

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

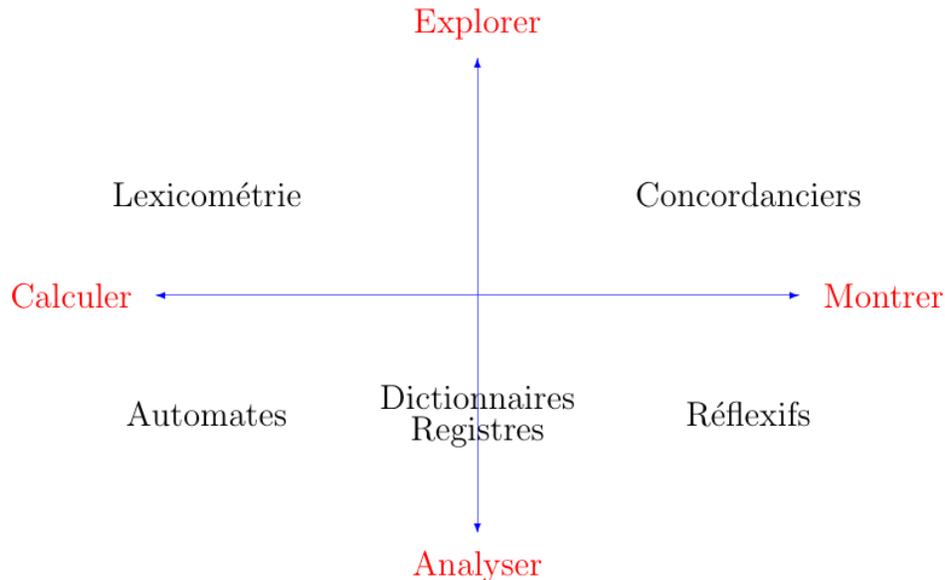


Figure 1. Familles de fonctionnalités

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

The screenshot shows a window titled 'Spécifs - Part : semaine'. It displays a reference corpus of 111-122 and 211-222, with 211 selected parts. The 'Spécificités' section has 'positives' selected. The table below lists terms with their total frequency, frequency in the selected parts, and specific frequency.

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

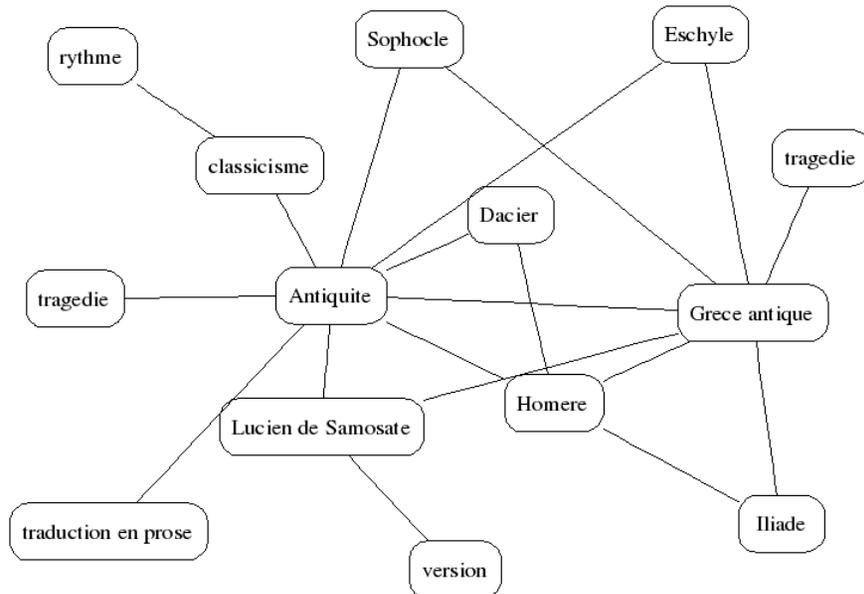


Figure 4. Automate (d'après Candide)

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).

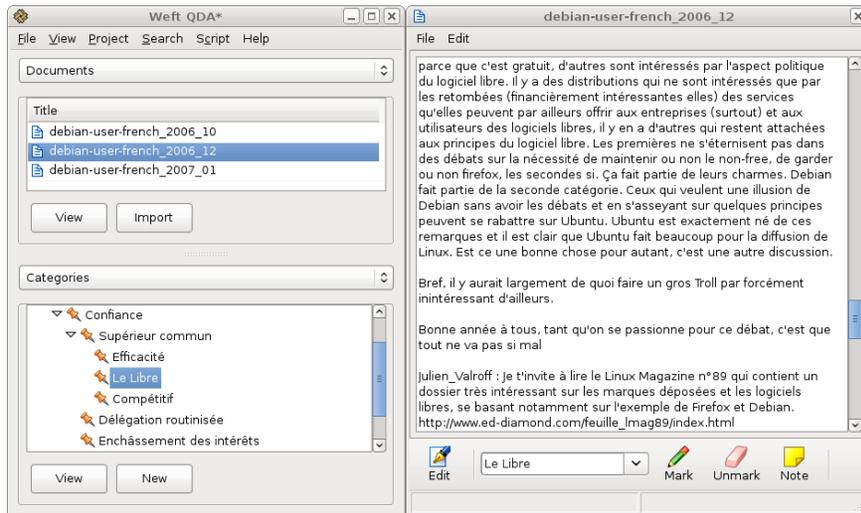


Figure 5. Outil réflexif (copie d'écran de Weft QDA)

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. Registres (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 Bénel 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>
- Bénel, 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.