



**Maurice Comte \***

## **Normalisation statistique et plasticité sociale**

Les statistiques sont le produit final d'un processus de production complexe. Cet article sera l'occasion d'en donner une vue synthétique, bien que sa perspective soit à la fois limitée et concrète.

Son point de départ est concret : c'est la constatation que les demandes en matière de statistiques se traduisent à la fois par une extension des domaines couverts, et par une diminution des tolérances en matière d'erreur. La contradiction entre ces exigences affecte différemment les divers stades de la production des statistiques et, finalement, la validité des données produites.

Sa démarche est abstraite et partielle : en effet, la réponse à la question posée est cherchée dans une analyse qui assigne à la statistique un objectif, supposé reconnu et accepté, de connaissance, alors que son rôle social excède largement cet aspect (Besson JL, Journet O, 1986 ; Besson JL, Comte M, 1986).

Bien que cette façon de voir ait un caractère "idéel", elle permet de formuler clairement les contradictions internes à la méthode lorsqu'elle est confrontée à une réalité qui ne respecte pas certaines hypothèses. Comme cette situation est fréquente, c'est la question de la pertinence de la méthode statistique elle-même qui est posée : l'affinement des techniques toujours possible, n'est-il pas générateur d'un supplément d'illusion ?

---

### **1. Adaptation croissante ou inadaptation croissante ?**

Les statisticiens, pour répondre aux questions de plus en plus précises et ardues qu'on leur pose développent leur appareil d'observation. Comme l'ont bien constaté les économètres qui tentaient de réduire les écarts entre leur modèle et la réalité en perfectionnant, i.e. en complexifiant leur modèle, au-delà d'un certain seuil de désajustement, une telle démarche a un effet pervers<sup>1</sup>. En effet, elle repose sur un postulat d'insuffisance, de manque de fiabilité de la mesure (du modèle) ; ce faisant elle ne s'interroge pas sur la validité de cette mesure, c'est-à-dire sur sa capacité (théorique et non plus technique), à fournir une image adéquate de la réalité. Cette "adéquation" est difficile à définir (cf. la notion de "convenance" dans la contribution de JL Besson), néanmoins, on peut observer de multiples symptômes de décalages

---

\* Maître de Conférences à l'Université Lumière Lyon 2. Cet article a été publié dans Abdelmalki L., Besson J.L. (ed.), 1989, *L'observé statistique*, Presses Universitaires de Lyon. pp 53-110.

<sup>1</sup> Artus (1988) exprime clairement le retour en arrière actuel, qui fait suite aux très nombreux travaux visant à établir des maquettes simplifiées ou à analyser les propriétés des gros modèles. Cf. les nombreux travaux de Muet, Bureau, Norotte etc.

qualitatifs entre les indicateurs habituellement produits et les besoins exprimés les "utilisateurs" <sup>2</sup>.

## **1.1. Un contexte nouveau**

---

Qu'il s'agisse de définir des populations cibles ou de tester des hypothèses de politique économique, les statistiques qui intéressent les entreprises ou les administrations, même si elles décrivent des faits, touchent aux comportements : les "restructurations" industrielles, le volume de l'épargne, la demande de produits nouveaux, résultent d'actes dont la signification ne peut être établie sans référence à des motivations. A l'heure actuelle, les liens entre les actes et les motivations, qui étaient assez étroits se distendent.

### **a Le comportement des individus**

#### *– De l'opportunisme*

Dans la société de classes moyennes, la plupart des actes de la vie quotidienne ressortent à un univers paradoxal : ils échappent à une codification très précise, tout en offrant une faible étendue de différenciation. Du point de vue des comportements, cela signifie qu'aucun système statutaire plus ou moins stable ne peut exister, parce qu'aucune structure n'est plus considérée comme contraignante (affaiblissement des normes) ou satisfaisante (insuffisance des différences). Les positionnements sociaux sont donc perçus comme l'enjeu de tactiques sans cesse renouvelées.

On peut qualifier celles-ci d'opportunistes, puisqu'elles poursuivent un but sans considération des moyens. Si elles s'ancrent dans le jeu social, elles ne s'y limitent pas : on connaît les jongleries pratiquées avec les définitions des agrégats monétaires, qui tiennent à l'impossibilité de les faire correspondre à des attitudes économiquement significatives, lorsque les produits financiers sont détournés de leurs usages "normaux" par les agents (par exemple, usage à court terme de fonds "placés" et inversement).

#### *– Des comportements aux attitudes.*

Des adaptations dans les utilisations et la production des statistiques semblent donc possibles, mais au risque de méconnaître que c'est la structure du champ des statistiques qui est transformée : alors que la mesure pouvait reposer sur des actes, elle doit désormais saisir des attitudes.

L'observation des comportements était associée à une méthode d'analyse des "tendances" (ex : évolution de la demande future en fonction de la demande passée). Ce mode de raisonnement et son hypothèse sous-jacente de stabilité, sinon des comportements, tout au moins des trajectoires, est caduc. Les actes n'ont plus de sens par eux-mêmes, puisqu'ils ne caractérisent plus un groupe (ou un individu), mais marquent transitoirement sa place. S'il est vrai que la structure de différenciation reste, elle, très stable, les formes qu'elle revêt ont une très grande importance pour la conduite des politiques publiques ou pour la gestion des entreprises<sup>3</sup>.

---

<sup>2</sup> Compte tenu du choix retenu des statistiques "de base", les utilisateurs peuvent être eux-mêmes des producteurs de statistiques plus élaborées ou synthétiques.

<sup>3</sup>Le fonctionnement du système éducatif français illustre parfaitement comment au travers de tentatives de brouillage positionnel (rôle du latin ou des mathématiques, création du collège unique etc.), les effets sélectifs se sont peu modifiés. Il n'empêche que du point de vue de la gestion du système, de la compréhension de son mode de fonctionnement etc., les conditions d'observation et d'analyse ont dû être profondément modifiées pour tenir compte de ces changements de "forme".

## **b Les entreprises et l'état.**

Cette mobilité nouvelle est encore plus évidente si l'on s'intéresse au mouvement des affaires : les mutations de la demande ont induit des changements organisationnels, mais le renouvellement de la concurrence dans un marché mondialisé a singulièrement complexifié l'élaboration des stratégies, en conférant une place privilégiée aux capacités d'innovation technique et sociale.

La tendance mondiale à la "déréglementation" a joué dans le même sens, en contraignant les entreprises à s'adapter à la fois à de nouvelles libertés et à de nouvelles règles du jeu. En effet, la déréglementation et la réglementation, loin de s'opposer, se nourrissent mutuellement. La déréglementation est la suppression de barrières anciennes, qui ouvrent de nouveaux domaines à l'action des agents. Ce faisant, ce sont aussi des espaces de réglementation qui sont inaugurés. Pour prendre une image, la déréglementation, c'est l'enlèvement d'une barrière sur une route, à laquelle se raccordent un très grand nombre de chemins. Parmi ceux-ci, il en est inévitablement qui sont dangereux et dont il faut réglementer l'accès. Les exemples des marchés financiers, des offres publiques d'achat etc. illustrent parfaitement cette synergie, entre l'extension du champ de la liberté d'action des agents économiques et la multiplication des règles visant à en limiter les effets néfastes.

Face à ces changements qui touchent aux structures, aux règles du jeu, les techniques de prévision routinière sont peu efficaces, d'où l'intérêt d'investigations statistiques spécifiques. Mais, là aussi, il faut analyser des attitudes, alors que l'on saisit plus facilement et avec de meilleures garanties "techniques", les comportements, et a fortiori les actes (dans ce dernier cas, la statistique se rapproche du dénombrement). Au contraire, pour étudier des attitudes, il ne suffit plus de définir un protocole de mesure, il faut reconsidérer la nature même de l'opération de mesure. La problématique de la "fiabilité" au sens technique étroit s'efface derrière celle de la fiabilité au sens étymologique (ce à quoi on peut se fier), que nous nommerons la validité. En bref, à la question "la mesure est-elle exacte".

### **1.2. Quels changements ?**

Face à ces nouveaux besoins, l'appareil de production des statistiques est-il adapté et adaptable ?

#### **a Un passé**

##### *- De nombreuses adaptations*

Si l'on observe le passé, la capacité d'adaptation de la production des statistiques semble non négligeable.

Le champ des investigations, qui se limitait pour l'essentiel aux variables démographiques et aux quantités, s'est considérablement étendu : statistiques sociales, valeurs, comblement partiel de l'énorme écart entre les statistiques globales et les dénombrements locaux etc.

Le processus de production a été assoupli et transformé. L'utilisation massive de la méthode des sondages a fortement contribué à cette évolution.

- Au niveau des structures de production : technique d'investigation légère, elle est accessible à des réseaux échappant à la centralité étatique (universités, instituts privés)

- Du point de vue organisationnel : la mise en oeuvre est confiée à des enquêteurs professionnels, alors que les "comptages" ou le recensement sont effectués par une main-d'oeuvre non spécialisée et peu qualifiée.

- Les tâches de conception prennent une importance plus grande, car les sondages ont permis d'étendre le champ d'investigation vers des constructions conceptuelles plus ambitieuses (pratiques sociales, prix, volumes, valeurs ajoutées etc.).

## *Normalisation statistique et plasticité sociale*

L'évolution des techniques a aussi largement contribué à ces changements : l'aspect purement matériel du rassemblement des faits nombreux a perdu de l'importance avec la mécanographie, puis l'informatique.

### *– Les limites*

Cette nouvelle organisation du travail, fondée sur une meilleure compréhension des processus d'information, a exigé une intense standardisation des procédures. La poursuite du processus de standardisation peut être comprise de deux manières contradictoires.

- La complexité et à la variabilité des situations à analyser supposent un affinement des instruments. Il suffit de voir les problèmes posés par les statistiques du commerce extérieur, de la production et du chômage au niveau macroéconomique, ou les décalages croissants entre les indicateurs traditionnels de gestion des entreprises et une réalité perturbée par les innovations financières ou technologiques.

- La course à la normalisation, entre dans une zone où l'on peut présumer des rendements décroissants, voire négatifs. En effet, la perfection formelle des instruments tend à occulter,

- la nature des indicateurs : à force de détailler la description, on oublie qu'elle fournit une des images possibles de la réalité.

- la forme spécifique de cette représentation, qui pour des motifs internes, est à la fois ordonnée et rigide.

Ces deux effets importent peu lorsque les phénomènes observés ont des structures élaborées, sont à maturité. Par contre, face à des changements, surtout s'ils sont rapides et plus ou moins erratiques, ils masquent l'infidélité de la représentation. Le processus de production continue alors à fournir des produits finis qui fonctionnent mais qui ne sont pas nécessairement fonctionnels : ce serait comme une usine qui fournirait exclusivement des automobiles standards ; l'habitant d'un hameau enneigé n'aurait que la médiocre satisfaction d'écouter le bruit du moteur dans son garage<sup>4</sup>.

Confrontée à des exigences nouvelles, instantanéité, variabilité, détail, la production des statistiques a répondu dans sa logique d'extension : l'ajout d'éléments analytiques, et la diffusion beaucoup plus large que par le passé de l'ensemble des éléments disponibles, y compris les statistiques "primaires". Le progrès technique aidant, on a vu apparaître des bases de données spécialisées, puis synthétiques, qui semblent affranchir l'utilisateur des cadres préétablis.

