

**CONSTITUTION ET GESTION DES FICHIERS
DE POPULATION ELABORES A PARTIR
DES REGISTRES D'ETAT CIVIL :
GENEALOGIE ET INFORMATIQUE.**

J. L. BONNIOL (*)

Le problème du dépouillement et de l'exploitation du matériel fourni par les registres d'état civil résulte de l'abondance et de la pulvérisation des données : on est en présence de milliers d'actes (la population d'un groupe est facilement multipliée par 10 lorsqu'on prend en compte plusieurs générations...) et ces actes sont épars, sans rapport entre eux... La première tâche consiste donc à introduire un certain ordre dans cette masse indifférenciée d'éléments émettés (naissance, mariages, décès...), afin d'obtenir des séquences individuelles significatives :

naissance - mariage(s) / divorce(s) - décès

séquence à laquelle il faut ajouter les différents épisodes de procréation.

La séquence de base est maintenant disponible dans les registres d'état civil, sous la forme de "mentions marginales" à l'acte de naissance (1892 : mention du mariage, avec date, lieu et nom du conjoint ; 1945, mention du décès (avec date et lieu), mais ce ne seront que les générations futures de chercheurs qui pourront pleinement profiter de ces améliorations.

(*) E.R. 221 du CNRS et Laboratoire D'Ecologie Humaine de l'Université D'AIX-MARSEILLE III.

Et, en tout état de cause, ils ne pourront toujours pas avoir accès directement à l'activité procréatrice des individus, dans la mesure où n'existe pas encore une mention des paternités ou des maternités sur l'acte de naissance des individus : ce type de renseignements n'existe que sur un document privé élaboré à la suite du mariage, le livret de famille... Cette activité procréatrice est certes recélée dans les registres, mais ne figure pas en regard d'un acte concernant l'individu, qui n'apparaît qu'en tant que géniteur dans l'acte de naissance d'un autre individu.

1 - LES METHODES DE LA DEMOGRAPHIE HISTORIQUE

Avant ces mentions marginales, l'état civil ancien apparaît comme un inventaire sélectif. Il retient les individus selon un rigoureux critère de localisation de l'évènement : une naissance, un mariage, un décès ne sont transcrits dans les registres que s'ils ont lieu dans la commune, et la migration est la grande absente... C'est la raison pour laquelle les méthodes mises au point dans les années cinquante par le démographe L. Henry (décrites dans le Nouveau manuel de dépouillement et d'exploitation de l'état civil ancien et dans le Manuel de démographie historique) ont été élaborées dans le cadre de monographies locales rendant compte de population anciennes relativement closes sur elles-mêmes.

Au delà de la restitution des séquences de base individuelles, la tâche essentielle de ce type d'études a été de raccrocher aux individus (en particulier aux mères) leurs différents épisodes de procréation : c'est la célèbre "reconstitution des familles", oeuvre énorme exigeant, pour la moindre cellule villageoise, des mois de travail minutieux et des manipulations innombrables.

Mais le développement de la démographie historique était à ce prix : le "noyau dur" de la discipline a résidé en effet dans le dépassement des taux classiques des démographes au profit d'une étude fine de la fécondité (l'un des principaux problèmes qu'elle a pu ainsi aborder est celui du passage d'une fécondité "naturelle" à une fécondité "contrôlée"). Et toute étude de fécondité, c'est-à-dire la mise en rapport des naissances avec les mères concernées, ne peut passer que par une telle reconstitution préalable.

2 - VERS L'INFORMATISATION DES PROCEDURES

Cet énorme travail de dépouillement et d'exploitation des données invitait bien-sûr à une informatisation des procédures, de manière à relier par une chaîne d'opérations deux séries d'éléments matériels : à un bout, les registres manuscrits consignants les diverses catégories d'actes, à l'autre les tableaux imprimés des divers résultats... Diverses tentatives témoignent de ce désir : proposition d'A. Chamoux, de R.S. Schoffield et de J.C. Perrot dans le numéro spécial Familles et sociétés, Annales E.S.C., 4-5 : 1972 ; mise au point au Canada d'une reconstitution automatique des familles pour l'étude exhaustive de la population québécoise (P. Beauchamp, R. Roy et J. Légaré, "Reconstitution automatique des familles sur le programme Hochelaga II, Population et famille, 33 (1974-3 pp. 1-40). Dans cette ligne, la constitution automatique de tableaux démographiques a donné lieu au programme (Calculs et analyses sur ordinateur appliqués aux reconstitutions, édité par la Société de démographie historique en 1981).

