

CLASSER ET MESURER

Les deux faces de l'argument statistique

Alain DESROSIERES

Même s'il est vrai que les ressources permettant de construire et de communiquer la science sont toujours à la fois cognitives et sociales, l'argument statistique occupe, parmi ces ressources, une place bien particulière. Dès ses origines et tout au long de son histoire, il a été associé aux questions soulevées par les relations entre les hommes et par l'organisation de la société, à travers le calcul des probabilités et la statistique descriptive (1). L'idée de probabilité est apparue au XVII^e siècle pour penser les conditions d'équité de contrats portant sur des événements incertains (2). Coumet, 1970. Puis il a été formalisé, au XVIII^e siècle, en tant que raison de croire, ou degré de confiance qu'il convient d'accorder à un énoncé, scientifique ou juridique. « L'erreur à craindre », ou intervalle à l'extérieur duquel l'énoncé a très peu de chance d'être vrai, fournit une garantie permettant sa mise en circulation. La statistique, de son côté, a d'abord

été liée à la description des Etats, puis à l'analyse des régularités (Quetelet) et des diversités (Galton) des populations, humaines ou non humaines. Les formalismes probabilistes et statistiques, produits initialement pour penser des questions très variées (jeux de hasard, assurances, astronomie, justice pénale, santé publique, eugénisme, agriculture, psychologie, cycle économique...) sont ensuite épurés, codifiés et partiellement unifiés pour devenir des branches des mathématiques appliquées.

Les spécifications logiques et rhétoriques de ces formalismes, explicitées et diffusées à travers des enseignements *ad hoc*, constituent des références mais aussi des contraintes, tant cognitives que sociales, pour leur déploiement ultérieur. Cependant, ces contraintes ne sont pas absolues. La circulation des schèmes probabilistes et statistiques à travers les domaines d'usages les plus divers des activités humaines (qu'elles soient ou non qualifiées de « scientifiques ») n'a pu se faire qu'au prix d'un grand nombre de transformations, retraductions et détournements. De nombreux exemples en sont fournis par la littérature historique, avec les transports des schèmes probabilistes, de la moyenne, de la méthode des moindres carrés, de la corrélation, de la régression linéaire. Chacune de ces notions a été intégrée dans une succession de constructions rhétoriques partiellement communes et partiellement différentes. L'étude fine et systématique de ces transferts et retraductions des outils probabilistes et statistiques, en liaison avec des réseaux d'usages spécifiques, reste à faire. Elle ne serait pas fondamentalement différente des recherches de sociologie des sciences et des techniques, qui analysent les usages variés d'objets réputés scienti-

(1) Cette double dimension de la rationalité statistique a été beaucoup étudiée par les historiens des sciences depuis les années 70. Les ouvrages les plus notables sur ce sujet sont : DASTON, 1988 (sur les probabilités au XVIII^e siècle), GIGERENZER et alii, 1989 (livre collectif sur la diffusion des probabilités), HACKING, 1990 (philosophique et sociologique), PORTER, 1986 (histoire sociale et cognitive), STIGLER, 1986 (statistique mathématique). Une histoire de la statistique et de certains de ses usages économiques et administratifs est présentée par DÉSROSIERES, 1993. Les conditions techniques et sociales de l'émergence d'une statistique démographique sont décrites par DUPAQUIER, 1985, et BRIAN, 1994. La perspective retenue dans ce texte doit par ailleurs beaucoup à des conversations avec Michel ARMATTE, Laurent THÉVENOT et Elisabeth ZUCKER.

(2) COUMET, 1990.

(3) LATOUR, 1993, CONEIN, DODIER et THÉVENOT, 1993.

fiques ou industriels (3). Elle pourrait même constituer un élément essentiel d'un éventuel « Traité de STS » qui viserait à généraliser et à systématiser de telles recherches, dans la mesure où l'argument probabiliste et statistique joue un grand rôle dans le « dur métier des travailleurs de la preuve » (4).

De quelles façons ces techniques contribuent-elles à façonner les pratiques d'écriture et les modes de représentation et de communication des savoirs ? Dans quelle mesure leur enracinement ancien et profond dans l'histoire des façons de penser les relations entre les hommes conditionne-t-il leur structure cognitive et leur efficacité rhétorique ? A l'inverse, en quoi celles-ci s'imposent-elles éventuellement lors du transport de ces outils d'un usage à un autre ? Nous esquisserons ici des réponses à ces questions en montrant que le raisonnement probabiliste et statistique peut être vu comme le lieu de rencontre et d'interaction entre deux traditions de pratiques cognitives et sociales : la qualification et la quantification. La première, issue plutôt du droit, de la vie administrative, de la médecine, construit des classes d'équivalence, des catégories, et affecte des cas singuliers à celles-ci à travers des jugements, des décisions, des diagnostics ou des procédures de « codage social » (5). La seconde provient des techniques du nombre et de la mesure du temps, des positions des planètes et des étoiles, des terres, des populations, des soldats, des quantités et des prix des marchandises échangées, des impôts... Bien sûr ces deux traditions sont étroitement entremêlées : pour compter ou mesurer des objets, il faut les identifier et les qualifier au préalable, ou définir des étalons de mesure à partir de propriétés qualitatives. Il est cependant fécond de distinguer et de confronter ces deux pratiques de savoir et d'action, à la fois conceptuellement et dans leurs trajectoires historiques, précisément pour formuler des pistes de réponse aux questions

initiales. Quelles restrictions les structures des formalismes font-elles peser sur leurs usages ? Comment les usagers jouent-ils de ces contraintes pour détourner et transformer, parfois complètement, ces formalismes ? Comment la position charnière des probabilités et des statistiques entre les sciences dites respectivement « de l'homme » et « de la nature », qui est liée à leur vocation à « maîtriser le hasard » (Hacking), contribue-t-elle à assurer une voie particulièrement fréquentée de circulation rhétorique (dans les deux sens) entre ces deux types de science ? L'exercice qui suit vise à pointer un certain nombre d'objets, de machines, de techniques graphiques, de métaphores, de formes standard (dont la reine, ici, est sans doute la « courbe de Gauss » ou « loi normale »), d'algorithmes de calculs, de modes de démonstration, en les mettant en relation avec les deux procédures de « mise en classe » et de « mise en mesure ».

Les contraintes de la feuille de papier : tableaux, graphiques, cartes.

L'origine de la statistique est en général rattachée, d'une part, à l'État, en tant que cadre de description plus ou moins standardisée (et non nécessairement quantitative) des royaumes et duchés allemands constituant le Saint Empire (6), et d'autre part, à l'état, en tant que liste, alignant des personnes ou des choses, (dont l'expression « état civil » porte la trace). La statistique allemande de Conring et Achenwall est d'abord une nomenclature, un cadre de classification pour présenter et faire circuler une description systématique des lois, des coutumes, des religions, du climat, de la géographie, des activités économiques d'un État. Par ailleurs, la construction d'un état, d'une liste énumérant des occurrences, est typiquement celle d'une classe d'équivalence : les *items* qui y figurent sont rapprochés, mis en perspective entre eux, et ainsi rendus équivalents d'un certain point de vue : individus sujets de

(4) Un bon exemple de cette démarche est proposé par GIGERENZER (1991), qui montre comment les procédures d'estimation et de tests issues de la statistique inférentielle se sont imposées en psychologie expérimentale.

(5) THÉVENOT, 1983.

(6) LAZARSELD, 1970.

même roi, baptisés mariés ou décédés dans la même paroisse (7), soldats obéissant au même général, familles habitant la même ville, objets appartenant au même propriétaire, pays concluant une alliance ou un traité. Tous ces « mêmes » constituent la classe d'équivalence et le point de vue, politique et cognitif, à partir duquel elle est dressée.

