

Article

« Traitement des données à facettes »

Pierre Verges

Sociologie et sociétés, vol. 25, n° 2, 1993, p. 37-46.

Pour citer cet article, utiliser l'adresse suivante :

<http://id.erudit.org/iderudit/001824ar>

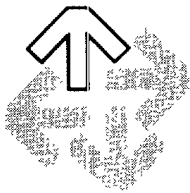
Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <http://www.erudit.org/apropos/utilisation.html>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : erudit@umontreal.ca

Traitement des données à facettes



PIERRE VERGES

« Il faut plusieurs filets pour avoir la pêche. »

Bien souvent le chercheur se contente d'associer une variable à une question, l'analyse de ses déterminations retenant toute son attention à travers des croisements et des analyses multivariés. L'aspect caricatural du procédé est visible dans les sondages d'opinion où le seul tri à plat des réponses à la question « Pensez-vous qu'il faut laisser les filles de 18 ans voir les films qu'elles veulent ? » suffit à rendre compte de l'état moral d'une société (ici de son libéralisme). Le croisement par catégories socioprofessionnelles, sexe, âge, etc., ne modifie pas profondément les résultats, car l'identification initiale de la variable dépendante (de la notion qui la sous-tend) à la question censée la représenter est toujours incertaine. Elle peut seulement être éclairée par le contexte contrasté des variables indépendantes (féminin / masculin, jeunes / vieux, etc.).

Nous désirons montrer comment cette manière de voir fait peu de cas de la forme de recueil des données, du rapport existant entre cette forme et l'information obtenue. S'interroger sur la forme des données paraît superflu tant le chercheur pense avoir traduit sa théorie en « énoncé de réalité ». Or ce rapport ne nous apparaît pas évident, il n'est pas immédiat. Ici, par le biais de quelques situations types, on se bornera à identifier le problème, à mettre l'accent sur une sorte de construction des données par rétroaction entre empirie et théorie. Une théorisation de ce problème pourrait à elle seule faire l'objet d'un livre.

Nous passerons sous silence la situation où ce rapport n'est que métaphorique, comme dans le cas où, dans une analyse de discours, on identifie l'agent principal de l'action au sujet verbal de la phrase. Nous nous intéresserons au rapport pluriel, qui nous semble la seule voie raisonnable. Plusieurs questions sont nécessaires à la définition des différentes facettes d'une même notion. Deux traitements se rapportant à la même question apportent des informations complémentaires, car une notion peut se présenter sous des formes expressives différentes, en particulier lorsque les formes de surface linguistique sont les indicateurs de notions sociologiques.

LA FORME DES OBSERVATIONS : À LA RECHERCHE DE L'IDENTITÉ LOCALE

Au cours d'une enquête portant sur l'image d'une localité, nous avons cherché les éléments les plus porteurs de l'identité locale. Nous avons voulu montrer comment certaines configurations d'images peuvent donner lieu à la formation d'un processus identitaire (ou y faire penser) (Bidart, Guidarini, Louis-Palluel, Vergès, 1991). Pour cela nous avons interrogé les habitants d'une ville nouvelle sur le pourtour de l'Étang de Berre (situé à côté de Marseille). Après une série d'entretiens préalables, nous étions en mesure de proposer aux sujets *une liste d'items* décrivant leur localité (cf. encart n° 1). Les choix retenus majoritairement dans cet ensemble d'items évoquant la ville font apparaître, d'une part, l'importance des équipements (nous sommes dans une ville nouvelle de l'agglomération de Marseille qui enregistre une croissance rapide) et, d'autre part, une vision négative : « une cité dortoir », « une ville qui a des problèmes sociaux ». Les items concernant l'urbanisme sont répartis tout au long de la hiérarchie des choix émis, montrant une vision mitigée, une incertitude sur le futur de cet aménagement urbain pourtant concerté.

Encart n° 1

Les expressions suivantes s'appliquent-elles bien à votre ville?
Choisissez par ordre de préférence les 3 expressions qui vous semblent définir le mieux votre ville (Bidart, C., Guidarini, V., Louis-palluel, J., Vergès, P., 1991).

