

De l'interprétation des statistiques en sociologie

Philippe Cibois, professeur émérite de sociologie de l'université de Versailles – St-Quentin en Yvelines

Résumé

L'interprétation des statistiques en sociologie, et en particulier de l'analyse des données est confrontée au problème de l'interprétation réaliste ou non des types-ideaux que le chercheur peut repérer. A propos d'une enquête sur la bonne volonté scolaire des enfants issus de l'immigration, on montrera comment on peut discuter entre une interprétation réaliste et une interprétation constructiviste des résultats. On élargira la question par un dialogue fictif entre des partisans des deux camps.

Mots-clés

Analyse des données, réalisme, constructivisme, philosophie spontanée des chercheurs

Parler de mathématiques en sociologie se faisait assez largement dans les années 70 quand par exemple Raymond Boudon, qui s'était formé auprès de Paul Lazarsfeld à Columbia intitulait sa thèse « l'analyse mathématique des faits sociaux » (Boudon 1967) ; Jacques Maitre à la même époque (1972) mettait encore en rapport « sociologie religieuse et méthodes mathématiques ». Mais il y avait eu des précurseurs comme Georges Guilbaud (1952) qui a fait redécouvrir l'effet Condorcet et qui fut à l'origine du « Centre de mathématiques sociales » de ce qui deviendra Ecole des Hautes Etudes en Sciences Sociales (qui a pris la suite de la 6^e section de l'EPHE) avec une revue toujours active « [Mathématiques et Sciences Humaines](#) ».

On rêvait à l'époque de mathématiser la sociologie : par exemple Raymond Boudon pensait que le travail de clarification effectué en se servant du langage mathématique permettrait à la sociologie d'accéder à la maturité scientifique (Boudon 1971:7). Cependant si l'on regarde ce qui se faisait, on s'aperçoit que hors la théorie des graphes, les problèmes étudiés tournaient autour de la corrélation, de la causalité et n'étaient finalement que des problèmes statistiques.

Depuis cette époque, les manuels ont abandonné l'usage du mot « mathématiques » pour lui substituer celui de « statistiques ». On y étudie l'analyse des tableaux de contingence, depuis Jean-Paul Benzécri sous une forme factorielle (Benzécri 1973), puis sous une forme log-linéaire. Sur tout cela se greffe, dans le cas du dépouillement d'enquêtes, l'usage des tests de significations.

Je voudrais ici montrer assez concrètement à partir d'un exemple réel d'utilisation (Cibois 2002) en quoi consistent les problèmes d'interprétations qui se posent quand on l'utilise.

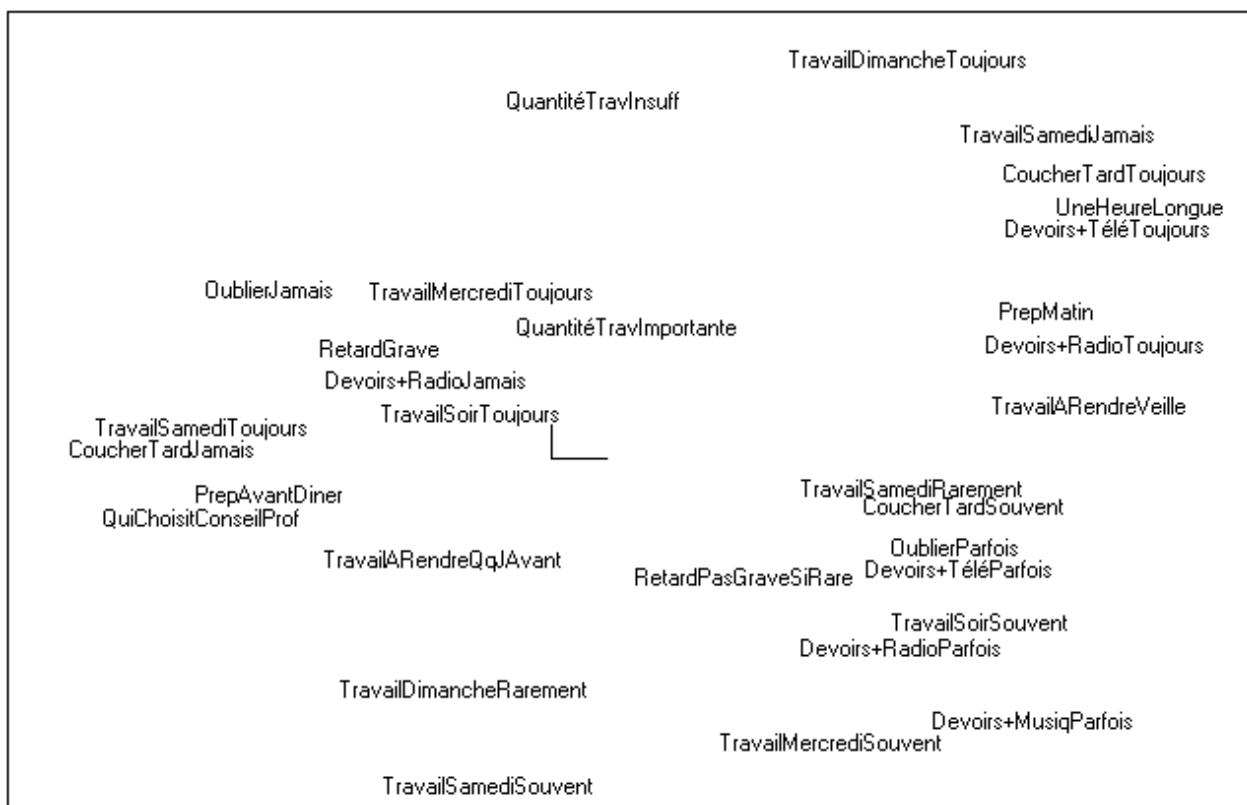
Le problème étudié est le suivant et concerne la carrière scolaire des enfants issus de l'immigration : certes le déroulement de leurs études les conduit à des situations inférieures en moyenne à la moyenne nationale mais on constate que si on compare leurs résultats avec ceux d'enfants non issus de l'immigration mais qui se trouvent dans le même environnement familial, ils réussissent comparativement mieux.