## **b Bases de données, mythe et réalité**

### *– L'apport des bases de données*

Une base de données s'apparente à l'archivage par le souci de rassembler l'essentiel de l'information sur un sujet déterminé et s'en distingue par son mode d'accès. La souplesse du langage d'interrogation permet d'échapper à une procédure hiérarchisée. Les relations horizontales sont tout aussi accessibles que les relations

---

<sup>4</sup>. Exemple. Qu'il s'agisse d'orienter les stratégies des groupes ou l'intervention de l'Etat, les statistiques industrielles traditionnelles reposant sur les notions pourtant bien établies de secteur ou de branche, fondées sur l'activité principale, se sont avérées peu adaptées à l'observation de "filières", qui nécessitaient l'introduction de nouvelles variables (technologiques, organisationnelles et commerciales). Si dans ce cas, on pouvait envisager un nouveau découpage (s'ajoutant, plus que se substituant au précédent), il n'en est pas de même lorsque les entreprises, dans le contexte de mobilité concurrentielle évoqué ci-dessus, cherchent systématiquement des créneaux porteurs, ou équilibrants. Les diversifications, parfois provisoires d'activité, les recentrages, reposent sur des "logiques" sous-jacentes variables, qu'il faut identifier, faute de quoi de nombreux faits ne seront pas observés (le cadre ne s'y prête pas) et, même s'ils le sont, ne pourront être analysés.

## *Normalisation statistique et plasticité sociale*

verticales. Cela se traduit par la possibilité de réaliser n'importe quel croisement : on peut étudier simultanément le cours d'une matière première et le prix des produits finis l'incorporant et tester instantanément leur liaison empirique. Dans les systèmes traditionnels, faute de suivre à la trace la matière première au sein du système d'information lui-même, il faut procéder à une réflexion analytique préalable et consulter des sources différentes, souvent difficilement accessibles. La base de données met à disposition une masse d'informations beaucoup plus considérable, au moment même où les techniques d'archivage subissent une révolution<sup>5</sup>. Un nouveau mode d'investigation est ainsi généré, la "navigation dans l'information". L'analyse de ses conditions montre que ce terme séduisant recouvre une liberté largement illusoire.

### **- Vers la bibliothèque de Borgès ?**

L'élément d'attraction essentiel d'une "base de données" réside dans sa capacité à rassembler un très important volume d'informations et à autoriser son accessibilité (rapidité, facilité)<sup>6</sup>. A la limite, elle contiendrait toute l'information disponible.

La bibliothèque de Babel de Borgès est une métaphore particulièrement intéressante pour la base de données. Elle contient tous les "mots" constitués par assemblage des lettres de l'alphabet, et toutes les combinaisons de "mots" qu'il est possible de réaliser. La question des bibliothécaires est alors celle-ci : en ouvrant l'un des innombrables livres, on trouve parfois un mot ayant un sens, exceptionnellement une phrase, mais existe-t-il un livre complet ayant une signification, et, comment le trouver ? Cette vision onirique (cauchemardesque) est révélatrice : les données "élémentaires" sont à l'image, disons des mots, ils ne prennent sens que dans leur association avec d'autres. Si les mots sont très nombreux, la navigation aléatoire a fort peu de chances de conduire à une phrase porteuse de sens. On comprend alors que l'élément fondamental de la base de données est la création des liens, qui servent de guide.

Même dans ce cas, la constitution d'une base encyclopédique est largement illusoire et finalement peu rationnelle. On retrouve ici la critique, à la fois technologique et sociologique faite par B. Lussato : le renforcement de la centralité, implique le développement d'un réseau lourd (et coûteux) de communication (télématique) et un traitement des demandes très sophistiqué et éloigné de l'utilisateur, là où on aurait pu concevoir un système décentralisé plus simple et plus efficace. Par exemple, il existe un stockage centralisé de données concernant individuellement chaque commune. Quel est l'usage d'un tel système ? En gros, il y a deux types d'utilisateurs : d'une part l'administration locale qui interroge de façon très détaillée sur un champ restreint, sa commune, éventuellement quelques autres voisines ; d'autre part, divers organismes centraux, qui cherchent un traitement systématique, beaucoup moins détaillé, mais portant sur l'ensemble<sup>7</sup>.

A l'évidence, il existe une forte probabilité de mauvaise utilisation du système : il est beaucoup trop riche pour les uns, beaucoup trop vaste pour les autres. La solution de sagesse serait d'avoir deux systèmes. L'un serait décentralisé, très

---

<sup>5</sup>La concomitance n'est pas fortuite : le potentiel de stockage des techniques modernes d'archivage (microfilms, mémoires magnétiques ou autres etc.) est tel qu'il est inutilisable par le biais des procédures classiques de sélection. Le développement des techniques d'investigation (langages, intelligence artificielle) est nécessaire pour rétablir un certain équilibre.

<sup>6</sup>C'est le critère privilégié de choix d'une base de données (cf. Chaumier 1986). Il correspond à la méthode associée à l'usage de ces bases, qui tend à privilégier l'exhaustivité.

<sup>7</sup>Un responsable d'ORACLE (Organisation Rhône-Alpes des collectivités locales) exprime clairement ce point de vue (in Sciences et Vie Micro n° spécial, "Les trésors du minitel", 1988) : "Les communes ont des besoins d'information ponctuels, pointus et urgents, or les banques de données existantes sont trop générales et trop compliquées. D'après nos tests comparatifs, le taux de satisfaction est actuellement plus élevé par les moyens papiers traditionnels."

## *Normalisation statistique et plasticité sociale*

détaillé, directement accessible, capable d'ajouter au langage standard d'interrogation ses propres notions ou concepts. L'autre, centralisé, reposant sur des procédures normalisées, mais au champ plus limité, composé d'indicateurs "synthétiques".

On peut cependant remarquer que l'organisation interne de la base de données peut pallier l'inconvénient de mal commodité (en laissant pour une large part subsister celui du coût social) grâce à la mise à disposition des utilisateurs spécialisés de clefs d'accès direct et "sur mesure", court-circuitant la voie hiérarchique.

Plus généralement, la prolifération des données accessibles ne devrait pas nécessairement aggraver le recours aux méthodes empiriques : le traitement statistique ou la modélisation sont plus commodes à réaliser, mais l'ampleur de la combinatoire autorisée par le corpus est telle qu'on utilise nécessairement des grilles théoriques pour l'investigation. En tout état de cause, on peut imaginer une phase de découverte, qui se terminera assez rapidement pour laisser place à des méthodes plus rationnelles. Comme le note bien Morgenstern (1950), lorsque la lunette astronomique a été inventée, on pouvait la braquer sur n'importe quelle portion du ciel et faire une découverte ; désormais, il faut une longue analyse afin de savoir dans quelle direction l'orienter pour en tirer des enseignements nouveaux.

### **c Une cage dorée ?**

C'est plutôt une autre question qu'il convient de discuter. La tendance à la centralisation est inscrite dans l'existence même du système de bases de données (statistiques ou non) et l'offreur est incité à s'y associer. Or, quel est le sens de ce rassemblement ?

#### *- Une expression trompeuse*

Le terme "base de données" contribue à dissimuler la vraie nature des informations que nous utilisons :

- le mot base suggère qu'il s'agit là d'un fondement, d'un point de départ.
- quant aux "données" (data), leur sens correct dans ce cas - "ce qu'on se donne"-, est occulté par : "ce qui se donne". Dans ce glissement, le produit disparaît derrière le résultat, le "fait", que l'investigation n'aurait qu'à saisir.

Le sens de "fait établi" des deux mots est renforcé par leur redondance ; de ce fait, l'expression définit une certaine place, super structurelle, à l'utilisateur, astreint à faire reposer son propre travail sur ce fondement, par exemple en y cherchant des preuves.

Or, il s'en faut de beaucoup que les "bases de données" soient dignes d'un tel statut, ne serait-ce qu'en raison du manque d'homogénéité des "données", qui devrait pourtant être leur qualité essentielle : à quoi bon rassembler des éléments en un même "lieu" si, du fait de leur hétérogénéité réelle et formelle, on ne peut pas les rapprocher dans le processus de connaissance<sup>8</sup> ? L'existence d'une centralisation est déjà en soi un postulat d'existence d'unité de l'information. Reste à savoir quelle est la valeur de ce postulat.

#### *- La tentation de l'empirisme*

Le coût de rassemblement des informations est négligeable par rapport à celui de leur production. Dans ces conditions, en dehors des bases de données spécialisées propres à un producteur, les bases synthétiques rassemblent des informations

---

<sup>8</sup>Des différences aussi banales que des différences de dates ne peuvent pas toujours être réduites à la source ou "corrigées". On connaît les multiples sources de divergences (champ, mode de recueil, délais etc.) qui existent entre des séries établissant le même indicateur. L'évidence même de ces divergences incite à la prudence ; il n'en est pas de même quand on associe des indicateurs différents. Pourtant, dans ce cas, des problèmes d'homogénéité similaires existent.

d'origines très diverses<sup>9</sup>. On peut imaginer la coexistence de statistiques établies à partir de l'annuaire téléphonique, des inscriptions électorales et du recensement : les complémentarités entre sources sont tentantes. Dans cet exemple, le rapprochement devrait logiquement avorter du fait des écarts quantitatifs existants, mais il existe des procédures de "redressement". A l'aide du recensement, on spécifie les non-inscrits et on reconstitue les fichiers incomplets. Les critères retenus, sont évidemment limités aux informations disponibles (âge, sexe, catégorie de commune) et l'opération repose sur le postulat habituel des procédures de redressement d'échantillon : la personne et son substitut statistique sont considérés comme identiques.

Derrière le bricolage ainsi effectué et sa technicité (plus ou moins grande selon les cas, mais se ramenant en fait à des règles de trois), on aura occulté les vraies différences entre sources : mode de recueil de l'information, champ couvert etc.

– *La condition de plausibilité, la normalisation*

C'est dire que l'homogénéité doit être produite, grâce à l'édition de normes ou standards, qui assurent une communauté de contenu : il s'agit essentiellement d'user de définitions (indicateurs, ratios, nomenclatures etc.), et de procédures de collecte similaires.

Les effets de cette homogénéisation sont tantôt réels, tantôt illusoires<sup>10</sup>, mais ils ont un point commun, la réduction du sens de l'information. La normalisation rapproche effectivement, mais c'est au prix d'un appauvrissement du langage (par extension de la synonymie), qui se répercute sur la production d'information, car la nomenclature est la syntaxe du système informationnel. La prédominance du discours centralisé (parce que centralisable), s'appuie sur sa rigueur formelle (le code), et ses usages pratiques (les comparaisons, ou croisements).

Un exemple particulièrement éclairant est celui du code des PCS (ex CSP)<sup>11</sup> propre à la sociologie française : bien qu'il ne soit assorti d'aucune obligation d'usage, sa forte présence dans les statistiques centrales en fait un point de passage quasiment obligé pour toute autre forme d'investigation, ne serait-ce que pour effectuer des contrôles. Il ne s'agit pas de nier son intérêt, mais de souligner qu'il n'est pas forcément le moyen efficace de rendre compte de certains phénomènes : son caractère synthétique tend à masquer que s'il fournit toujours de l'information, rien ne garantit qu'un de ses composants ne pourrait pas, à lui seul, en fournir davantage et, surtout, être plus explicatif.

On pourrait croire qu'il s'agit là d'une exemple particulier, propre aux statistiques sociales, mais les mêmes remarques peuvent être faites sur le cadre comptable actuellement utilisé par les entreprises, dont les "*concepts sont mal adaptés, en toute rigueur, à l'élaboration d'un système d'information à vocation générale*" et "*explique pour une part les difficultés que l'on rencontre dans la recherche d'une meilleure intégration entre les systèmes d'informations micro et macro-économiques.*" (A Benedetti, in PUH t.2, p 389).

