
Ce que dit un tableau et ce qu'on en dit

Remarques sur le langage des variables et l'interprétation dans les sciences sociales

Jean-Claude Passeron

CNRS

Résumé : La distance évidente qui sépare la catégorisation statistique de la conceptualisation sociologique est souvent traitée par les sociologues comme une distance hiérarchique qui séparerait une approximation opératoire et formelle de la saisie signifiante et substantielle des phénomènes, geste suprême réservé à la sociologie. Le contenu corporatiste d'un tel programme ne devrait pas dissimuler aux sociologues son risque essentiel, celui de pratiquer l'auto-suffisance théorique d'une sociologie qui n'aurait de comptes à rendre qu'aux constats qui la confortent et qui se réserverait de porter au passif des imperfections mécaniques de l'instrument statistique ce que celui-ci ne vérifie pas des constructions conceptuelles du discours sociologique. À la limite, la signification du raisonnement statistique ne pourrait lui advenir que de l'extérieur : ce serait toujours à lui de s'amender pour mériter de servir par ses « constats illustratifs » des énoncés sociologiques qui tirent d'ailleurs leur « évidence ». De leur côté, les statisticiens ne sont évidemment pas en reste en matière d'épistémologie corporative. Accoutumés aux exigences du recueil et du traitement de l'information et sachant ce qu'il en coûte d'arriver à homogénéiser les données économiques et sociales, ils sont inévitablement enclins, pour préserver l'univocité des énonciations portant sur des constats de recensement ou de corrélation si chèrement acquis, à une défiance généralisée envers tout changement du langage d'énonciation des constats de base, autrement dit à marquer une réticence de principe envers l'interprétation conceptuelle, toujours suspecte de surinterprétation polysémique. À la limite, les « langues artificielles », comme la langue tabulaire du tableau croisé ou la langue graphique des plans factoriels, seraient les seules à ne pas déformer les énoncés d'observation et de traitement, dont l'expression en « langue naturelle » majorerait toujours le sens de manière incontrôlée et incontrôlable. Il y a pourtant un accord latent entre ces deux épistémologies coutumières lorsqu'elles atteignent leur forme limite : elles semblent bien convenir que le discours statistique et le discours sociologique diffèrent intrinsèquement par leur nature assertorique. On part ici, tout au contraire, du postulat épistémologique que toutes les conceptualisations opérées à partir de l'observation du monde historique possèdent, en tant qu'abstractions scientifiques, une pertinence empirique commune ou, si l'on veut, que les énonciations des différentes sciences sociales ne peuvent avoir qu'une seule indexation de vérité : l'observation historique par quelque méthode qu'on l'opère, même si elles diffèrent par la logique des raisonnements qui mettent en œuvre les constats issus de cette observation.

Date de publication : 8/7/2021

Dossier : Sociologie et statistique

Comment citer : 10.52983/crev.vi0.35

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

Ce que dit un tableau et ce qu'on en dit

Remarques sur le langage des variables et l'interprétation dans les sciences sociales

Jean-Claude Passeron

CNRS

« — Tu sais pas ce que tu dis.
— Si on savait ce qu'on dit, on saurait tout. »
Raymond Queneau

Statistique et sociologie

La distance évidente qui sépare la catégorisation statistique de la conceptualisation sociologique est souvent traitée par les sociologues comme une distance hiérarchique qui séparerait une approximation opératoire et formelle de la saisie signifiante et substantielle des phénomènes, geste suprême réservé à la sociologie. Le contenu corporatiste d'un tel programme ne devrait pas dissimuler aux sociologues son risque essentiel, celui de pratiquer *l'autosuffisance théorique* d'une sociologie qui n'aurait de comptes à rendre qu'aux constats qui la confortent et qui se réserverait de porter au passif des imperfections mécaniques de l'instrument statistique ce que celui-ci ne vérifie pas des constructions conceptuelles du discours sociologique. Bien des discussions sur les rapports entre la grille socio-professionnelle et le concept de « classe sociale » trahissent ainsi la certitude inaltérable que c'est à l'instrument de mesure, et à lui seul, qu'il incombe de faire tout le chemin pour rejoindre les exigences d'une démarche conceptuelle *déjà* « parfaite » en son genre énonciatif. À la limite, la signification du raisonnement statistique ne pourrait lui advenir que de l'extérieur : ce serait toujours à lui de s'amender pour mériter de servir par ses « constats illustratifs » (*instantial statements*) des énoncés sociologiques qui tirent d'ailleurs leur « évidence ».