L'ontrée en scène des micro-ordinateurs, en particulier des portables, a d'autre part incité à restreindre au maximum les phases de travail intermédiaires entre la lecture de l'acte et son enregistrement informatique. C'est dans cet esprit qu'a été élaboré la brochure de J.P. Pellissier, Démographie, généalogie, micro-Informatique, éditée par la Société de démographie historique en 1983. L'auteur propose, au delà de la collecte codifiée des données, une possibilité de saisie directe sur support informatique ainsi que la constitution, par l'ordinateur, de fichiers successifs. Mais la machine ne fait qu'apporter une aide à la reconstitution des familles, qui continue à devoir être effectuée "manuellement".

Les comportements démographiques ne sont pas les seuls phénomènes auxquels permettent d'accéder les registres d'état civil. Dans la mesure où chaque individu est relié à son père et à sa mère, il est possible de dénouer des réseaux généalogiques et de faire apparaître des ensembles parentaux. Ce type de données peut être exploité à diverses fins. Signalons pour mémoire l'entreprise gigantesque de la Genealogical Society qui, dans les souterrains de Salt Lake City, organise l'entrée sur support informatique du contenu de millions d'actes systématiquement collectés à travers le monde, cela à des fins religieuses... Dans un tout autre esprit, l'anthropologie sociale, abordant l'étude des sociétés complexes, a compris tout le parti qu'elle pouvait tirer de l'utilisation de telles archives écrites pour approfondir un certain nombre de thèmes centraux de la discipline, comme les systèmes de parenté et les modes d'organisation des alliances.

Dans le même temps, le recours à l'ordinateur lui permettait de tester la validité de certaines hypothèses théoriques qui n'auraient pu être vérifiées autrement en raison de la masse du matériel à analyser : peut-on déceler des régularités dans le mode de choix de son conjoint, en l'absence de toute prescription énoncée, ainsi que des structures d'échanges et des processus de réciprocité entre des ensembles définis de consanguins ? Du côté de la génétique des populations enfin, les travaux fondés sur des données généalogiques provenant d'actes officiels se sont multipliés, ouvrant de nouvelles voies pour l'étude de l'évolution des patrimoines génétiques. Ils peuvent faire appel à divers programmes disponibles, utilisant le triplet individu-père-mère, qui concernent essentiellement la consanguinité (somme étendue à tous les chemins possibles reliant l'individu à ceux de ses ancêtres qui appartiennent à la fois au réseau d'ascendance de son père et au réseau d'ascendance de sa mère) et la probabilité d'origine des gènes (programmes élaborés par C. Garçon et A. Jacquard : "Programmes de calcul..." in Génétique et population, I. N. E. D., Paris, 1971).

3 - L'EXEMPLE DES RECHERCHES SUR TERRE-DE HAUT DES SAINTES ET LA DESIRADE

Nous avons pour notre part commencé à mettre au point une méthode qui s'appuie sur tout l'acquis de la démographie historique, mais englobe une finalité anthropologique et génétique. Le principe est d'aller à l'essentiel (en laissant de côté, de manière provisoire dans la phase d'exploitation, mais en se réservant la possibilité

d'y revenir ultérieurement, un certain nombre de renseignements qui alourdissent considérablement le travail en démographie historique : ceux touchant à la parenté rituelle, aux témoins, à la présence ou absence de signature) pour se concentrer sur les événements démographiques majeurs de la vie des individus, en constituant un fichier de biographies individuelles, reliées les unes aux autres par un codage raisonné. Chaque individu reçoit à la naissance un "numéro d'entrée" ; il est donc possible de reporter sur chaque enregistrement "biographique" individuel le numéro correspondant du père, de la mère et du ou des conjoint(s) éventuel(s).

Par là est rendue possible, de proche en proche, la connaissance de tous les réseaux de filiation (et d'alliance) de la population : c'est là la condition indispensable à toute étude ultérieure de génétique des populations.

