

# Logiciels d'analyse de données qualitatives ou d'analyse qualitative ?

Pièges, limites et questions liminaires

*Christophe Lejeune*

## Introduction

La généralisation de l'informatique et la multiplication des données numériques ont banalisé l'usage de logiciels d'analyse de textes<sup>1</sup>. Le nombre de logiciels disponibles, leur technicité et leur coût (non seulement monétaire, mais surtout en temps de formation) obligent le chercheur ou l'étudiant en santé à opérer un choix. Ce chapitre invite le lecteur à se poser les bonnes questions (et à se les poser tôt) avant d'ouvrir une fenêtre pour entamer des tests.

### Vignette 1. Un point de passage obligé

Les 21 et 22 octobre 2008, se tenait à Luxembourg une conférence sur l'usage des logiciels d'analyse de données qualitatives en sciences humaines. Au terme de la première session, je profite de la pause-café pour assister l'intervenant

---

1. Je tiens à remercier tous les étudiants et les chercheurs qui m'interrogent sur l'analyse qualitative. Ce chapitre leur doit beaucoup. Je suis en particulier redevable aux échanges entretenus avec Quentin Bullens, Cécile Delcourt, Hélène Langinier, Nadège Lorquet, Cécile Mathys, Julie Rijpens, Thierry Soulas et Viviane Vierset. Merci enfin à Joëlle Kivits et Myriam Winance de m'avoir invité à écrire ce chapitre et d'en avoir accompagné la maturation avec patience et bienveillance. Je suis évidemment seul responsable des erreurs ou des imperfections qui y subsisteraient.

suivant à brancher son ordinateur. Je découvre alors que tous les participants n'ont pas quitté la salle. Anne, Benoît et Claire y discutent entre eux. Ils sont les seuls à y être restés alors que les autres participants échangent informellement autour d'une tasse de café. Je les interpelle et, sur le ton de l'humour, leur signale que ce qui se passe à la pause-café est au moins aussi important pour eux que ce qui se raconte durant la conférence. Ils m'objectent que cette conférence ne porte pas sur une thématique ou un sujet proche de leur doctorat. Ses enjeux ne les concernant pas, y nouer des contacts ne présente pas d'intérêt pour eux. Je m'étonne dès lors de leur présence, étant donné le montant de l'inscription, relativement élevé pour les participants ne présentant pas de communication. Ils m'expliquent qu'ils sont envoyés par leur directeur de thèse. Ce dernier considère qu'un doctorant ne peut actuellement pas faire l'économie d'une bonne connaissance de tels logiciels. Or, les recherches d'Anne, de Benoît et de Claire progressent normalement sans que se manifeste le besoin de recourir à un logiciel. ■

Cette vignette illustre la place attribuée aujourd'hui aux logiciels d'analyse de données qualitatives. De plus en plus de chercheurs en santé considèrent que les logiciels constituent un point de passage obligé. Cela n'a pourtant rien d'évident. Bien entendu, si le chercheur travaille mieux avec un logiciel que s'il en est dépourvu, rien ne s'oppose *a priori* à ce qu'il y recoure. La situation est différente pour les chercheurs qui se sentent moins à l'aise avec l'informatique ou qui ont acquis des habitudes productives sans logiciel. Phénomène marquant : les services que ces outils sont susceptibles de rendre aux chercheurs n'ont pas été soumis à une réflexion méthodologique sur la valeur ajoutée de leur usage. S'y former se présente comme une étape nécessaire à l'apprentissage du métier de chercheur. Or, l'éventuel recours à un outil dépend nécessairement de choix méthodologiques mûrement réfléchis. Les questions qu'il faudrait se poser avant d'utiliser un logiciel sont relativement nombreuses.

1. Tout d'abord, même si cela a l'air trivial, il convient de se demander s'il est (vraiment) opportun d'utiliser un logiciel dans la recherche que l'on entend conduire.

2. Il importe ensuite de se demander pourquoi on souhaite utiliser un logiciel d'analyse de données qualitatives. Non seulement pour quoi faire, mais également pour quelles raisons, dans quel but ?

3. Il faudrait encore se poser la question du moment pertinent pour décider si l'on utilise un logiciel ou pas, pour choisir le logiciel en question et enfin, le moment opportun pour utiliser ce logiciel.

