

## **(En) - quête de l'autre: recherches qualitatives et corpus numériques**

**Huguette Rigot**

Fondation de la Maison des Sciences de l'Homme / Institut national de  
recherche Pédagogique

### **Résumé**

Les méthodologies qualitatives construisent leurs résultats à partir d'un double handicap. Tout d'abord, elles sont critiquées car elles mettent au centre de leurs analyses la relation enquêteur-enquêté. Cette relation, rendue visible à partir des échanges langagiers, des observations, etc. place au cœur de l'analyse produite la communication interpersonnelle qui cimente le lien social. Ce sont donc les mêmes productions sociales qui construisent la cohérence sociale et celle de l'enquête. Ensuite, ces données communicationnelles font trop peu l'objet d'un traitement rigoureux sur les plans de la collecte et de la conservation, situant ainsi la temporalité des enquêtes qualitatives dans l'immédiateté et l'éphémère. Les données issues de la société civile doivent faire l'objet de mesures spécifiques de collecte, de conservation pour constituer des corpus réutilisables par d'autres chercheurs autorisant ainsi comparaison et cumulativité. La réflexivité des recherches qualitatives se construit aussi par la préservation des données d'enquête.

### **Mots clés**

CORPUS D'ENQUÊTE – PROCÉDURES DE COLLECTE – ARCHIVAGE NUMÉRIQUE –  
DONNÉES QUALITATIVES

### **Introduction**

Les méthodologies qualitatives et les enquêtes qu'elles produisent font l'objet de jugements paradoxaux. D'un côté, elles représentent une part importante des recherches menées sur la société civile à cause de la reconnaissance de la richesse interprétative qu'elles apportent et de l'autre, elles sont dévalorisées au profit des analyses quantitatives, le chiffre et la mesure semblant être seuls susceptibles de conduire à des généralisations fondatrices d'éléments théoriques. De plus, cette relative incapacité à la généralisation s'accompagne d'une administration de la preuve qui est si différente de celle pratiquée dans les sciences exactes que se pose la question de son effectivité. Ces méthodologies font donc l'objet d'une double marginalisation, par rapport au quantitatif et par rapport aux sciences exactes.

Néanmoins, nombre de chercheurs en sciences humaines et sociales, par exemple des sociologues (Céfaï, 2003), des spécialistes des sciences de l'information et de la communication (Mucchielli, 2004) ont rassemblé, soit sous forme de textes choisis, soit sous forme de dictionnaire de notions, des matériaux théoriques permettant de caractériser la spécificité des méthodologies qualitatives dont l'objet est de comprendre les phénomènes sociaux à travers les réalisations langagières des individus et à travers l'observation et la description de leurs pratiques. Ces chercheurs, comme tous ceux pratiquant les méthodologies qualitatives, apportent la preuve, par leurs travaux, que la connaissance des phénomènes sociaux passe obligatoirement par le contact, la communication avec l'objet étudié, relation unique, non reproductible et non mesurable.

Dans cette filiation d'élaborations méthodologiques, si les résultats d'enquête sont souvent publiés et cités et si les éléments théoriques construits sont rassemblés et compilés sous forme de manuels universitaires, les corpus de données d'enquête sont à de très rares exceptions accessibles en même temps que les résultats qu'ils ont permis de produire. Parmi les exceptions, il faut citer *La misère du monde* de Pierre Bourdieu (1993), ouvrage composé d'entretiens précédés d'un texte décrivant le contexte de l'interview et un fragment de vie de la personne interviewée, avec cette précision et ce ton qui n'est sans rappeler le *style sociologique* de Richard Hoggart. Plus près de nous, *La Culture des individus* de Bernard Lahire (2004) fait alterner, dans des formats différents, textes de présentation documentant les entretiens et morceaux choisis plus ou moins longs des entretiens effectués. Le volume important de pages de ces ouvrages (plus de sept cents pages) montre d'une part la préoccupation de ces auteurs à restituer des données d'enquête à la fois auprès des enquêtés et d'un public de lecteurs, sociologues ou non et d'autre part le crédit de ces scientifiques auprès de leurs éditeurs, étant donné le risque financier lié à la production d'ouvrages scientifiques volumineux.

Ces deux exceptions scientifiques et éditoriales interrogent directement le traitement réservé aux données qualitatives dans la grande majorité des enquêtes, où les entretiens, les remarques, les notes issues d'observation, c'est-à-dire le corpus des données a disparu, absorbé dans les résultats publiés et n'apparaissant que sporadiquement sous forme de citations remplissant des fonctions diverses (illustration, étayage de preuves, etc.) par rapport au texte rédigé.

Ainsi, aborder une réflexion sur les corpus de données qualitatives, implique tout d'abord d'analyser les modalités de leur constitution, puis de comprendre le type de relation s'établissant entre les chercheurs et les données

qualitatives produites et enfin d'essayer d'évaluer l'impact que de la comparaison et la pratique du regard inter et pluridisciplinaire peuvent avoir sur le statut épistémologique des sciences humaines et sociales.

### **Constituer des corpus numériques de données qualitatives**

#### *Corpus de données quantitatives vs qualitatives*

*Stricto sensu*, il s'agit de données d'enquêtes qualitatives. D'évidence, cela suppose d'écarter les données quantitatives. Pourtant, caractériser rapidement l'actuelle et nouvelle situation de celles-ci en France... leur traitement en vue d'une accessibilité, n'est peut-être pas sans rapport avec le sort fait aux données qualitatives.