- une ville bien équipée
- une ville qui a des activités culturelles importantes
- une ville qui a des problèmes sociaux (chômage, délinquance...)
- une ville polluée
- une ville qui a des racines, une histoire
- une ville où il fait bon vivre
- une ville dispersée
- une ville où l'intégration des nouveaux venus se fait facilement
- une ville jeune et dynamique
- une ville où l'architecture et l'urbanisme sont ratés
- une ville commerciale
- une cité dortoir
- une ville industrielle
- une ville qui n'a pas de centre
- une ville angoissante
- une ville réussie

Nous avons par ailleurs effectué *une couverture photographique* des principaux espaces publics de cette ville (Windenberger, 1982) et posé aux sujets cette autre question : « Pour faire le portrait de votre ville, quelles photographies choisiriez-vous ? » Les réponses à cette question paraissent, au premier regard, bien différentes. En effet nous voyons émerger un « ancien » regretté, comme en témoignage ce commentaire : « Cette ville est un peu froide, parce que quatre tours de béton qui vous entourent [...] il manque le clocher, une vieille bâtisse, des vieilles maisons ».

La photographie la plus choisie représente non pas la ville nouvelle, mais une rue du vieux village portant le même nom et qui n'est plus maintenant qu'une partie du quartier. La seconde est une sorte de « carte postale » montrant le piton sur lequel se trouvent le vieux village et son église. Les commentaires faits sur ces deux photographies montrent que les personnes interrogées leur associent massivement l'identité locale.

Nous nous trouvons donc devant un résultat contradictoire. Les choix majoritaires expriment d'un côté une ville nouvelle envisagée de manière critique et, de l'autre, une ville ayant un passé villageois producteur d'identité au présent. Il faut en conclure soit que les habitants sont schizophrènes, soit que nous obtenons deux informations différentes. Un examen de la forme des questions nous fait pencher pour la seconde solution, que de toute façon nous aurions choisie au regard d'une perception a priori positive de nos semblables !

La notion théorique que nous nous efforçons de repérer est celle de l'identité locale. Cette notion veut synthétiser à la fois l'inscription dans le présent de la mémoire historique de la localité (les lieux témoins de cette histoire sont alors convoqués) et sa réappropriation par les habitants dans des pratiques identitaires et un discours pouvant prendre une distance critique. Cette double dimension renvoie à notre double questionnement. Mais elle n'explique pas la contradiction relevée.

Celle-ci doit être référée à la forme des données et à la nature de son traitement. En effet la photographie a la carte postale et la photographie de famille pour références, alors que le discours permet le balancement du pour et du contre. Or, ce qui apparaît dans un choix majoritaire, c'est l'expression la plus commune pour ne pas dire la plus triviale : d'un côté le passé (figé dans la photographie), de l'autre la discussion (inscrite dans les discussions présentes). Ces deux manières de recueillir l'information ne sont donc pas équivalentes. Elles suscitent deux approches de la réalité locale, exacerbées par un traitement des données se réduisant à comptabiliser la fréquence des choix. Elle renforce la particularité de chaque approche en ramenant l'information à un repérage de l'expression majoritaire.

Il convient donc de dépasser ce premier résultat pour montrer qu'il y a derrière cette contradiction une cohérence des réponses de chaque sujet. Pour cela nous avons étudié les choix simultanés des phrases et des photographies par une technique d'analyse multivariée dite « analyse de similitude » (cf. encart n° 2).

Ce traitement de l'information (Degenne, Vergès, Flament, 1981) nous permet de voir qu'il existe différentes images, ayant chacune leur cohérence, et que ces différentes images sont portées par des groupes sociaux identifiables. On calcule dans un premier temps la cooccurrence des phrases et des photographies et on compare cette valeur avec la valeur que l'on obtiendrait s'il y avait une indépendance statistique relativement aux occurrences de ces phrases et photographies. L'indice ainsi calculé (ici le Phi de contingence) sert d'indice de similitude (Hubalek, 1982).

Le graphique obtenu est intéressant, car il est constitué de zones assez bien différenciées. On obtient des coupures assez nettes entre les zones de fortes similitudes (identifiées par des sortes de cercles) et une organisation de ces zones autour de l'arbre maximum sous la forme d'un axe principal de gauche à droite et d'un axe secondaire vers le bas qui isole, en fait, un groupe bien particulier, comme nous le verrons plus loin.

L'analyse des zones de fortes similitudes fait apparaître cinq groupes (cf. graphique n° 1). Quatre de ceux-ci s'organisent sur un axe de gauche à droite sur le schéma. Cet axe nous fait passer d'une vision plutôt identitaire à une perspective plutôt distanciée de la ville.

— Sur la gauche on trouve une description valorisée du patrimoine ancien, des équipements et de la nature (zone n° 1 : L1 = « Une ville bien équipée », L9 = « Une ville jeune et dynamique », 022 = une vue des toits du vieux village). Elle correspond à la vision des habitants les plus anciennement installés dans la commune.

