

Des logiciels en recherche qualitative ?

Pièges, limites et questions liminaires

Christophe Lejeune

Ce chapitre est dédié à la mémoire de Mélodie Dieudonné et d'Evence Godfroid.

Introduction

La généralisation de l'informatique et la multiplication des données numériques ont banalisé l'usage de logiciels d'analyse de données qualitatives (Millet, Millerand, Myles, Latzko-Toth et 2021)¹. 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 choix². Ce chapitre invite le lecteur à se poser les bonnes questions (et à se les poser tôt) avant d'ouvrir une application pour entamer des tests.

Vignettes 1, 2 et 3. Lever la main de la souris

Dans la première édition du présent ouvrage, ce chapitre s'ouvrait sur une vignette relatant la réticence de trois doctorants à se former aux logiciels

1. Je tiens à remercier tous les étudiants et les chercheurs qui m'interrogent sur l'analyse qualitative. Ce chapitre leur doit beaucoup. Je tiens également à remercier Joëlle Kivits et Myriam Winance de m'avoir invité à écrire ce chapitre ainsi que d'en avoir accompagné la maturation et la mise à jour avec patience et bienveillance. Je suis également reconnaissant à Laetitia Godfroid, Beatrice Scholtes et Thomas Vangeebergen, qui en ont commenté une version antérieure. Je suis évidemment seul responsable des erreurs ou des imperfections qui y subsisteraient.

2. Les termes « analyste », « chercheur », « étudiant » et « scientifique » renvoient aux personnes qui conduisent des recherches qualitatives, indépendamment de leur sexe, de leur genre et de leur orientation sexuelle ou affective.

d'analyse de données qualitatives. Cette vignette a suscité la circonspection de plusieurs lecteurs, dont Ariane qui m'a indiqué que, de nos jours, les cursus en santé intègrent la formation à ces logiciels et que, même lorsque ce n'est pas le cas, on s'y forme par soi-même.

Bénédicte, quant à elle, m'a sollicité pour former les membres de son centre de recherche à l'usage d'un de ces logiciels, en particulier. Lorsque je lui propose de débiter mon intervention par un rappel méthodologique, elle m'objecte que, étant composée de professionnels, son équipe n'en a guère besoin.

Un autre jour, l'initiative d'une faculté voisine retient mon attention. Il s'agit d'une invitation à rencontrer une doctorante qui témoignera de son usage d'un logiciel de ce type. Je contacte Céline, qui a diffusé cette annonce, pour lui proposer ma contribution. Céline préfère éviter ma venue, car les ateliers sont appréciés pour leur dimension pratique. Selon elle, la présence d'un chercheur senior risquerait de brider les échanges.

Si leurs démarches diffèrent, les positions d'Ariane, de Bénédicte et de Céline concernent toutes la formation à un logiciel. Pour des raisons distinctes, toutes les trois privilégient l'usage d'un logiciel à la réflexion méthodologique. On peut se demander si leur posture ne confond pas l'outil (le logiciel) avec la méthode. En se bornant à évoquer le logiciel utilisé, la section « Méthode » de certains projets de recherche témoigne de l'existence d'une telle assimilation (Paillé, 2006). Gageons que cela ne soit pas leur cas.

Ariane, Bénédicte et Céline partagent vraisemblablement un présupposé défavorable envers la méthode. La méthode peut en effet être envisagée comme contraignante. En ce sens, elle empêcherait d'avancer. Dans le même ordre d'idée, la réflexion méthodologique est alors perçue comme rébarbative et superflue, ce qui incite à la postposer. Or, remettre à plus tard les lectures ou les réflexions méthodologiques peut amener à identifier *a posteriori* que l'on n'a pas respecté ses principes. Procrastinée, la méthode devient alors culpabilisante. Pourtant, la méthode scientifique relève moins de la conformité à des normes immuables que de la réflexion sur sa propre démarche. Ce chapitre entend inviter le lecteur à interroger ses pratiques, à opérer des choix réfléchis et à se poser les questions qu'évacue trop souvent le recours à l'informatique.

Les questions qu'il faudrait se poser avant d'utiliser un logiciel sont relativement nombreuses.

