

Direction de la Diffusion et de l'Action Régionale

H9305

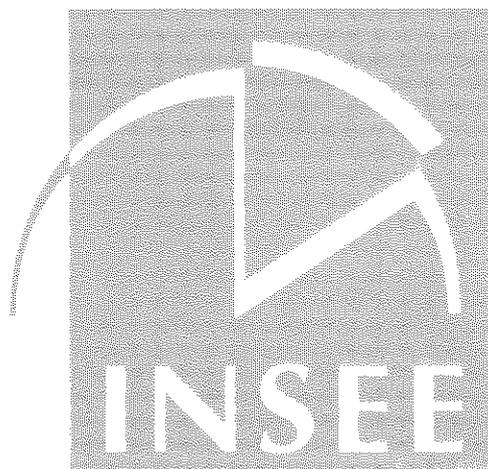
SAPHIR

UN SYSTEME D'ANALYSE
DE LA POPULATION PAR
L'HISTORIQUE DES RECENSEMENTS

Bernard AUBRY

Août 1993

Document de travail



Saphir : **un Système d'Analyse de la Population** **par l'Historique des Recensements**

Plan

INTRODUCTION

PREMIERE PARTIE :

Pourquoi un fichier historique des recensements ? Page 9

DEUXIEME PARTIE :

L'analyse d'un territoire Page 31

ANNEXE :

Les bilans des ressources humaines, par ZEAT Page 127

Une troisième partie est en préparation qui a pour thème l'analyse de catégories ainsi que les migrations. Par ailleurs une note technique sera rédigée qui contiendra notamment le contenu détaillé des codes du fichier SAP.1B qui sera proche du fichier SAP.1A, mais plus complet (cf. RAR n° 199).

Principales références :

RAR n° 14, 84, 136, 181, 199

Rapport d'étape : juin 1991

Communication au colloque de Luxembourg : le capital humain européen à l'aube du XXIème siècle (décembre 1991) et au colloque de démographie de Bordeaux (février 1993)

Articles parus dans les numéros de Chiffres pour l'Alsace n° 22-24-25

Strasbourg 1962-90 - Bilan des ressources humaines et de l'emploi (ADEUS, CUS, INSEE) - Document provisoire, 157 pages.

Principaux fichiers :

Alsace par commune :

Fichier A RP62 à 90 Exhaustif

Fichier B RP62 à 90 au 1/4

(Ces fichiers sont archivés, mission Constance, réf.: Compte-rendu de la 59ème réunion du Comité d'Archivage Historique du 21.12.1992.

France par département :

Fichier SAP.1A (référence RAR n° 199)

Fichier SAP.1B (début 1994 ?)

INTRODUCTION

Saphir se propose d'être un système d'analyse de la population et de l'emploi s'appuyant sur les cinq derniers recensements de la population. A ce titre, il se présente comme :

- 1- un fichier de référence (fichier historique), aisément accessible
- 2- un mode de consultation associé et une démarche d'analyse

---> 1 - Un fichier de référence

Il s'agit de rassembler en un même fichier un ensemble de données individuelles portant sur des caractéristiques peu nombreuses mais essentielles. Le contenu de la base ainsi créée se compose de quelques variables de structure (sexe, année de naissance, etc) ainsi que de variables géographiques relatives aux lieux de naissance, de résidence actuelle et antérieure, et de travail.

Chaque individu présent dans le fichier est décrit par un ensemble de modalités. Ces modalités se rapportent à des codes parmi lesquels figure l'identification des différents recensements (1962, 1968, 1975, 1982, 1990).

L'unité statistique est l'individu. Cela signifie qu'une personne qui aurait été recensée cinq fois, lors des cinq recensements, figurerait sous forme de cinq enregistrements contenant ses caractéristiques du moment. Pour ces personnes, certaines modalités sont fixes sur toute la période (l'année et le lieu de naissance, le sexe), d'autres pourront avoir changé en fonction du cycle de vie (la catégorie socio-professionnelle, la naturalisation). Quant à l'âge, il aura nécessairement varié.

