

## Compte rendu

---

### Ouvrage recensé :

Bouchard, Gérard, Raymond Roy et Bernard Casgrain. — *Reconstitution automatique des familles. Le système SOREP*. Chicoutimi, Centre interuniversitaire de recherche sur les populations, Université du Québec à Chicoutimi, Université Laval, Université McGill, 1985, 2 volumes, 521 p. et 224 p.

par Bertrand Desjardins

*Cahiers québécois de démographie*, vol. 16, n° 2, 1987, p. 306-310.

Pour citer ce compte rendu, utiliser l'adresse suivante :

URI: <http://id.erudit.org/iderudit/600619ar>

DOI: 10.7202/600619ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

---

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

---

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : [info@erudit.org](mailto:info@erudit.org)

\*\*\*

BOUCHARD, Gérard, Raymond ROY et Bernard CASGRAIN. - *Reconstitution automatique des familles. Le système SOREP.* Chicoutimi, Centre interuniversitaire de recherche sur les populations, Université du Québec à Chicoutimi, Université Laval, Université McGill, 1985, 2 volumes, 521 p. et 224 p.

Les démographes ne savent que trop à quel point l'étude des populations est à la merci des sources. Les registres de population ont à cet égard de quoi faire rêver. Consignant en principe pour l'ensemble d'une population les événements d'état civil et les changements de résidence au fur et à mesure qu'ils se produisent, ils constituent une observation suivie, théoriquement parfaite. Diverses administrations, passées et présentes, en ont mis sur pied, notamment en Scandinavie, en Belgique et en Hollande, et en quelques pays d'Asie; mais leur maintien n'a toujours pu se faire qu'à des degrés plus ou moins satisfaisants, de sorte que leur exploitation statistique pose de nombreux problèmes. Il est également possible de dresser rétroactivement de tels registres, pour autant que les documents nominatifs appropriés soient disponibles; il s'agit alors de relier à l'échelle des individus, des familles ou encore des ménages, l'ensemble des enregistrements, disséminés dans le temps et l'espace, qui les concernent.

Le Québec français est choyé à ce point de vue, puisqu'il dispose d'un bon enregistrement des faits d'état civil depuis l'arrivée des premiers colons, grâce aux registres de baptêmes, mariages et sépultures diligemment tenus par le clergé; bien plus, de nombreuses autres sources - recensements, greffes de notaire, archives de communautés religieuses, d'hôpitaux, de maisons d'enseignement, de compagnies, etc. - sont disponibles pour compléter et bonifier les informations de base tirées des registres paroissiaux. Cette richesse documentaire a suscité la mise sur pied de deux vastes projets visant à élaborer rétroactivement des registres de population québécois. L'un d'eux, le Programme de recherche en démographie historique (P.R.D.H.), créé par Hubert Charbonneau et Jacques Légaré à l'Université de Montréal, porte sur l'ensemble de la population depuis ses origines jusqu'au début de l'ère statistique au XIXe siècle; l'autre, qui nous intéresse ici, a été lancé il y a une quinzaine d'années à l'Université du Québec à Chicoutimi par Gérard Bouchard, dans le but d'étudier le développement d'une société régionale, celle du Saguenay, observée depuis le début de la colonisation blanche vers 1840 jusqu'à nos jours. Structuré depuis sous la forme d'un centre interuniversitaire de recherche sur les populations (SOREP), le projet saguenayen consacre maintenant une large part de ses activités à des questions de nature génétique; son aire géographique s'est élargie et inclura éventuellement l'ensemble des régions et sous-régions du nord-est québécois.

Une telle entreprise implique la manipulation d'éléments qui se chiffrent dans les millions : noms et prénoms, dates, lieux, caractéristiques diverses, liens filiaux et conjugaux, consignés à l'échelle de mentions, d'actes et de dossiers individuels et familiaux. Elle exige donc le recours à l'informatique et, en corollaire, l'élaboration d'un important appareil méthodologique correspondant à toutes les étapes d'une chaîne de traitement : l'extraction des éléments pertinents des documents et leur acheminement aux supports magnétiques; le couplage de l'information, dans le but d'identifier les individus ainsi que les liens filiaux et conjugaux qui les unissent, en même temps que les variables propres à chacun; la structuration, sous une forme commode pour l'exploitation, des informations ainsi rassemblées; l'analyse statistique des données. La phase de couplage, c'est-à-dire la mise en rapport des mentions de chaque individu dans les documents, pose des problèmes particulièrement redoutables, étant donné les aléas de l'identification nominative des individus, auxquels l'ordinateur est par essence extrêmement réfractaire. Le sort de toute l'entreprise repose cependant sur cette étape, et c'est pourquoi les chercheurs de SOREP ont jugé bon de rassembler en une seule publication les

éléments de la méthodologie qu'ils ont patiemment développée au fil des ans, et dont l'essentiel a déjà fait l'objet de documents de travail et d'articles particuliers.

