

Dumez Hervé (2006) "L'analyse de données qualitatives assistée par ordinateur (ADQAO)", *Le Libellio d'Aegis*, volume 2, n° 3, automne, pp. 36-38

---

## Sommaire

**2**

What Makes People Tip Motivation and Predictions  
*D. Gambetta*

**10**

---

### DOSSIER SPECIAL HANNAH ARENDT

---

Essai sur la théorie de l'action de Hannah Arendt dans ses implications pour la recherche en science sociale

Essai sur la théorie morale de Hannah Arendt dans ses implications éventuelles pour l'éthique d'entreprise  
*H. Dumez*

---

**32**

Academy of Management 2006 :  
six jours dans la ville de John Pemberton et de Martin Luther King Jr.  
*C. Curchod*

**36**

Notes de séminaires  
*H. Dumez*

**40**

Programme des prochains séminaires AEGIS

Les autres articles de ce numéro & des numéros antérieurs sont téléchargeables à l'adresse :

<http://crg.polytechnique.fr/v2/aegis.html#libellio>

---

## L'analyse de données qualitatives assistée par ordinateur (ADQAO)

Le 18 mai 2006, le séminaire AEGIS accueillait Catherine Voynnet Fourboul, maître de conférences à l'Université Panthéon-Assas Paris II.

La première remarque est fondamentale : l'analyse de données qualitatives assistée par ordinateur se distingue très nettement de l'analyse de contenu. Les deux démarches sont très différentes. L'ADQAO consiste à essayer de trouver du sens par delà le simple traitement des mots.

### Présentation générale de la démarche

L'approche ADQAO est traitée en référence à trois courants : la « *grounded theory* », les méthodologies de démarche qualitative et l'assistance par ordinateur. Il serait dangereux de se préoccuper du volet technique (assistance par ordinateur) avant de penser la méthodologie. Donnons tout de suite les références fondamentales dans ces trois domaines.

Côté « *grounded theory* », pour le chercheur débutant, la grande référence reste : Strauss Anselm & Corbin Juliet (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage Publications, et aussi Strauss Anselm & Corbin Juliet [eds.] (1997) *Grounded Theory in Practice*. London. Sage Publications.

Côté recherche qualitative, voir Huberman Michael A. & Miles Matthew B. (2003) *Analyse des données qualitatives*. Louvain-la-neuve, de Boeck, ainsi que Flick Uwe (2002) *An introduction to qualitative research*. London, Sage, qui est fondamental.

Sur le plan de l'assistance par ordinateur, deux références Fielding Nigel G. & Lee Raymond M. (1998) *Computer analysis and qualitative research (new technologies for social research)*. London, Sage, et Kelle Udo [Ed] (1995) *Computer-Aided Qualitative Data Analysis - Theory, Methods and Practice*. London, Sage.

Quels sont les objectifs de l'ADQAO ?

1. Structurer des données et rendre compte du caractère « vivant » de la recherche de façon organisée.
2. Trouver des régularités, des logiques d'articulation entre concepts ou idées pour produire une synthèse.
3. Isoler, bâtir, définir des concepts in vivo, la proximité des idées conduisant à la conceptualisation.
4. Traiter à la fois les cas et les variables. C'est ce que permettent de faire les logiciels. L'approche par le cas permet d'être proche de la réalité du terrain. En même temps, le chercheur doit aller plus loin et passer du discours incarné à une expression de la réalité plus conceptuelle. Cette prise de recul passe plus facilement par un travail sur les variables.

Qu'apporte le logiciel par rapport aux méthodes manuelles ?

Il est parfaitement possible d'opérer manuellement et beaucoup de chercheurs le font. Mais le volume des données est tel qu'il est souvent difficile de tout coder et que l'on perd beaucoup de données dans le traitement à la main. Beaucoup de temps est également perdu en rapprochement de données identiques.