On voit par là, grâce à cette identification codée des individus, combien il est facile, à l'aide d'un programme de tri classique, de reconstituer des fratries par rapport à chaque géniteur, qu'il s'agisse des pères ou des mères. Un tel procédé offre en outre l'avantage de s'écarter de la conception européo-centriste implicitement contenue dans les méthodes classiques de la démographie historique, mises au point à partir des populations européennes anciennes à fort taux de légitimité et monogames (les "fiches de famille" étant établies à la suite des mariages) et donc peu adaptées aux sociétés à fort taux d'illégitimité, comme celles de la Caraïbe, ou aux sociétés polygames...

a - Une progressive amélioration des
procédés de dépouillement

Cette méthode a été appliquée lors de deux études sur des petites populations insulaires isolées au large de la Guadeloupe, Terre-de-Haut des Saintes et la Désirade. Les procédés employés ont été progressivement améliorés.

La première tentative a exigé, à partir de "cahiers de collecte", le remplissage à la main de fiches individuelles, dont le contenu a été ensuite reporté sur cartes perforés (cf. notre ouvrage Terre-de-Haut des Saintes et l'article "Approche généalogique d'une population insulaire..." (1).

La deuxième tentative, en cours, sur l'île de la Désirade, a consisté à dépouiller immédiatement sur des feuilles standardisées de données de 80 cases, de manière à pouvoir saisir directement des séquences de caractères correspondant au nom de famille, au prénom principal et aux initiales des autres prénoms, au sexe, à la date de naissance, à la date de décès, à la date et au lieu de mariage. Les séquences de cases devant porter les identifications numériques d'individus (ego, père, mère, conjoint) n'ont pu à ce stade être remplies : ces données, non encore codées, ont été reportées au dessous des lignes individuelles.

(1) Jean-Luc BONNIOL, Terre-de-Haut des Saintes. Contraintes insulaires et particularisme ethnique dans la Caraïbe. Ouvrage édité avec le concours du C.N.R.S., Paris, Edition Caribéennes, 1980 et "Approche généalogique d'une population insulaire : Terre-de-Haut des Saintes, dans la Caraïbe, Annales E.S.C., 1980-6, pp. 1143-1170.

Une première saisie, suivie d'une numérotation séquentielle des individus, a permis l'édition d'un fichier numérique et d'un fichier alphabétiques, facilitant la recherche des ascendants et des conjoints, puis le report de leur numéro d'identification sur les enregistrements individuels. Dans les deux cas, le fichier définitif a été élaboré à partir des registres de naissance, puis complété à partir des données provenant des registres de mariage et de décès. De même un fichier complémentaire, rassemblant les individus qui n'apparaissent qu'en tant que géniteurs, et qui constituent donc l'arrêt de l'information généalogique, a été élaboré "manuellement" : c'est le fichier des "fondateurs" (individus nés avant la date du dépouillement mais aussi immigrants) qui a ensuite été concaténé au fichier central.

Le principe de base est donc resté constant. Contrairement aux méthodes classiques de la démographie, pérennisées même à l'heure du traitement informatique, imposant l'élaboration de plusieurs fichiers, le fichier définitif de population rassemblant les diverses biographies individuelles commence ici à être constitué dès le départ à partir des actes de naissance ; il est progressivement mis au point par l'adjonction de données codées concernant géniteurs et conjoints, ce qui permet de relier les individus les uns aux autres et de constituer un réseau qu'il est ensuite possible d'analyser.

b - Traitement des données

Une fois le fichier définitif constitué, il a fallu dans chaque cas en soumettre les données à des programmes de vérification pour éliminer, dans la mesure du possible, toutes les erreurs ayant pu se glisser dans les diverses phases de dépouillement et d'exploitation.

