

PIERRE LE QUÉAU

Un parcours interprétatif des données issues d'une question ouverte

Journal de la société française de statistique, tome 142, n° 4 (2001),
p. 121-138

http://www.numdam.org/item?id=JSFS_2001__142_4_121_0

© Société française de statistique, 2001, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

UN PARCOURS INTERPRÉTATIF DES DONNÉES ISSUES D'UNE QUESTION OUVERTE

Pierre LE QUÉAU *

RÉSUMÉ

L'étude des représentations sociales peut s'appuyer, depuis quelques années déjà, sur les techniques d'analyse statistique textuelle qui permettent de traiter des corpus volumineux recueillis au moyen de questions ouvertes. Outre le fait qu'elles en ont ainsi ouvert l'application à des textes significativement plus nombreux que ce à quoi on était accoutumé dans les enquêtes sociologiques, ces techniques ont largement contribué à clarifier un certain nombre des opérations entrant dans le cadre classique de l'analyse de contenu. Il reste toutefois que le passage entre l'analyse du discours et l'analyse sociologique n'est pas encore assez balisé. L'exposé présente un protocole d'exploitation des données issues d'une question ouverte articulant traitement statistique et analyse sémantique pour poser certains jalons vers l'interprétation.

Mots clés : Représentations sociales, analyse thématique de contenu, analyse lexicométrique, analyse lexicale, interprétation.

ABSTRACT

The study of social representations can now use techniques of statistical textual analysis that allow to treat large corpus collected with open-ended questions. Besides their applications to texts much more voluminous than usually in sociological surveys, these technics have contributed to set off and clarify several operations of the classic thematic contents analysis. But the transition from discourse analysis to sociological analysis remains yet opaque. The presentation set out how can be articulated statistical and semantic analyses to show a way toward interpretation.

Keywords : social representations, thematic contents analysis, lexicometric analysis, lexical analysis, interpretation.

1. Vers l'interprétation

Les usages qui sont faits de la question ouverte dans les enquêtes réalisées au CRÉDOC sont variés et se laissent mal catégoriser. Parmi ces différentes possibilités, il en est peut-être deux qui peuvent se révéler plus intéressantes que les autres en raison de l'intérêt du matériau qu'elles permettent de

* CSRPC/SHS/UPMF, 1251 av. Centrale, B.P. 47, 38040 Grenoble cedex 9. CRÉDOC, 142, rue du Chevaleret, 75013 Paris. E-mail : Pierre.Le-queau@upmf-grenoble.fr

recueillir et, donc, du traitement dont elles font alors l'objet. Une première catégorie de questions ouvertes contient celles qui permettent de recueillir les jugements et opinions qu'une personne peut librement formuler sur le sujet traité par un questionnaire. Un cas exemplaire est celui d'une enquête conduite auprès de 900 allocataires du RMI où, après une assez longue série de questions portant sur le parcours de la personne interrogée avant, pendant et après le dispositif, il s'est agi de lui demander son appréciation globale sur la mesure. L'exploitation approfondie de cette question a fait apparaître des réactions riches en affects qui ont éclairé sous un jour complètement nouveau les résultats de l'enquête (Aldeghi, 1998). Le second cas de figure comporte des questions ouvertes dont l'objet est d'explorer les représentations sociales qui informent *a priori* la perception que le public peut avoir de la problématique d'une enquête. C'est d'une telle question ouverte dont on va traiter dans le présent article. Son objet avait l'ambition d'éclairer sur les représentations de la santé et le cadre de la compréhension de cette notion par le grand public. Il s'agissait en outre de perfectionner la formulation d'une question fermée destinée à montrer les grandes structures de la perception de la santé. Les attentes liées aux résultats de cette question ont déterminé une attention particulière quant à son traitement et ce dont on se propose de rendre compte ici tient dans les principales étapes du parcours qui a été suivi pour exploiter les données. On insistera particulièrement sur ce qui constitue le moment crucial dans tout travail d'analyse de contenu, et le plus difficile à objectiver sans doute : celui de la compréhension du sens des réponses recueillies. Par « compréhension », on entend ici le double acte de connaissance, décrit par le sociologue M. Weber (1913), qui suppose de reconnaître la signification attribuée par son auteur à une « réponse » et d'établir des relations pertinentes entre cette « réponse » et l'« ensemble significatif » dans lequel elle s'inscrit. Le problème, comme l'ont fort bien explicité des philosophes comme H.-G. Gadamer (1960) ou P. Ricoeur (1965), est que l'acte de « restitution » du sens est avant tout un acte de « production » de sens. Cette proposition a fait les beaux jours de la critique qui renvoie souvent l'herméneutique à son « relativisme » tant il est vrai que ses arguments, pour fondés qu'ils soient en théorie, se déclinent assez mal sur le plan opératoire. Comment contrôler, en effet, l'affectation du sens ?... Le parcours interprétatif proposé, qui s'appuie sur des logiciels d'analyse des données textuelles, à défaut d'apporter des réponses définitives à de telles questions, tente de poser quelques jalons pour rendre moins opaque, à tout le moins, le passage du traitement statistique des données à leur interprétation.

2. Les contours de la santé

2.1. La problématique de la perception de la santé

L'expérience dont il va être question s'inscrit donc dans le cadre d'une réflexion menée depuis quelques années au sein du CRÉDOC visant à définir les meilleurs outils qui soient pour appréhender la perception qu'ont de leur santé les personnes interrogées dans divers types d'enquêtes. Cette thématique est

en effet de plus en plus présente dans les travaux destinés à mesurer les impacts des politiques sociales puisqu'un volet sanitaire accompagne désormais souvent les dispositifs d'aide à la réinsertion sociale et/ou professionnelle. Par ailleurs, d'autres travaux réalisés par le département Évaluation des Politiques Sociales ont plus spécifiquement porté sur une approche des comportements en matière de santé, qu'ils s'inscrivent ou non dans le cadre de l'évaluation de politiques publiques (Le Quéau & Olm, 1998).

De ces expériences, il ressort tout d'abord le constat de l'extrême intérêt que représente une approche des représentations de la santé pour comprendre le sens de certains comportements comme la consultation d'un médecin, les choix dans l'alimentation, la consommation de psychotropes, etc. L'approche des représentations de la santé peut se faire par plusieurs indicateurs mais, parmi ceux qui sont le plus fréquemment utilisés, il en est peut-être un qui dévoile mieux que les autres les enjeux de cette thématique. Il s'agit d'une question fermée, proposant un choix entre différentes conceptions possibles de la santé, qui a été construite à partir du résultat de différents travaux sociologiques (Herzlich, 1969; Vigarello, 1993)¹. L'exploitation de cette question, jointe à celle d'autres indicateurs, fait clairement apparaître un principe de structuration des représentations de la santé très fortement corrélé à certains *habitus*.