Mais, en même temps que naissent ces techniques de description standard des États et de constructions d'états à des fins d'administration, naissent aussi des procédures de comptage des baptêmes, des mariages et des décès, puis de la population dans son ensemble. L'arithmétique politique anglaise de Graunt et Petty offre des mesures de la population et des événements affectant les corps des personnes, ainsi traitées, au moins dans ces comptages, comme interchangeable et additionnables. La singularité de l'individu, qui subsistait dans la liste à travers son non propre, disparaît dans l'acte fondateur d'addition, et de fusion dans un tout, repérable par un seul nombre. Les controverses futures sur la statistique sont déjà inscrites dans cette réduction, caractéristique de l'enchaînement de deux opérations distinctes : mise en équivalence dans une liste, puis comptage et addition.

Une liste (par exemple une généalogie) peut être mémorisée et récitée oralement (cas des griots africains). Les nombres peuvent être manipulés et combinés par des techniques du corps parfois très complexes, ou par des bouliers. Le commerce direct du corps avec les classes et avec les nombres peut se faire de multiples façons (8), mais l'usage de l'écriture (9), puis de l'imprimerie (10), transforme radicalement les possibilités de combinaison et de conservation. La feuille de papier écrite ou imprimée soulage le corps et la mémoire. La liste, embrassée instantanément du regard, peut dès lors exercer son effet essentiel : traiter

chacun de ses *item* sur le même plan (au sens propre et au sens figuré), sur un pied d'égalité, au lieu de voir l'un du point de vue de l'autre. Cette « *vue d'ailleurs* », ou « *vue de nulle part* » est le trait essentiel de l'effet de liste, de tableau, de graphique ou de carte. La feuille de papier écrite offre le support matériel et la possibilité pratique de déployer les capacités logiques et cognitives de la notion de « *classe d'équivalence* », en créant une position d'extériorité par rapport à l'ensemble envisagé. Cette capacité est encore accrue et radicalement transformée par l'imprimerie, qui multiplie et transportant au loin des listes, les tableaux et les cartes, permet de les combiner, plus tard et ailleurs, et de créer d'autres listes, d'autres tableaux, d'autres cartes.

Mais la feuille de papier a encore une autre capacité, essentielle pour le travail statistique : son espace comporte deux dimensions. Le tableau croisé (dont le plus simple est celui à deux lignes et deux colonnes) permet d'enregistrer, de formaliser et d'exprimer des rapports de rapports, des comparaisons de comparaisons. De plus, la symétrie formelle des dispositions en lignes et en colonnes suggère la possibilité de les permuter en entrant dans le tableau dans un sens ou dans un autre. Ceci ouvre la porte à l'idée si féconde de dualité. Pourtant, cette mise en forme cognitive et les possibilités rhétoriques de « mise sur le même plan » qu'elle offre, heurtent d'autres représentations logiques et sociales. La statistique camérale allemande du XVIII^e siècle était centrée sur la description d'un Etat, à l'usage de son prince. Bien que le plan standard proposé par les « statisticiens » de Göttingen offre la possibilité d'une « mise en tableau » juxtaposant et comparant les descriptions, quantitatives ou non, des divers Etats, cette idée ne va pas de soi. Les « faiseurs de tableaux » sont violemment critiqués,

(7) LE MÉE, 1975 ; ZUCKER-ROUVILLOIS, 1994.

(8) LAVE, 1988.

(9) GODDY, 1979.

(10) EISENSTEIN, 1991.

et leur activité est accusée d'être « réductrice ».

Comment interpréter cette critique dans le cadre de la distinction entre qualification et quantification ? Porte-t-elle sur la « mise en tableau » et sur le fait de comparer l'incomparable, ou sur la « réduction » de la description à des nombres ? En apparence, c'est le nombre qui est attaqué : « *Les machinations auxquelles se livrent ces criminels de statisticiens-politiques dans leurs efforts pour exprimer tout par des chiffres sont méprisables et ridicules au-delà de toute expression (11)* ». Mais ce ne sont pas l'exactitude et la précision des mesures qui sont dénoncées, mais l'idée même de décrire un pays par des nombres. Le débat est inscrit dans l'espace cognitif de la qualification, et non dans celui de la quantification. La forme tabulaire, croisant plusieurs logiques de codage et de classification, que ce soit sous forme verbale ou numérique, multiplie les effets rhétoriques possibles de la « mise en catégorie », et ceux-ci ne sont pas de même nature que ceux de la « mise en nombre ». Cependant, c'est bien la structure même de la forme tabulaire qui incite à chercher et à comparer des nombres. Elle crée littéralement l'espace d'équivalence qui appelle la statistique quantitative.

D'autres techniques graphiques sont adaptées à la qualification des objets plutôt qu'à leur quantification. Elles reposent sur les propriétés topologiques (et non métriques) de l'espace à deux dimensions : représentation de réseaux par des graphes combinant des points et des traits, orientés ou non. L'un d'entre ces graphes est l'arbre, figurant une structure de classification emboîtée pyramidale, ou une généalogie, descendante ou ascendante. Ces représentations en réseaux, en graphes ou en arbres sont souvent présentées comme des alternatives aux descriptions reposant sur des totalisations statistiques, en permettant de décrire des relations, des connexions entre personnes ou entre objets, en amont

de la catégorisation. La figure très générale de la critique de la mise en équivalence, du type : « *Cette classe rassemble des choses en réalité différentes* », vise à dénoncer une hétérogénéité, quantitative ou, plus radicalement, qualitative. Elle trouve dans ces représentations en réseaux et en graphes des supports pour d'autres types de généralisation que ceux qui s'appuient sur des catégories agrégeant des choses supposées équivalentes (12). Ces graphiques, fondés sur les propriétés topologiques de la feuille de papier, sont en cela différents des graphiques statistiques développés depuis la fin du XVIII^e siècle, qui combinent des propriétés métriques : longueurs, largeurs, surfaces, pentes, angles.

Comment construire un réseau de sémaphore statistique

L'anglais William Playfair (1759-1823) est souvent cité comme l'inventeur de graphiques statistiques mettant en œuvre une large palette de ces propriétés métriques. Son nom ressort sans doute parce que non seulement il a présenté et commenté beaucoup de semblables graphiques dont certains assez compliqués et difficiles à dire, mais que, surtout il en a fait, simultanément, la théorie et la propagande. Analogie en cela à d'autres « grands noms » de l'histoire de la statistique, comme Quetelet, Galton ou Sauvy, il est remarquable pour avoir eu l'idée de transporter, de traduire, de combiner des schèmes variés venus d'ailleurs, puis de commenter et de populariser de façon opiniâtre quelques thèmes, en associant inlassablement des dispositifs visuels et verbaux avec une grande intelligence de l'efficacité technique et sociale de ces montagnes (13). Un grand nombre des idées de la « sémiologie graphique », développées par exemple par le Français Bertin, sont déjà explicitement formulées dans les écrits de Playfair. Le graphique permet de rapprocher et de

(11) Auteur anonyme de 1807, cité par LAZARFELD, 1970.

(12) GRIBAUDI et BLUM, 1990.

(13) COSTIGAN-EAVES et MACDONALD-ROSS, 1990.

mettre en relation, sous le regard, des grandeurs séparées tant par les conditions et les lieux de leurs enregistrements que par leurs natures. Un des graphiques les plus célèbres de Playfair, (figure 1) publié en 1801, associe les surfaces (figurées par des cercles) de divers pays, leurs populations et leurs revenus nationaux (figurés par des segments verticaux parallèles et tangents aux cercles) et leurs revenus par tête (représentés par la pente de la droite reliant les sommets des deux segments précédents). La graphique favorise ensuite la conservation et la réutilisation ultérieure de la mémoire d'une forme, et donc de nouvelles mises en perspective de formes *a priori* sans rapports. Par toutes ces propriétés, le graphique s'oppose trait pour trait au tableau de chiffres, difficile à regarder et à mémoriser. Playfair développe systématiquement ces arguments de psychologie cognitive en les inscrivant dans une perspective d'usage et d'action. La méthode graphique simplifie et rend attrayantes des statistiques ennuyeuses et complexes (14). Elle permet...