- 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.
- 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 ?

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

1. Les logiciels sont-ils nécessaires ?

Aujourd'hui, mobiliser un logiciel dans le cadre d'une recherche qualitative dans le champ de la santé semble aller de soi. Pourtant, cela n'a 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.

Vignette 4. Tribunal de la preuve

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

Je réponds à Daniel 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 Daniel 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 Daniel. Dès lors, que fera-t-il, à quelques mois de la soutenance ? Va-t-il, pour cette raison, invalider l'ensemble de son travail, auquel il 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 Daniel 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'un impétrant 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 Daniel les accepte tacitement en me contactant, ces présupposés sont toutefois loin de faire l'unanimité. Examinons le raisonnement que suggère le membre de son 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 du chercheur sont nécessaires à des conclusions scientifiques. Acter la différence entre le chercheur 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, p. 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 codeurs³. 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 machine⁴. Les analystes de contenu ne soumettent donc pas l'interprétation à la machine, bien au contraire.

3. 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).

4. Ce principe est affirmé dès le début de la préface du premier livre dédié à un logiciel d'analyse de contenu (Stone *et al.*, 1966, p. xi).

Le membre du comité de thèse de Daniel 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 Daniel 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. La résurgence de l'intelligence artificielle, ces dernières années, renforce certainement une telle croyance. Pourtant, alors même que les promoteurs de l'intelligence artificielle admettent que leurs algorithmes ne sont pas neutres, persiste cette idée selon laquelle le traitement opéré par une machine serait plus fiable que l'interprétation humaine. D'ailleurs, en me contactant, Daniel y adhère sans doute partiellement. Comme lui, 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 Daniel, l'interrogation, voire l'inquiétude de ces collègues trahit une croyance, répandue, selon laquelle 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 5. Génération spontanée

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

Éric souffre d'un syndrome que je qualifie de codomanie (Lejeune, 2019, p. 87). Dans une acception un peu particulière de l'analyse par théorisation ancrée, il considère que coder produit nécessairement de la théorie. Selon cette acception, la théorie apparaîtrait sans investissement intellectuel du chercheur, presque automatiquement, par « génération spontanée ». Or, une analyse ou une théorie n'émerge jamais d'elle-même (Strauss et Corbin, 2004, p. 179). Seule la construction patiente, délibérée et active de l'utilisateur du logiciel la produit.

Éric n'est évidemment pas seul à croire qu'une théorie émergerait via l'accumulation d'opérations non interprétatives. Au sein même de la recherche qualitative, cette conception a été encouragée par le « mantra de l'émergence » (Bryant, 2017). Celui-ci s'appuie sur les écrits de Barney Glaser (1992) qui opposent les démarches hypothético-déductives (qui « forcent » les données) aux approches dites « inductives » qui encouragent « l'émergence » de nouvelles théories. Ni Barney Glaser ni les autres spécialistes de la recherche qualitative n'adhèrent à ce « mantra » mais celui-ci continue de circuler. Aujourd'hui, sa vitalité est sans doute renforcée par les arguments développés autour des *big data*. Bien qu'ils soient étrangers à la recherche qualitative, ceux-ci reconduisent l'argument de l'émergence (Bryant, 2019, p. 54). Les promoteurs des *big data* propagent cette idée que l'agrégation de données peut produire quelque chose d'intéressant sans action délibérée de la part de l'humain. S'il peut se justifier dans le domaine industriel, cet argument est incompatible avec la rigueur scientifique et la réflexivité attendue en recherche qualitative.

2.2. Les théories aveuglent-elles ?

Vignette 6. Se passer de 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 de fouille de texte (*text mining*) 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 va cependant 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 prémisses 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 de fouille 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 risque de le faire disparaître, à terme, 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 ses faiblesses. 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 7. Compter ou dénombrer

Philosophe et biologiste, Gaëlle et Hélène conduisent une recherche sur le rapport aux antibiotiques. Formées aux méthodes qualitatives, elles réalisent une vingtaine d'entrevues et rédigent un article scientifique qu'elles soumettent à leur directrice de recherche. Habitée aux canons du secteur de la santé, cette dernière leur intime d'y préciser le nombre de personnes qui s'expriment sur tel ou tel sujet. L'article est modifié et soumis à une revue scientifique qui en refuse la publication au motif qu'une recherche qualitative devrait précisément exclure les comptages.

