

## PROGRAMMES APPLIQUES AUX RECENSEMENTS DU DEBUT DU XIXe SIECLE: LIEGE ET VERVIERS

Cl. DESAMA, Assistant à la Bibliothèque Centrale de l'Université de Liège

Dans une large mesure, l'analyse automatique du recensement de la population de Liège en l'an IX constituait une innovation par les méthodes qu'elle mettait en oeuvre. Le Colloque aura été l'occasion d'ordonner des démarches, fructueuses certes, mais le plus souvent empiriques. Aussi les quelques réflexions qui vont suivre sont-elles inspirées par la nécessité de passer le plus rapidement possible du stade de l'expérience à celui de la théorie ou, à tout le moins, de la méthodologie.

Nous ne croyons pas qu'il existe, en matière de recherches historiques sur ordinateur, un type de méthode fondé sur tel support d'information, impliquant une technique d'IN-PUT bien précise et débouchant sur un mode de traitement prédéterminé. C'est dire que nos considérations n'ont rien de normatif même si leur généralisation à d'autres travaux historiques menés au moyen de l'ordinateur ne poserait aucun problème technique. Le Professeur Et. Hélin ayant examiné par ailleurs les problèmes posés par la préparation des données et les opérations de contrôle des fichiers, nous nous bornerons donc au traitement proprement dit.

Sur base des enseignements tirés de l'analyse automatique du recensement, nous pouvons distinguer quatre phases.

### *L'exploitation du fichier primaire*

Cette exploitation consiste en opérations de comptage de deux types :

- a) compter la fréquence d'apparition d'un caractère ou d'un groupe de caractères (par exemple le nombre de femmes mariées ou le nombre de chefs de ménage propriétaires de la maison dans laquelle ils habitent) ;
- b) compter la fréquence des combinaisons de caractères isolés ou groupés. Il faut distinguer deux types de combinaisons, les associations de deux caractères (les dyades) ou groupes de caractères d'une part, les configurations plus complexes d'autre part.

Le premier type entraîne la confection de tableaux statistiques à double entrée. A titre d'exemple, citons l'association des variables démographiques (âge et sexe) ou socio-professionnelles (métier et statut à l'intérieur d'une entreprise).

Plus pénétrantes mais aussi plus délicates à interpréter sont les combinaisons complexes. Ainsi, par exemple, l'étude des caractères socio-professionnels par groupes d'âges de cinq ans en fonction de l'état matrimonial ou encore, la taille des ménages suivant leur composition et l'âge de la mère. Autant d'éléments qui répondent aux préoccupations les plus actuelles des économistes ou des sociologues et qui nous mettent à même de décrire une situation révolue à l'aide des normes en usage chez nos contemporains.

D'un point de vue technique, les opérations se déroulent de la façon suivante. Le fichier-cartes, enregistré préalablement sur disques, est lu par l'ordinateur qui en extrait les données à compter ou à combiner pour les soumettre ensuite au traitement adéquat défini par le programme.

D'une manière générale le fichier est fractionné en plusieurs éléments constituant des ensembles (les quartiers) qui sont ensuite traités séparément, les totalisations partielles ou générales étant effectuées ultérieurement. De ce fait, il nous était possible de repérer plus aisément les erreurs ou les négligences éventuelles et d'apercevoir à temps les compléments d'information souhaitables.

Quant aux résultats, ils étaient produits par l'ordinateur, le plus souvent sous forme de tableaux statistiques à double entrée.

La plupart des programmes conçus pour le recensement de l'an IX l'ont été par des historiens, ce qui constitue la meilleure des garanties quant à la conformité des résultats obtenus par rapport au but poursuivi. Le langage de programmation utilisé était le PL/I appliqué tantôt à un ordinateur IBM 360/20 tantôt à un IBM 360/25 suivant les circonstances.

Etant donné la nature du travail demandé à l'ordinateur, il importe de pouvoir compter sur des mémoires périphériques vastes et d'accès rapide. En revanche, les temps de travail et les possibilités de calculs interviennent comme des critères secondaires dans le choix du matériel.

#### *La constitution des sous-fichiers et leur exploitation*

Cette phase n'est pas négligeable lorsqu'on travaille, comme c'est souvent le cas en Histoire, sur des fichiers très importants. Il est en effet intéressant de constituer des sous-ensembles homogènes dont l'exploitation, semblable à celle que nous appliquons au fichier primaire, est menée séparément. Ainsi avons-nous fait perforer par l'ordinateur un fichier des "étrangers" qui nous a permis de traiter ce groupe d'individus indépendamment de toute division par quartiers et sans contraindre la machine à plusieurs relectures d'un fichier de plus de trente mille cartes.

Ces sous-fichiers ou fichiers dérivés sont préalablement constitués sous forme de cartes puis enregistrés sur disques ou tout autre support en vue du traitement.

### *Les calculs statistiques*

Sur la base des données fournies lors de la phase d'exploitation (cf. I), il convient d'effectuer une analyse statistique plus ou moins approfondie suivant l'objet du travail et l'importance quantitative des données. Ces calculs statistiques sont réalisés tantôt dans le cadre des programmes d'exploitation pour les indices les plus courants tantôt par des programmes particuliers traitant des informations spécifiques (par exemple : tests de  $\chi^2$ , calcul des coefficients d'aplatissement, d'asymétrie ou de corrélation).

Au delà de la signification statistique des résultats, cette troisième phase nous fait entrer de plain pied dans le domaine de l'analyse quantitative qui nous permet d'éviter les faiblesses du travail monographique et d'inférer une véritable typologie.

