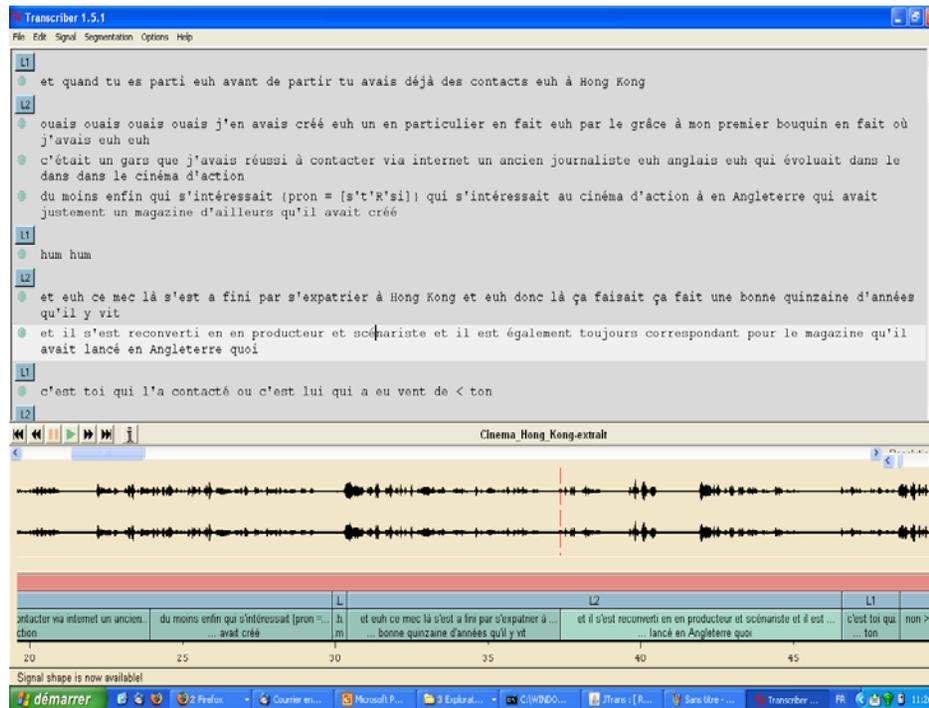


Figure 1. Copie d'écran du logiciel Transcriber

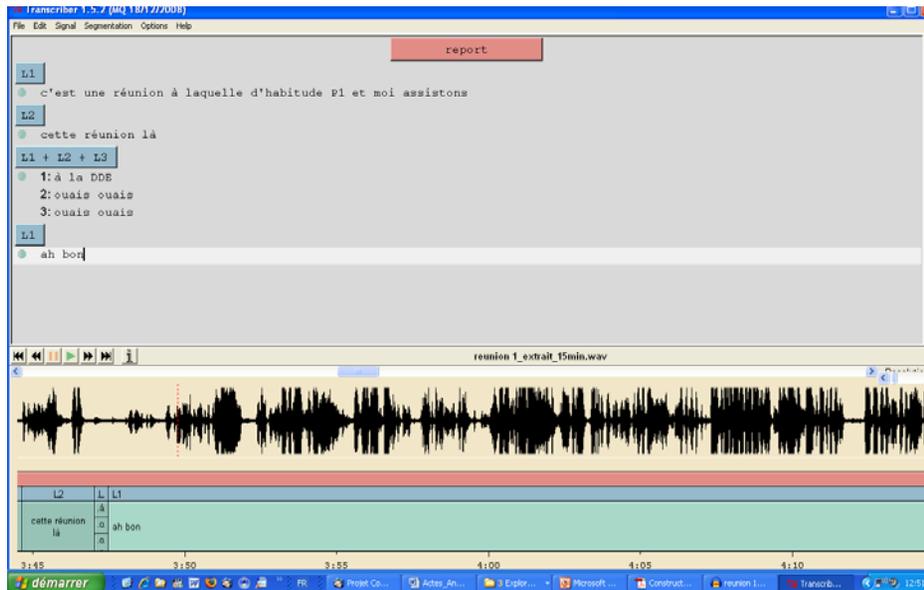


Figure 2. Version modifiée de Transcriber pouvant prendre en compte un nombre non limité de locuteurs dans le cadre de chevauchements de parole (réalisé par Matthieu Quignard)



Figure 3. Interface du logiciel JTRANS

Tableau 3
Extrait d'une fiche documentaire

Général	
Identifiant	Anthony2_Can
Ancien identifiant/Titre	
Responsable(s) du corpus	E. Canut
Droit d'accès	Libre
Contact (courriel)	emmanuelle.canut@univ-nancy2.fr
Nom(s) du/des corpus associé(s)	Anthony_Can
Type d'association	Longitudinal
Alignement texte-son	Oui
Nom du logiciel utilisé pour l'alignement	Transcriber
Support(s) anonymisé(s)	Son+transcription
Type d'anonymisation du son	Bip
Notes sur l'anonymisation	
Nombre de locuteurs	2
Relation entre participants	Cadre de l'animation
Type de corpus	adulte-enfant
Modalités de recueil des données	Chercheur participant
Canal de communication	Face à face
Cadre situationnel	Scolaire ou périscolaire
Degré d'interactivité	Très interactif

Ces deux problèmes nous ont poussés à choisir un format de transcription TEI (Text Encoding Initiative¹) qui permet :

- une intégration des méta-données et des données dans un même fichier;
- un encodage des fichiers alignés texte/son et des fichiers non alignés².

Le passage d'un fichier au format Transcriber accompagné de sa fiche documentaire vers un fichier au format TEI est en cours d'automatisation.

Un autre argument pour le passage au format TEI est celui de l'enrichissement du corpus a posteriori, par exemple le cas d'un fichier simplement transcrit que l'on voudrait enrichir par un étiquetage en partie du

discours ou en tours de parole. Dans la mesure où le format TEI est le fruit de plus de vingt-cinq ans de travail de communautés diverses, des modules permettant le type d'annotation que nous venons de mentionner existent. Par ailleurs, le format TEI étant largement documenté (3500 pages de recommandations avec référencement des annotations), l'échange de corpus est facilité.

Il est entendu que la constitution et la normalisation des corpus ne doivent pas nuire à leur lisibilité : pour pouvoir exploiter les corpus à des fins de recherche, la mise en forme doit être facilitée et correspondre aux besoins du chercheur.

L'exploitation linguistique des corpus alignés

Démarche d'exploitation pour les corpus adulte-adulte

Plusieurs démarches sont envisageables selon les objectifs du chercheur. Lorsqu'il s'agit d'analyser la distribution tant lexicale que syntaxique d'une construction, l'équipe a recours à l'analyse des concordances, grâce aux concordanciers existants et libre d'accès sur le net. Elle utilise par ailleurs le logiciel Contextes³ mis à la disposition de l'équipe depuis 2007.

Le recours aux concordanciers permet de mener des études quantitatives grossières sur la fréquence d'une construction ou les caractéristiques les plus évidentes de ses distributions, et ce en dehors de toute considération sur les types de textes concernés. Le logiciel de recherche Contextes est compatible, moyennant des travaux de programmation supplémentaires, avec la version alignée texte et son du corpus. Il permet donc d'observer les réalisations prosodiques des configurations recherchées. En l'absence de logiciels d'analyse, les observations prosodiques restent sommaires mais peuvent néanmoins mettre en évidence les réalisations particulières qui nécessitent une analyse spécifique. Nous illustrons ce point par un exemple.

Dans un corpus d'environ 350 000 mots (état de notre corpus dans les années 2005), une requête fait apparaître la fréquence d'une réalisation particulière du verbe « *penser* ». Ce verbe est réalisé, dans plus de 90% des cas, sous la forme conjuguée, à la première personne du singulier : « je pense ».

La construction avec « penser » se présente de deux façons :

- une construction en « que phrase » :
je pense que ça a apporté beaucoup de choses [athée]
- une construction en « je pense »
à peu près toutes les heures il y avait + euh je pense euh + un quart d'heure euh ou vingt minutes de récré [internat] (Voir Figure 4)

Lorsque la construction apparaît entre deux constructions verbales, comme dans l'exemple suivant, il y a deux possibilités de découpage :

*et mais ça n'a rien à voir je **pense** il y a de très bons chefs qui sont des des ouvriers*

[trombone]

On peut en effet regrouper la construction avec le segment gauche :

*mais ça n'a rien à voir **je pense***

Mais, on peut également regrouper la construction avec le segment droit dans l'énoncé :

*je **pense** il y a de très bons chefs qui sont des ouvriers*

C'est l'accès au son qui permet de déterminer à quel segment doit-on rattacher la construction.

Le corpus en cours de constitution n'étant pas étiqueté, la recherche se fait à partir de la stricte chaîne de caractères. Pour faciliter l'analyse quantitative des distributions, le logiciel prévoit un transfert direct vers un tableur Excel qui permet un comptage rapide et un marquage manuel des catégories par le recours à la fonction filtre. La méthodologie permet ainsi de croiser analyse quantitative et qualitative.

Le corpus peut donner lieu également à une exploitation de type exploration. Il s'agit en ce cas de repérer par une lecture exhaustive les phénomènes langagiers recherchés et pour lesquels des facteurs externes peuvent être prépondérants. Dans ce cas, l'exploitation des corpus passe par le choix des interactions à analyser. Elle peut se faire en fonction de la situation de communication dans laquelle se trouvent les interactants. Dans ce cas, le genre de discours (conversation, réunion de travail, transaction commerciale, etc.), le nombre de locuteurs, leur identité ou les relations qu'ils entretiennent, les thèmes abordés ou encore la longueur des échanges peuvent être des critères de choix pour la sélection des corpus à analyser. Cette première étape est très importante car elle a un fort impact sur les résultats dans la mesure où les analyses tiennent compte des conditions de productions des discours oraux.

Par exemple, on constate que les recherches de « dénomination » sont souvent construites à partir d'un schéma récurrent :

- euh c'est à dire qu'il y a des
- ce sont pas des immeubles +
 - ce sont des petites maisons hein
 - c'est en général en auto-construction [Chili]

The screenshot shows the 'Contextes' application window. The title bar reads 'Contextes'. Below the title bar is a menu bar with 'Fichier', 'Edition', 'Options', and 'Langage ?'. A toolbar contains various icons for file operations. The main area is a table with columns: 'N°', 'div', 'Contexte gauche', 'Occur', and 'Contexte droit'. The table contains 36 rows of search results for the word 'pense'. The first row is highlighted in blue. At the bottom of the window, a status bar shows '12433 paragr', '272 occur', '0,3 s', 'Version: Pro', and 'Fichier: C:\coppu2001\venso.txt'.

N°	div	Contexte gauche	Occur	Contexte droit	
1	escalade	au ventre mais au bout d'un au bout d'un moment on se	pense	plus euh on pense plus au vide euh on voit même plus	
2	escalade	out d'un au bout d'un moment on se pense plus euh on	pense	plus au vide euh on voit même plus la différence euh	
3	japon	super-beaux géniaux etc. euh - le métier d'artiste je	pense	le mot artiste n'a rien à voir avec euh avec euh avec	
4	japon	choses - qu'est le cinéma donc qui peut regrouper je	pense	euh vis-à-vis de choses dont on avait parlé de l'orth	
5	japon	c'est pas le premier film ce sera un court-métrage je	pense	mais avec une équipe parce qu'il y aura des moyens de	
6	japon	e les Thaïlandais ou avec les Japonais c'est euh - je	pense	c'est c'est c'est	
7	japon	est c'est une des questions qui traversera le film je	pense	celle de de - notamment qu'il y a aura des acteurs	
8	japon		12 je	pense	c'est un problème quoi là moi c'est une des questi
9	japon	comme c'est comme le fait de parler ou euh - c'est je	pense	l'art contemporain a a mis a mis en place certaines c	
10	japon	des questions de procédés de choses comme ça et je	pense	l'art contemporain ça a plus été un regard comme ça a	
11	japon	st une une c- c'est une culture insulaire quoi moi je	pense	peut être ça joue énormément de et qui euh - qui a a	
12	japon	je moi je sais pas trop ce que ça veut dire - mais je	pense	c'est important de de aujourd'hui de se poser ces que	
13	japon	euh - mais le - mais je veux pas être totalitaire je	pense	le le quoi le il y a beaucoup de problèmes des des vo	
14	japon	st quelque chose de la photo ou de la peinture moi je	pense	c'est quelque chose de la photographie et : - je pens	
15	japon	ense c'est quelque chose de la photographie et : - je	pense	pas qu'il y ait d'actualité ou d'invention à faire ou	
16	japon	n parce que le film je l'ai pas fait et - non mais je	pense	c'est important de d'abord on c'est un territoire d'e	
17	japon	es parce que c'est le moment où on fait les choses te	pense	qu'on est les choses et - après ouais il y aurait il	
18	japon	oui donc euh - même si elle est post-humaine quoi je	pense	c'est important de le dire et de le répéter mais euh	
19	japon	ça : si va utiliser un procédé parce que j'ai moi je	pense	c'est important d'utiliser des procédés pour raconter	
20	japon	un : un accord ou un désaccord entre eux parce que je	pense	ce serait un couple et même vis-à-vis d'eux-mêmes dan	
21	japon	et de voir comment comment ça réagit quoi sachant je	pense	qu'il y a des choses qui sont euh encore une fois de	
22	japon	lle(s) savoir aussi comment ça sera cadré et - moi je	pense	ça sera ça sera : le cadrage sera fait par un rag	
23	japon	ntéressant s- de faire - c'est pas un collage quoi je	pense	que : c'est ça qui me plaît quoi : parce qu'aujourd'h	
24	job	fesse vraiment attention pour me santé avant tout je	pense	ah - parce que c'est quand même ouais c'est plus de	
25	job		12 euh je	pense	out qu'il y a des patrons qui doivent en profiter euh
26	macdol		12 et puis je	pense	pas que tu montes comme ça aussi facilement tu vois
27	macdol		11 et euh enfin je	pense	qu'il y a des étudiants qui doivent travailler aussi
28	macdol	11 c'est comme tes vacances passées je	pense	pas	
29	Madonna	i c'était absolument le personnage donc euh et euh je	pense	qu'à l'époque il a eu quand même un petit succès et a	
30	Madonna	lus de deux millions d'exemplaires mais maintenant je	pense	qu'il est à plus de cinq millions d'exemplaires	
31	Madonna	bien jouer ce personnage donc elle s'est vraiment je	pense	qu'elle s'est vraiment euh abaissée pour pouvoir euh	
32	Madonna		11 heu je ne	pense	pas
33	Madonna		11 je ne	pense	pas étant donné que après Evita elle é - était enceint
34	Madonna	elle sera sur NRJ en direct ainsi que sur euh TF1 je	pense	à vingt heures parce que en général elle n'accorde qu	
35	Madonna	es interviewés à TF1 jamais aux autres chaînes donc je	pense	que c'est TF1 mais euh	
36	Madonna	stiquant bon maintenant qu'elle a ses deux enfants je	pense	qu'elle va ralentir le rythme il y a une tournée de p	

Figure 4. Aperçu des résultats de la requête « penser » avec le logiciel Contextes

il y avait quand même une espèce de comment dirais-je

- pas des seigneurs mais enfin
- une bourgeoisie qui possédait des terrains
 - qui possédait des maisons
 - qui possédait beaucoup de choses

Il peut être intéressant dans ce cas d'observer l'impact du type de texte concerné dans la constitution de ces définitions.

Ce dernier type d'analyse est possible sur la version papier du corpus à partir d'une exportation en HTML, de la transcription réalisée à l'aide de Transcriber. Une première lecture permet de repérer les phénomènes langagiers intéressants, et d'annoter « à la main » sur la transcription papier. Les corpus

étant alignés texte-son, il est aisé d'atteindre le segment sonore associé au passage choisi.

Les études pouvant être menées dans le cadre de cette approche sont nécessairement limitées à des données relativement restreintes. Il est en effet peu envisageable de traiter « à la main » plusieurs centaines de milliers voire plusieurs millions de mots.

Démarche d'exploitation pour les corpus adulte-enfant

Pour ce qui concerne les corpus d'interaction adulte-enfant, l'analyse recèle quelques spécificités. Nous procédons à une première analyse des énoncés de l'adulte et de l'enfant en grandes catégories syntaxiques : énoncés comportant maximale une construction « simple », plusieurs constructions « simples » juxtaposées ou coordonnées, une ou plusieurs constructions « complexes ». Le repérage des constructions dites complexes s'effectue en fonction d'une liste d'« introducteurs de complexité » (correspondant globalement à des éléments de subordination). L'annotation des énoncés selon ces catégories est manuelle, soit sur une grille, soit directement sur la version imprimée du corpus (voir Tableaux 4 et 5).

La phase d'analyse quantitative vient ensuite : pour chaque catégorie et dans chaque corpus, nous mentionnons son pourcentage afin de repérer s'il existe une évolution dans le temps (voir Tableau 6).

