

STATISTIQUE ET ANALYSE DES DONNÉES

ALAIN DESROSIÈRES

La partie pour le tout : comment généraliser ? La préhistoire de la contrainte de représentativité

Statistique et analyse des données, tome 13, n° 2 (1988), p. 93-112.

http://www.numdam.org/item?id=SAD_1988__13_2_93_0

© Association pour la statistique et ses utilisations, 1988, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LA PARTIE POUR LE TOUT : COMMENT GÉNÉRALISER ? LA PRÉHISTOIRE DE LA CONTRAINTE DE REPRÉSENTATIVITÉ

Alain DESROSIÈRES,
Administrateur de l'INSEE

Les enquêtes par sondage sur échantillons représentatifs ne sont apparues qu'au début du XX^e siècle, alors que les théories probabilistes qu'elles impliquent existaient depuis plus de 100 ans. Ce paradoxe peut s'expliquer par l'usage des enquêtes sociales pour la gestion des problèmes de pauvreté et de protection sociale. Le passage d'une gestion locale à un État Providence traitant ces questions au niveau national a nécessité une uniformisation du territoire et une standardisation des modes de description des personnes, lesquelles constituent les préalables indispensables à la mise en œuvre des méthodes probabilistes et à la définition d'une base de sondage.

Representative samplings surveys only appeared at the beginning of the 20th century, whereas the probabilistic theories they imply had been in existence for more than 100 years. This paradox can be explained by the use of social surveys for the management of poverty and social protection affairs. The transition from a local management to a welfare state dealing with those questions at a national level implied to make the territory uniform and to standardize the ways of describing the persons. Those were the indispensable preliminary for the utilization of probabilistic methods and for the definition of a sampling frame.

Un penseur allemand nous offre un modèle qui convient à nos besoins : je veux parler du système monadique de Leibniz. Chaque culture (ou société) exprime à sa manière l'universel, comme chacune des monades de Leibniz. Et il n'est pas impossible de concevoir une procédure (compliquée et laborieuse il est vrai) permettant de passer d'une monade ou culture à une autre par l'intermédiaire de l'universel pris comme l'intégrale de toutes les cultures connues, la monade-des-monades présente à l'horizon de chacune.

Saluons, en passant, le génie : c'est du milieu du XVII^e siècle que nous vient ce qui est sans doute la seule tentative sérieuse de conciliation entre individualisme et holisme. La monade de Leibniz est à la fois un tout en soi et un individu dans un système uni dans ses différences mêmes, disons le Tout universel.

Louis Dumont — Essai sur l'individualisme, pp. 195-196.

La technique des enquêtes par sondages probabilistes n'apparaît que vers la fin du XVIII^e siècle, et encore sous une forme très rudimentaire, plus intuitive que formalisée, avec le Norvégien Kiaer. Les premiers calculs d'intervalles de confiance, par l'Anglais Bowley, datent de 1906, et la formalisation détaillée des méthodes de stratification n'est présentée par Neyman qu'en 1934¹.

1. Ce texte a été présenté aux 19^e journées de statistique de Lausanne (mai 1987). Il est un élément d'une recherche en cours sur l'histoire des sciences sociales empiriques, menée en collaboration avec Michaël Pollak, dans le cadre du groupe de sociologie politique et morale de l'École des hautes études en sciences sociales. Il doit beaucoup aux remarques de Luc Boltanski et Laurent Thévenot, faites sur une version antérieure.

LA PARTIE POUR LE TOUT · COMMENT GÉNÉRALISER ?

Or des enquêtes auprès d'un petit nombre d'individus ont été conduites depuis beaucoup plus longtemps, en particulier tout au long du XIX^e siècle, et souvent par des personnes de haute culture scientifique (polytechniciens ingénieurs des Corps des Mines ou des Ponts) pour qui les éléments de calcul des probabilités nécessaires à l'« intuition » de la méthode des sondages n'auraient pas dû constituer une difficulté insurmontable. Laplace les avait d'ailleurs utilisés dès la fin du XIX^e siècle pour estimer la population française, mais cette expérience était restée sans suite pendant plus d'un siècle.

Le fait que les formalisations et les usages systématiques des méthodes d'échantillonnage probabiliste ne soient aujourd'hui anciens que d'à peine plus d'un demi-siècle montre que l'invention et la mise en œuvre d'une technologie supposent des conditions inséparablement cognitives et sociales : avant d'inventer la solution du problème, il fallait inventer le problème lui-même, c'est-à-dire ici la contrainte de représentativité, au sens donné désormais à ce mot par les statisticiens.

Or on ne saurait méconnaître que ce souci exprimé en termes d'homothétie, pour certains éléments précisément définis, de la partie et du tout, est extrêmement récent, en tout cas très postérieur au développement des grands recensements (Quetelet, à partir des années 1840) et des enquêtes monographiques (Le Play, à peu près à la même époque). L'examen des travaux déjà menés sur l'histoire des sciences sociales empiriques, des statistiques, et, plus précisément des techniques d'échantillonnage² donnent l'impression que l'on est passé directement d'une époque où la question de la représentativité ne se posait pratiquement pas (voir par exemple les publications, entre 1914 et 1916 par la Statistique Générale de la France, d'enquêtes sur les budgets des ménages, sous la plume du statisticien Dugé de Bernonville, ou du sociologue Halbwachs) à une autre où elle constituait une évidence non interrogée (débat de l'Institut International de Statistique, d'abord entre 1895 et 1903, puis entre 1925 et 1934).

Le débat, quand il a lieu, ne porte pas sur la contrainte de représentativité en tant que telle, mais se déroule en deux temps bien distincts. Entre 1895 et 1903, il s'agit de savoir si d'une part on peut légitimement remplacer *le tout par la partie* (comparaison avec les recensements exhaustifs) et d'autre part si, en procédant ainsi, on fait « mieux » qu'avec les « monographies » leplaysiennes encore très à l'honneur. Or, on le verra, ce « mieux » ne porte pas directement sur la contrainte de représentativité au sens de la *précision de la mesure*, mais sur la possibilité de rendre compte d'un espace *diversifié*. Puis, entre 1925 et 1934, le débat porte sur le choix entre les méthodes d'« échantillonnage aléatoire » et celles dites de « choix judicieux » : le développement par Neyman de la théorie de la stratification portera un coup fatal à ces dernières.

La chronologie sommaire rappelée ci-dessus ressort des quelques études citées sur l'histoire des enquêtes par sondage : elle décrit la façon dont, entre 1895 et 1935, se sont profondément transformées les normes sociales communément admises sur les exigences à attendre de descriptions du monde social visant à *généraliser* à une société entière des observations portant sur une partie de celle-ci. Comment passer de la « partie » au « tout » ?

Or les deux modes de généralisation successivement (et aussi simultanément) à l'œuvre dans les enquêtes sociales menées depuis un siècle et demi semblent rigoureusement hétérogènes l'un à l'autre, non comparables, comme si chacun avait son domaine de validité, sa logique propre, mais que leur confrontation ne pouvait se dérouler que sur le mode de la dénonciation mutuelle, dont on verra ci-dessous quelques exemples.

2. Les principaux articles sur l'histoire de ces techniques sont : Stephan 1948, Seng 1951, Chang 1976, Hansen et Madow 1976 et Kruskal et Mosteller 1980. Si chacun développe bien les définitions formelles de plus en plus précises de la notion actuelle de représentativité, de Kiaer à Neyman, aucun en revanche n'étudie l'histoire de la contrainte de représentativité en tant que telle.

LA PARTIE POUR LE TOUT : COMMENT GENERALISER ?

On voudrait suggérer ici que cette apparente incompatibilité peut être mieux comprise si on la resitue dans une opposition plus vaste entre les diverses façons de penser les liens entre les parties et le tout d'une société, qui se sont affrontées depuis le début du XIX^e siècle, à la suite des deux secousses sociales qu'ont été la Révolution française et le libéralisme économique anglais : on trouve des exemples d'analyses des transformations des relations entre les parties et le tout dans des recherches comme celles de Polanyi (1944), Dumont (1983) ou Nisbet (1966). Par ailleurs, la recherche de Boltanski et Thévenot (1987) est consacrée à une étude systématique de l'économie générale des différentes façons de construire le tout social, et on s'en inspirera ici.

On peut par exemple partir de l'opposition, formulée par Louis Dumont, entre « holisme » et « individualisme », parce qu'elle est souvent utilisée par les anthropologues et les historiens, tout en observant qu'elle donne du « tout » une définition insuffisante pour notre propos. Dans la vision « holiste » qui est pour Dumont celle des sociétés « traditionnelles » précédant les deux secousses révolutionnaires politique et économique, le tout social a une existence antérieure et supérieure à ses parties (et en particulier aux individus). En revanche, dans la vision « individualiste », qui est celle des sociétés « modernes », les individus, citoyens ou agents économiques, se groupent de diverses façons, mais sans que ces groupements ne les dépassent et ne les englobent totalement.

Mais cette manière de concevoir un tout englobant les personnes, opposé à des individus atomisés des sociétés modernes (que l'on retrouve aussi dans l'opposition « communauté-société » de Tonnies) ne rend pas compte d'une autre façon, *critérielle*, de construire le tout, qui est justement celle du statisticien préoccupé de bâtir un *échantillon représentatif*. Le « tout social » du holisme de Dumont et l'« exhaustif » de la statistique constituent deux façons complètement différentes de penser la totalité, et leur opposition nous aidera à comprendre ce qui distingue les deux façons implicites de généraliser sous-jacentes aux monographies et aux enquêtes par sondage (Thévenot 1987).