*une rencontre
de chiffres
dans l'espace
et dans le temps*

Insistons sur le fait que le fichier ne ressemble en aucun cas à un panel qui aurait également l'individu comme unité statistique. En effet, il n'y aura dans un panel qu'un seul enregistrement par individu; la taille de chaque enregistrement s'étend au fur et à mesure que se succèdent les recensements et les événements démographiques de la personne observée.

Si le fichier historique ne permet pas de suivre le parcours des individus, il permet de comprendre le comportement des groupes, par exemple ceux formés des personnes nées la même année (donc formant une cohorte) ou appartenant à des zonages (ensemble de communes) ou à une catégorie de population (les Portugais par exemple).

*Portraits
de
groupes*

La base de données servant de référence a été alimentée par des extractions effectuées sur les fichiers correspondant à des exploitations par sondage. Ainsi le fichier historique national actuellement disponible provient des sources suivantes :

- RP62, sondage au 1/20ème
- RP68, sondage au 1/4
- RP75, sondage au 1/5
- RP82, sondage au 1/4
- RP90, sondage au 1/4

En réalité on peut concevoir, et il est même souhaitable, qu'il y ait plusieurs fichiers de référence. C'est l'expérience qui doit déterminer le bon équilibre en fonction de l'environnement (taille des fichiers, capacités d'extraction).

*Un fonds
commun
où l'on
puise...*

Du reste dans l'expérience faite sur l'Alsace, on dispose de deux fichiers de base qui comprennent des informations communes (mais sous des codes différents) pour les recensements de 1962 à 1982, mais dont la composition diffère en 1990. Le premier fichier construit en fonction des caractéristiques de l'exploitation exhaustive du recensement de 1990 est plutôt un outil pour l'analyse des migrations alternantes. Le second fichier est plutôt destiné à l'étude des structures puisque celle-ci se réfère au dessin des enregistrements de l'exploitation par sondage au 1/4 du recensement de 1990.

---> 2 - Un mode de consultation associé et une démarche d'analyse

Le fichier de référence est nécessairement de grande taille. Pour l'Alsace il approche les deux millions d'enregistrements. Quant au premier fichier national, bien qu'il ait été cumulé (ce qui est un gain appréciable puisque le niveau géographique le plus fin est le département), il contient plus de dix millions d'enregistrements pour environ trente variables. Le cumul permet de regrouper l'ensemble des individus ayant les mêmes modalités de variables.

Une méthode rationnelle de traitement des données est donc conseillée. Celle-ci passe par une utilisation systématique des procédures d'extraction pour transcrire les seules données utiles à l'étude d'un domaine (les villes, les étrangers...) sur un nouveau support (fichier restreint) ou dans une base de tableaux.

*...la
seule
information
dont on
a besoin*

Pour éviter que les mêmes opérations ne soient trop souvent renouvelées et pour aider les néophytes dans leur investigation, on propose des tableaux standards qui peuvent

d'ailleurs eux-mêmes être rassemblés dans des fichiers informatiques. Ces tableaux doivent être construits en fonction de l'expérience, pour satisfaire des demandes types, par exemple pour une approche statistique d'une zone.

Un mode d'emploi pour éviter les pièges

Mais au-delà de la mise à disposition des fichiers et de la fourniture de tableaux, il est indispensable que le système propose une méthode d'analyse, ce qui implique un mode d'emploi. Outre la description détaillée des codes, le document doit contenir les éléments propres à aider à l'interprétation des résultats. Les principales anomalies de collecte, les incertitudes liées à la méthode (échantillons), d'une façon générale les pièges dans lesquels tombent inévitablement ceux qui utilisent les chiffres (en général et en l'occurrence ceux du recensement), doivent faire l'objet d'une présentation claire. Cinq chiffres disposés en série temporelle autorisent un examen dynamique d'une variable sur la période 1962-90. Si l'on sait être prudent, il peut être intéressant de prolonger cette tendance, du moins mentalement pour aider à explorer le présent (1993) ou l'avenir proche. Mais cinq chiffres en série permettent le cas échéant de repérer des valeurs aberrantes que l'on traitera comme il convient, l'opération sera plus facile si l'on associe des acteurs locaux à la réflexion.