Une seconde analyse vient alors compléter la première. On repère les constructions complexes pour lesquelles il y a une évolution générale (en termes quantitatifs et qualitatifs). Nous procédons à l'observation non outillée d'éléments qui semblent s'installer dans des contextes différents, mais aussi des reprises par l'enfant des verbalisations de l'adulte (soit sous forme d'essais c'est-à-dire de tentatives non abouties, soit sous forme de reprises immédiates ou différées, identiques ou avec des variations). Par la suite, une attention particulière peut être portée à une construction qui semble pertinente au regard de l'évolution générale en corrélation avec le langage adressé à l'enfant (voir Tableau 7).

Dans cette démarche, c'est essentiellement la phase d'annotation et de repérage des phénomènes qui pourrait être facilitée par des outils informatiques. Pour la phase de repérage, les concordanciers simples (comme Antcon) sont une première aide. Pour la phase d'annotation, des logiciels tels que N'Vivo sont à l'étude actuellement.

Tableau 4

Grille originale des introducteurs de complexité (d'après L. Lentin, 1998)

à + verbe à l'infinitif	17	Ex : elle donne à <i>manger</i>
Comme (= étant donné que)	<u>16</u>	Ex : <i>comme</i> la cloche a sonné, ils sont rentrés dans la classe
Comparative (= comme)	<u>21</u>	Ex : elle écrit <i>comme</i> elle parle
de + verbe à l'infinitif	18	Ex : elle n'est pas contente <i>de voir</i> sa maman
discours indirect (paroles rapportées)	<u>13</u>	Ex : elle <i>dit de</i> faire la vaisselle ; elle <i>dit que</i> je suis beau
divers	<u>24</u>	Opposition (tandis que, alors que), tellement que, sans que, surtout que, déjà que, sinon, etc
Extraction	1	C'est, voilà, il y a qui, que, où, etc
Gérondif	<u>20</u>	Ex : en marchant
il faut que	4	Ex : il <i>faut que</i> j'aille aider mes amis
interrogative indirecte	12	Ex : je sais ce que, je vais voir comment, je (ne) sais (pas) si je vais aller
introducteurs temporels (autres que « quand »)	<u>7</u>	dès que, après que, chaque fois que, pendant que, avant que, etc
Où relatif	<u>14</u>	Relatif ; ex : elle a trouvé le paquet <i>où</i> il y a les couches du bébé
parce que	<u>8</u>	Ex : le bébé ne boit pas <i>parce qu'il</i> a déjà bu son lait
pour + verbe à l'infinitif	19	Ex : papa sort sa clé <i>pour ouvrir</i> la porte
pour que	<u>15</u>	Ex : elle appuie un petit peu <i>pour que</i> le dentifrice sorte du tube
puisque	<u>23</u>	Ex : il va jouer tout seul <i>puisque</i> sa copine est partie
Quand (= lorsque)	<u>6</u>	Ex : ils sont contents <i>quand</i> ça roule
Quantitative (comparatif)	<u>11</u>	Plus, moins, autant...que ; ex : Paul est <i>plus grand que</i> Pierre
Que conjonction	<u>5</u>	Ex : elle trouve <i>que</i> c'est drôle (« que » après un verbe conjugué)
Que relatif	<u>22</u>	Ex : elle ouvre le cadeau <i>que</i> son papa lui a donné
Qui relatif	<u>3</u>	Ex : il sort du camion un canapé <i>qui</i> est très lourd
Si (supposition et condition)	<u>9</u>	Ex : <i>si</i> le chat griffe le ballon il peut éclater
verbe + verbe à l'infinitif (sauf le futur proche)	10	Ex : il <i>veut rentrer</i> dans sa chambre ; il <i>peut aller</i> le chercher

Tableau 5
Extrait d'une analyse en catégories syntaxique d'énoncés

Prénom et âge de l'enfant : Amel, 6ans 3mois								N° corpus : 2
Numéro de l'énoncé	Énoncés comportant une construction simple	Énoncés comportant des constructions simples multiples (juxtaposées ou coordonnées)	Énoncés comportant des constructions complexes (cf liste IC)	Énoncés comportant des essais (cf liste IC)	Fiches de commentaires			
E= Enfant			+					
A= Adulte			/ :					
	E	A	E	A	E	A	E	A/E
A21				X				
E21			X					
A22						12/10		
E22					<u>5</u> /7/4			Essai 7 ? Pb sémantique (après et non avant)
A23						1/12/1		Aspectuel « finir de »
E23					12 inc	0		Ou essai ?
A24						19/12 inc		
E24			X					
A25						<u>20</u> /12/ <u>13</u> inc		
E25					<u>20</u> inc			Essai <u>20</u> /12 ?
A26						12 inc		
E26			X					
A27		X						
E27							18 pour <u>5</u> ou 19	
A28						<u>5</u>		
E28	X							
A29		X						
E29					19			
A30						<u>13</u>		
Total	11	13	6	3	6	14	6	

Tableau 6
Évolution de la production des constructions syntaxiques
chez l'adulte et l'enfant

N° du corpus		1		2	
Adulte (A)		A	E	A	E
Enfant (E)		(/25)	(/24)	(/30)	(/29)
Constructions simples (%)		64	66,66	43,33	37,93
Constructions simples multiples (%)		4	8,33	10	20,69
Constructions complexes (%)		Total	25	46,66	41,38
	<i>Réalisées</i>	32	16,66	46,66	10,34
	<i>En incomplétude</i>	0	0	20	6,9
	<i>En essais</i>	-	8,33	-	20,68
	<i>Comportant 2 IC</i>	0	0	10	3,44
	<i>Comportant plus de 2 IC</i>	4	0	6,66	3,44

Tableau 7
Exemple de début d'analyse tirée des tableaux précédents

Commentaires sur le langage de l'enfant :

- Augmentation du pourcentage de constructions simples multiples et de constructions complexes (essais compris).
- Pour les constructions syntaxiques complexes : forte augmentation des essais (l'enfant tente plus de constructions) et dans le même temps, apparition des constructions comportant plus d'un introducteur de complexité. Phénomènes à observer : les tentatives sur « pendant que ».

Commentaires sur le langage de l'adulte :

- Evolution en parallèle de l'enfant : augmentation du pourcentage de constructions simples multiples et de constructions complexes.
- Hypothèses : les propositions de l'adulte se situent dans la zone proche du développement syntaxique de l'enfant et l'amène ainsi à tenter des constructions non encore maîtrisées comme les subordonnées temporelles : faire l'analyse qualitative des interactions (reprises et reformulations).

Prospectives

Nous aimerions en guise de conclusion poser quelques questions sur le potentiel des logiciels :

- faut-il un outil idéal? Pour faire quoi? Un logiciel qui remplacerait complètement l'annotation manuelle est-il envisageable? Souhaitable?

Les chercheurs interviennent toujours manuellement sur leur corpus car il paraît peu probable, dans l'état actuel des connaissances, que par exemple le repérage des questions ou encore des recherches de dénomination soit réalisable;

- n'est-ce pas illusoire d'attendre un outil générique adapté à tous les utilisateurs? Ne faut-il pas plutôt des outils plus performants qui serviraient à l'ensemble de la communauté travaillant aussi bien sur des données orales qu'écrites (un logiciel de type Contextes libre et gratuit)?
- cette question des logiciels est-elle essentielle? En effet, les chercheurs n'utilisent souvent qu'une partie ou détournent les fonctionnalités des logiciels pour leur usage propre et il en sera probablement toujours ainsi. N'est-ce pas plutôt la question du cumul des connaissances par le cumul des exploitations qui est à envisager comme priorité?

Notes

¹ Les formats Transcriber et TEI sont l'un et l'autre définis comme des vocabulaires XML. Sur le plan pratique, cela signifie qu'on passe de l'un à l'autre via des outils classiques dans le monde XML comme des feuilles de style XSL.

² Dans un fichier non aligné texte-son, on peut néanmoins savoir que la prise de parole d'un second interlocuteur débute après un mot précis émis par le premier interlocuteur; le format TEI permet également la prise en compte de ce type d'information.

³ Logiciel créé par Jean Véronis (Université de Provence). Nous le remercions.

Références

- Baude, O. (Éd.). (2006). *Corpus oraux. Guide des bonnes pratiques*. Paris : CNRS Éditions.
- Benzitoun, C., Campione, E., Deulofeu, J., Henry, S., Teston, S., Valli, A., & Véronis, J. (2004, 15 Mai). *L'analyse syntaxique de l'oral : problèmes et méthode*. Communication présentée aux Journées d'Études de l'ATALA, Evans : Méthodes et outils pour l'évaluation des analyseurs syntaxiques, Paris.
- Bilger, M. (Éd.). (2000). *Linguistique sur corpus. Études et réflexions*. Perpignan, France : Presses universitaires de Perpignan.
- Canut, E. (Éd.). (2008). Des lectures partagées aux recherches sur corpus. Aperçu de réflexions actuelles. *L'Acquisition du langage oral et écrit (ALOE)*, [Numéro thématique], 60-61.

- Cappeau, P., & Sejjido, M. (2005). *Les corpus oraux en français (inventaire 2005 v.1.0)*. Document consulté le 20 septembre 2008 de http://www.culture.gouv.fr/culture/dglf/recherche/corpus_parole/Presentation_Inventaire. Pdf.
- Debaisieux, J.-M. (2005). Les corpus oraux : situation, exploitation linguistique. Bilan et perspectives. *Scolia*, 19, 9-40.
- Équipe DELIC (2004). Autour du Corpus de référence du français parlé. *Recherches sur le français parlé*, 18, 11-42.
- Habert, B., Nazarenko, A., & Salem, A. (1997). *Les linguistiques de corpus*. Paris : Colin.
- Savelli, M. (Éd.). (2005). Corpus oraux et diversité des approches. *Lidil*, [Numéro thématique], 31.
- Véronis, J. (2000). Annotation automatique de corpus : panorama et état de la technique. Dans J.M. Pierrel (Éd.), *Ingénierie des langues* (pp. 111-129). Paris : Hermès.

Références complémentaires

Liste des sites internet des institutions, bases de données, corpus et logiciels :

- Analyse et traitement informatique de la langue française (ATILF), UMR CNRS : www.atilf.fr
- Base textuelle FRANTEXT : <http://atilf.atilf.fr/frantext.htm>
- British national corpus : www.natcorp.ox.ac.uk
- Centre de ressources pour la description de l'oral (CRDO) : crdo.risc.cnrs.fr/
- Centre national de ressources textuelles et lexicales : <http://www.cnrtl.fr/>
- Corpus de langue parlée en interaction (CLAPI) : <http://clapi.univ-lyon2.fr>
- Corpus de référence de l'espagnol actuel (CREA) : pheme.rae.es/java.ext/corpus.htm
- Corpus search, management and analysis system (COSMAS II) : <http://www.ids-mannheim.de/cosmas2/>
- Délégation générale à la langue française et aux langues de France : <http://www.culture.gouv.fr/culture/dglf/>
- Enquête sociolinguistique à Orléans (ESLO) : <http://www.univ-orleans.fr/eslo/>
- Logiciel ANTCONC : http://www.antlab.sci.waseda.ac.jp/antconc_index.html
- Logiciel JTRANS : <http://jtrans.gforge.inria.fr>
- Logiciel TRANSCRIBER : <http://trans.sourceforge.net/>
- Projet PFC (Phonologie du français contemporain) : <http://www.projet-pfc.net/>

Virginie André est maître de conférences à l'Université Nancy 2 et rattachée au laboratoire CNRS ATILF. Docteur en Sciences du langage, ses travaux portent plus particulièrement sur l'analyse des interactions verbales et la construction collaborative du discours notamment en situation de travail. Ses domaines d'intérêt concernent également la constitution, l'exploitation et la diffusion des corpus oraux ainsi que l'implication et l'application des sciences du langage, notamment dans le domaine de la formation linguistique.

Christophe Benzitoun est maître de conférences à l'Université Nancy 2 et rattaché au laboratoire CNRS ATILF. Docteur en Sciences du langage, diplômé de l'Université de Provence, il travaille plus particulièrement sur l'analyse syntaxique du français parlé et la question de l'informatisation des données orales.

Emmanuelle Canut est maître de conférences à l'Université Nancy 2 et rattachée au laboratoire CNRS ATILF. Docteur en Sciences du langage, elle s'intéresse à l'acquisition du langage par l'enfant et notamment aux interactions entre adulte et enfant dans la mise en fonctionnement du langage. Cette recherche a des implications dans le domaine scolaire et périscolaire avec la mise en œuvre d'actions et de formations à destination des professionnels de l'enfance.

Jeanne-Marie Debaisieux est maître de conférences au département de Sciences du langage de l'Université Nancy 2 et rattachée au laboratoire CNRS ATILF. Ses travaux portent sur l'analyse syntaxique et pragmatique de la langue parlée et sur la méthodologie de recueil et de traitement des données orales, tant à des fins de recherches linguistiques que d'applications en didactique des langues.

Bertrand Gaiffe est chargé de recherche dans l'UMR ATILF rattachée à l'Université de Nancy 2. Docteur en informatique, il travaille plus particulièrement la constitution, l'annotation, la normalisation, le versionnage de données textuelles de nature variée. Dans un cadre plus large, il est fortement impliqué dans le projet européen CLARIN dont l'un des objectifs est de fournir des données textuelles et sonores et des plateformes de traitements à l'échelle européenne.

Evelyne Jacquy est chargée de recherche dans l'UMR ATILF rattachée à l'Université de Nancy 2. Docteur en linguistique-informatique, elle travaille plus particulièrement sur l'analyse sémantique du français, écrit et parlé, par le biais de traitements et d'annotations de corpus. Elle s'intéresse aussi à la constitution de corpus écrits et oraux sur le français.

Méthodes d'analyse des communications fonctionnelles en situation de travail collectif

Guillaume Gronier, Ph.D.

Centre de recherche public Henri Tudor

Résumé

L'activité verbale en situation de travail constitue l'une des méthodes privilégiées pour l'analyse des processus coopératifs en psychologie ergonomique. Les communications sont ainsi considérées comme une trace externe révélant la mise en œuvre de processus cognitifs. L'analyse de ces communications s'avère d'autant plus nécessaire dans des contextes hautement coopératifs pour comprendre les mécanismes sous-jacents aux prises de décisions, comme en conception de produits. Cet article propose de comparer plusieurs méthodes d'analyse qualitative pour l'étude des communications fonctionnelles en réunion de projet. Ainsi, les méthodes COMET, ALCESTE et une grille d'analyse des confrontations de points de vue sont appliquées au même corpus textuel. Les résultats de chaque analyse sont comparés aux deux autres afin de dégager les complémentarités et les limites des trois méthodes.

Mots clés

MÉTHODES D'ANALYSE QUALITATIVE, COOPÉRATION, MÉTHODE ALCESTE, MÉTHODE COMET, CONFRONTATIONS DE POINTS DE VUE.

Introduction

Le travail collectif constitue l'une des thématiques les plus étudiées en psychologie ergonomique depuis le début des années 1990 (Navarro, 1993). Cette évolution résulte d'une modification progressive des contextes de travail où la coopération s'inscrit désormais au cœur de nouveaux modèles d'organisation du travail axés sur une plus grande flexibilité des activités productive, individuelle et collective. Dès lors, l'analyse des communications verbales constitue l'un des moyens privilégiés pour l'étude des situations de travail collectif. Les actes langagiers occupent en effet des fonctions importantes de synchronisation et de coopération entre les opérateurs.

Cependant, l'analyse des communications soulève de nombreuses questions d'ordre méthodologique, car les dialogues coopératifs se co-construisent au fur et à mesure de l'interaction. La technique du codage, à

partir d'une grille préalablement définie qui mettra l'accent sur certains aspects du dialogue coopératif (construction de sens, ajustement mutuel, processus argumentatifs, etc.), constitue l'une des méthodes les plus utilisées par les chercheurs en psychologie ergonomique. En effet, ces codages systématiques ont l'avantage de proposer une méthodologie rigoureuse pour l'analyse de l'activité coopérative, mais sont parfois trop restrictifs et perdent beaucoup d'informations qui pourraient aussi servir le chercheur. Un logiciel d'analyse qualitative, tel que ALCESTE (Analyse lexicale par contexte d'un ensemble de segment de texte) (Reinert, 1990), offre la possibilité d'étudier les aspects lexicaux des échanges verbaux coopératifs, sans se restreindre, *a priori*, à une analyse sémantique prédéfinie.

Les communications que nous abordons dans cette recherche s'inscrivent dans le cadre particulier de la conception de produits. En effet, les communications sont déterminantes et omniprésentes dans les activités coopératives de conception, car ce sont elles qui permettent l'évolution des représentations mentales et la redéfinition du problème de conception (Darses, 2006). Elles favorisent également l'intégration du point de vue des différents métiers (ingénieur et ergonome pour le cas de notre étude) travaillant sur un même projet.