Pour tenter d'expliquer cette situation, j'ai utilisé les données du Panel de l'éducation nationale : cette enquête a consisté, pour sa vague de 1989, à sélectionner un échantillon national tiré aléatoirement de tous les entrants en 6^e et de

les suivre tout au long de leur scolarité. Parallèlement à ce panel, sont venues se greffer des enquêtes spécifiques par questionnaire auprès de 1900 élèves du panel auxquels on a demandé des renseignements sur leur comportement vis-à-vis du travail scolaire ainsi que d'autres renseignements sur leurs loisirs par exemple.

On trouvera en annexe du texte [en ligne](#) la liste des questions utilisées avec les modalités de réponses possibles. On dispose donc d'un tableau de données où en ligne se trouvent les 1900 individus et en colonne le numéro de la modalité de réponse qu'ils ont choisi pour cette question. Pour explorer cet ensemble de données on utilise la technique de l'Analyse factorielle des correspondances : on croise deux à deux toutes les questions de l'enquête et on juxtapose tous les tableaux de contingence obtenus dans un tableau symétrique dont on recherche les projections des modalités sur les deux premiers axes factoriels.

On a le graphique ci-dessous où l'on a retenu que les modalités qui contribuaient le plus à la fabrication des axes.



A gauche du graphique on repère un certain nombre de modalités qui ont en commun de traduire un respect des normes scolaires : un retard c'est grave, on n'oublie jamais un livre ou un cahier, on travaille le soir, le mercredi ou le samedi toujours, on ne se couche jamais tard, on prépare ses affaires avant le dîner, on écoute les conseils de ses enseignants sur ce qu'il y a à faire, on ne fait pas ses devoirs en écoutant la radio, on fait ses devoirs quelques jours avant de les rendre.

Si ces modalités de réponses à diverses questions sont proches, c'est qu'elles ont été choisies ensemble d'une manière particulière par un sous-ensemble de répondants.

Du simple constat (à vérifier cependant car il s'agit évidemment, de par la méthode choisie, d'une approximation des données puisqu'on ne considère que les deux premiers facteurs d'une analyse qui en comporte beaucoup plus), on passe à l'interprétation : cette sous-population qui a le souci de respecter les normes scolaires ont peut la qualifier par un attribut généralisateur que l'on appelle ici "bonne volonté scolaire". Le statut de cet attribut, quel est-il ? S'agit-il de la découverte d'une réalité qui existerait indépendamment du chercheur de la même manière qu'une faille géologique est repérée en surface par des différences de végétation ? Ce qui existe sans aucun doute ici, c'est le fait que des individus ont donné des réponses à un questionnaire qui leur a évoqué des comportements auxquels ils n'avaient peut-être pas réfléchi au préalable. Ont-ils répondu en respectant leur pratiques réelles, probablement pas tous, mais même si leurs pratiques réelles sont différentes, ils ont marqué par leurs réponses un souci de ce qu'on peut appeler par une généralisation non abusive une "bonne volonté" vis-à-vis de l'institution scolaire, ce que n'ont pas fait tous les répondants.

