

LE PROCESSUS DE (DE)CODAGE DES DONNEES QUALITATIVES EN GESTION DES RESSOURCES HUMAINES

Catherine VOYNNET FOURBOUL

IAE

Université Lyon III

Les Grandes Terres

F-69210 Bully

Tél. (+33) 474 01 48 54

Fax. (+33) 474 01 07 46

c.voynnetfourboul@free.fr

Sébastien POINT

IAE

Université Lyon III

15 quai Claude Bernard

BP 0638

F-69239 Lyon cedex 02

Tél. / fax. (+33) 437 28 98 51

spoint@univ-lyon3.fr

Une analyse des contributions publiées dans les actes de l'AGRH ces cinq dernières années montre que les chercheurs explicitent rarement leurs éventuelles démarches de codage au sein de leur communication. C'est pourquoi cet article en propose une introduction. Selon nous le codage est un processus itératif, oscillatoire, facilitant l'abstraction des données en vue d'une construction théorique. A l'aide d'une illustration, nous essayons de montrer comment se différencient et s'articulent les différentes phases du codage. Pour cela nous avons utilisé l'exemple d'une recherche ayant fait l'objet d'un codage théorique de type « grounded theory » effectué à l'aide d'un logiciel. A partir d'une propriété (« l'information continue »), nous avons montré son articulation avec d'autres dimensions et propriétés ; nous avons donné un exemple de codage axial (intra-catégorie) pour cette catégorie-clé et un « scénario » pour le codage sélectif. L'intérêt des méthodes de codage est qu'elles facilitent la créativité du chercheur.

« Coding, as we have noted, can become tedious if you treat yourself as a sort of machine scanning the page methodically, picking out small segments of data and assigning categorical labels to them »

[Miles, Huberman, 1994 – p. 66]

« Coding is fun isn't it ? You never know what you might discover in a simple encounter »

[Strauss, Corbin, 1990 – p. 73]

Le codage est un processus d'analyse des données qualitatives, perçu de manières parfois opposées. Les citations en exergue de cet article illustrent ces deux visions: à la fois son côté fascinant, ludique mais également obsessionnel. Ce mode d'analyse des données qualitatives (ADQ) est défini par la littérature de différentes manières comme nous proposons de le voir au moyen d'un exposé des variations du codage en sciences de gestion.

La démarche méthodologique en recherche qualitative repose sur trois étapes : la collecte des données, le traitement de ces données et la production d'un rapport. La littérature abonde concernant les techniques de recueil des données qualitatives (par entretien, simulations ou observations) [Wacheux, 1996]. En revanche, elle présente des faiblesses sur le traitement de ces données, sans doute parce que les chercheurs, concentrés sur leur objet de recherche, disposent de peu de temps pour expliquer la manière dont ils ont échafaudé leur analyse. Le traitement et la présentation des données s'en trouvent donc affaiblis et même parfois occultés. En pratique, l'analyse suivante des contributions publiées dans les actes de l'AGRH ces cinq dernières années montre que les chercheurs explicitent rarement leurs éventuelles démarches de codage au sein de leur communication.

Nous proposons dans cet article une approche pratique à l'attention des chercheurs intéressés par le processus de codage, une partie cruciale de la recherche et pourtant souvent immergée. Les possibilités et techniques de codage existantes en sciences de gestion varient de l'« analyse de contenu » pratiquée en France, au codage théorique apanage du monde anglo-saxon ; or le codage théorique nous semble constituer une perspective très enrichissante.

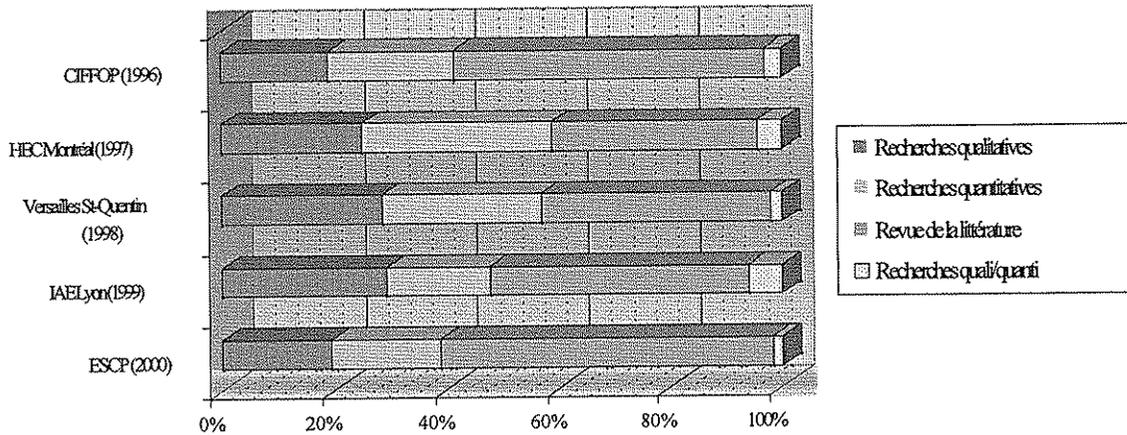
CINQ ANNEES DE CONTRIBUTION AU CONGRES DE L'AGRH

Une analyse approfondie des actes de l'AGRH de ces cinq dernières années (plus de 420 contributions) montre quelques tendances de recherche. Si les contributions n'initiant aucune analyse empirique (que nous appellerons les « revues de littérature ») représentent près de la moitié des papiers acceptés (48% exactement), la proportion de recherches qualitatives et quantitatives est relativement identique (près d'un quart des contributions, 25 et 24% respectivement). Enfin, ces cinq dernières années, une quinzaine de recherches utilisent à la fois une méthodologie qualitative et quantitative. Il s'agit généralement d'une triangulation des données [Denzin, 1988]) et non pas d'une quantification de données qualitatives (comme le suggère Ragin [1987]). Aussi, le souci de validité semble, pour certains auteurs, passer par un mélange de méthodologies quantitative et qualitative, de guide d'entretiens permettant de produire des questionnaires de qualité, administrés ensuite à grande échelle [Igalens, 2000 ; Igalens, Roussel, 1998] ou de questionnaires augmentés d'analyses documentaires.

Si le nombre de contributions par année a considérablement augmenté (d'une soixantaine de papiers en 1996 à plus d'une centaine au dernier congrès), la proportion des diverses orientations méthodologiques varie néanmoins d'années en années (voir figure 2). Par exemple, au 8^{ème} congrès de l'AGRH (Montréal, 1997), la proportion de « revues de littérature » représentait moins du quart des contributions (23%). Une proportion qui s'élève toutefois à près de 60% au dernier congrès (ESCP, 2000). Les thématiques des congrès et les lieux choisis peuvent en effet constituer des facteurs de contingences non négligeables sur la nature méthodologique des papiers proposés. Aussi, « l'internationalisation des ressources humaines » a engendré peu de recherches qualitatives ou quantitatives (moins de 20%

chacune). « Le contrôle (ou l'autonomie) des ressources humaines » a, quant à cette thématique, favorisé la proportion de méthodologies qualitatives (plus de 26% contre 17% de recherches quantitatives).