La recherche que nous présentons dans cet article rend compte d'une réflexion méthodologique sur l'analyse des communications fonctionnelles en conception. Nous avons retenu trois méthodes d'analyse appliquées au même corpus afin d'être comparées les unes aux autres.

Communications fonctionnelles et travail collectif

Dans un ouvrage sur le fonctionnement du média, Moles (1986) définit la communication comme :

l'action de faire participer un individu ou un système, situé en un point donné R, aux stimuli et aux expériences de l'environnement d'un autre individu ou système situé en un autre lieu et à une autre époque E, en utilisant les éléments de connaissance qu'ils ont en commun (p. 25).

Cette définition suggère notamment que la communication utilise des *éléments de connaissance* que les interlocuteurs ont en commun. Ces éléments (relatifs à l'expérience, aux compétences, aux métiers, etc.) sont regroupés en *répertoire*, c'est-à-dire un constitutif mnésique des savoirs d'un individu. Dès lors, l'entente et la compréhension entre plusieurs interlocuteurs dépendront de leur capacité à faire coïncider leur répertoire.

L'étendue des formes de communications possibles entre plusieurs individus nécessite de réduire leur champ d'analyse selon les intérêts du

chercheur. En ce qui nous concerne, nous nous limitons aux *communications fonctionnelles*, c'est-à-dire aux communications dans le travail, « regardant directement le contenu du travail réalisé, excluant ainsi celles qui sont prioritairement centrées sur les relations humaines dans l'équipe, la cohésion, les processus d'influence, etc. » (Savoyant & Leplat, 1983, p. 247).

L'élaboration d'un répertoire commun optimal se fonde sur une étroite coopération entre les interlocuteurs (Falzon, 1994). Plus précisément, il s'agira d'étudier les communications fonctionnelles en tant que *dialogues*, à travers lesquels sont mis en œuvre des processus de modélisation du partenaire qui favorisent l'adaptation même des dialogues. Falzon (1994) souligne que le dialogue suppose une construction collective qui nécessite la prise en compte de l'état de l'interlocuteur. Ainsi,

d'une part, chacun va prélever dans le discours de l'autre des indices permettant de s'assurer du bon fonctionnement de la communication et de construire un modèle approprié du partenaire. Ce modèle permet d'adapter sa production verbale aux connaissances, objectifs, intérêts du partenaire [...]. D'autre part, chacun va volontairement fournir des indices facilitant au partenaire la réalisation de cette tâche : accusés de (bonne ou mauvaise) réception, demandes de clarification, reformulations, etc. (Falzon, 1994, p. 303).

La faculté à faire coïncider plusieurs répertoires et le partage de connaissances ont aussi été étudiés par Grusenmeyer et Trognon (1997). L'analyse d'une séquence d'échanges verbaux, selon une méthodologie issue de la logique illocutoire, a montré que deux opérateurs se constituent, *in situ*, une représentation fonctionnelle partagée. Il s'agit alors d'une représentation circonstancielle, finalisée et transitoire, construite coopérativement au sein de (et grâce à) l'interaction verbale. Il y a donc une évolution des représentations des opérateurs entre le début et la fin de l'échange. Les répertoires individuels se recouvrent au fur et à mesure de la communication pour passer d'une compréhension partielle à une compréhension plus complète.

Les communications fonctionnelles occupent une place importante dans les activités collectives en conception de produit. En effet, des situations de co-conception émergent des processus de travail collectif, au cours desquelles « les partenaires de la conception développent la solution conjointement : ils partagent un but identique et contribuent à son atteinte grâce à leurs compétences spécifiques » (Darses, 1997, p. 52). Certains processus communicatifs prennent alors une importance majeure, comme l'argumentation, la construction et le maintien du contexte partagé,

l'intégration des points de vue, etc. (Darses, 2002; Martin, 2001). Ces processus permettent en effet « de construire et de développer un espace d'intersubjectivité, de compréhension réciproque » (Zarifian, 1998, p. 115), en vue d'établir des accords solides sur la nature des problèmes, l'identité des objectifs, le sens donné aux actions et les implications des mobiles individuels. Il s'agit également de co-construire un référentiel commun (de Terssac & Chabaud, 1990) afin de garantir l'efficacité de l'action collective.

Karsenty (2000) a également montré que, dans le travail collectif, l'explication qui s'exprime au cours de l'élaboration des points de vue contribuait à enrichir la représentation partagée du problème à résoudre. Dès lors, l'intégration des points de vue est considérée « comme le processus central de convergence vers la solution, et c'est de son efficacité que dépend le succès des organisations coopératives de la conception » (Darses, 2002, p. 291).

Méthode

Objectif de la recherche

L'objectif principal de cet article est de proposer une étude comparative de l'analyse d'un même corpus textuel, issu de l'enregistrement et de la retranscription d'une réunion de travail dédiée à la conception d'un produit, à l'aide de plusieurs méthodes d'analyse des communications fonctionnelles. Ainsi, deux grilles de codage et un logiciel d'analyse textuelle ont été retenus :

- la méthode COMET (Darses, Détienne, Falzon & Visser, 2001), qui a principalement été élaborée pour l'analyse et la modélisation des activités coopératives de conception;
- une grille de codage développée pour l'analyse spécifique des processus de confrontation de points de vue en ingénierie concurrente (Détienne, Martin & Lavigne, 2005);
- une analyse par le logiciel de traitement textuel ALCESTE (Reinert, 2001).

En appliquant chacune de ces méthodes au même corpus textuel, nous souhaitons mettre en évidence la complémentarité de chacune d'entre elles. Néanmoins, l'accent sera davantage porté sur l'analyse réalisée par l'intermédiaire du logiciel ALCESTE, afin de montrer dans quelle mesure cette méthode peut servir à l'analyse des communications fonctionnelles en conception de produit, au même titre que d'autres méthodes développées spécifiquement à cet effet.

Corpus et cadre d'analyse

Le corpus que nous étudions est extrait de l'observation d'un projet de conception dans le cadre d'un partenariat industriel. Ce projet, qui impliquait différents acteurs-métiers (ingénieurs et ergonomes), avait pour objectif le développement d'un module de confort initié par un grand équipementier automobile. Par module de confort, il faut entendre le dispositif informationnel qui regroupe un ensemble d'informations considérées non indispensables à la réalisation de la tâche de conduite. Nous y trouvons ainsi des informations qui ont pour principal objectif d'améliorer le confort du conducteur (climatisation, ventilation, autoradio, etc.). Soulignons toutefois que la complexité et le nombre croissant d'éléments qui viennent aujourd'hui s'intégrer au module de confort (système GPS, téléphonie, etc.), peuvent perturber la tâche principale de conduite et nécessitent l'intervention de l'ergonomie pour une meilleure adéquation du module à l'activité du conducteur.

Le projet intégrait neuf ingénieurs et deux ergonomes garants du facteur humain, qui avaient pour tâche de proposer à l'équipementier le prototype d'un concept de produit. Sur un total de 19 réunions organisées au cours du projet, nous avons sélectionné celle qui nous a semblé la plus riche en termes d'échanges et de débats argumentatifs. Celle-ci dura deux heures et 24 minutes, pour une retranscription de 15 682 mots.

Comparaison des méthodes d'analyse des communications fonctionnelles

Chacune des méthodes d'analyse des communications fonctionnelles rappelées ci-dessus (COMET, grille de codage pour la confrontation de points de vue, ALCESTE) sera présentée puis appliquée au même corpus textuel. Chaque analyse de corpus ainsi obtenu sera commentée et discutée. Nous nous attacherons ensuite à dégager la complémentarité de ces trois méthodes.

La méthode COMET

Présentation de la méthode

Définie dans le cadre de l'axe méthodologique du projet Eiffel (Coopération et cognition en conception) initié par l'Institut national de recherche en informatique et en automatique (INRIA), COMET est une méthode d'analyse des processus de conception collective (Darses, Détienne, Falzon & Visser, 2001). Elle vise à modéliser les activités coopératives mises en œuvre dans les situations de co-conception, c'est-à-dire lorsque des concepteurs, au cours de réunions, travaillent ensemble sur des projets de conception.

La méthode COMET permet de distinguer un niveau fonctionnel et un niveau coopératif. Le niveau fonctionnel examine la conception collective du

point de vue des actions et des objets mis en œuvre dans des réunions. Le niveau coopératif fait apparaître des séquences d'actions correspondant à des mouvements coopératifs.

L'application de la méthode COMET, résumée dans le tableau 1, consiste à découper les tours de parole des différents participants en une ou plusieurs *unités* selon un schéma de codage basé préalablement sur un *prédicat* (ACT). Chaque prédicat s'appuie ensuite sur un ensemble d'*arguments* (OBJ) possible, éventuellement modulée (MOD) par une *requête* (la valeur par défaut du mode est l'*assertion*).

Application de la méthode : résultat et discussion

La méthode COMET permet de dégager plusieurs caractéristiques de la structure des communications fonctionnelles. Pour commencer, il apparaît que chaque principe de solutions est systématiquement accompagné de justifications. On observe en effet que les ingénieurs et les ergonomes tentent en permanence de « se comprendre » (Zarifian, 1998) et invitent leur interlocuteur à étayer ses propos. Par exemple, on relève que l'ingénieur remet en question des propositions de solution faites par l'ergonome, ce qui encourage celui-ci à apporter de nouveaux arguments en vue de définir l'espace problème (Tableau 2).

L'analyse de la réunion par la méthode COMET révèle également que les ingénieurs reprennent régulièrement les objectifs fixés par le projet et rappellent les données du problème dans les termes du cahier des charges. Les ingénieurs reviennent ainsi régulièrement sur les contraintes techniques liées aux solutions qui sont proposées.

Une grille de codage pour la confrontation de points de vue

Présentation de la méthode

Martin, Détienné et Lavigne (2002) (voir aussi Détienné, Martin & Lavigne, 2005), dans une recherche concernant la confrontation de points de vue entre concepteurs, proposent une grille de codage qui permet d'analyser les points de vue mis en jeu dans la co-conception afin d'en dégager des règles pour les recommandations ergonomiques au niveau logiciel et organisationnel.

Cette grille de codage décompose les énoncés verbaux, échangés au cours de réunions de projet, en quatre niveaux (résumés dans le Tableau 3) :

- 1- le premier niveau concerne les éléments généraux sur lesquels reposent les débats entre les acteurs de la réunion (les *propositions*);
- 2- le second niveau se caractérise par les éléments illocutoires qui servent directement les éléments généraux. Autrement dit, le second niveau concerne directement les propositions;

Tableau 1
Éléments de codage de la méthode COMET

Modulation (MOD)	Prédicat (ACT)	Argument (OBJ)
Assertion	Génération (GEN)	Données du problème (DAT)
Requête (REQ)	Évaluation (EVAL)	Éléments de solution (SOL)
	Interprétation (INT)	Objets du domaine (OBJ)
	Information (INFO)	Règles ou procédures du domaine (PROC)
		But (GOAL)
	Tâche (TASK)	

Tableau 2
Illustration d'un échange entre ingénieur et ergonome

Acteur	Tour de parole	Codage
Ingénieur	Il y a une contradiction là.	[EVAL/DAT]
Ergonome	Pourquoi ?	REQ [INT/DAT]
Ingénieur	C'est valable pour l'autoradio mais pas pour la clim ?	REQ [EVAL/OBJ]
Ergonome	La clim ? Ramener au volant la clim, non. On va à l'encontre des stéréotypes là, complètement.	[INFO/DAT]
Ingénieur	Le fait de dire, euh, ça me dérange pas pour l'autoradio de faire des allers et retours à l'écran et sur la route, euh...	[EVAL/DAT]
Ergonome	L'autoradio, la plupart du temps, les gens ne le manipulent pas tant que ça en conduisant.	[INFO/OBJ]

- 3- le troisième niveau, constitué des *divergences* et des *convergences*, a pour fonction d'articuler le premier niveau avec le deuxième niveau décrits précédemment;
- 4- le quatrième niveau a pour objectif de conclure l'échange et les procédures interlocutoires depuis la proposition d'une idée.

Tableau 3
Résumé des niveaux de codage
pour l'analyse de la confrontation de points de vue

Niveau 1	Niveau 2	Niveau 3	Niveau 4
Proposition : expression d'une idée, proposition d'une solution	<ul style="list-style-type: none"> - Question : remise en question de la proposition - Argument : raison donnée pour faire admettre une proposition - Contre-argument : raison donnée pour rejeter une proposition - Réitération : réintroduction d'un élément déjà évoqué et traité au cours de la réunion - Répétition : redite d'un élément afin d'appuyer une idée - Confirmation : demande de répétition 	<ul style="list-style-type: none"> - Divergence : incompréhension ou mise en doute d'une proposition - Convergence : acceptation d'une proposition 	<ul style="list-style-type: none"> - Stabilisation positive : après une ou plusieurs convergences; les acteurs trouvent un accord sur une solution - Stabilisation négative : après une ou plusieurs divergences, les acteurs ne trouvent pas d'accord et restent sur un statu quo

Chaque tour de parole est codé par un élément d'un des niveaux. La structure par niveau qui se dégage de l'analyse permet d'appréhender les communications fonctionnelles en termes d'organisation hiérarchique.

Application de la méthode : résultat et discussion

De manière générale, on observe qu'il s'organise une co-construction de solutions : une idée émise par un acteur de la réunion en appelle une autre. Par exemple, sur la base de l'extrait présenté ci-dessous (Figure 1), on peut relever que la proposition faite par l'acteur A1 (ergonome) l'amène à proposer une autre idée, qui est ensuite affinée et reformulée en une autre proposition, puis argumentée jusqu'à ce que l'acteur A6 propose à son tour une nouvelle idée sur

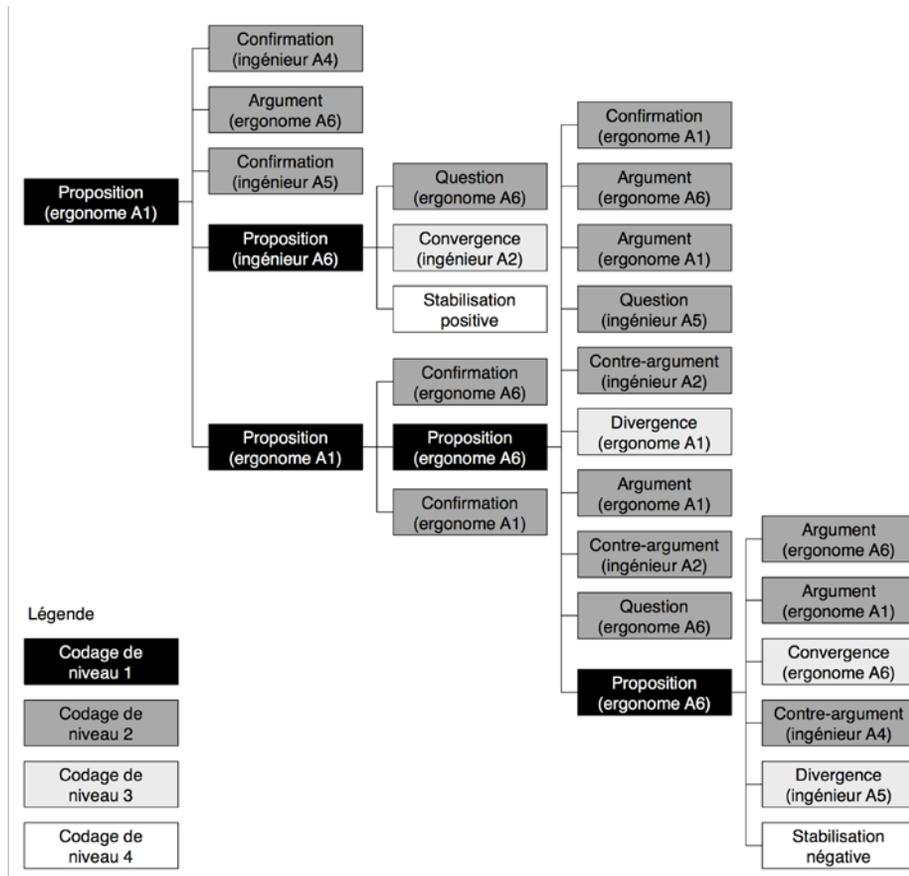


Figure 1. Extrait d'encodage et de représentation par arborescence des confrontations de points de vue.

la base des précédentes. Au cours de ces échanges, l'acteur A6 propose une idée connexe qui est acceptée (stabilisation positive) par l'ensemble des intervenants.

