

Réflexions sur la causalité en sciences sociales

Daniel Courgeau *

Si le chercheur ne peut induire une causalité au sens strict à partir de la seule observation des phénomènes, il doit cerner l'ensemble des faits observés dans son champ par des hypothèses suffisamment générales pour pouvoir tous les englober et conduire à une explication satisfaisante de ces divers événements. Deux paradigmes ayant prévalu dans certaines sciences sociales sont développés dans cet article. Le premier suppose que les faits sociaux ont une existence propre, indépendamment des individus qui les vivent et qu'ils ont une grande stabilité dans le temps. Mais ce paradigme peut conduire au risque d'erreur écologique, ignorant le rôle et l'effet des actions individuelles. Le second donne, au contraire, la priorité aux comportements individuels et conduit à une étude des actions du sujet en fonction de ses diverses caractéristiques. Il provoque le risque d'erreur atomiste si l'on ignore le contexte dans lequel les actions se produisent. L'analyse multi-niveaux permet de dépasser ces deux écueils en faisant intervenir divers niveaux d'agrégation, passant de l'individu à des unités plus complexes. Elle doit simultanément faire intervenir le temps pour prendre une approche des phénomènes humains plus satisfaisante, dans une analyse biographique multi-niveaux. Une discussion d'ensemble montre l'importance du mode d'observation sur les hypothèses posées et la façon dont de nouvelles observations, amenant des contradictions, conduisent à un changement de paradigme. Inversement, ce changement peut mener à de nouveaux modes d'observation et d'analyse des sociétés humaines.

Cet article part des difficultés rencontrées en sciences humaines, lorsque l'on cherche à mettre en évidence les « causes » des phénomènes et à reconstituer les relations qui lient ces causes aux effets qu'elles entraînent. Ainsi, dans ce dossier de *Recherches et Prévisions* sur les liens entre le chômage et la famille, il paraît utile de dépasser la simple description des faits observés et, par conséquent, chercher à établir des relations entre eux.

Si l'on veut montrer, par exemple, l'impact

du chômage sur la vie familiale, en particulier sur l'instabilité conjugale, il est nécessaire de mettre en évidence et de justifier les hypothèses sous-jacentes à une telle analyse et de montrer la façon dont les divers phénomènes agissent les uns sur les autres, pour conduire aux résultats finalement obtenus.

Ainsi, si on travaille d'abord sur des enquêtes qui relèvent les conditions dans lesquelles les individus vivent (appelées en démographie enquêtes du moment ou

* Directeur de recherche à l'INED.

transversales) et si ces enquêtes suivent les chômeurs sur une courte période de quelques mois, on pourra mettre en évidence un effet de courte durée : se trouver dans une situation de chômage met-il en cause l'existence et la vie du couple ? En utilisant l'enquête Suivi des chômeurs de l'INSEE – qui a suivi pendant dix-huit mois leur situation professionnelle et familiale – avec d'autres sources sur l'ensemble de la population, il a été possible de montrer que les personnes ayant connu le chômage n'ont pas une plus forte vulnérabilité à la dissolution du couple que les autres (Herpin, 1990).

En revanche, si on travaille sur des enquêtes permettant le suivi des couples depuis leur constitution et celui des individus tout au long de leur vie, appelées en démographie enquêtes biographique ou longitudinales, on pourra relever un effet de plus longue durée : la précarité durable de la situation professionnelle met-elle en cause la stabilité des couples ? En utilisant l'enquête Situations défavorisées de l'INSEE – qui comportait une partie biographique à partir de laquelle pouvaient être reconstitués plusieurs types de trajectoires familiales –, il a été possible de montrer que les risques de rupture conjugale sont très élevés chez les couples ayant eu une forte précarité professionnelle (Paugam, 1993).

Des approches différentes de l'explication scientifique

Selon la façon dont on réalise l'étude, on arrive à des résultats *a priori* et qui sont à rattacher aux problèmes liés à la causalité en sciences humaines, aux hypothèses sous-jacentes à chaque analyse et à la détermination des facteurs explicatifs d'un phénomène (Nezosi, 1999). Plusieurs types de causalité renvoient à des approches différentes de l'explication scientifique et ont été développés dans de nombreux ouvrages (Wunsch, 1988 ; Duchêne et alii, 1989 ; Franck, 1994). Il ne peut être question de présenter et de discuter en détail ces différentes approches, ce qui conduirait à des développements dépassant largement le cadre d'un article.

On écarte de l'analyse l'examen de la distinction entre causalité en sciences humaines et en sciences de la nature (Piaget, 1967), où l'approche expérimentale permet une meilleure imputation causale, même si celle-ci n'est jamais définitive (Popper, 1973). La causalité dans l'ensemble des sciences sociales n'est pas non plus étudiée : la démographie et la sociologie sont privilégiées, laissant de côté d'autres sciences telles que l'économie ou la psychologie pour lesquelles la conception de causalité peut être très différente. Seul un ouvrage réunissant des spécialistes de ces diverses sciences et des philosophes permettrait de répondre à ces questions.

Dans cet article, on examine plus précisément les formes de causalité à l'origine de certaines méthodes d'analyse utilisées en sciences humaines, en partant des analyses les plus simples et en les rendant successivement plus complexes et réalistes.

L'analyse du moment par des données agrégées

Supposons qu'il existe, dans un pays, des groupes d'individus homogènes vis-à-vis d'un phénomène donné. On peut penser que les membres d'un même groupe religieux (catholique, protestant, juif, etc.) ont la même propension au suicide quel que soit leur lieu de résidence, ou bien que les membres d'une même profession (agriculteurs, ouvriers, cadres, etc.) ont la même propension à la migration, etc. On peut considérer que ces tendances collectives ont une existence propre, indépendamment des individus constituant ces groupes : dans ce cas, ce sont des forces aussi réelles que les forces physiques, bien qu'elles soient d'une autre nature (Durkheim, 1930).