De leur côté, les statisticiens ne sont évidemment pas en reste en matière d'épistémologie corporative. Accoutumés aux exigences du recueil et du traitement de l'information et sachant ce qu'il en coûte d'arriver à homogénéiser les données économiques et sociales, ils sont inévitablement enclins, pour préserver l'univocité des énonciations portant sur des constats de recensement ou de corrélation si chèrement acquis, à une défiance généralisée envers tout changement du langage d'énonciation des constats de base, autrement dit à marquer une réticence de principe envers l'interprétation conceptuelle, toujours suspecte de surinterprétation polysémique. Le discours sociologique apparaît alors comme un tissu arbitraire de reformulations « majorantes » ou sélectives, tout juste bon à suggérer, par raccroc, des correctifs ou des additifs à la catégorisation ou à la mesure empirique. Plus généralement, si la plupart des statisticiens font droit à l'interprétation, c'est comme à une opération extrinsèque, à un exercice intuitif dont les normes resteraient distinctes de celles du langage dans lequel s'énoncent les constats empiriquement contrôlés.

À la limite, les « langues artificielles », comme la langue tabulaire du tableau croisé ou la langue graphique des plans factoriels, seraient les seules à ne pas déformer les énoncés d'observation et de traitement, dont l'expression en « langue naturelle » majorerait toujours le sens de manière incontrôlée et incontrôlable.

Le sens du vecteur scientifique entre énonciation statistique et énonciation sociologique est alors tout simplement inversé.

Il y a pourtant un accord latent entre ces deux épistémologies coutumières lorsqu'elles atteignent leur forme limite : elles semblent bien convenir que le discours statistique et le discours sociologique diffèrent intrinsèquement par leur nature assertorique, *i. e.* qu'ils ne sont pas justiciables, pour trancher de la vérité ou de la fausseté de leurs énoncés, de procédures de contrôle référant leurs expériences de dernière instance à la même « réalité ».

On part ici, tout au contraire, du postulat épistémologique *que toutes les conceptualisations opérées à partir de l'observation du monde historique possèdent, en tant qu'abstractions scientifiques, une pertinence empirique commune* ou, si l'on veut, que les énonciations des différentes sciences sociales ne peuvent avoir qu'une seule indexation de vérité : *l'observation historique par quelque méthode qu'on l'opère*, même si elles diffèrent par la logique des raisonnements qui mettent en œuvre les constats issus de cette observation. Il faut alors expliquer en quoi les constats et les traitements que l'instrumentation statistique permet d'opérer, dans les sciences sociales, conformément aux normes du raisonnement expérimental entendu au sens strict, se distinguent des opérations mises en œuvre par un raisonnement sociologique qui dit effectivement plus et autre chose, dans la moindre de ses énonciations, que ce que dit le tableau statistique correspondant, du moins dès que l'énonciateur entreprend de traduire ses énoncés tabulaires en syntagmes conceptuels, c'est-à-dire en phrases d'une langue naturelle se référant nécessairement par les mots employés à des données extérieures au tableau (interprétation conceptuelle).

La thèse proposée ici est que le raisonnement sociologique se distingue du raisonnement expérimental (qui ne peut procéder qu'à des comparaisons fondées sur des constats opérés « toutes choses étant égales par ailleurs ») non parce qu'il se référerait à des constats d'une autre nature que ceux qui sont accessibles à l'observation historique mais parce qu'il doit adopter une démarche spécifique de composition des « énoncés de base » qui lui est imposée par la nécessité d'énoncer ses généralités en prenant appui sur des constats *qui ne sont jamais comparables sous tous les rapports* : ce qui l'oblige à incorporer à l'interprétation des constats qu'il énonce un discours sur la variation de leurs contextes et de la production de l'information qu'il enferme. Autrement dit, le *raisonnement sociologique* se distingue du *raisonnement statistique* par la *logique* de composition de ses énoncés (agencement des références à l'empirie) mais non par son *sens assertorique* (critères de la validité empirique).

La différence entre raisonnement statistique et raisonnement sociologique tient, selon nous, aux deux usages que l'on peut faire du raisonnement expérimental dans les conditions de l'observation historique :

- (a) soit le pratiquer *inconditionnellement*, ce qui ne peut se faire dans les sciences sociales qu'en oubliant *momentanément* les conditions du prélèvement d'information permettant cet exercice (succession des choix d'enquête, d'échantillonnage, de nomenclature et de traitement; liens entre le contenu de l'information et les formes techniques ou linguistiques de son recueil; singularité historique du contexte des interactions entre variables traitées) ;
- (b) soit revenir sur ces *conditions de production de l'information* pour les incorporer à l'énonciation des constats, c'est-à-dire à la connaissance de « l'objet » qu'elles ont produit, seul moyen d'énoncer plus généralement

de manière empirique sur des *séries d'objets construits différemment*. L'observation historique, qui exclut la maîtrise formelle des contextes et, du même coup, l'univocité du sens assertorique des variables traitées, interdit au langage des variables d'être par lui-même un langage de description du monde historique; ces conditions n'excluent pas des moments de raisonnement expérimental mais elles obligent, si l'on veut appliquer les résultats d'un tel raisonnement au monde d'où provient l'information qu'il traite, à des raisonnements sur les conditions limitatives de chacun des raisonnements expérimentaux, à des constats sur les conditions des constats, dont la composition n'est pas autre chose qu'un raisonnement sociologique.