*Une aide
à
l'interprétation*

Décrire automatiquement et avec clarté une zone d'emploi est un objectif à atteindre. Non seulement le gain de temps est appréciable puisque l'on fait l'économie de la recherche, mais encore le débat est plus facile car les querelles de chiffres ont des grandes chances d'être évitées.

*Eviter
d'inutiles
querelles*

Bien entendu l'utilisation du fichier historique n'est qu'une étape avant de prendre en compte d'autres sources d'information nécessaires à l'approfondissement de la connaissance de la zone.

Une nécessaire pédagogie

C'est donc l'un des soucis de Saphir que de contribuer à la présentation des résultats des recensements de population. Cette démarche pédagogique est essentielle. On ne sera pas prisonnier de concepts très compliqués. En revanche on s'efforcera de fournir des données compréhensibles à tous ceux qui, sur le terrain, n'ont pas vraiment le goût des développements abstraits ou, l'auraient-ils, ne trouveraient pas le temps de le satisfaire. Car toute investigation simple n'en exclut pas d'autres plus compliquées. Mais c'est sur la

*Des cadres
d'analyses
accessibles
à tous..*

présentation de cadres d'analyse accessibles à tous, moyennant un effort limité, que l'on veut mettre l'accent. Et c'est sur cet aspect pédagogique que l'on devra d'abord juger Saphir, plus que sur les produits standards (extraction, tableaux-standards) qui ne sont pas encore parfaitement élaborés ne serait-ce que parce que les moyens de traitement informatique étaient limités.

...à
perfectionner
et à
développer

Seul un petit nombre de variables ont été sélectionnées dans le fichier de référence. D'autres pourraient leur être adjointes, mais l'objectif n'est pas d'étendre la capacité du fichier au point de risquer l'asphyxie. Il est, au contraire, de pouvoir accéder sans difficultés à quelques variables structurelles et géographiques de base, celles-ci ayant été harmonisées. Une parfaite cohérence des variables n'est pas possible, mais il est sans doute réaliste de vouloir s'en approcher.

Des tableaux standards

De ces données de base, on peut extraire un grand nombre de tableaux de référence, très utiles pour une première analyse d'une zone, pour réaliser un cadrage. Il suffit de songer au temps passé à la confection de ces tableaux de base, indispensable avant toute amorce d'analyse d'une zone pour un dossier de prime, pour un schéma directeur. L'idée est d'alléger le travail préparatoire de toute étude de zone par la mise à disposition en un ou plusieurs fichiers des variables principales que sont le sexe, l'année de naissance et quelques autres.

Des
données
de
cadrage...

En revanche l'accent est mis sur la nécessité de disposer d'une nomenclature géographique fiable et cohérente pour l'ensemble du territoire, de façon à constituer aisément un zonage adapté au territoire étudié.

Jusqu'à épuisement de l'information

Le fait de disposer d'enregistrements au niveau de l'individu permet de pouvoir accéder sans aucun problème à l'ensemble de l'information collectée ce que ne permet pas un fichier structuré sur la commune par exemple car, quelles que soient les décompositions effectuées, on ne saurait en ce cas prendre en compte la totalité des croisements potentiels de modalités.

...avant
de
pousser
l'investigation...

Ouvrir des pistes

Si dans de nombreux cas l'exploitation du fichier et la démarche de Saphir doivent suffire à satisfaire l'attente, on voudra aussi aller plus loin, c'est-à-dire examiner de plus près d'autres variables (la situation du logement, de la famille) ou prendre en compte des données plus récentes (sur l'emploi). Saphir aura alors fait gagner du temps en aidant à défricher le terrain et en ouvrant des pistes.

*...par l'emploi
d'autres
sources*

Marier les petits systèmes aux plus gros

Enfin le but sera atteint si l'on parvient à unir les qualités de souplesse et de convivialité d'un microordinateur avec les capacités considérables de stockage et de traitement des gros systèmes. Et l'on rêve d'une procédure qui permettrait à tout un chacun de sélectionner sur un écran un territoire composé d'un ensemble d'unités administratives. Pendant que l'unité centrale rassemblerait l'ensemble des enregistrements relatifs à la zone, l'utilisateur composerait à l'aide d'un menu déroulant des ensembles géographiques pertinents au regard de la zone d'étude. Ainsi le fichier serait recomposé et recodifié par rapport à la zone d'étude.