Ainsi, une première forme de causalité peut être définie : son origine est dans la société elle-même et non dans l'individu. Elle s'impose à lui, qu'il le veuille ou non. Un fait donné est la cause d'un effet social en fonction du système social dans lequel l'individu vit : cela permet d'éviter une analyse sur des données individuelles, des

données agrégées suffisant pour la faire. On peut également penser que ces forces ont une assez grande stabilité dans le temps et ne changent qu'avec une extrême lenteur : cela conduit à une analyse du moment.

Voyons plus précisément comment expliquer l'action causale exercée respectivement par le protestantisme et les autres religions (essentiellement le catholicisme) sur les taux de suicide d'un pays européen. Si l'on connaît, par exemple, les pourcentages d'individus qui se suicident ainsi que les pourcentages de protestants, dans chaque région du pays, il sera possible de porter sur un graphique la propension au suicide de cette religion, comparée aux autres. Dans ce cas, on va montrer que, si les hypothèses initiales sont vérifiées, on aura une relation linéaire entre ces deux caractéristiques.

En effet, si t_p et $t_{\bar{p}}$ sont les taux de suicide des protestants et des non-protestants, qui gardent la même intensité dans toutes les régions du pays considéré, si $N_p(r)$ et $N_{\bar{p}}(r)$ sont les effectifs des deux populations dans une région r donnée, alors le

nombre de suicides observé, $S(r)$ sera :

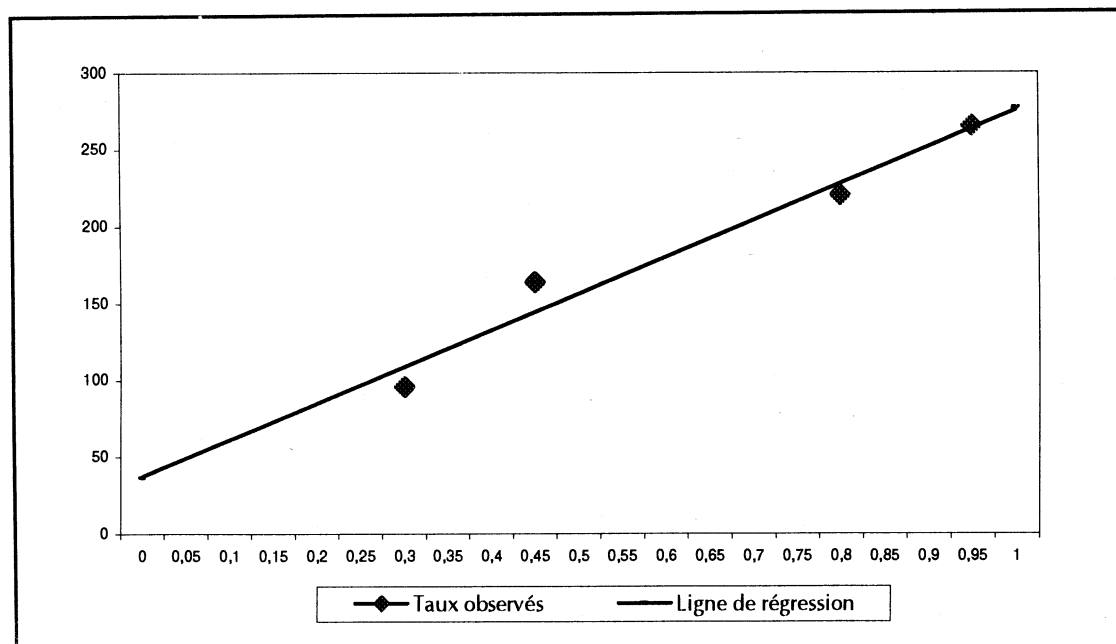
$$S(r) = N_p(r)t_p + N_{\bar{p}}(r)t_{\bar{p}} = N_p(r)(t_p - t_{\bar{p}}) + N(r)t_{\bar{p}}$$
où $N(r)$ est la population totale de la région r .

Il en résulte bien que le taux de suicide, $s(r)$, sera fonction linéaire de la proportion de protestants dans chaque région, $n_p(r)$:

$$s(r) = n_p(r)(t_p - t_{\bar{p}}) + t_{\bar{p}}$$

Emile Durkheim (1930) constate qu'en Prusse, entre 1883 et 1890, on a une augmentation régulière de la part des suicides lorsque la part des protestants augmente, d'une province à l'autre (graphique 1). On peut ajuster un modèle de régression linéaire avec un excellent coefficient de détermination ($R^2 = 0,97$) qui, par extrapolation, donne les taux de suicide théoriques que l'on observerait dans une population sans protestants, 37 pour 1 000 000, et dans une population entièrement protestante, 277 pour 1 000 000. Il en résulte que si les hypothèses à la base de cette analyse sont vérifiées en Prusse – ce qui semblerait vraisemblable selon le graphique 1 –, les risques de suicide des protestants sont près de huit fois plus forts que ceux des catholiques.

Graphique 1 - Suicide selon la part de protestants (Prusse, 1883-1990)



Source : Durkheim E., *Le suicide*, PUF, Paris, 1930 (première édition 1896).