¹ Les concepts sociologiques résument aussi des constats qui sont d'une nature autre que statistique, par exemple historiques ou anthropologiques. On ne développera pas ici les problèmes logiques posés par cette indexation simultanée des concepts sociologiques sur des formes différentes de l'observation historique.

Le premier corollaire

de cette thèse est que tout acte d'énonciation sur le monde historique implique nécessairement le passage au raisonnement sociologique et à ses risques (ses approximations logiques), y compris chez le statisticien. Du fait de l'impossibilité de pratiquer en toute rigueur l'induction dans les sciences de l'observation historique, le raisonnement statistique ne peut se confondre avec le raisonnement expérimental que dans ses moments d'*autonomisation*.

L'énonciation statistique devient *ipso facto* énonciation sociologique, fût-ce à son insu, fût-ce sur le tableau croisé le plus simple, dès qu'elle entreprend d'affirmer ou de nier quelque chose à propos du monde historique, en lui appliquant des énoncés statistiques dont le sens devient alors solidaire de l'énonciation du contexte de la mesure ainsi que des conditions du recueil et de la catégorisation de l'information traitée.

Le deuxième corollaire

est que l'énonciation sociologique ne dispose d'aucun droit régalien d'interprétation qui lui permettrait d'incorporer à ses reformulations d'énoncés statistiques des concepts synthétisant d'autres informations que celles procurées par des constats empiriques — qu'ils soient de nature statistique ou non¹.

Le raisonnement sociologique ne reste un raisonnement scientifique que dans la mesure où il s'astreint à ne transformer par l'énonciation conceptuelle ce que dit un constat statistique que par référence à d'autres constats empiriques, qui ne sont pas cumulables avec lui au sens strict de la composition logique, mais qui lui sont assez « apparentés » pour que le raisonnement naturel puisse contrôler cette parenté. Le contrôle de la parenté des contextes est d'autant plus sûr qu'il s'appuie sur une méthodologie du raisonnement naturel qui n'est autre que celle du raisonnement historique, *i. e.* de la construction de concepts typologiques.

Sociologues et statisticiens ont donc finalement en commun de devoir, les uns et les autres, se prémunir à la fois contre *l'illusion du statisticien* et contre *l'illusion du sociologue*.

Autrement dit, contre *l'illusion expérimentaliste* que la langue statistique pourrait à elle seule et sans emprunter des signifiés à la langue sociologique dire quelque chose de général à propos du monde historique. Mais aussi contre *l'illusion herméneutique* que la langue sociologique pourrait puiser le surplus de sens qui différencie la conceptualisation sociologique des catégorisations de la langue statistique dans d'autres constats que ceux de l'observation historique (statistiques ou non).

Faute de dire ce qu'un tableau statistique ne dit jamais tout seul quand on lui fait dire quelque chose du monde d'où provient l'information qu'il renferme, on lui laisse toujours dire sur ce monde plus et autre chose qu'il ne dit (induction rampante). Puisqu'aucune énonciation procurant une connaissance du monde historique ne peut fonctionner dans un langage purement formel (celui-ci interdisant, par définition, de formuler ce que signifient, en chaque cas, les conditions de la formalisation pour la construction de l'information historique), le sens formel d'un constat de relations entre variables qui, dans les sciences sociales, n'est pas défendu contre la « compréhension sauvage » par l'intelligibilité « nomologique » (validité universelle des relations) appelle nécessairement des opérations de remplissage sémantique et de majoration énonciative qui, lorsque le sociologue ne les prend pas en charge, s'opèrent de toute manière, mais dans son dos. Que la charge du contresens soit laissée, comme souvent, au lecteur n'en décharge pas l'énonciateur.

Les conventions du langage des variables et le raisonnement sociologique

Parmi les nombreuses opérations qui jalonnent la construction d'une information statistique par les choix d'enquête et les instruments de traitement, on n'en retiendra ici qu'une seule. Il s'agit de la convention analogique qui permet de traiter une distribution de propriétés sociales comme une « variable ». Dans le travail sociologique, une telle convention oblige à opérer toute une série de constructions pour mesurer une variation historique par une variable et, surtout, à opérer toute une série de manipulations lors de l'échantillonnage pour raisonner commodément sur le croisement entre variables. Elle oblige en conséquence l'énonciation sociologique à *parcourir ce chemin en sens inverse* pour se prononcer sur le sens historique de variations et de co-variations livrées par le traitement statistique en langage de variables. On prendra trois exemples successifs pour illustrer le rôle constitutif de l'*interprétation* dont le raisonnement sociologique n'est pas autre chose qu'un contrôle énonciatif.

Autrement dit, les exemples veulent illustrer quelques-unes des démarches par lesquelles le raisonnement sociologique vise à combler, pour formuler des propositions historiquement signifiantes, les discontinuités du raisonnement expérimental qui sont constitutives des sciences de l'observation historique.