Le contexte scientifique de la francophonie diffère de celui des recherches internationales publiées en anglais. La littérature anglophone désigne les logiciels d'analyse de données qualitatives par le sigle QDA ou CAQDAS (*Computer Assisted Qualitative Data Analysis*). Ce sigle renvoie à une série d'outils relativement comparables (comme *NVivo*, *Atlas.ti* et *MAXQDA*) qui souscrivent

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 prolixie à propos des logiciels d'analyse de textes (comme *IRaMuTeQ* et *TXM*). En première lecture, les données textuelles relèvent des données qualitatives. La différence peut dès lors 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'entrevues conduites 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, les arguments quantitatifs (plus de textes, de vitesse, d'objectivité...) concernent moins la recherche qualitative que l'analyse de textes et de discours, la fouille de texte ou le *text mining*, l'intelligence artificielle et les *big data*.

Dès lors, si ce n'est pas pour des raisons de quantité, pourquoi utiliser un logiciel en recherche qualitative ?

Vignette 8. Documenter et expliciter

Sage-femme expérimentée, Isabelle 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'étiquetage réflexif. Elle présente ses premières conclusions lors d'une journée d'étude en santé. Vraisemblablement mal à l'aise, certains membres de l'assistance lui demandent si ses conclusions sont scientifiques et représentatives ou subjectives et biaisées. En guise de réponse, elle expose patiemment et pédagogiquement les circonstances l'ayant amenée à construire ses conclusions, en montrant, pour chacune d'entre elles, les réflexions consignées par ses soins dans le logiciel.

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'entrevues, 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 entrevue. 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.

Des raisons plus profondes motivent les chercheurs à recourir à un logiciel d'étiquetage réflexif. Avec cette centralisation de 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. C'est notamment ici qu'intervient la différence décisive entre un recours dénué de fondement méthodologique et un usage instruit par la méthode. Les logiciels d'étiquetage réflexif peuvent être utilisés pour procéder à une analyse de données qualitatives. Mais le recours à ces logiciels ne suffit pas à garantir la scientificité de la démarche. Pour qu'un tel usage soit scientifique, il convient qu'il s'adosse à une méthode rigoureuse

(comme l'analyse de contenu ou de discours, l'analyse phénoménologie interprétative ou par théorisation ancrée).

Par ailleurs, toute analyse de données qualitatives n'est pas nécessairement une analyse qualitative. La nature qualitative des données ne suffit pas à garantir que leur analyse soit qualitative. Les principes de l'analyse de contenu classique la rendent rigoureuse et scientifique mais la positionnent clairement en dehors des analyses qualitatives. Pour que son analyse soit qualitative, le chercheur doit exclure toute quantification, exercer sa réflexivité à chaque étape et être en mesure d'explicitier les démarches accomplies durant la collecte et l'analyse de ses matériaux. L'informatique peut l'y aider, en rendant « traçable » l'ensemble de la 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, p. 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 aux comptages. 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 9. Au fil de l'interprétation

Psychologues de la santé, Julie et Karine conduisent des entretiens auprès d'adolescents présentant des troubles alimentaires. Julie rencontre des jeunes souffrant de boulimie et Karine des jeunes souffrant d'anorexie. 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 : après une importante revue de la littérature, 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 Julie et Karine. 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 une première entrevue, la 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 entrevue est conduite 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, p. 132).

Un logiciel d'étiquetage réflexif convient particulièrement bien à cette manière de procéder. Comme l'ethnographe des maisons de naissance (vignette 8), le chercheur rédige régulièrement des notes, des mémos, des comptes rendus ; le logiciel accompagne cette « analyse en mode écriture » (Paillé et Muchielli, 2021). 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. Par contre, les logiciels d'étiquetage réflexif n'ont initialement pas été conçus pour une démarche séquentielle et linéaire.