4. Enfin, la dernière question porte évidemment sur le choix du logiciel : quel logiciel utiliser ?

## 1. Les logiciels sont-ils nécessaires ?

### Vignette 2. Le tribunal de la preuve

Dominique termine une thèse en psychologie. Elle a étudié la construction de l'identité des couples homosexuels, sous un angle d'attaque psychanalytique. Lorsqu'elle me contacte, la recherche est pratiquement terminée ; elle est en train de rédiger sa dissertation doctorale. Un membre de son comité lui a néanmoins conseillé de soumettre ses entretiens au traitement d'un logiciel d'analyse automatique de textes afin de « confirmer » ses analyses et de valider ses résultats. Dominique m'explique que ses choix théoriques ne font pas nécessairement consensus, la psychanalyse étant aujourd'hui fortement décriée en faculté de psychologie. Elle ne souhaite donc pas faire l'économie d'un dispositif méthodologique susceptible de prouver la pertinence de ses interprétations. ■

Je réponds à Dominique de la manière suivante. L'utilisation du logiciel peut déboucher sur deux issues. Dans le premier cas, le logiciel produit des résultats qui ne sont pas incompatibles avec l'analyse déjà produite. Il sera alors aisé à Dominique d'argumenter dans le sens d'une « validation » de ses conclusions. Mais le scénario opposé s'avère tout aussi vraisemblable. Les résultats produits par la machine pourraient profondément diverger ou, à tout le moins, s'avérer difficilement conciliables avec les analyses de Dominique. Dès lors, que fera-t-elle, à quelques mois de la soutenance ? Va-t-elle, pour cette raison, invalider l'ensemble de son travail, auquel elle a consacré énergie, temps et soin, sous prétexte qu'un logiciel produisant des recoupements automatiques présente une mise en forme des données qui semble incompatible ?

Je conseille à Dominique de poursuivre sa rédaction sans tenir compte de la suggestion de son comité. En effet, dans les deux cas, que le logiciel produise des résultats conciliables ou inconciliables avec ses analyses, la valeur ajoutée sera maigre voire, dans le cas d'une incompatibilité, pourrait engendrer la démotivation d'une impétrante ayant pratiquement mené à terme un processus long et risqué.

Demander à l'ordinateur de valider, après coup, une analyse déjà réalisée est certes inhabituel. Cette démarche s'appuie en fait sur plusieurs présupposés concernant la recherche qualitative et l'informatique. Si Dominique les accepte tacitement en me contactant, ces présupposés sont toutefois loin de faire l'unanimité. Examinons le raisonnement que suggère le membre du comité de thèse. Ce dernier oppose le traitement automatique de la machine à l'interprétation subjective du psychologue. De fait, ces deux opérations diffèrent profondément : contrairement au chercheur, l'ordinateur n'interprète pas le matériau. En cela, son traitement est objectif. Cette objectivité

ne constitue pas pour autant un gage de scientificité. Par contre, les interprétations de la chercheuse sont nécessaires à des conclusions scientifiques. Acter la différence entre la chercheuse et la machine n'implique donc pas de soumettre la validité d'une interprétation à un algorithme déterministe. Une telle proposition procéderait d'un positivisme naïf, indigne de la recherche scientifique (Bourdieu, Chamboredon et Passeron, 1968 : 55-56).

La recherche qualitative n'est pas unanime sur la valence à accorder à l'opposition entre objectivité et subjectivité. Certains chercheurs entendent objectiver leurs analyses ; d'autres assument la subjectivité de leurs interprétations, moins comme une faiblesse que comme une orientation personnelle, réfléchie et délibérée, produisant des conclusions d'autant plus riches qu'elles sont originales.

N'écartons donc pas trop vite l'argument de l'objectivité des calculateurs. Celui-ci a en effet beaucoup compté dans l'histoire des méthodes qualitatives. Il a en effet offert une solution à un problème rencontré en analyse de contenu. Formalisée dès les années 1930, cette technique consistait à appliquer la statistique descriptive aux thématiques abordées dans la presse. L'identification de ces thématiques par des codeurs était validée selon un protocole strict : le relevé n'était réputé objectif que s'il emportait l'accord de plusieurs codeurs<sup>1</sup>. Or, les avis divergeaient souvent, même pour une opération aussi simple qu'un relevé thématique. Multiplier les consignes de codage n'y changeait rien. L'ordinateur a fourni aux analystes de contenu l'opérateur de codage stable, docile et prévisible dont ils avaient besoin. En partenariat avec IBM, Philip Stone mit au point le *General Inquirer*, une machine identifiant automatiquement les thématiques selon un critère objectif : la présence d'une série de mots-clés. L'objectivité de la machine apporte donc une solution (et un argument de poids) à l'analyse de contenu. Les analystes de contenu ne considèrent pas pour autant que l'ordinateur interprète ou analyse les textes. Son apport concerne uniquement le tri. Comme en statistique, c'est bien au scientifique qu'il incombe d'interpréter les traitements de la machine<sup>2</sup>. Les analystes de contenu ne soumettent donc pas l'interprétation à la machine, bien au contraire.