Cependant, si les hypothèses posées *supra* ne sont pas valides, le résultat de ces observations ne permet plus de conclure que les protestants ont de plus forts risques de se suicider que les autres religions. Ces observations permettent seulement de dire qu'un fort pourcentage de protestants dans la population conduit à un plus fort taux de suicide d'ensemble, qui peut concerner aussi bien les protestants que les autres religions. Bien plus, comme aucune des provinces n'est entièrement protestante ou non protestante, il se pourrait aussi bien que les non-protestants (principalement les catholiques) soient plus enclins au suicide dans les provinces à prédominance protestante. Il est possible que les membres d'une minorité religieuse aient une plus forte probabilité au suicide que les membres d'une communauté majoritaire. L'interprétation pourrait être alors inverse de ce que l'on concluait précédemment, sous l'hypothèse que les membres d'une communauté religieuse ont toujours la même propension au suicide quel que soit leur part dans la population. Cela revient à introduire un effet d'interaction au niveau individuel entre la religion et la composition religieuse du groupe où vit l'individu, dans une province donnée, qui enlève toute validité au postulat précédent.

L'« erreur écologique »

Ces divers types d'inférences peuvent conduire à ce que l'on a coutume d'appeler l'« erreur écologique », qui intervient lorsque l'on cherche à déceler des comportements individuels à partir d'observations agrégées. En 1950, W.-S. Robinson abordait déjà ces problèmes avec des arguments statistiques (Robinson, 1950). Il a montré que les corrélations entre deux caractéristiques mesurées de façon binaire sur des individus ou par des proportions sur des agrégats n'étaient, en général, pas égales entre elles. Ainsi, la corrélation entre les proportions de populations noires et illettrées aux Etats-Unis, en 1930, était de 0,95 en travaillant sur neuf divisions géographiques alors que la corrélation entre le fait d'être noir et le fait d'être illettré, pour un individu, n'était

que de 0,20. On peut affirmer qu'une corrélation ne permet en rien de conclure à une causalité entre deux phénomènes. Il peut y avoir une forte corrélation positive entre deux phénomènes en raison d'un troisième agissant de manière concomitante, alors qu'ils peuvent être entièrement indépendants entre eux.

Si E. Durkheim avait disposé des techniques de régression, il aurait eu la possibilité de vérifier s'il ne commettait pas une erreur écologique. Il disposait des taux de suicide pour chacune des communautés religieuses en Prusse. Or ses calculs conduisent à un taux de suicide, en 1890, de 240 pour 1 000 000 pour les protestants contre 100 pour 1 000 000 pour les catholiques – qui constituent l'essentiel des autres religions – soit un taux seulement 2,4 fois plus élevé pour les protestants, comparé au taux 7,6 fois plus élevé obtenu avec les données agrégées. Il n'avait pas relevé ce risque d'erreur écologique, à l'encontre de ses postulats, car il n'avait pas estimé l'importance numérique de ces effets dans les deux analyses.

Dès lors, il est nécessaire de lever l'hypothèse initialement posée, selon laquelle seule l'appartenance à un groupe donné peut influencer les chances qu'a un individu de connaître un événement, quel que soit son lieu de résidence, et essayer d'intégrer dans l'analyse des interactions plus complexes entre événements.

L'analyse du moment par des données individuelles et agrégées

D'autres raisons viennent maintenant compliquer l'analyse. En premier lieu, les sous-groupes d'individus étudiés peuvent ne plus être homogènes vis-à-vis du phénomène étudié. Selon la région, leur propension à le connaître est susceptible de varier. On dit alors que la population est hétérogène. Selon les régions, les individus ne faisant pas partie du groupe à étudier peuvent également avoir des propensions différentes à connaître ce même phénomène. Enfin, il

peut exister une corrélation non nulle entre les propensions de ces deux groupes. Comment formaliser ces différentes propositions ?

Disons que des données mesurées au niveau régional (pourcentages de protestants, par exemple) sont *agrégées*, alors que des données mesurées sur chaque membre de la population (le fait d'être protestant, par exemple) sont *individuelles*. Dans tous les cas précédemment cités, l'étude réalisée sur les données agrégées par région donne des résultats différents de celle faite au niveau individuel. Mais cette dernière étude risque de faire perdre une information non négligeable sur les comportements tenus dans les différentes régions.

Pour l'illustrer, prenons le travail réalisé en Norvège sur la génération née en 1958. On observe les migrations interrégionales entre 1980 et 1981, en faisant intervenir le fait que les individus sont des agriculteurs, à l'aide d'un modèle logit (Baccaïni et Courgeau, 1996). En travaillant d'abord sur les données agrégées au niveau régional, on montre un effet positif du pourcentage d'agriculteurs sur les chances de migration. Plus le pourcentage d'agriculteurs augmente, plus l'émigration de ces régions sera importante.

En revanche, en travaillant sur les données individuelles, on constate que les agriculteurs ont une bien plus faible chance d'émigrer que les autres professions. Ce résultat conduit à une conclusion apparemment inverse de la précédente, qui conforte la nécessité de poser des hypothèses plus complexes. Notons que cette approche ignore le contexte dans lequel les conduites humaines se produisent et conduit à ce que l'on appelle l'« erreur atomiste ». En fait, ce contexte doit certainement jouer sur les comportements individuels et il paraît fallacieux d'isoler l'individu des contraintes imposées par la société dans laquelle il vit.