L'ouvrage se présente en deux volumes, le deuxième étant constitué de vingt-trois annexes regroupant des éléments techniques reliés aux différents aspects de la méthodologie; fiches, algorithmes, listes et exemples. On pourrait s'étonner que la présentation porte sur la reconstitution des familles, alors que l'unité d'observation de base du registre de population est l'individu. C'est que la démarche s'inscrit dans une tradition, celle de la démographie historique, dont l'essor résulte principalement du développement de cette technique, créée dans le but d'étudier la fécondité. La contradiction n'est en réalité qu'apparente : lorsqu'on analyse cette opération classique, on constate qu'elle correspond en fait à établir précisément les biographies des individus. Mais en démographie historique, le concept de reconstitution des familles en est venu à recouvrir à la fois les opérations de couplage et la structuration de l'information pour la conservation et l'analyse, sous forme de «fiches de famille»; conditionnée par l'outil, l'analyse dans le cadre classique a ensuite privilégié l'unité familiale comme base d'observation. Dans le cas présent, il faut comprendre qu'il s'agit de l'établissement des biographies individuelles, que l'on effectue, sans préjudice pour la suite, par le biais de la reconstitution des familles, parce que l'identification des individus est facilitée lorsqu'on les situe dans leur cadre familial.

Le volume principal compte neuf parties. La première, qui fait office d'introduction, propose un survol rapide du domaine du jumelage automatique des données en général et de la reconstitution automatique des familles en particulier, et présente le projet SOREP dans son ensemble. La deuxième partie présente successivement les processus de dépouillement et de saisie des données, une critique de celles-ci, et les principes du système SOREP. La question des variations nominatives y mérite un développement à part, parce qu'elle est déterminante pour tout ce qui touche le jumelage automatique des données. Dans ce cas-ci, le découpage optimal des opérations menant à la reconstitution des familles étant déjà reconnu grâce aux travaux du P.R.D.H., le problème a consisté à définir une méthode visant à surmonter cette difficulté. Les parties trois à sept présentent donc la solution apportée au problème, soit la construction d'une série d'outils pour traiter les divers types de variations rencontrées, le tout assorti de tables de décision ad hoc. La huitième partie fournit les données critiques et techniques du système, et la neuvième conclut par la discussion

de quelques problèmes liés à la constitution et à la gestion du fichier, et par les perspectives d'avenir. Une bonne bibliographie et un glossaire des termes techniques coiffent l'ensemble.

L'annexe 16 donne des exemples qui montrent de façon particulièrement éloquente à quel degré d'imprécision peuvent conduire les caprices de l'identification nominative et, partant, la complexité du problème qui confronte l'entreprise. On ne s'étonne donc pas du caractère essentiellement empirique de la méthode qui, dans ses derniers retranchements, touche en quelque sorte aux limites de l'automatisation. On se demande à ce propos si le pari de l'automatisation maximale n'était pas tout compte fait trop optimiste, et s'il n'y a pas lieu de reconnaître à l'intervention humaine ponctuelle une part plus large dans le processus. Ceci est particulièrement vrai dans un contexte de migrations, qui pose à l'ensemble de la démarche un défi supplémentaire qui aurait probablement mérité plus ample discussion. Car, en définitive, est-il vraiment possible de statuer dans les cas comptant relativement peu d'observations, entre une identification nominative particulièrement déficiente et la mention d'un couple mobile, de passage dans la région ?

Tout au long de la présentation, les auteurs font preuve d'un remarquable souci critique face aux différentes composantes de leur méthodologie. À chaque instant une évaluation des résultats est soumise au lecteur. La chose est difficile, car la seule évaluation définitive - celle qui consiste à répondre à la question «les biographies du registre sont-elles exactement conformes à celles de l'ensemble des individus qui ont vécu dans la région ?»- est impossible à réaliser. Il faudrait en particulier connaître les gens et les événements de l'extérieur de la région, qui seuls peuvent donner la clé des situations les plus ambiguës. On cherche donc à l'approcher indirectement, en s'assurant de l'exactitude et de la fiabilité de chacun des éléments de la chaîne de traitement. Les chercheurs de SOREP ont fait à cet égard en effort remarquable.

La nature même de ces travaux méthodologiques confère au présent ouvrage un caractère extrêmement aride, propre à rebuter même l'éditeur le plus indulgent - pour ne pas parler des éventuels lecteurs. De ceci, nul ne peut être plus conscient que les auteurs eux-mêmes. À preuve, les deux phrases qui précèdent sont directement tirées de l'avant-propos de l'ouvrage !

On comprend donc pourquoi on est ici en présence d'une publication «maison», deuxième dans une série de «dossiers» du Centre. Les chercheurs de SOREP justifient ainsi leur démarche : «Dans l'esprit même où ces recherches ont été réalisées, l'idée qu'elles puissent un jour de quelque manière servir ou susciter d'autres recherches nous a conduits à en consigner la nature et les résultats de manière à leur assurer une certaine diffusion, si modeste soit-elle. En outre, par souci de transparence méthodologique et pour favoriser le regard critique sur notre démarche, nous avons tenu à en montrer à la fois le détail et les justifications».

Nous avons abordé ci-dessus ce deuxième point; quant au premier, il est déjà atteint, puisque nous avons lu avec un grand intérêt ce travail, qui nous a fourni ample matière à réflexion. Mais, est-il nécessaire de le dire, l'ouvrage ne s'adresse qu'à un public extrêmement restreint.

Bertrand DESJARDINS

\*\*\*