Figure: nature des communications publiées dans les actes de l'AGRH (période 1996 – 2000)



Considérons à présent les recherches qualitatives (soit un quart des contributions). Sur une centaine de papiers publiés, seules un tiers apportent une lumière sur leur démarche méthodologique en citant par exemple des auteurs dans leur bibliographie. Dans un cas sur dix, les ouvrages de Miles et Huberman ou de Yin sont cités. Sont plus rarement mentionnés la *Grounded Theory* de Corbin, Glaser et Strauss, ainsi que Wacheux, Bardin, Igalens et Roussel, Ghiglione, Mucchelli ou Denzin. D'autres auteurs sont également évoqués de manière isolée. La notoriété des ouvrages de Miles et Huberman [1994] ou encore de Yin [1989] peut s'expliquer par l'exposé de méthodologies qualitatives relativement proches des thématiques de gestion¹. Par exemple, la saturation théorique [Strauss, Corbin, 1998], principe selon lequel le chercheur ne détermine pas son échantillon à l'avance (nombre de cas) mais suivant la saturation des données recueillies, est difficilement applicable en sciences de gestion.

Peu d'informations sont délivrées à propos du codage. En effet, certains auteurs stipulent l'utilisation d'une méthodologie précise pour analyser leurs données, mais peu en donnent malheureusement les détails. En se focalisant sur ses résultats de recherche, le chercheur n'a pas l'espace nécessaire pour communiquer les techniques de codage utilisées. Si quelques auteurs affirment effectuer une analyse de contenu (un terme qui de toutes façons nous semble bien imprécis dans le processus de codage), rares sont les auteurs privilégiant des processus bien explicites telles l'analyse thématique ou la théorie fondée (« *grounded theory* »).

Les explications sur les démarches de codages demeurent peu explicites. Seules une poignée de recherches (4 en cinq ans) sur l'analyse lexicale (et l'utilisation de logiciel du type SPAD T ou ALCESTE) proposent un guide de leur codage automatisé (catégories identifiées). Outre

ces analyses « innovantes » utilisant l'ordinateur comme moyen de traitement des données, certains auteurs font preuve d'une certaine créativité dans les méthodes recensées. Ainsi les scénarios, les « histoires », l'ethnographie ou la méthode de l'incident critique sont ponctuellement utilisés.

Le lecteur a pourtant de quoi s'interroger sur la technique de codage mise en œuvre. Certaines recherches utilisent la théorie fondée et des démarches ethnographiques avec plus d'une centaine d'entretiens. En effet, bien des études utilisent un nombre considérable de cas et d'entretiens. Un tel volume de données qualitatives peut s'expliquer par le souci de validité mais aussi par une certaine réticence du chercheur à quitter son terrain pour se consacrer à l'analyse. En outre le codage d'une telle quantité de données apparaît pour le moins problématique. Fielding et Lee [1998] ainsi que Flick [1998] remarquent que les sociologues anglo-saxons travaillant sur des entretiens en profondeur diminuent progressivement la taille de leur échantillon (la médiane était de 50 en 1977 et de 35 en 1993).

En résumé, l'analyse de cinq années de contribution au congrès de l'AGRH aura permis de souligner les éléments suivants :

- Un quart des contributions optent pour une démarche qualitative,
- Peu de chercheurs spécifient véritablement le processus de codage adopté,
- Le souci de validité s'exprime de façon discutable par un nombre très élevé d'entretiens.

LES DEMARCHES QUALITATIVES EN THEORIE ET EN PRATIQUE

Que les données soient formalisées (rapports écrits), retranscrites (retranscription d'entretiens, d'observations ou de simulations), textuelles, sonores ou iconographiques, de nombreux processus méthodologiques sont à la disposition du chercheur pour analyser les données qualitatives. En effet, la littérature propose un éventail de techniques plus ou moins élaborées pour explorer, exploiter et comprendre des données qualitatives souvent riches et non structurées. L'ouvrage majeur de Miles et Huberman [1994] sur l'analyse des données qualitatives ne présentent pas moins d'une vingtaine de méthodologies dédiées à la compréhension, au décorticage et à la présentation des données qualitatives. L'ouvrage coordonné par Denzin et Lincoln [1998] illustre la nature de la recherche qualitative, c'est-à-dire un ensemble de méthodologies interdisciplinaires et transdisciplinaires contribuant à donner du sens ou à interpréter un phénomène à partir de diverses significations (p. 3).

• Aperçu rapide des méthodologies en recherche qualitative

Récemment l'ouvrage collectif dirigé par Thiétard [1999] consacre deux chapitres aux techniques de codage et autres méthodologies de l'analyse des données qualitatives. L'un [chapitre 16] propose de présenter le codage comme « un processus qui consiste à découper le contenu d'un discours ou d'un texte en unités d'analyse (...) et à les intégrer au sein de catégories sélectionnées en fonction de l'objet de la recherche ». Les auteurs fondent toutefois leur démonstration sur les ouvrages de Bardin [1998] mais surtout de Weber [1990] qui sont bien spécifiques en analyse qualitative.

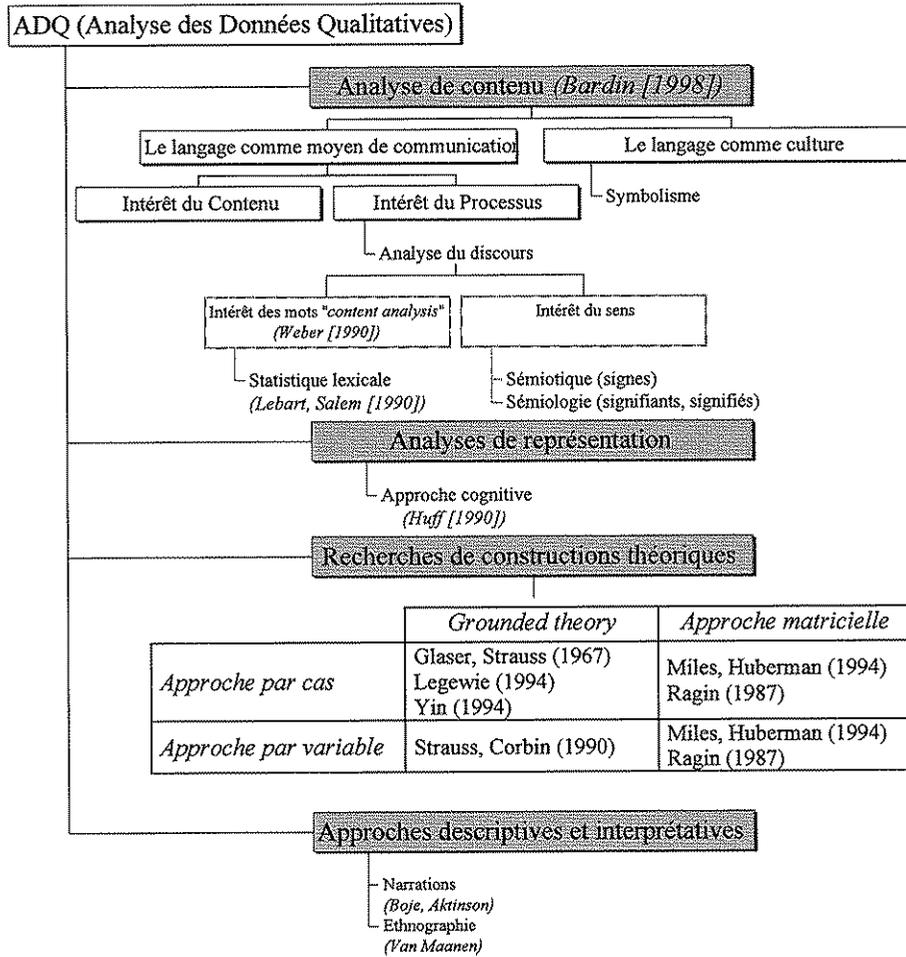
- **Sens de l'analyse de contenu en France et dans les pays anglo-saxons**