D'où l'idée de travailler simultanément à divers niveaux d'agrégation, afin d'expliquer un comportement toujours mesuré au niveau individuel et non plus régional comme dans

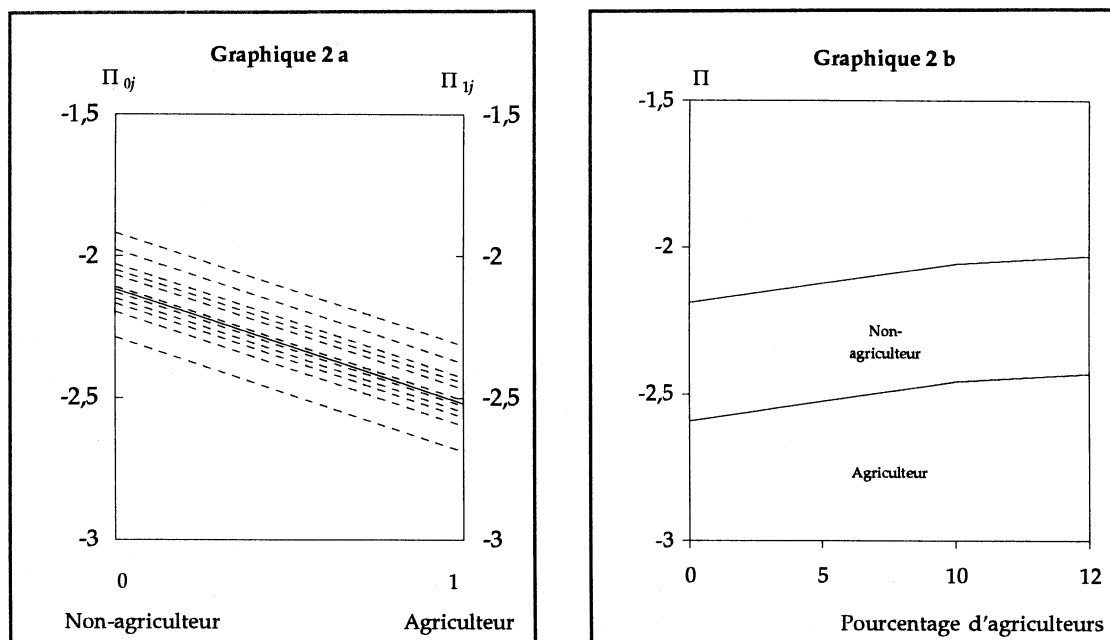
la partie précédente. Le risque d'erreur écologique est ainsi éliminé car la caractéristique agrégée va mesurer une construction différente de son équivalent au niveau individuel. Elle n'intervient plus comme un substitut, mais comme une caractéristique de la sous-population affectant le comportement d'un individu qui en fait partie. Simultanément, l'erreur atomiste disparaît à partir du moment où l'on fait intervenir correctement le contexte dans lequel vit l'individu. C'est ce que l'on peut appeler une analyse contextuelle, lorsque interviennent seulement des caractéristiques agrégées, et une analyse multiniveaux, lorsque l'on introduit une dépendance entre les comportements d'individus vivant dans une même région.

Le danger d'inférer certaines hypothèses

Un nouveau modèle permet une explication claire de cette apparente contradiction (Courgeau et Baccaïni, 1997). Appelons Π_{0j} la fonction logit de la probabilité d'émigrer des non-agriculteurs dans la région j et Π_{1j} celle des agriculteurs. Le graphique 2 a porte les valeurs de Π_{0j} et Π_{1j} reliées pour chaque région par des droites : ces droites sont parallèles entre elles. Ainsi, au niveau individuel, les agriculteurs ont toujours une bien plus faible probabilité d'émigrer que les autres professions et, au niveau des régions, les variances des logits des probabilités d'émigrer sont les mêmes pour ces deux groupes. En revanche, lorsque l'on fait intervenir, au niveau agrégé, le pourcentage d'agriculteurs présents dans ces régions (graphique 2 b), les chances de migrer tant des agriculteurs que des autres professions vont augmenter, les deux lignes étant parallèles entre elles. Mais les agriculteurs auront toujours une plus faible probabilité de migrer.

Cela montre le danger d'inférer certaines hypothèses sur le comportement individuel à partir de résultats obtenus au niveau agrégé. La présence de nombreux agriculteurs entraîne une plus forte probabilité d'émigration de toutes les catégories de la population, en partie, sans doute, à cause de la rareté des emplois non agricoles dans ces

Graphique 2 - Effet de la caractéristique « être agriculteurs » sur le logit de la probabilité de migrer en Norvège (génération 1958 entre 1980 et 1981)



Graphique 2 a : sans faire intervenir le pourcentage d'agriculteurs
Graphique 2 b : en faisant intervenir simultanément le pourcentage d'agriculteurs

Source : Courgeau D. et Lelièvre E., *Nuptialité et agriculture, Population*, 1986, vol. 40, n° 2.

régions. Mais cela n'implique en rien les plus fortes chances d'émigrer des agriculteurs par rapport aux autres : au niveau individuel, on observe l'inverse.

Pour diverses autres caractéristiques, toujours considérées séparément, on peut avoir des schémas différents selon les comportements considérés. L'effet agrégé peut venir accentuer l'effet individuel ou, au contraire, être entièrement indépendant de celui-ci. Mais une autre raison peut également influencer sur les résultats d'une telle analyse. Jusqu'ici on a seulement considéré les liens entre l'arrivée d'un événement donné et une seule caractéristique intervenant au niveau agrégé ou individuel. Or celle-ci peut agir en conjonction avec d'autres caractéristiques, modifiant son effet.

Il importe dès lors de voir si l'intervention de ces autres caractéristiques dans le modèle changera l'effet de la première caractéristique considérée et, si oui, comment. Si on reprend l'exemple norvégien, en plus du fait d'être agriculteur et du pourcentage

d'agriculteurs, on a la possibilité de faire intervenir dix autres caractéristiques (marié, actif, plus de douze ans d'études, au moins un enfant, faibles revenus, forts revenus, sans revenus, et certaines caractéristiques agrégées correspondantes). On constate que ces autres effets modifient peu les effets des caractéristiques antérieures qui restent toujours pleinement significatifs. Notre résultat est renforcé. Cependant, il ne permet toujours pas de parler de causalité car il n'est jamais possible de faire intervenir toutes les caractéristiques ayant un effet sur le phénomène étudié. Mais on peut dire que cet effet est conforté par une telle analyse. Cependant, jusqu'à présent, on a travaillé en transversal et on n'a pas examiné l'effet du temps sur ces liens. Il faut maintenant le faire intervenir.