En effet, du côté droit du graphique, on a l'attitude inverse, non pas "mauvaise volonté", mais une attitude "décontractée" vis-à-vis de l'école (on oublie souvent ses affaires, on fait ses devoirs avec la télé ou la radio, on fait ses devoirs la veille et on prépare ses affaires le matin, on se couche tard). Cette attitude pouvant être modulée entre ceux qui, en haut du graphique, assument fondamentalement cette attitude (diverses modalités "toujours", et qui trouvent qu'une heure de cours c'est long) et en bas ceux qui y mettent une réserve (modalités "parfois").

Comme le graphique est le fruit d'une approximation, un travail de vérification s'impose : pour le mener à bien on sélectionne cinq modalités qui expriment bien cette bonne volonté scolaire et qui se trouvent dans la gauche du graphique : un retard c'est grave ; on ne fait jamais ses devoirs avec la radio ; on prépare ses affaires pour le lendemain avant le diner ; on suit les conseils de ses enseignants pour les devoirs à la maison ; une heure de cours semble courte.

Pour vérifier que ce rapprochement n'est pas abusif, on fait un simple comptage sur la population des répondants en repérant l'effectif de ceux qui ont ces cinq modalités, ceux qui en ont quatre, etc. ou enfin aucune. On a la distribution suivante :

Nombre de modalités						
Tot.	0	1	2	3	4	5
N=1865	402	662	556	197	44	4
%= 100	21.6	35.5	29.8	10.6	2.4	0.2

Le résultat est à la fois contre-intuitif et rassurant. On s'attendait évidemment à rencontrer un effectif "suffisant" de répondants ayant les cinq modalités, or ils ne sont que quatre. Cependant, les 44 qui ont quatre modalités de bonne volonté scolaire peuvent être encore caractérisés comme appartenant à ce qu'il ne faut plus qualifier comme un groupe concret mais comme un type abstrait. Si l'on prend la distribution dans l'autre sens, il est évident que ceux qui n'ont aucune réponse de ce type n'y appartiennent pas et que ceux qui n'en n'ont qu'une non plus. Où se situe la frontière ? A trois modalités, on a dans ce cas 197 + 44 + 4 soit 245 individus qui représentent 13,1% de l'ensemble. A deux modalités on en ajoute 556 et le groupe représente maintenant 42,9% du total. Le choix de l'appartenance avec trois modalités au moins se justifie par le fait que dans un cas le groupe marque bien un petit nombre (3 modalités au moins sur 5 possibles), ce qui n'est pas le cas pour l'autre qui discrimine peu la population.

Ce qui est manifeste, c'est que le "type" que nous a manifesté le graphique factoriel n'est pas à prendre comme un type pur mais comme un "portrait-robot" ou ce que Max Weber a appelé un "ideal-type", c'est-à-dire un ensemble d'individus regroupés par des traits qui font sens ensemble (et dont on vient de voir qu'ils sont attestés).

Cette "stylisation" du réel est-elle utile ? On peut aller plus loin et l'utiliser dans de nouvelles analyses où l'on utilise la possession de ce type (3 modalités au moins) en fonction de diverses caractéristiques que l'on teste avec la régression logistique sur données discrètes dans une procédure "toutes choses égales par ailleurs".

Je reprends ici les termes de l'enquête citée :

"On veut donc maintenant expliquer l'appartenance au type "bonne volonté scolaire" en tenant compte de l'effet de plusieurs facteurs : la nationalité en opposant ceux qui sont d'origine maghrébine aux autres, le sexe, la catégorie sociale regroupée en "inférieure" (ouvriers, retraités, sans profession, en chômage ou non-réponse) contre tout le reste (agriculteurs, artisans commerçants chefs d'entreprise, cadres et professions intellectuelles supérieures, professions intermédiaires, employés) résumée par opposition en "supérieure". On ajoute enfin le niveau en mathématique qui oppose le niveau qualifié de "bon" à tout le reste (moyen, passable, insuffisant) qualifié de "non bon".