«... à des hommes de rang élevé ou actifs en affaires de porter leur attention sur des configurations générales..., sans la fatigue et l'ennui de l'étude des cas particuliers dont elles sont composées... » (Playfair, 1786).

Dans tous les cas, les tableaux de chiffres sont moins efficaces que les graphiques :

« Celui qui a soigneusement examiné un tableau imprimé se rend compte qu'il n'a conservé qu'une idée faible et partielle de ce qu'il a lu, et que, comme une figure tracée sur le sable, elle est vite complètement gommée et effacée... » (1786).

« Qui regarde un graphique obtient en cinq minutes ce qui aurait exigé des jours entiers à imprimer de façon durable dans la mémoire, avec un tableau de chiffres... »

« Le graphique fournit une idée simple, précise et permanente, en donnant une forme et un contour à un grand nombre

d'idées séparées, qui seraient, sans cela, abstraites et sans liens... »

« Les avantages de cette méthode sont de faciliter l'acquisition de l'information, et d'aider la mémoire à la retenir : ces deux points forment la grande affaire de ce que nous appelons apprendre... »

« Tout ce qui se présente vite et clairement à l'esprit met celui-ci en état de travailler, de raisonner, de penser, alors que si l'esprit enregistre beaucoup de faits séparés, il est purement passif et ne peut retenir un tel savoir... » (1801).

Ces remarques sur le fonctionnement de l'esprit et de la mémoire étayés et stimulés par les représentations graphiques peuvent être reliées à des développements contemporains des techniques de production et de circulations des faits scientifiques, auxquels Playfair a été associé d'une manière ou d'une autre. Adolescent, son frère aîné lui apprend à suivre la météorologie en enregistrant jour après jour une courbe des températures. Puis, jeune employé d'une entreprise dirigée par James Watt (l'inventeur de la machine à vapeur), il voit celui-ci construire des graphiques associant la pression et le volume d'une quantité gazeuse et leur produit représenté par une surface. Plus tard, l'idée que des faits scientifiques peuvent être construits et communiqués par le biais de surface, sera présente dans beaucoup de ses graphiques. Par exemple, des pays aux formes peu lisibles sont remplacés par des carrés ou des cercles.

En 1793, alors que Playfair voyage en Allemagne, un émigré français lui explique le nouveau système du télégraphe Chappe et de son alphabet transmis sur des centaines de kilomètres par le sémaphore. Il cherche aussitôt à l'introduire en Angleterre. Cette technique est comme un résumé de toute son activité : coder une information sous une forme visuelle aisément identifiable et transmissible au long d'un réseau *ad hoc*, franchissant montagnes et vallées. Sa démarche s'inspire aussi de celle des géographes et carto-

(14) Les citations et les graphiques de PLAYFAIR qui suivent sont issus des articles de ROYSTON (1956) et de COSTIGAN-EAVES et MACDONALD-ROSS (1990).

graphes qui confectionnent et font circuler des atlas. Dans ceux-ci, les cartes des divers pays sont réunies et plus ou moins standardisées. Elles associent, sur une même page, des représentations conventionnelles de l'espace, des légendes et des commentaires (15). De même, Playfair publie trois éditions d'un *Atlas commercial et politique*, en 1786, 1787 et 1801. Celles-ci contiennent deux grandes formes de graphiques. Les uns sont synchroniques ; ce sont par exemple des comparaisons entre pays (figure 1). Les conventions graphiques y sont des surfaces, des cercles, des pentes de segments de droite. Les autres sont des séries chronologiques inscrivant le temps sur un axe horizontal, une convention déjà utilisée en météorologie. De même que les cartes de l'époque, ces graphiques sont parsemés d'inscriptions verbales et d'explications. Les formes retenues visent à faire passer une idée ou à soulever une question : la balance des échanges commerciaux entre l'Angleterre et l'Allemagne (figure 2), l'évolution comparée des salaires et des prix du blé (figure 3). Ces graphiques sont conçus pour étayer les décisions des hommes d'affaires, ou pour apporter des arguments dans les débats entre les « hommes de rang élevé ». Ceux-ci doivent pouvoir prendre appui sur des choses solides et communément admises : le graphique statistique vise à produire du sens commun indiscutable.

Au moment où Playfair imagine un réseau socio-technique qui, à l'image de celui du télégraphe par sémaphore, comporte à la fois un système de codification des messages et un support matériel pour faire circuler ceux-ci, la production statistique susceptible d'être ainsi transportée est encore très réduite. Elle ne commencera à se développer qu'à partir des années 1830, en particulier avec l'activité militante de l'astronome belge Quetelet (1796-1874), qui organise et promeut tout à la fois l'enregistrement et la compilation systématiques de statistiques administratives (notamment les recensements), l'usage de certains

schèmes probabilistes (la loi de Gauss), la construction de réseaux internationaux de statistique, et enfin l'idée que les sociétés humaines sont soumises à des lois révélées par des régularités statistiques. C'est cet ensemble de procédures administratives, de formules mathématiques, d'administrations savantes et de philosophies sociales qui pourra, longtemps après que Playfair soit mort, constituer le réseau du sémaphore statistique capable de faire tenir des choses, et des gens par l'intermédiaire de ces choses.

Mais que transporte le sémaphore?

William Playfair a fait la promotion de plusieurs technologies qui préfigurent les autoroutes de l'information. Ce sont, d'une part, le réseau du télégraphe optique et, d'autre part, la mise en forme graphique des statistiques permettant de les transporter et de les utiliser aisément. Le rapprochement entre ces deux systèmes techniques est cohérent avec des arguments importants de la sociologie moderne des sciences. Celle-ci met l'accent sur le transport au loin de « mobiles immuables » dûment codifiés, rassemblés et combinés dans des « centres de calculs », puis redistribués par de longues chaînes d'usages. Ces images, souvent employées par Bruno Latour, sont parlantes, et rendent bien compte des conditions de réussite de projets futuristes comme ceux de Playfair, de Quetelet, ou, plus tard, de Karl Pearson, ou de Français comme Alfred Sauvy, René Carmille ou Claude Grusson. Mais pourquoi les réseaux sociaux, techniques et cognitifs proposés par chacun d'entre eux ont-ils ou non « tenu » ? Pendant combien de temps ? Pourquoi se défont-ils souvent, après une période de succès ? Comment analyser et décrire les liens tissés par les fabricants du réseau ? Comment intégrer à l'analyse la nature et les contraintes spécifiques au flux des objets qui circulent à travers les mailles de celui-ci ?

(15) JACOB, 1992

La sociologie moderne des sciences revendique un programme très ambitieux. Elle affirme prendre complètement au sérieux les arguments des « acteurs eux-mêmes », sans considérer *a priori* que leurs constructions rhétoriques sont extérieures à son champ, ou, pire encore, sont des illusions, interprétables en termes macrosociaux ou culturels. Dans cette perspective, une sociologie de la statistique ne peut être ni une sociologie des statisticiens (décrits par exemple comme des instruments de la gestion des frontières du monde social par l'Etat), ni même une sociologie du réseau des « centres de calcul », indépendamment du contenu des messages que ceux-ci échangent entre eux. L'abolition du « grand partage » est à ce prix. L'examen de la façon dont les deux opérations de « mise en classe » et de « mise en nombre » ont été, selon leurs usages, transmises, transformées, combinées ou opposées, au fil de la construction d'un « style de raisonnement statistique » (16) va dans ce sens. Qu'impliquent les contraintes proprement classificatoires et métrologiques, pour les usages sociaux de cette forme de la raison scientifique et politique ? Comment ces formalismes sont ils retraduits en circulant de main en main ? Ces questions sont analogues à celles que soulève une sociologie des objets. Quelles contraintes ceux-ci imposent ils à leurs usages ? En quoi l'objet est il à la fois le même et un autre selon les actions qui prennent appui sur lui ? (17).