De la variable à la variation

Le rapport d'une variation sociale à la variable qui permet d'attester rigoureusement ses co-variations n'est jamais de synonymie. Un exemple trivial suffira à montrer que l'énonciation sociologique de ce qui varie dans la mesure que procure une variable rencontre d'emblée la nécessité d'interpréter la variable pour restituer au moindre constat son caractère d'assertion sur le monde historique. Aucun énoncé formulé dans le langage des variables ne permet d'énoncer un algorithme qui contrôlerait le rapport sémantique d'un « indicateur » à un « concept », puisque ce serait un énoncé qui se prononcerait en langage formel sur le sens des rapports entre langage formel et langage de l'assertion historique.

L'exemple — bien connu des chercheurs, mais dont on suppose trop vite qu'il n'illustrerait que le cas de quelques variables « ambiguës » — est celui de l'énonciation des variations qui apparaissent dans un tableau croisé où l'âge des sujets est pris comme variable indépendante. On sait qu'une variable comme celle de l'âge peut prendre au moins deux sens différents (« effet de vieillissement » ou « effet de génération ») lorsqu'on l'interprète pour énoncer

la variation sociologique que révèle la liaison statistique, c'est-à-dire lorsqu'on entreprend d'énoncer en langue naturelle ce que dit statistiquement un tel tableau.

Que « veulent dire », en effet, des tableaux qui présentent des distributions croisées où x est l'âge de N sujets, répartis en plusieurs tranches d'âge, et y la variation d'une mesure susceptible de prendre plusieurs valeurs chez ces N sujets?

Pour énoncer ce que dit un tableau de ce genre, *faut-il vraiment interpréter?* Ne peut-on se contenter, pour formuler un énoncé de science sociale, de dire ce que dit statistiquement le tableau de fréquences sans prendre aucun risque énonciatif?

Assurément, on peut toujours mettre en énoncés syntagmatiques ce que le tableau dit dans ses énoncés tabulaires (*i. e.* selon le code graphique de la localisation des co-occurrences dans un « espace de propriétés » à deux dimensions). On peut, sans passer à l'interprétation, dire quelque chose comme : la mesure montre que — *dans les conditions où elle a été effectuée et où l'information a été catégorisée* — les éléments appartenant au sous-ensemble x_1 (défini par la possession d'une propriété p_1) de l'ensemble x possèdent proportionnellement plus souvent la propriété p_1 définissant l'appartenance au sous-ensemble y_1 de l'ensemble y que les éléments appartenant au sous-ensemble x_2 (défini par la possession de la propriété p_2) de x ; de même pour les éléments de x_2 par rapport à x_3 .

Cette formulation constitue bien une lecture de ce qui est énoncé par une colonne et des formules analogues diraient ce que disent selon leur code tabulaire les autres lignes et colonnes du tableau croisé.

L'insignifiance historique de cette assertion, qui reste vraie et univoque sans qu'il soit besoin d'interprétation, *i. e.* quoi que ce soit qui se manifeste dans l'âge des sujets, ne tient pas au langage formel dans lequel elle s'énonce. Le passage du « langage formel » au « langage matériel » peut encore s'accomplir sans interprétation, pourvu qu'il n'engage qu'à la nomination des propriétés telles que les a constituées *concrètement* l'opération de mesure. Tant qu'elle n'engage pas à se prononcer sur le *sens conceptuel* qu'on assigne à la variation statistique, une énonciation en langue naturelle peut encore esquiver l'interprétation. On peut toujours dire sans risques : *c'est un fait que, dans l'échantillon considéré et dans les conditions où la mesure a été opérée, les individus possèdent d'autant plus souvent la propriété considérée qu'ils étaient plus âgés.* Mais, ce disant, on ne formule pas encore un énoncé sociologique.

L'interprétation — et donc l'énonciation sociologique (qu'elle soit formulée par un statisticien ou par un sociologue) — commence lorsqu'on entreprend ici de choisir une formulation conceptuelle spécifiant ce que mesure « l'âge des sujets », seul chemin qui permette de mettre un constat de ce type en relation avec d'autres constats. C'est parce que ceux-ci diffèrent inévitablement par leur contexte assertorique que le raisonnement sociologique doit, sans jamais pouvoir atteindre à la nécessité probatoire du raisonnement expérimental, les faire entrer dans une chaîne de présomptions empiriquement fondées. Il n'y a pas d'autre moyen de formuler une proposition sociologique, *i. e.* de se prononcer sur le sens historique de ce dont parle le tableau statistique sans en rien dire. Dans le tableau lui-même, aucune information ne permettra jamais de faire dire à la variable « âge » si la variation de propriétés qu'elle distribue et dont elle atteste l'objectivité est (a) une variation qui resterait liée de façon constante aux différents âges si on la mesurait successivement à différentes époques ou (b) s'il s'agit de la variation d'un système de propriété qui, indépendant de l'âge tant que l'histoire ne le transforme pas, manifeste son évolution historique au travers de l'âge des sujets parce que ceux-ci deviennent alors de faux contemporains, juxtaposés par la coupe synchronique de l'enquête comme autant de témoins de ses états successifs.