Intervalles de confiance, méthodes d'interpolation, corrections d'erreurs, hypothèses fondées sur des modèles de simulation rétrospective, autant de problèmes que l'historien n'abordait, dans le meilleur des cas, qu'avec une extrême méfiance et que l'ordinateur permet de résoudre dans les meilleures conditions.

### *Le couplage de données*

En travaillant en accès direct sur le fichier primaire ou les sous-fichiers préenregistrés (sur disques par exemple), on peut traiter des données nouvelles, provenant d'autres documents, en référence à celles contenues dans le préenregistrement. C'est la technique du couplage qui nous permet d'intégrer différents travaux d'exploitation. Dans le cas du recensement de l'an IX, la première expérience consistera dans la comparaison entre les données des listes nominatives et celles contenues dans les registres d'état civil de l'époque. A cette occasion, nous pourrions nous rendre compte, expérimentalement, des difficultés inhérentes à cette technique du couplage qui suppose des procédures d'identification rigoureuses et automatiques.

L'importance du couplage de données est certaine. Pour nous en tenir à la démographie, nous pouvons considérer en effet une population recensée comme un stock soumis aux flux et aux reflux des mouvements naturel et migratoire. D'où la possibilité de constituer automatiquement des populations théoriques soumises ou non à l'hypothèse de stabilité. En considérant cette fois non plus une population recensée mais une génération ou une cohorte, nous pouvons confier à l'ordinateur des analyses longitudinales portant, par exemple, sur les probabilités de survie ou sur la fécondité.

Avant de proposer des conclusions, nous voudrions insister à nouveau sur le fait que les quatre étapes brièvement commentées ci-dessus ne constituent certainement pas l'alpha et l'oméga de tout traitement de l'information historique sur ordinateur. L'expérience d'autres travaux nous permet de constater le caractère spécifique de chaque étude mais aussi les nombreux points de convergence, notamment dans

la conception générale du traitement. En fait, l'écueil essentiel se situe au niveau de la préparation des données et c'est là précisément que se situe l'avantage décisif de la double formation d'historien et de programmeur. Comment décider par exemple, de la nature et de la composition des fichiers ou des articles ? Comment savoir s'il est utile ou superflu de coder au préalable certaines données si l'on n'a pas une notion bien nette de l'organisation d'un traitement et des techniques de programmation ?

Quant aux avantages de l'utilisation de l'ordinateur dans l'analyse du recensement, ils sont évidents. Ils portent non seulement sur la fiabilité des résultats et les ressources du calcul mais encore et surtout sur la possibilité de traiter globalement l'ensemble des informations.

### *CONCLUSION*

En cette matière, il n'est pas de conclusion définitive car toute tentative de structuration méthodologique d'une expérience est sans cesse remise en question. Aussi faut-il considérer cet exposé comme un essai provisoire qui ne tardera pas à être corrigé et développé ne serait-ce qu'en fonction du traitement d'autres sources (archives notariales notamment) ou de l'application des méthodes d'analyse statistique mises au point par le groupe d'études d'Histoire Quantitative et de Développement de l'Université de Liège.

S'il m'est permis toutefois de justifier le titre de cette communication, je dirai que l'introduction de l'ordinateur en Histoire entraîne une double mutation, non par hasard, mais par nécessité : celle de la recherche historique et celle de l'historien.

Dans le sanctuaire de la recherche historique, l'ordinateur bouleverse les traditions et les rites. Il oblige l'historien à insérer ses exigences critiques dans un cadre nouveau. La démarche classique fondée sur l'enchaînement de la critique externe et de la critique interne doit se compléter par la maîtrise des techniques de traitement de l'information et d'analyse statistique. En outre, l'ordinateur nous incite à renoncer aux études monographiques ou ponctuelles qui ont fait les délices de plusieurs générations d'historiens au profit de traitements en équipe qui visent à dégager une typologie des ensembles sociaux et des familles d'ensembles.

En ce qui concerne l'historien, l'introduction de l'ordinateur dans la recherche historique lui impose un changement profond dans ses modes de pensée, dans sa conception du travail historique et surtout dans sa formation. Contraint déjà par l'évolution récente de sa discipline à assimiler les méthodes des sciences économiques et sociales et les techniques probabilistes, il se trouve, à terme, dans l'obligation de s'initier au traitement de l'information, en assurant lui-même, dans la mesure du possible, les tâches d'analyse et de programmation.

Que l'ordinateur puisse accélérer ainsi la modernisation de notre enseignement et de nos méthodes de travail, ne serait certainement pas le moindre service qu'il rendrait à l'Histoire.

E. HELIN, CI. DESAMA, *La démographie historique au L.A.S.L.A.* dans *Revue* (Organisation Internationale pour l'étude des Langues Anciennes par ordinateur), n° 4, Liège, 1970.

W.A. ARMSTRONG, *The Interpretation of the Census Enumerator's Books for victorian Towns* dans *The Study of Urban History*, London, 1968.

H.J. DYOS, A.B.M. BAKER, *Computerising Census Data* dans *The Study of Urban History*, London, 1968.

M. COUTURIER, *Recherches sur les structures sociales de Châteaudun 1525-1789* (chapitre intitulé "Esquisse de méthodologie mécanographique"), Paris, 1969.

H.B. NEWCOMBE, *Couplage de données pour les études démographiques* dans *Population*, t. XXIV, p. 654-656, Paris, 1969.