Les configurations intellectuelles schématiquement stylisées ci-dessus s'affrontent et se combinent de façons variées au cours du XIX^e siècle chez les pères fondateurs des sciences sociales : Quetelet, Tocqueville, Marx, Tonnies, Durkheim, Pareto... (Nisbet 1966). Mais elles n'agissent pas comme des « deus ex machina » tirant alternativement les fils de telle ou telle pratique de recherche empirique. Elles constituent plutôt des constellations ayant chacune leur cohérence propre, relativement hétérogènes l'une à l'autre, et que l'on peut suivre dans les débats sur les méthodes d'enquête de la période considérée. Chacune d'entre elles implique à la fois des façons différentes de penser la gestion du monde social, la place des sciences sociales dans cette gestion, et la place des schèmes probabilistes dans ces sciences, de Quetelet à Fisher. Ceci nous servira de fil conducteur pour l'examen d'une histoire qui décrira successivement les enquêtes sociales avant que l'on ne parle de « méthode représentative », l'évolution de l'usage des raisonnements probabilistes au cours de la période, les débats de l'Institut International de Statistique entre 1895 et 1934, les premiers usagers de la méthode et les discussions entre « choix aléatoire » et « choix judicieux ».

La rhétorique de l'exemple

On peut reconstituer la philosophie des enquêtes visant à généraliser à partir d'observations effectuées sans la contrainte moderne de représentativité, à partir de trois cas en apparence bien différents : les monographies de Le Play et des leplaysiens entre 1830 et 1900, les travaux du durkheimien Maurice Halbwachs sur les budgets ouvriers entre 1900 et 1940, et enfin les enquêtes sur la pauvreté des Anglais Booth et Rowntree entre 1880 et 1910, ancêtres, par leurs questions, des enquêtes représentatives de Bowley.

Le point commun entre ces différentes enquêtes, par ailleurs fort différentes, est que les personnes interrogées y sont choisies à partir de réseaux de familiarité : familles désignées comme

« typiques » par des notables de village dans le cas de Le Play, personnes habituellement connues de « visiteurs scolaires » dans celui de Booth, ouvriers volontaires trouvés par l'intermédiaire des syndicats dans celui d'Halbwachs. Ces méthodes de sélection sont bien sûr stigmatisées comme génératrices de « biais » dans la période ultérieure, mais, dans le contexte où elles sont employées, elles sont cohérentes avec les buts de ces enquêtes, qui sont plutôt de décrire le fonctionnement (ou les dysfonctionnements) de communautés ouvrières soumises aux bouleversements des premières phases de l'industrialisation. Il ne s'agit pas encore de *mesurer* pour préparer les *mesures à prendre* comme ce sera le cas quand se développe l'« État Providence », mais plutôt de rassembler des éléments susceptibles de *camper les personnages* d'une histoire à raconter ou à organiser, en particulier grâce à des travaux typologiques : la classification, construisant des *acteurs collectifs*, est un des produits de ce type d'enquêtes, alors que, à la phase suivante, les individus atomisés étant devenus les acteurs essentiels (par exemple du vote, ou de l'acte d'achat dans les études de marché), il importera de les *compter* précisément.

On ne peut cependant pas affirmer que tout schème probabiliste soit absent de ces enquêtes et de leurs interprétations. Mais il s'agit d'une version plutôt « holiste » des probabilités, inspirée par Quetelet et par l'importance qu'il accorde à la *régularité des moyennes* calculées sur de grandes populations, opposée à la dispersion et au caractère apparemment imprévisible des comportements individuels. Cette régularité vient étayer avec vigueur une conception du tout social dépassant et englobant les parties qui le composent (Desrosières 1985).

Cette application de la loi des grands nombres à la stabilité de moyennes calculées (par exemple des taux annuels de natalité, de nuptialité, de crimes ou de suicides) frappe vivement les contemporains de Quetelet et constitue la principale armature d'une macrosociologie pour laquelle le « social » a une réalité extérieure et supérieure aux individus : c'est là la fil conducteur du « Suicide » de Durkheim, et aussi, on le verra, des analyses de la « conscience ouvrière » chez Halbwachs.

Mais c'est parce que le modèle probabiliste de la loi des grands nombres est tiré vers la régularité des moyennes et non vers des distributions, des fréquences de caractères, ou des dispersions, qu'il ne peut pas être à l'origine d'une utilisation en termes d'échantillonnage aléatoire. En effet, si les moyennes sont stables, il suffit de trouver des cas proches de ces moyennes, et, à eux seuls, ces cas « typiques » incarneront la totalité, non seulement ils la « représenteront », mais littéralement, ils *seront* cette totalité, puisque, dans le schème holiste, c'est la totalité qui est première, les individus n'en étant que des *manifestations contingentes*. Le modèle intellectuel de cette conception de la moyenne est en effet fourni par la théorie des *erreurs de mesure* des astronomes et des artilleurs : l'« ascension droite » d'une étoile n'est connue qu'à travers une série de mesures contingentes distribuées selon une loi normale autour d'une moyenne, qui en constitue la meilleure « estimation ». De la même façon, pour Quetelet, les individus contingents sont des manifestations aléatoires d'un « dessein divin » qui constitue la réalité supérieure.

Sachant que vers la fin du XIX^e siècle, c'est cette conception de la statistique qui est encore dominante (en France du moins, car en Angleterre ce n'est déjà plus vrai avec les premiers travaux de Galton et Pearson); on s'explique mieux la façon dont Émile Cheysson, un disciple de Le Play, décrit et justifie, en 1890, la « méthode des monographies » (Desrosières 1986), très peu de temps avant que le Norvégien Kiaer ne présente sa « méthode représentative », profondément différente de la précédente :

LA MONOGRAPHIE... fuit avec grand soin le cas particulier et poursuit le cas général; elle néglige l'accident, l'exception, l'anomalie, pour s'acharner après la moyenne, après le type. C'est le type qui est la véritable essence de la monographie. Hors du type, pour elle, point de salut; mais avec le type, elle acquiert vraiment le privilège d'éclairer d'une vive lumière les investigations économiques et sociales...

L'OBSERVATEUR... est guidé dans ses choix par les grandes statistiques synthétiques, par les enquêtes administratives, qui couvrent le pays de leur réseau et débroussaillent pour ainsi dire le terrain sur lequel va opérer l'auteur des monographies. Grâce aux données ainsi mises à sa disposition, il connaîtra d'avance la population qu'il veut étudier et choisira son type avec précision et sans crainte d'erreur. La statistique officielle va donc en avant-garde et dégage les moyennes qui conduisent le monographe à son type. A son tour, la monographie lui rend le service de vérifier par une étude détaillée les résultats généraux de l'enquête. Ces deux grands procédés se contrôlent ainsi l'un l'autre, mais en gardant chacun ses traits distinctifs.

Pendant que la méthode des enquêtes administratives s'étale en surface, la monographie creuse en profondeur. La statistique officielle, mettant en œuvre toute une armée d'agents plus ou moins zélés et expérimentés, accumule une masse de faits saisis un peu superficiellement et sous un seul aspect; elle les verse pêle-mêle sous ses meules pour les triturer ensemble; elle compte sur la loi des grands nombres pour éliminer les erreurs des observations élémentaires. La monographie vise au contraire la qualité bien plus que la quantité des observations; elle n'emploie que des observateurs de choix, à la fois artistes et savants, qui s'emparent puissamment d'un fait type, d'un fait universel, et s'acharnent après lui pour le disséquer jusque dans sa moelle (Cheysson 1890, pp. 2 et 3).

Le cœur de ces monographies est constitué par le relevé des dépenses et des recettes du budget de la famille, effectué grâce à un séjour plus ou moins prolongé de l'enquêteur dans le foyer. Mais, même si elles sont établies selon un « cadre uniforme », celui de Le Play, elles ne sont pas destinées vraiment à être comparées entre elles afin de dégager des structures de budgets typiques de milieux différents, comme le fera par exemple Halbwachs peu après, et ce sera une des critiques que ce dernier adressera aux monographies leplaysiennes (Halbwachs 1912). La question est donc : à quoi servent-elles ?

- Ces enquêtes semblent essentiellement orientées vers la défense et l'illustration d'une certaine conception de la famille et des rapports sociaux. Les arguments avancés par Cheysson ont trait aux besoins de l'administration et du législateur soucieux d'apprécier les effets de leurs mesures générales et abstraites sur les cas singuliers et concrets que constituent les familles, mais ce souci « gestionnaire » est aussi associée à une préoccupation morale :

« Cette connaissance est indispensable à la fois au moraliste qui veut agir sur les mœurs et à l'homme d'État qui dispose de l'action publique. La loi est une arme à deux tranchants : si elle a une grande puissance pour le bien, elle peut aussi faire beaucoup de mal entre des mains inexpérimentées... » (Ibid., p. 6)

De fait, sont évoquées des utilisations possibles pour l'étude de la répartition des charges fiscales sur les agriculteurs, les commerçants et les industriels, ou des effets de certaines réglementations du travail concernant l'interdiction du travail des femmes et des enfants. Mais, même dans ces cas, la préoccupation morale est explicite, et en revanche n'est exhibée, ni même suggérée, aucune technique permettant de faire la médiation pratique entre les quelques monographies disponibles et les totalisations étatiques : la question d'un éventuel cadre catégoriel à l'intérieur duquel trouveraient place les cas individuels n'est jamais évoquée.

Les monographies de Le Play et de ses disciples ont été critiquées puis oubliées pour des raisons à la fois techniques et politiques : d'une part elles n'offraient aucune garantie méthodologique quand au choix de l'échantillon, et d'autre part elles servaient à étayer un discours hostile à la Révolution française, au suffrage universel et au Code Civil, et cherchaient à rétablir les rapports sociaux de l'Ancien Régime. Pourtant le projet cognitif et le projet politique étaient très cohérents. La connaissance était produite par une familiarité prolongée de l'enquêteur avec la famille ouvrière. Elle

était présentée explicitement comme utile non seulement à produire un savoir, mais aussi à créer et entretenir des rapports personnels de confiance entre membres des classes supérieures et inférieures.