L'analyse de contenu, présentée dans la littérature française comme un ensemble de différentes méthodes [Bardin, 1998 ; Thiétart, 1999], n'est pas assimilable au terme anglo-saxon « *content analysis* ». En effet, selon l'ouvrage de référence en la matière [Weber, 1990], le « *content analysis* » demeure l'équivalent en France de l'analyse lexicologique. L'analyse lexicologique ou lexicométrique constitue une méthodologie pertinente pour détecter des structures à la fois apparentes et sous-jacentes d'un texte [Lebart, Salem, 1991]. La statistique lexicale semble de plus en plus employée en sciences de gestion mais parfois de façon critiquable.

Dans les ouvrages anglo-saxons de méthodologie qualitative, le « *content analysis* » est présenté comme une méthodologie, dont l'objet n'est pas de prendre en compte la contextualisation [Denzin, Lincoln, 1998, p. 248] ; le chercheur lit, compte et catégorise des mots issus d'un corpus textuel [Fielding, Lee, 1998, p. 52]. Weber [1990] insiste néanmoins sur la nécessité de considérer le contexte (signifié) afin de lever toute ambiguïté inhérente au discours. Cette démarche se résume ainsi essentiellement à identifier des éléments pour les classer par catégories. De ce point de vue le processus de codage est semblable à une approche thématique [Boyatzis, 1998].

Mais en France, l'analyse de contenu (traduction littérale de « *content analysis* »), renvoie à d'autres démarches. La **Figure**, qui recense les auteurs et les méthodologies évoquées dans cette communication, donne également un aperçu des démarches assimilables à l'analyse de contenu. Pour Thiétart [1999], l'analyse de contenu permet de dépasser la simple description du contenu d'une communication et d'expliquer certaines stratégies notamment au niveau comportemental (p. 463). A la lecture de l'ouvrage majeur de Bardin [1998] sur l'analyse de contenu, nous assimilons celle-ci à une méta-méthodologie, qui oriente le chercheur vers la construction de logiques à la fois au niveau de la structure du discours (occurrences et cooccurrences de mots) et au niveau du contenu de celui-ci (analyse thématique).

Figure:
Les approches privilégiées pour une analyse des données qualitatives en sciences de gestion



Les techniques proposées à travers l'analyse de contenu (telles la catégorisation, l'inférence) [Bardin, 1998] reviennent à rassembler des données dans des catégories (préexistantes ou non) et de veiller aux cooccurrences. Or comme le souligne Angot et Milano [in Thiétart, 1999], la difficulté majeure relève du processus d'abstraction (processus par lequel le chercheur passe de ses données empiriques à des concepts représentés par des codes théoriques).

• **Le codage comme construction théorique**

Nous souhaitons attirer l'attention du lecteur sur un problème majeur relatif au processus de codage et bien souvent occulté par la littérature : les relations entre codes. Si les méthodologies de cartographie cognitive [Huff, 1990] et l'analyse de similitudes [Silem, Martinez, 1990] sont présentées comme des méthodologies reliant les concepts entre eux, le lien reste peu explicite et rarement qualifié (hormis l'expression de signes lors de la recherche de causalité ainsi établie du phénomène). L'analyse de similitudes permet de dégager la

structure relationnelle d'un ensemble d'éléments et la traduit dans des représentations graphiques permettant de saisir à la fois la globalité et les relations entre ces éléments. Cette démarche permettant à la fois de traiter le mot et son contexte, suppose que la co-apparition de deux mots dans une même phrase traduit à elle seule la présence d'une représentation particulière [Silem, Martinez, 1983, p. 119].

Tableau

Les définitions du codage identifiées dans la littérature

Nom	Référence	Définition du « codage »
Coffey A., Aktinson P.	[1996 : 28-49]	... est un processus heuristique qui fournit au chercheur des façons de penser les données (un processus d'indexation). Le chercheur se pose lui-même des questions à propos des données. Le codage diffère de l'analyse des données.
Kelle U.	[1995]	... est à la fois un processus mécanique et conceptuel.
Miles M. B., Huberman A. M.	[1994 : 56]	... est un processus permettant au chercheur d'identifier des données pertinentes.
Strauss A., Corbin	[1990 : 61]	... est le processus d'analyse des données.

Le tableau illustre les différentes définitions relevées à partir de la littérature. Est accentuée la dimension progressive du codage (se poser des questions, un processus heuristique...). L'interprétation des données est au cœur de la recherche qualitative. Généralement deux stratégies se dessinent : une stratégie interactive dont l'objectif est la construction théorique (codage théorique) et une stratégie d'analyse séquentielle visant à reconstruire la structure du texte (analyse de contenu). Nous distinguons quatre grands types de codage (cf. tableau). Si l'analyse de contenu est la technique de codage la plus utilisée en France, elle est également la moins interprétative.

Nous considérons à présent la construction théorique (approche privilégiée par Glaser et Strauss [1967] et qui a évolué avec les travaux de Strauss et Corbin [1990]). Le chercheur souhaite ainsi :

- analyser ses données en procédant à un codage, en écrivant des mémos au sujet de ses idées, des codes, des interrelations et des nouvelles directions pour sa recherche,
- dessiner des diagrammes illustrant ses conceptions tirées des données (l'interaction permanente entre les données, le codage et la construction des mémos est essentielle, et permet de bâtir progressivement une structure conceptuelle, qui évolue vers une abstraction),
- procéder à des comparaisons constantes entre divers incidents (un incident est une expression ou un acte verbal achevé d'une attitude chez un individu ou un groupe d'individus), afin de vérifier si cette structure théorique est bien soutenue par les données.