– *Modèles et tautologie*

Les conséquences de la normalisation apparaissent très clairement si l'on prend l'exemple d'usage "extrême" des statistiques que font les modèles. La base de données est un préalable à la création de modèles. Ceux-ci fournissent une analyse et une quantification des mécanismes économiques et sociaux grâce à une formalisation

---

<sup>9</sup> Aux Etats-Unis, le marché des bases de données témoigne d'un dynamisme exceptionnel. Les grandes bases offrent des millions de séries statistiques. Compte Tenu des difficultés de l'homogénéisation, celle-ci ne concerne, au mieux, que des séries ayant la même source.

<sup>10</sup> cf. la comparaison des statistiques du chômage françaises et japonaises JL Besson 1988.

<sup>11</sup> Respectivement Professions et Catégories Socioprofessionnelles et Catégories SocioProfessionnelles. Pour une présentation récente cf. Desrosières A, Thévenot L 1988.

## *Normalisation statistique et plasticité sociale*

très poussée des "faits" contenus dans la base de données. Le modèle idéal est celui qui permet d'approximer de la façon la plus satisfaisante à partir d'un nombre relativement restreint d'équations et de variables, l'ensemble des relations existant entre les variables disponibles. Il est clair que, dans une telle procédure, le retour aux "faits", par lequel on teste la validité du modèle, est un retour à la masse des "données".

Nous avons vu que, quelle que soit sa richesse, cet univers est préalablement formalisé, quadrillé : il s'agit d'une image du réel qui, propulsée au premier plan par son statut, éclipse toutes les autres. Or, il arrive très fréquemment que, pour les besoins du modèle, on produise des données spécifiques (cf. le rôle de la demande). Sans aller jusqu'à l'exemple le plus outrancier, celui de la variable "ad hoc"<sup>12</sup>, le risque de tautologie propre au modèle est d'autant plus grand qu'est perdu le caractère représentationnel très partiel des "données".

### **d Conclusion ; quel réexamen critique ?**

Plus que par le passé, les "besoins" de statistiques sont changeants et tendent à échapper aux cadres établis. Pour tenter de les satisfaire, se développent des systèmes encyclopédiques à vocation générale, qui ajoutent à la tendance immanente à réifier les données indépendamment des usages qui en sont faits : la base de données se présente comme un gisement dans lequel chacun extrait de quoi faire ses assemblages analytiques. L'examen auquel il a été procédé ci-dessus montre le caractère erroné de cette représentation.

#### *– En l'état actuel, la combinatoire n'est pas possible*

Pour l'instant, les bases de données n'assurent qu'une simple "proximité" des renseignements, qui incite à les combiner, alors que leur homogénéité est faible, voire nulle. Cette tendance à la combinatoire est d'autant plus marquée que les producteurs sollicités fournissent désormais des "données élémentaires", et non plus seulement des synthèses. Or, celles-ci, réalisées par les producteurs eux-mêmes, tenaient compte des difficultés méthodologiques et des problèmes d'élaboration rencontrés<sup>13</sup> et en informaient les utilisateurs.

Avec l'accès aux sources primaires, ces considérations annexes, qui figuraient déjà trop souvent en note, ou dans le seul descriptif intégral du protocole, ne seront pas nécessairement éliminées, mais il est plus que probable qu'elles seront noyées par l'avalanche des chiffres engloutissables instantanément par des machines, lesquelles resteront, longtemps encore, incapables de comprendre les problèmes de production des statistiques.

#### *– Normalisation et perte de sens*

Au-delà, c'est à une gigantesque combinatoire que l'utilisateur est convié, sans égard à la spécificité éventuelle des combinaisons. C'est comme s'il était possible d'utiliser les mêmes "atomes" (des "êtres statistiques") pour créer une variété infinie de molécules différentes (des "indicateurs synthétiques"). Pourtant, il est évident que la définition d'un mot comme "concubin" est susceptible de grandes variations selon les principes qui guident l'observation (et a fortiori l'utilisation) : l'extension accordée au terme n'est pas la même si on s'intéresse à la démographie ou aux droits sociaux. La volonté d'homogénéiser, liée au caractère industriel de la production conduit à résoudre le problème de l'unité du sens en en imposant un.

En une image, la "base de données relationnelle", c'est une prison moderne, où des couloirs permettent de se déplacer librement du dortoir à la salle à manger ou au

---

<sup>12</sup> Dans ce cas, la base intègre des "données" qui font opportunément tourner le modèle.

<sup>13</sup> cf. par exemple les comparaisons entre les chiffres du chômage selon diverses sources, les multiples remarques sur les caractères spécifiques de telle ou telle enquête etc.



jardin, mais dont les décors sont immuables et la topographie prédéterminée. Les risques d'évasions sont faibles, tant parce qu'on ne peut imaginer un extérieur dans un aussi vaste univers, que parce que le nombre des promenades possibles (et aisées) est infini.

### *De nouveau la tautologie*

L'hypothèse ainsi formulée peut engendrer le scepticisme : malgré la centralisation, on dispose de plusieurs sources concurrentes, dont on peut comparer les qualités, et, surtout, la "réalité" ne se charge-t-elle pas de faire justice des déviations de la connaissance ?

On retrouve ici le problème de la "preuve", longuement débattu au cours des dernières années<sup>14</sup>. On sait que, pour l'essentiel, il est particulièrement difficile d'échapper à la représentation de la connaissance inspirée du rationalisme classique : la connaissance serait une construction progressive, la substitution à des paradigmes dépassés d'autres plus "puissants" (i.e. plus généraux).

En matière socio-économique, ce schéma ne convient pas. La procédure de contrôle "normale" est la confrontation de diverses représentations, mais elle conduit rarement à des éliminations ou des absorptions : l'issue est plutôt l'établissement d'une hiérarchie de fait. Les paradigmes coexistent, parce qu'ils représentent des aspects différents de la réalité ; si certains sont plus usités c'est parce que, quoique partiels, ils semblent mieux correspondre à la situation du moment (Kuhn 1962).

Les analyses en termes de "réfutabilité" ont eu le mérite de chercher un fondement à cette coexistence. Malheureusement, elles omettent le fait que l'observateur est, directement ou non, un intervenant direct sur la réalité. Un paradigme dominant tend à créer les moyens de la reproduction de sa dominance, dans la mesure où il crée des effets de réalité, aussi bien par sa structure, que par son langage, ou plus directement par les actions qu'il inspire. J'ai évoqué ci-dessus, les variables "ad hoc" ; on peut se demander dans quelle mesure il ne finit pas par y avoir des réalités "ad hoc", i.e. se conformant à la rationalisation logique qui en a été faite. L'exemple le plus frappant est celui du mot "cadres", qui a fini par se matérialiser en un véritable groupe social (Boltanski 1982), mais on peut interpréter de façon similaire<sup>15</sup> les cheminements sinueux de la notion de chômeur.

Sans aller jusqu'à ce stade extrême, la vérification par confrontation organisée à la réalité a un caractère tautologique. Elle s'effectue dans le référentiel des statistiques, et ne peut donc porter sur lui. C'est dire que toute approche de la "qualité" des statistiques doit emprunter d'autres voies, revêtir un caractère de méta-critique, traitant comme objet de connaissance le processus de production des statistiques et non pas ses produits.

---

## **2. Reconsidérer les conditions de production des statistiques**

Plusieurs approches sont possibles et ont été explorées au cours des dernières années. Si chacune a mis l'accent sur un aspect particulier de la production des statistiques, ce qui permet d'en donner une présentation heuristique sans les caricaturer outre mesure, elles ont un point commun, le refus du postulat de

---

<sup>14</sup> Pour un exposé général : Mingat A & alii, 1985.

<sup>15</sup> Outre l'article figurant dans cet ouvrage, on pourra consulter Besson JL, Comte M, Rousset P 1986. La similarité dont il est question est méthodologique, elle ne concerne ni le résultat (une certaine autonomie de la notion de chômage/vs la création d'un groupe social), ni le processus (une mutation consensuelle vs une véritable "stratégie"). L'exemple du chômage a une valeur beaucoup plus générale que celui des cadres, dans la mesure où il n'est porté par aucune force sociale clairement identifiable.

neutralité de l'observateur. C'est là une ouverture vers l'étude de la nature relationnelle (et donc fréquemment conflictuelle) de l'élaboration des statistiques.

Les statistiques sont une information<sup>16</sup> quantitative, obtenue grâce à un protocole d'observation. En conséquence, leur dispositif de quête tend à imposer un certain contenu, des formes d'existence (rationalisation, ordonnancement, logique formelle) et un mode d'interrogation, qui suscitent des réactions, des interactions. L'analyse consiste à mettre en évidence les lieux où elles se manifestent, et leurs particularités. Non seulement elle ne peut négliger aucune des différentes "phases" de la production statistique, mais elle doit restituer l'unité du processus. En effet, la description de la production en phases successives (demande, conception, recueil, mise en forme) est utile, mais elle fait perdre de vue l'étroite interdépendance qui caractérise l'ensemble. En particulier, en différenciant les niveaux, on occulte les interfaces qui les relient, où l'information est traduite, réinterprétée, avec tous les risques de distorsion que cela implique.

Dans cette perspective, trois points de contact paraissent essentiels.

- Les conditions de capture de l'information. Les statistiques concernent des phénomènes nombreux, dont la connaissance est dispersée entre de multiples informateurs. La relation de ceux-ci au processus de collecte est décisive pour le contenu et le sens des statistiques.

- Le travail d'élaboration. La collecte n'est pas effectuée au hasard : l'exploitation et l'utilisation impliquent des choix concernant ce qui est observé et comment. Dans ces conditions, le codage est l'occasion d'une confrontation permanente entre le référentiel construit (les nomenclatures) et le matériel "brut" issu de la collecte. La mise en forme de l'information est donc une phase critique où les principes mêmes d'organisation de la quête sont confrontés au "réel".

- La conception. Comment est construit ce "regard" particulier que manifestent le vocabulaire, les indicateurs, les nomenclatures, voire les méthodes. Les contingences propres à la quantification, semblent alors moins importantes que le contexte institutionnel dans lequel se formalisent les objets et les moyens de l'investigation.

On voit que les causes (principales) des déformations subies par l'information statistique diffèrent selon le stade de la production : elles sont relationnelles (sujet statistique), organisationnelles (codage) ou conceptuelles et institutionnelles (définitions).

## **2.1. Le travailleur involontaire**

On néglige souvent le fait essentiel que la source de l'information, au moins en matière socio-économique, est un participant plus ou moins volontaire au procès d'information. Dans tous les cas, des personnes fournissent, qu'elles en aient conscience ou non, des informations. Les exemples les plus divers pourraient être pris, du remplissage d'un formulaire administratif, à l'exécution de certains actes (paiement bancaire, passation de commande) ou, bien sûr, à la réponse à un questionnaire. Ces contextes divers peuvent faire l'objet d'une étude informationnelle, mais je ne retiendrai ici que le dernier :

- d'une part, il est celui qui se rapproche le plus des conditions de neutralité propres à l'observation : le remplissage d'un formulaire correspond à un but, ou à une contrainte, ce qui nuit à l'objectivité (c'est l'exemple caricatural de la déclaration d'impôts).

---

<sup>16</sup> Le terme n'est pas défini ici au sens de la théorie de l'information traditionnelle, pour laquelle l'information se réduit à des signes et les problèmes de transmission à des pertes ou à des ajouts de signes (bruits) ; cf. à l'opposé la conception purement relationnelle de la transmission de l'information (Watzlawick 1972).