— Vient ensuite la présentation d'éléments que l'on peut qualifier d'emblématiques (anciens ou modernes ; zone n° 2, 07 = la mairie moderne, 019 = le terrain de sport, 024 = la grande place et l'église du vieux village, L5 « une ville qui a des racines, une histoire »). Elle correspond à un groupe d'habitants plus récemment implantés et provenant de milieux socialement moins favorisés que les précédents.

— Au centre du graphe, la ville est caractérisée par les différentes facettes de son urbanisme (zone n° 3, L13 = « Une ville industrielle », 09 = une vue de la rue commerçante). Cette vision plus objectivante de la ville est celle de cadres ayant seulement un rapport d'usage à la ville.

— À droite, l'association de la pollution à une vision dévalorisante de l'urbanisme rassemble tous les mécontentements (zone n° 4, 05 = une vue d'une décharge sauvage, 05 = un chantier HLM).

— Un peu à l'écart de cet axe, la zone n° 5 associe, dans une évocation apaisée, la nature et la qualité de vie (L6 = « Une ville où il fait bon vivre », L16 = « Une ville réussie »). Cette image est proposée par des nouveaux venus issus de milieux ouvriers et

Encart n° 2
ANALYSE DE SIMILITUDE

L'analyse de similitude est une méthode d'analyse des données permettant, en particulier, l'étude des représentations sociales. C'est une méthode de description dont le but est de dégager une structure relationnelle ou de regroupements sur un ensemble d'éléments. Elle s'appuie sur les propriétés de la théorie des graphes.

L'analyse de similitude consiste à établir, entre les photographies et les phrases de notre questionnaire retenues par les interviewés, un indice de ressemblance basé sur le nombre de choix simultanés. Chaque personne choisissait six photographies et trois phrases parmi les 25 photographies et les 16 phrases proposées. La cooccurrence de ces choix est alors comparée à la valeur d'un choix aléatoire (hypothèse d'indépendance statistique) afin de calculer un indice de similitude : le «Phi de contingence», indice assimilable à un coefficient de corrélation.

Cette similitude a un caractère «local» dans la mesure où le calcul repose sur le nombre de personnes ayant choisi les deux photographies dont on mesure la similitude. Ce caractère local s'oppose au caractère global de l'indice classique de la distance du Khi2 utilisé dans l'analyse de correspondance. Ce dernier compare les «profils» de chacune de ces deux photos avec toutes les autres. Dans notre cas la valeur de la relation des deux photos (x et y) ne tient pas compte de ce profil global avec les autres photos, mais seulement du nombre de personnes ayant choisi ensemble deux photos.

Cet indice de similitude est calculé pour tout couple d'éléments. L'ensemble des indices s'organise en une matrice carrée, indiquant la similitude pour chaque couple.