La situation des données quantitatives a été récemment et peut-être momentanément réglée par le Rapport de Roxanne Silberman (Silberman, 1999) du Lasmas, laboratoire du CNRS (Centre National de la Recherche Scientifique), qui a constaté que les données publiques car produites avec des financements publics n'étaient pas accessibles directement aux chercheurs individuels et aux laboratoires. Les grands producteurs de données quantitatives comme l'INSEE (Institut National de la Statistique et des Études Économiques) font payer leur utilisation. Ainsi est né le réseau Quetelet dont la fonction est d'archiver les données des différents producteurs de la Statistique publique à des fins de diffusion et de réutilisation pour la recherche et la formation. Cette utilisation fait l'objet d'une réglementation stricte, pour écarter les réutilisations purement commerciales et se fait selon une procédure précise et à partir d'un engagement écrit des utilisateurs. Le problème de la préservation pour communication des données quantitatives semble réglé avec la création de l'ADISP (Archives de Données Issues de la Statistique Publique, <http://www.cmh.ens.fr/acsdm2/adisp.php> [consultée le 4 août 2006]) qui reprend le fonds des grandes enquêtes de l'INSEE, du CEREQ (Centre d'Étude et de Recherche sur les Qualifications), des directions statistiques des ministères de l'Éducation nationale, de l'Emploi et des affaires sociales. Cette politique de création d'archives publiques et d'encouragement à l'archivage a le mérite de dynamiser les recherches quantitatives. La plupart des laboratoires et des chercheurs, avant la création du Centre Quetelet, étaient obligés de se contenter de citer des données, sous forme de schémas, tableaux, etc., souvent reprises de publications précédentes, faute de crédits pour payer l'accès à ces données publiques. Ainsi, le lecteur averti avait quelque peu l'impression de lire les mêmes données chiffrées d'un ouvrage à un autre, sans avoir vraiment l'impression d'une réutilisation critique. Mais pour critiquer les sources... faut-il y avoir accès.

### Types de corpus numériques

Que ce soit dans les rapports Français ou émanant d'autres pays, quand on évoque les corpus numériques de données qualitatives, il est quelquefois difficile de savoir de quel type de données il s'agit. De façon, assez large, cinq types de données différents sont concernés.

- a) Des revues en lignes mettant à la disposition de lecteurs sur le net des résultats d'enquête publiés auxquels on accède, dans la majorité des cas, par abonnement. Il faut citer le travail fait par le consortium Couperin, en France, pour négocier les prix des abonnements de revues par bouquets, aux niveaux des bibliothèques universitaires.
- b) Des revues en ligne accessibles librement et pouvant être regroupées sur des portails thématiques ou disciplinaires.
- c) Des Archives ouvertes, disciplinaires ou institutionnelles, mettant en libre accès des textes finalisés à des degrés divers (textes déjà publiés dans des revues papier, working paper, etc.) de chercheurs.

Dans ces trois exemples, aucune différence n'est faite entre recherches qualitatives et quantitatives, puisque la mise à disposition sur le réseau ne porte que sur les résultats publiés. La politique de numérisation de ce type de textes fait l'objet actuellement d'une attention particulière de la part du CNRS à la fois avec sa plate-forme Hal (portail commun de publications scientifiques) et du Centre d'édition numérique scientifique, inscrit dans le TGE ADONIS (Très Grand Équipement pour Accès aux données et aux documents numériques des sciences humaines et sociales). La mise en place de cette nouvelle politique est liée au constat que 2 % des publications françaises dans ce domaine sont présentes sur le net visible, alors qu'elles représentent plus de 20 % de la production mondiale. Il faut certainement reconnaître dans cette prise de conscience un des effets du *syndrome de l'Université de Shanghai* et de son classement des universités mondiales, qui traite si mal la recherche française.

d) Des résultats d'enquête n'ayant pas fait l'objet de publication. C'est le cas de nombreux rapports rédigés à la demande de ministères et d'institutions publiques et composant pour partie ce qu'on appelait, il y a encore quelque temps, la « littérature grise ».

e) Des matériaux d'enquête (entretiens, notes de terrain, etc.). Même si le mode d'accessibilité souhaité est le net, l'état des données est à ce point préoccupant que le simple repérage de ces ensembles documentaires et leur consultation dans des dépôts d'archives publiques correspondraient déjà à une révolution dans le traitement des matériaux qualitatifs.

Le rapport de Françoise Cribier de 2003 en ligne actuellement sur le site du Lasmus (<http://www.iresco.fr/labos/lasmus/rapport/Rapdonneesqualita.pdf>, [consulté le 15 juin 2006], est particulièrement alarmant sur la situation de ces données en France, d'autant qu'en annexe se trouve une comparaison avec les politiques mises en place dans d'autres pays. L'exemple le plus intéressant à étudier est le site Qualidata anglais, (<http://www.esds.ac.uk/qualidata/about/introduction.asp>), [consulté le 4 août 2006]. D'autres rapports indiquent aussi clairement que les matériaux des enquêtes qualitatives sont à préserver en vue de leur communication. Parmi les principaux textes, il convient de citer : le rapport de Georgette Elgey avec la collaboration d'Annette Wieworka (2001) portant sur l'ensemble des archives sonores de diction, celui rédigé sous la direction de Claude Dubar (2001), s'intéressant aux entretiens d'enquête, le rapport rédigé par Marie-France Calas en 2001 sur la conservation des archives sonores et pour finir sur la même thématique et enfin le rapport sur les bonnes pratiques pour le traitement des corpus oraux, par Olivier Baude (2005).