Une sélection de tableaux standards (tel le bilan population active-emploi) seraient alors disponibles et immédiatement affichés sur l'écran.

Il resterait alors à l'utilisateur à examiner sur ces tableaux l'information la plus pertinente. Les séries correspondantes seraient alors sélectionnées (par surimpression).

Grâce à un menu déroulant, on décomposerait les séries selon un choix de modalités ce qui permettrait de privilégier l'information spécifique à l'information banale. Bien évidemment, on associerait à ce système des possibilités de représentation graphique et cartographique.

*Trouver le
bon chemin
pour accéder
à
l'information
pertinente*

* *

**

PREMIERE PARTIE

POURQUOI UN FICHIER HISTORIQUE DES RECENSEMENTS ?

Plan

1 - LES MOTIVATIONS

2 - LES PREOCCUPATIONS

3 - A QUI S'ADRESSE LE FICHIER HISTORIQUE DES RECENSEMENTS?

4 - LES LIMITES

"Car l'ordre ne peut être révélé que par les grands nombres : les inexactitudes des petits chiffres disparaissent d'autant plus que les nombres que l'on compare sont plus grands"

J.P. Süßmilch (1707-1767), L'Ordre divin

1 - LES MOTIVATIONS

Le besoin de disposer de résultats enchaînés s'est développé dans une région, l'Alsace, reconnue à bien des égards comme atypique. Mais au-delà de quelques caractéristiques alsaciennes propres à exciter la curiosité et qui trouvent une réponse par de telles investigations (l'Alsace perd-elle ses cadres, où vont les femmes, comment fonctionne le marché du travail des bassins d'emploi frontaliers?...), deux motivations plus générales justifient au moins la création de tels fichiers.

Cinq recensements de la population sont aujourd'hui retranscrits sur des fichiers informatiques. La grande taille de ces fichiers n'incite pas à les fusionner pour en créer un autre qui serait plus volumineux encore. Au handicap de la taille se cumule l'obstacle de la diversité dans la constitution des variables puisque les codes utilisés lors des différentes exploitations ne peuvent être harmonisés de façon satisfaisante.

Certes le chargé d'études qui connaît bien l'informatique est en mesure d'extraire de chaque fichier les informations dont il a besoin pour suivre les seules variables sur lesquelles il travaille et les suivre dans le temps. Mais cette démarche appliquée aux cinq recensements n'est pas fréquente pour diverses raisons : il faut interroger des fichiers très lourds et de plus ceux de 1962 et de 1968 présentent l'inconvénient d'être conformés dans une structure qui n'a plus cours et que seuls les informaticiens connaissent. De plus le travail ainsi réalisé au coup par coup est rarement accessible à d'autres qu'à leurs auteurs.

*Des
économies
de moyens*

Une sélection de variables

Et si cette démarche se généralisait, elle deviendrait très vite coûteuse et peu rationnelle. D'où l'idée de créer non pas un fichier unique et complet, mais un fichier qui comprendrait une sélection de variables, peu nombreuses, mais d'intérêt majeur. Et l'on se rend compte alors qu'une telle entreprise peut rendre de multiples services. Outre l'aisance dans l'accès aux données, le fichier ouvre des perspectives d'exploitation insoupçonnées.

Tel est l'objectif de Saphir. L'ambition du projet est limitée au rapprochement de quelques données de structure (année de naissance, sexe...) et de localisation géographique. La

création d'un tel fichier pour la région Alsace puis, d'une façon simplifiée, à l'ensemble de la France, n'a pas exigé un investissement considérable. Le gain marginal d'une telle exploitation est sans nul doute en sa faveur. Quant à la démarche associée au fichier, elle doit se construire peu à peu, en fonction de l'expérience acquise à la pratique du terrain. Le document encore provisoire qui est proposé décrit le travail réalisé à la DR d'Alsace. Cet essai n'aura pas été inutile si cette expérience peut servir à la définition de nouveaux projets puisque la fourniture d'informations élémentaires correspond à une demande extrêmement forte dans le domaine de la démographie comme dans d'autres, de la part des clients, institutionnels ou non de l'INSEE.