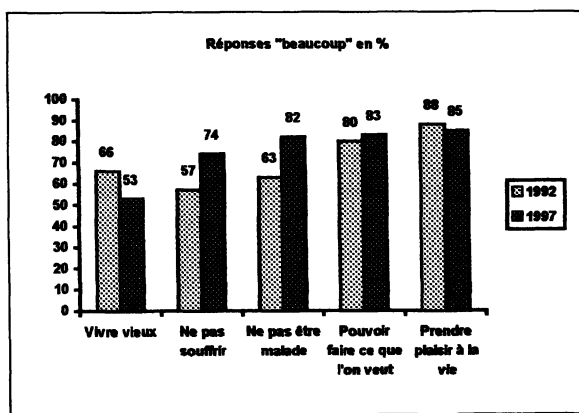


FIGURE 1. – La perception de la santé.

Schématiquement, la perception de la santé est organisée autour de deux pôles. Un premier ensemble de représentations, très positives, renvoie à ce que M. Foucault (1984) a jadis décrit comme un «souci de soi». La santé est conçue comme un capital de bien-être qu'il s'agit d'améliorer. Ce pôle

1. Intitulé exact de la question : « Pour vous, une bonne santé c'est avant tout... Prendre plaisir à la vie ? Pouvoir faire ce que l'on veut ? Ne pas être malade ? Vivre vieux ? Ne pas souffrir ? ». La réponse est formulée à partir d'une échelle d'accord : beaucoup, assez, peu ou bien pas du tout.

de représentations est logiquement associé aux items «prendre plaisir à la vie» ou bien «pouvoir faire ce que l'on veut». Un second ensemble est en revanche marqué par une aperception «en creux» de la santé dans la mesure où celle-ci se définit davantage comme l'absence de son contraire : la maladie. Ce pôle correspond assez bien, en principe, aux items «ne pas être malade» voire «ne pas souffrir». Ces deux univers de la santé apparaissent bien corrélés à la variable d'âge (les plus jeunes ont plus souvent une représentation positive de la santé encore qu'il ne s'agisse pas souvent pour eux de la préserver) et aux catégories socioprofessionnelles. Le «souci de soi» apparaît très caractéristique des catégories les plus aisées tandis que l'approche «négative» de la santé (comme «non-maladie») est beaucoup plus fréquente parmi les milieux populaires.

Ce qui est le plus intéressant dans cette observation, tient dans les liens qui peuvent ensuite être établis entre ces représentations et certains comportements de préservation ou de réparation de la santé. Certaines enquêtes montrent que, d'une manière idéale-typique, le «souci de soi» tend à favoriser les comportements de prévention. Une enquête menée dans le cadre de l'évaluation de la loi Évin, par exemple, a ainsi fait apparaître que le taux de tabagisme tend à décroître au sein des catégories les plus élevées de l'échelle sociale, particulièrement chez les femmes, tandis qu'il continue d'augmenter dans les couches ouvrières, surtout chez les femmes (Hébel et coll., 2000). Or le «souci de soi», et notamment celui de son apparence, y a été clairement identifié comme une composante importante du rejet du tabac.

Malgré l'intérêt de la question, elle ne répond que partiellement à toutes les interrogations qui portent encore sur le rapport qu'il convient d'établir avec un ensemble plus large de comportements sanitaires... sans parler des conditions de production et de reproduction de ces représentations (Pierret, 1996). Sur un plan opératoire, certaines données issues d'enquêtes, en particulier celles qui sont menées auprès de publics plus restreints que la population générale, montrent que les univers de représentations que permet d'identifier la question fermée ne sont pas toujours assez discriminants. Ainsi, au cours d'une enquête réalisée auprès des jeunes gens âgés entre 15 et 25 ans sur certains comportements à risque, les entretiens préparatoires avaient laissé penser que l'item «pouvoir faire ce que l'on veut» était insuffisant. S'il peut rendre compte de l'expérience de la plupart des jeunes, il ne permet pas de distinguer ceux qui s'imposent certaines limites de ceux qui n'observent aucune restriction dans la consommation qu'ils font de psychotropes tout spécialement (Le Quéau & Olm, 2000).

D'autres exemples de ce type qui montrent les insuffisances de la formulation de cette question pourraient encore être évoqués mais tous aboutiraient au même constat. C'est pourquoi la formulation de cette question a-t-elle varié d'une enquête à l'autre, selon le public auquel elle s'adressait. Cette interrogation, somme toute factuelle, conduit à un questionnement de fond qui porte sur les contours de la notion même de santé. Celle-ci renvoie, dans les faits, à des pratiques très diverses qui englobent non seulement des actes strictement médicaux comme la consultation d'un médecin mais aussi des comportements

touchant à l'hygiène de vie, l'alimentation, les loisirs, ou encore qui impliquent des choix concernant l'environnement, etc. Pour ces deux types de raisons, les contours de la santé perçue paraissent donc extrêmement variables et on dispose de peu de données d'enquêtes quantitatives pour se faire une idée plus précise de ce que recouvre concrètement et, si possible exhaustivement, cette notion. Pour commencer de répondre à ces interrogations portant sur le contenu des représentations de la santé et dans le but plus pratique de formulation d'un questionnaire fermé, il a donc été convenu de poser une question ouverte permettant aux personnes interrogées de livrer leur perception de la santé.

2.2. Méthodologie

La formulation retenue pour cette question a été extrêmement large. Une enquête récemment commandée par le Haut Comité de Santé Publique débutait donc, juste après la présentation de l'enquêteur, par la question suivante posée par téléphone à un échantillon national représentatif de 1 000 personnes âgées d'au moins 18 ans : « *quels sont tous les mots, images ou expressions qui vous viennent à l'esprit si je vous parle de santé ?* ». Il faut noter que cette question a donc été placée au commencement de l'interview : cette mesure avait principalement pour objet de préserver la spontanéité des réponses des personnes interrogées. On sait parfaitement ce que cette « spontanéité » peut avoir d'illusoire mais il s'agit plus modestement d'éviter les effets induits par le développement d'une thématique tout au long d'un entretien d'une vingtaine de minutes. Or la santé, comme le rappelle J. Pierret (1996), est un sujet « qui fait parler » (on a pu encore le vérifier puisque seulement 2 % des personnes interrogées n'ont pas répondu à cette question ouverte introductive) mais qui, lorsqu'il est abordé en tant que tel, suscite souvent une prise de parole marquée par un « discours public rationnel ». Il ne fait pas de doute que la situation d'une enquête téléphonique soit de nature à conforter la personne interrogée dans une attitude distanciée.

