

Le flou et le dur

Maurice Comte

Mariage, métier, département... Maintes catégories traditionnelles deviennent floues. La demande d'informations se fait de plus en plus pressante et la satisfaction plus difficile. Les statistiques entrent dans une zone de turbulences : comment observer ce qui n'est pas déjà normalisé ?

1

Les méthodes d'investigation statistique, pas plus que les autres, ne sauraient être universelles. Certaines conditions d'usage leur sont favorables, d'autres moins. Le domaine de **l'observable** s'est considérablement agrandi depuis le début du siècle (variables économiques, sociales etc.), ce qui pouvait laisser penser qu'un jour on connaîtrait "tout", ceci d'autant plus que les progrès des techniques d'information et de communication profitent largement aux statistiques.

Je me propose de montrer que les mutations récentes de la réalité économique et sociale mettent en cause le sens et la pertinence d'une part importante des statistiques, et que les remèdes "classiques" peuvent être pires que le mal, prolongeant indûment l'illusion d'une précision perdue.

En toute rigueur, la quantification n'est possible que si elle porte sur des "objets" identiques: elle se limite alors à un dénombrement. On compte des moutons ou des tonnes de pétrole, mais cela n'a pas grande utilité, sauf pour s'endormir. Si on raisonne en termes d'énergie, il faut additionner des barils de pétrole, de tonnes de

charbon, de Kw/h etc. La seule solution est de définir une relation d'équivalence, telle que la Tonne Equivalent Pétrole (TEP).

Cet exemple montre une propriété importante de la relation d'équivalence: elle n'a de sens que si on est capable de définir a priori son champ et son **principe**. Dans le cas de l'énergie, la théorie physique et des rendements mesurables fournissent un référent solide; il n'en est pas de même dans la plupart des champs décrits par les statistiques économiques et sociales. Ainsi, un sondage sur les intentions de vote n'a de sens que dans la mesure où il existe un lien stable entre des déclarations d'intention et les votes effectifs. De même, il y aurait grand risque à fonder une étude de mercatique sur les seuls chiffres des achats effectués par les ménagères: ils sont un indice du pouvoir d'attraction (au sens le plus global) du produit et peuvent varier au gré de conjonction passagère de facteurs sans que soient modifiées les attitudes sous-jacentes.

Toute statistique repose sur le postulat d'un lien étroit entre ce qui est observable et ce que l'on souhaite connaître. Si le postulat est valide, l'indicateur est une approximation satisfaisante. Or, depuis une dizaine d'années, la liaison entre les actes et les motivations a perdu de sa force.

De l'ordre au magma ?

Une société hiérarchisée repose sur des règles de comportements, sur des échelles de différences stables qu'illustrent bien les "lois de l'imitation". Le champ social est alors topographié: la possession d'une automobile vers 1955 constituait un solide indice de positionnement social et permettait une bonne approximation des comportements.

Au contraire, dans la société dite "de classes moyennes"¹, la plupart des actes de la vie quotidienne échappent à une codification précise. Aucune structure n'est plus absolument contraignante (affaiblissement des normes), ni satisfaisante (laminage des différences). Alors qu'elle n'était qu'un sous-produit "naturel" du statut, la différence est alors recherchée pour elle-même, et sa quête ne saurait avoir de fin. Dans ce contexte, les tactiques individuelles qui se développent peuvent être qualifiées d'**opportunistes**, puisqu'elles poursuivent des buts **sans considération des moyens**. Les événements enregistrés par les statistiques (les actes), qui ressortent aux seconds, perdent alors leur sens.

¹ ou selon l'expression de Weil "psychomatricielle" (Weil, *Et moi, émois*, Editions de l'organisation, 1986)

En matière de consommation par exemple, la satisfaction des besoins dits "élémentaires" et la diversification des produits (résultant en partie de la concurrence accrue) dilatent les sphères de choix. Ainsi la consommation de vêtements ne suit plus la logique sociale de la différenciation ou la logique symbolique de la mode (ou sa variante anti-mode) mais une logique individuelle qu'exprime la notion de "look"²: la panoplie remplace l'uniforme.

