

Réflexions sur la portée sociologique des diverses phases du travail statistique

Alain Desrosières

INSEE

Résumé : Les liens entre l'Insee et la recherche économique sont explicités et institutionnalisés depuis longtemps, en particulier au sein de l'Unité de Recherche de cet Institut, largement centrée sur des recherches économétriques. En revanche, l'apport de l'Insee à la recherche sociologique est plus rarement décrit, et encore moins revendiqué, dans la mesure où la plupart des enquêtes de l'Insee sont conçues à des fins d'éclairage de la gestion économique ou administrative, ou en vue des travaux de prévision utilisant les modèles de la comptabilité nationale, et non explicitement dans le cadre de recherches sociologiques. Pourtant, ces enquêtes et recensements ne constituent pas seulement une mine irremplaçable de données chiffrées pour les chercheurs, mais leur conception, leur préparation, leur réalisation, leur exploitation et l'analyse de leurs résultats sont autant d'occasions de poser, et souvent de résoudre, des problèmes sociologiques sur lesquels les « praticiens chercheurs » de l'Insee ont accumulé une expérience très grande : les difficultés rencontrées sont bien sûr « techniques », mais leur traitement implique une réflexion théorique que l'on voudrait suggérer ici très brièvement, en décrivant les six étapes qui découpent, en partie arbitrairement, les différentes phases d'une opération statistique.

Date de publication : 28/06/2021

Dossier : Sociologie et statistique

Comment citer : 10.52983/crev.vi0.67

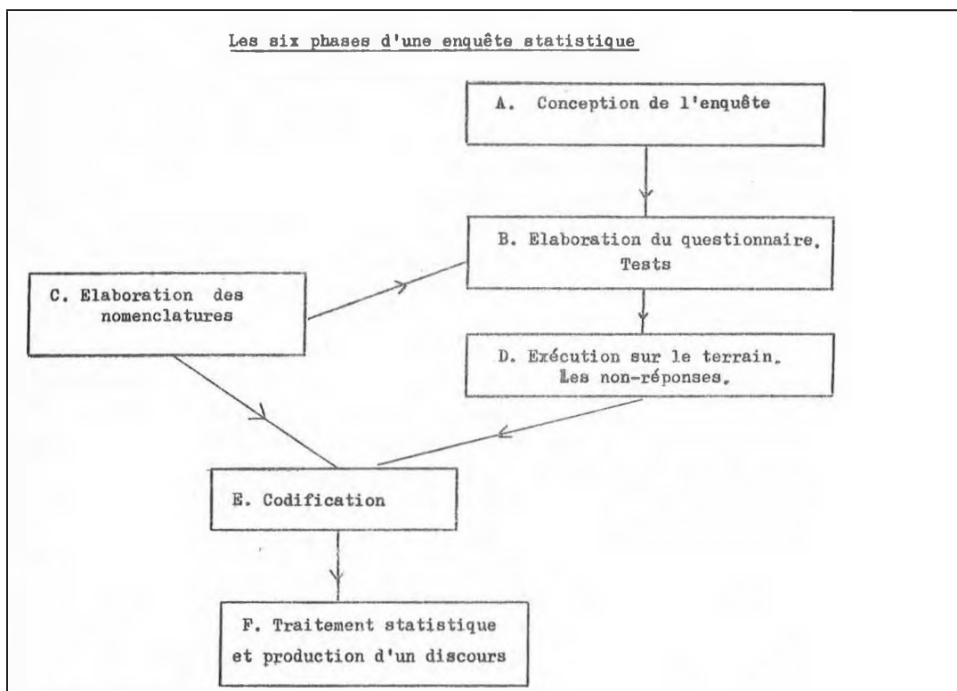
Licence : Cambouis publie ses contenus selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International. Les auteurices gardent leurs droits de propriété intellectuelle pleine et entière sur leurs articles.

Réflexions sur la portée sociologique des diverses phases du travail statistique

Alain Desrosières

INSEE

Les liens entre l'Insee et la recherche économique sont explicités et institutionnalisés depuis longtemps, en particulier au sein de l'Unité de Recherche de cet Institut, largement centrée sur des recherches économétriques. En revanche, l'apport de l'Insee à la recherche sociologique est plus rarement décrit, et encore moins revendiqué, dans la mesure où la plupart des enquêtes de l'Insee sont conçues à des fins d'éclairage de la gestion économique ou administrative, ou en vue des travaux de prévision utilisant les modèles de la comptabilité nationale, et non explicitement dans le cadre de recherches sociologiques. Pourtant, ces enquêtes et recensements ne constituent pas seulement une mine irremplaçable de données chiffrées pour les chercheurs, mais leur conception, leur préparation, leur réalisation, leur exploitation et l'analyse de leurs résultats sont autant d'occasions de poser, et souvent de résoudre, des problèmes sociologiques sur lesquels les « praticiens chercheurs » de l'Insee ont accumulé une expérience très grande : les difficultés rencontrées sont bien sûr « techniques », mais leur traitement implique une réflexion théorique que l'on voudrait suggérer ici très brièvement, en décrivant les six étapes qui découpent, en partie arbitrairement, les différentes phases d'une opération statistique.



Les six phases d'une enquête statistique

La carte et le territoire

Un fil conducteur de ce qui suit est l'examen de la tendance, inscrite dans la logique même du travail statistique, à autonomiser l'instrument, à l'extraire de ses conditions de production, et à le manipuler comme s'il existait par lui-même, remplaçant ainsi la vie économique par un schéma de comptabilité nationale et un modèle économétrique, la reproduction sociale par un tableau de mobilité sociale, l'histoire par des séries chronologiques, le territoire par la carte. Bien sûr, il ne s'agit pas de contester la nécessité et la fécondité de la construction d'outils effectuant des coupes dans le réel et d'en approfondir la logique interne, mais plutôt de prendre pour point de vue une réflexion sur les différentes étapes de cette opération d'abstraction que constitue la construction d'un objet statistique. C'est aussi une autre façon de poser la question que le statisticien se pose parfois sous le nom de « fiabilité » quand il s'inquiète sur les « erreurs » qui affectent ses mesures. Les « erreurs » d'observation sont la retraduction, dans le vocabulaire statistique, du système d'interactions complexes conduisant à une mesure plutôt qu'à une autre.