Les résultats sont les suivants :

Modalité à expliquer : Bonne volonté scolaire
 Situation de référence NBon Masc Inf Autr 14.3

Effets marginaux				
	Odds-ratio	Pourcentage	Significatif à	
Bon	0.62	-5.0	0.000	***
Fémi	1.23	2.7	0.038	**
Sup	0.80	-2.6	0.015	**
Magh	2.39	14.2	0.000	***

En prenant séparément l'effet propre de chaque caractéristique ("toutes choses égales par ailleurs"), on voit que le fait d'être bon élève (en mathématiques) ne pousse pas à avoir de la bonne volonté scolaire puisque l'effet est négatif (effet significatif comme d'ailleurs tous les autres). Un bon élève prend facilement ses distances vis-à-vis de l'institution, il n'est pas "scolaire", ne se croit pas obligé de respecter les conseils de ses enseignants (ou de ses parents).

L'effet féminin est positif : on retrouve là le comportement souvent attribué aux filles qui sont dit-on plus "respectueuses des consignes" que les garçons. A égalité d'autres variables, le fait d'être de sexe féminin fait augmenter la proportion de celles qui ont une bonne volonté scolaire.

Pour l'origine "supérieure", on retrouve un effet négatif ce qui veut dire que ces élèves, qui ne sont pas dans la catégorie inférieure (car cette catégorie dite supérieure se caractérise surtout par le fait de ne pas être dans la catégorie "ouvrier" ou dans une catégorie précaire), ressentent moins le besoin d'une adéquation aux normes scolaires. Ils ont vis-à-vis d'elles une certaine décontraction.

Par contre la nationalité maghrébine des parents a un effet très fort sur la proportion de bonne volonté scolaire qui est augmentée de ce fait de plus de 10%. Visiblement, pour cette population, l'obéissance aux normes scolaires est perçue

comme quelque chose d'important, quelque soit la réussite, le sexe, la catégorie sociale. Dans le projet migratoire de ces familles, l'école est perçue comme un moyen de promotion et la norme scolaire est vue comme devant être respectée par les enfants. Certes, on l'a vu, tous les enfants de ces familles ne jouent pas ce jeu mais jouer ce jeu est très fort dans cette catégorie de nationalité.

On peut donc dire en conclusion que parallèlement au désir d'une scolarité longue, qui avait déjà été repéré par Vallet et Caille (1995), la bonne volonté scolaire est un élément qui s'inscrit aussi dans le projet migratoire des enfants issus de famille de nationalité maghrébine. Cependant, comme l'a montré récemment Stéphane Beaud (2002), la confiance en l'école n'est pas forcément récompensée ; de plus le refus avéré qui est fait aux jeunes d'origine maghrébine d'emprunter les filières de l'apprentissage les condamne aussi à rester dans le système scolaire dont ils peuvent être vus comme prisonniers." (Cibois 2002)

Les résultats donnés ici sont significatifs à divers seuils mais ceci n'interdit pas, et c'est fait dans l'article de référence, de vérifier les résultats obtenus par d'autres méthodes pour s'assurer de leur stabilité.

Pour conclure sur cette utilisation de la statistique en sociologie, il faut bien voir que son utilisation s'inscrit dans une problématique plus large : au moment où l'on exploite les données du Panel de l'éducation nationale, on sait déjà des choses sur le statut scolaire des populations issues de l'immigration : toute enquête de sociologue se situe dans la longue durée et la cumulativité des résultats.

Reste le problème épistémologique de savoir ce qui fonde l'interprétation du sociologue : la "bonne volonté scolaire" appartient-elle au réel ou est-elle le fruit d'une démarche constructionniste ? Il y a des arguments dans les deux sens que je vais styliser dans un dialogue entre deux positions (idéales-typiques bien sûr) tenues par deux interlocuteurs :

Realismus : l'école génère des attitudes, des attentes qui sont bien réelles car l'enfant qui ne les respecte pas est sanctionné d'une manière dont il ne met pas en cause la réalité. De ce fait étudier l'intériorisation de ces pratiques chez des enfants et leur famille n'est que photographier la réalité.