Une telle connaissance, centrée sur le contact direct et la valeur de l'exemple, n'excluait d'ailleurs pas complètement des traitements comparatifs : une attention particulière était portée à la part *non monétaire* du revenu, enregistrées sous le nom de « subventions » ou « aubaines » : droits d'usages de terrains communaux, jardins familiaux, avantages en nature. Ne relevant pas du marché, ces formes de rétribution maintenaient des liens directs et personnels entre membres des diverses classes.

Inclus le plus souvent dans des plaidoyers pour les vertus familiales traditionnelles, les monographies servaient rarement à des comparaisons de structure, sauf dans un cas, celui justement de la part du revenu non monétaire, supposée refléter la persistance des liens sociaux de type patriarcal. Ainsi, en Europe de l'Est, société encore largement rurale et peu industrielle, cette part était-elle plus forte qu'à l'ouest, où les villes et l'économie marchande occupaient une plus grande place.

L'idée que les liens sociaux traditionnels, bouleversés par les transformations économiques et politiques de la société, ne peuvent être compris que par une proximité de longue durée entre observateurs et observés, et une prise de compte de la globalité des significations d'actes que l'enquêteur ne peut découper et coder en items a priori, se retrouve dans d'autres modes de connaissance qui se développeront ultérieurement, et qui impliquent d'autres façons de généraliser que celle de la représentativité statistique : l'ethnologie décrivant les sociétés non européennes à partir de séjours longs et patients du chercheur dans la communauté, la psychanalyse construisant un modèle des structures de l'inconscient à partir d'un matériel complètement singulier, recueilli au cours d'échanges personnels de très longue durée.

Cette même démarche de connaissance, selon laquelle un cas étudié de façon très approfondie peut à lui seul permettre de décrire un « homme générique », et les mécanismes d'une « commune humanité » se retrouve aussi, à la même époque dans les travaux des psychologues expérimentaux, comme l'Allemand Wundt qui, en laboratoire, tente de dégager de tels traits généraux. L'idée de dispersion des traits des psychismes humains n'apparaîtra que plus tard, grâce aux travaux de Pearson et Spearman.

On voit bien, dans la façon dont Cheysson justifie ci-dessus le choix de ses « cas typiques » par référence aux moyennes calculées grâce aux grandes enquêtes synthétiques, à quel point la théorie de la moyenne de Quetelet a pu fournir pendant près d'un siècle un schème intellectuel permettant de penser simultanément la diversité des cas individuels et l'unicité d'une espèce, ou d'un groupe social :

« ... grâce aux données de la statistique, le monographe connaît d'avance la population à étudier, et choisit son type avec précision et sans crainte d'erreur. La statistique officielle va en avant garde et dégage les moyennes qui le conduisent à son type. »

Cette manière de faire annonce, d'une certaine façon, la méthode du « choix judicieux », qui, entre les années 1900 et 1930, constituera une sorte de chaînon entre les deux types de méthode, par le biais du *territoire*.

Halbwachs : le groupe social et ses membres

Halbwachs est un élève de Durkheim mais, plus que celui-ci, il est soucieux de techniques d'enquêtes et d'observations des faits, dès sa thèse sur « La classe ouvrière et les niveaux de vie » (1912). Beaucoup plus que ne le sont les leplaysiens dont il critique vivement les travaux, il est sensible à la *diversité* des cas observés et cherche des moyens d'interpréter celle-ci, comme Durkheim l'avait fait dans « Le suicide », inventant la sociologie quantitative moderne. Par ailleurs, il est au fait de certains travaux des probabilistes : sa thèse complémentaire (1913) porte sur « La théorie de l'homme moyen;

essai sur Quetelet et la statistique morale », et il a même rédigé avec Maurice Fréchet un petit manuel de probabilité en 1924. Pourtant la discussion très serrée qu'il mène dans sa thèse sur les problèmes d'échantillonnage et d'enquête (pages 151 à 162) s'intitule : « le nombre des budgets » et porte sur l'équilibre « économique » à trouver entre le nombre des personnes enquêtées et le caractère plus ou moins approfondi des observations, c'est-à-dire sur la comparaison entre les méthodes dites « intensives » et « extensives », et non pas du tout sur les problèmes de *tirage d'échantillon* qu'il n'évoque pas.

Ce qu'il a retenu de la « loi des grands nombres », c'est qu'un ensemble de causes petites, nombreuses, aléatoires et de sens différents, se compensent pour produire une « moyenne » supposée révéler la vérité essentielle selon le schéma de Quetelet, et ceci est particulièrement vrai dans la méthode dite extensive, alors très utilisée aux États-Unis, où certains échantillons ont dépassé 10 000 personnes. Mais, de même qu'il a rejeté la méthode intensive leplaysienne qui ne fournit aucune indication sur la diversité et ne permet donc pas des *croisements de variables* suggérant des *explications*, il rejette aussi la méthode extensive des Américains, car, en sociologue durkheimien conséquent, il se méfie d'une interprétation microsociologique de la loi des grands nombres.

En effet, dit-il, les tenants des méthodes extensives ont bien senti ce que les réponses obtenues pouvaient impliquer d'erreurs, d'inexactitudes ou d'oublis mais

« ils ont cru qu'en multipliant les cas, ils obtiendraient, par le jeu de la loi des grands nombres, une compensation et une atténuation croissante de ces fautes ». (Halbwachs 1912, p. 152)

Mais Halbwachs suppose que, comme tout autre fait social, ces oublis ou imperfections ont des causes macro-sociales systématiques et ne sont pas aléatoires au sens probabiliste : il a l'intuition du « biais systématique » :

« ... on a observé que les oublis ne se produisent pas au hasard... Un oubli est la concomitance d'un état d'inattention du sujet et d'un devoir d'une importance sociale assez grande à remplir pour lui... Si les oublis sont périodiques, réguliers, il faut que les causes de l'inattention et les causes qui expliquent ces devoirs à remplir soient des forces constantes... Mais le fait de la périodicité des oublis fonde à croire que les forces qui les expliquent, en vertu de l'agencement de la vie sociale, s'exercent à leur tour, et que leur effet est dans une proportion exacte avec leur intensité constante. » (Ibid., pp. 152-153)

Il se livre, ensuite à une analyse détaillée des « effets d'enquête » liés à l'interaction, au rôle de l'enquêteur; montrant que, dans tous les cas, les erreurs ne sauraient être aléatoires et indépendantes les unes des autres. En effet dans ces enquêtes extensives américaines, il n'y a pas de tenue de carnet de compte et l'enquêteur demande a posteriori à l'enquêté d'évaluer ses dépenses pendant un laps de temps, ce qui, en fin de compte fait reposer les mesures sur des « opinions » :

« Le but de la science, ici comme ailleurs, est de substituer aux principes, aux opinions toutes faites, aux idées vagues et contradictoires, une connaissance précise, fondée sur les faits; mais voici que les faits sont déformés d'avance, vus à travers une opinion qui en atténue et noie les contours. On tourne dans un cercle de moyennes. » (Ibid., p. 157)

Toute cette critique de la méthode américaine est centrée sur l'idée que la « loi des grands nombres » est supposée annuler des erreurs nombreuses, petites et indépendantes, et non pas du tout sur le fait que cette même « loi » pourrait justifier une technique de tirage aléatoire d'un échantillon dans une population diversifiée *en réalité* et non pas à cause des erreurs d'observations. On est encore proche du modèle de Quetelet, plutôt que de celui de Galton et Pearson.

Pourtant Halbwachs évoque explicitement la diversité de la population ouvrière, et c'est même précisément à ce sujet qu'il s'oppose à Le Play, et à son choix des « cas typiques » guidé par les moyennes. Il se demande par quel trait extérieur visible on va pouvoir caractériser une famille ouvrière moyenne, en tout cas pour la structure de ses recettes et de ses dépenses :

LA PARTIE POUR LE TOUT : COMMENT GÉNÉRALISER ?

« Beaucoup de ménages ne savent pas eux-mêmes comment leur budget s'équilibre ou non. En tout cas, ils ne le font point savoir. Or cela ne se voit pas, ne se sent pas, ne se devine pas, comme, à la rigueur, la régularité ou la solidité du lien familial. Où trouver et comment chercher les cas moyens en ce domaine? » (Ibid., p. 159)

Il est donc nécessaire d'observer directement (avec tenue de carnets de comptes) une certaine diversité de familles, afin de pouvoir étudier les effets des variations de tel ou tel facteur : taille de la famille, présence d'enfants, etc. Mais cette diversité et ces variations restent macro-sociales, dans la ligne du durkheimisme du Suicide et des « causes constantes » de Quetelet. Les enquêtes allemandes utilisées par Halbwachs lui semblent n'avoir aucun des inconvénients des autres : suffisamment petites pour permettre une tenue soigneusement vérifiée de carnets de comptes, et suffisamment grande pour étudier des variations.

Mais il n'en reste pas moins que le but de ces recherches est de dégager les traits d'une « conscience ouvrière » commune, dont le caractère relativement homogène ne provient pas d'une essence divine comme chez Quetelet (Halbwachs est plutôt matérialiste) mais d'une communauté de conditions matérielles d'existence, qu'une adaptation quasi-darwinienne conduit à des comportements homogènes, à la fois dans les pratiques et dans les consciences. Et c'est bien parce que c'est, en fin de compte, cette conscience ouvrière qui l'intéresse, que les problèmes d'échantillonnage ne se posent pas du tout pour lui dans les mêmes termes qu'ils ne le feront pour ceux qui, un demi-siècle plus tard, utiliseront de telles enquêtes par exemple pour construire une comptabilité nationale.