Deux modèles issus de la science du XVIII^e siècle seront évoqués ici, celui de l'urne de Bernoulli et celui des erreurs de mesure. Le premier, cœur de la construction probabiliste, fournit un exemple typique de convention d'équivalence. Le second est le produit d'un épisode clé du développement de la métrologie scientifique. Les histoires de ces deux modèles

s'entrelacent et se séparent, au gré des associations inattendues auxquelles les soumettent des projets scientifiques et politiques toujours renouvelés. Elles peuvent être racontées en suivant les deux parcours verticaux plus ou moins entrecroisés des formalismes logiques et mathématiques s'engendrant les uns les autres selon les méthodes de l'histoire internaliste des sciences, en y incluant éventuellement quelques belles controverses théoriques. Une telle manière de faire est souvent critiquée par les courants plus « sociologiques » de l'histoire des sciences. Pourtant les recherches érudites des historiens internalistes sont incontournables, dès lors qu'est pris au sérieux le programme dit « fort » de sociologie des sciences : les apports de l'histoire interne font complètement partie des matériaux de cette sociologie. Mais, dans celle-ci, les fils diachroniques de l'histoire des théorèmes, des schèmes cognitifs et des outils techniques sont aussi regardés selon des coupes synchroniques. Celles-ci décrivent des montages complexes, dans lesquels ces schèmes et ces outils sont des personnages de théâtres changeants, fourmillant d'autres acteurs : Pearson, la corrélation et les inaptes héréditaires (18). Pasteur, les microbes et les hygiénistes (19), les pêcheurs de Saint-Brieuc et leurs coquilles Saint-Jacques (20).

L'idée même de probabilité, comme mesure d'un degré d'incertitude, provient du souci de fixer et d'unifier des conventions d'équité et d'équivalence. Equité entre contractants, lorsque des contrats impliquent des événements futurs et aléatoires, ou entre joueurs décidant d'interrompre leur jeu et souhaitant partager équitablement la mise (questions du Chevalier de Méré à Pascal). Equivalence entre espérances, combinant la probabilité d'un événement incertain et son utilité en cas d'issue favorable (21). Cette construc-

(16) HACKING, 1991.

(17) Ces questions sont discutées notamment dans : CONEIN, DODIER ET THÉVENOT (éds), *les Objets dans l'action. De la maison au laboratoire, Raisons pratiques*, n° 4, 1993.

(18) ARMATTE, 1994.

(19) LATOUR, 1984.

(20) CALLON et LAW, 1989.

(21) COUMET, 1970.

tion permet de comparer, de classer et de penser dans un cadre commun des faits de nature très différente : retour d'une expédition maritime, issue d'une partie de dés ou d'un tirage à « croix ou pile » par lancement d'une pièce de monnaie, tirage de boules dans une urne de composition inconnue, rapport des nombres de naissances masculines et féminines, réussite de l'inoculation de la variole, possibilité d'une erreur judiciaire envoyant un innocent à la potence. Dès son origine, cette idée de probabilité se présente sous deux faces (22). Hacking, 1975. L'une, subjective, ou épistémique, caractérise un état de l'esprit, une « *raison de croire* », degré de confiance assigné à un énoncé. L'autre, objective, ontique, ou encore fréquentiste, a trait à l'état du monde, à sa diversité et à la régularité tendancielle de certains événements : le soleil se lève tous les matins, le rapport des naissances de garçons et de filles est toujours proche de 1,05.

Les savants du XVIII^e siècle unifient ces diverses questions et ces deux conceptions autour du modèle de l'urne et de la première expression de la « loi des grands nombres » @MA23. (23), puis de la formulation analytique de la loi de distribution (dite plus tard « *gaussienne* » ou « *normale* ») spécifiant les probabilités des écarts par rapport à la tendance indiquée par la loi des grands nombres (24). De Moivre, 1738. Ce modèle de l'urne de Bernoulli est, fondamentalement, une convention fondant une classe d'équivalence entre événements assimilés à des tirages répétés d'une boule noire ou blanche dans une urne de composition constante. Cette formulation porte en elle à la fois l'efficacité rhétorique (l'idée de cause constante) et les contraintes spécifiques à ses usages ultérieures. Cette efficacité et ces limites sont reflétées par les controverses très variées suscitées par la mise en œuvre du modèle de l'urne dans les constructions successives et différentes de Bayes, La-

place, Quetelet, Poisson, Cournot ou Galton (25) que l'on n'évoquera ici qu'en tant que moments de débats portant sur la raison classificatoire.

L'urne de Bernoulli est efficace en ce qu'elle fournit un modèle pour traiter la question générale des conséquences d'une cause constante, éventuellement inconnue, mais dont les effets ne sont pas mécaniquement déterminés et prévisibles : résultat du lancer d'une pièce ou d'un dé, couleur d'une boule tirée de l'urne, sexe d'un nouveau né, mariage ou décès d'un individu dans l'année qui suit. Mais les contraintes du modèle sont liées à l'unicité de l'urne (c'est-à-dire de la cause constante) ou à l'irréversibilité et à l'asymétrie du temps, comme le montre le fait que Bayes ait dû, pour penser la symétrie des probabilités conditionnelles, inventer une autre machine probabiliste, la table de

Les métamorphoses de la causalité

billard. La question posée et formulée avec précision par le Révérend Bayes (1764) est essentielle dans la perspective d'une sociologie de la production et de la circulation des faits scientifiques. Cette question est inscrite dans la conception subjective de la probabilité, comme mesure d'un degré de certitude. Elle porte sur la façon dont l'occurrence d'un événement produit par l'action d'une cause constante mais inconnue transforme l'idée (subjective) que l'on peut avoir *a priori* de cette cause. Bayes renverse la question traitée par Bernoulli (1713) et De Moivre (1738). Ces derniers décrivent les effets (en termes de lois des grands nombres et de « *loi normale* ») d'une cause constante et connue. Bayes, en revanche, cherche à remonter des conséquences (connues) aux causes (inconnues ou mal connues), ou aux lois générales ayant produit ces effets. Il formule la probabilité d'une cause inconnue (par exemple la vraisemblance d'une hypothèse

(22) HACKING, 1975.

(23) BERNOULLI, 1713.

(24) DE MOIVRE, 1738.

(25) STIGLER, 1986 ; DESROSIERES 1993.

scientifique) en fonction, d'une part, de celle d'un de ses effets et, d'autre part, d'une probabilité de cette même cause postulée à priori avant de connaître cet effet.

L'intérêt de cette formulation est de faire éclater la contrainte spécifique au modèle de l'urne et de son idée de cause constante, en lui substituant un modèle plus complexe de succession de deux événements aléatoires, dont l'un est appelé cause et l'autre conséquence, puisque la « causalité » apparaît, dans l'axiomatique de Bayes, comme réduite à cette enchaînement temporel (26). Pour conserver le modèle de l'urne, il aurait fallu imaginer un tirage à deux degrés, tirage d'une (petite) urne dans une (plus grande) « urne d'urnes » (probabilité de la cause), puis tirage d'une boule dans cette petite urne déjà tirée. Bayes remplace cette machine bien compliquée par une table de billard sur laquelle sont lancées deux boules A et B. La question posée est celle de la position relative (à droite ou à gauche) de B par rapport à A, compte tenu de la position de A (assimilée à une « cause »). L'intérêt de ce montage est de permettre de penser facilement l'inversion de la « cause » et de la « conséquence », ce que le modèle de l'urne unique ne permet pas facilement. Ces deux exemples de l'urne et de la table de billard illustrent les idées « *d'efficacité rhétorique* » et de « *contraintes spécifiques* » d'un dispositif cognitif et matériel. De façon plus générale, ces machines (et celles dont il sera question ci-dessous) instrumentent de façon chaque fois différente l'idée de causalité, comme le montrent les exemples de la composition de l'urne (Bernoulli), de la position relative des deux boules de billard (Bayes), de la « vraie » position d'une étoile (Laplace), du partage entre « *causes constantes* » et « *causes variables* » (Quetelet), de la suppression (provisoire) de la notion, jugée métaphysique, de causalité, au profit de la corrélation statistique (Kearl Pearson), avant sa résurrection dans l'analyse écono-

métrique des séries temporelles (Granger).