Phase A : conception de l'enquête

Les travaux menés sur l'histoire de la statistique¹ ont montré que les opérations statistiques (enquêtes, exploitations secondaires et fichiers administratifs) ont surtout été impulsées par la croissance de la comptabilité nationale et de la planification. Celles-ci ont fourni à la fois questions et moyens, surtout à partir de la nomination de Claude Gruson à la direction de l'Insee en 1962, après toutefois une phase importante d'invention et d'expérimentation à petite échelle, dans les années 1950. Si cette particularité a contribué à accroître massivement le nombre et la taille des enquêtes (par exemple : budgets de familles, formation et qualification professionnelles (FQP), enquêtes emplois, enquêtes revenus...), les interrogations proprement sociologiques ont souvent été subordonnées au souci d'alimenter des schémas comptables ou des modèles de prévision économique. Dans ceux-ci, les catégories interviennent plus par leurs articulations mutuelles, comptables ou économétriques, qu'en tant qu'enjeux de conflits, de polarisations. La tendance à la réification des catégories est en partie inscrite dans la logique même de la modélisation. Cependant, la division du travail entre réalisateurs des enquêtes et utilisateurs des chiffres (opposition classique entre « statisticiens » et « comptables nationaux ») a contribué à maintenir chez les premiers une sensibilité spécifique à la dimension « pratique » de ces enquêtes, par les contraintes mêmes de leur gestion (le « contact avec le terrain » opposé à l'« irréalisme des modèles »). Ce savoir pratique constitue un ensemble de connaissances implicites résultant de la gestion rationalisée d'une activité de collecte périodiquement reproduite.

La transformation de cet ensemble de consignes ou conseils opératoires en un savoir organisé en tant que tel pourrait constituer un apport important de la statistique à la sociologie.

L'importance accordée dès 1950 au critère socio-professionnel, puis l'introduction de questions sur l'origine sociale dans l'enquête Emploi de 1953, dans l'enquête FQP de 1964, l'organisation d'enquêtes sur l'écoute de la radio, les loisirs ou les budgets-temps, témoignent d'interrogations ne correspondant pas directement aux demandes de la planification ou de la modélisation économique. Les démographes ont introduit très tôt le critère socio-professionnel dans l'étude de la fécondité ou de la mortalité, de la nuptialité ou du divorce. Le souci d'appréhender un espace social complexe, en tous cas plus complexe que ne le laisserait supposer l'image de l'échelle de revenus, ou de

¹ En particulier pour le Colloque de Vaucresson en 1976 (Insee, 1977 ; Insee, 1987).

prestige, est bien présent dans les enquêtes impliquant l'usage de multiples critères. Cependant, l'interrogation plus fondamentale sur la façon dont les « variables » s'articulent logiquement et statistiquement est en général renvoyée aux phases ultérieures du travail : construction et analyse de tableaux, rédactions et commentaires.

Phase B : préparation du questionnaire, tests, plans de sondages

Les différentes étapes de la préparation du questionnaire sont marquées de contraintes contradictoires : insérer le plus grand nombre de questions possibles, les adapter aux possibilités de réponses en contrôlant leur sens, en diminuer le nombre afin de parvenir à un temps de passation jugé possible compte tenu des réactions recueillies dans les tests successifs. Cette phase est l'une des plus riches sociologiquement car, à ce moment-là, tous les problèmes sont encore ouverts : les comptes rendus des tests permettent de juger si les questions et leurs formulations ont quelque pertinence par rapport aux situations concrètes des enquêtés, différentes selon les milieux sociaux, l'âge, le type d'habitat. Ainsi, par exemple, la signification de la relation au travail, au chômage, à la recherche d'un emploi diffère selon que la personne interrogée est salariée ou non, entrée dans la vie active, est une femme mariée ne travaillant plus depuis longtemps, ou proche de la retraite... et, plus fondamentalement, selon sa place dans l'espace social. C'est à ce stade qu'on peut voir si une question, par exemple imposée par un questionnement « omnibus », apparaît inappropriée pour certaines catégories (si ce n'est pas pour toutes), induisant des réponses (ou non-réponses) « artefacts », ou si elle correspond à une alternative inscrite dans la situation spécifique. La contrainte de passation de questionnaires identiques (ou quasi-identiques, aux filtres près) à toute une population impose souvent ce questionnement-artefact. La technique du filtre ne résout que très partiellement la question, car elle nécessite de trier de façon apparemment rigoureuse des situations justement floues, aux frontières par exemple entre activité et inactivité, entre travail ouvrier et travail non ouvrier, alors qu'un des objectifs sociologiques de l'enquête peut être justement de mieux connaître l'épaisseur et la consistance de cette zone floue.

Une autre question essentielle posée à ce stade est celle de la « transformation du qualitatif en discret », c'est-à-dire la « réduction » de situations complexes et multidimensionnelles en un nombre fini d'« items ». Ce nombre risque même d'être d'autant plus petit que l'enquête portant sur des sujets moins routinisés (et plus « qualitatifs ») dispose de crédits moins importants et donc d'un échantillon plus petit. Cette « perte » est au cœur de la problématique des rapports de la sociologie et de la statistique : plutôt que de la déplorer, ou d'opposer rituellement les démarches qualitatives (ou « ethnologiques ») et quantitatives, il est plus riche d'observer les « bords » de l'activité statistique, c'est-à-dire tout ce que révèle son déroulement social même : cette « perte » et cette « récupération de sens » constituent la matière de l'échange. La « frontière » entre les deux types d'activité est d'ailleurs mouvante, dans la mesure où des techniques permettent par exemple d'analyser statistiquement des « questions ouvertes » (Lebart, 1982).

La réflexion sur l'élaboration du questionnaire et les tests ne peut être disjointe de celle sur les nomenclatures : le questionnement dépend directement de la structure de celles-ci, de leur degré de précision. De même, des questions *a priori* ouvertes dans les préquestionnaires (de type « opinion ») peuvent être précodées dans le questionnaire définitif et, à ce stade, l'information pertinente perdue si la spécificité de la situation d'enquête a été oubliée².

2 Un exemple d'une telle méprise : dans une enquête portant sur des objets très variés, une question tente d'appréhender la fréquence de la fréquentation du café. Trois items sont prévus : une fois par jour et plus ; plusieurs fois par semaine ; moins souvent et jamais. Ils recueillent respectivement : 8, 12 et 80 % des réponses. Il est probable que cette formulation et la situation d'enquête en présence d'enquêteurs induisent, mais on ne sait comment, une réponse au 3^e item. La question était donc mal posée.