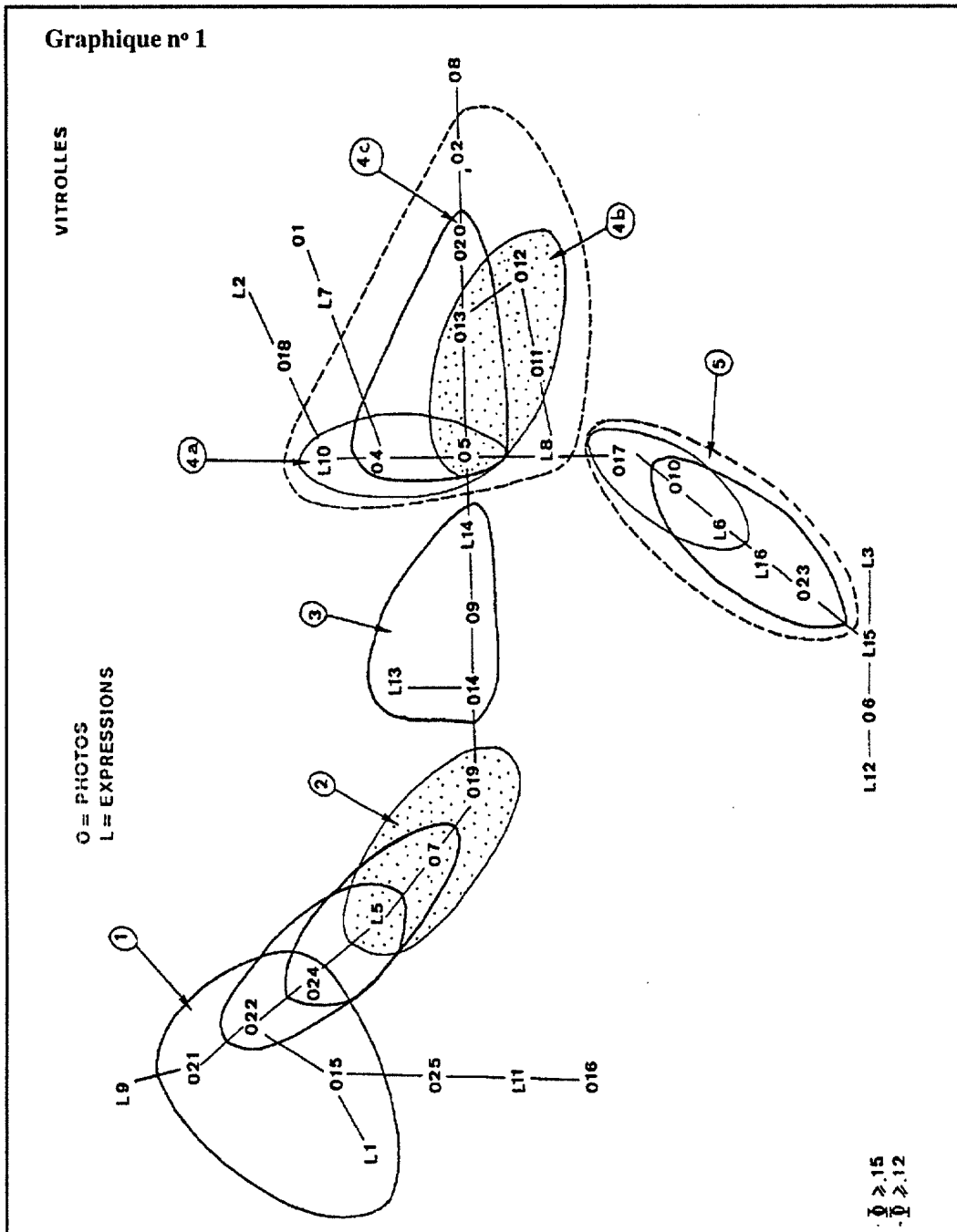
La visualisation des résultats peut être effectuée par un graphe : ensemble de sommets et d'arêtes. Le graphe connexe le plus simple est appelé arbre maximum : c'est la simplification extrême qui ne retient de l'ensemble des similitudes que celles localement les plus fortes.

À cet arbre maximum nous ajoutons une information supplémentaire — les cliques — afin d'indiquer les zones de regroupement des éléments analysés. La théorie des graphes définit la notion de clique comme un ensemble de sommets reliés totalement deux à deux. Nous avons, par des patatoïdes, indiqué les cliques dont les arêtes sont toutes supérieures à un seuil donné.

Notons que la disposition graphique du dessin qui représente un graphe ou un arbre n'a pas de valeur géométrique : seul compte le fait qu'entre deux éléments existe ou non une arête. La propriété «locale» de cette analyse fait que chaque zone du graphe a sa propre interprétation (ses propres critères de cohérence). Par exemple, d'un côté les photographies et les phrases s'organisent autour de la notion de patrimoine; en un autre endroit, elles se regroupent autour de l'urbanisation ou de la qualité de la vie... Il n'y a pas obligatoirement un axe général d'interprétation.

originaires des communes avoisinantes. Il semble que ces gens soient venus là volontairement (dans le cadre d'une stratégie résidentielle).

L'univers «représenté» ici décrit *ne se réduit donc pas* au seul rêve d'une ruralité préservée ou au rejet d'un lieu trop bétonné. *Il donne lieu à différents processus identitaires* (à une identité fondée sur le passé, à une identité urbaine) *ou à des rapports plus individuels* avec la ville (utilitaristes ou relevant d'une stratégie résidentielle). Ceux-ci n'ont pu être mis en évidence que par le croisement raisonné de deux ensembles d'observation. Ce croisement a été productif pour deux raisons : d'une part, on a identifié la nature de l'information apportée par chaque forme de questionnement; d'autre part en utilisant le rapport à l'indépendance statistique, on a neutralisé ce rapport majoritaire pour mettre en évidence les principes de composition propre à chaque sous-population.



LES FACETTES D'UNE NOTION : LES AGENTS ÉCONOMIQUES.

L'analyse d'une représentation sociale commence toujours par l'identification de son noyau central (Jodelet, 1989). On fait appel pour cela à différentes techniques de cueillette et de traitement des données. L'une d'elle s'inspire des thèses sur la prototypicalité (Rosch, Lloyd, Doise, Clemence, Lorenzioldi, 1992). Elle consiste à poser aux sujets une question d'évocation du type : « Qui, à votre avis, joue un rôle important dans l'économie ? » Le sujet doit fournir de trois à huit réponses. Le traitement de ces données suppose qu'est défini le critère de prototypicalité permettant de repérer les termes les plus

saillants dans la représentation sociale d'un objet donné. Or les écrits sur le sujet définissent plusieurs critères : la fréquence, le soulignement de l'importance ressentie par le sujet, l'ordre d'apparition de la réponse.