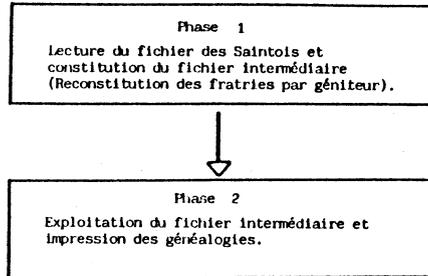
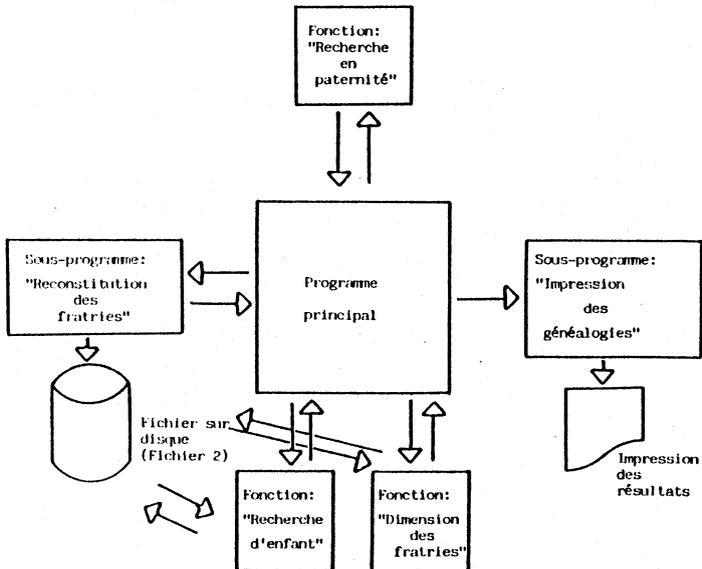
Il est nécessaire pour cela de vérifier la cohérence interne des enregistrements individuels (pas d'anomalie sur les dates, sur les âges au décès, au mariage...) mais aussi opérer des rapprochements entre enregistrements (les conjoints sont de sexe opposé, les numéros des conjoints correspondant, les enfants naissent du vivant de leur mère, entre l'âge de 15 ans et celui de 50 ans, les enfants naissent du vivant de leur père, ou tout au moins dans un intervalle de 9 mois après la mort de celui-ci...).

Le traitement des données a ensuite été le même pour les deux études : reconstitution des fratries, grâce à un tri sur les numéros de père puis de mère et à l'aide d'un programme d'édition suivi de l'enregistrement de tous ses enfants. Sont ainsi rassemblés les renseignements correspondant à la traditionnelle fiche "Henry" utilisée pour la reconstitution des familles. C'est à partir du fichier ainsi trié qu'il est possible de procéder à une étude de la légitimité et des comportements familiaux (présence ou absence de mariage, place du mariage par rapport aux naissances, absence du père, procréation avec plusieurs partenaires) ainsi que de la fécondité, à partir de l'âge des mères à la naissance de chaque enfant et de intervalles proto et intergénéraliques.

En ce qui concerne les autres comportements démographiques, le principe a consisté à conduire des analyses longitudinales en reconstituant des histoires de cohortes (quel est le destin, au regard d'un phénomène, d'un ensemble d'individus choisis en fonction d'un certain critère ?).

Le triplet numéro d'individu - de père - de mère a enfin permis de restituer les réseaux généalogiques au sein de chaque population. Des généalogies ascendantes peuvent être facilement effectuées à la main, à condition d'y prendre la temps (sauf dans le cas de paternité inconnue).

Mais seule la machine permet de retracer des généalogies descendantes à partir des individus "fondateurs", les seules à englober toute la population. La logique du programme utilisé consiste dans l'élaboration d'un fichier intermédiaire rassemblant les diverses fratries (chaque fratrie étant disposée sur un enregistrement, ses membres étant identifié par leur numéro) et d'une table des géniteurs correspondante. La restitution des généalogies repose alors sur un certain nombre de boucles emboîtées. Chacune permet une itération qui correspond pour la première génération au nombre total des fondateurs, desquels partent les chaînes généalogiques descendantes puis, pour les générations suivantes, au nombre d'enfants par géniteur desquels partent des arborescences. Les nécessités de l'édition finale imposent de descendre ces arborescences jusqu'au dernier individu sans descendance, puis de remonter à la génération précédente, et ainsi de suite jusqu'à ce que les différentes boucles soient saturées, de la plus interne à la plus externe.

ORGANIGRAMME GÉNÉRALARTICULATION DU PROGRAMME PRINCIPAL ET DES DIVERS
SOUS-PROGRAMMES ET FONCTIONS.