La phase du précodage implique donc déjà une bonne connaissance de la portée de la question posée, des enjeux impliqués par la situation d'enquête, de la perception de l'enquêteur et du questionnaire par les enquêtés, points qui seront évoqués au paragraphe D ci-dessous.

Plans de sondage, stratification et représentativité

L'objectif premier des grosses enquêtes par sondages auprès des ménages, conçues et expérimentées dans les années 1950, réalisées en grand à partir des années 1960, était de réaliser des « maquettes » aussi fidèles que possible de la population d'ensemble pour étudier les grandes oppositions pertinentes sur le plan socio-économique mais aussi, et de plus en plus à partir des années 1960, pour alimenter en chiffres les cadres comptables nécessaires à la modélisation économique, et donc procéder à des estimations. Pour cela, la *représentativité* des échantillons était supposée garantir à ceux-ci la meilleure homothétie possible par rapport à la population globale, et un moyen pour améliorer *a priori* cette représentativité était la *stratification*, puisque, en introduisant dès l'abord dans le plan de sondage les découpages estimés *a priori* les plus pertinents, on diminuait l'« erreur aléatoire » due à la procédure de sondage. Mais, paradoxalement, ceci supposait résoudre la question du choix de ces critères les plus pertinents et, surtout, reportait sur la strate de sondage l'hypothèse d'homogénéité, nécessaire à l'application de la théorie probabiliste. Et, pourtant, l'idée même de la stratification portait bien en germe l'intuition de l'hétérogénéité de l'espace social, mais dans la perspective comptable et économiste de l'amélioration statistique des estimations et non afin d'approfondir la logique structurelle de cette hétérogénéité. Dans cette logique d'estimation comptable, les diverses procédures de *redressement*, destinées à annuler les différences de taux de sondage par strates puis celles d'*estimations des non-réponses* (hotdeck...), contribuent à accroître la distance entre l'échantillon brut et le produit fabriqué et détournent donc l'attention du statisticien d'un examen proprement sociologique de l'activité d'enquête, sinon à des fins d'amélioration de la « fiabilité » : ceci est un des paradoxes de l'activité du statisticien d'enquête.

Phase C : élaboration des nomenclatures

Sur ce point, des réflexions nombreuses ont déjà été menées (Thévenot, 1981 ; Desrosières et Thévenot, 1979). S'il est clair que tout objet sociologique et donc toute enquête supposent une réflexion spécifique sur les nomenclatures *ad hoc*, les opérations statistiques de l'Insee, destinées d'une part à des usages multiples et conçues d'autre part pour être rapprochées les unes des autres, impliquent de ce fait des nomenclatures « omnibus », utilisées pendant une assez longue période, dont le code des catégories socio-professionnelles est le prototype. Leur élaboration nécessite donc tout un travail spécifique, intégrant et dépassant les usages variés prévus, dans l'institution statistique et au-dehors.

Le produit élaboré (par exemple : la nomenclature d'activités et de produits NAP, les CSP) doit simultanément avoir une cohérence logique d'ensemble et respecter les clivages et critères pertinents localement dans les diverses zones de l'espace analysé, qu'il est donc nécessaire d'explorer *a priori* minutieusement. La difficulté est donc de passer d'un ensemble de distinctions locales, indigènes, partielles, conflictuelles, reflétant des systèmes de représentations sociales complexes, à un découpage exhaustif, passe-partout, intégrant au mieux les clivages supposés les plus pertinents au vu de l'ensemble des recherches sociologiques déjà menées, et ceci dans un cadre

institutionnel dans lequel les agents eux-mêmes sont en partie consultés, à travers le filtre de représentations professionnelles. Si cette activité constitue une position d'observation hors pair, elle est aussi un lieu de contradictions quasi insolubles. Elle a néanmoins été le moteur d'une recherche très neuve sur le rapport entre la catégorisation statistique et les classements indigènes, sur les opérations mentales et sociales associées à la constitution et à la manipulation de catégories. Au moment de la fabrication des catégories, on a affaire *a priori* à une juxtaposition très partiellement cohérente de désignations souvent floues, enchevêtrées, de statuts sociaux et logiques de niveaux différents, porteurs d'enjeux sociaux plus ou moins marqués et dont les contenus sont plus ou moins solidifiés, inscrits juridiquement ou relevant seulement de l'« usage ». Tout le travail du taxinomiste se traduit en fin de compte par une arborescence d'« intitulés ». Chacun d'entre eux est structuré par un ou des cas-types exemplifiant la catégorie, et souvent lui donnant son nom, et une grappe de « bas voisins » ou « assimilés » jusqu'à des frontières, lieux conflictuels où les négociations institutionnelles ont pu être âpres, et où il a fallu trancher dans le vif. L'ensemble est organisé en « arbre » de niveaux différents, et chacun de ces niveaux pose à nouveau le problème des CSP-types, des intitulés de catégories et des frontières.

Par sa place dans le processus de production statistique, l'élaboration des nomenclatures est sociologiquement paradoxale : elle suppose résolu ou au moins largement explorés les problèmes mêmes de la recherche, c'est-à-dire la mise au jour des clivages pertinents. Grand est donc le danger de voir *a posteriori* restituées par les recherches les dimensions mêmes introduites dans la nomenclature et, plus en amont, les catégories sociales indigènes inscrites dans la pratique sociale historique. L'histoire de la catégorie « cadres » minutieusement analysée par Luc Boltanski (1982) en fournit un exemple. De même, la suppression, vers 1960, de la distinction « public-privé » du critère socio-professionnel a fait disparaître pour vingt ans l'analyse de cette opposition de maints travaux sociologiques.

Les problèmes évoqués ici ne se posent pas de façon abstraite, à l'occasion d'une réflexion « théorique générale » sur l'espace social, mais, très concrètement, à diverses phases des opérations statistiques.

Tout d'abord, les arbitrages toujours difficiles sur la lourdeur et la longueur des questionnaires (phase B) impliquent des choix sur les critères de tri utilisés ensuite, et donc sur la finesse des nomenclatures mises en œuvre. Ainsi la partie « professionnelle » d'un questionnaire sociologique portant sur de tout autres aspects de la vie sociale pourra être plus ou moins détaillée selon l'importance accordée à la finesse du découpage socio-professionnel et à l'importance de tel ou tel clivage.