Ainsi, se félicitant de ce que l'Union des Travailleurs des Métaux allemands ait pu rassembler 400 budgets, il note :

« C'est évidemment à la solidarité ouvrière, à l'influence que les organisations syndicales exercent de plus en plus sur leurs membres, qu'on doit d'être parvenu à ce résultat. » (Ibid., p. 138)

Ailleurs, commentant certains budgets rassemblés par C. Booth à Londres pour des familles très pauvres, il s'interroge sur la rigueur de la tenue des carnets de compte dans ces cas extrêmes, et conclut, de façon glaciale, en se demandant si ces fractions les plus démunies font bien partie de la classe ouvrière :

« ... en raison de l'état de dénuement où se trouvent ces ménages, peut-être une observation courte donne-t-elle une juste image de leur misère chronique : mais on n'en est pas sûr. Au reste, pour l'étude des classes sociales, cette couche sociale inférieure, qui ne s'élève pas à une conscience commune, n'est pas la plus intéressante, et peut à la rigueur n'être connue que superficiellement. » (Ibid., p. 470)

Malgré tout ce qui distingue les leplaysiens et les durkheimiens, dans leurs projets tant scientifiques que politiques, on peut cependant retenir quelques points communs à leurs démarches : soucieux avant tout de repenser (de façon il est vrai beaucoup plus subtile pour les seconds que les premiers) la nature de liens sociaux défaits par les changements révolutionnaires, leurs démarches empiriques visent à signaler la persistance d'anciennes formes ou la naissance de nouvelles formes de tels liens, en appréciant dans chaque cas leur portée morale.

Mais cette façon de mettre en œuvre une démarche empirique à des fins de reconstruction sociale, commune à tout le XIX^e siècle, y compris Quetelet, n'accompagnait encore aucune visée d'action sociale et politique *directe*, et n'impliquait donc en conséquence pour l'enquête aucune nécessité d'*exhaustivité* au sens territorial ou national. Le degré de généralité supposé des *cas* décrits suffisait à appuyer des développements politiques et moraux, qui ne supposaient aucune forme d'insertion territoriale (sauf à la rigueur dans le cas où Le Play comparait de façon très générale l'Europe de l'Est et celle de l'Ouest).

L'histoire de l'émergence progressive de la notion de représentativité au sens moderne peut être lue parallèlement à celle de l'extension et de la transformation des outils politiques et économiques de traitement des problèmes de la pauvreté, depuis la bienfaisance seigneuriale ou paroissiale du

xviii^e siècle, jusqu'aux diverses formes d'État Providence mises en place dès la fin du xix^e siècle. La statistique sociale va à ce moment changer de fonction : alors qu'auparavant, on l'a vu, elle venait plutôt illustrer des analyses compréhensives d'un monde social pensé à travers des schèmes holistes (modèle Quetelet), ou organicistes (modèle Auguste Comte), elle va peu à peu constituer un élément essentiel dans des politiques, par ailleurs très variées, visant à agir au niveau des individus, et la notion de représentativité prendra alors une importance décisive, pour évaluer les coûts et bénéfices des politiques mises en œuvre.

Parmi ces domaines trois d'entre eux vont jouer un rôle essentiel pour imposer la contrainte de représentativité au sens moderne : la mise en place des premières lois d'aide et de protection sociale, dans l'Europe du Nord dès les années 1890, puis le développement des marchés de biens de consommation nationaux (grâce aux chemins de fer) et des études de marché, et enfin la possibilité de mener nationalement des campagnes électorales (grâce à la radio), ceci dans les États-Unis de l'entre-deux-guerres.

Représentativité et délocalisation de la statistique sociale

Le point commun aux transformations dont ces trois domaines sont le théâtre est le passage de modes de gestion *locaux* centrés sur des relations personnelles (bienfaisance, petit commerce, artisanat, marchés ruraux, clientélisme électoral) à d'autres modes, *nationaux*, dans lesquels le territoire en tant que lieu de reproduction quotidienne des liens sociaux perd relativement de l'importance. Cette mise en équivalence générale et cette uniformisation du territoire³ ont été préparées, au moins en France, par l'organisation en départements, la diffusion du Code Civil, le suffrage universel, le service militaire obligatoire, l'obligation scolaire à travers l'école laïque, sans parler du système métrique et de l'horaire des chemins de fer... Elles étaient les conditions préalables pour penser les deux notions liées d'*exhaustivité* et de *représentativité*, absentes totalement, on l'a vu, des démarches leplaysiennes et durkhémiennes⁴.

Une bonne illustration du lien entre ces trois transformations qu'ont été : 1) la délocalisation de la statistique sociale, 2) la diffusion de la « méthode représentative » et de l'échantillonnage aléatoire, et 3) la mise en place de l'État Providence, est fournie par deux études (1976, 1987) de E.P. Hennock (Université de Liverpool) sur la série des enquêtes sur la pauvreté menées successivement par Charles Booth, Seeborn Rowntree et Arthur Bowley. L'examen des débats de l'Institut International de Statistique, à partir de 1895 vient aussi largement confirmer cette hypothèse.

Les pauvres : comment les décrire? qu'en faire?

Tout au long du xix^e siècle, l'histoire des enquêtes anglaises sur la pauvreté est liée aux interprétations et explications successives de celle-ci, et aux modes de traitement proposés. Dans un premier temps (Abrams 1968), un courant « amélioriste » a conduit des investigations locales généralement assorties de recommandations sur l'amélioration de la moralité de la classe ouvrière. La perspective ne diffère de celle de Le Play qu'en ce que la condamnation de l'économie marchande y est nettement moins vigoureuse.

3. Non pas nécessairement dans la réalité quotidienne des choses et des gens comme le suggèrent certaines descriptions apocalyptiques en termes de « normalisation » et de « quadrillage », mais au moins dans la façon d'en faire l'inventaire, de le voir.

4. Encore de nos jours, la présence ou l'absence de ces deux notions dans les recherches en sciences sociales pourraient servir de critère pour distinguer deux façons profondément différentes d'articuler ces sciences avec la gestion du monde social. Le critère pertinent n'est pas, comme on le croit souvent, leur caractère *quantitatif*, mais la contrainte de représentativité.

LA PARTIE POUR LE TOUT COMMENT GENERALISER!

Dans les années 1880, la crise économique fait rage, et la situation apparaît particulièrement dramatique à Londres, et surtout dans ses quartiers est (Hennock 1976, 1987). Une discussion s'élève sur la part de la classe ouvrière située en dessous d'un seuil réputé d'extrême pauvreté. Ce débat récurrent, de tous les temps et de tous les lieux, et surtout absurde tant sont arbitraires les définitions de ce seuil et les méthodes de mesure, est ici intéressant parce qu'il est à l'origine d'une typologie de la pauvreté, tout à la fois descriptive, explicative et opératoire. Booth déduit, de ses mesures de la pauvreté, des mesures à prendre pour la résoudre : expulser de Londres les très pauvres, qui le sont pour des raisons essentiellement morales (alcoolisme, imprévoyance...) afin de soulager le fardeau portant sur ceux qui sont un peu moins pauvres, et qui le sont, eux, pour des raisons économiques et en tous cas macro-sociales (la crise). Le découpage est en fait beaucoup plus complexe (il y a 8 catégories) et porte à la fois sur le niveau et la régularité du revenu. Les comptages sont opérés à partir d'impressions de « visiteurs scolaires » et conduisent à des statistiques détaillées.

Mais l'important réside dans l'association de la taxinomie et des résultats *géographiques*. Jusqu'à ce moment les problèmes de pauvreté étaient traités au niveau des municipalités. Mais le cas de Londres est particulièrement dramatique et il apparaît que, contrairement à ce que l'on pensait, la part des « très pauvres » est à peine moins élevée dans l'ensemble de Londres que dans les seuls quartiers Est (enquête de Charles Booth) : la question de la représentativité géographique des résultats par quartiers conduit peu à peu à des conclusions politiques. La suggestion de l'expulsion des très pauvres suppose bien sûr que c'est à Londres que la situation est la plus grave. Or l'enquête n'a été menée qu'à Londres, et les généralisations à la façon de Le Play ou même d'Halbwachs ne sont plus possibles; pour agir, il va bientôt falloir une *maquette*, un *modèle réduit*.

Quelques années plus tard, Rowntree entreprend dans d'autres villes d'Angleterre (York en particulier) des enquêtes comparables à celle de Booth à Londres. Il sent bien que les méthodes de celui-ci sont douteuses, et porte une attention beaucoup plus grande aux techniques de collecte de données. Il ne peut cependant les changer complètement car l'essentiel est de pouvoir comparer la part des très pauvres dans les deux villes. Or il apparaît que celle-ci est, à York, à peine inférieure à celle de Londres. Ceci appuie l'argument selon lequel la pauvreté ne peut être traitée localement. Le vote d'une nouvelle « poor law » en 1908 intervient dans ce contexte de prise en charge nationale de la protection sociale naissante, apportant ainsi des réponses macro-sociales à des problèmes qui ne peuvent plus être présentés comme relevant de la morale individuelle (sobriété ouvrière et bienfaisance bourgeoise). Entre le Booth des années 1880 et le Rowntree des années 1900, les problèmes posés initialement en termes locaux le sont dorénavant en termes nationaux, mais il manque encore l'outil permettant d'asseoir cette nouvelle statistique sociale nationale : celui-ci sera introduit par Arthur Bowley.

Concurrence internationale et statistiques lourdes

La relation entre la technique d'enquête, l'usage qui en est attendu et les moyens mis en œuvre change radicalement non seulement en raison du vote des nouvelles lois sociales, mais aussi en raison du poids pris dans le contexte anglais de cette époque par les discussions sur la concurrence entre les grands pays industrialisés et sur la question du libre-échange. Plusieurs enquêtes comparatives entre pays sont alors menées. Rowntree entre en contact avec Halbwachs, qui organise une recherche selon les méthodes de l'Anglais, publiée par la SGF (Halbwachs 1914). Puis surtout, le Board of trade anglais monte une très lourde opération dans différents pays, qui, si elle ne met pas encore en œuvre les méthodes probabilistes, est la première d'une telle ampleur en Europe, et porte surtout sur plusieurs pays. En France, elle porte sur 5 605 familles ouvrières, et les questionnaires sont distribués par les syndicats de salariés dans une trentaine de villes (Board of trade 1909).