La conception bayésienne (subjectiviste) de probabilité des causes est reprise par Laplace dans son traitement des questions d'erreurs de mesure en astronomie, qui vise à optimiser les évaluations de « vraies grandeurs », assimilées à des causes constantes mais inconnues. Puis le bayésianisme est ensuite rejeté avec mépris par toute la science du XIX^e siècle, pour qui l'idée d'une probabilité (subjective) *a priori* est une pure rêverie. La conception objective et fréquentiste des probabilités, qui l'emporte alors, prend appui sur les grosses machines de collecte statistique édifiées, à partir des années 1830, notamment sous l'influence de Quetelet. La régularité des fréquences observées par ces centres de calcul permet à Quetelet de transférer le modèle de l'urne vers les « *sciences morales* ». Le mariage, le suicide ou le décès d'un individu sont assimilés à des tirages aléatoires dans des urnes constantes dont la composition est révélée par les statistiques de l'état-civil. L'ampleur de ces statistiques garantit l'efficacité de la « loi des grands nombres », bien mieux, selon les détracteurs de Bayes, que ne le ferait tout exercice d'introspection d'une probabilité *a priori*.

Mais la démarche bayésienne resurgit, à partir des années 30, quand les sciences de l'esprit et de la décision cherchent à modéliser des processus de choix en situation d'incertitude. Les formalisations de l'intelligence artificielle et des systèmes experts y font largement référence. En effet, l'affectation d'un cas singulier de nature incertaine à une classe d'équivalence peut être vue comme une procédure bayésienne : les informations incomplètes disponibles sont assimilées aux occurrences passées et connues traitées par celle-ci. Une décision judiciaire, un diagnostic médical (27) ou même le codage d'un questionnaire statistique dans lequel certaines réponses sont floues peuvent être modélisés à l'aide de cet outil formel et linguistique proposé par Bayes, combattu par Cournot et par

(26) CLERO, 1988

(27) FAGOT-LARGEAULT, 1989,

presque tous les statisticiens du XIX^e siècle, puis réhabilité par un courant important (mais minoritaire) de ceux du XX^e siècle. L'argument bayésien est une modalité, non acceptée par tous, de l'argument statistique. Il joue, à ce titre, un rôle important dans la procédure et la circulation d'énoncés, dans les sciences de la décision, et plus généralement, dans tous les domaines où l'affectation, même incertaine, d'un cas à une classe est une obligation qui ne peut être différée.

Une ou plusieurs urnes : les contraintes du modèle de Bernoulli

Tout au long du XIX^e siècle, des controverses ont lieu à propos des conditions de validité et d'application du modèle de l'urne de composition constante et de la façon dont Quetelet s'appuie sur lui pour fonder une « physique sociale » régie par des lois aussi sûres que celles de l'astronomie et de la physique. La première de ces controverses porte sur la « *méthode du multiplicateur* », inventée par Laplace en 1780, pour mesurer la population française de façon indirecte, en multipliant le nombre national des naissances par un « multiplicateur ». Celui-ci est calculé seulement pour quelques paroisses tirées au sort, en divisant, pour chacune d'elles, leur population par le nombre de leurs naissances. Dans la publication résultant de cet ancêtre direct de la méthode des sondages, réinventée au XX^e siècle, Laplace associe à l'évaluation elle-même une « *erreur à craindre* » estimée à l'aide du calcul des probabilités, par un raisonnement, issu du modèle de l'urne (28). Le « *multiplicateur des naissances* » est supposé le même pour toute la France, assimilée ainsi à une urne unique. La critique contre cette méthode, exprimée en 1827 par Keverberg, un haut fonctionnaire belge, porte précisément sur cette hypothèse. Ce calcul « *approximatif* » repose, selon lui, sur une hypothèse indé-

fendable. Le taux de natalité (inverse du « multiplicateur ») n'a aucune raison d'être le même pour tout le pays, puisque...

« *La loi réglant la natalité ou la mortalité est composée d'un grand nombre d'éléments : elle est différente pour les villes et pour les plats pays, pour les grandes et opulentes cités et pour les villages plus petits et moins riches, ou selon que les populations sont denses ou éparses. Cette loi dépend du terrain (en altitude ou non), du sol (sec ou marécageux), de la distance à la mer, de l'aisance ou de la misère de la population, de son alimentation, de son vêtement et de sa manière générale de vivre, et d'une multitude de circonstances locales qu'omettrait toute énumération a priori. Il est donc difficile de déterminer par avance avec précision et en se fondant sur une connaissance incomplète et spéculative, la combinaison de ces éléments qui agit réellement. Ceux-ci semblent présenter une variété infinie quant à leur nature, leur nombre, leur intensité et leurs proportions relatives... Selon moi, il y a un seul moyen d'obtenir une connaissance correcte de la population et de ses éléments, c'est d'effectuer un recensement réel et complet, et de dresser un registre des noms de tous les habitants, avec leurs âges et leurs professions.* » (29). Ce texte énumère une liste complexe de critères de découpages et de classements, qui constituent autant de raisons d'hétérogénéité du territoire et de sa population, interdisant la référence au modèle de l'urne unique. Par ailleurs, la légitimité du travail des statisticiens n'est pas encore assez assurée pour que ceux-ci puissent s'écarter des méthodes purement administratives, le « recensement réel et complet », et le « registre des noms de tous les habitants ». Il faudra attendre plus d'un siècle pour que cette objection sociotechnique, complètement acceptée par Quetelet et ses émules, puisse être levée. Elle le sera grâce, d'une part, aux formulations des sondages aléatoires par Neyman (1934), et, d'autre part,

(28) BRU, 1988.

(29) KEVERBERG, 1827, cité par STIGLER, 1986.

(30) BLONDIAUX, 1994.

à leur promotion, assurée d'abord par les statisticiens officiels américains, pour la mesure du chômage, puis par Gallup (aux USA) et Stoetzel (en France), pour les sondages d'opinion et les études de marché, à partir des années 1930 (30).

L'histoire de la rhétorique statistique du XIX^e siècle peut être lue comme celle d'une série d'efforts pour lever la contrainte de l'urne unique, et instrumenter l'argument de la cause variable, c'est-à-dire des régularités des effets de la diversité. Le travail taxinomique (construction des découpages et description de leurs effets) occupe une place centrale dans ces efforts. Ainsi Poisson (1837), analysant la probabilité des erreurs judiciaires, selon la majorité requise pour les décisions des jurys d'assises, formule une « *loi forte des grands nombres* », qui démontre que la fréquence observée dans une série de tirages converge même si on ne suppose pas que la loi de probabilité du tirage est constante, mais qu'en revanche, elle varie selon une loi constante (31). Puis Lexis, retournant l'argument de Quetelet, montre que la régularité approximative d'une série temporelle ne prouve pas que les tirages soient issus d'une urne unique (et donc d'une loi constante). Il invente même un test pour défaire l'argument roi de Quetelet sur les « *causes constantes* ». Seule la régularité du sex-ratio échappe à l'épreuve implacable du test de Lexis (32).

Cournot (1843) travaille spécialement la place de l'argument probabiliste et statistique dans la construction et l'étayage de la connaissance scientifique. Il est notamment le premier à distinguer clairement...