Le membre du comité de thèse de Dominique n'a vraisemblablement retenu qu'une partie de l'histoire de l'analyse de contenu. Ne disqualifions pas son conseil pour autant. En l'occurrence, il est vraisemblable que les réserves adressées à Dominique devaient plus à la position hétérodoxe de la psychanalyse qu'à une foi aveugle dans l'objectivité des machines. Quoi qu'il en soit, l'argument de l'objectivité machinique reste tenace. D'ailleurs,

1. Connues sous le nom de « fidélité inter-codeurs » ou d'« accord inter-juges », différentes méthodes de calcul se proposent de mesurer ce consensus (Krippendorff, 2004).

2. Ce principe est affirmé dès les premières lignes de la préface du premier livre dédié à un logiciel d'analyse de contenu (Stone *et al.*, 1966 : xi).

en me contactant, Dominique y adhère sans doute partiellement. Comme elle, des chercheurs travaillant sans logiciel me consultent régulièrement et me demandent s'il est vraiment *nécessaire* de recourir aux logiciels d'analyse de données qualitatives. Bien évidemment, non !

Comme la suggestion du comité de thèse de Dominique, l'interrogation, voire l'inquiétude de ces collègues trahit une croyance, répandue, que les logiciels confèrent aux résultats une objectivité qui ferait défaut à une analyse non informatisée (Paillé, 2006). Or, recourir à un logiciel d'analyse de textes ne garantit ni la qualité de la recherche, ni la scientificité des preuves, ni surtout l'intérêt des conclusions. En fait, la question de savoir si l'on doit, peut ou souhaite utiliser un logiciel est indissociable de la question suivante, à savoir « pourquoi » on utilise un outil d'analyse des données qualitatives.

## 2. Pourquoi utiliser un logiciel d'analyse qualitative ?

La technique est rassurante. D'aucuns s'y réfugient, remettant à plus tard des choix méthodologiques décisifs ou évitant ainsi de se confronter à une littérature scientifique intimidante. Se demander ce qui motive le recours aux logiciels permet d'éviter que la technique se substitue à la consultation de la littérature scientifique ou à la construction patiente et rigoureuse d'une analyse documentée, personnelle et originale.

### 2.1. D'où viennent les interprétations ?

#### Vignette 3. La génération spontanée

Élise réalise une thèse en sciences de l'information et de la communication. Elle a conduit un grand nombre d'entretiens, les a transcrits et est en train de les analyser avec un logiciel réputé. Chaque année, elle rencontre son comité de thèse. Chaque année, ce comité lui demande de délivrer ses résultats intermédiaires. Élise refuse à chaque fois. Selon elle, le matériau doit être intégralement codé pour que des résultats soient effectivement délivrables. ■

Élise souffre d'un syndrome que je qualifie de codomanie (Lejeune, 2014 : 82). Dans une acception un peu particulière de l'analyse par théorisation ancrée, elle considère que coder produit nécessairement de la théorie. Selon cette acception, la théorie apparaîtrait sans investissement intellectuel du chercheur, presque automatiquement, par « génération spontanée ». Cette conception est sans doute encouragée par les développements de Barney Glaser (1992) opposant les démarches hypothético-déductives (qui

« forcent » les données) aux approches dites « inductives » qui encouragent « l'émergence » de nouvelles théories. Or, une analyse n'émerge jamais d'elle-même. Seule la construction patiente, délibérée et active de l'utilisateur du logiciel la produit.