Des molécules aux atomes

L'atomisation de la société va de pair avec une dissolution des cadres mêmes de l'observation statistique. Citons en deux: le ménage et l'espace.

i) Le ménage: on entend par là un ensemble de personnes ayant la même résidence principale. Il s'agit d'un ménage/logement. Celui-ci est le cadre traditionnel d'observation à la fois pour des raisons de commodité statistique et parce que maints comportements individuels ne peuvent s'expliquer qu'en référence à la "cellule de vie". Or, l'unité de celle-ci est remise en cause.

D'une part, le ménage ne forme plus une entité aussi bien identifiée: la prolifération des divorces, remariages, cohabitations etc., compliqués de la présence d'enfants "alternatifs", multiplie les relations entre "ménages"³, et il conviendrait dès lors de distinguer entre "noyau" et "périphérie" et de prendre en compte les consommations effectuées dans d'autres ménages.

D'autre part, l'homogénéité interne du ménage s'affaiblit: la satisfaction des besoins "collectifs" de la famille (équipement ménager etc) se traduit par un développement des consommations individuelles qu'illustrent bien la multi-possession (présence de plusieurs automobiles, et surtout l'équipement individuel en HIFI/vidéo), ou l'individualisation croissante de la consommation alimentaire⁴: le grignotage remplace le repas pris en commun.

Peut-on encore continuer de cumuler les dépenses au sein du ménage, lorsque celui-ci tend à n'être qu'une collection, d'étendue variable, d'individus?

ii) L'espace. Le découpage administratif du territoire (Etat, régions, départements, arrondissements, cantons, communes) offre

² La différence ne ressort ni à tel ou tel vêtement, ni à une norme d'assemblage prédéfinie, mais à un "arrangement" qui se doit d'être unique. Il est sans doute possible de reconstituer *a posteriori* des "règles d'arrangement", mais rien ne garantit leur stabilité. Yonnet, *Jeux, modes et masses*, Gallimard, 1985.

³ cf Annie Fouquet Rapport CNIS long terme, *Courrier des statistiques*, n° 50/51, 1989 ?

⁴ Herpin, Verger, *La consommation des français*, Repères, La Découverte, 1988.

longtemps un cadre tout prêt à l'observation statistique. Son principal avantage était d'être fortement **institué** et d'avoir de ce fait une évidence "naturelle" ⁵. Cette structure se délite à la fois par en haut et par en bas. Par en haut, avec les progrès de l'interdépendance internationale, et de la construction communautaire. Par en bas, car les espaces pertinents ne rentrent pas dans les découpages administratifs. La statistique tente alors de définir de nouveaux objets, aux recoupements complexes: de même que la notion d'agglomération s'est substituée à celle de ville, les bassins d'emploi ou de formation débordent les frontières administratives etc. Prospectivement, le remplacement des liaisons physiques par des liaisons immatérielles (télématiques) vide de sens la notion d'espace territorial et conduit à la remplacer par la notion de **réseau**, autrement difficile à saisir⁶.

Cette complexité nouvelle du monde n'interdit pas toute description et analyse statistiques, mais les rend problématiques en affectant **toutes** les conditions de l'observation.

Les défis adressés à l'observation statistique

L'investigation doit se porter **en amont** des faits, recréer un niveau d'analyse où la notion d'équivalence soit admissible. Ceci est d'autant plus difficile que la dissolution des structures anciennes ne fait pas apparaître une nouvelle grille⁷, une esquisse qu'il suffirait de passer à l'encre.

Au contraire, la stabilité conceptuelle des catégories n'est plus assurée. Il ne s'agit pas des glissements entre catégories, qui sont peu gênants, car les progrès de l'informatique permettent les raccordements de séries, les calculs rétrospectifs, et, dans une certaine mesure les recodages. Le vrai problème est de savoir si les phénomènes exprimés par des catégories (même inchangées) sont identiques. Par exemple, sans évoquer ici les flous de la notion de chômage, la comparaison des chiffres du chômage au début des années 70 et au début des années 90 (fût-ce dans un même pays) est-elle licite? 3 millions de chômeurs, est-ce bien six fois plus que 500.000? Il ne s'agit plus des mêmes chômeurs; ils ne sont pas dans le même environnement: la croissance du chômage et les politiques

⁵ Héran, j « L'assise statistique de la sociologie », *Economie et Statistique*, n°168, 1984.