Ces différents rapports abordent les matériaux qualitatifs du point de vue de leur commanditaire. L'état des archives sonores de diction - une place importante a été faite aux archives radiophoniques et télévisuelles - a d'abord été analysé, puis celui de l'archivage des entretiens faits à l'occasion d'enquêtes qualitatives et enfin, la conservation des fonds sonores a été abordée. En fait, il ne s'agit nullement d'une progression dans l'appréhension des supports et des données, mais la rencontre un peu fortuite de trois types de préoccupations qui ne demandent qu'à être articulées. Indiquer ce contexte est utile pour comprendre comment pourrait se mettre en place un partenariat interinstitutionnel ayant pour objectif de traiter des données d'enquêtes qualitatives. N'oublions pas qu'un entretien qualitatif peut être étudié du point de vue de l'usage de la langue et / ou de l'information scientifique qu'il a contribué à produire. De plus, sa communication peut se faire à partir d'un support sonore et / ou d'une transcription qui nécessite la maîtrise de procédures de codage ou de « nettoyage » par les chercheurs.

### **Principes de constitution des corpus**

Les modalités de constitution des corpus d'enquête se situent dans un contexte juridique, scientifique et institutionnel dont il faut connaître les principes d'organisation.

Quelques axes directeurs ressortent des conclusions des différents rapports sur les données qualitatives et les archives orales.

Les aspects juridiques constituent, pour l'instant, le domaine où il est à la fois relativement aisé d'intervenir en informant les chercheurs sur ce qui

constitue non seulement de *bonnes pratiques*, mais des usages obligatoires à respecter tant pour la constitution de ces corpus que pour leur communicabilité.

Il faut remarquer que si ces aspects ont une spécificité certaine, car se référant à des textes de lois précis, ils sont difficilement séparables des dimensions disciplinaires et institutionnelles et de ce fait, c'est l'articulation de ces différentes modalités qui sera évoquée.

La question juridique est un bon point d'entrée pour analyser les pratiques de constitution de corpus car elle est paradoxalement mal connue et représente un obstacle important et commun à tous les chercheurs et à toutes les disciplines. Les aspects juridiques portent sur les questions de droits moraux et patrimoniaux et de propriété des données : ils se retrouvent donc présents à toutes les étapes du travail sur le corpus.

Au niveau de l'étape du recueil des données et de l'enregistrement, les principaux problèmes portent sur le droit à l'image pour les enregistrements vidéos, le droit à la voix pour les enregistrements sonores et les autorisations nécessaires à l'obtention d'un consentement *éclairé* de la part de l'interviewé. Ce dernier point est particulièrement intéressant à analyser puisqu'il articule respect du droit des personnes interviewées et finalité de l'enquête. Demander une autorisation écrite aux interviewés est nécessaire pour que les données recueillies et leur analyse puissent être exploitées. Cet accord est obtenu sur la base de l'explicitation de la finalité de l'enquête et de la garantie de l'anonymat, Ces deux points, qui semblent relever du sens commun partagé, posent dans le détail de certaines enquêtes des problèmes juridiques. Une autorisation accordée dans le cadre de l'exploitation d'une enquête particulière couvre-elle les droits d'usage pour des possibles réutilisations de ces données dans des enquêtes futures dont on ignore la finalité. Ainsi, émerge une forte contradiction entre le fait d'obtenir un consentement éclairé donc limité et le souci actuel de conserver pour mettre à la disposition des chercheurs des données patrimoniales. D'autre part, l'anonymat est toujours lié au contexte de réception de l'enquête, une personne connaissant bien l'interviewé peut le reconnaître à des détails de sa vie privée qui font partie intégrante des informations recueillies, alors que le nom, l'adresse ont été masqués ou remplacés. Devoir faire disparaître ou opérer des transformations au niveau de ces éléments est souvent impossible ou rend les données incompréhensibles par rapport à leur contexte de production. Enfin, dernier point à envisager : le mode de recueil des données : s'agit-il de *données sollicitées*, comme dans le cas de situations expérimentales, de données de *parole continue* recueillie dans des situations plus ou moins provoquées par le chercheur. L'entretien peut être considéré comme une situation provoquée par les éléments qui le constituent

comme la prise de rendez-vous, la demande de consentement d'exploitation des données, l'explicitation des finalités de l'enquête, etc. Par contre, l'enregistrement d'une réunion relève d'une situation non provoquée par le chercheur. Dans ce cas, la présence de l'observateur et les dispositifs réglementaires modifient les conditions de l'interaction étudiée. Dernier type de parole recueillie, *le tout venant*, vise à réunir des corpus de *parole en interaction*, dans des situations non provoquées. Dans ce cadre aussi, le respect des dispositifs réglementaires est difficile à tenir.