## 2.2. Les théories aveuglent-elles ?

### Vignette 4. Le refus de la théorie

François est sociologue. Son orientation théorique doit beaucoup au courant dit pragmatique. Il est particulièrement sensible aux critiques formulées par ce courant à l'encontre des grandes théories sociologiques qui, saturées conceptuellement, fonctionnent indépendamment du terrain. Un tel argument rappelle évidemment la critique virulente que formulent Anselm Strauss et Barney Glaser à l'encontre des démarches hypothético-déductives. Assimilant la théorie sociologique aux préjugés du sens commun, donc à des œillères qui empêcheraient dès lors de voir ce qui se passe réellement sur le terrain, François conclut qu'il faut se passer de théorie. Dans cette optique, il utilise un logiciel d'analyse statistique des données textuelles pour identifier des structures discursives ou des topiques argumentatives, que la récurrence rend objectivables et donc, selon lui, objectives. ■

On peut difficilement reprocher à François d'adhérer à l'ouverture d'esprit préconisée par ses lectures. François cependant va plus loin que la sociologie pragmatique ou l'analyse par théorisation ancrée. Aucune ne suggère qu'il faille se passer de théoriser. Au contraire, dans les deux cas, la critique sert de prémisse à une invitation à produire de nouvelles théories.

Malheureusement, le problème ne se réduit pas à une question de lecture. La posture de François le rend en fait sourd aux arguments lui objectant que sa démarche est anti-sociologique, voire non scientifique. Pour lui, la génération précédente de sociologues se caractérise par une approche hypothético-déductive. Ses pères ont pris la mauvaise habitude de faire violence aux matériaux par l'application de théories préexistantes. Dès lors, chaque fois que ses aînés proposent à François de travailler à sa question de recherche, il écarte ce conseil comme résultant d'un manque d'ouverture d'esprit et d'un rapport pathologique à la théorie. Là où François interprète erronément les thèses de la sociologie pragmatique et de l'analyse qualitative, c'est que la critique des grandes théories n'implique pas l'absence de références théoriques. Son usage des logiciels d'analyse de textes se substitue à un questionnement théorique et risque bien de produire une description, peut-être fine, mais relativement peu productive dans la mesure où elle n'entend pas répondre à une question de recherche. Sa posture rompt tout dialogue avec ses pairs et le condamne, à terme, à disparaître en tant que chercheur.

Je suis particulièrement sensible à cette façon de travailler dans la mesure où, jeune doctorant, j'ai adopté une posture similaire et suis, aujourd'hui, d'autant plus conscient de mon erreur. Je comprends d'autant mieux les résistances des doctorants à qui je conseille de consulter la littérature scientifique, d'élaborer une question de recherche et de se poser des questions méthodologiques avant de se lancer dans le codage forcené et tous azimuts d'un épais corpus empirique.

## 2.3. Que font les logiciels ?

### Vignette 5. Compter ou expliciter ?

Sage-femme expérimentée, Gaëlle a décidé d'étudier les façons dont les maisons de naissance inspirent confiance aux futurs parents. Dans une démarche ethnographique, elle a quotidiennement consigné ses observations, ses questions et ses notes de lecture dans un logiciel d'analyse qualitative. Elle présente ses premières conclusions lors d'une journée doctorale en santé. Sur son diaporama, chaque observation est accompagnée du nombre de situations observées. Je l'interroge sur la pertinence de présenter des comptages, dont aucun ne dépasse la dizaine. Elle me répond que son directeur de thèse, médecin, considère que des conclusions scientifiques doivent être chiffrées. ■

Le contexte scientifique de la francophonie diffère de celui des recherches internationales publiées en anglais. La littérature anglophone désigne les logiciels d'analyse qualitative par le sigle QDA ou CAQDAS (*Computer Assisted Qualitative Data Analysis*). Ce sigle renvoie à une série d'outils relativement comparables et souscrivant à un seul et même type de fonctionnalités. Afin de ne pas les confondre avec les autres logiciels d'analyse de textes, je qualifie ces logiciels d'outils d'étiquetage réflexif (Lejeune, 2010). La littérature scientifique francophone parle peu de ces logiciels. Elle est par contre prolixes à propos des logiciels d'analyse de textes. En première lecture, la différence peut sembler négligeable et se réduire à une simple divergence de dénomination. Pourtant, les outils visés diffèrent profondément.