Essayons de voir s'il n'est pas nécessaire de dépasser les conditions du premier exemple de l'introduction : l'effet suivait de très près dans le temps la cause supposée. Or aucune raison empêche des événements survenus dans un passé lointain d'avoir

toujours une influence sur des événements actuels. Dans l'étude norvégienne, le fait que l'individu soit agriculteur juste avant sa migration masque les effets de sa carrière professionnelle antérieure qui peuvent avoir un rôle aussi important.

L'intervention du temps dans l'analyse

Il paraît donc nécessaire de faire intervenir des événements de la vie d'un individu, même s'ils sont très reculés dans le temps, pour mieux comprendre son comportement actuel. On est dans le cas du deuxième exemple donné dans l'introduction. Cette analyse peut être réalisée grâce à la théorie des processus stochastiques (1) (Andersen et alii, 1993). Elle fournit la base théorique indispensable à l'analyse démographique des biographies (Courgeau et Lelièvre, 1989 et 1992), et permet d'introduire le temps qui sépare les divers événements étudiés.

Pour l'illustrer, prenons le cas d'une sous-population qui débute sa vie dans le secteur agricole et qui est initialement célibataire (générations nées entre 1911 et 1936, en France). On va suivre ces individus tout au long de leur existence et s'intéresser aux changements dans leur situation : certains se marieront, d'autres quitteront l'agriculture... Ces événements peuvent se produire dans un ordre quelconque et certains ne jamais arriver (Courgeau, Lelièvre, 1986).

On suppose maintenant que le comportement de l'individu sera fonction de l'état dans lequel il se trouve – sachant que cet état peut changer au cours du temps –, de son état antérieur – lorsqu'il en a changé – et de la durée passée dans cet état. L'individu initialement agriculteur aura une certaine probabilité de se marier, supposée, pour commencer, identique pour tous les agriculteurs. Mais si cet individu décide de quitter l'agriculture, son comportement matrimonial changera, car il se sera libéré de toutes les contraintes financières et matérielles que lui imposait son statut. De même, un autre individu initialement célibataire aura une certaine probabilité de quit-

ter l'agriculture, supposée, pour commencer, identique pour tous. Si cet individu décide de se marier, alors son comportement vis-à-vis du monde agricole pourra changer, car il se retrouve dans de nouvelles conditions avec, par exemple, un apport de terres venant de son conjoint.

On pourra bien entendu faire dépendre le quotient de nuptialité des individus sortis du secteur agricole non seulement de leur âge mais également de la durée passée hors de l'agriculture et, de façon réciproque, faire dépendre le quotient de sortie de l'agriculture des individus mariés, non seulement de leur âge mais aussi de la durée écoulée depuis leur mariage. Il sera également possible de comparer le comportement des individus sortis de l'agriculture vers une autre profession à celui des individus de cette même profession n'ayant pas suivi ce trajet, etc.

Les méthodes d'analyse biographique permettent, par exemple, d'estimer des quotients de nuptialité des agriculteurs restés dans ce secteur et sortis de ce secteur, des quotients de départ de l'agriculture des célibataires et des mariés, etc. Elles rendent possible également l'estimation des variances des divers quotients par lesquels on évalue les différences de comportement selon l'état dans lequel se trouve l'individu. Elles permettent enfin de faire intervenir les dates d'arrivée de l'événement antérieur et de voir l'effet de la durée séparant les deux événements.

Dans le cas étudié ici, on trouve comme principaux résultats : pour les hommes, les quotients de nuptialité sont beaucoup plus élevés lorsqu'ils sont sortis de l'agriculture que lorsqu'ils y restent. Dans l'autre sens, il n'y a pas de différence entre les quotients de sortie de l'agriculture que les hommes soient célibataires ou mariés. Pour les femmes, les quotients de nuptialité restent les mêmes, qu'elles restent dans l'agriculture ou qu'elles en sortent. Dans l'autre sens, les quotients de sortie de l'agriculture sont très élevés lorsqu'elles sont célibataires, et sont fortement réduits si elles sont mariées.

(1) NDLR : un processus stochastique est un processus qui comporte la présence d'une variable aléatoire.

Cette analyse met en évidence des dépendances complexes entre deux événements, dans la mesure où l'on suit leur apparition tout au long du temps. Par exemple, des dépendances unilatérales (Schweder, 1970) apparaissent entre les deux événements étudiés, lorsque l'arrivée d'un des phénomènes influe sur les chances d'arrivée de l'autre, alors qu'il n'y a aucun effet du second sur le premier. Ainsi, il y a dépendance unilatérale de la sortie de l'agriculture sur la nuptialité, chez les hommes, puisqu'il y a différence entre les quotients de nuptialité des agriculteurs et de ceux qui sont sortis de l'agriculture, mais pas de différence entre les quotients de sortie de l'agriculture, qu'ils soient célibataires ou mariés. Pour les femmes, il y a dépendance unilatérale de la nuptialité sur la sortie de l'agriculture. Si aucune des deux égalités précédentes n'était vérifiée, il y aurait eu dépendance réciproque entre les événements. Si les deux étaient simultanément vérifiées, il y aurait eu indépendance totale.