Il ne s'agit pas là d'une indétermination occasionnelle du tableau ou de la variable — même si on a choisi cet exemple, où l'interprétation achoppe d'emblée sur le sens du constat, pour la commodité de l'analyse. *Tout constat opéré dans le langage des variables suppose pour passer au constat sociologique un pari de ce type*, dont seul le raisonnement sociologique peut restreindre le degré de liberté interprétative en faisant appel à des comparaisons enfermant plus d'information que les constats logiquement composables parce que la conceptualisation historique autorise les rapprochements typologiques de plus grande ampleur que la composition logique².

D'une variable à l'autre

Le langage des variables n'est pas seulement un lexique permettant de coder univoquement les variations historiques, mais aussi une grammaire dont l'opération centrale définit les formes de la mise en relations des variables et dont le croisement sur un « tableau de contingence » représente une figure simple. Examinons un des problèmes qui se posent dans l'interprétation d'un croisement de variables, lors même qu'on se contente d'entendre le terme de « relation » au sens peu exigeant qu'enferme une affirmation du genre : *les éléments d'une population ont, du fait de la possession de certaines propriétés, des « chances » plus ou moins grandes d'avoir telles ou telles autres propriétés*. Même si on minimise ainsi la charge de retraduire sociologiquement ce qu'énoncent des constats statistiques de corrélation entre variables, on voit immédiatement surgir dans l'énoncé sociologique une équivocité autour de la définition de la « population » dont on affirme quelque chose, équivocité qui n'existe évidemment pas dans l'énoncé statistique correspondant.

L'exemple topique peut être ici celui de l'interprétation des résultats procurés par le traitement en « analyse multivariée » des effets d'une ou plusieurs variables sur un processus temporel de sélection lorsqu'on opère les constats sur des populations prises à différents stades dudit processus de sélection. Ainsi, par exemple, la lecture des résultats d'une enquête en sociologie de l'éducation visant à établir l'action de facteurs de sélection comme l'origine sociale ou le sexe aux différents niveaux des études ne peut jamais interpréter *directement* les fréquences différentielles de réussite, propres aux sujets d'origine sociale ou de sexe différents, observées dans les étapes successives d'un cursus scolaire. Pour énoncer, sans déformer par cette énonciation même, la relation qui unit l'appartenance de sexe ou de classe à des chances relativement plus ou moins fortes de succès scolaire, le raisonnement sociologique doit prendre en compte le fait (absent des relations explicitement formulées par le tableau croisé) que les populations sur lesquelles on établit la variation de la relation entre appartenance à une catégorie (de sexe ou de classe) et réussite scolaire sont progressivement sélectionnées tout au long du *cursus* avec une inégale sévérité sous le rapport même des variables dont on entend énoncer les effets sur cette réussite. Le *contextualisateur* de l'interprétation est évidemment ici le concept de « sur- et de sous-sélection scolaire relative » qui, en résumant d'autres constats (statistiques ou ethnographiques), permet d'incorporer à l'énonciation des relations entre origine sociale ou sexe et réussite scolaire *le fait* que les mesures successives s'opèrent sur des populations qui, par rapport aux populations de départ, ont été progressivement biaisées dans leur structure selon les différentes valeurs de ces deux variables. Peu importe d'ailleurs le modèle de langage naturel selon lequel on décrit ici la contextualisation nécessaire à l'interprétation : on pourrait le faire aussi bien dans les termes de la théorie du « filtre scolaire » et des capacités « filtrées » ou « non filtrées », en disant par exemple que les filtres appliqués par la sélection scolaire n'ont pas, selon les catégories soumises à la sélection, le même « maillage ». Les-

2 On pourrait montrer que le raffinement du protocole de mesure — par exemple celui de l'enquête longitudinale — ne permet jamais de trancher dans une logique purement expérimentale ce qui reste ici en suspens du sens de la variable. La difficulté d'interprétation s'estompe derrière le surarmement du protocole mais reste de même nature : l'influence du changement historique se composera encore, dans le temps biographique de sujets d'âge égal, avec l'effet de leur vieillissement individuel. Même le protocole le plus « coûteux » (plusieurs tranches d'âge dans chaque génération), s'il augmente l'information recueillie, ne suffit toujours pas à rendre mécanique l'interprétation : c'est en effet un fantôme logique après lequel on court, celui d'un dispositif expérimental qui se décalquerait *par sa vertu propre* en interprétation historique.

sentiel reste de restituer aux énoncés tabulaires une information capitale sur la production, par le protocole même de l'enquête, de l'information donnée par le tableau de mesures.