⁶ Alors que le réseau des transports peut se définir à l'aide des flux, le réseau des télécommunications a une structure "fonctionnelle" qui peut être déconnectée de la structure spatiale (ex: le centre de Pleumeur Bodou). L'observation des flux de données n'est alors pas significative.

⁷⁷ soit par myopie, soit parce que la période de mutation n'est pas parvenue à son terme

publiques qui l'ont accompagné ont à la fois banalisé et adouci la position de chômeur.

Si le postulat d'identité de l'objet est infirmé, aucun protocole de mesure ne peut garantir la valeur des résultats.

Certes, ces problèmes ne sont pas nouveaux: ce sont mêmes les premiers qu'a rencontrés l'observation statistique au début du siècle. Ils n'ont pas été définitivement résolus: les progrès de la connaissance ne sont pas tous cumulatifs car certains sont imputables au contexte. De l'après-guerre aux années 70, la maturation d'un régime de croissance doté de règles de fonctionnement stables, a réduit les flous et évité les changements brusques, qui sont également préjudiciables à la signification et à la continuité des statistiques. Celles-ci bénéficiaient d'un "travail" (social et politique) de catégorisation externe qui leur donnait leurs cadres. Ce n'est pas par hasard que les problèmes fondamentaux des nomenclatures, de la signification ont resurgi dans les années 70 du sein même de l'institution statistique⁸.

Questions ou viol des consciences ?

On néglige trop souvent le fait que les "données" sont obtenues à partir d'informations pré-établies (les statistiques sont alors un sous-produit d'autres opérations) ou à partir d'informateurs.

Tout questionnaire recueille des **déclarations** et non des faits. Outre la présomption de bonne foi du déclarant, deux postulats sont nécessaires à la validité du questionnaire: i) la réalité est unique quelles que soient les circonstances, la situation du répondant etc. ii) l'informateur peut en transmettre une vision objective.

De manière générale, maints travaux de méthodologie ont discuté ces postulats et remis en cause la **neutralité** de l'observation ainsi réalisée⁹. Tant que l'observation portait sur des faits, il était admissible de négliger ce problème ou de le considérer comme un biais à minimiser. Si la possession d'une automobile par exemple est un fait discriminant, il suffit de poser la question. Même si celle-ci est parasitée par la relation d'enquête, les comportements de prestance etc., elle porte sur un fait objectif au sens où tout le monde peut être d'accord sur sa caractérisation.

⁸ cf Volle, Thévenot, Desrosières etc

⁹ Mellié, in Champagne et al. *Initiation à la pratique sociologique*, Dunod, 1989, reprenant Sardon in *Population*, 1986/1. Il faut préciser que le questionnaire est entièrement imprégné de l'idée que, soit par pudeur, soit par ignorance, les répondants ne répondent pas avec exactitude.

Au contraire, pour appréhender un phénomène complexe, il faut associer les "mots" que sont les questions au sein d'une "syntaxe", le questionnement, pour éliminer ou réduire les ambiguïtés. Ce faisant, la distance avec les "faits" augmente, le questionnaire s'épuise (et épuise l'informateur) en accumulant les questions.

Donnons un exemple: Merllié a bien montré comment un questionnaire portant sur les méthodes de contraception prenait un caractère policier du fait de la différence entre la représentation savante (celle du démographe) et la représentation spontanée.

Dans un autre domaine, le développement des situations intermédiaires entre chômage et inactivité, entre chômage et emploi et entre emploi et inactivité a conduit le questionnaire de l'Enquête Emploi à multiplier les questions pour détailler à l'extrême une réalité devenue floue.

Cette inflation du questionnement a des conséquences pratiques: le coût de l'enquête augmente (recueil et traitement des informations); le risque de rejet aussi, soit directement (non réponses), soit indirectement (dérobades, réponses peu sérieuses). En tentant de saisir "toute la réalité", l'enquête prend un tour **inquisitorial** et porte atteinte à la vie privée. Plus généralement, en incorporant très fortement, au travers de l'agencement des questions, l'idée directrice de l'enquêteur, tout questionnaire devient sondage d'opinion, avec les conséquences que l'on sait sur la valeur et la signification des indices recueillis¹⁰.