C'est parce que le gouvernement libéral anglais avait besoin d'arguments pour sa bataille contre le protectionnisme, que les moyens publics importants ont pu être dégagés pour le Board of trade, permettant d'enquêter dans un grand nombre de villes et de poser les problèmes cruciaux pour la mise en place ultérieure de l'infrastructure nécessaire aux enquêtes par sondages : organisation d'un réseau d'enquêteurs homogène, prise en compte des différences de « circonstances locales » (types de logement, habitudes de consommation, structures d'emploi). L'utilisation en vue d'une comparaison *entre villes* interne au pays a été un sous-produit non voulu a priori d'une opération de comparaison internationale (Hennock 1986).

L'étape suivante permise par cette opération de grande ampleur a été que Bowley a pu formuler de façon plausible les conditions de possibilité d'enquêtes « représentatives » (selon le vocabulaire de l'époque) en en organisant une, avec un taux de 1/20, dans 4 villes choisies de façon que deux soient du type « mono-industriel », et deux autres « pluri-industrielles ». Il avait vu aussi que les conditions d'exhaustivité et de représentativité en supposaient une autre : *l'obligation de répondre*, thème absent dans tout ce qui précédait. Ce faisant, il éloigne le processus de l'interaction entre l'enquêteur et l'enquêté du modèle de la familiarité confiante issue de réseaux de sociabilité, des enquêtes antérieures, pour la rapprocher d'une forme de devoir civique général, analogue au suffrage universel ou au service militaire obligatoire.

De même *l'erreur* et la *précision* changent de nature : alors que Rowntree était particulièrement pointilleux sur les modes d'enregistrement des informations, mais ignorant des questions d'échantillonnage, Bowley est moins soucieux du premier type de rigueur, et Rowntree lui reproche par exemple d'accepter des réponses sur les niveaux de salaire formulées par des *épouses* des salariés, en cas d'absence de ceux-ci, ce à quoi lui se refuse. Mais surtout Bowley a fait de l'imprécision, de la marge d'erreur, un objet respectable, propre (l'intervalle de confiance) et non plus honteusement dissimulé dans le silence pudique de la *faute*. Là encore la technique et la loi des grands nombres ont pris la place de la moralisation de l'individu.

Dans la même ligne, Bowley ne cherche plus à identifier la pauvreté à partir d'impressions visuelles liées à des visites, comme le faisait Booth, mais l'appuie sur des variables quantifiables constatées. Il n'a pas non plus l'ambition de distinguer une pauvreté liée à des « mauvaises habitudes » d'une autre résultant de causes économiques, bannissant là encore tout jugement moral de ses investigations.

Enfin, et ceci est la traduction de tout ce qui précède en termes *d'identité professionnelle*, il revendique de ne pas avoir à proposer lui-même les solutions des problèmes de pauvreté, de même que, à peu près au même moment, Max Weber réclame la distinction entre « le savant et le politique », et ceci est une position tout à fait nouvelle par rapport à tout ce qui s'est fait au XIX^e siècle :

« ... en tant qu'économistes et statisticiens, nous ne sommes pas concernés par les palliatifs ou les expédients [pour réduire la pauvreté], mais nous le sommes par une connaissance correcte et un diagnostic exact de l'étendue de ces maux, sur lesquels pourront être élaborés des remèdes raisonnés et permanents... » (Bowley 1906, p. 554)

La comparaison (inspirée du remarquable travail de E.P. Hennock) des travaux du trio formé par Booth, Rowntree et Bowley, qui se connaissaient et se référaient les uns aux autres, montre bien la cohérence des divers aspects cognitifs, techniques et politiques du basculement des moyens de savoir et de généraliser qui s'opère autour de 1900. Vont de pair : la mise en place de l'État Providence au niveau du territoire entier, la délocalisation de la production et de l'interprétation des données, la substitution de machineries techniques neutres aux jugements moraux, et enfin et surtout, l'apparition d'une nouvelle figure professionnelle, le *statisticien d'état*, différent à la fois du savant lettré du XIX^e siècle, tourmenté par la dissolution des liens sociaux, et du fonctionnaire gestionnaire directement responsable du traitement des problèmes sociaux.

Les congrès de l'IIS : de la monographie exemplaire au sondage bien réglé

Cette figure nouvelle s'affirme de plus en plus nettement dans les sociétés statistiques nationales et surtout, à l'Institut International de Statistique (IIS), créé en 1883, qui réunit les plus importants des statisticiens d'état. Dans ces divers lieux, les amateurs éclairés et éclectiques du XIX^e siècle sont peu à peu remplacés par des techniciens professionnels de la statistique, de culture de plus en plus mathématique et de moins en moins historique ou politique.

C'est dans ce cadre qu'est discutée, en deux temps, la « méthode représentative », à partir de 1895. Le coup d'envoi est alors donné par le Norvégien Kiaer, qui a organisé dans son pays un premier « dénombrement représentatif » en 1894, impliquant un tirage successif de localités et de personnes interrogées dans ces localités, et portant sur les professions, les revenus, les dépenses, les jours non travaillés, le mariage, le nombre des enfants...

L'initiative de Kiaer est largement discutée au cours de quatre congrès successifs de l'IIS, entre 1895 et 1903, et, à cette date, le congrès de Berlin adopte une motion favorable à cette méthode, sous réserve que soit spécifié nettement « dans quelles conditions s'est fait le choix des unités observées ». Le rapport alors demandé ne sera présenté qu'en 1925 par le Danois Jensen, et une motion (voir annexe, page 110) est alors adoptée, qui ne tranche pas encore entre les deux méthodes de « prise au hasard » et de « choix judicieux ». Cette dernière méthode ne sera éliminée qu'à la suite des travaux de Neyman en 1934.

Au cours de la première phase (1895-1903), l'aspect probabiliste de la nouvelle méthode et la nécessité du caractère aléatoire des tirages sont très peu vus, et d'ailleurs, au cours de sa première enquête de 1894, Kiaer n'est pas très exigeant sur ce point, dont il ne sent pas encore l'importance. Par exemple, après avoir soigneusement tiré des localités et des rues, il laisse aux enquêteurs le choix des maisons à visiter⁵ :

« ... ils devaient avoir soin de visiter non seulement les maisons moyennes au point de vue social, mais en général des maisons représentant les différentes conditions sociales ou économiques qui se présentent dans la commune. » (Kiaer 1895, p. 179)

En fait, en insistant fortement, et pour la première fois en un tel lieu, sur l'idée de *représentativité*, Kiaer voulait essentiellement montrer que, moyennant quelques précautions (encore rudimentaires) de choix de l'échantillon, on obtenait sur celui-ci, pour quelques variables *contrôlables* (car existant déjà dans les dénombrements exhaustifs), des résultats assez bons pour que l'on puisse supposer que, pour les autres variables, ces résultats soient aussi « assez bons », sans que l'on précise trop le sens de cette dernière expression. L'essentiel de l'idée de représentativité était bien là : *la partie peut remplacer le tout*, alors que dans les enquêtes antérieures, on ne songeait pas à comparer la partie et le tout, car on ne pensait pas le *tout* dans les mêmes termes. Ainsi, pour Quetelet, l'homme moyen résumait à lui seul une population, mais c'était sur ces traits moyens et leurs régularités que l'attention était dirigée, et non sur la population elle-même avec ses limites, ses structures, son exhaustivité.

Pour Kiaer, en revanche, un tel souci de description exhaustive et représentative dans ce sens nouveau est bien présent, même si l'outillage technique n'est pas encore disponible. Ce dernier, nécessaire pour *construire* un échantillon *aléatoire* et calculer des *intervalles de confiance* ne sera présenté qu'en 1906 par Bowley, en dehors de l'IIS (mais Bowley participera activement au Congrès de l'IIS de 1925 qui relancera le débat, avec Jensen, March, Gini...) [Bowley 1906].

5. Au cours de ces Congrès mondiaux de statistique, le Norvégien Kiaer s'exprimait uniquement en français. Les trois langues d'usage étaient, dans l'ordre de fréquence, le français, l'anglais et l'allemand.

Les débats techniques de ces réunions de statisticiens ont déjà été décrits en détail par des recherches importantes (voir par exemple Seng 1951, et Kruskal et Mosteller 1980), sous l'angle de l'intégration progressive des résultats du calcul des probabilités et de la statistique mathématique à la théorie de l'échantillonnage, l'étape décisive à ce point de vue étant le travail de Neyman (1934) sur la stratification, renvoyant le « choix judicieux » aux oubliettes. On ne les reprendra pas ici, mais on examinera plutôt comment Kiaer introduit sa méthode, et comment il éprouve le besoin de comparer celle-ci aux *monographies* à la Le Play, ce qui ne viendra plus à l'esprit des statisticiens du siècle suivant.

Les justifications qu'il donne d'emblée à son enquête sont significatives du basculement d'une période où les rapports entre classes étaient encore pensés en termes d'ordres et de places, et donc incommensurables, à une autre où les individus des diverses classes peuvent se comparer entre eux à une aune commune, où le thème de *l'inégalité*, non concevable dans l'autre système, devient fondamental, et où les problèmes de pauvreté ne sont plus pensés en termes de bienfaisance et de voisinage, mais de lois sociales votées par des parlements.