Le choix gagne donc à être opéré tôt, tout au début de la recherche voire avant que celle-ci ne commence. Décider en amont permet de se familiariser avec le logiciel, de s'habituer à son fonctionnement voire de s'y former avant que surviennent les enjeux décisifs de l'analyse proprement dite, ce qui est recommandé (Friese, 2019, p. 310). Si arrêter sa décision suffisamment tôt permet d'échelonner son apprentissage, le choix de l'outil dépend aussi de considérations méthodologiques, dépassant la logistique et l'apprentissage ; il ne peut donc 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. Soumise à un agenda serré, la recherche de Julie et Karine a progressé en supposant que ces questions pouvaient être postposées. Les conseiller à ce stade du processus est périlleux.

Mieux vaut arrêter ses décisions au début. Il devient alors plus facile de répondre à la question : quand utiliser le logiciel ?

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* et *MAXQDA* constituent des logiciels d'étiquetage réflexif courants. En face, quelques challengers comme *TamsAnalyser*, *QualCoder* ou *RQDA* offrent à la fois des solutions moins chères et une option politique particulière (celle des logiciels libres). D'autres (*Transana* ou *Sonar*) se spécialisent dans un certain type de matériau (vidéo, audio).

Plusieurs « collaboratoires » facilitent le travail en équipe : *Dedoose*, *Quirkos* et *Saturate* sont proposés par des entreprises privées. Pour sa part, *Cassandra* est développé en logiciels libres et est proposé par l'université de Liège, en tant que service à la collectivité. S'appuyant sur le *cloud*, ces plateformes présentent les avantages et les inconvénients de la décentralisation. Avant de déposer son matériau empirique (comprenant des données personnelles) sur un serveur, il convient de s'interroger sur le respect des mesures de protection et de confidentialité en vigueur, en particulier s'il s'agit d'une entreprise privée⁵. Il est aussi prudent de s'assurer que l'ensemble de son projet de recherche ne risque pas de disparaître si la société en question (ou un de ses sous-traitants) rencontre un problème technique (comme cela est arrivé à *Dedoose* en mai 2014).

Certains chercheurs recourent également à des outils de statistique textuelle déjà évoqués, soit via des fonctionnalités intégrées à des logiciels d'étiquetage réflexif, soit via des logiciels libres spécifiques comme *IRaMuTeQ* ou *TXM*. Bien qu'*a priori* non adaptés à l'analyse qualitative, ces outils peuvent assister une stratégie de parcours, d'exploration et d'approfondissement du matériau disponible. Un tel recours rencontre évidemment des besoins opératoires et remplit une fonction heuristique. Dans une démarche qualitative,

5. En Europe, le règlement général sur la protection des données à caractère personnel (RGPD) définit les normes minimales à respecter. Son application ne dispense évidemment pas d'une réflexion déontologique et éthique sur ce qui est collecté, conservé et transformé.

un tel usage appelle cependant la prudence. D'une part, il faut se rappeler que les mots et le langage importent moins que ce qu'ils expriment. D'autre part, il convient d'éviter avec précaution d'importer un raisonnement statistique dans une approche qui le proscrit. Loin de consolider les résultats d'une recherche qualitative, les quantifications la déforcent (Gaëlle et Hélène en ont fait les frais, dans la vignette 7).

Quelle que soit la stratégie adoptée, 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 la personne 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 créer soi-même un logiciel existant déjà par ailleurs. Si, personnellement, je me suis lancé dans le développement du logiciel libre *Cassandra* (Lejeune, 2016), c'était pour outiller un type de démarche particulière (autour du journal de bord), qui diffère de celles qu'assistent les logiciels existants (partant du codage du matériau). 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.

Une autre leçon de ce chapitre, c'est qu'il vaut mieux poser ces choix en amont de sa recherche. Mais qu'arrive-t-il si ces questions sont abordées plus tard, par exemple, en lisant le présent chapitre pour une recherche en cours ? S'interroger en cours de route peut rassurer si la réponse confirme les choix déjà posés. Dans le cas contraire, le questionnement peut s'avérer anxiogène ou culpabilisant. Dans ce type de situations, il convient de garder à l'esprit que toute recherche est itérative et réflexive. Il est toujours possible et admis de mesurer les écarts entre une intention initiale et les moyens mis en œuvre, de se raviser et d'adapter sa stratégie. Le récit de ce genre de réorientation a sa place dans la section ou le chapitre méthodologique de la future publication. En rendre compte avec transparence s'inscrit pleinement dans la perspective qualitative.