De plus, on remarque que les ergonomes argumentent ou contre-argumentent les propositions faites par les ingénieurs. Si, à l'aide de la méthode COMET, nous avons souligné que les ingénieurs remettaient en question les idées des ergonomes, et les encourager à apporter de nombreux arguments, la grille de codage pour l'analyse des confrontations de points de vue met en évidence que les ergonomes argumentent également les idées proposées par les

ingénieurs. En somme, les ergonomes enrichissent les dialogues d'éléments argumentatifs. À ce sujet, Darses (2006) a démontré l'importance des activités argumentatives au sein des tâches collectives. L'auteur relève en effet que les arguments sont

un moyen de comprendre le problème, d'analyser l'espace de conception, de construire des hypothèses communes, de développer un processus de confrontation de points de vue et de délibération aboutissant à la prise de décision. (Ainsi,) c'est au travers de l'argumentation que sont examinées les propositions de solutions, qu'elles sont évaluées et que des alternatives sont proposées (p. 319).

Cela montre l'importance accordée aux débats au cours des réunions de projet, ainsi que la dynamique des échanges coopératifs.

La méthode ALCESTE

Présentation de la méthode

Le logiciel ALCESTE repose sur la méthodologie d'Analyse Lexicale par Contexte d'un Ensemble de Segment de TEXte (ALCESTE) développée par Reinert (2001). Cette méthodologie consiste en un traitement statistique effectué sur un corpus textuel. Ce corpus est traité en trois grandes étapes successives qui permettent, tour à tour, d'identifier des *unités de contexte élémentaires (uce)* qui sont un découpage du corpus original, d'identifier les « formes réduites » du vocabulaire contenu dans le corpus, puis de calculer les occurrences des formes préalablement identifiées par rapport aux *uce*, à l'aide d'une analyse statistique par *classification descendante hiérarchique*. Ainsi, le logiciel repère les oppositions les plus fortes entre les mots, pour en extraire des *classes* d'énoncés représentatifs, c'est-à-dire des groupes de vocabulaire représentatifs du corpus analysé. Chaque mot est associé à une classe par un test de χ^2 (à un degré de liberté). Plus la valeur du χ^2 est élevée, plus le mot contribue à expliquer la classe à laquelle il est associé. C'est à partir de ces classes que le chercheur pourra apporter sa contribution en termes d'analyse et d'interprétation.

Pour le traitement par ALCESTE, le corpus textuel a été codé de manière à identifier chaque tour de parole en fonction du métier du concepteur (ingénieur ou ergonomiste) à l'aide d'un codage reconnu par ALCESTE (lignes étoilées : *Ingenieur; *Ergonome). Ce codage permet d'identifier les modalités de la variable indépendante (VI) qui nous intéressent pour notre analyse.

Application de la méthode : résultat et discussion

Sur la base de notre corpus textuel, ALCESTE a différencié trois classes lexicales, représentées par 47,56 % des *uce* sur l'ensemble de celles identifiées par le logiciel (1245 *uce* ont été classées sur un total de 2618). Nous présentons, dans le Tableau 4, les 10 mots (occurrences) et les variables hautement significatives ($p < .001$) de chaque classe dont le χ^2 est au moins supérieur à 15 (ddl = 1). Afin de faciliter notre analyse, nous avons choisi de différencier quatre catégories lexicales : les noms communs, les verbes et les adjectifs. Notons que les mots de chacune de nos catégories sont classés par ordre décroissant de χ^2 .

Analyse et interprétation de la classe 1

On remarque que l'univers lexical de la classe 1 regroupe des termes généraux relatifs au projet de conception (concept, préconcept, idée, réunion, premier et deuxième qui désignent chacun des préconcepts), ainsi que des termes techniques relatifs au module de confort (interface, clavier, commande, commande vocale). Le projet est donc abordé du point des éléments techniques du concept, avec une description des éléments qui le composent.

En termes de modalités de variable indépendante, c'est le *métier d'ingénieur* ($\chi^2 = 64.46$; $p < .001$) qui tient le rôle principal. En effet, les solutions techniques, reprises par l'ingénieur, s'inscrivent dans un modèle de la rationalité technique (Daniellou, 1994) à travers lequel le projet est conçu comme un processus de résolution de problème. Dans notre étude, l'ingénieur cherche à répondre à des questions d'ordre technique : comment intégrer une technologie dans le module de confort, ou comment réaliser les interfaces embarquées.

Analyse et interprétation de la classe 2

L'univers lexical de la classe 2 regroupe un vocabulaire spécifique à l'interface et aux commandes du module. En effet, les termes de cette classe font référence aux informations et aux actions qui pourront être effectuées sur cette interface et sur les commandes mécaniques du module de confort.

Ainsi, deux préconcepts sont retenus et font l'objet de débats entre les interlocuteurs. D'un côté, un préconcept de type interface informatique qui regrouperait les éléments du module de confort accessibles depuis des menus, des raccourcis et des zones tactiles. D'un autre côté, un préconcept plus « traditionnel » composé de boutons et de molettes. Cependant, en dehors des caractéristiques de chacun de ces préconcepts, il nous semble intéressant de relever que ces solutions sont discutées non pas du point de vue technique, comme nous l'avons relevé dans la classe 1, mais du point de vue de leur

Tableau 4
Les 10 premières occurrences et variables hautement significatives ($p < .001$)
des classes identifiées par ALCESTE

	Classe 1		Classe 2		Classe 3	
	<i>Occurrences</i>	χ^2	<i>Occurrences</i>	χ^2	<i>Occurrences</i>	χ^2
Noms com- muns	concept+	78.63	mode+	91.13	voiture+	169.6
	preconcept+	70.36	ecran+	81.17	climatisation	4
	idee+	47.84	informat+ion	63.70	temperature+	92.35
	reunion+	26.87	touche+	55.33	air+	72.14
	interface+	25.51	bouton+	49.56	auto+	57.56
	clavier+	23.48	menu+	39.68	systeme+	50.03
	commande+	17.58	fonction+	36.14	bagnole+	49.92
			raccourci+	35.16	pied+	44.43
			zone+	29.81	conducteur+	33.27
			molette	29.71	envi+e	32.34
					27.70	
Verbes	<i>Occurrences</i>	χ^2	<i>Occurrences</i>	χ^2	<i>Occurrences</i>	χ^2
	essa+yer	30.98	regl+er	36.81	regul+	66.86
	faire.	24.87	valid+	26.36	capt+er	38.85
	pens+er	20.63	appu+yer	25.98	comprendre.	38.36
	voir.	18.32	selectionne+	22.96	habitu<	33.27
		escamot+er	19.71	ventil+er	33.20	
Adjectifs	<i>Occurrences</i>	χ^2	<i>Occurrences</i>	χ^2	<i>Occurrences</i>	χ^2
	deuxieme+	25.60	principa+l	48.17	chaud+	50.03
	premier+	17.58	selecti+f	19.67	particulier+	38.85
	voca+l	17.26			automat+16	32.45
					long+	27.70
				neu+f	27.70	
				froid+	22.14	
V.I.	<i>Mots étoilés</i>	χ^2	<i>Mots étoilés</i>	χ^2	<i>Mots étoilés</i>	χ^2
	*Ingenieur	64.46	*Ergonome	28.30	-	-

utilisabilité, qui correspond « à la capacité des dispositifs techniques à être utilisés facilement, par une personne donnée, de façon à accomplir la tâche pour laquelle cet objet a été conçu » (Brangier & Barcenilla, 2003, p. 23).

Pour appréhender la définition des préconcepts du point de vue de leur utilisabilité, l'ergonome (représentatif de la classe 2 : $\chi^2 = 28.30$; $p < .001$) va introduire une réflexion sur l'*activité future probable* des conducteurs, qui « vise à rendre compte d'une *délimitation* progressive des formes possibles de l'activité future par les décisions prises au cours du projet » (Daniellou, 1988, p. 188). Les occurrences verbales décrivent ainsi les actions futures du conducteur : régler, valider, appuyer, sélectionner et escamoter. Ces actions sont appliquées sur les hypothétiques commandes du module de confort : touche, bouton ou molette. Ici, l'ergonome décrit précisément l'activité du conducteur en termes d'activités gestuelles, afin de guider le concepteur dans ses choix techniques. Il s'agit alors de simuler, de façon théorique, les situations d'activité fictives et de placer les acteurs de la conception dans des situations réalistes, proches de celles que rencontreront les conducteurs.

Analyse et interprétation de la classe 3

Pour finir, l'univers lexical de la classe 3 fait majoritairement référence au *contexte de l'activité de conduite automobile et aux éléments généraux qui composent le module de confort*. En somme, les dialogues échangés au cours de la réunion semblent avoir pour objectif de définir précisément la problématique du projet. Les éléments du module sont alors énumérés, comme pour délimiter l'*espace problème*. Il semble ainsi que les acteurs de la conception se mobilisent pour définir préalablement la tâche, et poser, de ce fait, les termes du problème :

uce n°690 ($\chi^2 = 26$; $p < .001$) : Bon alors, la climatisation dans une voiture, qu'est-ce qu'elle comprend? Alors, elle comprend la température plus la ventilation...

Néanmoins, l'*espace problème* n'est pas défini en termes de contraintes techniques, mais en termes de « contraintes humaines ». Ici, en accord avec Sagot et Gomes (2003), c'est « une délimitation progressive des formes possibles de l'activité future » (p. 63) qui est abordée, en vue de « fournir, à l'ensemble des acteurs-métiers, une évaluation plus large concernant les conséquences de leur choix de conception » (Sagot & Gomes, p. 63). Il s'agit alors d'élaborer une *représentation commune imaginaire* de l'utilisateur en situation de conduite, centrée sur l'activité et l'analyse des besoins de l'individu.

uce n°382 ($\chi^2 = 35$; $p < .001$) : Mais chaque conducteur a des besoins particuliers. Toi quand tu rentres dans un véhicule, ou S,

ou G, il y en a qui aime avoir une température à 19 degrés ou 26 degrés. Par contre, il y en a qui aime avoir beaucoup d'air dans la face.

Discussion générale

Intérêt de chaque méthode pour l'analyse des communications fonctionnelles

La méthode COMET a permis de fournir un ensemble de séquences verbales argumentatives, décrivant les phases principales d'identification du problème. Ainsi, en accord avec Darses (2006), il a été possible de montrer que, durant le travail de recherche de solution, la formulation de critères est le mécanisme préférentiellement choisi pour soutenir l'argumentation. Les acteurs de la coopération formulent alors une grande diversité de « raisons » pour justifier les choix retenus pour la résolution du problème de conception. L'application de la méthode COMET a également permis de dégager la nature des échanges entre les concepteurs, en précisant le rôle de la contribution de chaque métier dans le déroulement du projet. En somme, la différenciation des métiers s'est faite par le mode *assertion/requête* extrait du modèle proposé par COMET. Pour finir, cette méthode s'inscrit pleinement dans le courant épistémologique formulé par Simon (1969) pour une science de la conception, où le paradigme du traitement de l'information est appliqué à la résolution de problèmes de conception.

La grille de codage pour l'analyse de la confrontation de points de vue a quant à elle permis de dégager une arborescence des points de vue de chaque acteur. Cette arborescence a révélé un panorama des interactions verbales au cours des réunions, en mettant en évidence les propositions de conception qui n'ont pas fait l'objet d'une validation et celles qui ont été largement débattues. Grâce à l'utilisation de cette grille de codage, le chercheur peut dégager clairement les enjeux de la dynamique décisionnelle. De plus, la grille de codage a permis de relever la contribution de chaque métier à la définition des préconcepts. Aussi avons-nous observé que les ergonomes argumentaient les propositions formulées par les ingénieurs. Enfin, cette méthode pour l'analyse des confrontations de point de vue favorise la description des processus coopératifs, notamment en ce qui concerne la construction d'un référentiel opératif commun.

Pour finir, la méthode ALCESTE a proposé un découpage en 3 classes sémantiques qui ont permis d'illustrer la construction progressive d'une représentation partagée entre les acteurs de la conception. Cette représentation est une définition de l'espace problème élaborée à partir de l'activité future souhaitable des conducteurs qui seront les principaux utilisateurs du module de confort. Ainsi, cette dernière méthode a mis en évidence la nécessité, pour les

acteurs du projet, de co-construire une représentation partagée du problème qui leur permette de coopérer. ALCESTE a enfin permis de définir le contenu et l'opposition des communications par métier.

Inconvénients de chaque méthode pour l'analyse qualitative

Toutefois, les méthodes que nous avons utilisées par l'analyse des communications fonctionnelles nous ont posé certains problèmes tant du point de vue méthodologique que de l'interprétation des résultats qui en découlaient.

Ainsi, l'utilisation de la méthode COMET a tout d'abord posé quelques difficultés de codage. En accord avec les remarques formulées par Traverso et Visser (2003), nous pensons que l'analyste peut rencontrer des difficultés de choix entre différents codes. Il est en effet parfois difficile d'interpréter les actes verbaux et de les différencier en tant qu'actes d'interprétation, d'information, d'évaluation ou de génération d'idée. De plus, la méthode COMET ne permet pas d'établir clairement de lien entre des actions. Traverso et Visser (2003) soulignent ainsi qu'au lieu de lister uniquement deux actions consécutives, « on aimerait parfois pouvoir expliciter leur lien » (p. 244). Par ailleurs, la technique du codage de chaque tour de parole ne permet pas d'être appliqué à l'ensemble du texte, ce qui implique que certains extraits ne puissent pas être analysés alors qu'ils pourraient être la source de données précieuses pour le chercheur.

La grille de codage pour l'analyse des confrontations de point de vue se limite quant à elle à un codage parfois trop simplifié des communications. En conséquence, cette analyse perd une partie de la finesse des interactions verbales. De plus, tout comme l'application de la méthode COMET, le codage pour chaque tour de parole ne peut pas être appliqué à l'ensemble d'un corpus textuel issu de la retranscription intégrale d'une réunion de projet. Il nécessite que le chercheur se centre sur un échantillon de la réunion, perdant ainsi une partie de la source d'informations.

Enfin, le logiciel ALCESTE est souvent critiqué pour le manque de visibilité quant à la méthodologie d'analyse lexicale sur lequel il repose. Toutefois, Reinert (2001) précise que la méthode ALCESTE n'a pas pour objectif « d'analyser la signification linguistique d'un corpus, (mais de) représenter les fonds topiques d'un discours à l'aide des mondes lexicaux » (p. 43). Reinert précise ainsi qu'il « serait contradictoire [...] de penser qu'un calcul du sens pourrait être automatiquement effectué par un programme. Un calcul peut aboutir cependant à construire une représentation, une sorte de carte pour s'orienter » (p. 43). Il est donc important de souligner que la méthode ALCESTE n'est pas une méthode interprétative comme celles proposées par COMET ou la grille de codage des confrontations de point de vue, mais une

méthode de structuration et de représentation des informations présentes au sein d'un corpus textuel. Cela nécessite alors un travail d'interprétation ultérieur de la part du chercheur. Pour finir, on relèvera que la méthode ALCESTE n'est pas spécifiquement dédiée à l'analyse des communications en conception de produit. Cela qui lui permet d'élargir les éléments pris en compte dans l'analyse textuelle (notamment à travers la prise en compte d'autres communications que celles strictement *fonctionnelles*), tout en réduisant, *a contrario*, l'analyse plus détaillée qui pourrait être faite sur les communications qui servent uniquement la conception de produit.

Conclusion

La comparaison des méthodes destinées à l'analyse des communications fonctionnelles en situation de travail collectif devait nous permettre de faire un choix méthodologique pour la conduite d'une recherche sur le travail coopératif assisté par ordinateur dans le cadre de la conception de produit (Gronier, 2006). Après avoir appliqué et confronté certaines des méthodes présentées dans cet article (Gronier, Sagot, Gouin & Gomes, 2001), nous avons fixé notre choix sur l'utilisation du logiciel ALCESTE pour plusieurs raisons. Tout d'abord, nous avons souhaité recueillir et analyser des réunions de projet de manière exhaustive, sans devoir nous limiter à un extrait du corpus textuel. Nous avons ainsi retranscrit plus de 21 heures de réunions, puis différencié les tours de parole en fonction de modalités de variable sur lesquelles nous souhaitions travailler. Seule une analyse par un logiciel d'analyse qualitative comme ALCESTE pouvait nous permettre de conserver un corpus aussi conséquent.

Ensuite, nous avons préféré laisser au chercheur le maximum d'interprétation possible vis-à-vis du sens à donner au corpus textuel. Or, la méthode COMET et la grille de codage des confrontations de points de vue nous ont semblé trop directives. En effet, la construction de sens repose sur une interprétation préalable, liée au codage des tours de parole. Ainsi, l'interprétation du corpus que réalise le chercheur est-elle déjà orientée par le codage qu'il a appliqué en amont de son analyse. Cet aspect nous a semblé trop limitatif vis-à-vis des objectifs de recherche que nous nous étions fixés.

Pour finir, nous avons souhaité contribuer au développement de l'utilisation d'ALCESTE en psychologie ergonomique, largement initié par Navarro (voir notamment Navarro, 1990; Navarro & Marchand, 1994). ALCESTE a en effet démontré qu'il pouvait être considéré comme une méthode sérieuse d'analyse textuelle, au même titre que celles couramment employées en sciences humaines et sociales (Bardin, 2003).