On voit combien une telle analyse permet de mieux approcher la notion de causalité, même si l'on ne peut arriver à la démontrer dans un sens philosophique strict. Un tel concept de dépendance unilatérale formalise la notion intuitive selon laquelle un processus stochastique peut influencer le développement d'un autre processus stochastique, en les situant correctement dans le temps. Il en résulte que la dépendance unilatérale est une propriété dynamique se développant au cours du temps, et ayant une direction précise, puisque le second processus peut être dépendant du premier sans que la réciproque soit vraie. Il est bien entendu possible d'introduire l'hétérogénéité de la population initiale, à l'aide de modèles qui font intervenir diverses caractéristiques individuelles, certaines pouvant d'ailleurs changer au cours du temps à des instants précis. Ainsi, dans le cas étudié ici, on peut faire intervenir des caractéristiques de l'enquêté : le fait qu'il soit l'aîné de sa famille, son nombre de frères et sœurs, le fait que son père soit agriculteur, etc. On peut aussi prendre en compte les caractéristiques du conjoint, intervenant après le mariage, et des caractéristiques de

l'enquêté intervenant après son départ de l'agriculture lorsque celui-ci se produisait (Courgeau et Lelièvre, 1986).

Il est également possible de considérer divers niveaux d'agrégation, à condition que l'enquête porte sur un nombre suffisant de personnes pour que chaque zone soit suffisamment renseignée. Des méthodes d'analyse multiniveaux se mettent en place pour compléter une telle analyse biographique.

Une réflexion plus générale sur la causalité en sciences humaines

On a montré, tout au long de cet article, une évolution dans les hypothèses utilisées pour expliquer les comportements humains. Cette évolution s'explique par des enquêtes et des mesures toujours plus précises et détaillées, et qui permettent de tester ces hypothèses dans des cas de plus en plus divers. Lorsqu'une hypothèse n'est pas vérifiée sur un nouveau cas, il devient nécessaire de la modifier pour que la nouvelle hypothèse permette d'expliquer l'ensemble des observations. Pour aller plus avant et éclairer le problème de la causalité, on doit distinguer les concepts de déduction de ceux d'induction, qui correspondent à des types de causalité différents.

La déduction permet de dériver des conclusions de prémisses données. Elle part d'hypothèses supposées vérifiées et en dérive des conclusions en utilisant des règles logiques explicites. Ainsi, le *modus ponens* part du fait que l'hypothèse *H* est supposée vérifiée et déduit de façon logique, du fait que *H* implique la conclusion *B*, la vérité de *B*. Lorsque les prémisses sont correctement vérifiées, cette conclusion doit rester identique quelle que soit la façon dont la démonstration est faite.

Lorsque diverses conclusions – issues des mêmes prémisses – sont contradictoires, on peut alors être certain que ces prémisses sont incorrectes. Mais, même si ces

conclusions sont cohérentes entre elles, on ne pourra jamais conclure que les hypothèses de départ sont vraies, car de nombreuses autres prémisses peuvent conduire à la même conclusion. Enfin, une hypothèse vérifiée avec un certain type de données peut ne plus l'être avec des données plus détaillées.

Supposons que les agriculteurs aient tous la même probabilité de migrer, quel que soit leur position sur le territoire : ils sont donc homogènes vis-à-vis de la migration. Cela constitue notre hypothèse de départ, identique à celle d'E. Durkheim, qui considérait que les membres d'un même groupe avaient la même propension à connaître un événement donné, quel que soit leur province de résidence. Dans ce cas, si on ne dispose que de données agrégées, on peut chercher à vérifier qu'il y a bien une variation linéaire du pourcentage de migrants, lorsque le pourcentage d'agriculteurs d'une région croît. Cette linéarité est ici bien vérifiée. Par déduction, on peut conclure que les agriculteurs ont une plus forte probabilité de migrer que les autres professions, sous l'hypothèse d'une homogénéité des comportements sur l'ensemble du territoire.

Rendre cohérentes des conclusions contradictoires

Si on dispose de données individuelles, on devrait s'attendre à une confirmation de la plus forte probabilité d'émigrer des agriculteurs, toujours lorsque l'hypothèse de départ est vérifiée. Dans le cas considéré ici, les agriculteurs ont une bien plus faible probabilité de migrer que les autres professions, contredisant les résultats obtenus avec les données agrégées.

Ces conclusions montrent que l'hypothèse de départ n'est pas correcte. Dès lors, comment modifier cette hypothèse pour rendre cohérentes entre elles ces conclusions contradictoires ? Il faut alors se tourner vers l'induction, dont la définition, cependant, nécessite une discussion plus approfondie.

Si l'on suit l'approche de Karl-Raimund Popper (1973) et si l'on définit l'induction en disant qu'elle permet de répondre à « *la question de savoir comment établir des énoncés universels fondés sur l'expérience* », force est de reconnaître que cette induction ne peut jamais être considérée comme définitive. En effet, il est toujours possible qu'une expérience à venir contredise l'énoncé supposé universel, sans que l'on puisse le savoir *a priori*. Dès lors, toute tentative de fonder ce « *principe d'induction sur l'expérience échouera puisque celle-ci doit conduire à une régression à l'infini* ».

D'autres définitions de l'induction ont cependant, été proposées. Elles permettent d'avancer dans ce domaine. Si l'on se reporte aux définitions données dans différents dictionnaires, on voit que l'induction peut être considérée comme un « *raisonnement qui va du particulier au général* » (Larousse) ou qui permet de remonter « *de cas particuliers à une proposition plus générale* » (Robert). Ces définitions n'impliquent plus la multiplication à l'infini des expériences pour s'assurer de la régularité de la relation d'induction. Au contraire, celle-ci reçoit ici un sens différent de celui que lui donnait K.-R. Popper : il ne s'agit plus de mettre en place une hypothèse universelle, mais de chercher des hypothèses plus générales que les précédentes, qui permettent de réconcilier des conclusions divergentes.