Puis les critères ainsi introduits ne seront mis en œuvre que si, *sur le terrain* (phase D), la question est pertinente et suscite une réponse.

Ainsi, par exemple, la distinction entre ouvriers spécialisés et qualifiés, pertinente dans d'importants secteurs de l'industrie, n'en a plus du tout dans d'autres et suscite soit des « non-réponses », soit des réponses aléatoires en apparence et dont le décryptage est complexe et pose de nouvelles questions sociologiques. Au moment de la *codification* (phase E), la logique de construction est à nouveau rudement mise à l'épreuve, confrontée à l'afflux des cas ambigus, non ou mal prévus, susceptibles de plusieurs classements également plausibles au vu de l'ensemble des réponses au questionnaire. À ce point de vue, un atelier de chiffrage est, on le verra ci-dessous, un des lieux d'interrogation et de décision sociologiques les plus riches qui soient, posant concrètement des questions dont la diversité et la complexité dépassent largement tout ce qui avait pu être imaginé au moment de la conception de la nomenclature. On verra également ci-dessous le poids des décisions concernant

les nomenclatures sur les problèmes de tabulation et d'analyses statistiques (phase F) et sur la production du discours et du sens qui en résultent (phase G).

Phase D : exécution des enquêtes sur le terrain

Cette phase met en contact des enquêteurs munis d'un questionnaire³ et des individus extraits partiellement pour un moment de leur vie quotidienne. Comment se présente pour eux ce détournement de temps et d'informations, plus ou moins imposé, les plaçant dans une situation de contrainte, même si le refus est apparemment possible (et effectif dans des proportions allant de 10 % à 20 % dans beaucoup d'enquêtes) ? L'image qu'ils ont de l'enquêteur et de l'institution qui l'envoie, les conséquences supposées de leurs réponses sur leur propre situation ou sur l'image qu'ils livrent à l'enquêteur, la réactivation éventuelle pour eux d'une situation de type scolaire (avec la distinction entre « bonnes » et « mauvaises » réponses), le temps et l'énergie qu'ils consacrent à l'enquête, tout cela est étroitement lié à la place de la personne dans l'espace social, à sa trajectoire, à son image de lui-même, que la situation d'enquête, non choisie, lui renvoie parfois impitoyablement. Les *refus de réponse* sont la matérialisation la plus immédiate et visible de cette non-neutralité sociologique de la situation d'enquête. Ce refus peut se situer soit dès la prise de contact, de façon globale, soit au fur et à mesure du déroulement d'un questionnaire, face à certaines réponses. Une analyse menée par Jean-Louis Borkowski (1978) a montré que le premier type de refus est particulièrement marqué chez les petits patrons, artisans et commerçants et chez les employés de bureau. Dans le cas des petits patrons, ces refus peuvent être mis en relation avec une perception marquée de l'opposition entre les sphères « publique » et « privée » (la première étant ressentie comme menaçante) décrite par François de Singly (1982).

Pour les employés de bureau, pourrait être maximum la crainte de non-conformité à des normes ressenties comme extérieures et néanmoins très intériorisées.

Les refus de répondre à certaines questions d'un questionnaire accepté par ailleurs peuvent résulter de conflits entre deux postures perçues comme légitimes et néanmoins contradictoires. Ainsi des questions sur la tolérance sexuelle pourront susciter de tels refus. De même, une interrogation renvoyant, même indirectement, à un jugement sur sa propre position dans l'espace social peut déclencher une contradiction insoluble. Ainsi, dans une enquête portant sur un grand nombre de sujets (Desrosières et Gollac, 1982), les refus de réponse à des questions précises, une fois le questionnaire accepté, ont été très faibles, n'excédant presque jamais 3 %. Cependant, dans une série d'items portant sur des appréciations sur le travail, l'un d'entre eux a suscité près de 10 % de refus : « estimez-vous que votre travail est... mal payé ? » Si la réponse « Oui » reflète une attitude explicitement revendicative ayant une cohérence claire, la réponse « Non » reflète une contradiction entre la crainte d'une telle attitude revendicative et le souhait de gagner plus. Le refus de réponse peut être une solution, et il est significatif que ce refus soit maximum dans le cas des *contremaîtres*, dont la position charnière dans l'espace social rend encore plus difficile l'affirmation d'une attitude clairement revendicative : ils appartiennent par leurs trajectoires sociales au champ des ouvriers, et, étant plus payés qu'eux pour les surveiller et subissant de plein fouet les contradictions des contraintes de la production, il leur est plus difficile à la fois de s'estimer mal payés et de s'estimer bien payés.

3 Ou seulement le questionnaire, comme pour le recensement, où l'agent recenseur n'est pas toujours physiquement présent, ou pour les enquêtes postales. L'idée répandue est que ces dernières ont un taux de réponse beaucoup plus faible. Ce n'est pas toujours vrai, dans le cas où les personnes trouvent ou imaginent un « intérêt », dans les deux sens du mot, à répondre, comme cela a été le cas dans des enquêtes auprès de jeunes, et portant sur leurs trajectoires scolaires et professionnelles.

Comme on le voit, les commentaires ainsi faits ne sont probants que pour autant qu'ils renvoient à d'autres traits du comportement des agents, ce qui sera développé en F. Il est cependant nécessaire, dès cette phase, de mettre en relation les postures face à l'enquête et à l'enquêteur, avec ce que l'on sait par ailleurs des logiques de comportement des divers milieux sociaux, sans d'ailleurs que ces logiques d'ensemble induisent univoquement telle ou telle attitude ; mais chacune d'entre elles devrait pouvoir être réinterprétée par rapport, d'une part, à la logique de comportement du groupe social de la personne et, d'autre part, à la place relative et à la trajectoire (ascendante, descendante, stable) de cette personne dans son groupe. La combinaison de ces divers éléments et la variété des situations possibles expliquent la diversité des attitudes et l'absence de correspondance simple.

Ainsi, par exemple, en milieu populaire, cette attitude pourra osciller entre la bonne volonté et la coopération sans aucune trace de contestation, et l'acceptation du bout des lèvres d'une contrainte extérieure subie comme beaucoup d'autres et devant laquelle la seule attitude possible est de réduire au minimum la participation à l'interaction, de crainte de se dévoiler devant « eux » (Hoggart, 1967).