*Une
démarche
expérimentale...*

Il y a quelques années, des fichiers dénommés Migraine et Migralt avaient été mis à la disposition des utilisateurs pour analyser en particulier les résultats des recensements de la population. Ces fichiers et les logiciels qui les accompagnaient avaient rendu de grands services aux statisticiens régionaux. Toutefois ils n'ont pas été mis à jour de sorte que les tableaux qu'ils proposaient n'ont pu être disponibles pour le dernier recensement. De toutes façons, aussi précieuse qu'aurait pu être la fourniture de données sur des séries historiques, le logiciel aurait nécessité un toilettage ou plutôt une refonte adaptée aux changements fondamentaux intervenus dans le monde informatique par l'apparition des micro-ordinateurs.

*...adaptée
à la
micro-informatique*

Dans la cour des grands

L'irruption de la micro donne à tout un chacun la possibilité non seulement de se constituer ou d'avoir accès à des bases de données de plus en plus riches, mais encore de traiter ces données avec tous les logiciels qui naguère n'étaient accessibles qu'à un petit nombre d'utilisateurs, spécialistes de leur domaine, dont ils pouvaient d'autant jalousement garder l'accès que les obstacles pour s'en approcher étaient nombreux.

Aujourd'hui chacun peut suivre les chemins tracés par les spécialistes et peut librement faire les détours qu'il souhaite. S'il le veut il peut même se risquer à sortir des sentiers battus. En suivant une autre route que les spécialistes, il offrira probablement le flanc à la critique, mais peut-être aussi apportera-t-il une contribution modeste à la connaissance en proposant un regard nouveau sur des phénomènes à plusieurs facettes. Et si le risque n'est pas nul de certaines errances, c'est tout simplement le prix à payer de la liberté d'accès et de traitement des données. Cette remarque ne vaut pas seulement pour les recensements de la population.

*Des chiffres
pour
le plus
grand nombre*

La prise en compte de la dimension historique

La tentation, légitime, de vouloir coller au présent en fixant l'effort sur la collecte de données conjoncturelles nous amène parfois à négliger le passé. Or celui-ci ne doit pas être ignoré puisqu'il contribue largement à la compréhension du présent et à l'exploration du futur.

Les cinq derniers recensements, ceux de 1962 à 1990 permettent de couvrir une période qui, compte tenu de l'information disponible en 1962 sur le lieu de résidence antérieure, commence en 1954 et correspond à 36 ans, soit à peu près la durée d'une vie active.

*Presque la
durée
d'une vie
active*

Un exemple :

Soit l'ensemble des personnes nées en 1940. Elles ont 22 ans en 1962, premier recensement qui permet la description de la cohorte ainsi désignée. Au plan national, cela correspond à 600000 personnes, un effectif suffisant pour autoriser de nombreux croisements de variables. Donc cette population est décrite dans sa distribution spatiale et dans ses flux migratoires, depuis 1954. A l'époque, les personnes de la cohorte avaient 14 ans.

Nous retrouvons cette cohorte six ans après, en 1968. Les personnes qui la composent ont alors 28 ans. Ce sont presque toutes les mêmes qu'en 1962. Seuls les immigrants (qui ne représentent qu'un faible pourcentage et que l'on peut du reste traiter à part) se sont joints au groupe, cependant que disparaissaient les personnes décédées ou parties à l'étranger. Mais le noyau dur de la cohorte reste le même, de sorte que l'on peut comparer ses caractéristiques à celles qu'avait le groupe au précédent recensement. En particulier le taux d'activité a changé. Chez les femmes il est tombé de 63,6% à 47,9%.

*Suivre
des
cohortes*

Au recensement suivant, en 1975, les individus de la cohorte ont 35 ans. Le taux d'activité des femmes est remonté à 51,7%. Puis le groupe réapparaît au recensement de 1982 les personnes ont alors 42 ans. Enfin au dernier recensement, en 1990, elles atteignent le cap des 50 ans. Le taux d'activité des femmes s'est encore accru, atteignant 66,5% (62,0% en 1982). Osons penser qu'un nouveau recensement aura lieu en 1997. Alors ces personnes auront atteint ou approcheront le seuil de la retraite. Toute la vie active de la cohorte aura été alors décrite par les chiffres contenus dans le fichier historique.