Références

- Bardin, L. (2003). *L'analyse de contenu*. Paris : PUF.
- Brangier, E., & Barcenilla, J. (2003). *Concevoir un produit facile à utiliser*. Paris : Éditions d'Organisation.
- Daniellou, F. (1988). Ergonomie et démarche de conception dans les industries de processus continus. Quelques étapes clés. *Le travail humain*, 51(2), 185-194.
- Daniellou, F. (1994, Septembre). L'ergonome et les acteurs de la conception. *Ergonomie et ingénierie : XXIXe Congrès de la Société d'ergonomie de langue française* (pp. 27-32). Paris.
- Darses, F. (1997). L'ingénierie concourante : un modèle en meilleure adéquation avec le processus cognitifs de conception. Dans P. Bossard, C. Chanchevri, & P. Leclair (Éds.), *Ingénierie concourante. De la technique au social* (pp. 39-55). Paris : Economica.
- Darses, F. (2002). Éditorial : activités coopératives de conception. *Le travail humain*, 65(4), 289-292.
- Darses, F. (2006). Analyse du processus d'argumentation dans une situation de reconception collective d'outillages. *Le travail humain*, 69(4), 317-347.
- Darses, F., Détienne, F., Falzon, P., & Visser, W. (2001). *A method for analysing collective design processes*. Rapport de recherche de l'INRIA, No. 4258.
- Détienne, F., Martin, G., & Lavigne, E. (2005). Viewpoints in co-design : a field study in concurrent engineering. *Design Studies*, 26(3), 215-241.
- Falzon, P. (1994). Dialogues fonctionnels et activité collective. *Le travail humain*, 57(4), 299-312.
- Gronier, G. (2006). *Psychologie ergonomique du travail collectif assisté par ordinateur : l'utilisation du collecticiel dans les projets de conception de produits*. Thèse de doctorat inédite, Université de Franche-Comté, Besançon.
- Gronier, G., Sagot, J.C., Gouin, V., & Gomes, S. (2001, Juin). Étude exploratoire des activités coopératives de conception assistées par un collecticiel. *10^e Atelier du travail humain « Modélisation des activités coopératives de conception »* (pp. 93-109), Paris.
- Grusenmeyer, C., & Trognon, A. (1997). Les mécanismes coopératifs en jeu dans les communications de travail : un cadre méthodologique. *Le travail humain*, 60(1), 5-31.

- Karsenty, L. (2000). Cooperative work : the role of explanation in creating a shared problem representation. *Le travail humain*, 63(4), 289-309.
- Martin, G. (2001). *Intégration et confrontation des points de vue dans le cadre de la conception en ingénierie concurrente*. Thèse de doctorat inédite, Conservatoire national des arts et métiers, Paris.
- Martin, G., Détienne, F., & Lavigne, E., (2002). Confrontation of viewpoints in a concurrent engineering process. Dans P. Chedmail, G. Cognet, C. Fortin, C. Mascle, & J. Pegna (Éds.), *Integrating design and manufacturing in mechanical engineering* (pp. 3-10). London : Kluwer Academic Publishers.
- Moles, A. (1986). *Théorie structurale de la communication et société*. Paris : Masson.
- Navarro, C. (1990). Functional communication and problem solving in a bus traffic-regulation task. *Psychological Reports*, 67, 403-409.
- Navarro, C. (1993). L'étude des activités collectives de travail : aspects fondamentaux et méthodologiques. Dans F. Six, & X. Vaxevanoglou (Éds.), *Les aspects collectifs du travail. Actes du XXVIIe Congrès de la Société d'ergonomie de langue française* (pp. 91-106). Toulouse : Octarès.
- Navarro, C., & Marchand, P. (1994). Analyse de l'échange verbal en situation de dialogue fonctionnel : étude de cas. *Le travail humain*, 57(4), 313-330.
- Reinert, M. (1990). A.L.C.E.S.TE. : une méthodologie d'analyse des données textuelles et une application : Aurélia de Gérard de Nerval. *Bulletin de méthodologie sociologique*, 26, 24-54.
- Reinert, M. (2001). Alceste, une méthode statistique et sémiotique d'analyse de discours; application aux « Rêveries du promeneur solitaire ». *La Revue française de psychiatrie et de psychologie médicale*, 49(5), 32-46.
- Sagot, J.C., & Gomes, S. (2003). Intégration des facteurs humains dans la démarche de conception. Une approche ergonomique. *Cahiers de notes documentaires – Hygiène et sécurité du travail*, 191, 61-71.
- Savoyant, A., & Leplat, J. (1983). Statut et fonction des communications dans l'activité des équipes de travail. *Psychologie française*, 28(3), 247-253.
- Simon, H. (1969). *The sciences of the artificial*. Cambridge, MA. : MIT Press.
- Terressac, G. de, & Chabaud, C. (1990). Référentiel opératif commun et fiabilité. Dans J. Leplat, & G. de Terressac (Éds.), *Les facteurs humains de la fiabilité* (pp. 110-139). Marseille : Octarès/Entreprises.

Traverso, V., & Visser, W. (2003, Octobre). Confrontation de deux méthodologies d'analyse de situations d'élaboration collective de solution. *Deuxièmes journées d'étude en psychologie ergonomique, EPIQUE*, (pp. 241-246), Paris.

Zarifian, P. (1998). *Travail et communication*. Paris : Presses universitaires de France.

Guillaume Gronier est ingénieur de Recherche et Développement au sein du département Service Science and Innovation du Centre de recherche public Henri Tudor (Luxembourg), et membre du comité de pilotage de l'équipe Innovation Management : Organisation and People. Il est titulaire d'un doctorat de psychologie ergonomique de l'Université de Franche-Comté (France), durant lequel il s'est intéressé aux nouvelles formes de travail collectif assisté par ordinateur dans le cadre de la conception de produits. Ses travaux sur le travail collectif l'ont conduit à définir une grille d'analyse des activités collaboratives, qui a pour objectif d'accompagner les entreprises dans l'amélioration de leur performance organisationnelle. Ses thématiques de recherche actuelles concernent également la confiance numérique, la perception de la qualité des sites e-gouvernementaux et l'ergonomie des logiciels. Membre de la Usability Professionals' Association (UPA), il est l'un des fondateurs du chapitre local France-Luxembourg UPA (FLUPA), qui regroupe aujourd'hui plus de 80 membres.

L'objet d'analyse comme pivot de l'analyse qualitative assistée par ordinateur

Jacques Hamel, Ph.D.

Université de Montréal

Résumé

Sous ce titre, on envisagera l'objet de l'analyse afin de savoir en quoi et comment sa détermination — théorique, méthodologique et technique — fait office de vecteur de l'analyse basée sur l'exploitation de logiciel(s) d'analyse qualitative. Après avoir brièvement cerné ce à quoi correspond l'objet d'analyse dans la démarche propice à la formulation de la connaissance explicative, en sociologie par exemple, l'auteur cherchera à montrer dans la foulée qu'à ce stade, un logiciel d'analyse comme ATLAS.ti offre les moyens requis pour combiner la connaissance pratique de la vie sociale à la connaissance théorique destinée à en rendre raison.

En bref, on cherchera à montrer que la « programmation » du logiciel d'analyse sous forme d'opérations apparemment techniques traduit en réalité l'interprétation que l'analyste orchestre « dans son esprit » de la connaissance qu'affichent les artisans de la vie sociale afin de la transposer sur le plan théorique propice à son explication. De ce fait, « comprendre et expliquer ne font qu'un », selon la formule que Bourdieu propose pour fonder l'objectivation en sociologie et, plus largement, dans les sciences sociales. Dans cette voie, l'élaboration de l'objet de l'analyse assistée par ordinateur sera d'abord considérée par analogie avec l'objectivation participante chère à cet auteur puis, sur l'élan, sous l'optique des règles et des procédés proprement méthodologiques qu'imposent les logiciels susceptibles de donner corps à l'analyse qualitative. Sur cette base, on pourra constater que l'objet d'analyse se révèle être le pivot de toute cette entreprise que sous-tend la mise en œuvre de l'un ou l'autre des logiciels utilisés en recherche qualitative.

Mots clés

OBJET D'ANALYSE, DÉCRIRE/COMPRENDRE/EXPLIQUER, LOGICIELS D'ANALYSE QUALITATIVE

Introduction

Dans la veine de la *grounded theory*, on a vu naître et se développer des logiciels d'analyse qualitative susceptibles de donner l'éclat de la rigueur à la méthodologie du même nom, la méthodologie qualitative. C'est le cas du

logiciel ATLAS.ti, mis au point dans l'intention d'apporter à l'analyse des données qualitatives la touche de précision et de standardisation qui semblait apparemment lui faire défaut. Force est d'admettre qu'il suffisait à une certaine époque de se réclamer d'Howard Becker ou d'invoquer le fameux ouvrage de Glaser et Strauss (1967), *The Discovery of Grounded Theory*, pour prouver que l'analyse avait été produite — pour ne pas dire générée — directement à partir des données et qu'ainsi elle en possédait les qualités immédiates, la validité et l'éloquence. En d'autres termes, l'analyse et les résultats acquis dans la foulée se conformant à la connaissance pratique que recèlent d'emblée les « données » et, de ce fait, ne lui portant pas outrage, la *théorisation ancrée* se révélait forcément exacte sans que soient clairement et judicieusement précisées les règles et les opérations propres à lui donner corps.

En ma qualité d'évaluateur de thèses, il m'a été donné de lire des chapitres dits méthodologiques de moins de cinq pages, du fait que les candidats ne voyaient nullement le besoin de définir et d'explicitier les opérations méthodologiques sous-jacentes à leur analyse, puisqu'elles relevaient de cette entreprise qui produit la théorie à la lumière des données directement recueillies sur le terrain.

Le logiciel ATLAS.ti — dont la conception relève directement de la *grounded theory* (voir Strauss & Corbin, 2004) — vient toutefois rappeler que l'analyse conçue dans cette voie obéit forcément à des opérations et à des règles précises qu'impose son exploitation. Sans qu'on y ait porté suffisamment attention, les logiciels propres à l'analyse qualitative, à l'instar d'ATLAS.ti, révèlent l'importance cruciale de l'objet d'analyse, lequel correspond ici aux données qui font l'objet de l'analyse conduite en vertu d'opérations assistées par ordinateur. Voilà le thème abordé dans le cadre de cet article.

L'analyse en tant que chiasme épistémologique

L'analyse qualitative — comme du reste toute analyse en sociologie ou dans les sciences sociales — renvoie résolument à l'*objet de l'analyse* et celui-ci représente à bien des égards son point d'orgue, comme on le verra plus loin. Sans vouloir s'engager dans une vaste réflexion sur la sociologie et sur ce à quoi correspond l'entreprise qu'elle sous-tend d'office, on doit s'obliger à penser, pour reprendre la formule imagée de Pierre Bourdieu, qu'elle se révèle tout compte fait une « connaissance d'une connaissance » (Bourdieu, 1992, p. 103), à l'image de l'anthropologie que Clifford Geertz (1998, p. 80) conçoit pour sa part comme l'« explication d'explications ».

En effet, sans qu'on ait suffisamment noté le renversement de sa position sur le sujet, Bourdieu en est venu à affirmer que le sens commun recelait une

connaissance de la vie sociale par contraste aux « lieux communs », aux « évidences » et au « gros bon sens » auxquels il l'associait inmanquablement dans *Le métier de sociologue*, selon le « principe souverain d'une distinction sans équivoque entre le vrai et le faux » (Bourdieu, Passeron & Chamboredon, 1968, p. 47) pour démarquer la connaissance sociologique du sens commun.

La vie sociale relève dans le sens commun de l'évidence et du « c'est comme ça » du fait que cette connaissance est directement issue de la pratique, individuelle ou collective et, par conséquent, se forme sous le coup d'événements, de circonstances ou de faits que Bourdieu amalgame à des « réalités substantielles ». Le sens commun s'associe chez lui à la violence symbolique en raison de la nature routinière et répétitive de cette connaissance et de ce qu'elle prend pour objet, la « pratique sociale ». De ce fait, trop familière, cette dernière n'est jamais mise en question, ni évidemment expliquée sous l'optique sociologique.

Le sens commun ne se révèle donc pas forcément faux, mais se conçoit avec nuances comme connaissance *pratique*, c'est-à-dire en phase avec la « pratique sociale » que forment les individus à leur échelle.

La connaissance sociologique, quant à elle, s'élabore par contraste au moyen de notions, de concepts et de méthodes, dans le but d'expliquer en termes de « relations objectives dans lesquelles s'insèrent les individus », de relations « indépendantes de la volonté et de la conscience individuelle » (Bourdieu, 1992, p. 72). En d'autres mots, la sociologie s'emploie selon Bourdieu à élaborer une « topologie sociale », une espèce de géométrie fondée sur les positions sociales établies à la lumière de la mobilisation des ressources et des pouvoirs à la disposition des individus, que Bourdieu associe aux notions de capital et d'habitus. En effet, sous cette perspective, les individus jouent à leur échelle des différentes espèces de capital dont ils sont dotés ou qu'ils acquièrent selon leur habitus, lequel correspond à des « schémas mentaux et corporels de perception, d'appréciation et d'action » (Bourdieu, 1992 p. 24). La mobilisation de leurs pouvoirs et ressources sur cette base, propre à former la grammaire de leurs actions, détermine d'emblée leur position mutuelle dans l'espace des relations objectives qu'est le *champ*.

Sur l'élan,

les agents sociaux, et aussi les choses en tant qu'elles sont appropriées par eux, donc constituées comme propriétés, sont situés en un lieu de l'espace social, lieu distinct et distinctif qui peut être caractérisé par la position relative qu'il occupe par rapport à d'autres lieux (au-dessus, au-dessous, entre, etc.) et par

la distance (dite parfois « respectueuse » : *e longinquo reverentia*) qui le sépare d'eux (Bourdieu, 1997, p. 161).

Sous cette optique théorique, l'analyse ainsi orchestrée par la sociologie et la théorie à laquelle Bourdieu donne son nom oblige donc à circonscrire ce qu'est l'objet de la discipline aux yeux de cet auteur : en l'occurrence les « relations objectives dans lesquelles s'insèrent les individus ». Il s'agit plus exactement de les saisir telles qu'elles apparaissent et se conçoivent sous la forme de « réalités substantielles » à l'œuvre dans la connaissance pratique qu'affichent les individus à leur échelle et à leur niveau, celui de la « pratique sociale ». La sociologie se révèle ainsi véritablement une « connaissance d'une connaissance ».

L'analyse implique par conséquent de débusquer dans les « réalités substantielles » les « informations » susceptibles de porter au jour l'habitus et le capital grâce auxquels s'édifie la topologie sociale propice à l'explication sociologique. Bref, l'analyse consiste en ce sens à repérer les « informations », à les découper et à en extraire les éléments qui, dans l'optique de la théorie, donnent corps par exemple au capital conçu sur ce plan comme un « ensemble de ressources et de pouvoirs » de différentes espèces.

L'analyse sociologique, on le constate, s'opère en définitive en vertu d'une sorte de chiasme épistémologique que Bourdieu désigne pertinemment comme « connaissance d'une connaissance ». Or, ce chiasme épistémologique que sous-tend d'emblée l'analyse a pour vecteur l'objet d'analyse qui, à bien des égards, reste le point aveugle de l'ouvrage que constitue la sociologie.

L'objet d'analyse, le point d'orgue de l'analyse

Car, en effet, l'objet d'analyse — c'est-à-dire les « données » ou les informations qui deviennent l'objet de l'analyse ou, plus exactement, ce sur quoi se base l'analyse conduite sous l'optique sociologique — correspond à la « matière » que l'analyste associe dans son esprit aux notions qui donnent corps à la connaissance sociologique fondée sur l'habitus et le capital, par exemple. Bref, conçu de cette manière, l'objet d'analyse a trait aux « réalités substantielles » qui sont l'objet de la *connaissance pratique* et sur lesquelles l'analyste met l'accent pour les envisager sur le plan de la connaissance sociologique. Celle-ci, dans son principe, se formule sous le mode de la théorie susceptible d'expliquer, au sens que donne à ce terme l'épistémologie contemporaine : chercher à créer une « représentation d'une autre nature que celle de l'objet à connaître afin d'avoir sur lui un contact précis et pénétrant » (Granger, 1986, p. 120).

Non seulement l'objet d'analyse pointe-t-il les « réalités substantielles » qui, sur le plan de la connaissance pratique, correspondent à la « matière »

relative aux notions formulées dans le cadre de la connaissance théorique, sociologique en l'occurrence, mais il laisse également transparaître l'interprétation qui en est immédiatement donnée et qui sur l'élan détermine la combinaison des concepts grâce à laquelle est produite l'explication.