Strauss et Corbin [1990], proposent une approche assez opérationnelle du codage. Selon les auteurs, le processus comporte plusieurs étapesⁱⁱ :

- l'affectation d'un code à une unité de texte (le code étant le label conceptuel d'une catégorie reflétant une idée, une personne, un concept)
- le codage ouvert (*open coding*), consiste à développer les catégories en se concentrant sur leur nature, leur dimension et leurs relations ; il est toutefois important de souligner que les

codes ne sont pas toujours théoriques et émanants des réflexions du chercheur. Les codes *in-vivo* (c'est-à-dire issus directement des données recueillies) peuvent également être considérés.

durant le codage axial (*axial coding*), les données sont revues de façon à établir des liens entre une catégorie et ses sous-catégories ; le codage ouvert et le codage axial sont dans la pratique très imbriqués, fortement interactifs

le codage sélectif (*selective coding*), consiste à se concentrer sur la catégorie la plus importante, la catégorie clé (*core category*) de manière à étudier ses relations avec les autres catégories.

Miles et Huberman [1994] proposent une approche adaptée, procédant par itération inductive et déductive en employant non seulement les relations descriptives mais aussi des relations causales ; ils mettent l'accent sur l'attrait de la présentation en utilisant des synoptiques, des résumés structurés de cas, des diagrammes en réseau ainsi que des matrices. Ces auteurs distinguent également deux étapes dans la progression du codage :

l'élaboration des codes descriptifs voire interprétatifs (p57),

la répétition de ces codes en leitmotivs émergents (ce que les auteurs appellent des métacodes « *pattern codes* »).

Tableau

Les différents types de codage (inspiré de Flick [1998])

Type de codage	Contribution	Objectif de recherche	Limites
L'analyse de contenu Catégories théoriques But : réduire le matériau	Procédure codifiée, analyse réductrice comparaison des cas	METHODE NON LIMITEE A UN PLAN THEORIQUE Entretiens semi-structurés	Risque de perte de sens du fait de l'application de catégories théoriques fixées Risque quantitativiste
Analyse globale Legewie [1994] Vision globale des thématiques d'un texte	Donne une vue générale du matériau Peut être une étape du codage théorique	Vient en complément des méthodes de catégorisation classique	peu compatible avec des analyses séquentielles (de conversation, de discours, narrative)
Codage théorique Glaser et Strauss [1967] Strauss et Corbin [1990]	Induction et déduction Développe des catégories et des relations entre catégories	Développer des théories « <i>Grounded theory</i> »	Difficile à enseigner Difficulté d'appliquer le principe d'échantillonnage théorique Risque d'un grand nombre de codes
Codage thématique Boyatzis [1998] Etudes comparées Groupes définis <i>a priori</i>	Correspondances et différences entre cas Comparabilité	Les mondes sociaux Analyse de la variété et de la distribution des perspectives	Se limite à des comparaisons de groupes pré-définis Prend du temps

Seidel et Kelle [*in* Kelle, 1995] différencient deux manières de définir les codes en fonction des approches soit déductives soit inductives. Dans l'analyse de contenu « classique », un projet de codage précis établit au préalable une liste de codes, ensuite le chercheur attribue ces codes aux segments de texte. Le travail se poursuit à l'aide de procédures statistiques sur les codes exclusivement. Les codes sont utilisés comme des représentations d'un phénomène, ce qui nécessite une approche déductive. Cela nécessite que les variables pertinentes, et leur valeur doivent être déterminées avant que les données ne soient codées. Nous retrouvons également ce processus de codage déductif en analyse lexicale. En effet, avant tout codage

préalable, le chercheur peut définir un dictionnaire thématique où seront catégorisés les vocables du corpus considéré.

Le paradigme interprétatif privilégie la prise en compte des interprétations et du sens commun des acteurs. Ceci implique une orientation inductive de l'approche méthodologique. Les différentes parties des données sont analysées afin de trouver des points communs, des différences ou des liens entre elles. Dans une stratégie d'analyse qualitative de tradition plus interprétative, les codes ne sont pas utilisés comme dénominateur de certains phénomènes, mais comme dispositif heuristique pour la découverte.

Gauzente [1998] associe l'analyse de contenu à un codage de type déductif et la *Grounded Theory* à un codage inductif. Nous privilégions davantage une dimension itérative du codage, c'est-à-dire effectuer des « aller-retours » entre les données recueillies et les interprétations successives du chercheur (démarches à la fois inductive et déductive).

Seidel et Kelle [*in* Kelle, 1995] soulignent également le danger de perte du phénomène par réification des codes. Si l'analyste commence à travailler exclusivement sur ses codes, il en arrive à oublier les données de base. Un autre danger est de perdre le contexte de l'information. Or, en sciences de gestion, la littérature insiste sur le fait que le discours ne soit pas resté étanche à son contexte social [Alvesson, Kärreman, 2000]. Pour lutter contre ces dangers, l'emploi des logiciels peut aider le chercheur à rester en contact avec les données recueillies. Que ce soient des logiciels d'analyse lexicale ou d'aide à la construction théorique, ces outils facilitent la gestion d'un grand volume de données, tout en gardant la flexibilité nécessaire de ses interprétations. Bournois *et al.* [2001] proposent un recensement de ces logiciels et en soulignent les points forts et faiblesses inhérents à leur utilisation.

En conclusion de cette revue de littérature sur l'analyse de données qualitatives, le codage favorise la construction théorique. En effet, lors du processus de codage, le chercheur va particulièrement veiller à :

- Mettre en évidence un certain nombre de construits (processus d'abstraction, que les codes soient *in vivo* ou théoriques),
- Relier ces construits les uns aux autres par des liens émanant de ses interprétations (la relation entre les construits demeure fondamentale [Doty, Glick, 1994]),

En recherche qualitative, le codage ouvre des perspectives de construction théorique. Nous définissons donc le codage comme : un processus itératif, oscillatoire, facilitant l'abstraction des données en vue d'une construction théorique.

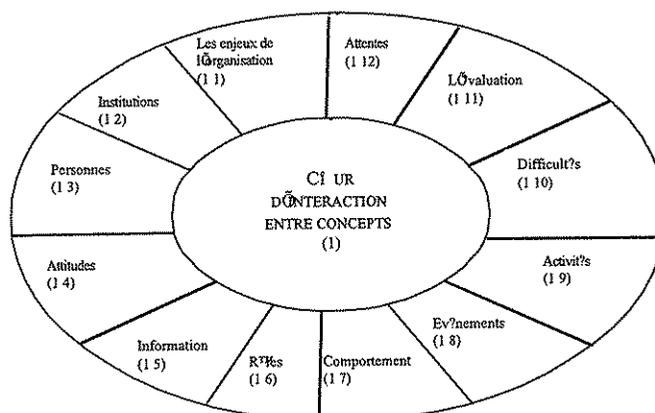
Le codage est :

- Un **processus** : le codage fait appel à des méthodes variées et progresse par étapes,
- Itératif** : l'aller-retour des données aux codes, et des codes aux données permet de transformer progressivement les codes en concept,
- Oscillatoire** : empruntant à la fois aux démarches inductive et déductive de façon dynamique et récursive,
- Une **abstraction** car à partir de ces données recueillies, le chercheur souhaite « mettre en oeuvre » progressivement des codes théoriques qui auront une valeur générale et non plus circonscrite,
- Une **entreprise de construction théorique** du fait de la combinaison entre l'abstraction et la mise à jour de liens entre codes.