Mais l'exemple des effets sociaux induits par une sélection qui opère différemment selon les catégories de la population de départ — ou, si l'on veut, le trompe-l'œil d'un échantillon qui se confond inévitablement avec une population naturelle lorsqu'on ignore que ses biais concernent la question qu'on lui pose — ne représente d'ailleurs qu'un cas particulier d'un phénomène plus général, sur lequel achoppe presque toujours l'interprétation sociologique des relations. La plupart des populations sur lesquelles s'opèrent des mesures sont des échantillons préfabriqués par un processus social, c'est-à-dire des populations qui ne s'offrent à l'enquête et à la mesure que parce qu'une action sociale (et souvent institutionnelle) les a immatriculées en les prélevant, d'une manière qu'on ne connaît pas, sur une population originaire à laquelle on aura toujours tendance à se référer implicitement pour donner le contexte de validité des relations attestées par un traitement de variables opéré sur la population accessible.

Bref, le raisonnement sociologique a toujours pour fonction de s'interroger sur les conditions sociales de constitution des populations apparemment les plus naturelles afin de restituer à l'énonciation le fait qui interfère en toute généralisation, même et surtout silencieuse, qu'on n'observe jamais avec une lorgnette, fut-elle la plus sophistiquée et la mieux réglée, que ce qui passe dans son champ. C'est par exemple totalement évident dans le cas d'un *prélèvement institutionnel*. Les relations qu'on y observe entre des groupes d'âge, de classe ou de sexe et des pratiques ou des attitudes ne doivent évidemment pas donner à penser qu'on pourrait les généraliser hors de l'institution aux groupes de « même » nom. Ce ne sont pas « les femmes » — ni même les femmes au foyer —, « les jeunes », « les retraités » ou les différentes classes sociales qui livrent ici leurs orientations et leurs propensions mais, dans ces différents groupes, les prélèvements différemment sélectifs et différemment biaisés par cet échantillonnage institutionnel qu'opère pour chacun l'inscription dans l'institution ou sa fréquentation. Ceci est évident, mais n'empêche jamais complètement une induction rampante dans les énonciations sur l'influence du sexe, de l'âge, de la classe, etc. que suggère au travers du langage du croisement la force mécanique du langage des variables.

Il suffit en tout cas que l'action de prélèvement institutionnel soit plus discrète ou qu'elle impose l'évidence sociale de sa « neutralité » comme recensement administratif pour engendrer des propositions radicalement fausses sur l'action des variables qui conditionnent les chances de présence dans la population recensée, lorsque ces chances sont mécaniquement interprétées en termes de facteurs favorisant l'apparition de la pratique que la présence au sein de cette population nomme administrativement (crime, maladie, etc.). C'est, on le sait, la croix méthodologique de la sociologie de la délinquance (ou de la maladie mentale), qui ne fut d'ailleurs découverte et soumise à expertise théorique qu'à l'occasion d'une distorsion un peu trop grosse, comme celle du *white-collar crime*. Le cas de la sociologie criminelle est d'ailleurs topique en ce que cette « découverte » a déclenché pour longtemps une obsession méthodologique, celle du recensement du « chiffre noir », parfait exemple de la course au rattrapage impossible des conditions d'une mesure qui se ferait enfin « toutes choses étant égales par ailleurs »: une criminalité non recensée socialement ne peut s'ajouter mécaniquement (« toutes choses étant égales par ailleurs ») à la criminalité institutionnellement recensée pour constituer un ensemble sociologiquement homogène. C'est ici la pré-construction de la population délinquante par l'institution de recensement qui constitue le fait social auquel ont affaire le raisonnement sociologique et l'interprétation des relations entre variables.

Jeux de variables et configurations historiques

On pourrait être tenté de croire que toutes les difficultés dans l'interprétation des relations entre variables tiennent au fait que les échantillonnages sociaux viennent sans cesse mêler leurs effets parasites aux démarches de l'échantillonnage contrôlé. Suffirait-il de ne travailler que sur recensements exhaustifs pour voir s'évanouir toutes les ambiguïtés propres à l'énonciation d'une relation entre variables? Tout au contraire, c'est dans le cas des recensements que l'on voit apparaître dans sa forme la plus générale la difficulté centrale de l'énonciation des relations lorsqu'on applique cette énonciation au monde des phénomènes historiques.

À peine en effet le statisticien (cela échoit plus rarement au sociologue artisanal) dispose-t-il d'une population réelle exhaustivement recensée que son interrogation sur la relation entre deux variables le conduit, *dans le droit fil du raisonnement expérimental*, à constituer artificiellement une « population fictive » combinant de manière raisonnée les variables dont on veut éliminer l'effet parasite avec la variable indépendante dont on veut isoler les effets propres afin de neutraliser par la composition de la population fictive le fameux « effet de structure », bête noire des statisticiens qui entendent le bannir de leur raisonnement. Si on soupçonne que derrière la relation entre une variable x et une variable y se cache aussi l'effet d'une relation entre x et t , il faut en effet, pour pouvoir énoncer univoquement sur x, y , neutraliser la variable t , *i. e.* se donner les moyens d'observer x, y , à valeur constante de t . La démarche est formellement irréprochable puisque l'exigence qui l'engendre (raisonner « toutes choses égales par ailleurs ») constitue un des deux pôles du champ d'exigences entre lesquels se meut le raisonnement sociologique, l'autre étant le pôle historique qui rappelle sans cesse le raisonnement à l'évidence que les co-occurrences ne sont données dans la réalité qu'ainsi et pas autrement.