## *Normalisation statistique et plasticité sociale*

- d'autre part, contrairement aux observations ou questionnements de type "administratif", fortement finalisés par des perspectives gestionnaires, l'enquête est une démarche de connaissance : elle s'efforce de saisir l'essentiel du comportement ou des actes<sup>17</sup>.

Le statut particulier de la personne interrogée est présent d'une certaine manière dans les études portant sur la méthodologie des enquêtes, mais il est formulé d'une façon partielle et partiale, comme source de biais éventuels, que diverses "astuces" permettent heureusement de tarir<sup>18</sup>. C'est une autre perspective qui est adoptée ici : tous ces biais sont inhérents à la méthode, et sont significatifs d'une relation, souvent conflictuelle, entre l'appareil d'observation et le "travailleur involontaire". Selon sa nature et son ampleur, le conflit apparaît d'une manière directe (dans des comportements), ou insidieuse, sous la forme d'un conditionnement de celui qui renseigne.

### **a Non-neutralité de la relation**

#### *- Diverses sources de biais*

Prenons l'exemple de la tendance à répondre oui plutôt que non, qui montre les limites de la participation à la production de l'information.

Elle est bien connue des hommes politiques qui l'utilisent, mais aussi des enquêteurs, qui cherchent à l'éviter. Elle s'explique par le contexte : la "neutralité" (entendue ici comme absence d'enjeu) de la situation d'enquête ne crée pas une motivation suffisante pour que le répondant mobilise l'énergie, importante, nécessaire à la relation d'opposition qu'est l'énonciation d'une négation.

D'autres biais, relevés dans les manuels consacrés aux enquêtes (Desabie 1965 ; Guiglione & Matalon 1976), sont justiciables de la même analyse : les réactions de prestance ou au contraire la tendance à se fondre dans la masse, et même l'effet de halo dû à la structure du questionnaire.

#### *- Une mise en cause plus profonde*

Tous ces comportements de réponse sont des signes de ce que la personne interrogée perçoit l'investigation comme un pouvoir s'exerçant ou susceptible de s'exercer sur elle. Mais, dans ce cas, la "présomption de bonne foi" qui sous-tend les enquêtes est une représentation pour le moins simpliste de la situation réelle. Deux postulats fort peu innocents lui servent de substrat : d'un côté, il existerait une seule vérité, de l'autre, l'individu pourrait être érigé en un juge objectif, susceptible de déterminer en toute indépendance cette vérité.

Que l'un ou l'autre de ces deux postulats, qui tendent à se renforcer mutuellement, soit faux, et l'hypothèse de neutralité du questionnement n'est plus admissible : au contraire, il faut considérer celui-ci comme un catalyseur, qui selon sa forme et les contextes, provoque des réactions<sup>19</sup>. Il est significatif que les erreurs (ou approximations) concernant une question de "fait" apparemment aussi incontestable que la taille, puisse être l'objet de variations selon les catégories sociales : le degré de connaissance de soi est lui-même une variable sociologique.<sup>20</sup>

---

<sup>17</sup> Par exemple, les banques effectuent un suivi des mouvements de compte de leurs clients ; quoiqu'il puisse être très précis et détaillé, il reste borné par le type de "données" disponibles, dont le champ est étroitement circonscrit (mouvements de fonds, soldes) et limité au fichier (rien n'empêche la personne d'avoir un autre compte).

<sup>18</sup> Cette interprétation "pragmatique" des erreurs est particulièrement nette dans l'ouvrage de J Desabie (1965) qui est resté le manuel de référence de l'INSEE en matière de sondage ; cf. aussi Chevy (1962).

<sup>19</sup> Garfinkel H 1967.

<sup>20</sup> Charraud A, Valdelièvre H 1981.

## **b** Violence du questionnaire

La nature de la relation créée conditionne fortement la circulation de l'information. Nous venons de voir comment cela se traduisait dans les comportements adoptés au médium de ce mode de collecte que sont les questions. Cette présentation a le mérite de fournir des illustrations simples, mais elle occulte un autre aspect : les questions sont moins importantes que la procédure d'ensemble ; disons que les questions sont les "mots", tandis que le questionnement est la "syntaxe".

### *– Le questionnement est partiel*

La quête vise à fournir des indications très précises sur les comportements des personnes : intentions de vote ou d'achat, décisions d'épargne ou de recherche d'emploi etc. La formalisation poussée des questionnaires est,

*"précise, limitée, planifiée, de telle sorte que si des indicateurs importants n'ont pas été répertoriés, le questionnaire ne permet pas d'en révéler l'existence." Chauchat 1982*

La solution pratique est en général la réalisation d'une pré enquête, voire d'entretiens accompagnant la présentation d'un pré-questionnaire : les personnes invitées à s'exprimer sur le questionnaire fournissent des appréciations, des avis, qui, faute d'avoir été prévus, apparaîtraient, mais de façon non systématique et inexploitable sous formes d'ajouts divers ou induiraient des comportements difficiles à interpréter (non-réponses, "ne sait pas" etc.)<sup>21</sup>.

Il est clair que, sauf à supposer que l'enquête soit complètement inutile <sup>22</sup>, il est impossible d'être certain de n'avoir omis aucun élément. Aussi, contrairement à une opinion très répandue, la pré enquête doit-elle moins avoir pour but de tester les questions (aspect technique secondaire), que la pertinence et l'exhaustivité de la conceptualisation, du questionnement et de la codification.

### *– Des mots non des actes*

Un questionnaire ne recueille que des réponses verbales<sup>23</sup>, d'où la différence possible avec les comportements. On connaît l'écart entre les opinions manifestées et leur traduction pratique (intentions d'achat, de votes).

Sans doute est-il possible, grâce à des formulations de questions appropriées de distinguer ce qui ressort de l'action de ou l'opinion. Mais il faut noter qu'on élimine alors celle-ci (ou l'attitude), qui est tout aussi importante, ne serait-ce que pour donner sens aux faits. Ainsi, si le choix de ne pas comptabiliser comme chômeurs<sup>24</sup> des personnes déclarant chercher un emploi, mais qui n'ont pas accompli un acte effectif de recherche au cours du mois qui précède est en général justifié, il existe de nombreux contextes où il ne l'est pas. L'utilisation de la statistique ainsi produite, pour quelque motif que ce soit, produira des interprétations ou décisions malencontreuses.

L'enquête n'est certes pas une procédure "démocratique". Sans être un plébiscite, elle est porteuse de modèles de comportements. On peut en induire que, si la

---

<sup>21</sup> Sur les non-réponses et autres "fuites" devant l'enquêteur, cf. Passeron JC (1982) et l'article de F De Singly (1982) au titre évocateur "La gestion sociale des silences" .

<sup>22</sup> Puisque l'on saurait ce qu'on cherche à apprendre.

<sup>23</sup> Le terme "réponse verbale" s'applique aussi aux réponses écrites. Il signifie que l'on demande de "verbaliser", ie d'exprimer sous forme linguistique. Comme la plupart des actes ne peuvent être observés directement (sorties, dépenses etc.), on demande à l'interviewé d'être lui-même l'observateur et de transcrire son propre regard.

<sup>24</sup> Cf. notre contribution sur le chômage dans ce même ouvrage et aussi Besson JL, Comte M, Rousset P 1981

recherche de la neutralité continue d'être souhaitable, elle prend une dimension nouvelle. En effet :

- d'une part la neutralité diffère selon les publics touchés et les thèmes étudiés ; il était admissible de négliger ces écart dans la mesure où on observait des actes<sup>25</sup> ; cela ne l'est plus si l'on s'intéresse aux attitudes. Sans doute l'enquête est-elle utilisée et efficace en psychologie, mais elle s'insère alors dans un schéma analytique adapté à la nature des variables, ce qui n'est pas le cas en économie (cf. la conclusion).

- d'autre part, il paraît clair qu'en cherchant à éviter les ambiguïtés, sur les mots, sur la position de l'informateur etc., on accroît le caractère normatif et inquisitorial du questionnaire (exemple : les multiples précisions sur la situation d'un chômeur).

Pour apprécier les conséquences de cette non-neutralité, il faut les mettre en relation avec les autres problèmes de l'investigation, et, en premier lieu, avec ceux rencontrés lors du codage.

## **2.2. La médiation du codage**

---

L'opération de codage consiste à classer les informations issues des questionnaires dans les nomenclatures qui ont été définies lors de la conception de l'enquête. Les nomenclatures sont une représentation organisée, dans laquelle le codage doit faire entrer la diversité extrême des images qui ont été recueillies. En bref, les nomenclatures définissent des cadres, une typologie, le codage étant l'opération d'affectation.

### **a Le rôle des codeurs**

Les nomenclatures sont la langue commune qui assure l'unité du processus de collecte. Or, elles sont soumises à des impératifs contradictoires : d'un côté, elles sont l'expression d'une certaine conception de la réalité, cohérente, précise et adaptée à des formalisations. De l'autre, elles s'ancrent dans le langage commun. En d'autres termes la nomenclature est une langue "intermédiaire", qu'on serait contraint d'inventer en raison des trop grandes différences existant entre deux univers. Certes, on peut appeler "catégorie 05" la strate de la nomenclature correspondant (principalement) aux "cadres supérieurs", distinguant ainsi clairement ce qui tient à l'usage commun et à la catégorie statistique. Les inconvénients sont évidents, car personne ne peut se situer dans l'univers de la nomenclature, et peu d'avantages : en fait, la distanciation deviendra rapidement illusoire, tant pour les codeurs (pour lesquels il y aura identification à 90%), que pour les analystes ou commentateurs, qui pour donner une certaine "réalité" à la catégorie en question lui feront désigner le phénomène observable correspondant, les cadres supérieurs<sup>26</sup>.

Cet exemple, poussant jusqu'à l'absurde la distinction entre les catégories communes et la représentation, a le mérite de mettre en évidence l'ambiguïté fondamentale du discours statistique, qui ne peut fonctionner que sur un mode analogique, alors qu'il crée son propre référentiel. Au stade où nous en sommes de l'étude de la production du chiffre, ceci a une implication très importante : la nomenclature est un point de contact et donc de conflit entre des niveaux de signification. Or, l'arbitre du conflit, le codeur, n'a pas une position "neutre" : il n'est pas un dictionnaire de nomenclatures, mais un traducteur, qui doit apprécier le sens du résultat. On pourrait imaginer une nomenclature qui détaillerait à tel point les modalités possibles de la variable qu'un ordinateur pourrait à lui seul établir le code.

---

<sup>25</sup> L'exemple des estimations de taille mentionné ci-dessus montre d'ailleurs les limites de l'hypothèse.

<sup>26</sup> Le cas des "professions intermédiaires" de la nouvelle PCS est très différent : il s'agissait dans ce cas d'éviter une absorption de la catégorie par une de ses composantes fortement connotée (cadre moyen), absorption rendue possible par le manque d'homogénéité apparent de la catégorie.

Ce serait faire bon marché de la principale difficulté rencontrée lors du codage, qui n'est pas l'excès de précision des intitulés, mais leur insuffisance. L'intitulé "secrétaire technique au service commercial" est bien plus aisé à coder (dans une nomenclature de professions ou de PCS) que "fonctionnaire" ou "agent de l'EDF", deux réponses fréquentes, qui traduisent la prégnance du statut sur la profession ou la position hiérarchique. C'est dire qu'une bonne nomenclature ne peut pas éviter de prévoir, et donc de connaître, les associations spontanées, les "abus" éventuels du langage courant, et qu'elle doit fournir des instruments logiques permettant de juger de son sens, toutes choses qui supposent que sa production s'appuie sur une compréhension de la place, technique et sociale, de ceux qui sont chargés de l'interprétation.