Compte tenu de ces observations, nous souhaitons à présent proposer au chercheur une technique de codage favorisant la construction théorique et proposant ainsi d'autres horizons que ceux proposés par l'analyse de contenu.

EXPERIENCES ET ILLUSTRATION D'UNE DEMARCHE DE CODAGE

Une recherche portant sur le comité d'entreprise européen nous sert de support pour illustrer une démarche de codage ; 36 Directeurs des Ressources Humaines, membres de syndicats permanents ou de différents comités d'entreprise européens, dans différentes langues ont été interrogés. Compte tenu du volume important des données recueillies (retranscription de plus de 1000 pages), nous avons utilisé des logiciels d'aide à l'analyse des données qualitatives : NUD*IST (version 4) et DECISION EXPLORER. Une étude récente sur ces outils logiciels situe NUD*IST comme un logiciel d'aide à la construction théorique et dont la fonctionnalité principale est de produire des matrices [Bournois *et al.*, 2001]. Nous allons détailler le processus de codage utilisé dans cette recherche qui a abouti à l'identification de plus de 700



codes (sur la base de recherches identiques et à titre de comparaison, la littérature considère généralement une moyenne de 500 codes [Fielding, Lee, 1998]).

Figure: Les 12 catégories principales au centre de notre processus de codage

La figure 3 présente la première étape de notre recherche : la segmentation en codes thématiques principaux (ou catégories-mères). Nœuds et catégories peuvent être employés pour désigner le même support ; toutefois il existe une différence fine entre ces deux désignations : si les nœuds ou codes constituent des raccourcis les plus proches possibles du discours brut, la catégorie, elle, est une notion plus abstraite, qui est pensée en terme d'articulation. Nous emploierons toutefois les termes indifféremment.

Le choix de ces catégories-types résulte d'un processus et d'une réflexion.

1- Les premières catégories-mères suivent les éléments-clés du guide d'entretien (en particulier les éléments théoriques de départ, les questions de recherche, selon l'approche déductive préconisée par Miles et Huberman [1994]).

2- Les autres catégories sont nées de l'interaction entre les données recueillies, des interprétations des codes et des relations des nœuds entre eux (phase inductive). Lorsque nous lisons notre texte, nous relevons un fait, un thème, une idée que nous estimons avoir de l'importance dans l'interprétation future des données. Nous lui attribuons un code (ou nœud), à son tour placé dans l'une des catégories principales [Araujo *in* Kelle, 1995]. Chaque nœud ou catégorie possède alors une adresse spécifique, correspondant à la place de ce nœud dans l'arbre des catégories organisé de manière hiérarchique. Si certains nœuds, au début de notre processus, étaient positionnés comme des codes secondaires (ou sous-catégories), ils se sont constitués en codes principaux au fur et à mesure que nous leur accordions de l'importance.

Les 12 catégories principales représentées constituent un résultat intermédiaire de la recherche. Après avoir codé une dizaine de transcriptions, nous avons pu stabiliser les 12 catégories principales de l'arbre des index. Rappelons que le codage est un processus progressif et itératif. L'articulation des codes les uns par rapport aux autres ne cesse d'évoluer dans cette première phase d'analyse.

• Le codage ouvert ou l'identification des codes

Au cours de cette première étape de notre processus de codage, il s'agit de repérer les différents codes dans le corpus analysé. Les codes ainsi identifiés viennent se greffer autour de catégories principales identifiées en amont de notre travail (cf. la démarche théorique et conceptuelle de Huberman et Miles [1998]). Au fur et à mesure de la lecture des transcriptions, de nouveaux codes sont créés et s'ajoutent aux codes pré-existants. Ces codes sont ainsi attribués à des unités de texte (mot, phrase, ligne, paragraphe ou document). Bien évidemment, une unité de texte peut ainsi être affectée par plusieurs codes (voir exemple ci-dessous). Dans notre exemple, nous naviguons entre données brutes et catégories. Ces itérations nous permettent d'obtenir une certaine complétude des codes (ce que Miles et Huberman [1998] appellent la saturation, c'est-à-dire lorsque de nouvelles données n'ajoutent rien à la signification de la catégorie).

Nous prendrons un exemple pour illustrer concrètement la démarche de codification. L'exemple ci-dessous (cf. figure 4) nous permet :

- D'introduire le nœud « *information continue* » (de voir quelle définition en fait le répondant, quels sont les autres nœuds sous-jacents (nombre de réunions) ou nœuds explicites (information compréhensible) qui sont reliés), (entretien #1)
- De comparer ce nœud d'une situation de gestion à une autre (il s'agit toujours du nœud « *information continue* », mais dans des contextes différents),
- D'interpréter le nœud (l'information continue est ressentie comme une attente du secrétaire, elle a pour but d'actualiser la nature de l'information et elle s'accompagne de la dimension relationnelle entre les différents acteurs, elle nécessite communication et relation régulière entre les personnes (entretien #2),
- D'aborder par exemple les contraintes multi-culturelles et linguistiques. Sont ici exprimées la nécessité de la régularité, mais aussi la difficulté de la mise en œuvre [de l'information continue]. Cet entretien conforte notre idée que la qualité de l'information va être un facteur de différenciation entre les entreprises (celles qui jouent le jeu véritablement et celles qui se contentent d'appliquer la loi). Cet entretien nous oriente également sur la dimension de continuité de l'information, propriété jusqu'ici non intégrée (entretien #3).