Il y a donc ici, à partir d'une même question, la possibilité de traiter plusieurs informations. Ces dernières sont-elles cohérentes, en corrélation ? Le croisement des informations ne relève pas de l'utilisation de deux formes de données aussi différentes que sont les photographies et les textes, mais d'une réexploitation des informations initiales sous différents angles d'approche d'un point de vue méthodologique. L'analyse des mots ou des syntagmes énoncés peut ainsi prendre deux directions : celle de *l'analyse des fréquence d'occurrence*, celle de *l'analyse du rang* dans l'ensemble des mots cités par un même sujet. L'hypothèse initiale est qu'il doit y avoir corrélation entre ces deux informations : la fréquence et le rang moyen (Vergès, 1992 ; Guimelli, Rouquette, 1992). Cette hypothèse est souvent postulée, on calcule alors un coefficient agrégeant les deux informations (Bourgeat-Carter, 1993).

Or on constate qu'il n'en est rien. En effet un terme peut être important pour le sujet qui l'énonce (et donc placé dans les premières évocations), mais n'être pas présent chez beaucoup de sujets. C'est le cas des mots *investissement* ou *industrie* ou encore de mots moins génériques comme *Bérégovoy* (lequel était le premier ministre français à l'époque où l'enquête a été menée). Inversement, certains mots sont très fréquemment utilisés, mais ils sont considérés secondaires par les sujets ; ils apparaissent au dernier rang comme, par exemple, les mots *bourse* ou *banque*.

Par le croisement de ces deux critères, on arrive donc à créer un tableau à quatre cases (cf. tableau n° 1) où il est possible d'identifier quatre sortes de terme. En haut et à gauche on trouve les termes vraiment saillants et significatifs pour toute la population (le mot *entreprise* et son chef, puis le mot *État* avec son gouvernement et ses ministres, enfin le mot *consommateur*). En bas et à droite on trouve les mots les moins fréquemment utilisés et ceux cités en dernier (ici le monde du travail, l'étranger et certaines notions économiques abstraites). On peut établir une relation hiérarchique entre ces deux cases, car les deux critères sont en situation de congruence.

Mais, dans les deux autres cases, ces critères sont en situation de contradiction. On trouve en haut et à droite les agents de l'économie monétaire et les ménages, considérés comme des agents secondaires ; en bas et à gauche, une reprise des deux principaux agents (l'entreprise et l'État) avec leur reformulation en des termes particuliers. Il n'y a pas de stricte hiérarchisation des cases, et par conséquent des mots apparus. L'approche prototypique est mise partiellement en défaut.

Le croisement de ces deux critères et la constatation de leur non-corrélation sont cependant producteurs d'une analyse beaucoup plus pertinente que la seule référence à la fréquence d'apparition des termes, laquelle est souvent le principal indicateur de prototypicalité. Cette analyse fait apparaître deux formes de réponses : celles qui mettent en évidence *les choix majoritaires / minoritaires* et celles qui indiquent *l'importance / la moindre importance* de l'évocation *pour les sujets* qui l'énoncent. Il s'agit bien de deux dimensions ontologiquement différentes : d'une part, une dimension collective (le nombre de fois où un mot a été énoncé) ; d'autre part, le résumé (rang moyen) d'une dimension individuelle, alors que le sujet a établi un ordre entre les mots qu'il a énoncés.

La possibilité d'exprimer une contradiction entre ces deux critères est, à notre avis, essentielle, car elle montre qu'au sein de chaque notion une structure de déséquilibre de changements est présente. Il ne suffit pas de connaître les éléments les plus évidents ou ceux qui apparaissent secondaires ; il faut aussi connaître les éléments qui expriment une ambiguïté. Par exemple, dans une étude sur la représentation des « nouvelles technologies » (Grize, Vergès, Silem, [éds.], 1987) en milieu ouvriers et employés, les réponses faites à une question semblable sur le « travail » situaient le caractère nécessaire et contraint du travail dans les cases de l'antidiagonale (de l'ambiguïté), alors que le revenu (gagner sa vie) devenait un élément saillant et que les aspects humains du travail passaient en seconde position.