Constructivus : si le type observé s'appuie sur des pratiques effectives, l'observation du type, sa photographie si l'on veut, est bien le fruit d'une construction dont nous avons suivi les périples et qui repose sur un échafaudage que certains pourraient qualifier comme relevant de la subjectivité du chercheur, de ses intérêts de recherche, voire de ses croyances en matière politique et sociale. En effet croiser la question de la bonne volonté scolaire avec l'appartenance au monde de l'immigration est spécifiquement une préoccupation marquée au niveau social comme appartenant à ce qu'on pourrait là aussi une "bonne volonté politique de gauche" (pour ne pas dire "politiquement correcte"). Le chercheur pourrait chercher des choses toutes différentes et en projetant l'attention sur la question qui le préoccupe, il la fait socialement exister. Dans d'autres environnements politiques ou sociaux, cette question n'existerait même pas et on peut donc dire qu'elle est socialement construite par l'intérêt du chercheur.

Realismus : vous mélangez les perspectives. Je vous accorde que la focalisation d'un objectif photographique est bien fixée par celui qui tient

l'appareil. Le cliché est le fruit du croisement d'une préoccupation mais aussi d'une réalité qui était préexistante.

Constructivus : est-ce bien sûr ? Ce n'est pas parce que vous isolez dans votre viseur des traits qu'ils existent indépendamment de votre regard : les sanctions scolaires pour non respect de la bonne attitude ne définissent une "attitude" que parce que vous les regroupez pour les besoins de votre démonstration. L'imprécision de votre type-idéal le manifeste d'ailleurs : vous avez isolé des pratiques qui peuvent être reliées à d'autres univers, celui de l'école, celui de la famille, celui des loisirs, chacun de ces univers ayant sa logique propre.

Realismus : mais ces logiques propres dont vous faites état ont bien, elles, une existence réelle ? Les faits sociaux sont des "choses", comme le disait Durkheim, qui sont aussi présentes que cet arbre que je vois ou que cette table sur laquelle je travaille.

Constructivus : votre table est typiquement une construction sociale de type "menuiserie" et votre arbre est souvent une création artificielle, plante d'ornement, de parc, qui n'existerait pas toute seule à l'état naturel.

Realismus : l'arbre qui est devant moi appartient à une espèce scientifiquement repérable.

Constructivus : une espèce est aussi un découpage, variant à chaque instant comme le montre l'Evolution et dont les définitions actuelles sont pragmatiques comme celle de celui qui fait autorité aujourd'hui E. Mayr (1974) qui définit l'espèce comme un "groupe de populations naturelles effectivement ou potentiellement interfécondes, isolé par rapport aux groupes similaires au plan de la reproduction".

L'espèce est un découpage local et temporel qui regroupe des individus présentant des similarités qui permettent l'interfécondation, elle n'est pas basée sur des caractéristiques morphologiques "objectives" comme aux origines chez Cuvier. Ce sont les individus qui portent le matériel génétique qui a d'ailleurs une certaine variabilité.

Realismus : donc vous niez la réalité des découpages scientifiques ?

Constructivus : absolument pas, mais même le langage courant est bien conscient de la priorité du point de vue. Reprenons le problème de l'arbre et supposons une forêt à l'état naturel et dans une expérience de pensée, voyons des arbres isolés, perceptibles individuellement.

Pourquoi sont-ils individualisables ? C'est parce qu'ils sont à l'échelle humaine, ce qui manifeste l'importance du point de vue. Dans une prairie, les brins d'herbes sont tout autant individualisables que les arbres à l'échelle humaine mais si nous considérons la prairie comme un tout, c'est bien pour cette raison d'échelle. D'ailleurs quand nous survolons une forêt, nous avons le même réflexe et parlons de bois, de bosquet, de futaie, de forêt, sans éprouver le besoin d'analyse plus précise.

Un arbre existe-t-il seul ? Pas plus qu'une herbe, il ne peut se passer d'un milieu écologique, il suppose l'existence de la terre, de l'eau, du gaz carbonique. Selon le point de vue, on peut l'envisager seul, si l'on s'intéresse à ses qualités pour la construction, ou dans son milieu, si l'on s'intéresse à

sa croissance. Si nous sommes réticents à refuser l'existence propre, situons-nous au niveau du brin d'herbe où le changement d'échelle nous facilite la démarche intellectuelle. Le vocabulaire est d'ailleurs explicite : l'herbe est collective et pour l'individualiser il faut en saisir le brin, la partie. L'arbre individuel existe ainsi que son collectif, la forêt.

C'est le regard, l'intérêt humain qui entraîne l'opération de découpage qui est fondamentale, pas la réalité qui accepte tous les points de vue. Quand à l'individu, arbre, brin d'herbe, il peut tout autant être analysé en ses composantes, qui existent aussi, que regroupé en collectifs qui existent également.