Figure

Illustration de la première étape de codage

ENTRETIEN #1	
<p>...</p> <p>Moi je crois que la première des choses, c'est d'avoir joué la carte de l'information continue, j'insiste là dessus, parce que pour moi c'est fondamental. C'est-à-dire qu'il n'y a pas un projet, qui arrive comme ça, que cela tombe dessus, sans qu'ils comprennent quelle est la logique de l'entreprise, d'où ça vient, à quoi ça correspond, à quoi ça répond etc. on n'a jamais dit : voilà cela va se passer dans tel endroit, on ne le savait pas forcément à cette époque là, mais que cela s'inscrive dans une logique d'évolution dont ils comprennent le sens. Qu'on ait pu suffisamment en amont, dans une certaine mesure, une fois que nous-mêmes, commençons à dire voilà à peu près où on voudrait aller, travailler avec eux. Cela me paraît le critère tout à fait fondamental. C'est pour ça que cela nécessite effectivement de se voir beaucoup. Le deuxième, c'est un peu ce que l'on vient de dire, c'est la proximité de la relation. C'est sûr que, si on entretient de bonnes relations humaines, avec un certain nombre de gens, on arrive à dire les choses plus facilement, plus simplement. Je crois que cela permet que même si on traverse des périodes difficiles, le contact ne soit jamais rompu et qu'on puisse faire un peu la part des choses.</p> <p>...</p>	<p>(1 5 2 17) /Données/l'inform. /car.info/ information continue (1 5 2 26) /Données/l'inform. /car.info/ inform. Compréhensible (1 9 2) /Données/Activités / les travaux CEE (1 11 1) /Données/évaluation /satisfait (1 2 2 2 1) /Données/ institutions/ moyens /réunions/ Nbr réunions (1 7 9) /Données /Comportement membres /Les relations</p>
ENTRETIEN #2	
<p>...</p> <p>Non, une autre voie très facile est qu'il y ait plus de contacts entre le comité restreint composé de 6 personnes et la Direction, avec des réunions plus fréquentes, des meilleures informations et des informations mises à jour, pas des informations de l'année écoulée.</p> <p>...</p>	<p>(1 2 1 3) /Données/ Institutions/ struct. C.E.E./ instance restreinte (1 5 2 17) /Données/ l'inform. / car. info/ information continue</p>
ENTRETIEN #3	
<p>...</p> <p>Mais, l'important aussi, c'est qu'il y ait communication régulière et c'est sans doute une des choses très difficiles. C'est déjà difficile à faire quand on est sur un pays, que l'on est monolingue et quand on est mono-culturel. C'est éminemment beaucoup plus difficile, quand c'est multi-lingue, multi-culture et sur des données, qui ne sont pas apparemment toujours l'immédiat quotidien. Cela suppose tout un travail de captage de l'information, de circulation, de traitement de l'information dans les 2 sens. C'est l'un des problèmes redoutables pour les comités. D'ailleurs, tous les comités ont tous conscience, que ce soit du côté de la Direction, quand la Direction veut jouer le jeu ou que ce soit du côté des travailleurs.</p>	<p>(1 12 1 2) /Données/ Attentes/ Amélior. possibles/ informations (1 2 2 2 1) /Données/institutions/ les moyens/ les réunions/ Nbr réunions (1 12) Attentes (1 5 2 17) .../ Information continue (1 11) Evaluation (1 10 1) Difficulté du C.E.E./ Langues (1 10) Difficulté du C.E.E.</p>

Ces exemples de codage nous montrent comment les nœuds s'enrichissent dans leur définition, qui intègre la relation avec d'autres idées et deviennent progressivement des notions plus larges, peu à peu assimilables à des concepts :
dans notre premier cas l'information continue a une valeur explicative, pédagogique, et nécessite une relation régulière,

dans notre second cas, l'information continue est une attente des participants et une façon d'actualiser les données,
 dans notre troisième cas l'information continue est à la fois une difficulté et un critère de segmentation des entreprises.
 C'est ainsi que le processus de codage permet après comparaison des textes affectés par les codes de délivrer le sens lié au code.

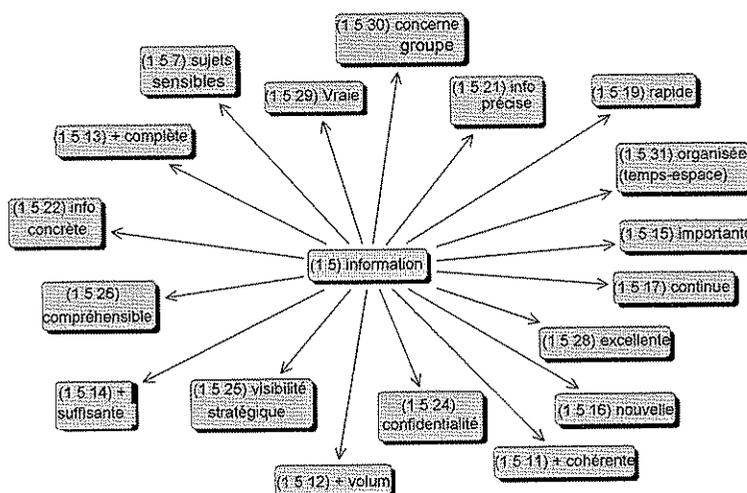
• **Le passage au codage axial**

Dans notre exemple, l'information continue se rattache à une catégorie plus générale : « l'information ». D'autres codages portant sur d'autres propriétés de l'information ont été menés de la même manière. « L'information » apparaît dans notre cas comme une catégorie-clé que nous avons déclinée avec ses propriétés et ses dimensions. Une propriété est par exemple : « une information complète » et deux dimensions dichotomiques de cette propriété ont été considérées : la dimension qui constate l'existence ou la non existence de la propriété. On dira qu'une information est « faiblement » complète versus qu'elle est « fortement » complète pour exprimer son existence ou pour exprimer le contraire. La catégorie information en devenant un axe du codage s'interprète et se représente comme une catégorie axiale. Nous faisons figurer à la suite ses propriétés (sans toutefois mentionner les dimensions par souci de clarté).

Si la démarche de codage présentée dans cette communication s'apparente au processus élaboré par Strauss et Corbin [1990], le chercheur peut toutefois multiplier à ce stade du codage la création de matrices (l'ouvrage de Miles et Huberman [1994] permet en effet la confection de bon nombre de matrices suivant l'objectif et le processus de recherche suivi).

Figure

Un exemple de codage axial sur la catégorie "information"



Tableau

Nombre d'occurrences de la catégorie "information" déclinée par propriétés et dimensions selon les fonctions des répondants

Information	Représentant au comité d'entreprise européen	Syndicat permanent	Coordinateur Europe	Direction relations sociales	Total	Représentant au comité d'entreprise européen	Syndicat permanent	Coordinateur Europe	Direction relations sociales	Total
Dimensions	Information faiblement :					Information fortement :				
Propriétés										
Complète	4	4	1		9	1	2		4	7
Suffisante	4	3	2		9		1			1
Nouvelle	7	1	1	1	10	2	1			3
Précise		3		1	4	1			1	2
Non sensible	10	4	1	2	17	1		1	4	6
Confidentielle		2		1	3	3	5		4	12
Volumineuse	2				2	3	2	1	2	8
Cohérente	1				1	1				1
Concrète			1		1	2		1	1	4
Organisée	2		1		3				3	3
(temps-espace)										
Compréhensible	1				1	2	2		2	6
Continue						2	1		3	6
Vraie						1		1	3	5
Excellente						2		1	3	6
Unilatérale	1			2	3					
Importante						1	1			2
Superficielle	1				1					
Visibilité stratégique	3	3			6	5	3	1	4	13
Rapide						1				1
Total	36	20	7	7	70	28	18	6	34	86

Le logiciel avec lequel nous avons effectué notre analyse (NUD*IST) permet de dresser des matrices d'occurrences pour les propriétés de la catégorie axiale « l'information ». Il est ainsi possible d'obtenir les occurrences à chaque niveau de notre codage (catégorie principale [information], secondaire : propriété [information continue], dimensions [degré de complétude]. Nous avons en effet choisi de faire des distinctions selon 4 groupes d'appartenance de nos répondants (Représentant au comité d'entreprise européen, Syndicat permanent, Coordinateur Europe, Direction relations sociales). Le tableau 3 croise les propriétés de l'information avec l'appartenance des répondants à l'origine de ce type d'idées.