L'objet d'analyse ou, en d'autres mots, l'objet que cible l'analyse témoigne du fait qu'en sociologie « comprendre et expliquer » ne font qu'un, selon une autre formule éloquent de Bourdieu (1993, p. 910). Il représente donc à l'œuvre le chiasme épistémologique qui s'opère dans la tête de l'analyste, ici le sociologue (voir Figure 1). En effet, il correspond à l'objet recherché tel que formulé dans les termes de la connaissance pratique et repéré en vertu de la connaissance théorique qui gouverne l'entreprise du sociologue.

L'objet d'analyse conçu sous ce chef restait à bien des égards le point aveugle de l'analyse qualitative axée, par exemple, sur la *grounded theory*. On avait en effet peine à connaître exactement les opérations et les règles qui président à la « connaissance de la connaissance » en vertu de laquelle s'opère l'analyse, d'une part, et la formulation de l'explication à laquelle la théorie donne corps, d'autre part. Le talon d'Achille de l'analyse qualitative résidait dans le manque flagrant de précisions sur les rouages à l'œuvre dans l'esprit de l'analyste quand il décrit et interprète les « données » et, simultanément, sur la base de cette interprétation, quand il les associe à la théorie afin d'en élaborer l'explication. La difficulté, notée par Bourdieu, est que dans le feu de l'analyse « comprendre et expliquer ne font qu'un ».

Les logiciels d'analyse qualitative remédient aux ratés de l'analyse en obligeant l'analyste à « mettre cartes sur table ». En effet, les rouages techniques qui, sous le mode informatique, commandent leur exploitation obligent à déterminer exactement les opérations et les règles qui donnent ici corps à l'analyse.

Sans faire la démonstration technique d'ATLAS.ti, disons simplement qu'il permet d'observer, voire de déterminer exactement toutes les opérations mentales de l'analyste quand il s'évertue à décrire et à interpréter les données qui sont l'objet d'analyse et à les intégrer au cadre de la connaissance sociologique qui, on l'a noté, s'établit sur le registre de la pensée formelle que représente la théorie.

L'analyse se déroule au gré de la segmentation, de la focalisation et de la construction des données, que l'analyste opère sur le vif, mais qu'il se voit obligé de coordonner à des catégories associées à des cadres ou des « fenêtres », ce qui nécessite l'apport systématique de « renseignements » ou de

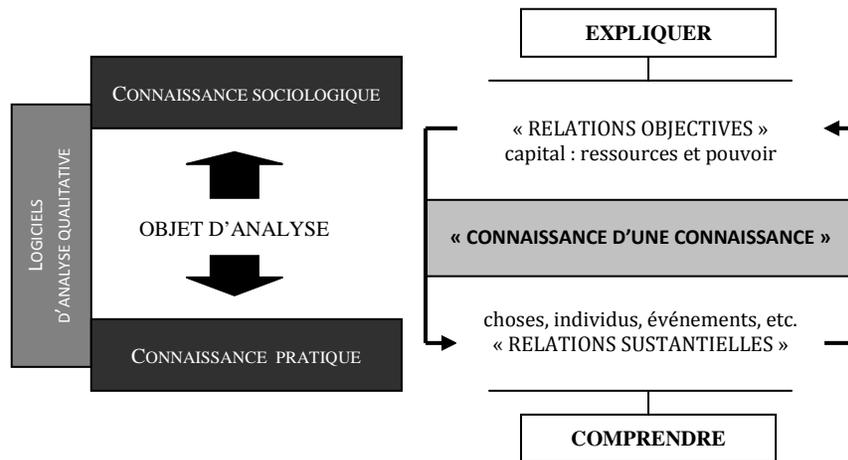


Figure 1. Objet d'analyse et chiasme épistémologique

définitions qu'il se doit de précisément élaborer. Sur cette base, l'analyse en vient à être gouvernée en vertu de règles et d'opérations relevant de la « manipulation formelle », que l'épistémologie contemporaine conçoit comme le prélude à la connaissance explicative ou à la théorie toujours fondée sur une représentation caractérisée par une « image d'une autre nature que celle de l'objet à connaître » (Granger, 1988, p. 144; voir également 1993). En d'autres termes, l'exploitation d'un tel logiciel permet de « retrouver dans une intuition — en dernier ressort sensible — un terme abstrait exprimé dans un énoncé » qui, aux yeux de Granger (1992, p. 179), se révèle être le fer de lance de toute entreprise se réclamant de la connaissance explicative associée à la théorie, et qui donne tout son éclat à la science.

Ainsi, les intuitions sur lesquelles s'érige la catégorisation, par exemple, doivent être énoncées dans les fenêtres qui s'ouvrent automatiquement pour y inscrire des mémos. Les formules inscrites, certes rectifiables, maintiennent constantes les définitions attachées à chacune des catégories, lesquelles sont donc ouvertes à d'autres que l'analyste qui les a produites.

Dans cette voie, il devient possible d'*indexer* — le terme importe — les données à des catégories de plus en plus abstraites, qui dès lors les transposent sous la forme de la représentation propre à la connaissance explicative que formule la théorie. En effet, les données sont alors manipulées, au bon sens du terme, en vertu d'opérations précises qui les coordonnent aux catégories dorénavant répertoriées dans l'index du logiciel d'analyse. L'affaire correspond à ce que Jean-Claude Passeron (1993, p. 8) nomme la « description définie ». Par « définie », il entend la qualité qu'a la description d'indexer des données,

toujours associées à des configurations historiques, à des notions abstraites qui les préparent à une manipulation formelle au sein de la représentation associée à la connaissance proprement théorique.

Les logiciels d'analyse qualitative viendraient ainsi remédier selon Passeron aux ratés de la sociologie, ces descriptions basées sur des « actes d'interprétation » que leurs propres auteurs sont incapables de formuler en une « méthode de discursion » (Passeron 1996, p. 94), c'est-à-dire une interprétation élaborée selon un « discours » précisément réglé et voué à l'argumentation. Les logiciels d'analyse assistée par ordinateur fournissent les moyens de « stabiliser, fût-ce provisoirement, la théorie » qui chez cet auteur donne corps à une langue protocolaire de description et d'interprétation (Passeron, 1991, p. 59) capable de régler la connaissance issue d'une entreprise comme la sociologie.

Les logiciels d'analyse qualitative comme vecteur d'objectivation participante

L'objet d'analyse, on le constate, est au cœur de l'analyse assistée par ordinateur et, à bien des égards, fait office de vecteur des opérations conduites au moyen de logiciels comme ATLAS.ti. En effet, les données ciblées qui sont objet de la manipulation formelle orchestrée par le logiciel renferment dans l'esprit de l'analyste les informations de sens commun susceptibles de révéler les éléments du social qui, représentés sous la forme de catégories et de notions, donnent lieu à l'analyse capable de produire la connaissance explicative attendue de la sociologie.

Sans qu'on lui voue fidélité outre mesure, la théorie des champs de Bourdieu permet ici d'illustrer le propos. L'objet de la sociologie correspond chez cet auteur aux « relations objectives » dans lesquelles les individus s'insèrent pour évoluer en société. La théorie de Bourdieu les conçoit à la lumière du *capital*, c'est-à-dire l'ensemble des ressources et des pouvoirs de différentes espèces, que mobilisent les individus selon leur *habitus*, conçu comme un ensemble de schémas mentaux et corporels de perception, d'appréciation et d'action. La combinaison du capital et de l'*habitus* détermine en théorie la position de l'individu dans l'espace qu'est le *champ*, susceptible de la représenter sous la forme d'une géométrie sociale qui prend corps au gré de points unis par des « relations objectives ».

La connaissance sociologique a donc pour enjeu de puiser les éléments à même de traduire le capital et l'*habitus*, en décrivant et en interprétant les données formulées sous le mode de la connaissance pratique des individus en termes de réalités substantielles : individus, contextes, événements et autres éléments tangibles. Si par exemple en entrevue un enquêté confie avoir

bénéficié d'une « entrée » ou d'une « connexion » pour l'obtention d'un emploi, l'analyste doit d'abord s'employer à décrire cette expression à la lumière des propos tenus afin de savoir exactement ce qu'elle désigne, par exemple un avantage ou un pistonnage, ce qui pourra alors être interprété comme ressource ou pouvoir utile pour décrocher l'emploi en question. Ces ressources et pouvoirs peuvent être conçus en théorie comme capital, voire comme capital politique ou relationnel, qui dans l'axe théorique de Bourdieu correspond à l'influence propre à expliquer la position sociale acquise grâce à l'emploi décroché.

L'entreprise donne forme à l'objectivation des données, selon l'expression consacrée, et précède la manipulation formelle en vertu de laquelle se forme la connaissance explicative orchestrée sur la base de la représentation que gouvernent les catégories et notions mises en jeu dans l'analyse assistée par ordinateur. Sous ce chef, elle correspond à bien des égards à ce que Bourdieu, encore lui, nomme « l'objectivation participante », si on a toutefois soin de nuancer la définition qu'il en donne.

L'expression, « objectivation participante », traduit éloquentement ce que représente l'analyse assistée par ordinateur. Les logiciels d'analyse qualitative constituent sans conteste un moyen propre à objectiver les données et à les élaborer dans l'optique de la connaissance théorique, sociologique en l'occurrence. Même si leur exploitation dépend des contraintes techniques sous-jacentes à leur mise en œuvre, il n'en demeure pas moins qu'elle requiert inmanquablement la « participation » de l'analyste qui mobilise cet outil.

Sans adopter le lieu commun voulant que l'usage des moyens informatiques soit d'emblée ludique, il s'avère que l'exploitation des logiciels d'analyse qualitative se conforme bien à un jeu, celui que sous-tend l'élaboration de la connaissance théorique, conçu par l'épistémologie contemporaine comme un « travail qui, indépendamment de toutes les connotations et de toutes les conséquences que lui confère son insertion obligée dans une réalité sociale », correspond à un « effort de rationalisation à partir de principes établis » (Granger, 1986, p. 115).

Dans cette voie, l'exploitation des logiciels d'analyse assistée par ordinateur se rattache à l'objectivation participante chère à Bourdieu. Chez cet auteur, l'exercice cherche à objectiver l'analyste en retournant de manière réflexive la théorie issue de la « géométrie sociale » sur la personnalité du sociologue lui-même quand il se propose d'objectiver le monde social qui constitue l'objet de son analyse. Il peut ainsi prendre en compte ses dispositions théoriques et méthodologiques élargies aux déterminations sociales qui s'exercent sur sa pratique en vertu d'« adhésions » et

d'« adhérences » à la société qu'il étudie (Bourdieu, 2003, p. 45), au métier qu'il exerce avec ses problématiques et notions « obligées » du fait qu'elles apparaissent comme « allant de soi » sous le mode de l'évidence, du « naturel » que Bourdieu associe irrémédiablement à la violence symbolique.

Bourdieu nomme cet exercice « objectivation participante » à juste titre puisque l'objectivation requiert ici la participation du chercheur désireux de prendre conscience des catégorisations auxquelles est sujette sa pensée, pensée qu'il veut néanmoins mobiliser pour objectiver le « monde social » en vertu de la géométrie sociale capable de le concevoir comme une configuration de « relations objectives ».

Cette participation est toutefois rigoureusement bornée par la théorie composée notamment des notions d'habitus et de capital propres à déterminer la position sociale du chercheur, lequel est susceptible d'éclairer, voire de contrôler son inconscient intellectuel dans l'intention d'en prendre acte afin de s'en libérer pour produire enfin la connaissance à même d'objectiver le « monde social », dans lequel il évolue également.

Or, si en prendre conscience lui est sans doute utile pour objectiver, l'analyse qu'il produit outrepassa singulièrement sa position sociale et se fonde bien davantage sur les règles et les opérations qu'elle sous-tend et que porte au jour l'exploitation de logiciels d'analyse qualitative comme ATLAS.ti. L'innovation n'est pas strictement d'ordre technique, elle est également méthodologique, permettant de montrer patte blanche en termes de rigueur.

En guise de brève conclusion

Voilà pourquoi il faut être infiniment reconnaissant aux créateurs de ces logiciels (voir Demazière, Brossaud & Trabal, 2006) de concevoir par ces moyens l'analyse qualitative comme un jeu, lequel, d'un point de vue épistémologique, fait preuve de sérieux et de rigueur pour peu qu'il se forme comme le veut l'épistémologie contemporaine « à partir de principes établis ». L'analyse qualitative assistée par ordinateur progresse grâce à ces logiciels qui permettent enfin de mettre au jour ce que sous-tend l'objet d'analyse. En effet, ils conduisent à dévoiler, en obligeant à les expliciter, les règles et les opérations en vertu desquelles les « réalités substantielles » à l'œuvre à l'échelle individuelle ou collective dans la connaissance pratique sont décrites et interprétées au moyen de la connaissance théorique, dans le but d'expliquer la « pratique sociale » en la représentant sur un autre registre que celui des événements, des circonstances, etc.

Si en sociologie décrire, comprendre et expliquer ne font qu'un, pour reprendre la formule de Bourdieu, c'est-à-dire jouent simultanément dans le feu de l'analyse, les logiciels d'analyse qualitative obligent à énoncer et à préciser

ce que recouvre chacun de ces trois mots. Sous ce chef, l'analyse assistée par ordinateur représente une innovation indéniable, car elle permet dorénavant de connaître exactement ce sur quoi se basent la description, la compréhension et l'explication dans la démarche qualitative. Voilà longtemps, Clifford Geertz (1986), anthropologue adepte de l'analyse interprétative, avouait que cette dernière était « l'une des choses qui comme la bicyclette [était] plus aisée à faire qu'à dire » (p. 16). Les logiciels comme ATLAS.ti invitent à penser aujourd'hui qu'il est possible d'élaborer l'analyse qualitative tout en étant capable de dire ce que l'on fait et de faire ce que l'on dit.

Références

- Bourdieu, P. (1992). *Réponses*. Paris : Seuil.
- Bourdieu, P. (1997). *Méditations pascaliennes*. Paris : Seuil.
- Bourdieu, P. (2003). L'objectivation participante. *Actes de la recherche en sciences sociales*, 150, 43-57.
- Bourdieu, P., & Accardo, A. (1993). *La misère du monde*. Paris : Seuil.
- Bourdieu, P., Passeron, J.-C., & Chamboredon J.-C. (1968). *Le métier de sociologue*. Paris : Mouton.
- Demazière, D., Brossaud, C., & Trabal, P. (2006). *Analyses textuelles en sociologie. Logiciels, méthodes, usages*. Rennes : Presses Universitaires de Rennes.
- Geertz, C. (1986). *Savoir local, savoir global*. Paris : PUF.
- Geertz, C. (1998). La description dense. *Enquête, La description*, 6, 73-105.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago : Aldine.
- Granger, G.-G. (1986). Pour une épistémologie du travail scientifique. Dans J. Hamburger (Éd.), *La philosophie des sciences aujourd'hui* (pp. 111-122). Paris : Gauthier-Villars.
- Granger, G.-G. (1988). *Pour la connaissance philosophique*. Paris : Odile Jacob.
- Granger, G.-G. (1992). Définir, décrire, montrer. *Alfa*, 5, 3-16.
- Granger, G.-G. (1993). *La science et les sciences*. Paris : PUF.
- Passeron, J.-C. (1991). *Le raisonnement sociologique. L'espace poppérien du raisonnement naturel*. Paris : Nathan.
- Passeron, J.-C. (1993). Anthropologie et sociologie. *Raison présente*, 108, 1-34.

Passeron, J.-C. (1996). La constitution des sciences sociales. *Le débat*, 90, 93-112.

Strauss, A., & Corbin, J. (2004). *Les fondements de la recherche qualitative*. Fribourg : Academic Press.

Jacques Hamel est professeur titulaire au département de sociologie de l'Université de Montréal où il est responsable des enseignements liés à l'étude de la jeunesse. Il est également membre de l'assemblée des chercheurs associés à l'Observatoire jeunes et société. Ses domaines de recherche touchent actuellement l'insertion de la génération numérique dans la « nouvelle économie » et les valeurs des étudiants à l'ère de la « société du savoir ». Il est également l'auteur de nombreux écrits sur l'épistémologie et la méthodologie qualitative. Ses recherches en la matière concernent présentement l'objectivation en sociologie. Sur le sujet, il publiera prochainement deux ouvrages intitulés *Woody Allen au secours de la sociologie* (Paris, Économica) et *La qualité de l'analyse qualitative interdisciplinaire* (Paris, L'Harmattan).

(Ne pas) Enseigner les méthodes qualitatives et (ni) former aux logiciels d'analyse des données qualitatives : une approche constructiviste de l'enseignement de la recherche qualitative

Diógenes Carvajal, M.A.