Mais le statisticien a normalement tous les moyens d'apercevoir la dérive à laquelle risque de l'entraîner la tâche interminable de neutraliser le « parasitage » d'une relation par les interactions entre variables. Disposant de la connaissance de la population réelle, il est bien placé pour apercevoir que ce « parasitage » est aussi un autre nom de la *configuration historique* qui constitue, en tant que constellation de co-occurrences observées, la seule réalité empirique dont on ne peut extraire, par déformation des relations réelles, une relation « pure » qu'au prix d'une abstraction formelle. L'obtention d'effets statistiquement « purs » suppose une décomposition des interactions qui, au bout du processus de purification, deviendrait un non-sens historique (« population fictive où la Lozère comprend en partie une grande métropole urbaine et où Paris est peuplé d'une fraction appréciable d'agriculteurs ou de salariés agricoles » – Desrosières, 1982). On voit en effet, quand on a les moyens de pousser jusqu'au bout cette entreprise statistique, qu'à mesure que le raisonnement s'améliore en tant que raisonnement expérimental, il s'affaiblit en tant que raisonnement historiquement pertinent puisqu'il éloigne les constats de relation de la situation sociologique où ces relations opèrent réellement : il est alors facile de montrer que pour atteindre à la *perfection logique*, le raisonnement expérimental devrait atteindre à l'*absurdité sociologique* en se donnant, pour pouvoir réaliser envers et contre tout la condition « toutes choses étant égales par ailleurs », des représentations suffisantes de co-occurrences inexistantes ou hautement improbables.

Dans le travail sociologique, c'est le « plan d'expérience » souvent conçu comme une forme idéale de la vérification des hypothèses qui illustrerait le même paradoxe épistémologique. L'« échantillon raisonné », où les quotas définis par les critères stratificateurs sont lestés d'un poids suffisant pour

permettre les croisements de variables appelés par les hypothèses et autoriser l'exercice statistiquement correct de la comparaison des fréquences d'apparition des différentes co-occurrences, a pour contrepartie d'éloigner le raisonnement des configurations réelles. Soit une enquête visant à étudier l'effet sur des attitudes ou des pratiques mesurées par des variables de type (y) de la classe sociale (x) et de l'activité professionnelle des femmes (t) , qui veut se donner les moyens de faire apparaître l'influence croisée de x et de t sur les variables de type y . On peut considérer que le « plan d'expérience » a été commodément conçu si non seulement l'échantillon raisonné contient des quotas égaux de femmes appartenant aux différentes catégories sociales $(x1, x2, x3)$, mais aussi dans chacun de ces quotas autant de femmes au foyer que de femmes actives $(t1, t2)$. C'est en tout cas minimiser les tracas dans la lecture des tableaux doublement croisés et dans l'interprétation de leur significativité statistique.

Mais si, dans les différentes catégories sociales, il en est par exemple une où les femmes sont presque toujours au foyer, l'abstraction qui permet de formuler comme croisement de trois variables (x, y, t) la variation des comportements entre femmes au foyer et femmes actives dans les différentes catégories sociales doit, pour se donner les conditions logiques d'une assertion générale sur les effets croisés du statut professionnel et de l'appartenance de classe, oublier momentanément l'inégale probabilité des deux statuts féminins dans les différentes catégories sociales. Autrement dit, le raisonnement statistique doit, pour pouvoir énoncer univoquement la *différence des effets*, supposer l'*identité de sens* d'une valeur donnée de chaque variable quels que soient les croisements dans lesquels elle entre. La distorsion que l'exigence expérimentale impose à la réalité sociale engendrerait, dans le cadre d'un raisonnement purement statistique qui ne peut que l'oublier, un contresens sociologique majeur : le sens de la distribution du trait mesuré par y ne peut s'énoncer sociologiquement sans interpréter l'improbabilité sociale pour une femme d'être professionnalisée quand elle appartient à cette catégorie sociale. L'irréalité de l'énonciation formelle que cet exemple grossit à dessein est évidemment de portée générale.

L'énonciation statistique des effets croisés à laquelle tout plan d'expérience travaille à fournir une base commode peut bien se faire en oubliant la structure de la population de référence, l'interprétation proprement sociologique des variations ne peut, elle, s'effectuer qu'en restituant aux différentes valeurs prises par une variable les sens différents que lui marquent, dans le contexte réel des co-occurrences, les raretés relatives gommées par le croisement. Il faut en effet accepter de penser, pour raisonner sur le sens sociologique des différents degrés de rareté, qu'une « même » valeur de l'indicateur qu'on traite comme une variable voit son sens varier selon les croisements auxquels il est soumis. Cela entraîne évidemment à référer à d'autres constats le sens d'une *rareté sociale* qui ne vient figurer qu'en y perdant son sens sociologique dans une case d'un tableau de comparaison, c'est-à-dire dans un tableau où toutes les co-occurrences ont un droit de vote énonciatif *formellement égal*. Bref, la mécanique comparative stricte tend toujours par ses exigences statistiques à faire oublier l'objet propre du raisonnement sociologique.