Le logiciel permet une plus grande proximité avec les données, d'autant que le travail de retranscription fait écho au codage assisté par l'ordinateur. Il offre une grande flexibilité : par exemple, les catégories peuvent être facilement fusionnées ou au contraire dégroupées. Surtout, il permet de gérer un plus grand nombre d'idées (les chercheurs américains considèrent qu'il ne faut pas dépasser 500 codes – ce qui est déjà énorme – mais on peut gérer de l'ordre de plusieurs centaines de codes). Un autre avantage est que l'on peut assez facilement passer du qualitatif au quantitatif. Les résultats sont quasi-instantanés. Enfin, les concepteurs de logiciels affirment que le travail de groupe autour des données est facilité. Mais dans la pratique se posent des questions d'ordre organisationnel qui placent cette ambition au rang du défi. On peut être un peu sceptique sur ce point, mais il s'agit sans doute d'une ligne de front pour le développement des logiciels d'autant que cet aspect constitue une attente réelle des chercheurs qui n'est pas suffisamment développée au plan de la conception des logiciels.

### **Le codage**

Le codage a une finalité théorique : il s'agit de développer des catégories et de repérer des liens entre ces catégories. Ceci est difficile à expliquer, à enseigner, mais facile à comprendre dès qu'on le pratique. Insistons sur un point : le logiciel fournit une « assistance » – c'est le chercheur qui conceptualise. Notamment, le codage est dépendant de l'identification d'une problématique. L'idée est de faire émerger des catégories-clefs. Il est possible d'opérer des codages thématiques qui consistent à constituer des groupes définis a priori. En ressources humaines, par exemple, on peut faire des groupes de catégories de personnel, ou regrouper par sites. Ne dérive-t-on pas alors dans une démarche hypothético-déductive ? Oui et non. On dispose d'un cadre théorique (ce qui rapproche de l'hypothético-déductif) et, en même temps, la pratique montre que beaucoup de choses inattendues émergent, qu'il y a des surprises, ce qui est l'intérêt de la démarche qualitative.

Concrètement, chaque « incident » apparaissant dans les transcriptions d'entretiens est codé. Le terme d'« incident » est assez peu défini dans la littérature, mais il est assez clair en pratique : il recouvre « une expression verbale achevée d'une attitude ou d'actes achevés d'un individu ou d'un groupe ». Le travail de codage est déjà un travail comparatif puisque lorsqu'on est tenté de coder un incident avec un code déjà utilisé, on compare cet incident à ceux qui ont été codés de la même manière et on s'interroge à la fois sur l'incident et sur le code : cette double interrogation amorçant le travail de catégorisation. Ce travail se concrétise par l'écriture de « mémos » : le chercheur met sur le papier les idées qui lui viennent à l'occasion de ses interrogations. Le logiciel intègre les mémos.

Le codage simple se double d'un codage hiérarchique ou codage axial : on peut relier des catégories « filles » à une catégorie « mère », un axe auquel se rapportent les autres catégories.

Le travail théorique se poursuit lors du codage sélectif sur la concentration autour de la catégorie-clef, sur le travail des liens entre la catégorie clé et les autres catégories.

### La publication

Elle reste un problème. Le chercheur dispose de données codées, de mémos et de théories, et il doit présenter le tout. Une solution consiste souvent à publier deux papiers : un de méthodologie sur le maniement des données, le processus de conception, d'analyse, d'organisation, le choix du logiciel et la catégorisation et l'autre sur les résultats et le contenu théorique de la recherche. On peut considérer qu'un chercheur ou un doctorant devrait toujours fournir une grille de saturation de ses catégories, ce qui est assez simple à faire mais nécessaire : pendant le codage, lorsque aucun concept nouveau n'apparaît, on peut considérer qu'on a saturé les catégories. Il convient également de travailler par codage extraction selon l'indexation des catégories obtenues et de faire apparaître certains verbatim à l'appui des catégories (mais sans surcharger, on risquerait de s'appesantir sur une forme de codage primaire au détriment de la conceptualisation, de la construction théorique).