Université des Andes

Résumé

La formation des chercheurs à l'approche qualitative est un défi pour l'enseignement supérieur, notamment parce que ce n'est pas bien perçu par une partie des chercheurs. D'une part, il y a un risque que cette approche soit considérée comme une alternative simple par les étudiants dans un processus d'analyse parce que « tout est intéressant », et d'autre part, que certaines méthodes soient mobilisées comme une recette pour donner une validité au projet, ou pire, en utilisant un logiciel comme seule recette. Comment éviter que le processus d'enseignement et de formation des méthodes qualitatives et de l'utilisation des logiciels ne débouchent sur de tels résultats? La théorie constructiviste, comme théorie de l'enseignement, peut éclairer ces questions. Je vais présenter une façon d'introduire le constructivisme dans le processus d'apprentissage et de formation à la recherche qualitative.

Mots clés

RECHERCHE QUALITATIVE, ENSEIGNEMENT, FORMATION, CONSTRUCTIVISME, LOGICIEL

Introduction

Comment enseignons-nous la recherche qualitative aujourd'hui? Pourquoi est-il si important d'enseigner les fondements épistémologiques? Quel est le rôle des méthodes de recherche qualitative dans le parcours d'apprentissage des nouveaux chercheurs et qu'en font-ils? Comment les logiciels d'analyse qualitative des données peuvent-ils être incorporés dans le processus d'enseignement et d'apprentissage des nouveaux chercheurs? Ces questions ont émergé au fil de la pratique des chercheurs qui enseignent la recherche qualitative. Mais existe-t-il une réponse à celles-ci?

Le champ de l'enseignement de la recherche qualitative a été longtemps mis de côté dans la formation des enseignants-chercheurs; tous, nous

enseignons la méthode traditionnelle telle que nous l'avons apprise. Nous répétons le même plan qui commence par les fondements épistémologiques et termine par une série de méthodes. Peut-être certains d'entre nous ont été confrontés, de temps à autre, aux questions d'étudiants lorsqu'ils travaillent sur leur travail de recherche : « Bon, j'ai collecté mes données, je connais les méthodes, qu'est-ce que je dois faire avec mes données maintenant? » Cette question m'a interpellé : Où avons-nous échoué? Est-ce notre faute? Qu'avons-nous raté dans le processus d'apprentissage – formation qui amène les étudiants à poser ce genre de questions?

Je n'ai pas de réponse mais dans ma recherche d'une solution, j'ai imaginé que peut-être, nous enseignons la recherche qualitative d'une mauvaise manière ou, pour le dire autrement, d'une façon qui ne permette pas aux étudiants de comprendre en quoi consiste la recherche qualitative. Mais ce type de questions d'étudiants n'est pas le seul souci; dans la plupart des cas, elle est suivie d'une affirmation. Maintenant je dois apprendre à utiliser un logiciel d'analyse qualitative de données pour les traiter et sortir mon rapport final. Elle constitue une autre source d'inquiétude.

Dans cette réflexion, je pense que le constructivisme peut nous aider. Je voudrais partager une proposition pour enseigner la recherche qualitative et former à l'utilisation des logiciels afin de permettre aux étudiants de comprendre en quoi consiste la recherche qualitative, au-delà des méthodes, et en quoi les logiciels soutiennent le processus d'analyse.

Cet article est divisé en quatre parties. La première s'intéresse à la manière traditionnelle d'enseigner la recherche qualitative. La seconde se concentre sur les idées fausses des nouveaux chercheurs concernant le soutien que les logiciels peuvent leur apporter dans leur analyse. La troisième introduit quelques principes constructivistes utiles dans le cadre de l'enseignement de la recherche qualitative. Le dernier concerne la proposition en elle-même. Celle-ci n'a pas encore fait l'objet d'application dans son entièreté mais certaines parties ont été mobilisées avec des individus ou des petits groupes d'étudiants, du premier degré ou de maîtrise et de doctorat, et également avec des chercheurs indépendants.

La recherche qualitative dans le cursus traditionnel

L'enseignement traditionnel de la recherche qualitative : fondements épistémologiques et méthodes

La conception existante de l'enseignement de la recherche qualitative est double. Tout d'abord, chaque étudiant doit connaître les fondements philosophiques de la recherche qualitative : ses origines, l'histoire, les différentes approches, les bases épistémologiques, etc. Ensuite, les étudiants

doivent connaître les méthodes. Ce que j'appelle méthodes est ce que d'autres auteurs considèrent comme des types de recherche qualitative; par exemple, ethnographie, analyse de contenu, recherche-action, phénoménologie, parmi d'autres. Dans son livre *Qualitative research. analysis types and software tools* (1990), Renata Tesch a non seulement listé 46 types de recherche qualitative (voir Appendice 1), mais a également soulevé un problème majeur concernant cette liste :

The problem with this list is not only that it is far too long, but that some of the terms overlap with or are synonyms for others, and not all terms are on the same conceptual level. Some terms describe the perspectives qualitative researchers adopt (for instance « naturalistic », « interpretive », « experiential », « clinical »), or the tradition in the field on which they base their stance (such as « ethnography »). Other terms refer to the research approach used (« discourse analysis », « case study », « action research », etc.), or merely to the type of data, method, or research location (« document study », « participant observation », « field research », « oral history »).

Even this grouping of terms is forced. Sometimes it is difficult to distinguish clearly between labels that denote an epistemological stance and those that refer to method (p. 58).

D'autres chercheurs considèrent que les méthodes sont les moyens de collecter des données qualitatives, par exemple, les entretiens approfondis, les focus-group, l'observation directe, et d'autres; toutefois ce sont moins des méthodes que des techniques de collecte qui peuvent être mobilisées dans différentes méthodes. Une rapide enquête auprès de certains enseignants en recherche qualitative concernant les méthodes enseignées m'a permis d'identifier trois méthodes principales : l'ethnographie, les « cultural studies » et l'analyse de discours, avec de temps en temps, la théorie enracinée (grounded theory).

Certaines questions sont soulevées à ce moment : S'il existe une grande variété de méthodes, qui décide quelles méthodes enseigner? Quels sont les critères qui permettent de sélectionner deux ou trois méthodes parmi l'ensemble? Est-ce que ces trois méthodes suffisent pour permettre aux étudiants de réaliser une recherche qualitative?

Une fois que les étudiants ont appris les fondements philosophiques et épistémologiques, ainsi que quelques méthodes qualitatives, ils sont prêts à faire de la recherche qualitative. Toutefois, c'est tout le contraire que j'observe : lorsqu'ils ont terminé les cours et qu'ils se retrouvent confrontés à

leur travail de fin d'étude, ils n'ont aucune idée de comment s'y prendre! Une autre question apparaît : dans l'ensemble des méthodes, laquelle est adaptée à leurs besoins? Et lorsqu'ils trouvent une réponse, une autre question surgit : Comment dois-je m'y prendre concrètement?

Ce que je souhaite souligner ici n'est pas que les étudiants n'ont pas besoin d'apprendre les fondements épistémologiques ou les différentes méthodes de la recherche qualitative. Au contraire, je crois que l'enseignement traditionnel rempli bien sa tâche à cet égard mais est insuffisant. Les étudiants n'ont pas seulement besoin de connaître l'épistémologie ou les méthodes, ils ont surtout besoin de savoir comment FAIRE une recherche qualitative, en particulier l'analyse qualitative de données, d'être confronté à une masse de documents et de décider ce qui est important au point d'être identifié comme texte pertinent (Auerbach & Silverstein, 2003), et qu'en faire, spécialement pour pouvoir dire quelque chose de nouveau à partir des données analysées.

Conséquences de l'enseignement traditionnel de la recherche qualitative

Qu'est-ce que nos étudiants apprennent lorsque nous leur enseignons la recherche qualitative de façon traditionnelle? De la théorie. Ils connaissent les fondements philosophiques de la recherche qualitative, ce qui est important, ainsi que quelques méthodes, ce qui est nécessaire également. Mais la plupart d'entre eux ne savent pas comment faire de la recherche qualitative. Ils ne savent pas à quoi cela correspond en pratique. Ils peuvent concevoir un projet de recherche qualitative, mais une fois qu'ils ont collecté les données, ils ne savent pas quoi en faire. Ils n'ont jamais été formés à cela. Tout au plus, leur a-t-on enseigné à collecter de l'information, en faire une sorte de systématisation, et en tirer quelques conclusions dans un temps relativement court.

Certains étudiants cherchent de l'aide à ce moment de leurs études. Et ce tant pour les étudiants du premier cycle que du second (master ou doctorat) qui sont confrontés à la recherche qualitative pour la première fois. Ils se souviennent des savoirs à propos de la recherche qualitative mais pas « comment faire » une recherche qualitative.

Et dans leur quête du « comment faire » une recherche qualitative, beaucoup d'étudiants se familiarisent avec les logiciels d'analyse de données qualitatives, et beaucoup d'entre eux croient que cela répond à toutes leurs questions. Bien plus que nous le pensons! Certains pensent que l'utilisation d'un logiciel dans leurs projets de recherche est suffisant pour traiter les données, obtenir les résultats et être considéré comme un chercheur qualitatif. A tout le moins, c'est ce que j'observe à travers mon expérience. C'est ici que commence l'interaction entre faire de la recherche qualitative et utiliser un logiciel pour soutenir le processus de recherche qualitative.

Les logiciels d'analyse qualitative de données

Le contexte d'usage des logiciels dans la capitale colombienne

Tout d'abord, j'aimerais donner un bref aperçu de l'endroit d'où j'écris. En Colombie, il était impensable, il y a une décennie, d'utiliser un logiciel pour assister l'analyse qualitative de données. En fait, j'ai commencé à entendre parler de logiciels QDA, en 1998, après quatre années de travail comme chercheur qualitatif. Mes attentes, ainsi que celles des autres membres de l'équipe de recherche à laquelle j'appartenais, étaient similaires à celles de nombreux novices : comment le logiciel peut-il faire l'analyse qualitative de données? Comment introduire les données dans le logiciel et quel type de rapport va-t-il nous donner? Comment devons-nous interpréter les résultats? Est-ce qu'il suffit de prendre les rapports pour les présenter dans les résultats de la recherche? Est-ce que le rapport du logiciel peut constituer le rapport de recherche en lui-même? Nous avons été très surpris de voir le logiciel et les actes à poser. Nous n'étions pas déçus, à tout le moins, pas moi.

Si les logiciels sont disponibles depuis les années 70 et bien plus encore dans les années 80, pourquoi commencent-ils à être connus seulement dans le milieu des années 90 dans mon pays? Parce qu'il s'agit d'un pays du Sud, en périphérie du savoir académique international, et parfois (d'autant plus à l'époque), qui reçoit les progrès du Nord avec des années voire des décennies de retard. Je ne développerais pas plus avant ce point. En Colombie, l'utilisation généralisée de logiciels commence donc il y a une décennie. Par généralisée, je veux dire qu'ils deviennent utilisés dans les universités privées et publiques les plus importantes, principalement à Bogota, capitale, où la plupart des universités se situent.

Au départ, l'utilisation de logiciel était limitée aux chercheurs et aux étudiants engagés sur des projets de recherche importants. Il y a sept ou huit ans, les étudiants du premier et surtout du second degré ont commencé à demander à être formés aux logiciels pour les assister dans leur analyse qualitative. Les universités se sont alors tournées vers les nouveautés dans le secteur et en particulier vers ce qui était utilisé dans le monde académique réputé. C'est ainsi que les logiciels ont été inclus dans les universités comme moyen potentiel pour soutenir l'analyse qualitative de données. Ce « moyen potentiel » n'est pas accessible à tous les étudiants. Aujourd'hui, certains sont formés à l'utilisation de logiciels et quelques-uns les utilisent dans leurs travaux. De façon limitée.

Les efforts des enseignants et les demandes des étudiants

Quel est le paysage actuel d'utilisation des logiciels QDA parmi les étudiants de Bogota? Comme mentionné précédemment, certaines universités incluent la formation aux logiciels dans leurs programmes de cours. Toutefois, cela s'organise de façon limitée et lorsque les étudiants le réclament. Et les enseignants ont du répondre à cette situation.

En ce qui concerne les enseignants, l'inclusion de la formation aux logiciels QDA dans leur cours est curieusement limitée parfois à une simple mention de leur existence, voire à quelques heures de formation aux fonctionnalités de base. Les raisons avancées par les professeurs sont soit qu'ils ne connaissent pas eux-mêmes le logiciel ou qu'ils ne disposent pas du temps nécessaire pour former les étudiants. Lorsqu'ils disposent du temps nécessaire, la formation se résume bien souvent à l'utilisation mécanique du logiciel, loin de la formation à l'analyse qualitative en elle-même. Cette pratique débouche sur ce que j'ai nommé « la pensée mécanique », plutôt que sur une utilisation analytique du logiciel en relation avec la méthode et la systématisation des données qualitatives (Carvajal, 2002).

Au moins les enseignants incluent les logiciels dans leurs cours et au plus les étudiants sont intéressés par ceux-ci. Ceci a donné lieu à l'organisation de groupes d'étudiants qui, de façon indépendante, recherchent et paient pour être formés aux logiciels. Il s'agit principalement d'étudiants de maîtrise en sciences sociales qui considèrent comme un manque important le fait que les universités ne forment pas à l'utilisation des logiciels dans le cadre de leur curriculum. Par exemple, une étudiante en Anthropologie m'a dit « nous, en tant que chercheur qualitatif en sciences sociales, devons être préparés à l'utilisation de tels logiciels, si c'est ce que les chercheurs d'autres pays font », elle a organisé un groupe de 12 étudiants pour financer une formation aux logiciels. Ou encore « c'est fondamental d'utiliser un logiciel pour réaliser le travail de fin d'étude, pourquoi les professeurs ne nous en parlent pas avant? » m'a dit un autre étudiant. Je voudrais répondre : pourquoi les étudiants croient que c'est fondamental d'utiliser un logiciel pour effectuer l'analyse qualitative des données?

La raison principale donnée par les étudiants est que ce n'est pas juste de devoir faire l'analyse qualitative des données à l'ancienne, c'est à dire à la main, alors qu'il est possible de gagner du temps en utilisant un programme dédié à cette tâche. Cette perception est liée à une autre : utiliser un logiciel d'analyse qualitative de données est équivalent à l'utilisation d'un logiciel d'analyse quantitative. Les chercheurs, tant novices qu'expérimentés qui sont confrontés pour la première fois à un logiciel QDA s'attendent à ce que le

programme *fasse* quelque chose équivalent à ce que les programmes statistiques *font*. Ils croient qu'il suffit simplement d'intégrer les données dans le logiciel et de procéder par quelques clics pour obtenir un résultat. En réalité, lors d'une formation récente donnée à des étudiants de maîtrise, même lors de la seconde session de quatre heures, certains continuaient de demander ce que le logiciel *fait* pour créer le réseau et lier les différents éléments à travers des relations sémantiques...

Je crois que la principale raison de la demande des étudiants est, à nouveau, la volonté de faire de la recherche de la même façon que dans les pays développés; si les chercheurs des grandes universités prestigieuses utilisent les logiciels QDA, alors nous, dans les pays émergents, devons calquer nos pratiques sur les leurs.

Les risques d'une utilisation des logiciels QDA déconnectée de l'apprentissage de la recherche qualitative

En conséquence de la formation limitée, si pas pauvre, donnée aux étudiants pour l'utilisation des logiciels QDA, et la requête de formation des étudiants, ces derniers les utilisent de la manière dont ils pensent qu'ils doivent être utilisés. Ceci débouche sur différentes utilisations des logiciels QDA; certaines sont de nouvelles façons de systématiser et d'analyser les données qualitatives, mais la plupart sont de mauvaises manières de faire de l'analyse qualitative de données, sans rapport avec des méthodes ou des principes fondamentaux.

Dans ces cas de mauvaises utilisations des logiciels, la raison principale est que les étudiants n'ont pas une vision claire de ce qu'est une analyse qualitative de données, et comment le logiciel peut soutenir cette démarche. D'autant plus lorsque les étudiants ne sont pas formés de façon critique à l'utilisation des logiciels. J'entends par là, lorsque les étudiants ou chercheurs novices sont formés à une utilisation mécanique du logiciel, sans recevoir d'information sur la relation entre le logiciel QDA et l'analyse qualitative de données. En d'autres mots, ils ne savent pas en quoi les fonctionnalités des logiciels sont utiles à ce qu'ils doivent faire de leurs données. Ici, le processus ou la méthode, est remplacé par le logiciel.

Deux cas pour illustrer cette situation.

- Dans le premier cas, un chercheur expérimenté utilise un logiciel QDA à partir d'un raisonnement quantitatif : il code le corpus en plusieurs centaines de codes. Chaque code couvre un, parfois deux, segments pertinents de texte. Chaque élément particulier identifié devait être codé sous un code spécifique. Le rapport entre les codes et les segments était quasi de 1 :1. Le fait que les autres chercheurs du projet n'utilisent que les six codes a priori et quelques codes émergents supplémentaires lui

paraissait incroyable, lui qui avait plus de 200 codes émergents des 50 documents analysés. Le problème : une approche quantitative de l'utilisation des logiciels QDA.