Le raisonnement sociologique, à qui le projet historique donne toujours pour tâche de restituer *in situ* les relations observées sur des populations réelles (d'individus ou de propriétés), exige donc de construire en chaque cas comme « objet » le sens particulier que prend un système de relations quand il est restitué à son contexte, puisque *la distance même à la situation expérimentale* (ou, si l'on veut, le sens irréductible conféré aux associations de propriétés par une configuration qui les associe ainsi et non autrement) fait partie de l'objet sur lequel on énonce.

La *majoration sémantique* invisible à laquelle les constats de relation sont sans cesse et comme mécaniquement entraînés par la grammaire des variables livre en creux son fonctionnement dans les perplexités des chercheurs néophytes qui découvrent, en multipliant les comparaisons entre leurs tableaux croisés ou leurs lectures de « plans factoriels », les proximités entre variables indépendantes traitées par l'enquête comme autant d'obstacles insurmontables à l'énonciation univoque du sens des corrélations. « L'effet de structure » est en effet le pain quotidien de l'interprétation : est-ce bien l'âge (ou le sexe ou l'appartenance sociale) dont on mesure « l'influence », s'il apparaît par exemple que les catégories d'âge (ou de sexe ou d'appartenance socioprofessionnelle) recouvrent aussi, dans la population soumise à l'enquête, des catégories découpées à peu près de la même manière par d'autres critères (scolarisation, travail salarié ou niveau du diplôme) ?

Devant cette provocation statistique, on voit souvent les sociologues qui ne peuvent se résigner à cette impuissance du raisonnement expérimental rêver, pour leur prochaine enquête, d'un « plan d'expérience » qui, en multipliant les croisements entre « critères stratificateurs », procurerait un échantillon idéal, celui qui serait capable de représenter par un nombre suffisant de sujets toutes les catégories nécessaires à la neutralisation successive des variables qui entrent en interaction. Le *leurre expérimental* livre ici en clair, en même temps que ses impossibilités, la plupart de ses implications épistémologiques. On voit qu'il conduit, lorsqu'on lui laisse développer sa logique, à un protocole d'échantillonnage aussi paradoxal que la *fiction borgésienne* de « la carte du territoire » superposable au territoire – avec ici cette aporie supplémentaire que la « représentation » du territoire devrait le déformer, selon des « échelles » incompatibles entre elles – pour autoriser les comparaisons de fréquences indispensables au plein effet de l'analyse multivariée. Il révèle à la limite que la plupart des catégories exigées par une application sans failles du raisonnement expérimental à des variations sociologiques deviennent rapidement des catégories quasi vides ou impossibles : enfants non scolarisés en deçà du terme de la scolarité obligatoire (pour séparer l'effet de l'école et celui de l'âge), hommes au foyer (pour séparer l'effet du sexe et celui du travail domestique), ouvriers sortis des Grandes Écoles, étudiantes mères de famille nombreuse, etc., etc. Sans aller aussi loin dans la soumission à une exigence logique sociologiquement absurde, on aperçoit à la lumière de cette contradiction que les distributions appelées par l'efficacité du raisonnement statistique sont plus ou moins éloignées des catégories découpées par les interdépendances réelles entre propriétés et, par conséquent, les effets tenant au fait qu'elles sont *plus ou moins improbables* dans la réalité historique disparaissent ou se déforment quand la ténacité de l'expérimentateur indifférent à ce « biais » réussit (on arrive à tout) à en trouver des représentants.

Le raisonnement sociologique ne peut donc être défini que comme un raisonnement condamné à se mouvoir et à opérer des compromis entre les exigences du raisonnement expérimental et celles du raisonnement historique. Le travail sociologique qui recourt au langage des variables dans l'analyse des variations sociales laisse en effet apparaître un *dilemme* ou, au moins, une *tension méthodologique* entre le raisonnement historique qui impose la connaissance des configurations réelles comme systèmes singuliers et non reproductibles de co-occurrences de propriétés et le raisonnement expérimental qui entraîne toujours à formuler les constats de relation fondés sur ces co-occurrences comme des assertions générales portant sur des corrélations entre variables pures. Il n'est évidemment pas davantage possible de se priver des services que rend à l'établissement des constats de relation le recours au langage des variables que d'ignorer les limitations et les tâches qu'impose à cette méthode de reconstruction et de traitement des faits le raisonnement sociologique. Ce n'est pas seulement parce qu'il étend le sens mathématique

ou physique de la notion de « variable » jusqu'à y comprendre des « ordres » ou de simples « partitions d'ensembles » que l'emploi sociologique de ce terme est analogique, mais surtout parce que l'*interprétation* des constats procurés par le croisement des variables renvoie à la connaissance du contexte, à chaque fois singulier, de ces corrélations, y compris dans le cas des « variables quantitatives » qui opèrent une mesure au sens strict.