

Note de présentation à L'EDTSS

Introduction

Bonjour à tous. Je m'appelle Didier Fripiat et je suis actuellement assistant en sociologie à l'UCL. En tant que tel et comme beaucoup ici sans doute, je ne suis pas encore docteur. J'ai entamé ma dernière année de thèse en novembre ce qui me place aujourd'hui un peu en porte-à-faux : d'une part je n'ai pas l'impression que mon expérience générale de la recherche en sociologie soit suffisante pour donner des conseils méthodologiques mais d'autre part cela fait maintenant plus de 6 ans que j'expérimente une voie méthodologique relativement nouvelle. Qui plus est, la thématique que j'ai choisie de traiter dans ma thèse ouvre à elle seule toute une série d'interrogations comme je l'introduirai brièvement par la suite.

Mon objectif ici sera donc d'essayer de vous faire profiter de certaines réflexions qui me sont venues au fil de ces 6 années et je vous demanderais d'être indulgents si mon manque d'expérience m'amenait m'égarer quelque peu (que ce soit à propos du logiciel Nvivo ou d'autres choses). Je vous encourage d'ailleurs à me les faire remarquer si cela arrivait.

Entrée en matière

Comme le mentionne le titre de mon intervention, je me concentrerai sur Nvivo en tant que logiciel d'analyse de données qualitatives, même si à travers celui-ci ce sont bien les logiciels de QDA en général que je questionne. Nvivo est le logiciel que j'ai le plus employé ces dernières années, même si j'ai eu l'occasion de jeter un regard plus ou moins approfondi sur Atlas Ti, Modalisa ou MaxQDA.

Je ne vais pas m'attarder sur les spécificités de Nvivo étant donné que François Mélard l'a présenté bien mieux que ce que je pourrais faire. Tout au plus j'en détaillerai quelques fonctions.

Ce logiciel, je l'ai principalement utilisé dans le cadre de ma thèse, celle-ci portant sur les forums de discussion sur Internet. Cette thèse repose sur différents éléments qu'il m'a semblé intéressant de vous présenter dans un premier temps car ils sont en lien direct avec les perspectives méthodologiques que j'aimerais discuter avec vous aujourd'hui.

Investigation du Net

Tout d'abord : le boum Internet. Même si on ne s'intéresse pas spécialement au sujet, il est relativement difficile de ne pas se rendre compte que les communications médiatisées par ordinateur (autrement dit les CMO) ont littéralement explosé en l'espace de 10 ans.

Un nombre croissant d'auteurs s'intéresse au phénomène en l'abordant par parties : sites persos, blogs, chat, messenger, forums... Les moyens d'interaction présents sur le Net sont divers. Mais au-delà de cette diversité, on assiste surtout à une véritable cacophonie d'expression de soi. Plusieurs théories apportent un certain éclairage sur l'explosion de ces pratiques : individualisme, communautarisme, dilution de repères normatifs, reconstitution de groupe de pairs.... Si bien que petit à petit une base théorique tangible est en train d'émerger. Mais du point de vue de la méthodologie les choses sont beaucoup moins avancées. Or c'est précisément ce que nous allons aborder.

Mon propos se concentre sur la masse d'informations qui devient disponible et sur son traitement via des logiciels d'analyse de données qualitatives. Ces informations, que je reprends sous le terme de données digitales, présentent une double spécificité :

- A. Une accessibilité étendue
- B. Un archivage renforcé

Une accessibilité étendue d'abord en ce que chaque information produite et rendue publique est virtuellement consultable par le monde entier. Quiconque sachant chercher pourra tomber sur cette information. Mieux encore, les moteurs de recherches divers et variés et leur fonctionnalités de plus en plus abouties permettent à ceux qui s'en donnent la peine de développer de véritables stratégies de récolte d'information.

Ce référencement m'amène à la deuxième spécificité : l'archivage renforcé. Car le référencement et l'accessibilité ne seraient pas grand-chose s'il n'existait pas une certaine pérennité de l'information. Or, sur le net, l'information reste consultable telle qu'elle a été produite tant qu'on ne la modifie pas sciemment.

Ces deux éléments offrent ensemble au sociologue un matériau potentiel proprement incalculable. Autant en termes de récolte de l'information que de production de celle-ci par les médias du net.