Kiaer observe en effet que les enquêtes antérieures portaient sur les seuls *ouvriers* (ou les *pauvres*), puisqu'il n'était pas encore concevable de mettre en équivalence, à l'intérieur d'un tout supérieur, les diverses classes. Il est ainsi un des premiers à poser dans de tels termes le problème des « inégalités sociales », et il est frappant que ceci soit dit au début du premier texte d'un statisticien d'état portant sur la représentativité :

« Une chose m'a surtout frappé, c'est que les investigations détaillées concernant les revenus, les habitations et autres conditions économiques ou sociales, qui ont été faites relativement aux classes ouvrières, n'aient pas été étendues d'une manière analogue à toutes les classes de la société. Il me semble évident que même en n'envisageant que la question ouvrière proprement dite, on doit comparer la situation économique, sociale, morale, etc., des ouvriers à celles des classes moyennes et des classes riches. Dans un pays où les classes supérieures sont très riches et les classes moyennes très aisées, les prétentions des classes ouvrières relatives à leurs salaires, à leurs habitations, etc., se mesurent d'après une autre échelle que dans un pays (ou dans une localité) où la plupart des personnes appartenant aux classes supérieures ne sont pas riches et où les classes moyennes se trouvent dans la gêne.

De cette proposition, qui me paraît tout à fait claire, il s'ensuit que pour bien apprécier les conditions de la classe ouvrière, il faudra aussi connaître, outre celle-ci, les éléments analogues dans les autres classes. Mais il faut faire un pas de plus et dire que, puisque la société ne consiste pas seulement dans la classe ouvrière, on ne doit dans les investigations sociales, négliger aucune classe de la société. » (Kiaer 1895, p. 177)

Il explique, immédiatement après, que cette enquête va jouer un rôle pour la création d'une caisse de retraite et d'assurance sociale, garantissant une péréquation sociale et un traitement statistique de risques divers :

Depuis le commencement de cette année, il s'est opéré et il s'opère dans notre pays, un dénombrement représentatif ayant pour but d'élucider diverses questions concernant le projet de création d'une caisse générale de retraite et d'assurance contre l'invalidité et la vieillesse. Ce dénombrement s'opère sous les auspices d'un Comité parlementaire chargé de l'examen de ces questions et dont je suis l'un des membres. (Ibid. p. 177)

Deux ans plus tard, en 1897, au cours d'une nouvelle discussion à l'IIS, le débat tourne sur ce que la « méthode représentative » apporte par rapport à la « méthode typologique » alors préconisée, au sein de l'IIS, par des statisticiens leplaysiens comme Cheysson. Kiaer insiste alors sur l'aspect *territorial*, en évoquant « une miniature du territoire total », montrant non seulement des types, mais aussi la « variété des cas qui se trouvent dans la vie ». Il n'aborde pas encore la question du tirage aléatoire, mais en revanche insiste sur le contrôle des résultats par la statistiques générale :

LA PARTII POUR LE TOUT : COMMENT GENERALISER?

Je ne trouve pas que la terminologie employée dans notre programme, c'est-à-dire « procédés des études typologiques », soit conforme à mes idées. J'aurai occasion de démontrer la différence qui existe entre les investigations par types et les investigations représentatives.

Par investigation représentative je comprends une exploration partielle où l'observation se fait sur un grand nombre de localités éparses, distribuées sur toute l'étendue du territoire d'une telle manière que l'ensemble des localités observées forme une miniature du territoire total. Ces localités ne doivent pas être choisies arbitrairement mais d'après un groupement rationnel basé sur les résultats généraux de la statistique; et les bulletins individuels dont on se sert, doivent être arrangés d'une telle manière que les résultats en puissent être contrôlés à plusieurs égards à l'aide de la statistique générale. (Kiaer 1897, p. 180)

Quand il oppose sa méthode, permettant de décrire la « variété des cas » à celle qui ne montre que des « cas typiques », il souligne une mutation tout à fait parallèle à celle que Galton et Pearson viennent d'opérer par rapport à la vieille statistique de la moyenne de Quetelet : en portant désormais l'attention sur la *variabilité des cas individuels*, avec les notions de variance, de corrélation et de régression, les eugénistes anglais ont fait passer la statistique du stade de l'examen de *tous* résumés par une moyenne (holisme), à celui de l'analyse de distributions d'individus à *comparer* (Mac Kenzie 1981) :

L'institut a recommandé l'investigation par types choisis. Sans contester l'utilité de cette forme d'investigation partielle, je pense qu'elle présente certains désavantages, comparée aux investigations représentatives. Même si l'on connaît les proportions dans lesquelles les différents types entrent dans le total, on est loin d'arriver à un résultat plausible pour l'ensemble; car le total comprend, non seulement les types, c'est-à-dire les rapports moyens, mais toute la variété des cas qui se trouvent dans la vie. Il est donc nécessaire, afin que l'investigation partielle donne une vraie miniature de l'ensemble, qu'on observe non seulement les types, mais toute espèce des phénomènes. Et c'est cela qui pourra se faire, sinon complètement, à l'aide d'une bonne méthode représentative qui ne néglige ni les types, ni les variations. (Kiaer 1897, p. 181)

Puis, tentant de se situer entre les deux modes de connaissance fondamentalement différents que constituent la monographie individuelle et le dénombrement exhaustif, il insiste de façon surprenante sur le fait qu'il pense pouvoir faire aussi bien que les monographies sur *leur terrain* (« le sang, les chairs, les nerfs »), alors qu'ensuite les sondages seront comparés aux recensements exhaustifs (en matière de coût et de prévision) mais sûrement pas aux monographies. Ceci montre la prégnance d'un mode de connaissance fondé sur l'intuition de la totalité de la personne :

En discutant les rôles réciproques des monographies et de la statistique partielle, on a dit que la monographie s'occupe d'objets qu'on ne peut ni compter, ni peser, ni mesurer, tandis que la statistique partielle s'occupe « d'objets qui par eux-mêmes pourraient être comptés dans leur totalité mais qui, de propos délibéré, ne sont comptés qu'en partie »...

En général, je crois pouvoir appliquer aux investigations partielles et surtout aux investigations représentatives les paroles éloquentes que notre très honoré collègue, M. Bodio, a prononcées à Berne concernant le travail de notre regretté Dr Engel sur les budgets de familles ouvrières. « La monographie statistique et le dénombrement sont deux manières d'investigation des faits sociaux qui se complètent l'une par l'autre. Le dénombrement, à lui seul, ne peut donner que les profils généraux des phénomènes, la silhouette, pour ainsi dire, des figures. La monographie » — et moi j'y ajoute l'investigation partielle en général — « permet de pousser l'analyse dans tous les détails de la vie économique et morale du peuple, donne le sang, les chairs, les nerfs au squelette bâti par la statistique générale, et à son tour le dénombrement complète les notions fournies par la monographie ».

LA PARTIE POUR LE TOUT COMMENT GENERALISER ?

En y insérant le mot « d'investigation partielle », je trouve dans ces paroles une démonstration excellente des rôles réciproques des investigations partielles et de la statistique générale. (Ibid., pp. 182-183)

Il décrit alors son outil idéal, aussi riche que la monographie et aussi précis que le dénombrement exhaustif, pour peu que soit respectée la *contrainte de représentativité* (dont il a l'intuition juste, mais pas encore l'outillage) :

La valeur scientifique des investigations partielles dépend beaucoup plus de leur caractère représentatif, que du nombre des données. Il arrive souvent que les données qui sont faciles à obtenir représentent plutôt une élite que les types ordinaires. (Ibid., p. 183)

On vérifie a posteriori que la procédure suivie est bonne si les variables contrôlables ne diffèrent pas trop entre l'échantillon et le recensement :

Dans la même mesure, dans laquelle l'investigation partielle s'est montrée correcte en ce qui concerne les points qui ont pu être contrôlés, elle est probablement correcte aussi en ce qui concerne les points qui ne peuvent pas être contrôlés à l'aide de la statistique générale. (Ibid., p. 183)

L'idée que les théorèmes probabilistes formulés dès le début du XIX^e siècle pourraient permettre d'en dire plus sur les « erreurs probables » qu'entraînent des tirages d'échantillons aléatoires (et donc sur la significativité des écarts observés par Kiaer) ne vient alors à personne. Sa procédure est encore ancrée dans une solide connaissance du terrain, le contrôle n'intervenant qu'après coup. La déterritorialisation et la mathématisation des procédures ne viendront qu'ensuite.

La mise en œuvre des schèmes probabilistes n'apparaît timidement qu'en 1901, dans une nouvelle discussion autour de la méthode de Kiaer, quand l'économiste allemand Bortkiewicz affirme avoir utilisé des « formules déduites pour des cas analogues par Poisson, pour savoir si la différence entre deux nombres est fortuite ou non », et constate que, dans les cas présentés par le Norvégien, ce n'était pas le cas, et que *les écarts étaient significatifs*. Ainsi l'échantillon de ce dernier n'était pas aussi représentatif qu'il le pensait. En apparence, Bortkiewicz a porté un rude coup à Kiaer. Pourtant, curieusement, dans la suite du débat, personne ne reprend l'argumentation de Bortkiewicz et on ne connaît même pas la réaction de Kiaer. Peut-être Bowley en a-t-il eu vent, puisque cinq ans plus tard, en 1906, il présente devant la Royal Statistical Society les premiers calculs d'intervalles de confiance (Bowley 1906).

Comment articuler « ce que l'on sait déjà » avec le hasard ?

L'idée selon laquelle on peut garantir la représentativité d'un échantillon grâce à des « variables de contrôle » subsiste néanmoins une trentaine d'années à travers la méthode dite du « choix judicieux » (« purposive selection »). Celle-ci continue à privilégier le découpage territorial (héritage de l'époque précédente) de l'espace national en un ensemble de districts, dont on sélectionne un sous-ensemble, non pas aléatoirement, mais de façon qu'un certain nombre de variables essentielles (les variables de contrôle) aient les mêmes valeurs pour ce sous-ensemble et pour le territoire complet ⁶.