Les argumentaires en faveur de l'analyse de textes proviennent d'utilisateurs et de concepteurs de logiciels n'ayant pas grand-chose à voir avec la recherche qualitative. Ceux-ci s'inscrivent dans une autre tradition méthodologique : l'analyse statistique de données textuelles, parfois articulée à une tradition particulière de l'analyse de discours (Lejeune, 2017). Les plaidoyers en faveur des logiciels d'analyse de données textuelles relèvent du raisonnement statistique. Ils vantent notamment les possibilités de traitement massif qu'offrent les outils informatiques. Ainsi, à la question de savoir s'il faut utiliser ou non un logiciel, les promoteurs de l'analyse statistique des données

textuelles plaident pour le traitement d'un nombre sans cesse grandissant de textes. Le traitement statistique étant facilité par l'ordinateur, un deuxième argument porte sur la rapidité de l'analyse. Ainsi, les logiciels permettraient de traiter plus de matériau qu'une recherche ne faisant appel à aucun logiciel. Le traitement serait accéléré, voire instantané. En outre, puisqu'un automate n'est sujet ni à l'humeur ni au jugement, les résultats obtenus seraient plus scientifiques. Ces questions de quantité sont effectivement légitimes pour qui entend identifier les structures récurrentes d'un large corpus de textes, y compter le nombre d'occurrences d'un terme ou recenser les différentes formes d'un même mot.

Ces arguments n'ont, en fait, pas grand-chose à voir avec la recherche qualitative. Le nombre de textes, leur longueur et la vitesse de leur traitement ne constituent pas des questions pertinentes pour la recherche qualitative. De fait, celle-ci ne s'intéresse pas aux textes pour eux-mêmes. Lorsqu'un matériau est textuel, il s'agit soit d'observations, prises en note par le chercheur, ou de transcriptions d'entretiens conduits par le chercheur. La longueur de ces textes et l'éventuelle récurrence de certains mots sont secondaires par rapport au phénomène qui est étudié. Or, les recherches qualitatives en santé étudient des phénomènes qui relèvent de la perception, du vécu, de l'informel (Becker, 2002). Ces phénomènes diffèrent de ceux étudiés en analyse de discours : l'analyse de discours s'intéresse effectivement à la façon dont les discours sont produits, aux structures qui caractérisent les différents types de discours, voire au vocabulaire qui est mobilisé dans tel ou tel contexte. En recherche qualitative par contre, les mots, le vocabulaire et la construction des phrases importent moins que ce qu'ils signifient pour les acteurs qui les émettent. En conséquence, si les arguments quantitatifs (plus de textes, de vitesse, d'objectivité...) peuvent s'avérer pertinents pour l'analyse de textes et l'analyse de discours, ils concernent moins la recherche qualitative.

Dès lors, pourquoi utiliser un logiciel en recherche qualitative ?

Les raisons sont, en fait, d'un tout autre ordre. Utiliser un logiciel d'étiquetage réflexif peut tout d'abord aider le chercheur à rassembler en un même lieu son « laboratoire », des notes d'observation, des transcriptions d'entretiens, des réflexions, des guides d'entretien ou d'observation, des croquis ou des plans des lieux dans lesquels il s'est rendu, des ébauches de rédaction qui pourraient être utiles dans un article, un mémoire ou une dissertation doctorale, des références théoriques... Le logiciel occupe alors la fonction très pragmatique du journal de bord, du dossier ou du classeur rassemblant en un même lieu toutes les ressources relatives à une seule et même recherche.

Rassembler ces documents aux statuts très divers au sein d'un même programme évite de passer un temps précieux à courir de la cave au grenier pour retrouver le dernier article lu ou une formulation lumineuse apparue en entretien. S'agissant de documents numériques, le logiciel permet également au chercheur en déplacement d'emporter son laboratoire avec lui, sans



pour autant alourdir ses valises. Si cette fonction importe d'un point de vue pratique, elle ne constitue pas pour autant une justification épistémologique satisfaisante.

De fait, des raisons plus profondes motivent les chercheurs à recourir à un logiciel d'étiquetage réflexif. En centralisant des documents hétérogènes, le logiciel facilite des parcours de lecture originaux, des croisements et des recoupements qui contribuent à structurer l'analyse. Cette aide informatique ne modifie cependant pas la nature des opérations intellectuelles effectuées. Ces dernières restent celles posées sur papier. En fait, ceux qui mobilisent un logiciel souhaitent être en mesure d'explicitier les démarches accomplies durant la collecte et l'analyse de leurs matériaux. L'informatique rend « traçable » l'ensemble de leur démarche. Le logiciel rencontre ici la fonction épistémologique du journal de bord. Il n'est pas simplement un lieu de mémoire et de rassemblement, mais joue le rôle décisif de validation des recherches qualitatives. L'outil permet au chercheur de montrer et de démontrer l'ensemble du processus de recherche, ses inflexions et les justifications de celles-ci.