Tableau n°1		
Étudiants français 1992		
Tableau des mots les plus fréquents (plus de 6 fois)		
Il comprend 31 mots (soit 10 % des mots). Ces mots représentent 604 évocations, soit 58 % des évocations.		
FRÉQUENCES	RANG MOYEN	
	Inférieur à 3,0	Égal ou supérieur à 3,0
Supérieur ou égal à 15	112 Entreprise 15 Chef d'entreprise 93 État 37 Ministre-f 25 Gouvernement 15 Ministère-f 46 Consommateur	8 Banque 15 Bourse 28 Ménage
Inférieur à 15	12 Producteur 11 Industrie 10 Grandes (entrep.) 9 Entrepreneur 8 Patron 9 Investissement 7 Production 11 Premier ministre 9 Président République 8 Bérégovoy	11 Syndicat 9 Travailleur 7 Salariés 7 Emploi 7 Marché 6 Concurrence 12 Pays 6 Étranger 7 Ministre-a 7 Média 7 Économistes

Cet exemple montre aussi que l'analyse de la thématique des termes de chaque case permet de caractériser celles-ci et d'apprécier dans quelle mesure la notion soulevée par la question a ou non une représentation structurée. Ainsi le « travail » isolait bien trois domaines : le gain, la contrainte et les aspects relationnels et humains. Il les hiérarchisait et indiquait par où la mise en cause du travail pouvait se faire (responsabilité, autonomie, etc.). La représentation du travail est bien établie sur des siècles de la mémoire humaine, aussi sa représentation est-elle bien structurée. Inversement, on observait, dans le cadre de la même enquête, que la notion de « nouvelles technologies » faisait référence à deux domaines distincts : d'une part, le progrès et, d'autre part, la technique et ses machines. Or ces deux domaines se retrouvaient à part égale dans les quatre cases du tableau, montrant par là que cette représentation n'était pas encore autonome, mais se situait à la jonction de deux représentations plus ancrées dans l'imaginaire collectif.

En comparant ces résultats avec ceux de notre premier exemple nous pouvons remarquer :

a) la valeur d'une analyse mettant en évidence la présence (ou l'absence) des diverses modalités de la réponse ; cette analyse décrit un phénomène qui peut être qualifié de *majoritaire* ;

b) l'existence d'une autre dimension de la réalité sociale : la *structure des différences individuelles*. Elle est construite à la fois par une information initiale plus diversifiée (phrases / photographies, fréquences / rangs) et par un traitement des données qui, neutralisant le phénomène « majoritaire », fait apparaître une autre dimension : dans le premier cas, construction de sous-population par le rapport à l'indépendance statistique (Phi de contingence), dans le second cas, rapport individuel / collectif par la prise en compte du jugement personnel du sujet.

Si nous cédon à l'analogie, nous pouvons dire que nous sommes en présence de phénomènes comparables à ceux mis en évidence par la moyenne et par l'écart type dans la statistique descriptive.

FORMES DE SURFACE ET OPÉRATIONS ARGUMENTATIVES

La nécessité de recourir à plusieurs types d'informations pour décrire une même variable, ou, plus exactement, pour découvrir les différentes facettes d'une même variable, nous paraît nécessaire au regard du respect des données et de la déontologie de l'interprétation. Cette précaution ne concerne pas seulement les questionnaires, mais aussi les analyses de textes. On sait depuis longtemps combien l'analyse de discours cherche à se définir comme une méthode rigoureuse, sans avoir vraiment trouvé son chemin. Aussi, ce que nous allons présenter ici n'est-il qu'une réflexion portant sur un pan de ce vaste chantier et dont le seul mérite est de montrer comment des surfaces linguistiques différentes peuvent renvoyer à des éléments logiques semblables et qu'il n'y a pas d'isomorphisme entre la forme et le contenu, même si ces éléments ne sont pas sans rapport.

Nous nous sommes interrogés sur la possibilité d'exprimer une même idée avec des formulations différentes, car il nous semblait que le discours du citoyen « moyen » pouvait bien souvent exprimer des raisonnements relativement complexes en ayant recours à des formes linguistiques assez simples. Ainsi on peut dire la même chose de trois manières différentes :

- 1- l'inflation salariale ;
- 2- l'inflation dépend de l'augmentation des salaires ;
- 3- si les salaires augmentent alors on observera une inflation.

Dans le premier cas, on a affaire à un syntagme nominal ; dans le second, à une phrase (prédicat) ; dans le dernier, à un raisonnement logique (Si [...] alors).