Hélas, la ponctualité étant quelque chose de peu répandu chez les sociologues, d'autres avant nous ont senti l'aubaine. C'est pourquoi vous trouvez depuis plusieurs années beaucoup de spam, d'enquêtes marketing et autres tentatives de production d'information toutes aussi violentes les unes que les autres. Ainsi, au moment où les sociologues s'intéressent à la production d'information par CMO, d'autres en ont déjà tellement abusé d'une manière négative que le terrain en devient peu réceptif.

Mais laissons cela de côté pour le moment et intéressons nous aux données digitales. Face à cet océan d'expression de soi qu'est le net, le sociologue peut adopter essentiellement deux positions : investiguer les données existantes ou se servir des CMO pour en produire de nouvelles. La manière dont j'ai développé ma thèse m'a poussé à pratiquer successivement ces deux positions. Ce qui va me permettre de les aborder plus concrètement.

Investiguer des données existantes

Comme on l'a vu, la masse de matériau déjà disponible est proprement énorme et traite de quasiment tout et n'importe quoi : tout le monde ici a sans doute déjà été surpris de ce qu'il découvrait en surfant. Mais en plus d'être considérable et diversifiée, elle présente des avantages a priori tout à fait utiles pour le sociologue. Parmi les principaux avantages que je pointerai :

- L'invisibilité du chercheur : pour les données qui ont déjà été produites évidemment mais surtout pour celles en train de se produire car plusieurs media du net permettent à un simple observateur de n'avoir virtuellement aucun impact sur l'interaction tant sa présence passe inaperçue s'il ne s'en mêle pas.
- Un format digital : toutes les informations sont d'ores et déjà sous un format récupérable (avec plus ou moins de difficultés on le verra) sur l'ordinateur du chercheur. Aucune retranscription n'est donc nécessaire. Plus encore, l'interaction dans son entièreté est

sauvegardée telle qu'elle a eu lieu. C'est donc sur tous les détails de l'interaction que le chercheur peut revenir à loisir comme si elle apparaissait à chaque fois en direct devant lui.

Au vu de ces avantages, travailler sur le net semble donc vraiment alléchant pour le chercheur et ce à plus forte raison s'il utilise un logiciel de QDA puisqu'il lui suffirait de transférer le matériau digital vers le dit logiciel. Mais, car il y a toujours un mais, les problèmes techniques ne disparaissent pas pour autant. Premier exemple : un matériau de ma thèse a été le contenu de différents messages publiés sur un forum. Or il faut savoir que, actuellement, pour importer un matériau dans Nvivo (ou dans un autre logiciel de QDA) il faut que celui-ci soit le plus clair possible et soit surtout « mono-média » donc pas d'images dans le texte ou d'hyperliens... Pour information, voici un petit aperçu d'une page d'un forum sur lequel je travaille: Ceci n'est pas importable directement dans Nvivo à cause des images, de la mise en page et d'autres détails... ce contenu doit donc être retravaillé. Or, en le retravaillant pour garder le contenu textuel je perds tous les smiley qui sont évidemment une part importante de l'interaction. Les limitations en termes de mise en forme des logiciels de QDA obligent donc le chercheur à se séparer d'une partie parfois cruciale du matériau digital qui est, par essence, multimédia¹.

Deuxième problème, les sites internet et autres données digitales sont de plus en plus souvent fondés sur des bases de données et une répartition en pages dynamiques. Autrement dit, si je veux récupérer l'information, je dois passer chaque page en revue pour en copier le contenu et le coller dans un document et ensuite traiter ce document au complet. Sachant que sur certains forums francophones, des sujets dépassent allègrement les 2000 pages, le travail devient rapidement très fastidieux et peut prendre énormément de temps.

Bien entendu, plusieurs solutions peuvent être trouvées pour résoudre ces problèmes mais elles passent systématiquement par un travail intermédiaire. Pour ma part, cette solution a résidé dans l'utilisation de logiciels robots pour le problème des pages et de logiciels travaillant sur le texte par expressions régulières pour la réduction des médias présents dans la page. Ces logiciels font la jonction nécessaire entre le matériau brut et l'importation dans Nvivo qui, à cause de règles de mises en forme strictes ne permet pas de les traiter directement. Cette limitation de Nvivo ne porte donc pas trop à conséquence, mais la mise en place de ces logiciels intermédiaires nécessite un laps de temps variable qui ne peut-être négligé dans le timing du chercheur. La retranscription n'a donc pas vraiment disparu, elle a juste pris une nouvelle forme et cette nouvelle forme impose l'adaptation du chercheur et l'apprentissage de nouvelles techniques. Il serait donc faux de penser que parce que l'on désire traiter un matériau digital via un logiciel de QDA, le passage de l'un à l'autre va se faire sans heurt.