Au niveau de l'utilisation et de l'exploitation informatisée des données deux problèmes se posent principalement. Le premier concerne l'exploitation informatique de ces données qui se trouve relever de la loi Informatique et liberté. Le deuxième problème porte sur la notion d'instances propriétaires. Qui est l'auteur des données : l'interviewé, l'interviewer et/ou l'institution pour laquelle il travaille ? De plus, les corpus oraux sont rarement exploités sous leur forme d'origine, sauf pour certains chercheurs en phonologie, il est donc nécessaire de les transcrire. Là aussi, les pratiques suivant les disciplines et les chercheurs sont extrêmement variées. Les types de transcription sont à la fois dépendant de la dimension des corpus et de la situation disciplinaire de l'enquête. Au niveau des dimensions des corpus, plusieurs ensembles d'unités de temps sont identifiés. Les très petits éléments de corpus (d'une moyenne de quatorze secondes) sont principalement utilisés dans la recherche appliquée aux conversations téléphoniques. Les pauses, les chevauchements de parole en sont les principaux éléments structurant. Les petits éléments durent en moyenne dix minutes, ils sont la base des analyses de conversation, de façon à apprécier les rites d'entrée et de sortie de l'interaction. Les grands corpus correspondent généralement à la situation d'entretien.

Les procédures de transcription sont évidemment très différentes selon la dimension et la finalité de la recherche. Hormis dans le domaine linguistique, journalistes, historiens et sociologues ont des pratiques de transcription qui s'apparentent plus à la transposition ou à l'adaptation. Ils favorisent le contenu par rapport à la forme, supprimant répétitions, hésitations, tout ce qui peut gêner la relecture des données écrites, notamment pour l'interviewé. Les linguistiques cherchent à conserver les particularités des productions orales, alors que la plupart des autres chercheurs les font disparaître. La transcription fidèle de la parole des interviewés provoque souvent son rejet, les fautes de français, les répétitions, les hésitations, c'est-à-dire ce qui relève directement de la forme parlée constituant un obstacle à l'auto-analyse du contenu de l'entretien. Ainsi, la forme donnée par la transcription a une incidence directe sur le rendu des données et sur leur analyse auprès des publics interrogés comme auprès des publics qui y accèdent sans avoir participé à l'enquête.

Les aspects juridiques concernent aussi la diffusion, la mise en circulation des données : le droit de citation et le droit de diffusion doivent être respectés.

Au niveau de la conservation des données, l'articulation des dimensions juridiques, scientifiques, et institutionnelles est aussi très forte puisqu'il faut décider à la fois de ce qu'il convient de conserver, de l'instance responsable de la pérennisation et de la limitation possible de l'accessibilité.

De fait, la mise en place de données textuelles ne peut se faire qu'à partir de corpus d'enquête, corpus qui sont documentés, situés, finalisés et qui ne sont ni *sac de mots*, ni *archives de textes*, selon l'expression de F. Rastier (2005) et dans l'objectif de permettre à des chercheurs de construire de nouveaux corpus pour des enquêtes à venir. Les bases de données textuelles correspondent à des réservoirs de matériaux qualitatifs à l'intérieur desquels les chercheurs peuvent construire leur corpus selon leurs propres objectifs et la finalité de leur enquête. D'un côté donc un réservoir, de l'autre, un objet construit pour un usage spécifique. La dimension institutionnelle prend ici toute son importance car c'est l'instance institutionnelle qui devient le garant juridique et informatique de la conservation, de la sélection et de l'accessibilité aux données.

Le *Guide des bonnes pratiques* (Baude, 2005) constitue une première étape de la réflexion à mener sur les modalités de constitution des corpus numériques par son entrée principale sur les aspects juridiques. Il est nécessaire de le compléter par des analyses plus fines et en situation des pratiques de recherche qualitatives. Il semblerait plus opportun de partir du travail de terrain et des questions posées et à résoudre par les chercheurs que de partir d'une formalisation des modalités de constitution des corpus à partir d'une approche disciplinaire. Interrogé sur ses pratiques et méthodologies d'enquête, chaque chercheur affirme sa particularité. Que faut-il entendre derrière cette déclaration ? La spécificité irréductible des recherches qualitatives et / ou les solutions adaptatives que le chercheur a dû mettre en œuvre pour mener à bien sa recherche et produire des résultats ?

### **Singularité et spécificité des relations chercheurs / données**

#### ***L'engagement du chercheur comme personne***

Pourquoi la grande majorité des chercheurs interrogés sur leurs pratiques de recherche répondent-ils que leur manière de procéder est tout à fait spécifique et n'est en rien exemplaire des recherches menées dans leur discipline ? Pourquoi cette déclaration de principe d'une singularité, alors que dans la rédaction de leurs résultats de recherche, ils n'ont de cesse de situer leurs discours dans un cadre théorique et de spécifier leur méthodologie et leur

démarche à partir d'éléments théoriques précis, permettant de contextualiser les idées qu'ils avancent ? Faut-il reconnaître une différence de perception entre l'ordre du faire, des savoirs en action et l'ordre de l'écrire, du rendre compte, de l'argumenter et de prouver ?

Ces questions qui dépassent de loin la problématique de la constitution des corpus de données d'enquête, indiquent néanmoins que les chercheurs ont en commun d'appréhender leur rapport au terrain et notamment les interactions provoquées avec des interviewés comme étant de l'ordre du singulier. Alors, pourquoi et comment passer du singulier des matériaux issus des situations d'enquête à la mise en commun et au partage de données qui portent la trace de l'expérientiel ?

Les différents rapports portant sur la conservation et l'accessibilité des données qualitatives parlent de données en déshérence, de données abandonnées dans des placards, des caves ou carrément perdues. De toute évidence, le passage de la singularité du terrain et de ses données descriptives au partage et à la communication de corpus de matériaux qualitatifs pour l'usage d'une communauté scientifique ne se fait pas, et à entendre certains chercheurs, il n'a pas à se faire ! Les données, non seulement ne se partagent pas, mais elles font peu l'objet de réutilisation par le chercheur lui-même.