Ces règles sont d'autant plus importantes qu'elles s'insèrent dans la problématique plus générale de l'efficacité organisationnelle du processus de collecte. De ce point de vue, il faut distinguer au moins deux contextes : dans l'un la quête statistique est greffée sur d'autres tâches, de gestion ou d'administration, alors que dans l'autre la collecte est une tâche spécialisée.

### **b Les à-peu-près du codage de fait**

Dans le premier cas, le problème essentiel est que la plupart des tâches informationnelles, y compris celles qui se traduisent par une quantification, n'apparaissent pas en tant que telles, mais sont considérées comme le sous-produit de l'activité "normale". Par exemple, à l'occasion de la prise de commande, les employés du service commercial d'une entreprise recueillent de nombreux renseignements sur les clients. Nombre d'entre elles ne sont pas indispensables au traitement de la commande, qui n'exige que le nom, l'adresse et une quelconque référence financière ou bancaire. Dans ces conditions, les autres informations apparaissent comme des appendices, facilement jugés superflus et les employés les omettent ou sont négligents dans le remplissage. Certes, on peut leur imposer d'être exhaustifs, mais le risque est grand qu'ils cèdent au formalisme, difficile à détecter et à contrôler.

La seule issue possible est l'adoption de formes adaptées de motivation, périodiquement renouvelées. Un moyen efficace est la mise en évidence des résultats du travail effectué. Malheureusement, si cela est parfois simple (dans l'exemple ci-dessus : exactitude des prévisions, maintien à un niveau exceptionnellement bas des stocks etc.), bien souvent il n'existe pas de lien aussi direct entre collecte et résultats opérationnels et l'incitation ne pourra se faire que par des voies plus générales (formation aux objectifs et instruments de la quête de l'information)<sup>27</sup>.

### **c Le codage organisé**

La création d'un service (ou d'un institution) spécialisé dans la collecte donne à l'information un véritable statut, favorable à l'émergence d'une dynamique de motivation ; la question du codage peut alors faire l'objet d'un traitement spécifique.

#### *– Les limites de l'automatisation*

J'ai évoqué ci-dessus celles qui sont inhérentes à la nature des problèmes habituellement rencontrés, et celles qui tiennent au flou des déclarations.

D'autres tiennent à l'organisation du travail, tels les effets pervers des automatismes. Trop développés, ils finissent par restreindre le rôle du codeur au traitement des seules fiches rejetées par la procédure automatique, soit qu'elles soient purement intraitables, soit que la machine ne puisse trancher entre diverses solutions, soit encore que son programme bute sur une anomalie interne au questionnaire. Le codeur doit alors prendre une décision sur le cas litigieux. Pour ce

---

<sup>27</sup>J Penneff (1984), montre de façon précise les effets désastreux d'une mauvaise organisation de l'enregistrement de la profession du père lors de l'inscription des étudiants à l'université (questionnaire mal conçu, tâche peu valorisée, confiée à des vacataires sans formation etc.).

faire, il recourt à un examen de la totalité du contenu de la fiche, son expérience lui permettant d'en déduire une solution (la plus plausible, ou la plus probable), ou, à défaut, de conclure à la nécessité d'adopter une procédure aléatoire. Le tri automatique a l'avantage de concentrer l'attention sur les problèmes les plus difficiles (d'ailleurs tranchés hiérarchiquement dans une procédure non-automatique), mais aussi l'inconvénient de dissoudre le référent, ie l'ensemble des cas normaux. Dans ces conditions, on peut craindre que des recettes de cuisine diverses ne se substituent à un principe classificatoire. D'un pur point de vue organisationnel, il est souhaitable que soit maintenue une véritable responsabilité de l'opérateur sur le processus qu'il contrôle et, qu'à l'image de ce qui se passe, par exemple, dans la conduite des trains, l'automatisme n'élimine pas la participation active. En l'occurrence, le programme rejeterait un certain nombre de fiches normalement traitables, dans le seul but de maintenir une attention suffisante des codeurs à la finalité de la classification, plus qu'à l'opération de classification.

– *Tendance à la routine*

Plus généralement, le système informationnel est affecté par une tendance structurelle à l'ossification due à la prégnance des normes, que l'organisation combat en créant des champs dans lesquels les exécutants ont un rôle actif. Dans les zones, souvent peu étendues, où le codage est ardu, la tâche du codeur est alors de procéder à un choix raisonné. Il est souhaitable, pour la suite des opérations de production et de traitement de l'information, qu'il recoure le moins possible aux catégories fourretout (non-réponse, autres etc.), mais un excès de "renormalisation" est tout aussi nuisible qu'une insuffisance. Par exemple, un observateur des statistiques scolaires [Briand 1984], s'interrogeant sur une anomalie évidente (gonflement des catégories moyennes), constate que les "codeurs" (non professionnels), répugnent à ne pas coder. Interrogés sur leurs motivations, les codeurs déclarent craindre que ce comportement n'apparaisse comme une négligence dans leur travail. En conséquence, ils placent les cas douteux dans les catégories "moyennes", ce choix leur paraissant minimiser le risque d'erreur. Ce comportement "de bon sens"<sup>28</sup> génère pourtant à un biais structurel important, là où une méthode en apparence beaucoup plus arbitraire (répartition aléatoire) aurait donné des résultats plus satisfaisants. Mais, la question ne se résume pas à un problème de technique : la procédure aléatoire est aussi une échappatoire (trop) commode, alors que ses conditions d'efficacité sont strictement définies : les allocations faites de cette façon doivent avoir une structure aléatoire et être peu fréquentes.

Les statistiques sont produites par une organisation. Les problèmes de gestion des ressources humaines prennent d'autant plus d'importance que les produits statistiques sont élaborés et portent sur une matière informationnelle plus mouvante. La standardisation est un moyen efficace et économique d'assurer l'unité de la communication dans le processus (H Simon 1964). Par contre elle rigidifie le système, dilue les responsabilités et engendre des comportements routiniers.

### **2.3. Offre et demande de statistiques**

Le contexte de la production des statistiques a une influence considérable sur leur contenu. La confrontation entre les producteurs et les utilisateurs est souvent d'ordre institutionnel (B), mais elle s'effectue sur un terrain "neutre", celui de la finalité et des moyens de l'investigation (A).

---

<sup>28</sup>Il repose sur une représentation implicite, d'ailleurs suggérée par la nomenclature (cf. moyens, supérieurs), de la hiérarchie sociale. Sur cette base, le critère utilisé de fait est "minimiser l'erreur moyenne sur chaque questionnaire". Au niveau global, le biais est important, dans la mesure où la procédure agit simultanément sur les extrêmes (dégonflés) et le centre (hypertrophié).

a **Que mesurer ?**

– *Signification des indicateurs*

On peut aborder le problème à partir des idées assez différentes qu'ont le producteur et l'utilisateur quant au contenu des statistiques : d'un côté, une approche pragmatique et "concrète", de l'autre une représentation beaucoup plus abstraite, voire théorique. Le demandeur cherche une assurance, voire une réponse précise à une question qui lui semble très précise : par exemple, quelle proportion d'électeurs vont voter pour le candidat X, quel est le nombre de clients qui vont acheter le produit Y, quel est l'effectif des "pauvres" ou des mères célibataires.

Prenons le cas le plus simple, celui du marketing du produit Y. Le statisticien se préoccupe de la fiabilité de ses chiffres ; or, il connaît ses limites, qui tiennent à la méthode et aux techniques. Dans notre exemple, il ne recueille que des indicateurs d'intérêt pour le produit, des intentions d'achat etc. qui ne correspondent pas aux conditions réelles d'exercice de l'acte. Dans ces conditions, il existe un écart substantiel entre le résultat de la mesure et les conclusions fournies à l'utilisateur.

Sans doute, l'expérience a-t-elle permis d'accumuler des observations grâce auxquelles on apprécie l'ampleur des décalages en fonction des circonstances, mais la correction opérée est moins quantitative qu'analytique : elle suppose en fait une connaissance très précise du type de marché du produit, c'est-à-dire aussi bien de son segment technique/fonctionnel, que de la clientèle à laquelle il s'adresse ou des produits complémentaires ou concurrents. En d'autres termes, la garantie d'une réponse correcte repose sur une perception très fine de problèmes de commercialisation (voire de conception) du produit, qui permet d'entourer le noyau du questionnement quantitatif (relativement simple dans le principe) d'indicateurs qualitatifs, sans lesquels le résultat ne pourrait pas être calculé.

Bien entendu, la question se complique encore lorsque l'objectif est formulé en termes conceptuels : le nombre des "demandeurs d'emploi" ou de "chômeurs BIT" n'est pas le chômage.

On comprend pourquoi l'élaboration de la définition est une phase délicate : d'une part, elle requiert une collaboration étroite du demandeur et du producteur ; d'autre part, elle oblige les parties à expliciter les conceptions du phénomène observé (les pauvres, les chômeurs, les mères célibataires), ce qui ne peut manquer d'être une source de conflits.

Au total, surtout dans le cas de phénomènes aux contours diffus ou à la structure complexe, l'objectif fixé n'est simple que dans la mesure où il est le produit d'une conception simpliste du "fait", réduit à l'une de ses dimensions. Il n'est évidemment pas indispensable, et cela est d'ailleurs le plus souvent impossible, de connaître toutes celles-ci ; mais la sélection des plus importantes et leur prise en compte ne correspond pas à un dépassement abusif des objectifs assignés à la mesure, mais à une nécessité interne.

– *Contingences dues à la mesure*

Les contradictions (apparentes) entre le demandeur et l'offreur tiennent aussi aux contingences de la mesure. La mesure requiert des formulations à la fois précises, pour éviter des "flottements" sémantiques, et suffisamment simplifiées pour se prêter à l'interrogation ou à l'observation. Cela signifie que, même dans le meilleur des cas, le résultat obtenu ne prend sens que par une connaissance simultanée des faits et des procédures de mesure, qui suppose, là aussi, une collaboration périodiquement réactivée entre producteurs et utilisateurs.

Cela est d'autant plus vrai que l'exaspération des possibilités d'information renforce la nécessité d'une activité d'orientation préalable, de filtrage, sans laquelle l'utilisateur va être noyé. En effet, on ne doit pas oublier l'excellent principe de H. Simon (1964, 1983) : une information est d'autant meilleure qu'elle est synthétique, ie que le rapport entre l'output (ce qui est transmis à l'étage hiérarchique supérieur et in



fine au décideur) et l'input est faible<sup>29</sup>. Or, cette condensation ne peut être obtenue sans une très grande efficacité de la communication préalable à la circulation de l'information, sous peine de voir la "synthèse" se résumer à une simple agrégation d'informations disparates : des éléments qui s'avèrent pourtant nécessaires à la synthèse n'ont pas été recueillis, ou, dans le cas d'un balayage très large, ne sont pas connus avec précision et correspondent mal aux besoins.

L'idée que les statistiques seraient un produit prêt à consommer est l'un des facteurs les plus nuisibles lors de leur élaboration<sup>30</sup>. C'est aussi la source de bien des déboires quand on les utilise comme telles, puisque cela revient à considérer comme analogues le contenu (informel, spécifique) de la demande d'information et le produit calibré fourni par le statisticien.

## **b Un regard organisé**

La spécificité d'une investigation statistique donnée se manifeste par le choix des variables et des modalités, grâce auxquelles on construira par exemple des tableaux. L'ambition et les limites de l'observation sont inscrites dans les nomenclatures. Pour comprendre les problèmes essentiels et les enjeux qui accompagnent l'élaboration d'une nomenclature, on peut approcher celle-ci de façon générale en tant qu'activité de classement.