4 - MISE AU POINT D'UNE PROCEDURE STANDARDISE

Il apparaît désormais possible de mettre au point une procédure standardisée valable pour toute étude de petite population, dans une perspective combinant démographie historique, anthropologie sociale et génétique des populations. Sa logique se rapproche par certains de ses aspects de la démarche proposée par P. Lamaison et M. Selz-Laurière dans un tout récent article ("Généalogies, alliances et Informatique", Terrain 4, mars 1985). Il s'agit en effet dans chaque cas de mettre sur pied un fichier standard définitif, dont chaque enregistrement doit présenter la forme suivante :

I P M

"I, P, M, étant les noms des variables dont les valeurs sont les codages respectifs de chaque individu, de son père, de sa mère... Chaque enregistrement peut comprendre également un nombre V de variables annexes, caractérisant l'individu I". On aboutit ainsi à un "fichier non ambigu, composé d'enregistrements sous la forme :

I P M V

même si, pour certains d'entre eux, une ou plusieurs des variables n'ont pas de valeur connue". L'élaboration de ce "produit" doit passer par deux améliorations essentielles par rapport à nos précédentes tentatives :

- début de constitution directe du fichier sur support informatique, dès la phase de dépouillement, grâce à l'utilisation en salle d'archives d'un micro-ordinateur portable et l'appui d'un programme de saisie permettant l'entrée des données dans trois fichiers : un fichier "naissances" devant devenir le fichier définitif, un fichier "mariages" et un fichier "décès" ; ces données peuvent être ensuite transférées dans un gros

centre de calcul pour la poursuite de l'exploitation. Il faut garder la possibilité de collecter des informations parallèles (témoins, parenté rituelle....) mais on ne peut les coder (sauf si on désire les analyser de manière spécifique) : il faut se restreindre en effet à un nombre défini de paramètres.

- réalisation semi-automatique du codage (et donc de l'identification) des générateurs et des conjoints : une procédure entièrement automatique se heurte en effet à ce problème de l'identification formelle de chaque individu, par suite d'ambiguïtés propres aux registres : homonymes, variations et erreurs dans la transcription de l'identité, approximations dans la notation des âges... Nous ne pouvons que souscrire aux constatations de P. Lamalson et de M. Selz-Laurière : "Comment... traiter par programme l'ensemble des opérations d'évaluation des informations, de déduction, voire de discernement par lesquelles le chercheur parvient "à la main" à corriger les inexactitudes et à préparer un corpus généalogique susceptible d'être ensuite analysé ?... La gamme des situations est immense et il n'est pas possible à priori de les imaginer toutes, ni même de définir la façon de solutionner certaines des ambiguïtés qui ont pu être décelées dès la collecte des informations...". Personne effectivement ne semble avoir apporté de solution générale à ce problème : rappelons cependant que dans l'étude de la population désiradienne le codage a été facilité par l'édition de listages alphabétiques pour retrouver les individus. C'est la raison pour laquelle il est proposé une solution médiane fondée sur une interaction permanente entre le chercheur et l'ordinateur. Il est demandé à celui-ci de balayer le fichier "naissances", recherchant les individus sur le patronyme, les prénoms et une approximation de l'année de naissance, et de signaler au

bout de l'opération les cas résolus et non résolus (individus non retrouvés dans le fichier) : il appartient dans ce dernier cas au chercheur de poursuivre lui-même la recherche ou de verser l'individu au fichier provisoire "fondateurs" (individus nés avant ou ailleurs...). La même procédure est valable pour l'incorporation au fichier "naissances" (qui devient progressivement le fichier définitif) des données provenant des registres de mariages et de décès : le fichier "mariages" et le fichier "décès" sont confrontés tour à tour à ses enregistrements. Là encore l'ordinateur propose l'adjonction de faits concernant la nuptialité, l'alliance ou la mortalité pour les individus communs aux deux ou aux trois fichiers et la liste des individus mariés ou décédés sur la commune mais ne figurant pas sur la même variable, "seule une intervention manuelle permet alors de trancher : le chercheur force alors la variable à la valeur de son choix." (formule de P. Lamaison et de M. Selz-Laurière).

Une telle recherche, qui vise à l'établissement d'un progiciel valable pour toute petite population, s'inscrit également dans une recherche personnelle finalisée où l'on cherche à vérifier un certain nombre d'hypothèses anthropologiques sur les populations du social au biologique et la dynamique bioculturelle. On est en effet dans une région du monde où, à partir d'une confrontation originelle de divers apports humains et culturels, la société gère de manière permanente un phénomène biologique, dans la mesure où la diversité phénotypique est utilisée comme matériau de la différenciation sociale... Et, tout au long des généalogies, il est possible de voir apparaître, comme in vitro, des processus de mélanges ou de barrières...