Lorsqu'un chercheur fait des observations inattendues au vu des théories prévalantes, il est amené à mettre en place d'autres hypothèses pour expliquer ces divergences. On peut qualifier ce raisonnement de processus de « *rétroduction* » (Greenland, 1998). Sans ces observations précises, ces nouvelles hypothèses n'auraient pas émergé. Les données les ont générées. Pour les chercheurs qui identifient induction et rétroduction, le problème n'est plus de savoir si cette dernière existe. Depuis longtemps elle s'est révélée être très féconde pour la recherche.

Devant l'incohérence des résultats obtenus avec l'hypothèse que les agriculteurs ont tous la même propension à l'émigration, de

nouvelles conditions doivent être posées pour permettre de résoudre ces contradictions, et ce de la manière la plus simple possible.

Le modèle contextuel

Un modèle contextuel fournit une des solutions, en faisant intervenir une propension individuelle à la migration, qui sera modifiée par les conditions particulières de chaque région dans lesquelles vivent ces individus. C'est le cas notamment dans des régions à fort pourcentage d'agriculteurs. La probabilité pour l'ensemble de la population d'en émigrer sera plus importante en raison de la grande rareté des emplois agricoles.

De même, lorsque l'on suppose que l'action d'événements individuels est seulement immédiate, les problèmes rencontrés conduisent à une approche biographique qui étend cette action tout au long de l'existence, tout en permettant un effet du moment de certaines caractéristiques. Enfin, pour accorder les deux généralisations précédentes, il est nécessaire de mettre en place une approche biographique en multiples niveaux.

Ces divers résultats peuvent dès lors conduire à modifier certains paradigmes sur lesquels diverses sciences humaines sont basées. Un paradigme définit ce qui est légitime dans le champ scientifique qu'il couvre. Par nature, il résiste donc à une définition précise, mais peut être délimité par des principes généraux. Ainsi, le paradigme antérieur selon lequel le démographe ne peut étudier que l'arrivée d'un événement et d'un seul, sous l'hypothèse qu'il est indépendant des autres phénomènes et qu'il se produit dans une population supposée homogène, ne paraît plus tenable au travers des diverses observations. On a proposé un nouveau paradigme (Courgeau et Lelièvre, 1996), qui permet d'appuyer l'approche synthétique préconisée ici : un individu parcourt tout au long de sa vie une trajectoire complexe dépendant, à un instant donné, de sa trajectoire antérieure, des

informations qu'il a pu acquérir dans son passé et des conditions qui prévalent dans la société où il vit.

Ce paradigme permet de faire intervenir les divers événements d'une vie, comme des processus en interaction les uns avec les autres, et constitue une approche intégrant les différents niveaux d'agrégation, allant du micro au macro. Ces divers événements peuvent maintenant être dépendants les uns des autres et il est possible de faire intervenir l'hétérogénéité de la population sur laquelle on travaille. Il permet également de compléter cette approche de la vie individuelle par une étude des changements survenant dans les règles en vigueur dans les divers niveaux d'agrégation. Ainsi, des actions individuelles isolées effectuées dans une communauté donnée peuvent amener la prise de conscience d'un problème qui affecte l'ensemble de cette collectivité. Dès lors, cela peut conduire à des mesures politiques prises à un niveau plus agrégé. Bien entendu, ces mesures affecteront les conduites individuelles, provoqueront de nouvelles actions pour contrebalancer leur effet pervers, etc.

Mieux structurer les sciences sociales

Enfin, parallèlement à ces nouvelles hypothèses, induites pour répondre aux besoins des sciences humaines, des méthodes d'analyse permettent d'en déduire les effets observés. Il s'agit de l'analyse contextuelle ou multi-niveaux (Courgeau et Baccaini, 1997) et de l'analyse biographique (Courgeau et Lelièvre, 1989).

A partir de la seule observation de phénomènes, on ne peut induire une causalité au sens strict en sciences sociales. Une régularité empirique ne comporte aucun caractère de nécessité et conduit à une régression à l'infini. En revanche, il est possible d'induire par « rétroduction » de nouvelles hypothèses qui permettent, à l'aide d'un paradigme, de mieux structurer ces sciences et d'éviter des conclusions contradictoires. Sa validité devra être tes-

tée en permanence, pour voir s'il couvre bien tous les événements étudiés.

Ce paradigme permet d'appréhender la vie individuelle comme un tout, où les divers domaines de l'existence vont influencer les uns sur les autres, sans séparation arbitraire entre vie familiale et vie professionnelle, et où des événements survenus dans un passé lointain peuvent être actualisés pour jouer un rôle dans une nouvelle situation. L'analyse biographique permet d'approcher la causalité, sans cependant jamais l'atteindre, et relie entre eux les divers événements de la vie, en les replaçant les uns par rapport aux autres. Cette première approche fait intervenir la vie des individus dans toute sa complexité, à la fois temporelle et spatiale. Elle autorise également la mise en relation avec la vie de leurs proches (Lelièvre et alii, 1998).

Au-delà de cette première approche, ce paradigme montre qu'il peut y avoir des effets d'un niveau plus agrégé sur les comportements individuels. Cependant, de tels effets n'apparaissent plus, tels qu'E. Durkheim les voyait, comme des formes causales extérieures aux individus mais, au contraire, comme l'émergence de formes, qui paraissent organisées à un niveau d'agrégation supérieur à celui de l'individu et qui résultent d'un ensemble d'actions individuelles parfaitement rationnelles. Comme la théorie des systèmes complexes le montre, l'agrégation des comportements individuels ne se fait pas de façon additive et conduit à de nouvelles formes de causalité. De tels effets peuvent apparaître comme irrationnels pour les acteurs eux-mêmes et même contradictoires avec leurs propres actions. Ces effets non intentionnels sont, cependant, aussi importants à considérer que les effets intentionnels de ces mêmes actions individuelles.