Produire des données

Deuxième position empruntable par le chercheur : la production d'informations sur le net. A ce niveau, d'autres avantages viennent titiller le sociologue :

- Réduction des coûts : produire de l'information sur le net semble largement plus économique que la plupart des méthodes de production d'information utilisées en

¹ A noter que ceci n'est plus vrai depuis la version 7 de Nvivo même si les possibilités offertes sont encore limitées (pas de codage partiel d'une photo)

sociologie, surtout si l'on tient compte des déplacements nécessaires aux méthodes « classiques ».

- Zone d'investigation élargie : avec une bonne stratégie de publicité de l'enquête c'est tout le net francophone qui s'offre au chercheur avec la possibilité de toucher certaines populations dispersées.
- Mise à distance du terrain : la production online offre le confort de garder le terrain à distance et mine de rien, cela rassure beaucoup de jeunes chercheurs que la confrontation au terrain angoisse fortement.

A nouveau, succomber sans réfléchir au chant des sirènes ne serait pas raisonnable. D'abord et pour raison principale que trop peu d'enquêtes se sont véritablement posé la question du statut de l'information récoltée par les médias du net. Les mails ressemblent à des lettres mais n'en sont pas, la visioconférence semble être du face à face mais n'en est pas... Or dans beaucoup d'enquêtes utilisant des données produites sur le net, l'auteur ne se pose pas la question de la différence et estime a priori que ça revient au même.

D'un point de vue plus pratique, Nvivo et les logiciels d'analyse qualitative accusent à nouveau certains manques lorsqu'il s'agit de faire la jonction avec les outils de production d'information online. Cette jonction nécessite à chaque fois de transformer le matériau pour le mettre dans un format spécifique, cela nous l'avons vu il y a quelques minutes. Mais surtout, les outils de collecte online proposent des fonctions très intéressantes qui, faute d'adaptation de la part des logiciels de QDA, ne peuvent pas être exploitées. Un exemple concret réside dans le mélange de données quantitatives et qualitatives. Les solutions de collecte online permettent de récolter un grand nombre de réponses à des questions ouvertes ce que l'on ne s'autorise habituellement pas dans des questionnaires papiers. Nvivo permet d'importer ces informations et de les traiter en masse mais celles-ci ne peuvent être directement intégrées qu'en les découplant du reste des données quantitatives, ce qui constitue naturellement une perte dont on peut s'attrister.

A cela on peut ajouter d'autres options limitées du logiciel quant au traitement des fréquences d'apparition de certains mots ou quant à la mise en place d'une chronologie dans le matériau. En gros, Nvivo souffre de fonctionner sur le principe de la base de données sans ouvrir ce fonctionnement au chercheur. Si Nvivo permettait d'accéder à la base de données qu'il crée cela neutraliserait bon nombre d'écueils et permettrait sans doute une meilleure jonction avec les solutions de production online via, par exemple l'importation de bases de données extérieures.

Le seul logiciel de QDA que je connaisse permettant en partie cela est Modalisa mais celui-ci souffre des défauts de ses avantages : d'une part il intègre directement l'outil de collecte d'information ce qui limite les possibilités et d'autre part il ne propose que peu de choix du point de vue de l'infrastructure mise en place (ici je parle des serveurs récoltant l'information, de la disponibilité du questionnaire sur le net...)

A ce stade, les logiciels de QDA ne semblent donc pas encore prendre en compte l'évolution qu'ont connu les outils online. C'est en partie normal étant donné que d'une part cette évolution est rapide et d'autre part, que les utilisateurs des logiciels de QDA ne semblent que commencer à investir le Net. Il n'en reste pas moins que de par leur conception, ces logiciels ne facilitent pas beaucoup le lien avec les outils de production de données digitales.