Un usage remarquable de cette méthode est présenté en 1928 par l'Italien Corrado Gini. Ayant eu à se débarrasser, pour des raisons d'encombrement, des bulletins individuels du recensement de 1921, il eut l'idée d'en conserver une partie, portant sur 29 arrondissements (sur les 214 que comptait

6. Une trace de cette idée subsiste quand, au début des grandes soirées électorales, on livre les résultats de tel village qui, dans le passé, a toujours voté « comme la France entière ».

LA PARTIE POUR LE TOUT : COMMENT GÉNÉRALISER ?

l'Italie), pour laquelle les moyennes de 7 variables soient proches de celles du pays entier (natalité, mortalité, nuptialité, part de la population agricole, part de la population agglomérée, revenu moyen, altitude moyenne au-dessus du niveau de la mer). Le choix des 29 arrondissements respectant au mieux ces contraintes fut fait à l'issue d'un laborieux tâtonnement, et Gini en fait lui-même la critique, montrant que, à moins d'hypothèses très particulières de linéarité des corrélations entre variables contrôlées et non contrôlées, il n'y a aucune raison pour que cet échantillon soit un bon substitut de l'Italie entière.

Toute cette discussion, qui s'étend du rapport de Jensen en 1925 à l'article de Neyman en 1934, porte en fait sur la question de l'articulation entre le tirage aléatoire pur et « ce que l'on sait déjà par ailleurs » (par exemple par le recensement). Ceci a conduit successivement aux méthodes de « variables de contrôle » de Kiaer, puis du « choix judicieux », chacune d'elles étant successivement rejetée pour que soit finalement adoptée la technique de l'échantillonnage *stratifié* selon des découpages a priori de la population, supposés résumer *ce que l'on sait déjà*, c'est-à-dire qu'il existe des différences significatives de moyennes entre les classes, et qu'on améliore donc la précision des estimations globales en stratifiant a priori. Ceci suppose donc que de telles nomenclatures, véritable dépôt du savoir antérieur, existent, aient une certaine pérennité et solidité, et inspirent confiance : on sait comment, après 1950, les catégories socioprofessionnelles, les niveaux de formation, les catégories de communes, les types de familles, etc. ont pu jouer un tel rôle. La construction de la machinerie de la représentativité passait donc d'une part par celle de l'édifice mathématique, peu à peu épuré des vieilles « variables de contrôle » (par Neyman), et d'autre part par celle d'un système de nomenclatures inscrivant les qualités des personnes dans des cadres garantis par un État dépositaire de l'intérêt général et élaborés par une institution suscitant la confiance.

Ce point est bien mis en lumière dans le rapport de Jensen de 1925 sur la méthode représentative. Observant que cette méthode suscite encore de la méfiance parce qu'elle ne porte que sur une partie de la population, il se demande si il suffit, pour lever ce type de critique, que l'administration statistique inspire confiance. Le fait que les deux catégories de problèmes, techniques et socio-politiques, soient évoqués en même temps est un élément de réponse par rapport à la question initiale (pourquoi la méthode représentative n'a-t-elle pas été utilisée plus tôt?) :

« Le cœur de cette objection est que la plus grande importance doit être attachée à la confiance mutuelle entre les institutions statistiques officielles et la population qui, à la fois, fournit le matériel pour la statistique, et pour qui tout ce travail est fait. La statistique officielle doit naturellement être extrêmement attentive à sa réputation : « il n'est pas suffisant que la femme de César soit vertueuse, mais tout le monde doit être convaincu qu'elle l'est ». Mais il serait difficilement justifiable, en dehors de toute question de prestige, d'empêcher un progrès technique, dont la justification en soi est pleinement reconnue. On ne s'interdit pas de construire un pont qui a un plan précis sous prétexte que le public, dans son ignorance n'a pas confiance en ce plan; on le construit quand l'ingénieur peut garantir sa résistance, et le public l'utilisera alors normalement et aura confiance dans sa solidité. » (Jensen 1925, p. 374) [traduit par nous]

Le problème de Jensen dans ce texte est d'articuler la solidité technique de l'objet et sa réputation : c'est cette articulation qui fait la force de la statistique d'État.

État Providence, marché national, prévisions électorales

Mais, à partir des années 30, ce cas de figure typiquement européen, de l'utilisation de la méthode représentative à des fins de gestion des problèmes sociaux par l'État, ne va plus être le seul, dès lors que, au moins dans deux autres cas, surtout aux États-Unis, certaines totalisations portant sur tout le territoire national deviennent directement pertinentes : les études de marché de biens de consommation et les prévisions électorales.

LA PARTIE POUR LE TOUT · COMMENT GÉNÉRALISER ?

Dans ces deux cas, il fallait qu'une standardisation et une mise en équivalence nationales des produits soient faites préalablement. Pour les biens de consommation il était nécessaire que des entreprises de grande taille puissent diffuser régulièrement des produits standard sur tout le territoire grâce à un réseau de transport national, et que ces produits soient clairement identifiés (Eymard-Duvernay 1986) : il devenait alors possible de faire un sondage national pour savoir si les consommateurs préféraient le Coca ou le Pepsi.

Pour les prévisions électorales, il importait que les candidats soient les mêmes sur tout le pays (ce qui était le cas pour les élections présidentielles américaines, mais non pour le scrutin d'arrondissement français) et que leurs images soient relativement diffusées et unifiées, ce qui commençait à être vrai grâce à la radio. Il fallait aussi que la base de sondage soit la plus proche possible de l'électorat : on connaît les mésaventures des sondages par téléphone, qui, s'adressant aux personnes aisées possédant cet équipement, annonçaient à tort la victoire des Républicains.

Mais, ce qui est commun à ces différents cas, c'est que les enquêtes sont faites pour être *livrées à d'autres*, qui les utilisent à des fins opératoires (administration, grandes entreprises, chaînes de radios ou de journaux). La représentativité est alors la condition permettant d'articuler la limitation du *coût* de ce savoir et sa *pertinence*, à la fois technique et socialement reconnue. Dans tous ces cas, ce sont aussi des *individus* qui sont concernés (secourus, consommateurs ou électeurs) et non plus des totalités comme l'ordre divin de Quetelet, la lignée de Le Play ou la conscience ouvrière d'Halbwachs.

La monade et les lacets de chaussure

Le passage d'un mode de pensée à un autre, à l'aide duquel, dans ce texte, on a tenté d'interpréter l'apparition, au tournant de ce siècle, d'une nouvelle conception de la représentativité, a été présenté, dans le passé, de maintes façons. L'une d'elles, dont Louis Dumont se fait l'écho dans le texte cité en exergue, oppose des « cultures nationales » réputées incommensurables les unes aux autres (tradition allemande) à une « civilisation » universelle ou du moins universaliste (tradition anglaise pour la version économiste, ou française pour la version politique). Dumont indique que, pour résoudre cette contradiction, Leibniz avait imaginé un « système monadique » dans lequel « chaque culture exprimait à sa manière l'universel », façon habile de généraliser.

L'histoire de la notion de représentativité nous a renvoyé à celle de *généralisation* : qu'est-ce que la partie ? qu'est-ce que le tout ? On a vu que les deux opérations, tant politiques que cognitives, de délimitation d'un tout pertinent (la Nation ou le marché, selon les cas) et de définition et de mise en équivalence des unités de ce tout (les citoyens ou les consommateurs, selon les cas) étaient indispensables pour que l'idée moderne et l'outillage de la représentativité puissent apparaître.

On observe, sans y insister, que l'on a dédoublé, ci-dessus, Nation et marché, citoyen et consommateur. En effet, dans tout ce qui précédait, on avait assimilé, pour simplifier l'exposé, deux façons de faire équivalence qui sont en fait très distinctes (Boltanski et Thévenot 1987).

Le problème posé par les transformations de la notion de représentativité, et en particulier par les discussions de la période 1895-1935, est bien de savoir comment articuler entre elles des connaissances produites selon des registres différents. C'est à ce problème que tentaient de répondre successivement les « variables contrôlées » de Kiaer, le « choix judicieux » de Jensen et Gini, la « stratification » de Neyman. Ceci renvoie aussi au problème théorique, vieux comme le calcul des probabilités, de choix entre l'acceptation ou le rejet de « lois a priori ».

Au cours de la discussion de 1925, Lucien March, alors âgé de 66 ans, et qui a vu basculer toute la pensée statistique, évoque ce problème en termes qui semblent à la fois fins et archaïques, car toutes ces choses ont été, depuis, formalisées, c'est-à-dire, d'une certaine manière, figées :

LA PARTIE POUR LE TOUT : COMMENT GÉNÉRALISER ?

Le système qui consiste à se rapprocher autant que possible du tirage au sort des unités de l'enquête n'est pas nécessairement celui qui correspond le mieux à l'idée de représentation. Comme nous l'avons remarqué, il suppose que les unités ne présentent entre elles aucune différence. Or, la méthode représentative a pour but de mettre en relief les différences. L'hypothèse semble donc quelque peu contradictoire avec le but poursuivi. De là des préférences pour un choix mieux compris, plus intelligent. Par exemple, certaines personnes écartent les cas extrêmes réputés anormaux, ou bien elles opèrent le choix du spécimen en se basant sur quelque critérium jugé essentiel. (March 1925, p. 449)

Quand March parle de « différence », en soulignant l'apparente contradiction entre la nécessité de mise en équivalence des unités et la recherche de leurs différences, on pourrait lui répondre que les formalisations ultérieures ont parfaitement résolu ce problème. Mais ses hésitations, pour savoir s'il faut faire un choix « mieux compris, plus intelligent », s'il faut éliminer les « cas anormaux », reflètent bien un problème que le statisticien, *s'il connaît déjà bien son terrain*, rencontre toujours.