- Dans le second cas, un étudiant en sociologie décide de construire des réseaux pour montrer comment les différents éléments qui ont émergé de son analyse sont reliés entre eux et comment ces relations donnent une compréhension potentielle des différentes situations et du phénomène. Son rapport est un document de 120 pages. Soixante pages sont des réseaux avec entre 20 et 30 éléments reliés entre eux. Chaque réseau est suivi d'une, parfois une page et demi, en doubles interlignes qui expliquent le réseau. C'est un espace réduit pour expliquer comment 30 à 30 éléments sont reliés entre eux. En fait, la lecture montre que seuls quatre ou cinq éléments sont mobilisés et le reste n'est pas même mentionné. Le problème : croire que les réseaux ou les cartes sont suffisantes pour expliquer les résultats obtenus par le logiciel.

La recherche qualitative, les logiciels d'analyse qualitative de données et... constructivisme?

Comment la recherche qualitative, les logiciels d'analyse qualitative de données et le constructivisme peuvent-ils être associés? Tout d'abord, j'aimerais rappeler quelques-uns des problèmes qui surviennent lorsque nous enseignons la recherche qualitative de façon traditionnelle et que nous formons à l'utilisation des logiciels : les étudiants apprennent l'épistémologie, les méthodes, mais ne savent pas comment faire, concrètement, dans la vie réelle, une analyse qualitative, ce qui est un des aspects les plus importants si nous voulons faire avancer la théorie. Et lorsque les étudiants commencent à utiliser les logiciels QDA, le risque est grand, dû notamment au manque antérieur de connaissance sur comment réaliser une analyse qualitative, qu'ils le prennent pour la méthode et fassent l'amalgame entre faire une analyse qualitative et utiliser un logiciel, en mobilisant les fonctionnalités qu'il propose. Il est possible également que l'étudiant ou le chercheur novice en recherche qualitative aborde l'utilisation de logiciels QDA de la même façon qu'il l'a fait ou le ferait avec des logiciels d'analyse quantitative. Est-ce possible que nous, chercheurs qualitatifs enseignant la recherche qualitative et l'utilisation des logiciels, provoquions cela?

Je suppose qu'une réponse positive potentielle à cette question peut être donnée à partir du constructivisme. Quelques principes constructivistes utiles pour l'enseignement de la recherche qualitative et la formation aux logiciels sont proposés.

Quelques principes constructivistes

Parmi de nombreux principes constructivistes que nous pouvons trouver et utiliser en éducation, j'en retiens trois que je considère utile dans le contexte décrit. Ces principes sont : connaissance antérieure, savoir situé et activité authentique.

En référence au dernier, l'activité authentique, Brown, Collins et Duguid (1989), identifient quatre éléments fondamentaux en éducation :

1. apprendre des concepts sans leur contexte spécifique est comme apprendre du vocabulaire en lisant un dictionnaire;
2. les disciplines sont des réseaux de connaissances construits socialement et sont indispensables pour comprendre ce qui est réalisé sur et à partir d'elles;
3. très souvent, il est demandé aux étudiants d'utiliser des outils d'une discipline sans en adopter la culture;
4. l'apprentissage doit être vu comme un processus de socialisation dans une nouvelle culture.

Pour les auteurs, les activités authentiques sont des pratiques ordinaires d'une culture et sont le seul moyen pour les apprenants d'appréhender la manière, le sens et l'objectif pour lesquels les professionnels d'une discipline agissent.

Concernant le savoir situé, un étudiant apprend mieux (ou peut apprendre quelque chose) s'il est activement impliqué dans un contexte réaliste : le savoir émerge de l'activité développée par lui dans le contexte et la culture qui mobilise cette connaissance. Dès lors, un étudiant va développer la connaissance d'une discipline seulement s'il est immergé dans le contexte social où cette connaissance va être mobilisée.

Si les activités des chercheurs dans le cadre d'un projet sont des activités authentiques, sous-entendu réalisées dans un contexte où elles prennent sens et utilité, et si la connaissance qui émerge du projet est située, c'est à dire ayant un sens dans un contexte et une culture particulière, ne devrait-on pas former nos étudiants de la même façon? En ce, je veux dire que le processus d'apprentissage – enseignement en recherche qualitative doit être basé sur des faits réels et avoir des applications réelles dans un contexte particulier. Pour se faire, les connaissances antérieures des étudiants doivent être prises en compte.

La connaissance préalable est le point de départ : c'est ce que les étudiants savent à propos de quelque chose et qui peut être utile pour les aider à apprendre quelque chose de nouveau. Dewey (1925; 1945) considère l'expérience comme l'élément de base du processus d'apprentissage. Les expériences significatives vont contribuer au processus d'apprentissage si elles

sont appuyées par un processus de réflexion à leur propos; c'est à dire, il n'est pas question de laisser l'étudiant seul avec son expérience, le professeur doit intervenir pour indiquer les variations de l'expérience à l'étude et motiver les étudiants à réfléchir à ce qui pourrait évoluer dans l'expérience en cas de variation. C'est une façon de transférer une connaissance acquise à d'autres situations, ce qui est toujours présent dans les processus de recherche. Nous pouvons trouver ici une relation entre la connaissance antérieure et l'enseignement pour comprendre un principe. Perkins (1999) suggère que la compréhension d'une situation ne sous-entend pas seulement la compréhension des idées mais qu'il s'agit d'un processus plus complexe : il s'agit d'apprendre à agir d'une manière flexible.

Ritchhart, Stone Wiske, Buchovecky et Hetland (1999) considèrent la pratique de l'enseignement à la compréhension sous trois aspects :

1. l'arrière-plan : pour interpréter et adapter les éléments théoriques de l'enseignant au contexte particulier afin de leur donner du sens;
2. à l'intérieur des frontières : pour concevoir les ressources que l'enseignant va utiliser dans le cadre de son enseignement;
3. l'action : pour appliquer et intégrer concrètement la connaissance. Cet aspect se divise en trois étapes : l'exploration, la recherche guidée et le projet de synthèse final.

Le rôle des logiciels d'analyse qualitative dans le processus de recherche

À partir de ce qui est déclaré ci-dessus, comment les logiciels QDA peuvent-ils être inclus dans l'enseignement de la recherche qualitative? À ce niveau, je souhaite présenter ma première proposition publiée précédemment sans mentionner l'approche constructiviste en elle-même dans l'article *The artisan's tools. Critical issues when teaching and learning CAQDAS* (Carvajal, 2002). Les logiciels peuvent être intégrés dans l'enseignement de la recherche qualitative à partir de trois règles résumées ci-dessous (voir Carvajal 2002 pour de plus amples informations).

1. *Donner plus d'information que de développer des compétences* : la plupart des programmes de formation aux logiciels QDA sont focalisés sur l'utilisation du logiciel; par exemple, les étapes de base pour que le chercheur travaille avec ses documents, crée des codes, réalise des rapports et, dans certains cas, réalise une carte ou un réseau. Les programmes devraient toutefois inclure de l'information sur la recherche qualitative et l'analyse; sur la façon dont les outils du logiciel sont en rapport avec ce qui se fait en recherche qualitative selon la méthode choisie. Il ne s'agit pas seulement de savoir utiliser le logiciel. Il s'agit

de savoir quoi faire avec le logiciel pour qu'il soit utile au processus d'analyse qualitative.

2. *Promouvoir une pensée critique plutôt qu'une pensée mécanique* : la discussion sur les implications méthodologiques de ce que l'utilisation d'un logiciel entraîne dans la méthodologie qualitative est limitée aux académiques. C'est devenu un sujet quasi théorique non abordé dans les programmes de formation. Pourtant chaque utilisateur de logiciel QDA a quelque chose à dire à ce sujet si on lui en donne la possibilité. La discussion académique doit s'ouvrir au public et inclure les différents types de chercheurs. En introduisant des aspects comme les faiblesses des logiciels, la façon dont ils organisent et systématisent les données, les participants ne tomberont plus dans une néophilie, une utilisation non critique de nouvelles technologies, et seront plus prudents en utilisant un logiciel pour analyser leurs données.
3. *Analyser leurs propres données* : Cet aspect concerne directement les principes constructivistes abordés ici. Donner aux étudiants qui sont formés aux logiciels l'opportunité de travailler sur leur propre matériau, leur apporte un intérêt complémentaire dans le processus analytique. Une fois qu'ils savent utiliser les outils de base, il est important de leur permettre d'expérimenter le logiciel par eux-mêmes. Ce faisant, ils trouveront le meilleur moyen de segmenter et coder leur donnée, d'organiser les codes et les catégories selon leurs options méthodologiques.

En suivant ces règles, ce que j'ai fait ces dernières années, les étudiants peuvent comprendre pourquoi il est important d'utiliser les logiciels d'une façon qui leur permette de soutenir l'analyse qualitative de données, quand il est intéressant de les utiliser et quand il est préférable d'en revenir aux formes traditionnelles d'analyse des données.

Cette proposition est une partie de la proposition concernant l'enseignement de la recherche qualitative.

Une approche constructiviste de l'enseignement de la recherche qualitative

En conservant en tête les éléments présentés plus haut, je propose trois étapes à travers les étudiants devraient passer afin de les impliquer dans ce qu'est une recherche qualitative, comment elle peut être réalisée et pourquoi. Ces étapes ont des durées différentes.

Étape 1 : la recherche qualitative et la vie de tous les jours

Chaque jour, à chaque moment, nous classifions et analysons différents types d'information à travers un processus similaire à celui qui est réalisé, plus rigoureusement, dans la recherche. Les étudiants aussi. Dès lors, leurs expériences personnelles de tous les jours peuvent servir de point de départ pour identifier le processus de classification qu'ils utilisent et les critères utilisés pour analyser l'information classée avant de prendre une décision. Ces expériences montreront également comment certaines décisions impliquent des processus plus complexes, parce que mobilisant des aspects qui ne peuvent être mesurés. A ce moment, l'objectif est de reconnaître les différents aspects que nous prenons en compte pour prendre une décision et comment ces dimensions sont classifiées; il est même possible d'identifier certaines taxonomies utilisées pour classer l'information. En abordant cela, les étudiants peuvent se rendre compte de ce que sont des données qualitatives et comment traiter ces données à différents moments et par différents moyens.

Étape 2 : La relation entre l'expérience personnelle et les fondations théoriques de la recherche qualitative

Une fois que les étudiants comprennent ce que sont des données qualitatives et comment ils les systématisent de façon brute dans leur vie de tous les jours, il est temps d'introduire certains aspects théoriques de la recherche qualitative concernant les techniques traditionnelles de collecte : observation (qu'est-ce qui doit être observé et comment l'enregistrer); interviews (que demander pour recueillir de l'information pertinente); and focus groups (quand les utiliser). La question d'objectivité de la recherche qualitative rentre également dans le processus. Les étudiants réfléchissent à propos de la réflexivité et du rôle de la subjectivité dans la recherche qualitative. C'est ainsi que les fondements philosophiques sont introduits dans le cours; lorsque les étudiants sont déjà impliqués dans le processus de compréhension d'en quoi consiste la collecte de données qualitatives et de leur analyse.

Étape 3 : la recherche qualitative comme pratique et non comme théorie ou méthode

« Apprendre la recherche en faisant de la recherche » guide tout le processus. Dans cette étape, ce principe va être d'autant plus évident : les étudiants définissent une question de recherche qui peut être étudiée en collectant de l'information grâce aux techniques apprises. Ils vont analyser leurs données également. L'utilisation d'un logiciel va être autorisée, mais uniquement si elle est souhaitée par l'étudiant. La formation des étudiants aux logiciels se fait donc sur demande : de petits groupes d'étudiants intéressés par le soutien potentiel des logiciels à leur processus d'analyse vont recevoir une formation

respectant les principes présentés plus haut. Bien qu'ils reçoivent un accès à un logiciel, ils sont libres de décider non seulement quand l'utiliser, d'en utiliser un autre, voire de ne pas en utiliser du tout.

Cette étape commence par la conception d'un projet de recherche, la collecte des données qualitatives et l'étude du processus de systématisation et d'analyse de ces données; aucune méthode n'est abordée ici, seulement les aspects principaux impliqués dans une analyse qualitative de données. L'attention est portée au processus en lui-même, pas dans les méthodes particulières. Une fois que ce processus est compris par les étudiants, ils peuvent eux-mêmes se renseigner sur différentes méthodes et sélectionner celle qui convient mieux à leur question de recherche.

À ce moment, les étudiants continuent leur processus de réflexions sur la recherche qualitative, avec une attention sur deux aspects; un élément mineur : y a-t-il besoin d'un logiciel dans leur processus d'analyse qualitative, et un élément majeur : la place d'un cadre théorique dans l'analyse qualitative de données. Ce dernier élément est soutenu par le professeur qui donne aux étudiants l'opportunité de lire différents rapports de recherche qualitative pour qu'ils puissent les analyser et trouver du matériau théorique sur leur sujet.

L'avenir de l'enseignement

Actuellement, cette proposition est mise en œuvre avec des étudiants du premier et du second degré. Il est trop tôt pour tirer des conclusions sur la façon dont les étudiants apprennent la recherche qualitative et si il y a une différence parmi eux et ceux qui reçoivent un enseignement traditionnel. Un projet de recherche est nécessaire pour évaluer cette différence et voir si elle est due aux changements introduit par cette proposition. J'invite d'autres chercheurs à réaliser certain changement dans leur façon d'enseigner la recherche qualitative, dans le respect des principes épistémologiques, mais sans se limiter aux méthodes ou à l'utilisation de logiciels. Ces deux aspects ont tendance à avoir trop de poids dans le processus de formation des nouveaux chercheurs. Certains enseignants ont oublié que les étudiants ont besoin de savoir comment FAIRE de la recherche qualitative. Pas seulement la théorie.

Références

- Auerbach, C.F., & Silverstein, L.B. (2003). *Qualitative data. An introduction to coding and analysis*. New York : University Press, New York.
- Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.

- Carvajal, D. (2002). The artisan's tools. Critical issues when teaching and learning CAQDAS [47 paragraphs]. *Forum Qualitative Sozialforschung / Forum : Qualitative Social Research*, 3(2). <http://nbn-resolving.de/urn:nbn:de:0114-fqs0202147>
- Dewey, J. (1925/1981). Nature, mind and the subject. Dans A. Boydston, *John Dewey. The Later Works, 1925-1953*, vol. 1 (pp. 162-190). Carbondale, IL : Southern Illinois University Press.
- Dewey, J. (1945). *Experiencia y educación* (4^e éd.). Buenos Aires, Argentine : Losada.
- Perkins, D. (1999). ¿Qué es la comprensión? Dans M. Stone Wiske (Éd.). *La enseñanza para la comprensión. Vinculación entre la investigación y la práctica* (pp. 69-92). Buenos Aires : Paidós.
- Ritchhart, R., Stone Wiske, M., Buchovecky, E., & Hetland, L. (1999). ¿Cómo se ve en la práctica la Enseñanza para la Comprensión? Dans M. Stone Wiske (Éd.). *La enseñanza para la comprensión. Vinculación entre la investigación y la práctica* (pp. 169-212). Buenos Aires : Paidós.
- Tesh, R. (1990). *Qualitative research : Analysis types and software tools*. Bristol : The Falmer.

Diógenes Carvajal est psychologue avec maîtrise en éducation. Il a une grande expérience dans l'usage et l'enseignement de logiciels pour l'analyse qualitative. Il a enseigné la recherche qualitative dans diverses universités de la Colombie et a publié de nombreux articles critiques sur le sujet. Il est actuellement professeur adjoint au Centre de recherche et de formation en éducation de l'Université des Andes à Bogotá, en Colombie. Il y combine l'enseignement de la recherche qualitative avec la recherche sur divers sujets éducatifs, dont l'enseignement de l'usage de logiciel pour une analyse qualitative à partir d'une perspective constructiviste.

Appendice 1

Liste des types de recherche qualitative, selon Tesch (1990, p. 58)

Action research	Focus groups research
Case study	Grounded theory
Clinical research	Hermeneutics
Cognitive anthropology	Heuristic research
Collaborative inquiry	Holistic ethnography
Content analysis	Imaginal psychology
Dialogical research	Intensive research
Conversation analysis	Interpretive evaluation
Delphi study	Interpretive interactionism
Descriptive research	Interpretive human studies
Direct research	Life history study
Discourse analysis	Naturalistic inquiry
Document study	Oral history
Ecological psychology	Panel research
Educational connoisseurship and criticism	Participant observation
Educational ethnography	Participative research
Ethnographic content analysis	Phenomenography
Ethnography	Phenomenology
Ethnography of communication	Qualitative evaluation
Ethnomethodology	Structural ethnography
Ethnoscience	Symbolic interactionism
Experiencial psychology	Trascendental realism
Field study	Transformative research

Voir également Tesch (1990, p. 72) pour un aperçu graphique de types de recherche qualitative selon les objectifs de recherche.