Ces deux postures seront à associer à la pente de la trajectoire dans le monde ouvrier, à l'accumulation ou non d'échecs scolaires ou professionnels. Dans les classes moyennes, le souci de se couler dans la logique formelle apparente du questionnaire, assimilé à un questionnement scolaire, fréquent dans les groupes salariés (cadres moyens et employés) s'opposera à la méfiance et à l'hostilité de petits patrons, réticents, on l'a vu, à toute incursion dans leur « vie privée » assimilée à une intervention étatique ou à une inquisition.

Enfin, en milieu supérieur (ou moyen plutôt intellectuel) pourra apparaître une attitude dédaigneuse ou critique plus théorisée ou explicitée en termes idéologiques, sans que l'attitude respectueuse de l'ordre disparaisse pour autant. Dans ce dernier cas, la position de classe pourra transparaître par une interprétation du questionnaire, un commentaire, une critique du caractère « réducteur » des items proposés. C'est alors que des questions ouvertes, posées en fin de questionnaire, du type : « Manque-t-il des questions dans ce questionnaire ? » « Que pensez-vous de cette enquête ? » susciteront le maximum de réponses.

La position et la trajectoire de l'enquêté dans l'espace social sont d'un poids décisif dans sa manière d'être vis-à-vis de l'enquête (un exemple est analysé de façon détaillée dans (Thévenot, 1981, p. 18-19). Par ailleurs, celle de l'enquêteur et les situations relatives de l'un et de l'autre modifient également la nature de l'interaction. Ainsi, dans l'enquête évoquée ci-dessus, effectuée par une dizaine d'enquêteurs, une moitié d'entre eux environ était des enquêteurs Insee « classiques », issus plutôt des classes moyennes ou supérieures, femmes inactives ou instituteurs en retraite, formés à une application très scrupuleuse de consignes soigneusement codifiées et dotés d'un grand « savoir-faire » pratique pour situer leurs interlocuteurs, et éventuellement trouver les arguments convaincants en cas de refus de réponse. Leur comportement, proche d'un questionnement administratif classique, induit chez l'enquêté une attitude soumise et de faibles développements « à la marge », qui expliciteraient tel ou tel cas ambigu ou complexe. En revanche, l'autre moitié, formée d'étudiants confirmés ou de jeunes chercheurs en sciences sociales, dotés d'une apparence extérieure moins conventionnelle et d'une plus grande curiosité par rapport aux objectifs de l'enquête, mais moins à l'aise pour « forcer la main » aux enquêtés réticents, pouvaient rapporter des questionnaires plus riches d'informations et, par-là, parfois plus faciles à codifier, mais pouvant aussi poser des problèmes insolubles pour cette codification, dans la mesure où ce fonctionnement plus artisanal est contraire à la logique de la chaîne et de la division du travail impliquée par la gestion de grosses enquêtes décentralisées.

De cette diversité des relations d'interactions entre enquêteur et enquêté peut résulter le « flou de déclaration », c'est-à-dire le fait qu'une personne enquêtée, soumise plusieurs fois au même questionnement, peut varier ses réponses. Ceci est fréquemment le cas pour une question « ouverte » comme la profession, mais peut l'être pour une question fermée, d'opinion par exemple. Comme l'a déjà montré Laurent Thévenot (1981), ce flou est un indicateur à la fois du degré de solidification sociale (juridique, verbale...) de la catégorie envisagée, et de la posture dans laquelle se trouve l'enquêté vis-à-vis de l'enquêteur, ou aussi de son environnement à ce moment : conjoint, enfants, etc.

4 La situation n'est qu'en partie différente dans le cas de procédures automatiques telles que COLIBRI, car les cas où la chiffruse doit prendre une décision restent nombreux.

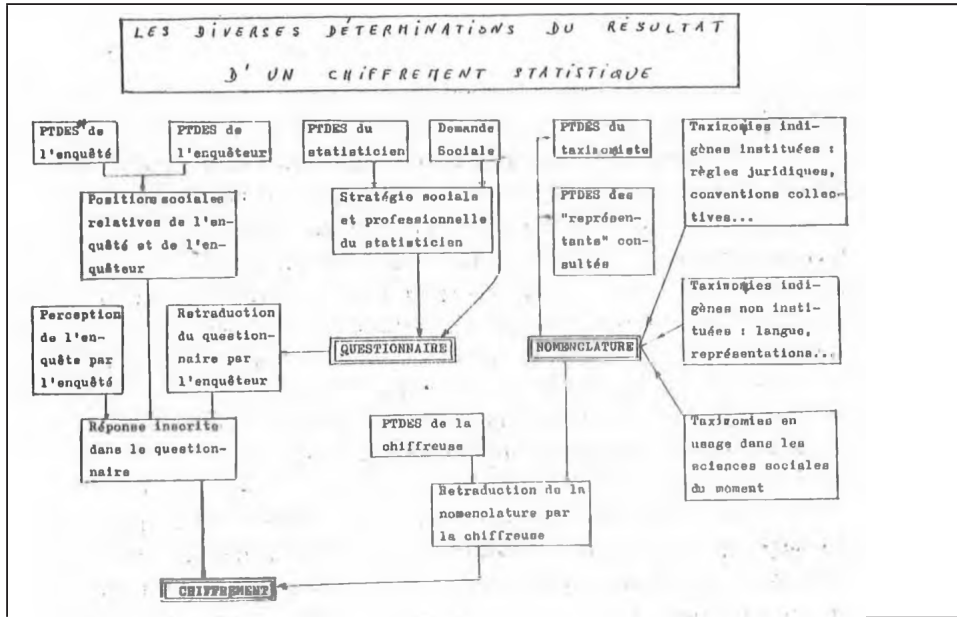
Phase E : la codification

La codification est le moment douloureux où sont mises à l'épreuve tant les nomenclatures que les questions posées. Le réel observé, même réduit dans la grille du questionnaire, est toujours plus complexe que les cas prévus et, surtout, à ce stade, la diversité et la complexité des cas, encore perceptibles à la lecture d'un questionnaire, vont être fondues en « catégories » dotées d'un intitulé et de frontières discrètes. Mais ce qui est « perdu » n'est pas seulement de l'ordre du détail, de la finesse, comme si une nomenclature plus détaillée pouvait sauver idéalement le foisonnement du réel.