• Le passage au codage sélectif

Avec le codage axial, nous avons vu comment s'organisait une catégorie-clé « l'information ». Une troisième étape consiste à élargir la recherche de liens entre nœuds quelle que soit leur position hiérarchique. Les codes ne doivent plus être pensés d'une

manière structurée ou ordonnée pour privilégier une approche plus flexible par réseau (on parle alors de « réseau conceptuel »). Plusieurs manières d'analyser les liens entre les codes sont à la disposition du chercheur.

Il peut au fur et à mesure de la lecture des transcriptions en se référant au sens, décider de comptabiliser les liens entre chaque code ; dans ce cas il utilise une approche globale. L'unité choisie est la transcription complète. Cette méthode s'avère lourde si elle est exécutée manuellement ; il peut toutefois se concentrer sur les seuls codes qui intuitivement lui apparaissent importants. L'intérêt réside dans l'expression d'une logique mais les risques de perte de liens constitue la grande faiblesse de cette approche.

Une autre solution consiste à systématiser la procédure et à automatiser les requêtes de recherche de liens grâce à un outil informatique. Les liens sont détectables grâce aux cooccurrences entre nœuds ; pour cela il s'agira de définir l'unité de texte dans laquelle la cooccurrence pourra être relevée par programme. Dans ce type d'exercice, l'unité choisie doit être suffisamment large pour que l'on puisse relever des cooccurrences. Dans la pratique les chercheurs peuvent choisir le paragraphe ou la phrase comme unité de traitement afin de capturer un grand nombre possible de liens ; il s'agit ensuite de relire les unités de texte pour éliminer les cooccurrences fortuites et pour qualifier la nature du lien entre codes. Le chercheur peut vouloir se concentrer sur les liens les plus récurrents et les représenter à l'aide d'une carte. Nous avons retenu cette solution.

Le lien est évidemment qualifié par le chercheur, qui dans un souci de représenter au mieux le phénomène, retourne examiner chacun des segments de texte concerné par le lien établi entre les deux codes. Dans notre exemple sur l'information continue, nous avons choisi de présenter une cartographie à partir de la catégorie-mère (ou principale) : l'information. Celle-ci possède des liens avec de nombreuses sous-catégories. Dans un souci de clarification, retenons uniquement les sous-catégories clefs.

A partir de la cartographie ainsi représentée (voir figure 6), nous extrayons tous les segments de texte illustrant les liens représentés. Pour la compréhension du lecteur de la cartographie, il est alors nécessaire de produire des explications (comme suit pour notre exemple) :

« L'information dans sa conception générale, est liée à l'idée de consultation lors des restructurations. Les participants de Renault ont évoqué le conflit de Vilevoorde dont l'origine tenait au défaut de fonctionnement de l'information - consultation. Un manque d'information, un défaut de fonctionnement du processus de consultation peut provoquer un conflit. Un premier sens conféré à l'information est l'information dite de consultation. Lorsque les sujets sont sensibles, et que les Directions acceptent de les évoquer, elles les qualifient de confidentiels pour éviter le phénomène de divulgation. Les sujets sensibles concernent les projets de cession, d'acquisition d'activité, de restructuration qui, s'ils étaient annoncés prématurément risqueraient de ne pas aboutir du fait des divulgations possibles. La stratégie des Directions est donc de ne pas donner une information complète en n'abordant pas franchement les questions de restructuration. Les entreprises d'origine britanniques, nordiques, asiatiques et nord-américaines sont plus enclines à ne pas transmettre une information complète et, pour cela, diffusent une information fragmentée, relative à des activités que l'entreprise a pris soin de séparer au préalable afin de ne pas rendre

possible la vision globale de la stratégie de l'entreprise. Même pour les entreprises européennes, subsiste un doute et l'information présente un caractère vague, flou, non concret. Plus grave, est l'impression des représentants syndicaux que l'entreprise n'a pas une vision claire de sa stratégie. Le point de référence pour juger si une information est nouvelle est la presse : si l'information est connue avant la presse, l'information est considérée comme nouvelle. L'information continue est une notion, qui est liée au fonctionnement de l'instance restreinte. »

Par ailleurs, le passage de l'analyse de la recherche qualitative à la partie rédactionnelle peut s'effectuer en utilisant les données extraites structurées en fonction de l'arbre. Il est alors facile de suivre le déroulement arborescent des codes et de produire des synthèses des discours des répondants et de commentaires (les mémos) attribués lors du codage (voir également l'ouvrage de Miles et Huberman [1994] qui illustre de nombreuses stratégies de rédaction).

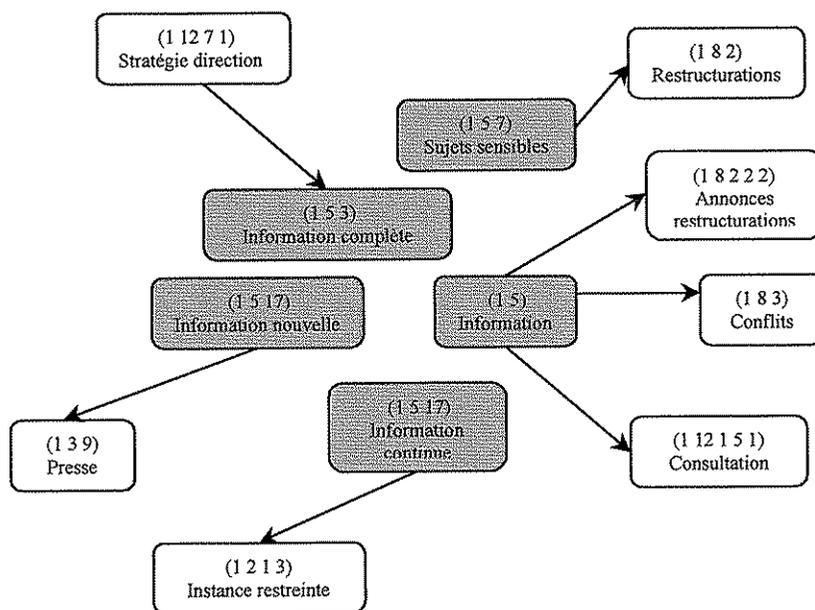


Figure
Codage sélectif – Un réseau conceptuel autour du code "information"

CONCLUSION

Les méthodes de codage de la recherche qualitative sont peu explicitées dans la littérature. Nous avons essayé de montrer comment se différencient et s'articulent les différentes phases du codage. Pour cela nous avons utilisé l'exemple d'une recherche ayant fait l'objet d'un codage théorique de type « grounded theory » effectué à l'aide d'un logiciel selon une

catégorisation hiérarchique. A partir d'une propriété (« l'information continue »), nous avons montré son articulation avec d'autres dimensions et propriétés; nous avons donné un exemple de codage axial (intra-catégorie) pour cette catégorie-clé et un « scénario » pour le codage sélectif.