Pour conclure, notons qu'il existe plusieurs types de logiciels disponibles sur le marché, et que le choix dépend des orientations épistémologiques retenues. La présentation qui a été faite repose plutôt sur l'utilisation de NUD\*IST. Mais, d'autres approches, portant par exemple sur la recherche de cartes cognitives, sont possibles (logiciel Decision Explorer).

*Question : l'approche que vous avez présentée n'est pas complètement fidèle à l'orientation première de la théorisation ancrée (orientation première qui a d'ailleurs été abandonnée au fil du temps et par Strauss dans ses travaux avec Corbin, et par Glaser de son côté) qui reposait sur l'idée de coder, donc de catégoriser à partir du matériau brut ; vous présentez en effet une approche qui repose sur des sortes de préclassifications, de pré-catégories. Ce faisant, est-ce que l'on ne se retrouve pas in fine avec des catégories qui devraient être issues du matériau mais qui sont en fait issues du chercheur ?*

*Réponse* - La question est compliquée, mais la réponse finalement assez simple. Tout d'abord, l'idée que l'on puisse aborder le matériau sans pré-catégories, vierge de toute idée, est naïve. Ce n'est jamais comme cela que les choses se passent. Le chercheur a toujours des idées quand il commence une recherche et quand il va sur le terrain. Cependant ses idées sont beaucoup plus réduites qu'il n'y paraît, si l'on prend l'exemple du nombre des pré-catégories par rapport à celui de l'ensemble des catégories finales obtenues, on obtient généralement un ratio de 10 %. Cela montre l'ampleur de la découverte. On se situe davantage à un niveau d'exploration qu'un niveau d'exploitation. Il y a toujours – fort heureusement – des surprises venant du codage du matériau, des catégories qui émergent. Donc, la question posée est plus théorique que pratique : en pratique, le travail de théorisation repose toujours sur une tension, une articulation, entre des catégories, des interrogations, apportées par le chercheur, et des catégories qui émergent du matériau. Beaucoup de chercheurs, en France surtout, estiment que pour éviter les biais, sortir de cette situation dans laquelle le chercheur risque de retrouver dans son matériau les catégories qu'il y a lui-même placées sans trop s'en rendre compte, il faut pratiquer le double codage. Je suis personnellement assez sceptique. La seconde personne ne va pas tout coder. Elle va se contenter de coder des petits bouts de matériau et de confronter ce second codage à celui effectué par le chercheur. C'est un investissement dérisoire et très superficiel. En effet, comme

je l'ai dit, le travail de codage suppose de la part du chercheur une imprégnation de son matériau, un travail de comparaison fine au moment du codage entre les incidents qui sont codés ou non de la même manière, une interrogation qui naît entre cette comparaison des incidents entre eux et leur regroupement dans une même catégorie ou dans des catégories différentes. L'imprégnation de l'ensemble du matériau et ce travail de comparaison fine, qui est à la base même du codage, sont impossibles au second codeur. Le double codage ne me paraît pas être une réponse très adéquate au problème posé.

Le double codage me semble en revanche très pertinent dans le cas d'un véritable travail en équipe, où deux chercheurs progresseraient concomitamment du début à la fin du codage. Même si le travail est ralenti dans ces conditions, il me semble qu'une telle démarche peut présenter une fiabilité intéressante avec toutefois le risque qu'une influence réciproque entre les deux chercheurs suscite une interprétation biaisée.

Le travail de catégorisation a toujours un aspect subjectif et, en même temps, les surprises qui viennent du codage du matériau permettent une objectivation sans qu'il y ait besoin de recourir à un point de vue donné par un autre codeur, qui lui a le désavantage de ne pas connaître le matériau aussi bien. ■

*Hervé Dumez*

*PREG — CNRS / École Polytechnique*