«...les probabilités qui ont une existence objective, qui donnent la mesure de la possibilité des choses, et les probabilités subjectives, relatives en partie à nos connaissances, en partie à notre ignorance, variables d'une intelligence à une autre, selon leurs capacités et les données qui

leur sont fournies... » (Cournot, 1843).

Discutant le statut d'une preuve appuyée sur un découpage de la population, il aperçoit clairement le fait que les choix de celui-ci (« *l'orientation de la coupe* ») peut conduire à sélectionner les arguments favorables à la démonstration d'une thèse, au détriment des arguments défavorables. Il est alors proche d'une sociologie de la rhétorique statistique, dans la mesure où il ne verrouille pas complètement les normes de la bonne démonstration scientifique, comme le feront les épistémologues qui suivront. Le statut du raisonnement probabiliste est encore suffisamment ouvert pour que Cournot puisse montrer la diversité de ses usages.

L'instrumentation statistique de la variabilité des effets d'un ensemble de causes variables sera faite, à partir des années 1880, par Galton, Edgeworth, Pearson et Yule, à travers les méthodes de la corrélation, de la régression et des moindres carrés. Les idées de « probabilité conditionnelle » et de « loi de probabilité à plusieurs dimensions » permettent de lever les contraintes rhétoriques du modèle de l'urne unique, tout en conservant cependant l'essentiel de l'argument probabiliste issu des formulations de Bernoulli, Laplace, Gauss et Quetelet. Un espace rhétorique nouveau est ainsi ouvert, dont découleront la statistique inférentielle (33), la psychométrie (34), l'économétrie (35). Mais cette modalité de l'argument statistique ne recouvre pourtant pas tout le champ des usages des statistiques mathématiques, comme le montrent des courants (par ailleurs très différents) comme la statistique bayésienne, ou l'analyse des données (36).

Rhétorique statistique et réalisme métrologique

Un spectre hante le travail du statisticien, celui de la réalité des objets dont il exhibe les mesures. La production et la

(31) HACKING, 1990.

(32) PORTER, 1986.

(33) GIGERENZER, 1991.

(34) GOULD, 1983.

(35) MORGAN, 1990.

(36) BENZECMI, 1982.

mise en circulation des résultats de ce travail ne peut s'affranchir d'une visée réaliste, inspirée de la rhétorique de la métrologie scientifique. Pourtant la mise au point de la boîte noire de l'instrument de mesure ne peut être qu'une succession d'inscriptions conventionnelles et de manipulations de classes d'équivalence, postulées pour les besoins du raisonnement et de l'action collective. Comment faire tenir ensemble ces deux contraintes largement contradictoires ? Comment l'argument statistique gère-t-il au jour le jour la tension entre une visée réaliste indépassable et les conventions d'inscription et de mise en classe ? Cette gestion est-elle fondamentalement différente dans la construction et la communication des faits scientifiques et dans le débat plus directement lié à l'action et à l'argumentation politique ?

Une gamme de stratégies rhétoriques composant ces deux exigences peut-être déployée entre les plus réalistes et les plus constructivistes. La première est celle du biais, de l'approximation, de l'ordre de grandeur, de l'à-peu-près (37). Directement inspirée de la métrologie, elle insiste cependant sur l'écart irréductible entre la mesure « *inévitablement approximative* » et la réalité inconnaissable. La maîtrise de cet écart, par la théorie probabiliste des erreurs, tient compte des spécificités de l'action et de l'argumentation. Une saine hygiène de l'à-peu-près est la marque d'une métrologie statistique consciente des limites des instruments de mesure et de leurs manipulateurs. Cependant cette rhétorique de l'à-peu-près, dans la ligne de la pratique de mesure des sciences de la nature (physique, astronomie), ne problématise pas la « réalité » et les conventions d'équivalence qui sous-tendent la définition et la mesure de celle-ci.

Une seconde stratégie de gestion de la tension entre visée réaliste et conventions de mesure s'appuie sur les idées complémentaires de variables latentes et d'indicateurs. L'écart entre la réalité et sa mesure ne résulte plus seulement de l'imperfection conjointe des instruments d'observations

et des observateurs humains, mais de l'idée que l'objet évoqué n'est pas par lui-même susceptible de mesure (le bonheur, la santé, la pauvreté). Mais il constitue néanmoins une réalité latente, cause de manifestations susceptibles d'inscriptions, dont la synthèse constitue des indicateurs, sociaux ou économiques. Cette rhétorique, assez proche de la cause constante (mais inconnue) de Quetelet, suppose une permanence et une consistance de la réalité latente. Par exemple, une convention de permanence dans le temps d'un objet intitulé « *population active* » permet à Marchand et Thélot (1991) de décrire « *Deux siècles de marché du travail* », à travers des indicateurs construits par une suite compliquée de convention de substitutions et de mesures, calculant, par exemple, un taux d'activité à partir d'un taux d'urbanisation. Dans ce cas, les auteurs utilisent l'expression « *ordre de grandeur* » pour justifier ces conventions, prenant ainsi implicitement appui sur la rhétorique métrologique précédente, bien que leur présentation de la nature de la réalité latente soit nettement différente de celle des sciences de la nature. Le compromis rhétorique entre la visée de réalité et la présentation des conventions de mesure ne postule plus que la réalité et sa mesure sont homogènes, et que leur écart est purement quantitatif. La réalité existe, mais elle est radicalement inconnaissable, et sa mesure est construite par des moyens plus ou moins détournés. L'idée d'ordre de grandeur est dès lors ambiguë, mais elle crée un lien langagier entre les deux rhétoriques précédentes.

Celles-ci se distinguent pourtant par la nature des controverses qu'elles suscitent. Par exemple, la rhétorique métrologique est à l'œuvre dans les disputes sur la « *figure de la terre* », dans les années 1730 et 1740, qui conduisent aux expéditions de Maupertuis et aux critiques des théories de Cassini. En revanche, celle des indicateurs se déploie dans les débats sur la pauvreté et l'exclusion sociale. Les discussions sur le chômage oscillent entre la première et la

(37) GUILBAUD, 1985.

seconde, mais les arguments des deux types peuvent être distingués, combinés par ailleurs avec des accusations de manipulation délibérée.

Dans le cas des statistiques sociales, un détour banal pour mesurer une réalité inconnaisable est offert par l'action même instituée pour gérer cette réalité politiquement constituée comme problème : délinquance et pauvreté (première moitié du XIX^e siècle), chômage (années 30), avortement (années 70), toxicomanie (années 80), harcèlement sexuel et mauvais traitements à enfants (années 90). Dans ces cas, la question classique est celle du lien entre la réalité visée et la mesure statistique issue de l'action suscitée par le problème social : la croissance de la statistique de la délinquance reflète-t-elle un accroissement des actes délictueux, connus ou non, ou une activité plus énergique de la police ? La statistique résultant de l'interaction de ces deux ordres de causes variables et inconnues est alors un indicateur d'on ne sait quoi (38). Cette dualité fournit matière à des dénonciations récurrentes des données statistiques issues de l'action administrative.

Une réponse classique aux critiques soulevées par l'incommensurabilité entre la réalité latente et son indicateur est d'affirmer que leurs relations sont stables, et que les variations et les différences observées sur l'indicateur reflètent celles de la réalité inobservable. Si les conventions d'enregistrement sont durcies, par exemple par des institutions, des règlements et des procédures standardisées, alors les indicateurs ont une consistance sur la période et dans l'espace couverts par ces conventions institutionnelles. Par exemple, la mesure de la population active et du chômage en France entre 1960 et 1990 aura une certaine consistance interne, pour autant que les modes de fonctionnement des agences d'aide aux chômeurs et des institutions gérant les recensements et les enquêtes statistiques sont stabilisés.