C'est aussi ce qui fait qu'une catégorie statistique est en même temps un espace relationnel, auquel on peut appartenir « plus ou moins » selon les points de vue, polarisé dans diverses directions, doté de « cas assimilés » dont la logique de l'assimilation disparaît à jamais dans la codification, comme par exemple ces « étalagistes » qui, par une série d'« assimilations » en cascade dont chacune peut se justifier, sont groupés avec les « instituteurs », par l'intermédiaire des « publicitaires » et « professions intellectuelles diverses », dans le code des CSP utilisé de 1954 à 1981.

Les difficultés et ambiguïtés rencontrées dans le chiffrage ne sont pas seulement un obstacle à une gestion efficace de cette activité nécessairement industrialisée et routinisée, mais aussi un indicateur précieux du degré d'existence sociale et de fixation des catégories utilisées. À ce titre, le flou observé à ce stade, de même que le flou observé au moment des déclarations elles-mêmes, sont, on l'a vu (Thévenot, 1981), des révélateurs à la fois de la topographie et des rapports symboliques inhérents aux divers champs analysés.

Est à ce moment en jeu la façon dont les personnes responsables du chiffrage se représentant l'espace social, ou plutôt la façon dont elles retraduisent le document « nomenclatures » à travers leur représentation⁴. On pourrait représenter la série des effets conduisant à un chiffrage par le schéma suivant, résumant en partie ce qui précède.



Les diverses déterminations du résultat d'un chiffrement statistique
PTDES = Position et Trajectoire dans l'Espace Social

Les opérations de *recodification* se trouvent à la charnière entre cette phase et la suivante : en combinant, recomposant, regroupant des modalités d'une ou plusieurs variables, on fabrique de nouvelles variables supposées améliorer la mise en évidence de tel système de différences ou constituer un regroupement d'agents ayant une cohérence sociale (par exemple : les CSP comme code calculé complexe à partir de profession, statut, qualification, activité économique). Tranchant, par exemple, dans des variables à une dimension (âge, revenu, taille de commune), la recodification peut faire apparaître (ou dissimuler parfois) telles variations non monotones (courbes en U), présomptions de logiques causales non univoques le long d'un axe (Desrosières et Thévenot, 1979). Inscrite dans la logique des opérations statistiques, par exemple en raison de la petite taille d'un échantillon, la recodification peut gommer des différences locales ou des effets de « queue de distribution », pertinents même s'ils portent sur peu d'individus.

Phase F : traitement statistique et production d'un discours

La lourdeur et la longueur (souvent plusieurs années) des opérations nécessaires à la réalisation d'une enquête statistique créent une relative autonomie pour chaque phase. Souvent, même les responsables, dont les rythmes moyens de mobilité dans l'institution sont du même ordre de grandeur, sinon inférieurs à la durée complète d'une opération, de la conception à son commentaire, ont changé au cours de son déroulement. L'énergie et la mobilisation intellectuelles nécessaires à chaque phase rendent difficiles les « boucles » et, au moment où commence l'exploitation, certains passages sont presque imposés par la tradition. Quels sont-ils ?

La croissance des enquêtes ayant été liée historiquement aux demandes des comptables nationaux et des modélisateurs économiques, certains modes de traitements statistiques, à l'origine conçus pour ces besoins, sont devenus presque nécessaires, comme par exemple des calculs de divers types d'élasticité de consommation par rapport au revenu, pour les enquêtes « budgets de famille », ou les tables de mobilité professionnelle pour les enquêtes FQP : la prévision économique et sociale induit une demande de « régularités statistiques » dans des liaisons entre variables, susceptibles d'être prolongées,

« extrapolées » dans le futur. De même, la logique de la modélisation économique induit une demande de « causalité », c'est-à-dire d'enchaînements de variables, telles que certaines d'entre elles « expliquent » statistiquement (au sens des liaisons économétriques) certaines autres, l'ensemble de ces liaisons s'ajoutant à celles de « cohérences comptables » pour former une architecture dont les piliers sont des *variables* et des *agents* définis moins par la cohérence de leurs comportements et leurs *différences* internes et mutuelles que par la régularité de flux, monétaires ou humains, qui les relie. Peut-on, par rapport à ce type de traitements à finalité macroéconomique, reconstituer une logique de traitement statistique de finalité plus proprement sociologique ? Si cette logique existe, quelle est son autonomie, et comment a-t-elle été influencée par la logique économique, plutôt dominante, et en tous cas beaucoup plus liée à une demande sociale évidente ?

Par opposition à la première, le cœur de la logique sociologique pourrait être la recherche de différences et, plus précisément, de la cohérence entre des systèmes de *différences* (Desrosières et Thévenot, 1979) compte tenu de ce que les cohérences globales et synthétiques des comportements des agents, résultants de leurs positions dans un grand nombre de champs, apparaissent indirectement, dans l'enquête statistique, sous la forme de ces systèmes de différences, révélés par le jeu des diverses opérations de tris, à plat puis croisés, ou d'analyses plus sophistiquées de plus en plus synthétiques comme, par exemple, les analyses de variances, les études d'effet de structure, les analyses de correspondances, les constructions de typologies... Le travail sociologique du statisticien consiste donc souvent à inférer des cohérences d'agents ou groupes d'agents, elles-mêmes relationnelles, à partir de séries discontinues⁵ de variations, relatives à des observations inscrites dans des chiffrements qui résultent eux aussi, on l'a vu, d'un processus social très complexe.

Parmi les traitements que le statisticien opère sur son fichier, des plus simples aux plus complexes, tous ceux qui relèvent de la logique ci-dessus sont de portée sociologique, même s'ils sont moins pensés de façon systématique que les traitements à fins économiques.