### *– Ambiguïtés et problèmes de l'activité de classement*

Pour en saisir la difficulté d'un classement, choisissons un exemple simple : le banal découpage en classes d'une variable quantitative déjà recueillie (âge, taille). Une division purement arbitraire se traduit, au mieux par une perte d'information, au pire par une désinformation pure et simple lorsque la série n'est pas continue et monotone. Une classification empirique fondée sur l'observation de la distribution garantit une certaine cohérence, mais sa signification peut toujours être discutée<sup>31</sup>.

Tout autre est l'enjeu d'une nomenclature : il ne s'agit pas d'un découpage a posteriori visant à identifier des particularités distributionnelles, mais d'un instrument d'observation, qui sélectionne les éléments à étudier et les hiérarchise<sup>32</sup>. Rappelons qu'il ne s'agit pas là d'un coup de force théorique, mais d'un impératif absolu, dans la mesure où le regard porté sur la réalité ne saurait être universel. Le paradoxe du classement est qu'il faudrait qu'il repose sur la connaissance de ce qu'il permet de connaître. La réalité par sa complexité est irréductible à un, voire plusieurs

---

<sup>29</sup> "... la ressource rare n'est pas l'information, mais la capacité de traitement pour s'occuper de cette information" (H Simon, 1983, p 264).

<sup>30</sup> Idée fort répandue chez les décideurs, et finalement partagée en fait, de guerre lasse, par bien des statisticiens.

<sup>31</sup> Le critère de regroupement (proximité, distance) choisi influe sur le contenu des classes... Or, le choix du critère ne peut que rarement reposer sur des motifs analytiques.

<sup>32</sup> L'analyse de l'industrie par branche apporte une incontestable connaissance, fondée sur des caractéristiques communes aux entreprises d'une même branche. On pourrait très bien imaginer un découpage des entreprises dont le critère serait la capacité à générer du profit. Il faudrait alors rechercher une information de nature tout à fait différente, et elle conduirait à un autre mode de regroupement. Les deux regards ainsi obtenus contiennent une certaine part de réalité ; ils ne sont pas incompatibles (le profit peut être une variable auxiliaire de l'étude par branche et inversement), mais si l'une a un rôle dominant, elle va conditionner toute l'observation (organisée autour d'elle) et les commentaires portant sur les résultats. Une nomenclature inadéquate finira par apparaître en tant que telle, mais une "moyennement adéquate" peut durer longtemps.

Comme le fait judicieusement remarquer M Volle, un conducteur qui n'aurait pas le concept de "feu de circulation" pourrait voir le feu au rouge et juger qu'il s'agit d'une lumière comme une autre. La connaissance des modalités n'a de signification que rapportée au concept.

classements<sup>33</sup>. Le problème est donc de constituer une grille, qui permette de rendre compte des dimensions de la réalité, sans imprimer sa marque sur elle. Cette grille se trouve prise dans les contradictions inévitables entre la formulation théorique (fournissant des principes constitutifs) et les observations empiriques, au degré de complexité très supérieur, puisqu'en elles s'entrecroisent de multiples déterminations et leurs relations, lesquelles n'ont pas toutes été formalisées par la théorie.

Au-delà de ces questions théoriques/techniques qui compliquent son élaboration, la nomenclature souffre d'être largement unilatérale :

- Elle peut être un obstacle à la communication : on sait qu'une question à choix multiples induit beaucoup de non-réponses (ou de réponses aléatoires), lorsque les alternatives proposées semblent arbitraires ou contraignantes aux répondants.

- Elle ne contribue pas toujours, loin de là, à l'unité sémantique, qui est pourtant le postulat sur lequel repose toute la collecte d'information :

- le concepteur a une finalité qui implique une forte logique interne et manipule des instruments qui exigent la rigueur. C'est là un contexte très différent de celui plus varié et plus flou des informateurs.

- si l'organisation et le support logique qu'est la nomenclature forgent une certaine identité au sein des producteurs, ces procédés ne jouent pas dans l'univers des répondants qui reste très hétérogène.

L'unité du sens est un postulat nécessaire, mais plus ou moins bien assuré, que ce soit en amont ou en aval de la quête statistique. Les conditions les plus défavorables sont, encore une fois, celles où les termes utilisés sont mal définis. Dans ce cas, le rigorisme propre à l'investigation statistique est peu efficace. Au-delà d'un certain point, il est même nuisible : en imposant ses standards, il élimine des "fluctuations" de réponse, mais aussi interdit tout échange de points de vues qui, donnant réellement la parole à celui qui renseigne, ferait éclater les divergences.

#### *– Les échappatoires*

Le bon sens suggère des solutions : créer les conditions d'une unanimité autour du sens des catégories, soit en limitant l'ampleur du champ, soit en utilisant des nomenclatures de fait.

Sur le premier point, je me contenterai ici d'un exemple. Même en interprétant de façon assez stricte le bornage de son champ théorique, l'économiste est contraint de s'intéresser aux attitudes<sup>34</sup>. Or, ses instruments s'avèrent pauvres en la matière, d'où sa préférence pour ce qui est monétarisé. Les avantages sont évidents :

- la monnaie fournit une base objective, quantifiée, et théoriquement fondée : le mécanisme du marché fait que le prix traduit la valeur objective et subjective des biens.

- l'exploitation est plus aisée (variables quantitatives) et la collecte simplifiée, puisque portant sur des actes. Par ailleurs, si le recueil se fait par voie d'enquête, on constate que les répondants se coulent facilement dans le moule de ce mode d'évaluation.

Les insuffisances nées de cet a priori sont malheureusement nombreuses. L'uniformité de l'espace monétaire est une fiction commode et trompeuse dans laquelle la spécificité du système de relations au monde qu'est la position sociale se

---

<sup>33</sup> On peut prendre une image simple : comme décrire une plante ? Un botaniste, un peintre, un chimiste etc. peuvent définir chacun un protocole d'observation extraordinairement complexe ; quoique leur quête individuelle soit infinie, ils ne saisiraient pour autant qu'une partie de la réalité de la plante.

<sup>34</sup> L'analyse néo-classique repose même totalement sur la définition d'une logique de comportement ; cependant, elle élimine la difficulté en créant un modèle idéal, rationnel et universel.

dissout. Par exemple, en ramenant la consommation à son aspect monétaire, on néglige totalement les modes de consommation ; ainsi, une enquête sur la consommation est réduite à une enquête sur les dépenses de consommation. L'identité des montants de dépenses est assimilée à une équivalence des pratiques, alors qu'une observation plus étendue révèle des différences très significatives, tout aussi décisives au niveau de l'observation macro-économique que de la pratique microéconomique.

C'est dire qu'il y a là une restriction abusive du référentiel, qui fait disparaître une large part de la substance du phénomène, et rend impossible le suivi des phénomènes actuels de différenciation qualitative rappelés ci-dessus (section 1).

Pourquoi ne pas utiliser les catégories communément utilisées par les personnes, ou catégories "indigènes" ? Elles ont l'avantage d'être "réalistes" et faciles à renseigner, mais il est clair que la disparition des catégories délimitées par le statisticien laisse le champ libre à d'autres qui le sont par le langage ou les habitudes.

Pour reprendre l'exemple des loisirs ou de la consommation, le travail des enquêteurs est singulièrement compliqué (gros effort de persuasion) lorsqu'ils prétendent obtenir une comptabilité du temps, alors que les carnets de dépense ne posent que peu de problèmes. C'est que les enquêtés ne perçoivent pas l'unité de temps comme un moyen d'appréhension de leur activité. C'est aussi, comme le montrent les recherches sur l'évaluation du travail domestique (Chadeau, Fouquet 1981), que la valorisation monétaire des services rendus est largement préférée au décompte des heures d'activité. Une réflexion sur la méthode de mesure permet de comprendre ce choix : la valeur du travail domestique, c'est son résultat ("une maison bien tenue") ; le temps de travail est associé à une "peine" et n'est pas affiché (ou sous la forme vague de "je passe mon temps à faire le ménage"). L'enquêteur, en proposant un substitut monétaire à l'appréciation qualitative, ne fait que modifier les termes de la problématique des répondants (passage du qualitatif au quantitatif), alors qu'en recourant au temps, il les inverse (de la valeur au coût).

On pourrait certes arguer que la valorisation monétaire fait partie du phénomène et justifier ainsi par le poids que lui confèrent les enquêtés l'abandon de toute autre perspective. C'est là une facilité inadmissible, car la préférence marquée pour des réponses en termes de dépense est elle-même un phénomène signifiant à deux niveaux au moins :

- pour les besoins de l'échange propre à l'enquête ce mode de communication paraît plus simple (effet relationnel).
- l'univers formulé des gens est rétréci à l'univers marchand (effet structurel).

#### *- Langage et catégories instituées*

L'unité sémantique ne peut résulter d'un coup de force. Le statisticien cherche à la favoriser, mais, pour l'essentiel, elle ne dépend pas de lui, mais de son existence de fait dans le champ étudié.

C'est là une situation plutôt rare. Il suffit d'observer les variations géographiques et historiques du sens des mots pour constater le faible consensus qui le caractérise. Et l'unité de lieu et de temps n'y change rien, même s'il s'agit seulement de définir, avec le minimum de rigueur indispensable à la mesure, n'importe quel terme couramment utilisé : dans quelles conditions (durée, type de relation, âge etc.) pourra-t-on parler de cohabitation juvénile ; une étrangère épousant un français est-elle une "immigrée" ; que signifie un emploi "stable" (un contrat à durée indéterminée peut être rompu) etc. (ML Levy 1987).

Faute de pouvoir s'assurer de cet accord sur un sens commun, on explícite dans la question le contenu préalable à la classification. Le résultat n'est pourtant pas garanti : il faut que ce contenu soit reconnu et adopté par le répondant et qu'il corresponde aux intentions du questionneur...

## Normalisation statistique et plasticité sociale

Les formes juridiques, qu'il s'agisse d'institutions ou de statuts (la famille, le mariage, la paternité etc.) ou simplement de droits (à indemnisation du chômage) "fixent" le cadre d'observation, lui donnent une objectivité qui naît de l'évidence. Or, l'histoire, même récente, montre des évolutions marquées de ces catégories "évidentes" ; a fortiori pour une notion beaucoup plus complexe, comme celle d'individu.

Plus généralement, les catégories instituées imprègnent profondément le regard porté sur la société par les personnes interrogées, servent de point de référence. Ceci présente l'avantage d'ancrer la mesure dans la réalité : d'une part, l'interrogation est légitime (elle se sert de catégories en usage), d'autre part, le flou sur la réponse est restreint par la connaissance (au moins globale) du référentiel qu'ont les individus.

A contrario, cela veut dire que nombre de phénomènes qui ne sont pas institutionnalisés se prêtent mal à la quantification. Or, ceci est particulièrement gênant dans la mesure où ce manque de définition est caractéristique des phénomènes changeants. Se trouvent ainsi éliminées les phases de constitution ou de dissolution. En d'autres termes, l'analyse quantitative tend à donner une vision statique du fonctionnement de la société, tant parce qu'elle se détourne de ce qui change, que parce qu'elle "immobilise", par la rigidité de ses catégories, les modifications de contenu que celles-ci ont subies.