Realismus : vous admettez la réalité, ce qui me rassure pour la vie pratique et pour l'existence de la réalité matérielle. Ce qui serait artificiel serait le point de vue en fonction des objectifs de celui qui fait le découpage. Je le veux bien pour la table ou même pour l'arbre, mais pour la structure intime de la matière ? Les atomes et les molécules ne sont pas des vues de l'esprit.

Constructivus : des physiciens qui réfléchissent sur ces questions, par exemple Espagnat (1994) ou Bitbol (1998), sont plus dubitatifs, le premier parlant du "réel voilé" et le second de "l'aveuglante proximité du réel", ce qui dans les deux cas montre bien que le réel n'est pas si facilement à notre portée...

Le dialogue pourrait continuer longtemps : il est important de voir que dès que la pensée s'est reportée sur son fonctionnement, c'est à dire au moment où le *concept* a été l'objet d'une réflexion, son existence indépendamment de celui qui le pense a été posée. Pour Platon le concept a une existence indépendante puisque si nous l'acquerrons c'est que nous nous remémorons des *idées* pré-existantes. Ce réalisme des idées n'a été remis en cause qu'au Moyen Age par le nominalisme qui s'est opposé à la synthèse thomiste de la vérité comme *adequatio rei et intellectus* : cette adéquation de l'intelligence et du réel continue d'être la philosophie spontanée des chercheurs. Il faut qu'une théorie "colle au réel", il ne faut pas "aller contre les faits", faits qui sont "durs, incontournables".

Cette position est intenable : comment pourrait-il y avoir identité entre le réel et l'intelligence sinon par le biais d'une représentation, d'un concept, d'une théorie. Dans ce cas, la théorie n'est plus le réel, mais le réel pensé, ramené aux cadres de notre intelligence. Ce que les sciences produisent, ce sont des modèles du réel, c'est à dire des concepts qui font des hypothèses sur le comportement du réel : l'expérience, en confrontant le modèle et le réel, viendra nous dire si le modèle était plus ou moins correct.

Ma conclusion est que dès qu'on réfléchit au problème, on doit abandonner "le réalisme spontané du savant" qui est une démarche pragmatique qui signifie simplement qu'on croit en ce qu'on fait. Il est naïf d'ignorer la controverse philosophique actuelle sur la question et d'en rester à un simple gros bon sens. Le réel ordinaire existe bien mais peut être découpé de plusieurs manières selon le point de vue adopté ; quand au réel de la démarche scientifique, plus on s'écarte du monde ordinaire, plus il devient problématisé (mais non *problématique* au sens de douteux). Comme le souligne Ian Hacking (2001), même les pierres, symboles de la

dureté, de la fixité, ont une histoire, comme on le voit dans le cas de la dolomite où le vivant vient peut-être jouer son rôle (Hacking 2001 : 251-276).

Références

- Beaud, S., *Les enfants de la démocratisation : 80% au bac, et après ?*, Paris, La découverte, 2002
- Benzécri, J.-P., (dir.), *L'analyse des données, I, La taxinomie, II, L'analyse des correspondances*, Paris, Dunod, 1973
- Bitbol, M., *L'aveuglante proximité du réel*, Paris, Champs/Flammarion, 1998
- Boudon, R., *L'analyse mathématiques des faits sociaux*, Paris, Plon, 1967
- Boudon, R., *Les mathématiques en sociologie*, Paris, PUF, 1971
- Cibois, Ph. "La bonne volonté scolaire. Expliquer la carrière scolaire d'élèves issus de l'immigration" dans Philippe Blanchard et Thomas Ribémont (Dir.), *Méthodes et outils des sciences sociales. Innovation et renouvellement*, Paris, L'Harmattan, 2002, coll. "Cahiers politiques", p.111-126.
- Espagnat, B. d', *Le réel voilé*, Paris, Fayard, 1994
- Guilbaud, G.-Th., « Les théories de l'intérêt général et le problème logique de l'agrégation », *Économie appliquée* 5, 1952, p. 501-551
- Hacking, I., *Entre science et réalité. La construction sociale de quoi ?*, Paris, La découverte, 2001
- Maitre, J, *Sociologie religieuse et méthodes mathématiques*, Paris, PUF, 1972
- Mayr, E., *Populations, espèces et évolution*, Paris, Hermann, 1974
- Vallet L.-A et Caille J.-P., "Les carrières scolaires au collège des élèves étrangers ou issus de l'immigration", *Education et Formations*, n°40, 1995, p. 5-14