Soit par exemple l'analyse d'un tableau croisé distribuant une population selon deux critères additifs, ou bien encore décrivant une mesure effectuée sur une population répartie selon deux critères croisés de tri. Les diverses procédures possibles susceptibles de provoquer les commentaires de ces tableaux tendront à apprécier l'*écart* entre un tel tableau et un autre jugé ne contenir aucune « information », qui pourra être, selon le cas, le tableau produit des marges ou le tableau réduit à une diagonale. Ces notions d'écart et d'information peuvent recevoir un contenu mathématique (Volle, 1974), mais leur référence sociologique est toujours de l'ordre de la différence, ou plutôt des *différences de différences* (Passeron et de Singly 1981). Ainsi, la comparaison la plus simple de deux ratios (calculs de pourcentage) implique la vérification de la significativité de cette différence, par un test du χ^2 appliqué à un tableau à deux lignes et deux colonnes : un tel calcul porte bien sur une différence de différences. Sur un tableau plus complexe, une *analyse de variance* visera à pondérer l'effet statistique de divers modes de découpage de la population étudiée sur une observation, sans préjuger, pour autant, de causalités éventuelles, mais orientant vers des présomptions de causalité, issues de coïncidences répétitives.

Une variable peut en cacher une autre

L'analyse des *effets de structure* apparaît spontanément dès que sont postulées de telles causalités, dont certaines sont jugées trop immédiates et susceptibles d'en dissimuler d'autres, plus latentes. Soit par exemple une recherche visant à décrire l'effet de l'appartenance à un milieu social sur tel comportement.

5 Discontinues dans la mesure où chacune est relative à une observation originale, résultant d'une procédure, d'un questionnaire qualitativement distinct des autres, même si leurs relations sont étroites.

L'échantillon a permis de décrire ce comportement selon le milieu social, l'âge et la taille de la commune de résidence. Le premier tri à plat selon le milieu social révèle bien sûr des différences, mais la tournure d'esprit du statisticien implique immédiatement l'objection : les catégories sociales sont inégalement constituées de personnes des diverses tranches d'âge ou habitant les divers types de communes. Il y a donc lieu d'éliminer les « effets » de ces deux dernières variables, jugés ici parasites, en construisant une *population fictive* dont la structure par catégorie sociale serait la même que la population réelle, mais dont chacune d'entre elles aurait les mêmes répartitions par tranche d'âges et par type de communes, en l'occurrence celles de la population totale. Les différences entre milieux sociaux alors observées sur cette population fictive, en général plus faibles que les premières concernant la population réelle, sont jugées plus « pures », reflétant plus strictement un effet « milieu social ». Ce raisonnement, dont il existe maintes variantes, est inscrit dans une logique de décomposition statistique d'effets statistiques, c'est-à-dire portant sur des *catégories* dont chacune enfermerait substantiellement et exhaustivement sa signification, même si la logique de la recherche d'effets de structures repose justement sur l'idée qu'*une variable peut en cacher une autre*. La posture du statisticien le pousse donc à traquer toujours plus les « effets de structure » parasites, quitte à fabriquer des populations fictives où la Lozère comprend en partie une grande métropole urbaine et où Paris est peuplé d'une fraction appréciable d'agriculteurs ou de salariés agricoles. Ces derniers exemples sont bien sûr irréels : aucun statisticien ne commentera de telles populations fictives, mais le paradoxe est néanmoins fécond. La recherche d'effets de structure ne tend-elle pas à estomper de fait les « co-occurrences », sociologiquement aussi pertinentes que les « effets croisés », même si, paradoxalement, elles en sont le point de départ ? Ainsi, par exemple, les non-salariés, petits patrons et surtout agriculteurs sont en moyenne plus âgés que les salariés et habitent beaucoup plus fréquemment les communes rurales ou les petits bourgs. Parmi les salariés, les ouvriers sont un peu plus jeunes et habitent plutôt dans des communes petites ou moyennes. Ces co-occurrences sont essentielles pour comprendre les constellations de propriétés caractéristiques des diverses catégories et la façon dont opère sociologiquement (et non statistiquement) « l'effet de la catégorie ». Le vocabulaire statistique est d'ailleurs trompeur : ce n'est pas l'appartenance à une catégorie qui « opère », mais la position (et aussi le mouvement) dans des champs variés et multivoques, définis non pas substantiellement par des frontières ayant une objectivité matérielle mais à travers les systèmes de perceptions, d'assignations que leur confèrent les agents appartenant aux uns et autres champs. Plus précisément, cette position implique en *probabilité* un ensemble de comportements liés entre eux, et les divers tableaux statistiques, à plat ou croisés, ou même les analyses de correspondance plus sophistiquées, sont des coupes selon des directions variées de ces espaces multidimensionnels.

L'analyse factorielle : instrument de cumul ou artefact ?

L'analyse des correspondances, ou certaines analyses typologiques, fournissent des images partielles de ces multiplicités et de ces co-occurrences, même si elles risquent de réifier non plus des variables ou des catégories mais les objets plus mystérieux et plus difficiles à critiquer, parce que plus synthétiques et plus opaques, que sont les *axes* ou *a fortiori* les regroupements *ad hoc* des analyses typologiques. Les axes des analyses de correspondances, dont certains sont assez stables d'une analyse à l'autre (à commencer par le « grand axe » opposant les puissants aux dépossédés), risquent à leur tour d'être naturalisés et traités comme des objets ayant une substance dépassant les variables de

base et leurs techniques de mesure. Et, pourtant, cette stabilité constitue un remarquable instrument de cumul et d'articulations de connaissances, dans la mesure où elle permet justement de rapprocher des comportements très variés.

Ainsi, ces axes peuvent avoir un extraordinaire pouvoir de mise en ordre et de confrontation d'observations, mais présentent aussi le danger de sembler constituer de nouvelles variables « explicatives » par elles-mêmes, d'autant plus efficaces qu'elles sont plus synthétiques. Ainsi l'opposition, maintenant bien acquise, entre catégories « à capital économique » et catégories « à capital culturel » (Bourdieu, 1979) n'a pas été dégagée immédiatement à partir des premiers travaux utilisant la nomenclature socioprofessionnelle fine, entre 1955 et 1975. Pourtant, étant inscrite dans la logique même de constitution de cette nomenclature (ou plutôt dans les « positions et trajectoires dans l'espace social » et les schémas mentaux des taxinomistes des années 1950, et plus généralement dans les institutions et représentations de cette époque), elle peut sembler à certains quasi tautologique, n'« expliquant » rien à elle toute seule, mais « mettant en perspective » remarquablement et permettant de reconstituer les cohérences locales de comportements des catégories par un ensemble de procédures sociologiques dont l'enquête statistique n'est qu'une, parmi d'autres.