Le choix de la formulation de cette question a par ailleurs été déterminé en fonction des objectifs de l'expérience qui était dès le départ orientée vers une analyse thématique de contenu. On a donc accordé moins d'importance au développement d'une argumentation formelle qu'à la partie la plus « émergente » du raisonnement quotidien que chacun peut faire sur le thème imposé. La plupart des évocations recueillies ont donc valeur d'« images » qui condensent des significations, des affects et des expériences qui peuvent parfois paraître assez paradoxaux. Les réponses plutôt courtes formulées par la plupart des personnes interrogées (quatre mots en moyenne par réponse) ont par ailleurs largement facilité leur recueil fidèle par l'enquêteur².

2. L'enquête a été réalisée à partir d'un central téléphonique piloté par le système CATI. Toutefois, les réponses aux questions ouvertes ont été notées intégralement par les enquêteurs sur papier, puis ont été saisies dans les mêmes termes.

3. Classer

Le protocole d'analyse comportait donc deux grandes étapes dont la première consistait à effectuer certains classements sur le matériau recueilli. Elle a impliqué l'utilisation d'un logiciel, Alceste, au moyen duquel il est possible de regrouper entre elles les réponses ayant le même profil lexical³. L'hypothèse que l'on peut alors formuler consiste à supposer que les réponses classées dans le même ensemble, sur la base de leur vocabulaire commun, évoquent le(s) même(s) thème(s) qu'il s'agit ensuite d'interpréter, c'est-à-dire : de nommer sous un terme générique, sur un plan sémantique, et de resituer dans son (leur) « contexte significatif », sur un plan sociologique. Une classification descendante hiérarchique réalisée sur le corpus permet ainsi de produire 3 grandes classes, chacune étant caractérisée par un vocabulaire spécifique. Ces trois groupes permettent de classer 83 % des UCE⁴.

3.1. Les champs lexicaux des classes

La première classe regroupe 23 % des UCE classées et semble à première vue se caractériser par un ensemble de mots qui traduit une vision clairement positive de la santé⁵. Ces évocations correspondent assez bien à l'item de la question fermée « prendre plaisir à la vie ». Cet ensemble de réponses est cependant assez complexe dans la mesure où s'y confondent l'hédonisme et un certain holisme. La santé, en effet, est non seulement quelque chose qui renvoie à un état physique mais aussi moral, voire mental. Par ailleurs, elle est non seulement une disposition « intérieure » mais également un rapport à un certain environnement, tant naturel que social. En réalité, et en toute première analyse, la notion d'« équilibre », assez spécifique de ce champ lexical (cf. tableau 1, ci-après), paraît assez bien résumer les différentes idées contenues parmi ces réponses : la santé y est présentée comme le résultat d'une interaction entre des facteurs physiologiques et moraux, intérieurs et extérieurs

TABLEAU 1. – Vocabulaire spécifique de la première classe.

(listing Alceste)

forme+(45), joie+(33), sport+(37), phys+3(18), sain+(18), equilibre+(15), soleil+(12), vivre.(34), aliment<(25), plein+(15), air+(13), humeur+(9), repos(8), vacance+(6), boire.(8), mang+er(12), port+er(9), mora+l(9), bien-etre(41), campagne+(5), gaiete(6), liberte+(4), nature+(8), peche+(5), vie+(29), vitalite(8), voyage+(5), esprit+(3), etat+(4), genera+l(5), loisir+(4), plaisir+(2), fum+er(4), souffrir.(2), act+ion(7), bonne+(44), dynam+3(5), menta+l(3), eau+(3), mer+(3), nourriture+(6), stress(6), correct+(3), corps(7), tabac+(5), sentir.(3).

3. Pour une présentation un peu plus complète d'Alceste, voir par exemple dans ce journal Brugidou (2001).

4. Le texte d'une réponse est découpé en « Unités de Contexte Élémentaires » : UCE.

5. Rappelons que les mots sont réduits à leur racine (stemmatisation) d'autre part, le vocabulaire « spécifique » l'est au sens du Chi-2 et le chiffre entre parenthèses représente l'occurrence du mot dans la classe (cf. tableaux 1, 3 et 4).

UN PARCOURS INTERPRÉTATIF DES DONNÉES

Chacune de ces formes identifiées comme spécifiques par le programme peut être resituée dans le contexte d'une réponse, permettant ainsi de préciser la compréhension qu'on peut en avoir. En l'occurrence, Alceste extrait 19 réponses caractéristiques du contenu de la première classe, parmi lesquelles on trouve la suivante⁶ : « *Grand #air, #eau, blanc, #soleil, beauté, lumière, promenade, #nature, #forme, #dynamisme, #bonne #humeur, #joie, rires, travail, le travail c'est la santé* ».

La seconde classe regroupe 35 % des UCE classées. Elle réunit entre elles des évocations qui traduisent une appréhension manifestement plus problématique de la santé. Le mot le plus fréquent, et parmi les plus spécifiques de cette classe, est en effet « maladie ». Cette réponse est par ailleurs redoublée par un ensemble de termes qui désignent diverses affections. Parmi ceux qui sont caractéristiques de la classe, il est en effet question du « sida » (43 citations), du « cancer » (77), de la « tuberculose » (7), de la « grippe » (11), de « rhumatisme » (8), etc. Ce thème est par ailleurs associé à l'univers des soins : qu'il s'agisse des institutions chargées de la prise en charge des malades, des personnels soignants, des remèdes ou bien encore d'actes médicaux. La santé apparaît ici envisagée au travers de son envers, la maladie, dans un balancement précaire entre la « guérison » (15 citations) et la « mort » (18) ou le « décès » (5).

TABEAU 2. – Vocabulaire spécifique de la seconde classe.

(listing Alceste)

Hopita+l(148), maladie+(182), medecin<(164), medica<(124), cancer+(77), securite sociale(47), sida(43), docteur+(42), mort+(18), preventi+f(22), soin+(43), infirm+e(20), mutu+el(14), operat<(22), pharmac<(31), guerison+(15), pique+(11), vaccin+(14), hospitali<(9), blan+c(10), clinique+(8), dentiste+(8), recherche+(17), tubercul+eux(7), fol+(8), publi+c(9), vasculaire+(5), ambulanc<(9), analyse+(5), cure+(6), deces(5), fer+(5), grippe+(11), microbe+(6), rhumatisme+(8), rhume+(6), enf+ant(22), epidem+3(6), urg+ent(6), homeopathi+(5).