Toutes les disciplines n'ont pas le même rapport aux données qualitatives.

Psychologues et ethnologues ont appris à conserver la parole d'autrui, parce que cette préservation fait partie intégrante de leurs pratiques de recherche. Pour les premiers, la conservation des entretiens cliniques constitue la base à la fois du suivi d'un patient et des comparaisons nécessaires à établir entre les différents cas cliniques. Les analyses produites par les psychologues sont de fait des études de type longitudinale nécessitant un ensemble d'actions spécifiques permettant, d'une part de conserver les traces des entretiens et d'autre part de les réutiliser pour comparaison plusieurs années après. Cette conservation n'est pas faite en vue d'une communication de ces matériaux, pour lesquels il peut y avoir difficilement un consentement éclairé à la divulgation à des tiers de leur contenu.

Pour les anthropologues, la dimension patrimoniale est différente, bien présente depuis les débuts de la discipline qui trouvent ses racines dans les récits de voyageurs. La prégnance des sentiments d'altérité, d'exotisme et d'éloignement domine et débouche sur la création de matériaux qualitatifs souvent volumineux - notes de terrain, carbets de bord, journaux - consignants à la fois des descriptions précises des phénomènes sociaux rencontrés et des réflexions personnelles des chercheurs. Ces matériaux, rassemblant le *ici et le là-bas* mettent au centre des préoccupations méthodologiques l'écriture de

l'anthropologue, tout en maintenant l'objet étudié à une distance telle que certains terrains n'ont pu être analysés qu'une fois. La préservation des matériaux d'enquête débouche, dans ce cas, nécessairement sur le partage, la discipline se construisant sur la comparaison des terrains et donc des matériaux descriptifs et analytiques qui en sont la trace. Le *Colloque Sources et Ressources pour les Sciences Sociales. Ressource complémentaire : Plaidoyer pour les archives orales ou comment faire des archives parlées des archives parlantes*. <http://www.ethno-web.com/evenements.php?action=archive&id=6&numeve=5> [consulté le 4 août 2006], est un exemple de cette obligation pratique et méthodologique de la conservation et du partage de l'écriture anthropologique.

Le rapport aux matériaux qualitatifs change totalement dans les enquêtes qui s'inspirent plus ou moins de la démarche sociologique. En fait, il est tentant de mettre en corrélation situation d'enquête, donc implication du chercheur comme personne dans la constitution et la restitution des matériaux, et traitement des matériaux qualitatifs permettant de les conserver, de les communiquer et donc de les partager avec d'autres chercheurs.

Cinq situations types d'enquête peuvent être déterminées, par l'articulation de deux paramètres. Ces situations d'enquête correspondant à des cas exemplifiés, définissent un continuum situationnel sur lequel les pratiques réelles trouvent leur place et non des situations qu'on peut opposer les unes aux autres. Le premier paramètre concerne le contexte qui va de l'expérimental à l'écologique, alors que le deuxième paramètre concerne le style d'intervention du chercheur qui se décline de l'observateur au participant à l'interaction.

La *situation expérimentale* est surtout mise en œuvre par certains psychologues avec un protocole précis supposant peu de relation entre l'observateur, le chercheur, et l'objet de la recherche, le cobaye. Le matériel issu des observations est généralement abondant et, sur le mode des sciences exactes, est communiqué aux autres équipes de chercheurs, accompagné de son protocole, de façon à pouvoir reproduire les conditions d'enquête et à vérifier et à comparer les résultats obtenus.

La *situation clinique* vient d'être évoquée plus haut et concerne les psychologues cliniciens. Quel que soit le type de thérapie et le mode d'interrelation entre le patient et son thérapeute et la plus ou moins grande implication de celui-ci dans la relation, les matériaux produits sont conservés, mais sans possibilité de communication à des tiers.

La *situation d'interaction verbale*, qui elle aussi connaît des degrés divers dans l'implication du chercheur – de l'entretien directif avec questions ouvertes ou fermées, à l'entretien compréhensif voire explicatif – se trouve être

à la base méthodologique de nombreuses enquêtes et produire des matériaux souvent abondants qui, jusqu'à présent n'ont pas été traités pour être conservés ni communiqués à d'autres chercheurs, ni même faire l'objet d'un retour auprès des interviewés. C'est dans ce type de situation que se déroulent la plupart des enquêtes qualitatives plus ou moins inspirées des méthodologies sociologiques.

*La situation écologique* se rapproche du travail, lui aussi évoqué plus haut, effectué par les anthropologues.

Enfin, *la situation d'observation* des pratiques langagières est le fait des linguistes.

Il est tout à fait intéressant de remarquer que c'est à partir des situations d'observation que les linguistes ont pu produire des éléments théoriques servant à définir un corpus, en développant récemment un nouvel axe de leur discipline, la linguistique de corpus. Par contre, les enquêtes dont la méthodologie de production et de traitement des matériaux qualitatifs est basée sur l'interaction plus ou moins forte entre individus interrogés et chercheur - celui-ci se trouve en situation de co-auteur ou de co-producteur des matériaux produits – les corpus sont peu conservés et accessibles. L'engagement physique, intellectuel dans les échanges lui fait considérer ceux-ci comme étant du domaine du singulier voire du particulier et du privé. Les données qu'il construit pour son enquête révèlent autant d'informations sur sa personne et son mode opératoire qu'elles en fournissent sur les enquêtés.