La représentation spatiale issue de l'analyse factorielle n'est cependant pas tautologique si elle suggère que les topographies de positions sociales ainsi décrites reflètent l'existence de champs (Bourdieu, 1980) c'est-à-dire des ensembles à l'intérieur desquels les agents sont en relations ou en compétitions par rapport à des enjeux spécifiques, dont la définition constitue le champ lui-même⁶. Cette façon d'utiliser la notion de « champ » a montré sa fécondité dans l'étude de nombreux domaines, mais pour lesquels l'ensemble envisagé était souvent de petite taille et les intersections internes assez denses. Les variables étudiées par le démographe ou le statisticien relèvent-elles toujours de cette problématique des interactions à l'intérieur de champs ? Ainsi les régularités et les évolutions constatées de la fécondité, de la mortalité ou de la nuptialité peuvent sembler se situer dans des espaces si vastes que l'on n'y reconnaît pas aisément des logiques de champ. Mais, en dehors du fait que certains démographes utilisent spontanément des logiques d'interactions (Easterlin, 1966 ; Halbwachs, 1935), on peut se demander si le statisticien ne commence pas par *constater* certaines régularités bien avant que la science sociale ne parvienne à formuler les hypothèses décrivant les interactions qui sont à l'origine de ces régularités. À cet égard, la théorie des champs peut constituer un point de vue à partir duquel le statisticien et le sociologue tentent de donner sens aux tableaux de chiffres. Il faudrait pour cela que soit approfondie le problème de l'emboîtement, de la hiérarchisation des champs les uns dans les autres, les interactions étant de plus en plus denses quand on descend vers des champs de petite taille, et de plus en plus lâches, difficiles à percevoir et « statistiques » quand on envisage des champs très vastes, ceux justement qu'étudie en général le démographe. Ces champs ne sont d'ailleurs pas, à l'instar des catégories emboîtées d'une nomenclature, des zones munies de frontières, préexistant au système des interactions, mais ils sont définis par ce système lui-même. Il s'agit donc bien d'une *façon* de regarder le monde social, en décidant de s'intéresser aux diverses manières ou aux divers enjeux par l'intermédiaire desquels les agents entrent en relations ou en compétition mutuelles. Ainsi, par exemple, l'étude « économique » d'un « marché » est une façon d'isoler, parmi les interactions possibles des producteurs et des consommateurs d'une branche, celles qui sont marchandes alors que d'autres, passant par des institutions (syndicats, chambres de commerce...), des réglementations (conventions collectives, normes de sécurité...) ou même des rapports de domination beaucoup plus implicites (accès au crédit...), peuvent

6 L'interaction directe n'est pas nécessaire et la simple existence des autres définit l'espace dans lequel l'agent se définit et se meut.

être très importantes (Bony et Eymard-Duvernay, 1982 ; Eymard-Duvernay et Thévenot, 1982).

Ainsi, le statisticien décrit en général le résultat des multiples systèmes d'interactions : le sociologue a la tâche de les identifier.

Bibliographie

- AFFICHARD Joëlle, 1987, *Pour une histoire de la statistique. Tome 2*, Paris, Insee/Economica.
- BOLTANSKI Luc, 1982, *Les cadres : la formation d'un groupe social*, Paris, Éditions de Minuit.
- BONY Daniel & EYMARD-DUVERNAY François, 1982, « Cohérence de la branche et diversité des entreprises : étude d'un cas », *Économie et Statistique*, n° 144, p. 13-23
- BORKOWSKI Jean-Louis, 1981, « Enquête "Situations défavorisées" (1978). Qui a refusé et pourquoi ? », *Note de travail interne Insee*, n° 299/453 du 30/01/1981.
- BOURDIEU Pierre, 1979, *La distinction : critique sociale du jugement*, Paris, Éditions de Minuit.
- BOURDIEU Pierre, 1980, *Le sens pratique*, Paris, Éditions de Minuit.
- DE SINGLY François, 1982, « La gestion sociale des silences », *Consommation*, n° 4, p. 37-63.
- DESROSIÈRES Alain et GOLLAC Michel, 1982, « Trajectoires ouvrières, systèmes d'emplois et comportements sociaux », *Économie et Statistique*, n° 147, p. 43-66.
- DESROSIÈRES Alain et THÉVENOT Laurent, 1979, « Les mots et les chiffres : les nomenclatures socio-professionnelles », *Économie et Statistique*, n° 110, p. 49-65.
- EASTERLIN Richard, 1966, « On the relation of economic factors to recent and projected fertility changes », *Demography*, vol. 3, n°1, p. 131-163
- EYMARD-DUVERNAY François et THÉVENOT Laurent, 1982, « L'économiste et son modèle », dans *Sociologie et statistique. Tome 2*, Paris, INSEE/ Société française de sociologie, p. 1-35
- HALBWACHS Maurice, 1935, « La nuptialité en France pendant et depuis la guerre », *Annales sociologiques*, Série E, fascicule I, p. 1-45 [reproduit dans Halbwachs Maurice, 1972, *Classes sociales et morphologie*, Paris, Éditions de Minuit].
- HOGGART Richard, 1967, *La culture du pauvre*, Paris, Éditions de Minuit.
- INSEE, 1977, *Pour une histoire de la statistique. Tome I, contributions*, Paris, Insee.
- INSEE, 1987, *Pour une histoire de la statistique, Tome II, matériaux*, édité par Joëlle Affichard, Paris, Economica
- LEBART Ludovic, 1982, « L'analyse statistique des réponses libres dans les enquêtes socio-économique », *Consommation*, n°1, p. 39-62.
- PASSERON Jean-Claude et DE SINGLY François, 1981, « Différences dans la différence : segmentation, profils et rythmes de la socialisation », *Documents du GIDES*, n° 2 [repris sous forme d'article : PASSERON Jean-Claude et DE SINGLY François, 1984, « Différences dans la différence : segmentation, profils et rythmes de la socialisation », *Revue française de science politique*, vol. 34, n°1, p. 48-78].
- THÉVENOT Laurent, 1981, « Les catégories socio-professionnelles et leur repérage dans les enquêtes », *Études méthodologiques. Archives et documents*, n° 38.
- VOLLE Michel, 1974, « Une méthode pour lire et commenter automatiquement des grands tableaux statistiques », *Économie et Statistique*, n° 52, p. 46-50.