La valse des étiquettes

Les difficultés précédentes, qui concernent la conception du questionnaire comme grille de saisie de la réalité engendrent ultérieurement des problèmes pour classer les informations dans des nomenclatures¹¹.

Le cas idéal est celui où on affecte sans ambiguïté telle réponse à telle case: une machine peut effectuer le codage. Prenons la nomenclature la plus simple, celle des sexes. sexe masculin=1, féminin=2. Pour autant que les sexes soient clairement attribués, la traduction est immédiate, et, en cas de non-réponse à la question, le reste du questionnaire permet de trancher. Imaginons que l'ambivalence sexuelle se développe, d'une part, la nécessité de choisir entre 1 et 2 sera ressentie comme une violence et les non-

¹⁰ Cf. l'article de P. Champagne dans ce même ouvrage.

¹¹ La nomenclature est une **classification**. Le codage répartit les informations dans les catégories de la nomenclature. C'est une opération de **classement**.

réponses seront plus nombreuses, d'autre part la ré-interprétation sera problématique.

Ce cas est exagérément simple¹², mais il montre la nécessité de l'**unité de sens**, tout le long de la chaîne statistique, du concepteur à à l'utilisateur final, en passant par les enquêteurs, les informateurs, les codeurs. Il y a toujours un problème de traduction et d'interprétation, ne serait-ce que parce que la définition stricte que doit se donner le concepteur se heurte au sens commun, plus intuitif, ou, pire, à l'absence de sens commun.

Prenons la déclaration de profession qui est l'atome de base de la construction des catégories socio-professionnelles. Sans évoquer ici le problème général de ces déclarations¹³, notons qu'il est en principe soluble s'il y a, à la base, une profession réellement exercée et présente dans la nomenclature. Dans la période actuelle le flou porte à la fois sur les catégories et sur la notion d'activité professionnelle.

i) De nouveaux métiers apparaissent dans le secteur des services et en particulier dans les activités liées à l'information. D'une part, la nomenclature retarde toujours sur la réalité ¹⁴; d'autre part et surtout, les nouveaux métiers sont souvent à cheval entre plusieurs anciens métiers. Je pense en particulier aux nouveaux postes industriels qui correspondent davantage à une activité immatérielle qu'à une activité matérielle: quelle différence entre un employé et un ouvrier lorsqu'ils utilisent le même ordinateur? De même que deviennent les anciennes dichotomies (salariés/indépendants, privé/public etc.) lorsque des indépendants sont *de facto* dans un rapport quasi-salarial (dernier maillon d'une chaîne de sous-traitance), tandis que des salariés sont en fait quasi-indépendants (auto-salariés)? Quid encore lorsque le secteur public découvre les vertus du marché?

ii) L'activité professionnelle a un sens lorsqu'elle a un caractère habituel, prolongé, qu'elle façonne les mentalités et les

¹² La différence sexuelle au sens strict n'a guère d'intérêt en dehors de la démographie. En dehors de ce champ, elle est une commodité théorique, qui traduit de multiples différences de positionnement et d'attitudes. L'affaiblissement de l'indicateur objectif "sexe" oblige alors à déterminer de manière beaucoup plus précise les éléments d'analyse qui sont recherchés.

¹³ Cf. Desrosières Thévenot 1989). Plusieurs cas. Agent EDF, Michelin, ou au contraire appellations contrôlées (médecin, ingénieur).

¹⁴ Elle est prise entre l'exigence de se tenir à jour et celle de permettre la continuité chronologique. D'autre part, l'établissement et la mise en oeuvre d'une nouvelle nomenclature relève de l'industrie lourde, tant par le coût des investissements que par la durée du cycle de production (7/8 ans)

comportements. C'est ce que traduit le schéma simple: actifs **ou** inactifs. Lorsque l'exercice d'une profession perd son caractère statutaire avec la montée des situations intermédiaires, son pouvoir discriminant s'effrite. Un chômeur qui a déjà travaillé est classé dans son ancienne profession, quelles que soient ses chances de retrouver un emploi dans cette profession. Une personne qui tente de s'insérer dans l'emploi et va d'emploi précaire en emploi précaire change continuellement de profession. Certes, tous les emplois ne sont pas en libre service: beaucoup sont soumis à des conditions d'accès strictes. Mais parmi les autres, l'hétérogénéité et le brassage rendent douteux les classements.