On peut faire l'hypothèse que l'absence de travail préparatoire suivant un protocole précis de construction d'un corpus, impliquant une attention particulière aux aspects juridiques, disciplinaires et institutionnels, renforce l'engagement personnel du chercheur dans les données produites, bloquant toute velléité de mise à distance *a posteriori* d'un corpus, en vue de le conserver pour le communiquer.

La mise à distance par rapport aux matériaux qualitatifs s'effectue par l'oubli et l'abandon des données produites au cours d'une enquête et leur non-réutilisation pour celles à venir. Ainsi, la linguistique de corpus, comme discipline d'observation des pratiques langagières et la sociologie, comme discipline d'interactions, possédant une forte valeur communicationnelle - et toutes les sciences sociales et humaines utilisant comme méthode de recueil de données l'entretien - se trouvent actuellement sur des positionnements opposés quant à la production de corpus de données d'enquête.

Cette opposition basée sur des modes d'approche spécifique des réalisations langagières permet de comprendre, d'une part pourquoi les données qualitatives sont à ce point « abandonnées » et servent d'argument à la dévalorisation des résultats obtenus par les méthodologies qualitatives et

d'autre part comment il est possible de modifier le statut épistémologique des sciences humaines et sociales en renouvellement le traitement des paroles recueillies auprès de la société civile.

### **Le désintérêt des communautés scientifiques et des instances institutionnelles**

Si à la demande de différentes institutions gouvernementales, en France, plusieurs rapports ont été produits, dans la dernière décennie, tous dénonçant le « contexte alarmant de l'archivage des sources orales » (Dubar, 2001), plusieurs niveaux de responsabilités peuvent être identifiés pour expliquer cette situation.

En premier lieu, il faut citer les communautés scientifiques elles-mêmes et le rapport qu'entretiennent leurs méthodologies avec la parole des interviewés. Il est bien évident qu'en sociologie, par exemple, le fait de pratiquer une sociologie critique mène à développer méfiance vis-à-vis des représentations exprimées par les acteurs sociaux et discours scientifique surplombant et à distance. Cette situation a fortement contribué à la mise en place de pratiques de recueil et de traitement des matériaux qualitatifs visant à accorder plus d'importance au discours sociologique, seul capable d'accéder à la réalité du social, qu'à la parole des interviewés. Par contre, l'ethnométhodologie repose, quant à elle, sur un autre type d'exploitation des matériaux qualitatifs qui met l'accent sur les descriptions effectuées par les acteurs et la volonté des chercheurs de les prolonger le plus longtemps possible, afin que leur analyse ne soit pas un dévoilement, mais une explicitation, une reformulation du travail réflexif mené par les acteurs sociaux. Dans ce cas, on peut supposer que la valeur accordée aux données d'enquête peut déboucher sur une volonté de les rendre transparentes et accessibles.

En deuxième lieu, les structures communautaires institutionnelles – laboratoires, équipes de recherche, etc. – ont aussi leur part de responsabilité dans cette déshérence des matériaux qualitatifs. Le manque de place pour le stockage des données, le manque de personnel formé à l'archivage des documents, le manque de budget pour pérenniser leurs activités de recherche, correspondent à des modalités de fonctionnement qui privilégient l'individuel par rapport au collectif, l'immédiateté par rapport à la durée.

En troisième et dernier lieu, la responsabilité est aussi au niveau des instances gouvernementales qui, tout en commanditant différents rapports, n'ont pas su mobiliser les chercheurs et leur donner les moyens techniques et budgétaires pour suivre les recommandations qu'elles avaient contribué à faire émerger. Il est intéressant de remarquer que les rapports ayant fait l'objet d'une publication - ce qui peut être considéré comme un indice de la reconnaissance officielle de leurs conclusions - concernent d'avantage les corpus oraux et les

problèmes de conservation liés aux supports d'enregistrement que les matériaux issus d'enquêtes qualitatives. Les linguistes ont relativement rapidement réagi à la constitution de corpus oraux, s'en servant pour configurer le nouveau champ de recherche de la linguistique de corpus. Par contre, la communauté des sociologues a peu réagi aux rapports de Françoise Cribier (2003) - disponible sur le réseau – et de Claude Dubar (2001), nulle part accessible.

### **Pour une évolution de la recherche qualitative**

L'état de ces données perdues, oubliées, non réutilisables, détruites donne une image dévalorisante des recherches qualitatives et du travail de terrain, tout à fait contraire à celle que les chercheurs veulent donner de leurs activités tant auprès du grand public qu'auprès des jeunes chercheurs qu'ils forment au traitement des données et au travail de terrain à grand renfort d'ouvrages méthodologiques.

De plus, cette situation installe le travail d'enquête dans une temporalité trompeuse, celle de l'immédiateté, qui est aussi celle de l'oubli. Le temps de l'enquête, celui nécessaire à la lecture des documents théoriques au travail de terrain (formulation des hypothèses, préparation des entretiens, passation, analyse) et à la rédaction des résultats de recherche se trouve fortement déconnecté de la temporalité nécessaire au déploiement des méthodologies disciplinaires.