L'analyse contextuelle met en évidence un effet des caractéristiques agrégées à différents niveaux sur les comportements, tandis que les caractéristiques individuelles correspondantes ont un tout autre effet indépendant du précédent. Cette analyse montre l'existence de ces effets non intentionnels et

détermine leur influence. En faisant intervenir à la fois les caractéristiques individuelles et agrégées, on est à même de souligner la complexité du rôle des divers phénomènes auprès du groupe ou de l'individu, et s'ils l'assurent efficacement.

A l'aide en particulier d'analyses multi-niveaux, ce paradigme devrait enfin permettre de dévoiler la structure complexe du système social dans lequel vivent les individus, en déchiffrant les régularités qui existent entre les divers niveaux d'agrégation et les changements qui peuvent intervenir dans cette structure. Il est nécessaire de détecter les niveaux d'agrégation les plus pertinents, de mettre en évidence la force des contraintes qu'ils imposent aux individus, et d'analyser la façon dont ils peuvent se modifier et changer leurs effets. Il s'agit d'une troisième forme d'action réciproque du groupe sur l'individu et de l'individu sur le groupe, leur permettant une évolution permanente.

Pour une autre théorie des comportements sociaux

Un tel paradigme est encore rarement utilisé dans toutes ses dimensions, car il nécessite des données correspondant à la fois au niveau individuel, issues d'enquêtes biographiques explorant le maximum de domaines de la vie individuelle (famille, profession, migration, éducation, etc.), et à différents niveaux agrégés, de telle sorte que l'on puisse disposer de suffisamment d'individus dans chaque unité d'un niveau – qui n'est pas nécessairement géographique – pour pouvoir réaliser une analyse multi-niveaux avec des résultats significatifs. Toutes les possibilités de ce paradigme n'ont pu être démontrées car le recueil de telles données est encore trop restreint, en particulier pour des raisons de coût.

Finalement, cette approche devrait mettre en place une théorie plus satisfaisante des comportements sociaux, dont les bases méthodologiques et épistémologiques restent encore largement à établir mais pour-

raient être tirées du paradigme que l'on a proposé. Les recherches futures diront la richesse d'une telle piste, qui aborde simultanément les divers niveaux d'agrégation

apparaissant dans les sciences sociales et permet de mieux préciser ce que l'on peut y entendre sous le terme de causalité.

Références bibliographiques

Andersen P., Borgan O., Gill R. et Keiding N., *Statistical Models based on Counting Processes*, Springer-Verlag, New York, 1993.

Baccaïni B. et Courgeau D., *Approche individuelle et approche agrégée : utilisation du registre de population norvégien pour l'étude des migrations*, in Bocquet-Appel J.-P., Courgeau D. et Pumain D. (sous la dir.), *Analyse spatiale de données biodémographiques*, John-Libbey/INED, Paris, 1996.

Courgeau D. et Baccaïni B., *Analyse multi-niveaux en sciences sociales*, *Population*, 1997, vol. 52, n° 4 ; *Multilevel Analysis in the Social Sciences*, *Population. An English Selection*, 1998, vol. 10, n° 1.

Courgeau D. et Lelièvre E., *Nuptialité et agriculture*, *Population*, 1986, vol. 40, n° 2.

Courgeau D. et Lelièvre E., *Analyse démographique des biographies*, Editions de l'INED, Paris, 1989 ; *Event History Analysis in Demography*, Clarendon Press, Oxford, 1992.

Courgeau D. et Lelièvre E., *Changement de paradigme en démographie*, *Population*, 1996, vol. 51, n° 2, *Changing Paradigm in Demography*, *Population. An English Selection*, 1997, vol. 9.

Duchêne J., Wunsch G. et Vilquin E. (sous la dir.), *L'explication en sciences sociales : la recherche des causes en démographie*, Editions Ciaco, Bruxelles, 1989.

Durkheim E., *Le suicide*, PUF, Paris, 1930 (première édition 1896).

Franck R. (sous la dir.), *Faut-il chercher aux causes une raison ? L'explication dans les sciences humaines*, Librairie philosophique Vrin, Paris, 1994.

Greenland S., *Induction Versus Popper : Substance Versus Semantics*, *International Journal of Epidemiology*, 1998, n° 27.

Herpin N., *La famille à l'épreuve du chômage*, *Economie et Statistique*, 1990, n° 235.

Lelièvre E., Bonvalet C. et Bry X., *Analyse biographique des groupes. Les avancées d'une recherche en cours*, *Population*, 1998, vol. 52, n° 4 ; *Event History Analysis of Groups. The Findings of an on-going Research*, *Population. An English Selection*, 1999, vol. 10, n° 1.

Nezosi G., *Chômage et famille. Synthèse des connaissances*, in « Famille et chômage » (sous la dir. de Commaille J.), rapport pour le Haut conseil de la population et de la famille, Paris, 1999.

Paugam S., *Précarité et risques d'exclusion en France*, *Document du CERC*, La documentation Française, Paris, 1993, n° 109.

Piaget J. (sous la dir.), *Logique et connaissance scientifique*, Encyclopédie de la Pléiade, Editions Gallimard, Paris, 1967.

Popper K.-R., *La logique de la découverte scientifique*, Editions Payot, Paris, 1973 (première édition 1935).

Robinson W.-S., *Ecological Correlations and the Behavior of Individuals*, *American Sociological Review*, 1950, n° 15.

Schweder T., *Composable Markov Processes*, *Journal of Applied Probability*, 1970, n° 7.

Wunsch G., *Causal Theory and Causal Modelling*, Leuven University Press, Leuven, 1988.