Bien sûr, chacune de ces formes permet des développements différents. Par exemple, dans le premier cas, l'inflation salariale peut être qualifiée de galopante par simple ajout d'un adjectif, alors qu'il faudrait une toute nouvelle phrase pour le dire dans les autres cas. Le fait de poser l'inflation en termes salariaux permet difficilement de lui opposer un argument, alors que, dans le second cas, il est facile de contredire le verbe (elle dépend / elle ne dépend pas) ou le second terme (les salaires ouvriers / les profits patronaux). Dans le dernier cas, on fait appel à la force du raisonnement logique pour convaincre son partenaire.

L'existence de différentes variantes de mise en discours peut donc être référée à une intention d'orienter sa parole et celle de son interlocuteur. Mais on peut aussi considérer ces trois formes comme tout à fait contingentes à la situation de communication. Si vous laissez au sujet deux lignes pour répondre à une question sur les raisons de la stagnation économique actuelle, il peut se contenter de la première forme. S'il s'agit d'un journaliste qui écrit un article didactique, il peut choisir la dernière.

Dans le cas où le sociologue met seulement l'accent sur le contenu des discours, laissant de côté son argumentation ou les effets de la situation d'interlocution, il convient de remarquer qu'une même information (l'existence d'une inflation dont on attribue la cause aux salaires) peut se trouver exprimée sous trois formes de surface. Or chacune d'elle a donné lieu à une méthode spécifique d'analyse du discours. Si on se contente d'une analyse fréquentielle des syntagmes, on mettra en évidence la première forme et on laissera de côté les deux autres. Si on effectue une analyse syntaxique, on repérera seulement le rapport entre

inflation et salaires dans le second cas. Si on fait une analyse relationnelle des étayages entre les thèmes, on ne pourra observer que le troisième cas de figure. Chaque mode d'analyse ne retiendra qu'une forme linguistique donnée.

La surface n'a jamais été la sémantique, n'en déplaise aux producteurs de logiciels automatiques d'analyse textuelle qui ne laissent au chercheur que la possibilité d'interpréter un graphique final d'analyse de correspondances. Il convient donc de se donner un outil d'analyse de discours qui établisse des équivalences entre les formes de surface.

— En utilisant la première forme (les syntagmes nominaux) le locuteur effectue *des opérations sur les objets du discours*, c'est-à-dire sur les notions exprimées par des mots. Dans notre exemple, le sujet, en parlant d'inflation salariale, établit une relation entre inflation et inflation salariale. Il prend en charge l'idée d'un processus d'inflation ayant pour cause l'augmentation des salaires, et, par là même, l'existence d'agents externes pouvant produire cette augmentation (ici ces agents sont assez abstraits, voire obscurs : ce sont ceux qui définissent les salaires).

— Avec la seconde forme, on a affaire à une prédication à une place : l'inflation est le thème du prédicat « dépend de l'augmentation des salaires ». Ce prédicat est *une propriété affectée par le locuteur* à l'inflation, de la même manière que l'adjectif « salariale » était une propriété affectée à l'inflation. L'agent externe est sous-entendu à travers la nominalisation de l'acte « augmenter » : l'augmentation.

— Avec la troisième forme, le locuteur établit *un rapport d'étayage* de l'objet inflation par l'objet salaire. On peut établir une équivalence formelle entre le premier cas et ce dernier. En effet, dans le premier cas, le récepteur doit extraire, à partir d'un objet complexe (l'inflation salariale), un nouvel objet (les salaires) qui est, par cette extraction même, cause de l'inflation. Il en est de même à travers la forme « Si [...] alors ».

Les instruments de la logique naturelle (Grize, 1991) nous paraissent un bon point de départ pour établir des équivalences et repérer, derrière une nominalisation, le verbe et donc, son sujet; derrière une prédication, l'expression d'une propriété; derrière une connexion logique, le rapport entre objets de discours.

La question n'est donc pas d'identifier les différentes facettes d'une notion par différentes formes de données, mais, au contraire, de repérer la même mise en relations à travers des formes *a priori* différentes. Le rapport entre formes des données et significations ne se pose donc pas en des termes toujours semblables et dont on connaîtrait les modèles canoniques. À chaque analyse de données doit correspondre une réflexion sur ce rapport.

Les géographes nous ont toujours appris que la carte n'était pas le territoire. Il conviendrait que le sociologue ne prenne pas ses données pour des objets sociologiques immédiatement interprétables. Le traitement, sophistiqué ou pas, ne change pas profondément la question du rapport entre les méthodes de cueillette ou de traitement et les objets construits pour l'interprétation. Les facilités du traitement informatique ont seulement obscurci ce rapport. Si le géographe qui établit la carte de l'industrie européenne nous dit bien qu'elle n'est pas celle de l'agriculture ou du peuplement, il n'en est pas de même du sociologue, qui associe une question ou le traitement varié d'un ensemble de questions à un concept qu'il appellera « axe du capital social » ou « comportement libéral ».