En définitive, une recherche qualitative n'est pas évaluée en fonction des outils informatiques mobilisés voire produits, 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. Ces résultats dépendent moins du logiciel que de la réflexion méthodologique de l'analyste, que les questions soulevées dans ce chapitre entendent stimuler et instruire. Au point de départ comme à l'arrivée de ce parcours, cette réflexion implique de s'interroger sur sa question de recherche. Les méthodologues ne le répéteront jamais assez ! Une bonne analyse dépend de la clarté de la question de recherche et de la capacité du chercheur à s'interroger, tout au long de ses investigations, sur sa pertinence et sur l'adéquation des moyens (notamment informatiques) avec ses fins.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ALBER A. (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 H. S. (2002), *Les ficelles du métier. Comment conduire sa recherche en sciences sociales*, Paris, La Découverte.
- BÉNEL A. (2014), « Quelle interdisciplinarité pour les "humanités numériques" ? », *Cahiers du numérique*, 10 (4), p. 103-132.
- BOURDIEU P., CHAMBOREDON J.-C., PASSERON J.-C. (1968), *Le métier de sociologue*, Paris, Mouton/Bordas.
- BRYANT A. (2017), *Grounded Theory and Grounded Theorizing. Pragmatism In Research Practice*, New York, Oxford University Press.
- BRYANT A. (2019), *The Varieties of Grounded Theory*, London, Sage Publications.
- FENTON A. (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.
- FRIESE S. (2019), "Grounded Theory Analysis and CAQDAS: A Happy Pairing or Remodeling GT to QDA?" in A. Bryant, K. Charmaz (ed.), *Current Developments in Grounded Theory*, London, Sage Publications.
- GLASER B. (1992), *Emergence vs Forcing. Basics of Grounded Theory Analysis*, The Sociology Press, California.
- KRIPPENDORFF K. (2004), *Content Analysis. An introduction to Its Methodology*, Thousand Oaks, Sage Publications.
- LEJEUNE C. (2010), « Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative », *Recherches qualitatives*, Hors Série 9, p. 15-32.
- LEJEUNE C. (2016). « Le blog de recherche comme journal de bord informatique. Un soutien à la réflexivité, à l'analyse, à la communication et à la scientificité ? » *Recherches qualitatives*, Hors Série 20, p. 402-415.
- LEJEUNE C. (2017), « Analyser les contenus, les discours ou les vécus ? À chaque méthode ses logiciels ! » in M. Santiago-Delefosse, M. del Rio Carral (dir.),

Manuel pratique de méthodes qualitatives en psychologie et sciences humaines, Paris, Dunod.

- LEJEUNE C. (2019), *Manuel d'analyse qualitative. Analyser sans compter ni classer*, Louvain-la-Neuve, De Boeck.
- LOFLAND J., LOFLAND L. (1984), *Analysing Social Setting. A Guide to Qualitative Observation and Analysis*, Belmont, Wadsworth.
- MILLETTE M., MILLERAND F., MYLES D., LATZKO-TOTH G. (2021), *Méthodes de recherche en contexte numérique : Une orientation qualitative*, Montréal, Les Presses de l'Université de Montréal.
- PAILLÉ P. (2006), « Lumières et flammes autour de ma petite histoire de la recherche qualitative », *Recherches qualitatives*, 26 (1), p. 139-153.
- PAILLÉ P., MUCCHIELLI A. (2021), *L'analyse qualitative en sciences humaines et sociales*, Paris, Armand Colin.
- STONE P., DUNPHY D., SMITH M., OGILVIE D. (1966), *The General Inquirer: A Computer Approach to Content Analysis*, Cambridge, MIT Press.
- STRAUSS A. L., CORBIN J. (2004), *Les fondements de la recherche qualitative. Techniques et procédures de développement de la théorie enracinée*, Fribourg, Academic Press Fribourg.