### *- Mesurer c'est instituer*

Il y a inévitablement dans l'opération de mesure une tendance à instituer. Ses conséquences pourraient être assez restreintes : les décalages entre la réalité et le contenu des catégories statistiques seraient absorbés par des réajustements périodiques. Ce raisonnement suppose la "neutralité" de l'opération de mesure. Or, on ne doit pas oublier que les statistiques servent à agir/réagir tout autant qu'à connaître. Dans ces conditions, des décisions sont prises en fonction de représentations plus ou moins adéquates, avec des conséquences diverses : parfois les "faits" résistent, parfois l'intervention est suffisamment puissante, pour modifier les représentations sans toucher au réel, voire pour influencer sur le réel (conformément aux intentions premières ou non). Quel que soit le résultat, la grille d'analyse ne peut éviter de laisser une marque, ne serait-ce qu'en formalisant des groupes (d'appartenance, d'intérêt, de revendication), qui polarisent les réactions.

On pourrait objecter que nul n'est tenu de porter les "lunettes" propres à la statistique. Ce serait faire peu de cas des relais divers par lesquels les statistiques exercent leur influence. J'ai évoqué ci-dessus certains effets, directs (émergence d'une catégorie à la suite d'une intervention administrative ou législative, ou même d'un groupe social), mais il en existe un autre, plus insidieux parce que méconnu : les statistiques influent sur la production de la connaissance, quoiqu'elle soit dotée d'une autonomie bien plus grande que l'action.

En effet, les sciences sociales sont fortement tributaires de la quantification, en raison de la difficulté qu'elles ont à se fonder dans un corpus théorique incontestable. F Héran [1984] a bien souligné l'importance de la tradition de Durkheim en sociologie : en quête d'une objectivation, Durkheim introduit l'usage des statistiques (cf. le "Suicide") sans s'interroger sur leur contenu implicite.

*"De fait, la statistique entre en scène dans le Suicide comme un instrument qui va de soi, sans qu'apparaisse aucune critique sur la confection des données, ni aucune interrogation sur l'institution des catégories statistiques et, en particulier les effets de sélection qu'elles impliquent" (op. cit. p 25)*

Or, les catégories statistiques sont tout à la fois instituées et instituantes : les mieux instituées sont justement celles qui ressortent à la problématique, souvent pratique, des statisticiens. Comme le dit bien l'auteur :

*"[...] alors que la statistique tire toute sa force de l'exploitation d'un registre limité et sélectif, la sociologie joue sur un registre plus étendu et plus mal défini" (id p 27)*

Dans ces conditions, la démarche sociologique, mais aussi économique, en quête de justification est tentée de recourir à l'ensemble des "preuves" que lui propose la statistique, non pas parce qu'elles sont plus satisfaisantes du point de vue de l'analyse, mais parce qu'elles sont plus convaincantes (parce que mieux instituées, moins contestables) dans l'argumentation que les discours théoriques. Il se produit alors un certain enfermement de la démarche, une banalisation du débat théorique et méthodologique, puisque les armes les plus décisives sont celles, communes et "omnibus", de la pratique statistique.

Si à la dimension "passive" de la non-neutralité de la statistique (il ne s'agit pas d'une chambre d'enregistrement), on ajoute sa dimension active d'intervention sur la réalité, c'est tout un nouveau champ d'investigation qui s'ouvre, celui des conditions de sa mise en oeuvre .

### **c Le contexte institutionnel**

Quelle est l'importance des conditions de production des statistiques pour leur contenu<sup>35</sup> ?

L'analogie avec la production de biens est, de ce point de vue, assez frappante. Un produit n'existe que dans la mesure où s'exprime une demande que les offreurs voudront satisfaire. Chacun sait que l'offre contribue largement à créer et à modeler la demande et que les formes de marché ont des incidences non négligeables.

Si nous étudions le cas du produit "statistiques", il apparaît immédiatement que la relation entre producteurs et utilisateurs ne s'apparente que rarement à un marché concurrentiel. Décrivons quelques-unes des situations possibles.

#### *- La statistique au service de l'Etat ; l'exemple français*

La statistique est, dès ses origines, associée aux préoccupations gestionnaires de l'Etat. Même si son champ s'est désormais largement étendu, cet aspect reste essentiel. Les administrations publiques, en lien avec l'interventionnisme étatique, sont devenues les principaux producteurs et consommateurs de données.

Il est clair que la production interne, directement issue de préoccupations de gestion publique a des effets significatifs sur la réalité observée, ne serait-ce que par ses liens "organiques" avec la réglementation en vigueur. Aussi, je m'intéresserai surtout au cas apparemment plus "neutre", où le produit statistique est fourni par des organismes publics ou parapublics spécialisés.

Ceux-ci étant tenus de répondre aux demandes qui leur sont adressées, on peut penser à un monopsonne ; mais l'ampleur même des demandes, leur proximité du Politique, la centralisation de la production etc., ont conduit à une autonomie de fait, à une "magistrature du chiffre", dotée d'une assez large indépendance<sup>36</sup>. L'INSEE français est un exemple caractéristique. Les conséquences sur l'appareil d'observation et la méthodologie sont tout à fait évidentes :

- D'une part, la demande publique très importante a eu un effet mécanique sur la définition des champs observés et la conceptualisation.

- D'autre part, la prépondérance des demandes publiques, l'autonomisation relative, la traditionnelle méfiance des entreprises envers une administration

---

<sup>35</sup>Comme toute production, celle des statistiques est soumise à des contraintes multiples entre lesquelles il faut arbitrer : fiabilité bien sûr, mais aussi rapidité et coût. Les aspects "gestionnaires" de la production statistique, qui mériteraient une étude spécifique, ne sont pas étudiés ici.

<sup>36</sup>La proximité de la politique se traduit par de fréquentes contraintes, qui agissent en "aval" de la production, sur la publicité données aux résultats (choix des indicateurs retenus, restrictions diverses sur d'autres). Au contraire, en amont, la nature fortement conflictuelle du Politique, rend indispensable l'existence d'une vision gestionnaire "neutralisée". Un très bon exemple de problèmes institutionnels se trouve dans Volle M 1980.

## *Normalisation statistique et plasticité sociale*

publique, ont produit un effet d'éviction touchant aux autres demandes, qui n'ont pas été considérées ou l'ont été tardivement, incomplètement, et le plus souvent dans la perspective générale propre à l'établissement<sup>37</sup>.

### *– La statistique privée*

Sans doute existe-t-il des entreprises privées de production de statistiques. Elles se trouvent très directement confrontées à la nécessité de dépasser l'univers trop restreint des gestionnaires, alors même que la collecte qu'on pourrait qualifier de fondamentale reste en dehors de leurs possibilités. D'où finalement le fait que cette production de données privées concerne des secteurs très spécialisés (marketing en général) ou des variables associées directement à la gestion et issues de sources internes (formes diverses de comptabilités analytiques, de contrôles de gestion etc.).

Quel que soit leur intérêt propre (voire leur nécessité), les produits ainsi obtenus se prêtent mal aux comparaisons (avec d'autres entreprises), ou à l'ouverture sur des sources externes. Par ailleurs, leur conception, régie par des principes érigés, en limite singulièrement l'intérêt face à des situations mouvantes ou à la prise de décision stratégique. Enfin, il est rare que les statistiques cherchées ne soient pas directement issues d'une volonté de contrôle interne, ce qui a des incidences sur leurs conditions de recueil et leur fiabilité.

### *– Vue d'ensemble*

En bref, il existe une statistique générale, assurant une certaine cohésion entre ses divers aspects, orientée vers la réflexion à terme, la comparaison, et des statistiques locales (opérationnelles), adaptées à la gestion interne, à la fixation d'objectifs à échéance limitée. On pourrait imaginer une certaine complémentarité entre ces niveaux. Les contextes évoqués ci-dessus montrent qu'il n'en est rien. Du point de vue du contenu, les objectifs de connaissance et de gestion ne peuvent qu'avoir des rapports assez distendus et une compatibilité limitée, même sur des champs identiques. L'isolement dans des réseaux de fonctionnement particularisés (public/privé) a élargi le fossé : d'un côté, un développement autocentré, un expansionnisme d'origine interne, tempéré par la contrainte de la demande publique ; de l'autre, un univers éclaté de fait, même si pour l'essentiel il reste prisonnier d'une méthodologie simpliste et répétitive<sup>38</sup>.

C'est ainsi que les disjonctions propres à la société française s'inscrivent dans la structure même de l'appareil d'observation. Il s'agit là d'une source de biais, touchant au regard que la société porte sur elle-même. L'aspect principal du biais est de surévaluer les divergences de points de vue entre acteurs publics et privés<sup>39</sup>.

---

<sup>37</sup> Cf. dans le n° 100 d'Economie et Statistique (1978) le constat de Malinvaud (ancien directeur général de l'INSEE) lui-même, où l'étude sur la clientèle de l'INSEE, qui montre qu'elle s'intéresse principalement... au recensement (Marc G, Templé P).

<sup>38</sup> Le rapport Riboud 1987 (p 56-61) met fort justement l'accent sur l'information insuffisante des petites entreprises. Ces 200.000 clients potentiels se heurtent à une information inadaptée ("le système travaille essentiellement pour les besoins de l'administration et des grands groupes"), ou inaccessible (ensemble lacunaire, dispersé).

<sup>39</sup> La surévaluation tient à ce que la différence des points de vue, inscrite dans la structure, est nécessairement démultipliée par la systématisation propre aux procédures statistiques. A la limite, on arrive à une pathologie classique où, du fait d'une mauvaise communication, les discours des interlocuteurs se referment sur leur logique jusqu'à ne plus fournir aucune ouverture réciproque.

### 3. Conclusion

A l'issue de cette exploration, il convient de réexaminer son point de départ. L'approche retenue s'est avérée fructueuse, mais elle a aussi, chemin faisant, montré ses limites.

#### 3.1. L'adaptabilité limitée de la méthode statistique

---

L'adaptation des statistiques à l'observation de la société actuelle, et plus généralement de toute société où se produisent d'importants remodelages internes, met fortement en cause la méthode et le processus d'élaboration.

##### a Une souplesse accrue...

Il est possible d'imaginer un système de production bien plus souple que celui qui existe actuellement et de nombreuses indications qui précèdent vont dans ce sens. Cependant, il est remarquable que cette réponse de bon sens se heurte aux caractéristiques de la méthode elle-même. En particulier, il faudrait envisager une adaptation permanente des cadres et nomenclatures, ce qui aboutirait, selon la rapidité d'adaptation et la nature des évolutions en cause, à diverses erreurs :

- en cas de modification lente, un effet "mécanique" de déformation des changements du fait d'une certaine inertie des adaptations.
- dans le cas contraire, à une endogénéisation des glissements sémantiques, qui enlèverait tout sens à l'observation comparative des catégories: le "suivi" du phénomène reviendrait à amalgamer changement symptomatique et instabilité structurelle.

La statistique suppose un minimum de stabilité des notions qu'elle utilise. Sauf à admettre son incapacité totale, c'est à une redéfinition de ses objets qu'elle est confrontée. En simplifiant, le niveau des stabilités structurelles a changé, s'est déplacé vers des schémas de détermination moins factuels, plus stratégiques. Pour les comprendre, il faut traduire cette structuration dans l'investigation: à l'observation des actes élémentaires, il faut substituer la reconstitution de schémas combinatoires plus complexes. C'est pourquoi la recherche de détails supplémentaires, le compartimentage plus fin des champs, qui semblent améliorer la fiabilité, courent à l'échec, dans la mesure où ils accroissent l'instabilité des champs observés et les vident de sens. En d'autres termes, la seule issue possible semble être une "théorisation" accrue.