La troisième classe représente 42 % des UCE. La lecture des stemmes spécifiques de cette classe laisse difficilement comprendre ce qui, sur le plan thématique, pourrait en faire l'unité... s'il en est. La notion de pouvoir (le verbe à l'infinitif ou bien conjugué) est certes la plus importante quantitativement (en particulier associée avec d'autres verbes comme « aller », « rester », « marcher »...). Mais, par ailleurs, de nombreuses réponses traduisent une intention de valoriser la santé en la désignant, tout d'abord, comme « importante », « primordiale », etc. Mais la santé est également appréciée en termes quasi monétaires puisque sont associés à sa représentation des mots comme « riche », « richesse », « argent », « précieux », « cher », « moyens »...

Une lecture, même rapide, des premières réponses caractéristiques de la classe permet d'illustrer une des difficultés attendues liées à la formulation de la question, en même temps qu'une limite de l'analyse, lorsqu'un thème, comme celui de l'argent, est traité de façons différentes, voir opposées. La santé est en

6. Le signe # désigne les mots relevant d'un stemme spécifique de la classe.

UN PARCOURS INTERPRÉTATIF DES DONNÉES

TABLEAU 3. – Vocabulaire spécifique de la troisième classe.

(listing Alceste)

Pouvoir.(25), malade+(31), aller.(27), import+ant(23), meilleur+(11), primordia+l(13), soign+er(16), vouloir.(10), pouvoir+(10), faire.(20), rest+er(9), problem<(18), riche+(12), cher+(6), argent(10), bonheur+(24), chose+(10), moment+(6), arriv+er(6), fait(7), possi+ble(10), devoir.(3), falloir.(9), dire+(12), monde+(7), part+(5), esper+er(5), evit+er(5), march+er(6), profit+er(5), trouv+er(4), voir.(7), chanc+e(6), travail<(16), grand+(9), preci+eux(2), annee+(4), moyen+(4), niveau+(2), peau+(9), personne+(5), conserv+er(2), pens+er(4), prendre.(5), cardiaque(5), jeune+(4), gens(4), savoir+(3), occup+er(3), suivre.(3).

effet une « richesse » parmi les plus importantes, mais se garder en bonne santé, voire se soigner, suppose d'en avoir les « moyens ». Comme dans la première classe, en d'autres termes, la santé est donc perçue de manière plutôt positive (c'est « pouvoir »...) mais elle est ici mise au regard de son coût : *« la santé, elle s'en #va ; il y a beaucoup de #choses qui ont évolué, c'est une question d'argent, comme la fabrication #faite par des #grands groupes, les produits sont #chers, c'est pas légal, il-y-a beaucoup plus de modernisme mais je ne sais pas où on #va... »*.

3.2. La structure des champs lexicaux

La projection sur un plan des résultats de l'analyse (AFC) réalisée sur les formes lexicales fait mieux ressortir encore les différences entre les classes. Le premier axe oppose clairement les termes contenus dans la classe 2 aux autres, comme le montre également le dendrogramme de la classification. Cette opposition dessine une ligne de partage entre des réponses marquées, d'un côté (à l'est du plan), par une présence remarquable de mots désignant la maladie et son traitement et, de l'autre, des notions évoquant davantage la bonne santé. D'une part, la santé est donc envisagée comme l'absence de son contraire ; d'autre part, elle est décrite positivement en rapport avec tout ce qu'elle permet d'être, de faire ou d'avoir. Le second axe oppose entre elles les formes des classes 1 et 3 et, pour désigner ce qui les distingue, on peut pointer le fait que les réponses rangées dans la première (nord-ouest) se caractérisent par une évocation du plaisir ; tandis que les autres (sud-ouest) associent la bonne santé au travail. La notion d'« argent », absente du registre de la première classe, est par ailleurs très spécifique de la troisième. Les termes relevant de la première classe ont donc en commun la référence à un certain hédonisme que l'on ne retrouve pas dans la troisième classe davantage problématisée autour de thèmes plus concrets et matériels comme ceux du travail et de l'argent.

Les opérations de classement effectuées sur le corpus simplifient considérablement le travail qui reste à accomplir pour interpréter les données. Toutefois, au stade où nous fait parvenir cette première exploitation, on peut éprouver quelque difficulté à comprendre le « sens » des représentations qui sous-tendent les verbatims ainsi rangés en trois groupes. Ceux-ci apparaissent encore un peu trop larges, sans doute, et chacun d'entre eux contient une pluralité de thèmes qui n'en rend pas toujours très aisée l'appréhension. L'analyse peut

UN PARCOURS INTERPRÉTATIF DES DONNÉES

(listing Alceste)	
Cl. 1 (187 UCE) -----Plaisir----->	
4	-----Santé positive (bien être)----->
Cl. 3 (178 UCE) -----Travail----->	
5	
Cl. 2 (458 UCE) -----Santé négative (maladie)----->	

FIGURE 2. – Dendrogramme des classes stables.

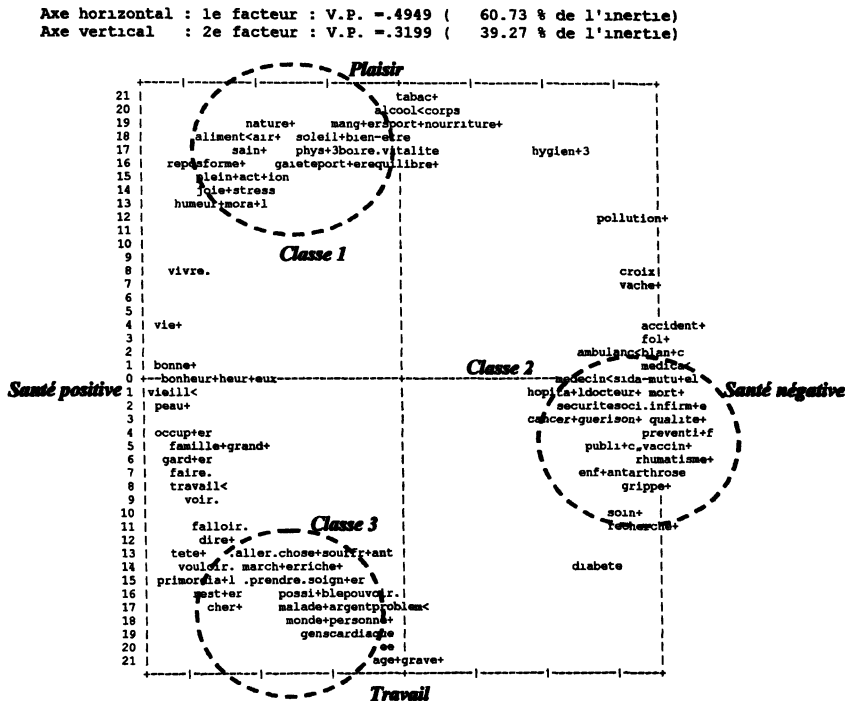


FIGURE 3. – Plan principal.