Nomenclature ou nomenklatura ?

La validité de la nomenclature repose donc sur l'existence d'une armature institutionnelle et linguistique fondée sur des pratiques habituelles. Cette condition n'est pas satisfaite en période d'incertitude, de réarrangement ou de désagrégation. Lorsque les catégories sont solides, elles leur contenu fait l'objet d'un consensus général.

Lorsqu'elles sont floues, on voit se développer des **interprétations** diverses tout au long de la chaîne de recueil de l'information. Les personnes chargées du classement ont des consignes moins précises, ce qui accroît les risques de biais systématiques (e.g aspiration par les catégories moyennes¹⁵) ou d'hétérogénéité des normes de classification.

On peut être tenté de multiplier les consignes de codage pour résorber de la même manière toutes les ambiguïtés, mais l'homogénéité ainsi produite est le résultat d'un appauvrissement du langage: la catégorie du statisticien ne recouvre plus qu'une petite partie du phénomène, l'arbitraire de la procédure est finalement plus préjudiciable que les flottements des codeurs.

En d'autres termes, on résout alors le problème de la multiplicité du sens en en imposant un. Cette démarche volontariste du statisticien ne saurait pallier l'absence de référent commun. Illustrons ce point.

Marier les concubins?

Longtemps ignoré par une nomenclature matrimoniale centrée sur le mariage (marié, célibataire, veuf, divorcé), le concubinage est entré dans les moeurs et, de là, dans les statistiques.

¹⁵ Peneff, « La fabrication statistique », *Sociologie du travail*, 3/1984.

Mais le concubinage est, par essence, une position non définie qu'on essaye de ramener à la norme (vivant maritalement, déclaration de concubinage). Selon qu'on envisage la démographie, les droits sociaux ou la sexualité, la notion est plus ou moins étendue (elle est encore absente de la fiscalité). En tentant de définir le concubinage, on fait apparaître une autre notion qui condense le flou qu'on avait éliminé: la **cohabitation**, avec ses variantes (cohabitation temporaire, cohabitation partielle, cohabitation alternative).

On devine, sur cet exemple, quels détails choquants devraient figurer dans le questionnaire, s'il voulait saisir toutes les modalités du phénomène. Et ensuite, que ferait le statisticien? Inventerait-il des socio-styles aux noms ésotériques ¹⁶? Et pour qui? La seule solution n'est-elle pas de marier les concubins?

Le dilemme du statisticien

Le statisticien ne peut normaliser que ce qui est déjà formalisé par des institutions (exemple type : les variables d'Etat) et/ou dans des représentations sociales. Les catégories socio-professionnelles françaises s'enracinaient dans une tradition du métier et du statut, ce qui explique leur fortune en France. La nomenclature d'activités/produits l'organisation professionnelle de l'industrie et reprend encore aujourd'hui les frontières des organisations de branche¹⁷. Je pourrais multiplier les exemples.

Notons en passant que cette immersion dans la réalité quotidienne explique pourquoi la neutralité du statisticien (absence de jugement personnel etc.) ne suffit pas à assurer celle des statistiques. Beaucoup de critiques faites au statisticien sont imprégnées de fétichisme statistique¹⁸ ; elles lui reprochent de fournir de fausses représentations parce qu'elles méconnaissent sa position exacte: le statisticien ne fait que **durcir** le regard toujours un peu vague que la société lance sur elle-même. Regarder un Pissarro ou un Monet à 50 cm ne les rendront pas plus nets, au contraire.

Les statistiques sont entrées dans une zone de turbulences. L'opération statistique proprement dite bénéficiera de moyens techniques de plus en plus puissants susceptibles d'augmenter leur

¹⁶ Socio-styles qui eux-mêmes présentent, ces dernières années, une fâcheuse tendance à proliférer. On ne compte pas moins d'une quinzaine de types de base !!