Gestion de la recherche

En dehors de ces considérations purement méthodologiques, une autre utilisation fort appréciable peut-être faite de Nvivo : celle d'un centre de gestion de recherche.

D'abord dans la gestion des différents matériaux que cette recherche rassemble. Nvivo propose une vision globale et claire de toutes les sources mais également de tous les cas rencontrés (cas auxquels on peut associer des caractéristiques que l'on juge pertinentes) ou encore de toutes les catégories indigènes et analytiques qui ont été créées.

Mais Nvivo peut également être utilisé pour la gestion bibliographique. Par exemple, dans le cadre de ma thèse j'y ai également intégré toutes sortes de fiches de lectures, d'ouvrages ou d'articles dont je disposais en version digitale. Ceci est particulièrement utile si comme moi vous êtes pourvus d'une mémoire exécrationnelle. En codant ces ressources comme on code un matériau je me suis composé une base de données permettant de retrouver rapidement les ouvrages qui traitent de telle théorie ou de tel concept.

Malheureusement, encore une fois, Nvivo n'étant pas directement prévu pour cela, on se retrouve confronté à certains problèmes dont par exemple la non-prise en charge des formats pdf. Il n'empêche qu'avec un peu d'imagination ce logiciel permet bien plus de choses que celles aperçues au premier abord.

Conclusion

Que faut-il en tirer comme conclusion ? Que les logiciels de QDA ne sont pas encore parfaitement adaptés à la recherche sociologique online ? Oui certainement. Qu'il faut donc éviter de les utiliser dans ce sens ? Non certainement pas. Même si mon intervention s'est focalisée sur différents inconvénients qu'il m'a fallu surmonter, à aucun moment je n'ai envisagé de ne plus utiliser Nvivo. Car les avantages restent très nombreux et François Mélard les a bien présentés. Parmi les plus cruciaux je mentionnerais celui-ci. Le recours à des logiciels de QDA impose plus que jamais au chercheur une démarche de catégorisation systématique.

Face au matériau et à son codage progressif, le chercheur ne peut plus faire l'impasse sur une réflexion approfondie quant aux catégories qu'il crée et utilise. Chaque nouveau passage significatif oblige le chercheur à se poser cette question : s'agit-il d'une nouvelle catégorie ou non ? Chaque codage ramène le chercheur à percevoir l'enjeu existant entre perte d'information et gain de compréhension. Il y a tout un équilibre à trouver sans quoi le chercheur risque soit de se noyer dans une infinité de codes proprement ingérable soit de ne garder que trop peu de codes et verser dans un biais totalisant. Cette approche demande un travail significatif de la part du chercheur mais est à mon sens garante d'un meilleur travail de catégorisation et donc d'un travail empirique plus pertinent.

Si une impression négative pouvait malgré tout sortir de mon intervention, il faudrait la tempérer par ceci : ces logiciels et les données digitales partagent le même support mais à ce stade, la recherche sociologique online est balbutiante. Je suis convaincu que dès que les sociologues ou autres chercheurs de sciences sociales multiplieront les recherches pertinentes, les logiciels s'adapteront à leurs besoins.

Ce qui m'amène à un dernier point que je voudrais aborder et qui touche plus à la recherche online qu'aux logiciels de QDA. Il est impératif que les chercheurs en sciences sociales ne fassent pas n'importe quoi lorsqu'ils abordent une recherche prenant place sur le net. Je l'ai dit auparavant, beaucoup de disciplines ont déjà investi les médias du net d'une manière négative. Au moment où nous, chercheurs en sciences sociales voulons entrer dans la danse, il est primordial que nous réfléchissions à la manière de nous y prendre. Ceci afin de créer une image qui se démarque de tous ceux qui y produisent des données de manière agressive. J'ai trop souvent vu des étudiants en sociologie vouloir profiter de l'aubaine en spammant des listes d'adresse mail, en bâclant leur propre présentation ou en ne se posant tout simplement pas la question de la spécificité du matériau ainsi récolté. Cela a pour conséquence de produire des travaux médiocres et de contribuer à constituer une mauvaise image de nos disciplines. Or, le stigmaté est quelque chose d'extrêmement puissant sur le net.

La recherche online offre de nouveaux horizons tout-à-fait fascinants pour le sociologue, il serait dommage de s'en priver en agissant de manière précipitée et irréfléchie.