##### b ...remet en cause le procès de production

En procédant ainsi, on ne peut qu'accroître les tensions internes à la production, dans la mesure où l'information cherchée s'éloigne des conditions favorables à la vérification de la condition d'unité sémantique, dont nous avons vu qu'elle est d'autant plus probable que le décalage avec les divers vécus est faible et peu évolutif. Or, la conjonction de ces deux hypothèses ne se réalise pas au hasard, mais quand le fait social analysé est clairement désigné (i.e. institutionnalisé) ou lorsqu'il est possible de fixer précisément un protocole de mesure (observation d'actes). Cela explique que la statistique tend à suivre une certaine ligne de pente, à préférer les catégories bien établies ou aisément formalisables. L'élévation du niveau d'appréhension, suggérée ci-dessus, implique des formalisations beaucoup plus abstraites, la création d'indicateurs plus efficaces pour apprécier les attitudes, mais qui ne correspondent pas à une catégorisation admise par tous. Les PCS qui ont cette caractéristique sont un bon exemple, mais il en existe d'autres (passage du mariage à la relation de couple).

L'émergence d'une "ethnométhodologie" des statistiques est tout à la fois l'une des conséquences d'un certain état de crise interne, un élément de son accélération et,

l'occasion de s'engager dans des directions très différentes. En particulier, si l'on voit bien les effets positifs de ces réflexions (abus de l'usage des statistiques comme "preuve", attention portée aux modes d'organisation), on en perçoit aussi les perversions probables. Car, comme l'avait noté Héran, aux problèmes qu'elles rencontrent les statistiques peuvent toujours opposer une solution apparente incorporant davantage de statistiques. De ce point de vue, la réflexion sur les erreurs à l'enregistrement (dont on trouve un exemple très conséquent dans Merllié 1983), peut tout à fait être prolongée par une série d'études statistiques comparatives sur les comportements des personnes face aux classifications, permettant, dans une certaine mesure de limiter ou corriger les biais les plus flagrants<sup>40</sup>.

A défaut d'éliminer la présence de l'observateur (de son regard, des nécessités de la construction des observations etc.), on pourrait "simuler" son absence. Les techniques de redressement de biais auraient de beaux jours devant elles, et, sans doute, les statistiques s'éloigneraient-elles encore un peu plus de la "réalité". Après avoir fait répondre ceux qui ne répondent pas, elle rectifierait a posteriori les flous et les vagues de l'enregistrement, au risque de gommer l'importance de ces incertitudes.

C'est dire qu'en restant dans la perspective de connaissance définie ici, on ne peut éviter de poser la question des présupposés épistémologiques de la méthode.

### **3.2. Les statistiques comme système rationnel**

---

#### **a De l'irrationalité du rationnel**

En effet, on a pu remarquer dans ce qui précède qu'à chaque niveau, on peut associer un type de problèmes et des esquisses de solutions, qui vont toutes dans le sens d'un renforcement de la logique interne. Les déchets de la méthode sont réingérés par la méthode. On assiste alors à une hyper rationalisation, dont E Morin (1987, p 97) donne une description parfaite :

*"La rationalisation enferme l'Univers dans un système cohérent, construit à partir de postulats jugés indiscutables, et rejette tout fait, toute idée, tout argument non conforme avec le système. La rationalisation est une machine à justifier de façon cohérente toute croyance ou idée qui veut échapper à la critique rationnelle."*

Alors, la rationalisation s'éloigne de la rationalité, puisque celle-ci n'est pas une valeur transcendante, mais "un dialogue entre l'esprit et le réel : l'esprit veut correspondre avec le monde en élaborant des conceptions qui lui correspondent." (id p 96).

#### **b Volonté de connaissance ou de contrôle ?**

Dans ce cadre, on peut mieux comprendre les relations entre économie et statistique, en se fondant sur celles, mieux observées, qui existent entre sociologie et statistique: si celles-ci sont si proches, si, en apparence, la sociologie ne peut que pousser la statistique vers davantage de statistiques, c'est-là le fait d'une sociologie particulière, cherchant au sein d'une vision réifiée de la structure sociale un principe explicatif. La tradition post-Durkheimienne n'y est pas pour rien<sup>41</sup>.

Comme le souligne F Dubet (1987), la question fondamentale est le statut de l'acteur, qui n'est conçu que comme "réagissant" dans des situations de contrainte. Dans ces conditions, on définit des stéréotypes, mais on voit s'échapper de la "grille",

---

<sup>40</sup> L'étude de Peneff 1984, bien qu'elle ne soit pas faite dans ce sens, pourrait fort bien être utilisée pour cela.

<sup>41</sup> Qui semble avoir passablement caricaturé Durkheim. Cette déformation n'est pas le fruit du hasard : c'est bien une approche rigoureuse qu'il recherchait, mais, alors qu'il en soulignait de facto les limites, ses successeurs se sont emparés de son système.



### *Normalisation statistique et plasticité sociale*

---

tout ce qui dépasse la simple adaptation, ie tout ce qui est création sociale. Le monde des statistiques est un monde où le mouvement est toujours saisi historiquement, rétrospectivement, lorsqu'il a pour une part détruit le cadre de saisie. Ce sont bien les statistiques qui sont en position d'adaptation aux contraintes, et qui, pourrait-on dire, prêtent leur comportement à l'acteur, le sommant de fonctionner comme elles (observer, s'observer); pis, tout se passe comme si elles détruisaient le produit original de son action sur la société puisqu'elles l'intègrent dans leur référentiel, lui conférant le statut de l'évidence et de la banalité. Condamnée à fonctionner par sauts la statistique se prétend d'autant plus le gardien de la mémoire sociale, qu'elle en a effacé les acquis principaux, ces modifications de forme et de contenu des phénomènes, qui sont les créations qualitatives du social. Dans un monde en voie d'institutionnalisation, tendu vers la croissance quantitative, elle a bénéficié de privilèges exorbitants. Les nouvelles remises en cause qu'elle connaît sont le signe de la crise interne qu'elle subit, crise conjoncturelle, mais aussi structurelle : c'est l'inadaptation fondamentale de son formalisme à rendre compte de la plasticité interne au social qui est en jeu.



## Bibliographie

- ANCORI B 1983 "Communication, information et pouvoir" in LICHNEROWICZ(éd)"*Information et communication*" Maloine, pp 59-84.
- ARTUS P 1988 "Brève histoire de la modélisation" in MEURS D (ed) "*L'information économique*" Cahiers français, n° 235, mars-avril.
- BENEDETTI A 1977 "L'évolution des cadres et des nomenclatures comptables" in INSEE 1977 *PUH t2* pp 389-420.
- BESSON JL 1988 "Mesure et réalité du chômage au Japon et en France" in PUL, "*Japon, stratégies industrielles et enjeux sociaux*", pp 155-216
- BESSON JL, COMTE M, ROUSSET P 1981 "*Compter les chômeurs...*" Presses universitaires de Lyon, 401 p.
- BESSON JL, JOURNET O 1986 "Le nombre et son ombre" in "*Des Mesures*" BESSON, COMTE (éd) 1986, pp 13-56.
- BESSON JL, COMTE M 1986 "Du compte à l'affabulation: les statistiques du chômage", *id*, pp 289-324
- BESSON JL, COMTE M (éd) 1986 "*Des Mesures*" PUL, 331 p.
- BESSON JL, COMTE M, ROUSSET P 1987 "*Evaluation des politiques de chômage. Elements de Problématique*" Rapport pour le Commissariat General du Plan 2t, 401p.
- BOLTANSKI L 1982 "*Les cadres. La formation d'un groupe social*" Ed de Minuit.
- BRIAND JP 1984 "Sur quelques conséquences des différents emplois du code des CSP" *Economie et Statistique*, n° 168, juil-août, pp 45-58
- CHADEAU A, FOUQUET A 1981 "Peut-on mesurer le travail domestique ?" *Economie et Statistique* n° 136, sept.
- CHARRAUD A, VALDELIEVRE H 1981 "La taille et le poids des français" *Economie et statistique*, 132, avr, pp 23-38.
- CHAUCHAT H 1985 "*L'enquête en psychologie*" PUF, 253 p.
- CHAUMIER J 1986 "*Systèmes d'information. Marché et technologies*" Entreprise moderne d'édition, 117p.
- CHEVRY G 1962 "*Pratique des enquêtes statistiques*" PUF
- CICOUREL AV 1964 "*Method and measurement in sociology*" Free Press of Glencoe, 247p.
- DE SINGLY F 1982 "La gestion sociale des silences" *Consommation, revue de socio-économie*, n° 4, pp 37-63.
- DE SINGLY F 1984 "Les bons usages de la statistique dans la recherche sociologique" *Economie et Statistique*, n° 168, juil-août, pp 13-22
- DESABIE J 1965 "*Théorie et pratique des sondages*" Dunod.
- DESROSIERES A 1986 "La portée sociologique des divers phases du travail statistique", in "*Des Mesures*", BESSON, COMTE (éd) 1986 pp 247-274.
- DESROSIERES A, THEVENOT L 1988 "*Les catégories socio-professionnelles*", La Découverte, 122 p.
- DUBET F 1986 "Les modèles modèlent-ils ?", in "*Des Mesures*", BESSON, COMTE (éd) 1986, pp 275-288.
- GARFINKEL H 1967 "*Studies in ethnomethodology*" Englewood Cliffs, Prentice Hall
- HERAN F "Résistance à la statistique, résistance à la sociologie" Ronéoté.
- HERAN F 1984 "L'assise statistique de la sociologie" *Economie et Statistique*, n° 168, juil-août, pp 23-36
- KUHN TS 1962 "*La structure des révolutions scientifiques*" Flammarion, 1972.
- LEMEL Y 1984 "Le sociologue des pratiques du quotidien entre l'approche ethnographique et l'enquête statistique" *Economie et Statistique*, n° 168, juil-août, pp 5-12.
- LEVY ML 1986 "Nommer pour compter", *Populations et sociétés*, mai.
- MALINVAUD E 1978 "Répondre à la demande" *Economie et Statistique*, n°100, mai
- MARC G, TEMPLE P 1978 "L'INSEE et son public" *Economie et Statistique*, n° 100, mai.

*Normalisation statistique et plasticité sociale*

- MARCH JG, SIMON HA 1964 "*Les organisations. Problèmes psycho-sociologiques*" Dunod, 244 p
- MERLLIE D 1982 "Sur la production de données statistiques: l'origine sociale des étudiants" in "*Etudes dédiées à M Grawitz*", Dalloz, p 77-97.
- MERLLIE D 1983 "Une nomenclature et sa mise en oeuvre. Les statistiques sur l'origine sociale des étudiants." *Actes de la Recherche en sciences sociales*, n° 50, nov. 3-47
- MINGAT A, SALMON P, WOLFELSPERGER A 1985 "*Méthodologie économique*" PUF, 576 p.
- MORGENSTERN O 1950 "*L'illusion statistique. Précision et incertitude des données économiques*" rééd, Dunod, 1972, 288 p.
- MORIN E 1987 "*Penser l'Europe*" Gallimard, 222p
- PASSERON JC 1982 "Le langage de l'abstention" *Consommation*, n° 4, p 3-11.
- PEAUCELLE JL 1981 "*Les systèmes d'information. La représentation*" PUF, 249 p.
- PENEFF J 1984 "La fabrication statistique ou le métier du père" *Sociologie du travail*, n°2, pp 195-211.
- PUH "*Pour une histoire de la statistique*", *Economica/INSEE*, t1 1977, t2 1986.
- RIBOUD J 1987 "*Modernisation mode d'emploi*", UGE 10/18.
- SIMON H 1983 "*Administration et processus de décision*", *Economica*.
- VOLLE M 1980 "*Le métier de statisticien*" Hachette, 298 p.
- WATZLAWICK P & alii 1972 "*Une logique de la communication*" Le Seuil, 277p.