donc se prolonger dans des conditions identiques, c'est-à-dire que le même traitement de classification peut être appliqué sur le corpus de chaque classe considéré indépendamment des autres. On peut ainsi espérer constituer au sein de chaque groupe un certain nombre de sous-classes, plus homogènes sur le plan lexical, permettant d'y lire de façon plus tangible les thèmes autour desquels s'articulent les topiques de la représentation. On ne peut rendre compte de ce travail dans le cadre limité de cette présentation qui doit également montrer comment s'effectue le passage du traitement statistique à la sémantique.

4. Nommer

Le second moment de l'analyse consiste à effectuer un saut épistémologique significatif entre les plans statistique et sémantique. C'est souvent ce passage qui, dans toute analyse thématique du reste, est le plus difficile à objectiver : il suppose d'effectuer un certain nombre de classements et de regroupements non pas en fonction de la forme des expressions recueillies mais de leur sens. Or l'analyste dispose souvent de peu d'outils pratiques pour rendre sa démarche plus transparente en la matière. Pour ce faire, nous avons utilisé le logiciel Tropes qui propose les moyens de résoudre certaines des difficultés liées à ce passage (Ghiglione et coll., 1995, 1998 ; Le Quéau & Brugidou, 1998). Pour présenter sommairement cet outil, il convient seulement de préciser qu'il fonctionne comme un logiciel d'analyse de contenu qui permet de reconstituer le réseau des liaisons sémantiques existant entre différentes notions. Son approche est basée sur la notion de « classes d'équivalents » ou « univers de références » en ce sens qu'elle ramène différentes catégories de termes identifiés dans un corpus quelconque à une méta catégorie sémantique.

4.1. Les univers de références

L'« univers de référence » est donc un ensemble qui regroupe des « équivalents » sémantiques. « Santé », par exemple, contient des notions qui, toutes, renvoient en réalité à la maladie : sa désignation ou bien son traitement. Parmi les 81 mots qu'elle rassemble dans la classe 1 (cf. tableau 4, ci-après), on trouvera donc en premier lieu les termes assez génériques de « maladie(s) » ou « malade(s) ». Mais, de façon plus précise, on lira aussi le nom de diverses pathologies comme celles qu'on a détaillées plus haut comme caractéristiques de la seconde classe. Ces notions sont, en quelque sorte, mises en balance avec l'institution médicale dont c'est évidemment la vocation de les combattre : « docteur(s) », « infirmier(s) », « ambulance », sont quelques-uns des termes qui y sont donc associés. La notion d'équivalent sémantique, autrement dit, est assez large et déborde en tout cas le champ des synonymes.

En l'occurrence, cet ensemble d'équivalents (« santé ») n'est guère spécifique d'une classe puisqu'il regroupe l'essentiel des termes qui les composent toutes les trois. Son importance quantitative est cependant tout à fait remarquable dans la seconde. De ce fait, les équivalents qu'on y rencontre sont plus nombreux et diversifiés sans que, toutefois, le contenu général de l'univers change de manière significative. Les pathologies évoquées sont certes plus nombreuses et on y trouve des maladies éventuellement moins graves que celles qui ont été évoquées dans la première classe. Les termes de la technique médicale y sont aussi plus variés (« seringue », « vaccin », « scanner », « antibiotique »...) comme ceux de l'institution (« maison de retraite », « SAMU », « pharmacie », « chirurgie »...). L'exploration des équivalents relevant de cet ensemble dans la troisième classe, où ils sont moins nombreux, n'ajoute rien de nouveau au contenu de cet univers.

La notion d'univers de référence permet donc de rendre compte de quoi il est question dans une réponse et, ainsi, de décrire les différents thèmes évoqués

UN PARCOURS INTERPRÉTATIF DES DONNÉES

TABLEAU 4. – Les univers de références des classes.

Classe 1	Classe 2	Classe 3	
Santé (81)	Santé (512)	Santé (116)	1
Sentiment (54)	Enfant (18)	Vie (24)	2
Sport (52)	Mort (10)	Emploi (20)	3
Alimentation (41)	Souffrance (10)	Sentiment (20)	4
Vie (41)	Biologie (8)	Gens (17)	5
Air (14)	Sentiment (6)	Richesse (17)	6
Climat (14)	Sport (6)	Famille (10)	7
Loisir (8)	Alimentation (6)	Pouvoir (8)	8
Boisson (8)	Vie (6)	Enfant (7)	9
Emploi (6)	Assurance (6)	Chance (6)	10
Environnement (5)		Souffrance (6)	11
Souci (5)		Souci (5)	12
		Sport (5)	13

dans une classe. La comparaison des différents univers des classes doit donc permettre de faire apparaître les spécificités de leurs réponses, du point de vue de leur contenu⁷.

Si l'on excepte donc l'univers de la «santé» qui est non seulement le plus important dans toutes les classes mais aussi le même, quant à son contenu, l'examen des autres univers permet de faire émerger quelques différences significatives. Certains univers caractérisent ainsi la première classe parce qu'ils y sont sur-représentés par rapport au reste du corpus. L'univers des «sentiments», par exemple, est moins important dans les classes 2 et 3. En revanche, l'univers «enfant», totalement absent de la première classe, semble constitutif de la seconde classe⁸.

4.1.1. Les principaux univers spécifiques de la première classe

● Sentiment

Cet univers regroupe 54 équivalents qui, tous, renvoient à une perception extrêmement positive de la santé davantage, à la limite, envisagée comme un «état d'esprit» que comme un état physiologique. Parmi les équivalents appartenant à cet univers on trouve notamment : «joie», «bonne humeur», «sourire», «plaisir», etc. Une réponse typique que le programme extrait de

7. Les classements sont effectués sur la base d'un dictionnaire de langue française intégré au programme mais des options de traitements permettent de tenir compte de l'usage et du sens que peut avoir un mot dans un contexte particulier, non prévu par le dictionnaire, voire d'ajouter des mots inconnus par le dictionnaire (argot, termes techniques, etc.) et de les situer dans leur «univers» spécifique.