Certains logiciels d'analyse qualitative autorisent en outre de nouvelles façons de travailler. Par exemple, une plate-forme informatique permet à une équipe d'analyser collectivement un matériau partagé ou de tenir un journal de recherche commun. Il serait ardu de conduire ce type de démarches sur papier. L'informatique rassemble les conditions de possibilité de certaines formes de « collaboratoires », ces lieux de collaboration entre chercheurs. En définitive, les logiciels ne permettent pas de faire plus ni mieux, ils permettent de faire d'autres choses (Bénel, 2014 : 114-115). En tant qu'outils, ils enrichissent donc les possibilités de la recherche.

C'est pour mémoriser, expliciter et collaborer que la recherche qualitative recourt à des logiciels. Mobilisées dans la recherche internationale publiée en anglais, ces raisons spécifiques sont rarement évoquées en francophonie. Les chercheurs engagés dans des recherches qualitatives reprennent plutôt l'argumentation développée par les promoteurs des logiciels d'analyse de textes et d'analyse de discours.

Cela produit un raisonnement paradoxal. D'un côté, le chercheur dit s'inscrire dans une approche ne s'intéressant pas à la fréquence. D'un autre côté, il compte des mots, interprète des fréquences hautes ou basses et utilise une panoplie d'outils statistiques. La raison est, comme je l'ai mentionné plus haut, liée à la position « dominante » de l'analyse statistique des données textuelles et de l'analyse de discours en francophonie.

### 3. Quand ?

Après le pourquoi vient le moment. Quand décider d'utiliser un logiciel d'analyse qualitative ? Quand le choisir ? Quand l'utiliser ?

### Vignette 6. Au fil de l'interprétation

Psychologues de la santé, Hélène et Isabelle conduisent des entretiens auprès d'adolescents présentant des troubles alimentaires. Hélène rencontre des boulimiques et Isabelle des anorexiques. Elles me contactent pour me demander conseil dans le choix d'un logiciel d'analyse de textes. Lorsque je les rencontre, la recherche est déjà bien avancée : leurs questions de recherche respectives ont été mûrement réfléchies, les observations ont été conduites, les entretiens ont été transcrits. Ayant terminé la collecte, elles se posent la question de l'analyse et s'interrogent légitimement sur l'outil à mobiliser. ■

Je suis régulièrement consulté par des chercheurs comme Hélène et Isabelle. Ces interpellations me mettent souvent mal à l'aise. En effet, les logiciels d'étiquetage réflexif (les « CAQDAS ») se revendiquent systématiquement de l'analyse par théorisation ancrée. Or, en analyse par théorisation ancrée, la collecte, l'analyse et l'écriture se réalisent de concert. Le chercheur rédige tout d'abord sa question de recherche puis son premier guide d'entretien. Il conduit un premier entretien, le transcrit et l'analyse dans la foulée. Les résultats de cette première analyse conditionnent la formulation du deuxième guide d'entretien, qui sera donc très différent du premier. Ainsi, de manière itérative, le chercheur reformule très régulièrement sa question de recherche ; celle-ci s'affine, se précise. Chaque entretien est conduit avec un guide d'entretien différent et porte sur un phénomène de plus en plus précis, ce qui constitue la « productivité » de la recherche qualitative (Lofland et Lofland, 1984 : 132).

Un logiciel d'étiquetage réflexif convient particulièrement bien à cette manière de procéder. Comme l'ethnographe des maisons de naissance (vignette 5), le chercheur rédige régulièrement des notes, des mémos, des comptes-rendus ; le logiciel accompagne cette « analyse en mode écriture » (Paillé et Mucchielli, 2016 : 183-205). Il est délicat de conseiller l'outil adéquat à des chercheurs qui ont déjà conduit un grand nombre d'entretiens. En effet, les logiciels d'étiquetage réflexif ne sont pas censés être utilisés dans un cadre méthodologique qui conçoit la collecte et l'analyse comme des étapes séparées. Cette tension met en évidence le problème du « quand » ! La question du choix non seulement du logiciel, mais aussi de la méthode peut difficilement se faire en cours de route. Il convient de réfléchir dans les tout premiers moments d'une recherche au type de méthode qui convient à la question de recherche que l'on entend explorer et au type d'outils susceptibles d'assister cette démarche méthodologique. De fait, certaines méthodes séparent les différentes activités en étapes : d'abord, la problématisation ; ensuite, la collecte ; enfin, l'analyse, l'interprétation et l'écriture. Comme déjà évoqué à propos de l'objectivité des machines, c'est notamment le cas de l'analyse de contenu qui mobilise, dans l'analyse de

matériau textuel, un protocole strict (la fidélité inter-codeurs) et qui sépare les différentes étapes, devant être réalisées successivement. De manière comparable, une analyse de discours peut être conduite en étapes successives, qu'elle recoure ou non à des outils statistiques au moment de l'analyse. Mais les logiciels d'étiquetage réflexif n'ont initialement pas été conçus pour une démarche séquentielle et linéaire.