*La mise à
jour par
le prochain
recensement*

La mise en évidence de la transformation des caractéristiques des cohortes est sans aucun doute l'un des atouts majeurs d'un fichier historique des

recensements. L'information de base étant l'individu, il est possible de procéder à n'importe quel croisement de variables structurelles ou géographiques (pour autant que l'information soit fiable quand il s'agit de données issues de sondages).

Suivre des groupes et non des individus

Précisons que le fichier historique tel qu'il a été constitué ne ressemble en rien à un panel de population à l'instar de celui qui a été créé par l'INSEE (Département de la Démographie) pour suivre, depuis 1968, la trajectoire des personnes nées les quatre premiers jours du mois d'octobre. Les principaux événements démographiques (naissance, mariage...) de chaque individu appartenant au panel sont inscrits au fur et à mesure qu'ils surviennent, de même que les caractéristiques figurant dans le recensement. Il y a un seul enregistrement par individu, mais cet enregistrement est enrichi en permanence.

*Saphir
n'est pas
un panel*

S'agissant du fichier historique, les données individuelles ne peuvent pas être raccordées entre elles. Ainsi les caractéristiques d'une personne qui aurait été présente aux cinq recensements apparaîtraient dans cinq enregistrements individuels disjoints, repérés chacun par leur code de recensement. Avec Saphir, et son intérêt réside bien là, il s'agit de suivre l'évolution de comportements de groupes (cohortes, catégories...) et non de suivre l'évolution de comportements individuels.

Le panel qui porte sur environ 600 000 enregistrements permet de comprendre des comportements collectifs pour l'ensemble du territoire ou pour de grandes zones géographiques. En revanche la faible taille du fichier ne permet guère de découvrir des caractères spécifiques pour des petites régions et a fortiori pour des zones plus petites. De son côté Saphir apporte certes quelques regards nouveaux au plan national, mais il ouvre plutôt un débouché intéressant pour la compréhension des petites zones géographiques puisqu'il en décrit l'histoire démographique contemporaine.

*Son utilité
pour de
petites zones*

Des cadres d'analyse

Une présentation qui s'impose consiste en l'élaboration de quelques tableaux simples, des fiches d'identité de zone, qui permettent d'un coup d'oeil de comprendre le développement de la zone d'étude. On propose ainsi

notamment deux tableaux de base sur lesquels d'autres peuvent se greffer. L'un qui s'intitule "de la population active à l'emploi" compare l'évolution de ces deux composantes, souvent divergentes dans des zones ouvertes sur l'extérieur.

Des fiches d'identité...

L'autre est un bilan des ressources humaines et compare, en variations annuelles moyennes, les principales variables intéressant le marché du travail sous la forme d'un cadre comptable où s'équilibrent ressources humaines et emplois. Les tableaux d'une zone ne prennent en fait véritablement leur sens que si les chiffres qu'ils contiennent sont comparés à d'autres, établis par exemple sur une entité géographique plus vaste qui peut être la région, l'ensemble du pays, ou une catégorie de communes pour lesquelles le rapprochement avec la zone d'étude a un sens. On voit alors se dessiner d'emblée les comportements spécifiques de la zone et leur évolution dans le temps. Les éléments ainsi présentés constituent un tableau de synthèse utile à toute première approche d'un bassin d'emploi, d'un SIVOM, d'une unité urbaine puisqu'au-delà des caractéristiques communes à l'ensemble d'une zone, ils en traduisent les singularités. Ce devrait être notamment un outil indispensable à toute personne nouvellement installée dans une circonscription qu'elle ne connaît pas mais sur laquelle pourtant elle est appelée à jouer un rôle.

...Pour une commune, un SIVOM, un bassin d'emploi

Le vieillissement de la population est une donnée générale bien connue, qui affecte l'ensemble du territoire. Un recensement dira l'état d'une zone à un moment donné. Saphir en décrira immédiatement l'évolution. En particulier on pourra se rendre compte comment certaines zones "jeunes" risquent de "vieillir" plus vite que les autres. En ce sens, le fichier historique doit donc être un instrument complémentaire aux projections de population dans la mesure où il fait ressortir aisément les tendances de fond qu'une projection réalisée sur la base de deux ou trois points peut ne pas faire apparaître.