8. Le seuil de 5 mot a été fixé au départ pour déterminer un «univers de référence».

cette classe et désigne comme une des 10 « propositions remarquables » est la suivante⁹ : « *Grand air, eau, blanc, soleil, beauté, lumière, promenade, nature, forme, dynamise, bonne humeur, joie, rires, travail, le travail c'est la santé* ». Dans la seconde classe, cet univers ne compte que 6 équivalents qui connotent des expressions souvent opposées : « tristesse », « crainte », « peur », « horreur », « joie », « confiance »... Les deux premières classes sont donc parfaitement symétriques sur ce thème. La troisième classe, quant à elle, constitue le moyen terme exact entre les deux premières puisque parmi les 20 équivalents qu'elle contient, on trouve à part égales les mêmes notions, positives et négatives.

• Sport

Cet univers se présente comme très homogène parce qu'il ne connaît que peu d'équivalents différents du terme qui le désigne... Tout au plus peut-on trouver « sportif(s) », « gymnastique » et « footing ». Parmi les propositions remarquables de cette classe, on trouve la réponse suivante : « *sport, dynamisme, activité physique* ». Dans les autres classes, chaque fois que cette notion apparaît, c'est bien dans les mêmes termes, rigoureusement.

• Alimentation

Plus divers, cet univers ne compte pas moins de 41 équivalents qui dénotent l'attention qu'il convient d'accorder à son alimentation, condition de cette vitalité. Rien d'étonnant donc à trouver le substantif « alimentation » lui-même, ou bien son synonyme « nourriture », souvent associés à des adjectifs comme « bonne », « saine » ou « équilibrée ». Cet univers contient cependant d'autres termes désignant certains aliments particuliers : la « viande », qui doit être de « bonne qualité » ou bien le « yaourt ». C'est également en des termes semblables que ce thème apparaît dans la classe 2.

• Vie

Cet univers ne comprend en réalité qu'un seul terme qui renvoie à ce vitalisme que connotait l'univers des sentiments exprimés. Toutefois, un examen des contextes dans lesquels s'inscrit la notion permet de retenir plusieurs sens. La vie, tout d'abord, en tant qu'énergie, vitalité, joie : « *C'est le soleil, la vie, la gaieté, jaune, bleu, couleur pétarde* » ; la restriction de l'hygiène de vie, ensuite : « *Les plantes, je suis contre les tranquillisants, une vie saine, bien équilibrée, sans écarts, sans faux pas* » ; le minimum de la qualité de vie, enfin : « *Qualité de la nourriture, recherche de développement des produits pharmaceutiques, qualité de vie, hospitalisation* ». Comme l'indique le tableau 4, cet univers se retrouve dans toutes les classes, quoique dans des proportions variées. En tout cas, son contenu ne change pas sauf, peut-être, dans la troisième où la notion de vie est aussi associée à trois autres types de contexte. La vie y désigne alors une durée quand il est fait allusion à l'existence

9. Une proposition remarquable est extraite du corpus en raison des indices qu'elle contient et qui correspondent aux descriptifs de la classe, en termes de catégories de mots fréquents (verbe, modalisation, adjectif, joncteur, pronom personnel) et en termes de style. Notons à cet égard que les deux logiciels tombent souvent d'accord sur les mêmes propositions remarquables quoique, évidemment, sur des bases totalement différentes.

particulière de la personne interrogée : « *J'ai beaucoup de chance parce que je n'ai jamais été malade de ma vie* » ; ou bien l'existence envisagée d'un point de vue plus générique : « *Il arrive un moment, dans la vie, où il y a des maladies graves* ». Il arrive encore que, dans la troisième classe, la notion renvoie à la vie humaine en général : « *Don du sang, sauver des vies humaines* ».

• Environnement

On ne peut finir d'évoquer le sens des évocations classées dans le premier type sans livrer quelques indications de plusieurs univers qui pourraient être regroupés en un seul et qui sont très caractéristiques : « air », « climat » et « environnement ». Ces trois univers sont extrêmement proches et contribuent à renforcer la spécificité de la première classe, non seulement parce qu'ils sont absents des autres, mais aussi parce qu'ils sont très liés à certains thèmes de l'univers des sentiments positifs (le « plaisir ») et aux activités physiques (le « sport »). Il faut préciser que l'univers « climat » ne comprend qu'un seul terme qui est « soleil ». Cette notion est connexe d'un autre univers caractéristique de la classe « loisir », qui ne contient en réalité qu'un seul terme : celui de « vacances ». En effet, lorsqu'il est question de « soleil », il est aussi souvent question de vacances et de repos.

La vision de la santé qui semble se dégager de cette classe apparaît donc non seulement extrêmement positive mais également « holistique », dans la mesure où elle est liée à des éléments d'environnement, comme on l'a dit plus haut. Par rapport aux résultats de l'analyse faite avec Alceste, Tropes permet en outre de mieux faire apparaître ici la dimension d'« état d'esprit » que sous-tend une certaine perception de la santé.

4.1.2. Les principaux univers spécifiques de la seconde classe

• Enfant

Cet univers recouvre des significations paradoxales que dit bien cette réponse : « *un enfant qui joue dehors, un enfant sur un lit d'hôpital* ». L'enfant est souvent cité comme une image cristallisant toute la vitalité et l'énergie que l'on souhaite évoquer dans une réponse. Par ailleurs, lorsque la maladie atteint un enfant, c'est comme si le scandale en était redoublé : « *Je ne supporte pas la maladie pour les autres, surtout pour les enfants* ». Sur le plan lexical, l'univers est donc très homogène, tout au plus ne peut-on trouver qu'un seul synonyme d'enfant : « gosse ». Dans la classe 3, c'est également sous cette forme équivoque et imagée qu'il apparaît. Dans ce contexte, la notion y renvoie cependant à un autre univers caractéristique : celui de la « famille ».

• Mort

Très caractéristique, également, cet univers renforce bien la structure pessimiste qui semble structurer l'univers de perception de la santé de la classe (cf. plus haut l'univers « sentiment »). Cet univers ne comprend que les formes, substantives ou adjectives, associées au radical « mort », comme l'indique cette réponse extraite des propositions remarquables de la seconde classe : « *Cancer, sida... des maladies pénibles comme il y en a eu dans ma famille. Une personne est morte de cancer* ».