Le choix gagne donc à être opéré tout au début de la recherche. Mais il ne peut être laissé au hasard. Chaque question de recherche écarte certaines méthodes et en favorise d'autres. Une question compréhensive, visant un phénomène implicite, informel, tacite, se prête plus à une analyse par théorisation ancrée qu'à une analyse de contenu ou de discours. À l'inverse, une question portant sur la réputation ou la propagande se prête mieux à une analyse de contenu puisque les phénomènes respectifs de la réputation et de la propagande comportent une dimension intrinsèquement quantitative : la propagande fonctionne d'autant mieux qu'un message se diffuse de manière répétée dans l'opinion publique ; la réputation dépend de la répétition d'une série d'actes ou de paroles accessibles à une audience donnée. Il importe donc, au moment de formuler la question de recherche, de s'interroger à la fois sur la tradition méthodologique et subsidiairement sur le type d'outil informatique susceptible d'assister la recherche en question. Si, dès lors, les décisions se prennent au début, le troisième « quand » (quand utiliser le logiciel) trouve aisément sa réponse.

Ainsi, dans une recherche par théorisation ancrée, le logiciel mérite d'être utilisé de bout en bout, de la formulation de la question de recherche à la rédaction du rapport final ou de la dissertation doctorale, en passant par la collecte et l'analyse des matériaux.

Par contre, dans une démarche séquentielle, différents logiciels pourront être impliqués à chacune de ces étapes. Un logiciel pourra assister la transcription (Alber, 2010) ; un autre sera dédié au traitement des données collectées. Dans ce cas de figure, certaines décisions peuvent être tranchées en cours de route. Plusieurs techniques qualitatives s'accommodent d'une démarche linéaire. Leur choix peut donc s'opérer au moment d'analyser. Rien n'empêche non plus de détourner un logiciel de l'usage prescrit par ses concepteurs. Les logiciels d'étiquetage réflexif conviennent ainsi autant à l'analyse phénoménologique interprétative qu'à une recherche par théorisation ancrée. Changer de technique en cours de route n'est cependant possible que dans une démarche linéaire. Lorsque la collecte est clôturée, il est impossible d'opter pour une analyse par théorisation ancrée. Autrement dit, les possibilités d'intervention ou de détournement ne modifient pas la préconisation précédente : puisqu'elles limitent les choix suivants, la question de recherche et la méthode doivent être définies dès le début de l'investigation.

## 4. Que choisir ?

Finalement, la question qui taraude les chercheurs (donc les utilisateurs potentiels de logiciels), « quel logiciel utiliser », devient beaucoup plus simple une fois que l'on a répondu aux questions précédentes, à savoir pourquoi et quand recourir à un logiciel. Si plusieurs dizaines de logiciels sont disponibles, la plupart d'entre eux se proposent d'outiller une méthode particulière. Avoir opté pour une méthode facilite donc le choix en ce qu'elle le restreint. Le choix répond alors à des critères méthodologiques, épistémologiques ou scientifiques, ce qui est toujours préférable.

Quoi qu'il en soit, même ainsi restreint, le choix persiste. En effet, plusieurs logiciels se proposent d'assister chacune des méthodes déjà citées. Par exemple, plusieurs sociétés commerciales proposent une solution informatique pour assister une recherche par théorisation ancrée. *NVivo*, *Atlas.ti*, *MAXQDA* et *HyperRESEARCH* constituent des logiciels d'étiquetage réflexif courants. En face, quelques challengers comme *WeftQDA* ou *TamsAnalyser* offrent aux chercheurs des solutions moins chères ou une option politique particulière (celle des logiciels libres). D'autres encore (*Transana* ou *Sonar*) se spécialisent dans un certain type de matériau (vidéo, audio).