En recherche qualitative, l'objectif primordial est de donner du sens à ses données. Aussi cette démarche est similaire à l'assemblage d'un puzzle (métaphore empruntée à Kelle [1995, p. 57-58]). Si un arbre, un immeuble et le ciel composent l'image du puzzle, l'utilisateur utilise des codes qu'il nomme « arbre », « ciel » et « immeuble ». Il range ainsi les pièces du puzzle suivant ses catégories (illustration du codage ouvert). Il affine ensuite son premier tri : les formes et les couleurs vont être méticuleusement examinées, comparées (illustration du codage axial) puis assemblées une à une, les unes aux autres (illustration du codage sélectif). Contrairement à un jeu de puzzle, le chercheur ne dispose pas d'une image originellement imprimée sur la boîte d'emballage. Sa démarche n'en sera que plus créative.

L'intérêt des méthodes de codage tient dans les critères de qualité de la recherche qualitative. Les critères de qualité d'une recherche qualitative dépassent le simple fait de reporter les citations illustratives, ce que Flick [1998] appelle la « plausibilité sélective ». Les avancées dans la recherche qualitative permettent de relever certains critères qu'une analyse menée à l'aide des techniques de codage exposées peut satisfaire. En pratique la fiabilité se constate lorsque le travail de recherche explicite le moment de l'interprétation [Flick, 1998, p. 224]. La fiabilité se mesure avec ce type d'analyse lorsque l'on passe du codage ouvert au codage axial et sélectif.

Egalement la technique de codage associée à la puissance d'un logiciel d'aide à l'analyse des données qualitative permet d'assurer une meilleure validité en minimisant les risques d'erreurs principales : voir une relation, un principe, alors que celui-ci est incorrect (erreur type 1), le rejeter alors que celui-ci est correct (erreur type 2).

Enfin, la triangulation d'investigation est facilitée avec cette méthode dans la mesure où des chercheurs peuvent travailler en groupe ; Fielding et Lee [1998] donnent des conseils intéressants à ce sujet. La triangulation méthodologique peut consister à décliner des questions sous des angles différents et variés, afin d'obtenir une meilleure représentation des situations évoquées par le répondant. La triangulation risque toutefois d'être utilisée comme une combinaison de méthodes susceptibles d'accroître la richesse des résultats plutôt que de chercher à valider les résultats [Kelle, 2001].

On ne peut que constater le lien étroit et grandissant entre le codage et les outils logiciels qui lui sont désormais étroitement associés. Cette synergie est fructueuse dans certaines limites : tant que le chercheur n'est pas tenté de jouer abusivement avec toutes les fonctionnalités offertes par ce type d'outils. D'où la nécessité de formaliser les procédures de codage théorique. C'est en ces termes que Miles [1979] qualifie la recherche qualitative de « nuisance attractive »...

BIBLIOGRAPHIE

- BARDIN L. (1998), *L'analyse de contenu*, Presses Universitaires de France, Paris, 9^{ème} édition.
- BOURNOIS F., POINT S., VOYNNET FOURBOUL C. (2001), L'Analyse des Données Qualitatives Assistée par Ordinateur, *Revue Française de Gestion*, à paraître.
- BOYATZIS R. E. (1998), *Transforming Qualitative Information : Thematic Analysis and Code Development*, Sage Publications, Thousand Oaks.
- COFFEY A., AKTINSON P. (1996), *Making Sense of Qualitative Data : Complementary Research Strategies*, Sage Publications, Thousand Oaks.
- DENZIN N. K., LINCOLN Y. S. (1998), *Handbook of Qualitative Research : Collecting and Interpreting Qualitative Materials*, London, Sage Publications.
- DOTY D. H., GLICK W. H. (1994), Typologies as a Unique Form of Theory Building : Toward Improved Understanding and Modeling, *Academy of Management Review*, vol. 19, n° 2, pp. 230-251.
- FIELDING N. G., LEE R. M. (1998), *Computer Analysis and Qualitative Research*, Sage Publications, Londres.
- FLICK U. (1998), *An Introduction to Qualitative Research*, Sage Publications, Londres.
- GAUZENTE C. (1998), Grounded Theory et Techniques d'analyse de contenu : comparaison et illustration, *actes des XIV^{ème} journées nationales des LAE*, Nantes, tome 3.
- GLASER B. G., STRAUSS A. L. (1967), *The Discovery of Grounded Theory : Strategies for Qualitative Research*, Aldine de Gruyter, New York.
- HUBERMAN M. A., MILES M. B. (1998), *Data management and analysis method*, in Denzin N.K., Lincoln Y.S., *Collecting and Interpreting Qualitative Materials*, Sage Publications
- HUFF A. S. (1990), *Mapping Strategic Thought*, Wiley & Sons, Chichester.
- IGALENS J., (2000), *Audit des ressources humaines*, éditions Liaisons.
- IGALENS J., ROUSSEL P. (1998), *Méthodes de recherche en gestion des ressources humaines*, Economica
- KELLE U. (2001), Sociological Explanations between Micro and Macro and the Integration of Qualitative and Quantitative Methods, *Forum : Qualitative Social Research*, vol. 2, n° 1, février, <http://qualitative-research.net/fqs/fqs-eng.htm> (accédé le 11/02/01).
- KELLE U. (1995), *Computer-Aided Qualitative Data Analysis: theory methods and practice*, Sage Publications, Londres
- LEBART L., SALEM A. (1994), *Statistique textuelle*, Dunod.
- LEBART L., SALEM A. (1994), *Statistiques Textuelle*, Dunod.
- LEGEWIE H. (1994), Globalauswertung, in Böhm T., Muhr T., Mengel A., *Texte Verstehen : Konzepte, Mehtoden, Werkzeuge*, Universitätsverlag, Konstanz, pp. 100-114.
- MILES M. B. (1979), Qualitative Data as an Attractive Nuisance : The Problem of Analysis, *Administrative Science Quarterly*, vol. 24, n° 4, pp. 590-601.
- MILES M. B., HUBERMAN A. M. (1994), *Qualitative Data Analysis : An Expanded Sourcebook*, Sage Publications, Thousand Oaks.
- RAGIN C. C. (1987), *The Comparative Method : Moving beyond Qualitative and Quantitative Strategies*, California Press
- RICHARDS T.J., RICHARDS L. (1998), Using Computers in Qualitative Research, in Denzin N.K., Lincoln Y.S. [ed.], *Collecting and Interpreting Qualitative Materials*, Londres, pp. 211- 245, Sage Publications.
- SILEM A., MARTINEZ G. (1983), *Information des salariés et stratégies de communication*, Editions d'Organisations.

- STRAUSS A., CORBIN J. (1990), *Basics of Qualitative Research : Grounded Theory Procedures and Techniques*, Sage Publications, Newbury Park.
- THIÉTART R.-A. [coord.] (1999), *Méthodes de recherche en management*, Dunod, Paris.
- WACHEUX F., (1996), *Méthodes qualitatives et recherches en gestion*, Economica.
- WEBER R. P. (1990), *Basic Content Analysis* (Sage Publications, University Paper Series on Qualitative Applications in the Social Sciences, 07-049), Sage Publications, Newbury Park.
- YIN R. K. (1989), *Case Study Research : Design and Methods*, Sage Publications, Londres.