¹⁷ cf Volle *Histoire de la statistique industrielle*, Economica, 1982.

¹⁸ Selon heureuse expression de Volle, « Statistique fétichisée et statistique réelle », *Le mouvement social*, n°104, 1978.

potentiel de traitement et de diffusion, mais c'est en amont et en aval que croissent les difficultés.

Le flou de l'observé d'une part, la demande d'indicateurs hautement élaborés d'autre part, soumettent à de fortes tensions tout le processus d'élaboration de l'information, l'éloignent des formes institutionnelles stables, elles-mêmes de plus en plus rares.

Existe-t-il des remèdes? Les enquêtes traditionnelles ne suffisent plus et il faudra user d'autres méthodes d'investigation. Certaines sont déjà connues (ethnométhodologie, formalisation d'itinéraires etc.¹⁹) et d'autres sont à inventer. Compte-tenu de la lourdeur de ces méthodes et des problèmes de traitement qu'elles suscitent, leur extension est suspendue à l'élaboration de techniques de traitement automatique sophistiquées (analyse automatique de contenu, par exemple).

Il est cependant impossible que le progrès technique à lui seul résolve les problèmes de vision qui frappent aujourd'hui la société et dont témoignent maints autres indices (ainsi dans "Le Monde" du ** : de moins en moins de Français font la différence entre "la droite" et "la gauche"). La seule voie consiste à approfondir la conceptualisation, à rechercher comment le **sens** vient aux chiffres, à s'intéresser autant aux difficultés de chiffrage qu'à ses résultats, autant, et même plus, au flou qu'au dur.

Bref, c'est la redécouverte, avec des moyens supérieurs, de la problématique des pionniers du début du siècle. Cette évolution est à la fois difficile, coûteuse et incertaine: elle remet en cause "l'habitus professionnel du statisticien"²⁰, les appareils, les formations professionnelles et les hiérarchies correspondantes.

Depuis les premiers travaux de Volle (1970), une série de recherches et d'analyses²¹ ont exploré cette voie. Ce courant se heurte à la

¹⁹ L'ethnométhodologie repose sur l'observation directe d'un très grand nombre de faits et gestes "significatifs", dont on tirera ensuite une analyse. Non dépourvue de préjugés, cette méthode ne s'applique bien que dans des conditions particulières (unité de lieu).

L'étude des itinéraires établit son principe d'équivalence sur les cheminements. On voit alors qu'elle s'adapte bien à des situations mouvantes, à condition d'avoir établi un schéma théorique d'interprétation qui permette de les déchiffrer.

²⁰ Volle, 1982, « Ensemble de modes de vie et de modes de faire acquis dans l'institution statistique et perpétués par elle [...] l'habitus des statisticiens nous parut tellement inébranlable qu'il prit à nos yeux les dimensions d'un phénomène sociologique - propre à une profession, mais caractéristique aussi de l'insertion de cette profession dans la société ».

²¹ Sans pouvoir rendre justice à tous les auteurs, citons Desrosières et Thévenot sur les CSP, Thévenot et Boltanski sur les "investissements de forme", Salais sur la notion de chômage, les journées "Pour une Histoire de la Statistique" (19xx), les journées "Statistiques et sociologie" (19xx), les numéros du Courrier des Statistiques dirigé par JP Girard. Mentionnons aussi, à

représentation des statistiques façonnée par leur triomphe antérieur, en même temps qu'il contribue à la modifier en sapant le "fétichisme statistique". Le pire serait que cette mise à jour ne soit pas ressentie comme indispensable ou qu'elle soit bloquée par des structures institutionnelles ou organisationnelles inadaptées ²².

Les décisions tant publiques que privées, prises en fonction d'informations inadéquates, dériveraient alors insidieusement jusqu'à ce que brutalement le système d'informations éclate. Tout le système de décision serait alors atteint par le chaos.

1

l'extérieur de l'Institution statistique, les recherches de D. Merllié, et les nôtres (Centre Analyse Epistémologie Histoire, Université Lumière Lyon 2)

²² Desrosières 1989 laisse deviner une telle possibilité, « Les spécificités de la statistique publique en France », *Courrier des statistiques*, n° 49, 1989.