Cet argument de la consistance interne peut à son tour fournir la base d'une troisième rhétorique, partiellement différente de la précédente. Dans ce cas, le statut logique du référent des indicateurs est mis entre parenthèses, et l'accent est mis sur la cohérence, la continuité et, éventuellement, la capacité prédictive des systèmes de mesures, mis à l'épreuve par les techniques de la statistique inférentielle et des probabilités. Le réalisme des objets résulte alors des procédures statistiques elles-mêmes. Les moyennes de Quetelet, les régressions des économètres, l'axe principal de l'analyse factorielle des tests des psychomètres (Spearman), les axes factoriels de l'analyse des données (Benzecri) produisent des effets de réalité. Dans les controverses, ceux qui souhaitent défaire les réalités des autres parlent de réification : ce mot suggère un durcissement artificiel, obtenu par l'efficacité spécifique, jugée alors trompeuse, du formalisme mathématique. Dans ce cas, la rigueur formelle est retournée contre son objectif normal : loin de consolider l'objet, elle le rend suspect. Mais tant que ne surgissent pas ces situations de controverse, le système des mesures statistiques est « jugé à ses fruits », c'est-à-dire à ses capacités à tenir ensemble des choses et des gens dans un dispositif cognitif et politique cohérent. La statistique est alors un outil de coordination parmi d'autres, permettant à chacun de référer son action à une réalité constituée précisément par le système des mesures, au double sens du mot, statistiques et administratives. Un réalisme interne a remplacé un réalisme externe devenu problématique.

Classes ou variables : quels sujets précèdent les verbes ?

La distinction entre la logique des classes et celle des variables peut éclairer l'analyse des formes grammaticales utili-

(38) Cette problématique rappelle celle des premiers économètres qui, cherchant à construire des lois d'offre et de demande à partir de séries observées de prix et de quantités échangées sur un marché, butaient sur le fait que les variations enregistrées résultent des variations distinctes, simultanées et inconnues, de l'offre et de la demande. L'idée de résolution de « modèles à équations simultanées » est issue de cette question (MORGAN, 1990).

sées dans les textes accompagnant et interprétant des résultats statistiques. Elle constitue un aspect souvent méconnu de la comparaison, par ailleurs classique, entre deux familles de méthodes statistiques, celle de l'analyse des données (ou analyse des correspondances), et celle résultant des calculs de régression linéaire (notamment en économétrie). Les premières, très utilisées par les sciences sociales en France, le sont moins dans les pays anglo-saxons, où, très souvent, ces sciences ont adopté les outils formalisés de façon de plus en plus complexe depuis les années 40 par les économètres.

L'analyse des correspondances (due à J.-P. Benzécri) classe des populations d'individus selon une multiplicité de critères, puis étudie les proximités entre eux du point de vue de la batterie de ces classements. Elle offre un outil d'exploration et de visualisation de la variabilité interne d'une population caractérisée par de nombreuses variables. Les graphiques de l'analyse des correspondances sont des projections orthogonales conservant le maximum de la variance d'un nuage de points construit dans un espace à un grand nombre de dimensions, sur l'espace à deux dimensions de la feuille de papier. La perte d'information résultant de cette projection peut être réduite par l'examen des axes successifs (troisième, quatrième...) conservant le maximum de l'information orthogonale au plan des deux premiers axes.

Tels que les graphiques de Playfair, ces schémas résument et offrent au regard l'information énorme et impossible à maîtriser contenue dans un tableau comportant plusieurs dizaines ou centaines de lignes et de colonnes. Ils permettent aussi de stabiliser et de mémoriser des formes, pour les rapprocher et les combiner ensuite à d'autres formes issues d'autres analyses. Les visualisations de l'espace des classes sociales en France ont beaucoup tiré parti, dans les années 70, de ces diverses ressources des graphiques de l'analyse des correspon-

dances. Leurs justifications semblent reproduire, presque mot pour mot, celles des graphiques de Playfair formulées près de deux siècles auparavant. La lecture de ces graphiques fait apparaître des proximités entre certains des individus et certaines de leurs pratiques, et des constellations de propriétés cooccurrentes, selon une logique statistique et probabiliste, et non pas déterministe. Dans les commentaires écrits, les sujets des verbes sont des groupes sociaux, des classes d'individus, liés entre eux par une communauté probable de comportements dans une perspective holiste de reconstitution de la globalité d'une personne, d'un groupe, ou d'une localité. Ces analyses tentent d'approcher, à partir d'un fichier statistique, la richesse sémantique d'une description littéraire.

Les méthodes économétriques, en revanche, mettent en œuvre des formes grammaticales centrées sur le langage des variables. Le fichier statistique est lu du point de vue de ces dernières, mises en relation par des équations, les unes expliquant les autres selon un modèle comportant une partie déterministe et un résidu aléatoire non expliqué. La distinction entre variables dites endogènes et exogènes, ou expliquées et explicatives, s'oppose à la symétrie des analyses de données, qui ne visent pas l'explication mais la description. Dans l'analyse économétrique, des facteurs agissent, influencent d'autres variables, dont les valeurs résultent de ces déterminations. Ces métaphores ne sont souvent que des notations sténographiques, redoublant une logique entièrement inscrite dans la modélisation. Il y a, dans ce cas, peu d'autonomie de la partie textuelle de l'analyse (39).

La rhétorique des analyses économétriques est souvent proche de celle, suggérée ci-dessus, qui met de côté la question de la référence réaliste de la statistique au profit de la consistance et de la cohérence interne des constructions édifiées à partir de celle-ci. Le langage des variables, ex-

(39) Il est significatif que les analyses économétriques se prêtent facilement à une présentation dans une autre langue (principalement l'anglais) puisque le langage y a peu de relief spécifique, l'information essentielle étant inscrite dans les formulations mathématiques et statistiques. Les analyses de données s'appuient plus sur des mots et donc sur la langue propre à l'analyse.

cluant la présence d'acteurs distincts, se prête bien à des combinaisons verbales qui redoublent les combinaisons exprimées par les équations.

Cette comparaison sommaire de deux rhétoriques stylisées devrait être complétée par une étude de leurs usages, c'est-à-dire de leurs inscriptions dans des constructions argumentatives, scientifiques ou politiques, plus générales. Quels sont les contraintes et les effets spécifiques à chacune d'entre elles ? En quoi les langages des classes et des variables participent-ils à des façons différentes de mettre en avant des arguments statistiques pour construire et communiquer des énoncés ? Le langage de la régression économétrique, distinguant les variables exogènes (explicatives) et endogènes (à expliquer) est adapté à la construction de modèles d'action, liant des effets (recherchés) à des causes (actions possibles). La modélisation macroéconomique, largement utilisée entre les années 50 et 80, est fondée sur les combinaisons de plus en plus complexes de telles relations économétriques construites à partir de régularités observées dans le passé. Elle vise à la fois à prévoir et à tester les effets de mesures de politiques économiques. Ce langage des variables est organisé selon des métaphores mécaniques. L'écart entre le déterminisme complet d'une machine et l'indétermination partielle du système économique modélisé est formalisé en termes de distribution de probabilité. Le modèle est une machine déterministe brouillée par un aléa, mais celui-ci est apprivoisé par des lois de probabilité.

L'analyse des données met en œuvre une toute autre rhétorique. Elle simplifie, classe et ordonne la diversité du monde pour l'offrir au regard. Cette diversité n'est pas supposée produite par une loi de probabilité *a priori*. Les cooccurrences décrites et classées se prêtent à des métaphores plus organicistes que mécaniques. Une espèce se caractérise par la juxtaposition de divers traits, sans que les uns « ex-

pliquent » les autres. Le darwinisme décrit des cas singuliers et des constellations de propriétés, mais ne donne, comme explication générale de l'évolution, qu'un principe de sélection agissant au cas par cas. Le langage de l'analyse des données est adapté à la description et à la narration. Comme son ancêtre lointain de la statistique camérale allemande du XVIII^e siècle, il offre au regard des cartes et des voyages possibles, en dotant l'espace social de points cardinaux (les axes factoriels), de zones et de territoires partiellement cohérents, de points d'accumulation qui, telles des villes remarquables, méritent que l'on s'y attarde. Ce langage de la pérégrination statistique s'oppose ainsi au machinisme du modèle économétrique. Pour l'un, l'indétermination y prend la forme de la contingence d'associations inattendues, alors que, pour l'autre, elle est enserrée par des urnes probabilistes constantes ou variant selon des lois constantes.

