

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 questionnaire 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 & Jacquey) 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.

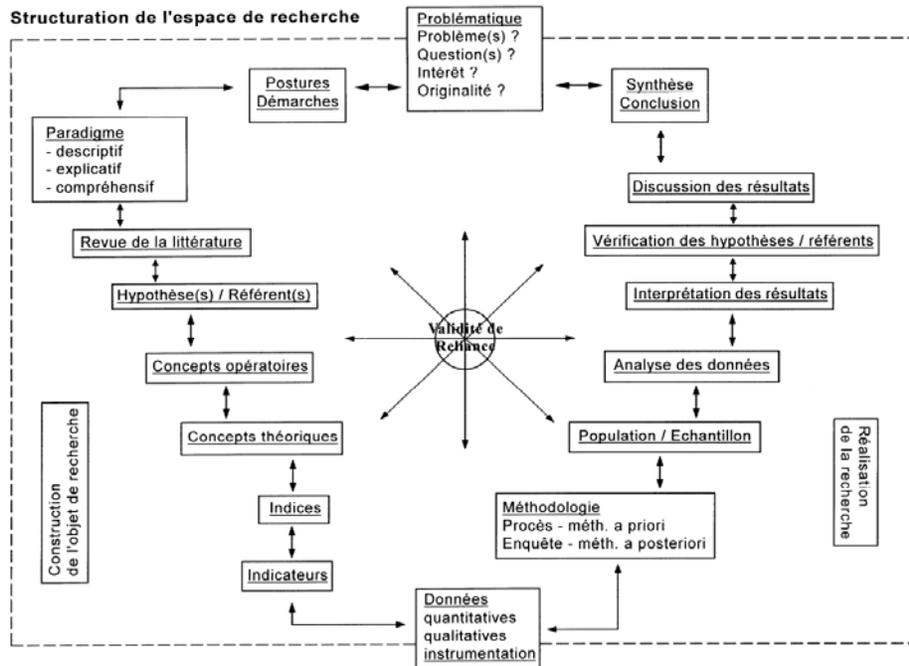


Figure 1. Les points charnières de la recherche (extrait de Pourtois et al., 2006)

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.