Enfin, ces dernières se trouvent formalisées paradoxalement, non à partir du traitement des matériaux d'enquête, mais des résultats qu'ils ont contribué à produire. Il s'opère ainsi une inversion des causes et des conséquences qui aboutit à la réification des objets sociaux étudiés. Cette réification peut être, certes, un objectif scientifiquement revendiqué, comme dans le cas du travail de Durkheim fondant la sociologie comme étude des faits sociaux, elle peut être aussi une simple conséquence d'une lecture de l'enquête sociologique qui inverse l'ordre des discours, plaçant la parole des interviewés sous la tutelle autorisée et légitimée du chercheur.

Si certains chercheurs, individuellement sont conscients et sensibilisés à ces problèmes de conservation et d'accessibilité des données d'enquête qualitatives, pour l'heure aucune prise de position collective et disciplinaire n'est encore venue soutenir et confirmer les conclusions de F. Cribier. Il est intéressant de constater que, si celle-ci dresse un état des principales objections apportées à l'encontre des méthodologies qualitatives pour tenter d'analyser le désintérêt des communautés scientifiques pour la préservation des données d'enquête, les critiques faites à ces méthodologies ne les empêchent pas d'exister et de produire la grande majorité des résultats de recherche en

sciences humaines et sociales. Alors comment gérer le paradoxe qui fait de ce type d'approche un outil d'exception pour analyser la société civile tout en ne préservant pas ce qui en constitue les fondements empiriques ?

Comment et pourquoi transformer les pratiques scientifiques ? Pourquoi conserver ces données ? Tout d'abord, elle constitue les archives de la société civile. Les corpus d'enquête ont une valeur patrimoniale, ils sont uniques et irremplaçables, leur destruction implique une perte qui ne peut être compensée par une autre enquête. Les paroles des interviewés recueillies et perdues ont une valeur singulière. Il est absurde et irresponsable tant aux niveaux scientifique et budgétaire de penser pouvoir les remplacer par d'autres.

De plus, la réutilisation de ces matériaux permet de faire perdurer le travail interprétatif. La lecture par un autre chercheur au moment de l'enquête ou plus tard ouvre sur de nouvelles explications, ne serait-ce qu'en rendant possible la comparaison avec d'autres travaux de la même époque ou plus récents.

Cette réutilisation, en permettant de faire retour sur les matériaux, sur les méthodes et sur les conditions de production des enquêtes, pourrait servir d'une part à construire une autre histoire des champs disciplinaires et à développer la pluridisciplinarité, donnant ainsi la capacité aux sciences sociales d'explorer ensemble des objets communs et de fédérer les savoir-faire, de capitaliser les résultats, pour les revisiter, pour les réinterpréter.

Enfin, il serait possible de rendre compatibles et complémentaires les recherches qualitatives et quantitatives. Le travail accompli, il y a peu, sur la conservation et l'accessibilité des données quantitatives doit ouvrir la voie à un travail de même type pour les données qualitatives. La recherche en sciences humaines et sociales ne peut continuer à se satisfaire de l'opposition existant entre ces deux approches, alors que sur le terrain, de nombreux chercheurs utilisent les deux méthodologies.

Il est désormais évident que l'ère du numérique construit une relation différente à l'écrit. Les changements ont une dimension à la fois sociale, cognitive et sémiotique. Pourtant, assez paradoxalement, la plupart des pratiques de recherche en sciences humaines et sociales ne sont concernées qu'à la marge par l'évolution ou la révolution numérique. Ce rapport différent à l'écrit, aux textes et aux documents bouleverse assez peu les pratiques de diffusion des données et des résultats d'enquêtes qualitatives. Si aucune difficulté technique n'empêche, actuellement les chercheurs de mettre sur le réseau non seulement leurs textes plus ou moins finalisés dans des Archives ouvertes institutionnelles ou communautaires, mais aussi les matériaux qui ont permis leur production, pourquoi cet ensemble de données (données

qualitatives et analyses) est-il encore inaccessible sur le réseau ? Qu'est-ce que *La Misère du monde* a donné à voir de particulier du travail sociologique ? La parole des interviewés ! Celle-ci a été considérée comme un matériau ayant une valeur suffisamment forte pour qu'elle soit diffusée auprès des lecteurs, sociologues et les non-sociologues. Or, ce type de matériau peut être rendu accessible sur le réseau, sous certaines conditions juridiques.

Si ce type de démarche peut être reproduit, qu'apporte-t-il au travail des chercheurs ? Un début de réponse se trouve exprimé par les chercheurs eux-mêmes. La préface de *Casseroles, amour et crises* (2006) de Jean Claude Kauffman fournit un exemple intéressant. Cet ouvrage se termine par un chapitre consacré à la méthode et aux données biographiques des personnes interrogées. Or, le sociologue fait un aveu d'importance : peu de questions dans les entretiens effectués concernaient les repas qui, lors de la phase de dépouillement, se sont révélés être une thématique de première importance, montrant le lien existant entre la cuisine et *la fabrique de la famille*. Deux solutions s'offraient à lui : relancer une seconde campagne d'entretiens ou à partir d'un travail documentaire, découvrir des enquêtes plus anciennes, traitant des repas et ainsi réutiliser leurs matériaux. « Il aurait été inconcevable de les ignorer. Il aurait été aussi inconcevable de ne les citer qu'allusivement, sans restituer un peu de leur chair la plus concrète et vivante » (Kauffman, 2006, p. 303-304). Ces deux exemples apportent quelques arguments en faveur de l'accessibilité des données d'enquêtes qualitatives. J.-C. Kauffman, dans son texte, avance d'autres raisons. Procéder par entretiens compréhensifs est une méthode incomparable pour rendre compte des processus sociaux et qui impose une grande souplesse pour ne pas prendre le risque de voir les idées du chercheur orienter les conclusions. Si la principale façon d'échapper à ce biais est bien identifiée et réside traditionnellement dans l'auto-contrôle, une autre est moins pratiquée, du fait de l'actuelle organisation de la recherche. Cette autre solution consiste à se confronter avec d'autres sources sur le même sujet. « Il n'existe pas deux chercheurs qui traitent leur matériau exactement de la même manière et dans la même optique. Croiser les regards s'avère très précieux. » (Kauffman, 2006, p. 304)