• Souffrance

Cet univers, très proche du précédent, ne contient qu'un seul terme, décliné au singulier et au pluriel : « douleur ». Le terme arrive seul ou bien dans un contexte qui désigne la lutte ou bien les traitements « contre la douleur » : *« Prévention, tabac, alcool, médecin référent, soins palliatifs, traitements contre la douleur, bilan de santé »*.

Cette seconde classe, très largement dominée par le thème de la « santé » comme défini strictement plus haut (en l'occurrence essentiellement une énumération impressionnante de pathologies) se caractérise par une perception problématique de la santé, dans le sens simple où celle-ci pose toujours un problème, est le résultat d'un équilibre fragile. Très révélatrice à cet égard est la présence du couple antagoniste « vie » et « mort » qu'illustre, à sa manière, l'évocation ambiguë de l'enfant.

4.1.3. Les principaux univers spécifiques de la troisième classe

• Emploi

Si la troisième classe contient beaucoup d'univers de référence qui ont déjà été détaillés auparavant (« vie », « sentiment », « enfant », « souffrance »...), elle ne s'en caractérise pas moins par quelques thèmes tout à fait spécifiques au premier rang desquels figure le « travail ». Cet univers regroupe des évocations assez différenciées. Si l'on excepte le célèbre lieu commun *« le travail, c'est la santé »* qui revient quelques fois, le travail est tout d'abord envisagé comme une conséquence de la santé : elle est ce qui permet de travailler. Les réponses qui peuvent être classées sous ce thème rappellent très explicitement l'item de la question fermée selon lequel la santé, c'est « ne pas être malade » : *« Etre en bonne santé, ne pas être malade, pouvoir se lever tous les jours vaquer à ses occupations, se rendre à son travail tous les jours »*. Dans une acception très proche, la santé est bien ce qui permet d'accéder au travail et, partant, aux ressources qu'il procure : *« Quand on est en bonne santé, on peut travailler. La santé, c'est la fortune la santé, c'est tout »*. Moins fréquemment, le travail peut être associé à la maladie par le biais des affections professionnelles : *« Mon mari a une maladie liée à sa profession de prothésiste dentaire... »*. Enfin, le travail en tant que signe de bonne santé, est évoqué avec d'autres activités, un « faire » plus global : *« La vie, le travail, l'amour, la fête, c'est la santé »*.

• Richesse

La santé se présente, pour un certain nombre de personnes interrogées, comme une des principales richesses que peut offrir la vie : *« C'est posséder quelque chose de très important, bien dans sa tête, sérénité, zen, la plus belle richesse »*. On pourrait peut-être rapprocher cet univers d'un autre qui caractérise également cette classe qui est celui de « chance ». Dans les deux cas, il est donc question d'un capital qu'il s'agit d'essayer de conserver ou d'augmenter : *« C'est capital, médecin traitant, alimentation, sport, information, image de ma fille malade actuellement. C'est un capital qu'on se doit de conserver. C'est pas évident au vu d'agressions extérieures. C'est pas évident de rester en bonne santé, indépendant de la volonté »*. Le thème de la richesse

rejoint accessoirement le thème du travail («...cela amène aux soins») dans une évocation du coût de la santé. On trouve en effet ici, et c'est encore une fois tout à fait caractéristique, plusieurs allusions au fait que la santé (mais aussi les «soins»), cela peut être «cher»... voire «inaccessible».

• Famille

Le thème de la famille peut se présenter comme extrêmement proche du précédent dans la mesure où la famille, elle aussi, peut être perçue comme une richesse. D'autre part, cette évocation, qui pourrait ressortir du thème «environnement» de la première classe, permet de souligner les implications collectives de la santé : la famille et les relations qu'elle sous-tend est une condition de la santé, en même temps que celle-ci conditionne un état du bien-être collectif.

Il est assez difficile de réduire à une formule assez simple l'ensemble des thèmes qui structure cette classe. Nombre d'entre eux recoupent ce qui a déjà été évoqué dans les autres classes. Ainsi la dimension affective y est-elle très présente mais les réponses soulignent ici surtout le rapport qu'entretiennent les relations sociales et surtout familiales avec la santé («sentiment», «famille», «enfant»...). La dimension «insertion sociale» est également présente, de différentes manières d'ailleurs, avec l'évocation du thème du travail et de l'activité au sens plus large auquel renvoie le thème de «pouvoir».

4.2. Remarques sur le style

Le logiciel Tropes permet de donner quelques indications générales sur le style des textes analysés. On doit cependant considérer que cette option d'analyse est peu pertinente, en l'occurrence, compte tenu à la fois de la question posée (et donc du type de réponses suggérées) et des contraintes de la prise de note. D'une manière générale, les réponses présentent assez peu souvent l'aspect d'une phrase : verbe, sujet, complément. D'ailleurs, lors de la description du «style général» du texte des réponses, peu de différences apparaissent entre les indicateurs. Dans les trois classes, le style est désigné comme «ancré dans le réel», l'énonciation est prise en charge par un «je», etc. Dans cette catégorie d'analyse un seul type d'indicateur se révèle en réalité d'une relative utilité : celui qui permet la comparaison quantitative des différentes catégories de mots employés (verbe, modalisation, joncteurs, etc.). Une seule catégorie fait apparaître des différences significatives entre les 3 classes, c'est celle des adjectifs : des qualificatifs «subjectifs» caractérisent les classes 1 et 3 (respectivement 75 % et 83 % des adjectifs employés) alors que les réponses de la classe 2 contiennent, en majorité (54 %), des adjectifs «objectifs».

4.2.1. Les adjectifs subjectifs

L'adjectif subjectif dénote une prise en charge ou une valorisation tout à fait particulière, c'est-à-dire plutôt impliquée, de l'énoncé de la réponse. Ceci constitue donc un trait commun aux classes 1 et 3 qui peut par ailleurs correspondre à l'importance que revêt, dans ces deux classes, l'univers des «sentiments». Il semble cependant que les adjectifs considérés dans cette

catégorie soient assez différents d'une classe à l'autre. Dans la première, la quasi-totalité des adjectifs subjectifs surdéterminent la perception positive de la santé et, même, ce qu'on a appelé plus haut la dimension «holistique» de cette perception. C'est ce que dénotent les adjectifs qui suivent (liste non exhaustive) : «bon(ne)», «sain(e)», «raisonnable», «équilibrée», «vital(e)», «naturel(le)», «heureux(se)»...