Ce « savoir préalable » est souvent acquis par contact direct, par familiarité (mais pas toujours, comme on le voit dans le cas des « variables contrôlées » par la statistique générale). Il est alors du même ordre que la « culture » dont parle Dumont, la « communauté » de Tönnies. Il y a toujours, pour l'observateur l'intuition d'une connaissance globale de la totalité d'une situation, d'une personne, d'un groupe.

« Il faut recomposer ce qui a été décomposé » dit parfois le chercheur à l'issue de son travail analytique. C'est encore le problème de la généralisation : refaire du tout dans son unité. Remonter à la totalité à partir d'une de ses parties : ce fantasme ne cesse d'agiter le savant, et, dans les développements les plus récents des méthodes de sondage, des techniques de simulation, permises par les gros ordinateurs, permettent de feindre cette reconstruction, en générant un grand nombre de sous-échantillons à partir de l'échantillon, et en étudiant la distribution des statistiques ainsi engendrées. Cette méthode a été baptisée « bootstrap », car elle évoque le rêve de se soulever soi-même de terre en tirant sur ses lacets de chaussure. Ce rêve de reconstruction de l'univers à partir d'une de ses parties n'est pas sans analogie avec la « procédure compliquée et laborieuse » suggérée par Leibniz pour passer d'une culture singulière à l'universel. Ce qui a changé, c'est la machinerie cognitive et politique.

ANNEXE

La motion de l'IIS sur la méthode représentative (Rome 1925). Rapport de Jensen.

Méthode représentative.

L'Institut International de Statistique :

Considérant que, dans bien des cas, il est impossible de recueillir des données statistiques complètes sur un sujet d'étude, et que l'on est, par conséquent, amené à formuler des conclusions générales d'après des enquêtes partielles;

considérant que, même dans les cas où l'on dispose de données complètes pour l'étude principale que l'on a en vue, on peut obtenir des enseignements supplémentaires en mettant en œuvre une partie de ces données, pourvu que l'on opère avec méthode; et

LA PARTIE POUR LE TOUT. COMMENT GÉNÉRALISER ?

considérant que l'économie de travail, de temps et d'argent que procure la limitation de l'enquête à une partie des données, permet souvent d'utiliser les renseignements recueillis plus complètement que si l'on opérait sur l'ensemble de ces données, et ainsi d'étudier plus profondément le sujet en question;

I. rappelle la résolution adoptée lors de la session de Berlin en 1903, et appelle de nouveau l'attention sur l'avantage considérable que l'on peut obtenir par l'application de la méthode représentative, dans les conditions ci-après;

pour que les résultats d'une enquête partielle puissent être légitimement généralisés, la fraction retenue comme un spécimen de l'ensemble dont il fait partie doit être suffisamment représentative de cet ensemble. A cet effet, le spécimen peut être pris dans la masse de différentes manières; cependant il convient de distinguer deux cas principaux :

A) *Prise au hasard*. On prend un certain nombre d'unités dans des conditions telles que les unités de l'ensemble soient également susceptibles d'être comprises dans ce nombre. Dans ce cas le degré de précision dépend en grande partie du nombre adopté, celui-ci devant être suffisant pour que les déviations accidentelles puissent être insignifiantes.

B) *Choix judicieux*. Les unités de l'ensemble étant réparties par groupes, le choix porte sur des groupes tels qu'ils présentent, au total, à peu près les mêmes caractéristiques que l'ensemble. Afin d'obtenir quelque idée de la précision des estimations qui résultent de ce choix, le nombre des groupes doit être suffisant pour que l'on puisse mesurer la variation des éléments caractéristiques d'un groupe à l'autre. Comme la précision des estimations dépend beaucoup de la prudence appliquée au choix du spécimen, les précautions suivantes sont recommandées :

1. Comparer deux ou plusieurs spécimens obtenus par l'application des mêmes procédés, et, le cas échéant, joindre les spécimens choisis. (Cette manière de procéder peut être également recommandable en cas de prise au hasard.)

2. Réexaminer minutieusement, en faisant des enquêtes répétées de temps à autre, la relation du spécimen retenu avec l'ensemble;

II. recommande de diriger l'étude de façon à permettre autant que possible le calcul de la précision des résultats. Ces résultats devraient être fournis avec l'indication des limites de l'erreur dont ils sont susceptibles;

III. rappelle, en insistant sur son importance, le vœu formulé dans la résolution de 1903 tendant à ce que la publication des résultats d'une enquête représentative comporte obligatoirement dans tous les cas un compte-rendu détaillé des procédés employés pour le choix du spécimen.

Bulletin de l'IIS — 1925.
pp. 212-213

BIBLIOGRAPHIE

- ABRAMS P. (1968). *The origins of British Sociology. 1834-1914*. University of Chicago Press. Chicago and London.
- BOARD OF TRADE (1909). *Cost of living in French towns. Report of an inquiry by the Board of trade*. Darling and Son-London.
- BOLTANSKI L. et THÉVENOT L. (1987). *Les économies de la grandeur*. Cahiers du CEE. Série Protée. N° 1.
- BOWLEY A. (1906). *Presidential address to the economic section of the British Association*. *Journal of the Royal Statistical society* (Sep.), pp. 540-558.
- CHANG Wei-ching (1976). *Statistical theories and sampling practice*. In : Owen D. (1976).
- CHEYSSON E. (1890). *Les budgets comparés de cent monographies de famille*. Avec A. Toque. Rome, Botta.

LA PARTIE POUR LE TOUT : COMMENT GÉNÉRALISER ?

- DESROSNIÈRES A. (1985). Histoires de formes : statistiques et sciences sociales avant 1940. *Revue française de sociologie*. Vol. XXVI, pp. 277-310.
- DESROSNIÈRES A. (1986). L'ingénieur d'état et le père de famille : Émile Cheysson et la statistique. *Annales des mines. Série Gérer et comprendre*. N° 2. pp. 66-80.
- DUGE DE BERNONVILLE L. (1916). Enquête sur les conditions de la vie ouvrière et rurale en France en 1913-1914. *Bulletin de la Statistique Générale de la France*. Tome IV.
- DUMONT L. (1983). Essai sur l'individualisme. Seuil. Paris.
- EYMAR-DUVERNAY F. (1986). La qualification des produits. In : R. Salais et L. Thévenot, eds : Le travail. Marchés, règles, conventions. INSEE Economica — Paris.
- FRLCHET M. et HALBWACHS M. (1924). Le calcul des probabilités à la portée de tous. Dunod. Paris.
- GINI C. (1928). Une application de la méthode représentative aux matériaux du dernier recensement de la population italienne (1^{er} décembre 1921). *Bulletin de l'Institut International de Statistique*. Vol. 23, tome 2, pp. 198-215.
- HALBWACHS H. (1912). La classe ouvrière et les niveaux de vie. Alcan. Paris.
- HALBWACHS H. (1913). La théorie de l'homme moyen. Essai sur Quetelet et la statistique morale. Alcan. Paris.
- HALBWACHS H. (1914). Budgets de familles ouvrières et paysannes en France, en 1907. *Bulletin de la Statistique Générale de la France*, 4, fasc. 1, pp. 47-83.
- HANSEN M.H. et MADON W.G. (1976). Some important events in the historical development of sample surveys. In Owen D. (1976).
- HENNOCK E.P. (1976). Poverty and social theory in England : the experience of the eighteen-eighties. *Social History* (1), pp. 67-9.
- HENNOCK E.P. (1987). The measurement of poverty : from the metropolis to the nation, 1980-1920. *Economic History Review*, 2nd ser. XL, 2, pp. 208-227.
- JENSEN A. (1925a). Report on the representative method in statistics. *Bulletin de l'Institut International de Statistique*. Vol. 22. Livre 1, pp. 359-380.
- JENSEN A. (1925b). The representative method in practice. Annexe B. *Bulletin de l'Institut International de Statistique*. Vol. 22. Livre 1, pp. 381-489.
- KIAER A.N. (1895). Observations et expériences concernant les dénombrements représentatifs. *Bulletin de l'IIIS*. Vol. 9, pp. 176-78.
- KIAER A.N. (1897). Sur les méthodes représentatives ou typologiques appliquées à la statistique. *Bulletin de l'IIIS*. vol. 11, pp. 180-185.
- KIAER A.N. (1901). Sur les méthodes représentatives ou typologiques. *Bulletin de l'IIIS*. Vol. 13. Livre 1, pp. 66-70.
- KRUSKAL W. et MUSTELLER F. (1980). Representative Sampling IV : the history of the concept in statistics. 1895-1939. *International Statistical Review*, 48, pp. 169-195.
- MAC KENZIE D. (1981). Statistics in Britain. 1865-1930. The social construction of scientific Knowledge. Edinburgh University Press. Edinburgh.
- MARCH L. (1925). Observations sur la méthode représentative et sur le projet et de rapport relatif à cette méthode. Annexe D. *Bulletin de l'IIIS*. Vol. 22. Livre 1, pp. 444-451.
- NEYMAN J. (1934). On the two different aspects of the representative method : the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, pp. 558-606. Discussion 607-625.
- NISBET R.A. (1966). La tradition sociologique, traduction en français : P.U.F. 1984. Paris.
- OWEN D. (1976). On the history of statistics and probability. Marcel Dekker ed. New York.
- POLANYI K. (1944). La grande transformation : aux origines politiques et économiques de notre temps. Traduction en français : Gallimard. 1983.
- SENG You Poh (1951). Historical survey of the development of sampling theories and practice. *Journal of the Royal Statistical Society*, series A, 114, pp. 214-231.
- STEPHAN F. (1948). History of the uses of modern sampling procedures. *Journal of the American Statistical Association*, 43, pp. 12-39.
- THÉVENOT L. (1987). Forme statistique et lien politique : éléments pour une généalogie des statistiques sociales (note INSEE, Unité de Recherche, n° 112/930 du 9-4-1987). A paraître.