Ces exemples conduisent à faire l'hypothèse que rendre accessibles sur le net les données d'enquêtes qualitatives devrait ouvrir à de nouvelles perspectives les méthodes qualitatives. Cette action permettrait d'une part de faire (re)connaître leur valeur des matériaux de base par les acteurs des sociétés civile et politique et par ceux de la communauté scientifique et d'autre part, d'améliorer les « performances » épistémologiques de ces enquêtes par la pratique de la comparabilité des données.

C'est aussi faire l'hypothèse que l'actuelle situation de *privatisation* des matériaux qualitatifs a une incidence directe sur le statut épistémologique des sciences humaines et sociales conduisant à leur dévalorisation par les autres types de sciences. Statut de disciplines dominées et désintérêt relatif des chercheurs en sciences de la culture, pour constituer des bases de données textuelles, sont liés et s'expliquent principalement par le fait que la plupart des chercheurs sont peu formés à l'utilisation et à la production de corpus numériques et qu'ils entretiennent une relation d'intimité, de personnalisation avec les données d'enquête.

En abordant le traitement des matériaux qualitatifs, on touche directement à la méthodologie des enquêtes qualitatives. Opérer des changements a donc une implication sur les pratiques d'enquêtes et sur la façon d'évaluer leurs procédures et leurs résultats.

En reconnaissant et en traitant spécifiquement de l'aspect communicationnel dans la recherche qualitative, on reconnaît l'engagement du chercheur et on l'intègre dans le processus de recherche. Ainsi, le partage des données qualitatives peut être accepté et l'interprétation démultipliée.

Ainsi, les sciences humaines et sociales pourraient ne plus être épistémologiquement et institutionnellement des disciplines dominées.

### Références

- Baude, O. (2005). *Le Guide des bonnes pratiques pour la constitution, l'exploitation, la conservation et la diffusion des corpus oraux*. Rapport à la Délégation générale à la langue française et aux langues de France. Ce guide vient d'être édité sous le nom de *Guide corpus oraux* (2006), Paris : Ed. du CNRS.
- Bourdieu, P (Éd.). (1993). *La Misère du monde*. Paris : Seuil, 956 p.
- Calas, M-F. (2001). Rapport pour le Ministère de la culture et de la communication aborde les problèmes de politiques de conservation et de valorisation menées en France pour l'ensemble des fonds sonores.
- Céfaï, D (Éd.). (2003). *L'Enquête de terrain*. Paris : La Découverte. 621 p.
- Consultation nationale sur les archives de résultats de recherche*. (2001). Conseil de recherches en sciences humaines du Canada. Phase un : rapport d'évaluation des besoins. [http://www.sshc.ca/web/about/publications/da\\_phase1.pdf](http://www.sshc.ca/web/about/publications/da_phase1.pdf) [consulté le 4 août 2006]
- « Corpus et recherches linguistiques » (2002). *Corpus*, 1, novembre 2002.
- Cribier, F. (2003). *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »*. Rapport présenté au Ministère délégué à la Recherche et aux nouvelles

technologies.

<http://www.iresco.fr/labos/lasmas/rapport/Rapdonneesqualita.pdf>,

[consulté le 15 juin 2006].

Dubar, Cl, (2001). *L'archivage et la capitalisation des entretiens en sciences sociales*. Rapport remis au CNRS-SHS.

Elgey, G. (2001). Rapport remis au Conseil économique et social sur les archives sonores de diction.

Kauffman, J-C. (2006). *Casseroles, amour et crises. Ce que cuisiner veut dire*. Paris : A. Colin. 342 p.

Lahire, B. (2004). *La Culture des individus. Dissonances culturelles et distinction de soi*. Paris : La Découverte. 777 p.

Mucchielli, A (eds.). (2004). *Dictionnaire des recherches qualitatives en sciences humaines*. Paris : A. Colin. 304 p.

Rastier, F. (2005). « Enjeux épistémologiques de la linguistique de corpus » dans *La Linguistique de corpus*, sous la dir. de Geoffroy Williams. Rennes : Presses universitaires de Rennes. P. 31.

Silberman, R. (1999). Rapport au Ministre de l'Éducation nationale, de la recherche et de la technologie. *Les sciences sociales et leurs données*. <http://www.education.gouv.fr/rapport/silberman/table.htm>

**Huguette Rigot** est enseignant-chercheur en Sciences de l'Information et de la communication à la Fondation de la Maison des sciences de l'homme(Paris) et à l'Institut National de la Recherche Pédagogique. Ses actuels axes de recherche portent sur l'écriture des sciences humaines et sociales et plus précisément les actes bibliographiques des chercheurs en sciences humaines et sociales (SHS) et l'analyse des stratégies des acteurs (institutions, éditeurs publics et privés, bibliothèques, chercheurs, etc.) mises en œuvre pour construire une offre de diffusion numérique des revues en SHS.