On note cependant aussi la présence de quelques adjectifs «négatifs» («triste», «grave»...) en général associés à la maladie mais ils restent relativement peu nombreux dans cette classe. Ils représentent en revanche une part beaucoup plus importante des adjectifs subjectifs dans la classe 3. À côté d'autres connotant positivement, et qui sont peu ou prou les mêmes que dans la classe 1, on trouve en effet, non exhaustivement : «affreux», «grave», «triste», «terrible», «malheureux»... Comme la classe 2, structurée autour d'une opposition («santé»-«vie»/«maladie»-«mort»), la classe 3 apparaît donc ici organisée autour d'une autre paire antagoniste : «santé»-«bonheur»/«maladie»-«malheur».

4.2.2. Les adjectifs objectifs

Les réponses composant la classe 2 ne sont certes pas exemptes de tout adjectif subjectif mais ce sont ceux qui sont plus strictement dénotatifs qui, numériquement, l'emportent. En général, les adjectifs objectifs contribuent à la description des affections évoquées. Les maladies citées peuvent être : «infantiles», «vasculaires», «cardiaques», «cérébrales», «chroniques»... Ces adjectifs peuvent encore concourir à préciser d'autres termes du corpus technico-médical comme «sanguin» (comme un «bilan» voire un «groupe»); «corporelle» (comme l'«hygiène»); ou bien ils contribuent à qualifier les institutions médicales : «public(que)», comme l'«hôpital» ou la «santé»; «blanche», comme la «blouse»; «verte», comme la «croix»; «primaire» comme la «caisse d'assurance maladie», etc.

5. Bilan de l'expérience

Parce que les approches statistiques et sémantiques se complètent très bien, la méthode permet d'aller assez loin dans l'exploration du corpus. On peut bien entendu discuter chacun des choix qui est fait à chaque étape mais c'est précisément l'intérêt des outils et techniques proposés que d'y autoriser. Dans le détail, on peut ainsi remettre en question certains regroupements proposés par Tropes comme celui qui consiste à ranger dans l'univers santé des évocations très diverses qui touchent à la maladie, à l'institution médicale et ses personnels, et aux traitements. C'est d'ailleurs le dépouillement statistique initial qui peut nous avoir mis en garde quant à la pluralité réelle de cet univers. Cette catégorisation, aussi large soit-elle, a cependant cet intérêt de faire émerger d'autres thèmes que l'approche statistique n'aurait peut-être pas permis de saisir avec la même évidence, le thème étant traité par un grand nombre de mots différents, par exemple, chacun revenant avec une faible fréquence. On peut également poursuivre et perfectionner encore l'analyse en

évaluant le « poids » statistique réel des différents thèmes identifiés sur un plan sémantique. Mais on considérera, provisoirement, qu'il n'est pas évident que cet exercice soit d'un apport très significatif pour les objectifs assignés à cette expérience.

En tout état de cause, à défaut de réellement repousser « plus loin » le moment de l'interprétation (l'usage de logiciels présuppose d'emblée l'adoption d'options techniques précises qui en relèvent déjà pleinement), la méthode donne des points de repère pour objectiver la construction d'une « représentation des représentations » de la santé. Pour ce qui est de la question de fond posée à cet égard, l'intérêt de l'exploitation de cette question ouverte est d'élargir la connaissance qu'on avait de ce que recouvre la notion dans l'esprit des personnes interrogées. Il reste bien entendu à établir les relations significatives entre ces éléments de représentations avec l'expérience vécue par les acteurs. Mais, pour ce qui est de la question plus pratique, touchant à la formulation d'une question fermée, un examen de l'ensemble des différents thèmes identifiés permet déjà d'entrevoir les pistes d'un enrichissement thématique important. L'item « pouvoir faire ce que l'on veut », par exemple, mérite d'être scindé pour distinguer ce qui relève de la possibilité de travailler, de la recherche du plaisir au travers diverses activités de loisir. Il faudra aussi statuer sur l'intérêt qu'il y aurait à inclure des thèmes nouveaux portant sur le rôle des environnements naturel, social et familial ; l'alimentation ; etc. Le problème économique, enfin, qui structure l'univers de la classe 3, devrait sans doute être intégré. Mais c'est une autre expérience-test qui peut nous assurer de parvenir à une formulation pertinente, sociologiquement, de cette question fermée.

Bibliographie

- ALDEGHI I. et coll. (1998), *Étude complémentaire sur les nouveaux entrants au RMI*, Collection des rapports du CREDOC, n° 188, Paris.
- BRUGIDOU M. (2001), La combinaison des inférences statistiques, linguistiques et sociologiques dans l'analyse d'une question ouverte. *Journal de la Société Française de Statistique*, 142, 4.
- FOUCAULT M. (1984), *Le souci de soi. Histoire de la sexualité*, Tome 3, Gallimard, Paris.
- GADAMER H.-G. (1960), *Vérité et méthode*, Seuil, Paris.
- GHIGLIONE R. et coll. (1995), *L'analyse cognitivo-discursive*, Presses Universitaires de Grenoble.
- GHIGLIONE R. et coll. (1998), *L'analyse automatique du discours*, Dunod, Paris.
- HÉBEL P., LE QUÉAU P. et coll. (2000), *Les Français et la prévention de l'alcoolisme et du tabagisme*, Collection des rapports du CREDOC, n° 210, Paris.
- HERZLICH C. (1969), *Santé et maladie. Analyse d'une représentation sociale*, Mouton, Paris-La Haye.
- LE QUÉAU P. et BRUGIDOU M. (1998), *La dynamique du récit, contribution au traitement et à l'analyse des entretiens non directifs de recherche*, Cahier de recherche du CREDOC n° 124, Paris.
- LE QUÉAU P. et OLM C. (1998), *La perception de la santé en France*, Collection des rapports du CREDOC n° 185, Paris.

UN PARCOURS INTERPRÉTATIF DES DONNÉES

- LE QUÉAU P. et OLM C. (2000), *Les jeunes et le risque routier*, Collection des rapports du CREDOC n° 211, Paris.
- PIERRET J. (1996), *Cultures ou production de la santé dans la famille : différenciations et inégalités sociales*, CERMES, rapport pour la MIRE, Paris.
- RICOEUR P. (1965), *De l'interprétation*, Seuil, Paris.
- VIGARELLO G. (1993), *Le sain et le malsain. Santé et mieux-être depuis le Moyen Age*, Seuil, Paris.
- WEBER M. (1913, 1965), *Essai sur quelques catégories de la sociologie compréhensive, Essais sur la théorie de la science*, Plon, Paris.