À ce stade, il importe de mettre de côté les logiciels qui ne conviennent pas au type de matériau ou à la question de recherche en cours. Même ainsi, des choix subsistent. Dans ce cas de figure, opérer un choix revient moins à savoir quel est le « meilleur » outil qu'à identifier l'outil qui permettra d'avancer efficacement. Chaque chercheur décidera des priorités qui sont les siennes. À ce moment, il ne faut pas négliger la disponibilité d'un collègue maîtrisant déjà l'un des outils entre lesquels on hésite. Ce dernier pourra en effet conseiller ou aider le novice qui apprend à utiliser le logiciel ou qui bute sur une difficulté. Les logiciels libres ou gratuits présentent évidemment un certain attrait, économique ou philosophique. Toutefois, il convient de peser le pour et le contre. Je rencontre régulièrement des utilisateurs ayant arrêté une solution gratuite, mais ayant perdu beaucoup de temps soit à se former à l'outil soit à en contourner des limitations. Finalement, déboursier quelques deniers leur aurait peut-être permis de gagner de précieuses semaines. S'il s'avère économique, leur calcul sous-estime le prix de leur temps.

La pire des extrémités consiste sans doute à développer soi-même un logiciel existant déjà par ailleurs. Si, personnellement, je me suis lancé dans le développement d'un logiciel libre d'analyse qualitative (*Cassandra*), c'était pour outiller un type de démarche différant de celles qu'assistent les logiciels existants. Il ne s'agissait pas de « réinventer la roue ». D'autres développeurs de logiciels (libres ou non) témoignent en ce sens et encouragent les chercheurs envisageant de s'engager dans cette voie à bien y réfléchir auparavant (Fenton, 2008). Il y a toujours mieux à faire que de recréer quelque chose qui est déjà disponible.

En définitive, en effet, le chercheur en sciences humaines sera rarement évalué sur la production d'outils informatiques ou même sur son usage d'outils existants, mais bien plus vraisemblablement sur sa contribution à sa discipline scientifique, qu'il s'agisse du champ de la santé, de la psychologie, de la géographie humaine, de l'anthropologie, de la sociologie ou de toute autre science humaine.

## RÉFÉRENCES BIBLIOGRAPHIQUES

ALBER Alex (2010), « Voir le son : réflexions sur le traitement des entretiens enregistrés dans le logiciel Sonal », *Socio-logos. Revue de l'association française de sociologie*, 5.

BECKER Howard Saul (2002), *Les ficelles du métier. Comment conduire sa recherche en sciences sociales*, Paris, La Découverte.

BÉNEL Aurélien (2014), « Quelle interdisciplinarité pour les “humanités numériques” ? », *Cahiers du numérique*, 10 (4), p. 103-132.

BOURDIEU Pierre, CHAMBOREDON Jean-Claude et PASSERON Jean-Claude (1968), *Le métier de sociologue*, Paris, Mouton/Bordas.

FENTON Alex (2008), “Open Source Development as a Stimulus for Innovation in Qualitative Data Analysis Software”, *AnaLogiQual'2008. Software for Qualitative Analysis : Social and Technical Innovations*, Luxembourg.

GLASER Barney (1992), *Emergence vs Forcing. Basics of Grounded Theory Analysis*, The Sociology Press, California.

KRIPPENDORFF Klaus (2004), *Content Analysis. An introduction to Its Methodology*, Thousand Oaks, Sage Publications.

LEJEUNE Christophe (2010), « Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative », *Recherches Qualitatives*, 9, p. 15-32.

LEJEUNE Christophe (2014), *Manuel d'analyse qualitative. Analyser sans compter ni classer*, Louvain-La-Neuve, De Boeck.

LEJEUNE Christophe (2017), « Analyser les contenus, les discours ou les vécus ? À chaque méthode ses logiciels ! » in Marie Santiago-Delefosse et Maria del Rio Carral (dir.), *Manuel pratique de méthodes qualitatives en psychologie et sciences humaines*, Paris, Dunod.

LOFLAND John et LOFLAND L. (1984), *Analyzing Social Setting. A Guide to Qualitative Observation and Analysis*, Belmont, Wadsworth.

PAILLÉ Pierre (1994), « L'analyse par théorisation ancrée », *Cahiers de recherche sociologique*, n° 23, p. 147-181.

PAILLÉ Pierre (2006), « Lumières et flammes autour de ma petite histoire de la recherche qualitative », *Recherches Qualitatives*, 26 (1), p. 139-153.

PAILLÉ Pierre et MUCCHIELLI Alex (2016), *L'analyse qualitative en sciences humaines et sociales*, Paris, Armand Colin.

STONE Philip, DUNPHY Dexter, SMITH Marshall et OGILVIE Daniel (1966), *The General Inquirer : A Computer Approach to Content Analysis*, Cambridge, MIT Press.