La distinction suggérée ici entre un langage de l'explication, tourné vers l'action et le contrôle, et un langage de la description, tourné vers la visualisation et la narration, ne recouvre pas celle proposée auparavant entre quantification et qualification, puisque ces deux processus de mise en forme sont à l'œuvre dans chacun des deux langages. L'idée de « contrôle » implique ici à la fois un suivi politique de l'action et un encadrement cognitif de celle-ci dans des urnes probabilistes précisément « spécifiées ». Les mesures statistiques permettent de suivre les effets des mesures politiques. Par ailleurs, la description et la narration identifient des personnes, les qualifient et les comparent selon des critères et des mesures variés. L'action comme la narration font flèche de tout bois, classificatoire ou métrologique.

RÉFÉRENCES

- ARMATTE M. (1994), *Invention et intervention statistiques*. Une conférence exemplaire de Karl Pearson (1912), *Politix*, 25, mars 1994, p. 21-45.
- BAYES T. (1764), An essay toward solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London*, 53, p. 370-418.
- BENZECRI J.P. (1982), *Histoire et pré-histoire de l'analyse des données*, Dunod, Paris.
- BERNOULLI J. (1713), *Ars Conjectandi*, Basil, Thurnisorium.
- BLONDIAUX L. (1994), *Le chiffre et la croyance*. L'importation des sondages d'opinion en France ou les infortunes d'une opinion sans publics, *Politix*, 25, mars 1994, p. 117-152.
- BRIAN E. (1994), *La mesure de l'État*. Administrateurs et géomètres au XVIII^e siècle, Albin Michel, Paris.
- BRU B. (1988), *Estimations Laplaciennes*. Un exemple : la recherche de la population d'un grand Empire, 1785-1812, in Mairesse J. (éd.), *Estimation et sondages*. Cinq contributions à l'histoire de la statistique, Economica, Paris, p. 7-46.
- CALLON M., LAW J. (1989), « La proto-histoire d'un laboratoire ou le difficile mariage de la science et de l'économie », *Cahiers du Centre d'études de l'emploi*, n° 32, p. 1-34.
- CLERO J.P. (1988), Un instrument de mesure des croyances. *La règle de Bayes, Histoire et mesure*, Editions du CNRS, vol. III, n° 4, p. 491-513.
- CONEIN B., DODIER N., THEVENOT L. (éds) (1993) Les objets dans l'action, série « *Raisons pratiques* », n° 4, Editions de l'EHESS.
- COSTIGAN-EAVES P., MACDONALD-ROSS M. (1990), William Playfair (1759-1823), *Statistical Science*, vol. 5, n° 3, p. 318-326.
- COUMET E. (1970), La théorie du hasard est-elle née par hasard ?, *Annales ESC*, n° 3, mai-juin 1970, p. 574-598.
- COURNOT A. (1843), *Exposition de la théorie des chances et des probabilités*, ré-édité en 1984, in *Oeuvres complètes* de Cournot, Bernard Bru (éd.), Vrin, Paris.
- DASTON L. (1988), *Classical Probability in the Enlightenment*, Princeton University Press, Princeton.
- DE MOIVRE A. (1738), *The Doctrine of Chances*, Woodfall, London.
- DESROSIERES A. (1993), « La politique des grands nombres ». *Histoire de la raison statistique*, La Découverte, Paris.
- DUPAQUIER J. et M. (1985), *Histoire de la démographie*, Perrin, Paris.
- EISENSTEIN E. (1991), *La révolution de l'imprimé. A l'aube de l'Europe moderne*, La Découverte, Paris.
- FAGOT-LARGEAULT A. (1989), *Les causes de la mort. Histoire naturelle et facteur à risque*, Vrin, Paris.
- GIGERENZER G. et AL II (1989), *The Empire of Chance – How Probability Changed Science and Everyday Life*, Cambridge University Press, Cambridge.
- GIGERENZER G. (1991), From tools to Theories : A Heuristic of Discovery in *Cognitive Psychology*, *Psychological Review*, 98, 2, 254-267.
- GOODY J. (1979), *La raison graphique*, Minuit, Paris.
- GOULD S.J. (1983), *La mal-mesure de l'homme*, Ramsay, Paris.

- GRIBAUDI M., BLUM A. (1990), « Des catégories aux liens individuels » : l'analyse statistique de l'espace social, *Annales ESC*, n° 6, novembre-décembre 1990, p. 1365-1402.
- GUILBAUD G.T. (1985), *Leçons d'à-peu-près*, Bourgois, Paris.
- HACKING I. (1975), *The Emergence of Probability*, Cambridge University Press.
- HACKING I. (1990), *The Taming of Chance*, Cambridge University Press.
- HACKING I. (1991), *Statistical Language, Statistical Truth and Statistical Reason : The Self-authentication of a Style of Scientific Reasoning*, in McMullen J. (éds), *Social Dimension of Sciences*, Université de Notre-Dame Press.
- JACOB C. (1992), *L'empire des cartes. Approche théorique de la cartographie à travers l'histoire*, Albin Michel, Paris.
- KEVERBERG, baron de (1827), *Notes sur Quetelet*, Nouveaux Mémoires de l'Académie royale des sciences et belles-lettres de Bruxelles, 4 p. 175-192.
- LATOUR B. (1984), *Les microbes. Guerre et paix, suivi de Irréductions, Mé-taillé*, Paris.
- LATOUR B. (1993), *La clef de Berlin, et autres leçons d'un amateur de sciences*, La Découverte, Paris.
- LAVE J. (1988), *Cognition in Practice*, Cambridge University Press.
- LAZARSFELD P. (1970), « Notes sur l'histoire de la quantification en sociologie : les sources, les tendances, les grands problèmes », in *Philosophie des sciences sociales*, Gallimard, Paris, p. 75-162.
- LE MEE R. (1975), « La réglementation des registres paroissiaux en France », *Annales de démographie historique*.
- MARCHAND O., THELOT C. (1991), *Deux siècles de marché du travail*, Insee, Paris.
- MORGAN M. (1990) *The History of Econometric Ideas*, Cambridge University Press.
- NEYMAN J. (1934), *On the Two Different Aspects of the Representative Method : the Method of Statistical Sampling and the Method of Purposive Selection*, *Journal of the Royal Statistical Society*, 97, p. 558-606, discussion p. 607-620.
- PLAYFAIR W. (1786), *The Commercial and Political Atlas*, 1st ed., J. Debrett, London.
- PLAYFAIR W. (1801), *The Statistical Breviary*, T. Bensley, London.
- POISSON S.D. (1837), *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*, Bachelier, Paris.
- PORTER T. (1986), *The Rise of Statistical Thinking*, Princeton University Press.
- ROYSTON E. (1956), *A note on the History of the Graphical Presentation of Data*, *Biometrika*. vol. 43. december, p. 241-247.
- STIGLER S. (1986), *The History of Statistics. The Measurement of Uncertainty before 1900*. Harvard University Press.
- THEVENOT L. (1983), « L'économie du codage social, Critiques de l'économie politique », n° 23-24, avril-septembre 1983, p. 188-222.
- ZUCKER-ROUVILLOIS E. (1994), *La chœur statistique ou l'Etat garant de l'ordre du vivant. Les déclarations d'état-civil et d'IVG*, *Politix*, 25, mars 1994, p. 73-103.