La rigueur suppose un minimum de recherche sur les formes d'expression d'un même objet. L'identité locale ne se réduit pas à la publication de maximes, la prototypicalité à l'analyse fréquentielle, et le texte à sa syntaxe. Il convient, dans chaque cas de figure, de se demander comment il est possible d'obtenir plusieurs informations relatives au même objet sociologique et comment leur croisement nous éclaire sur sa nature.

Pierre VERGES
CRES-CNRS,
Centre de recherche en
écologie sociale,
3-5 Avenue Pasteur
13617 Aix-en-Provence Cedex 1
France

RÉSUMÉ

Nous désirons montrer comment il convient de construire des modes de cueillette de l'information qui donnent plusieurs versions d'une même variable par le croisement d'observations ou de traitements. Ce croisement va, à notre sens, bien au delà du croisement classique des variables en analyse de questionnaires ou en analyse de données. L'article s'appuie sur des données de natures diverses pour montrer la généralité du propos. L'auteur en conclut qu'il convient, dans chaque cas de figure, de se demander quelle est la forme du rapport entre les informations recueillies et l'objet sociologique construit, et comment le croisement des informations nous éclaire sur la nature de cet objet.

SUMMARY

The author aims to show how modes of information collection which provide several versions of a variable by means of cross observations and analysis should be constructed. This cross analysis, in our opinion, goes far beyond the classical crossing of variables in questionnaire analysis or data analysis. The author uses data of diverse kinds to illustrate the generality of this thesis. He deduces that in each case one should ask oneself what form the relationship between the information collected and the constructed sociological object takes, and how the cross analysis of information sheds light on the nature of the object.

RESUMEN

Nosotros deseamos mostrar que la manera más conveniente de construir los modos de colecta de la información que dan varias versiones de una variable es por medio del cruzamiento de observaciones o de tratamientos. Este cruzamiento va, según nosotros, mucho más allá del cruzamiento clásico de variables en el análisis de cuestionarios o en el análisis de datos. Nos apoyaremos en datos de naturaleza diversa para mostrar la generalidad del propósito. Se deducirá que conviene en cada caso preguntarse cual es la forma de la relación entre las informaciones recogidas y el objeto sociológico construido y de que manera el cruzamiento de la información nos aclara sobre la naturaleza de este objeto.

BIBLIOGRAPHIE

- BIDART, C., V. GUIDARINI, J. LOUIS-PALLUEL et P. VERGÈS (1991), *Structuration d'une aire urbaine : l'impact des images locales (la zone ouest de Marseille)*, CRES (Centre de Recherche en Écologie Sociale), Marseille.
- BOURGEAT-CARTER, G. (1993), *Le bien-être chez soi : représentation sociale complexe ou complexe de représentations sociales*, thèse de doctorat, Université de Provence.
- DEGENNE, A., P. VERGÈS, (1973), « Introduction à l'analyse de similitude », *Revue française de sociologie*, XIV, pp. 471-512.
- DOISE, W., A. CLEMENCE et F. LORENZI-CIOLDI (1992), *Représentations sociales et analyses de données*, PUG, Grenoble.
- FLAMENT, C. (1981), « L'analyse de similitude : une technique pour les recherches sur les représentations sociales », *Cahiers de psychologie cognitive*, 1, pp. 375-395.
- GRIZE, J. B. (1991), *La logique naturelle*, Ophrys, Paris.
- GRIZE, J. B., P. VERGÈS, A. SILEM (1987), (éds), *Salariés face aux nouvelles technologies*, Éd. du CNRS, Paris.
- GUIMELLI, Ch. et M. L. ROUQUETTE (1992), « Analyse structurale des représentations sociales », *Bulletin de Psychologie* n°405, tome XLV, janvier-février, pp. 196-202.
- HUBALEK, Z. (1982), « Coefficients of association and similarity, based on binary data : an evaluation », *Biological Review*, 57, pp. 669-689.
- JODELET, D. (1989), (éd), *Les représentations sociales*, PUF, Paris.
- ROSCH, E. et B. LLOYD (1978), *Cognition and catégorization*, Erlbaum, Hillsdale.
- VERGÈS, P. (1992), « L'évocation de l'argent : une méthode pour la définition du noyau central d'une représentation », *Bulletin de Psychologie* n°405, tome XLV, janvier-février, pp. 203-209.
- WINDENBERGER, J. (1981), *Des rapports entre la photographie documentaire et des sciences humaines*, CACES (CRES), Aix-en-Provence.