Dégager des tendances

La prise en compte des mouvements migratoires

Les disparités géographiques relatives à la fécondité et à la mortalité par âge se sont plutôt estompées au cours des dernières décennies. En revanche les mouvements migratoires ont pris une certaine ampleur et sont grandement à l'origine des spécificités observées dans les pyramides des âges. Ils conditionnent donc indirectement le devenir démographique des espaces régionaux.

Au sein des agglomérations et alentour se sont développés de forts courants migratoires dont la résultante est presque nulle pour un observateur extérieur. Dans l'aire d'attraction plus ou moins grande d'une ville, la migration définitive est souvent en relation avec une migration alternante. Prenons l'exemple d'un jeune issu d'une commune rurale. S'il trouve un emploi dans la ville voisine, il commencera sans doute par effectuer une migration alternante de son domicile à son lieu de travail.

*De la
migration
définitive à
la migration
alternante
et vice-versa*

Une fois son emploi assuré, il voudra s'établir en ville (migration résidentielle) jusqu'au jour où, ayant fondé un foyer, il achètera un pavillon dans une commune suburbaine. Cette seconde migration résidentielle sera probablement suivie par une nouvelle migration alternante. Les cycles de vie ont fait l'objet d'enquêtes spécifiques (deux numéros de Population ont consacré naguère des articles à ce thème), mais les recensements n'ont pas encore sans doute été assez exploités pour donner, à tous les niveaux pertinents, l'information qu'ils recèlent en ce domaine.

En effet, pour chaque individu, à chaque recensement, les fichiers contiennent quatre localisations : le lieu de naissance, les lieux de résidence actuelle et antérieure, le lieu de travail. Certes il n'est sans doute pas utile de chercher à combiner systématiquement, dans une même investigation, ces quatre caractéristiques qui, si elles sont effectivement liées entre elles, ne le sont pas directement. Une exception, peut-être : le marché du travail des zones frontalières. Le développement du travail frontalier (navettes) a créé un appel de main d'oeuvre sur des régions lointaines (migrations naissance-résidence et intercensitaire). Mais la juxtaposition judicieuse de deux ou trois de ces localisations suffit généralement.

*Quatre
localisations*

* * *

**

2 - LES PREOCCUPATIONS

Les essais de création d'un premier fichier historique ont été réalisés dans un triple souci de simplicité, de souplesse et de transparence.

La simplicité

Il est apparu que si l'on voulait donner au projet une grande chance de réussite, il fallait limiter les ambitions immédiates. C'est pourquoi, dans le tout premier essai, le travail a consisté simplement à constituer un jeu d'essai sur quatre départements de l'est de la France formant ainsi un modèle réduit du système à développer. Cette expérience réalisée au début de l'année 1990 a permis de vérifier le contenu des variables intégrées dans le fichier, à en insérer de nouvelles, d'élaborer quelques tableaux pour juger de l'intérêt de la démarche. Elle a permis également de prendre la mesure du travail à réaliser par la suite pour aller jusqu'au bout. La mise en évidence des singularités dans les échanges migratoires entre l'Alsace et la Lorraine est alors apparue d'emblée comme très intéressante et encouragea à poursuivre l'expérience.

*Une
progression
par étapes*

Une fois cette première phase achevée, il suffisait de généraliser l'application de la démarche à l'ensemble du territoire. En fait il s'agissait de constituer un fichier informatique regroupant l'ensemble des personnes qui, d'une façon ou l'autre, ont eu un lien avec l'Alsace: personnes, qui y ont vécu (naissance ou résidence antérieure ou actuelle) et qui y travaillent. Pour les quatre premiers recensements, la taille du fichier ainsi créé n'était pas considérable (1360000 enregistrements comprenant chacun environ vingt variables). Il est vrai que les données de base provenaient des exploitations par sondage.

Puis furent annoncés les résultats du recensement de 1990, sous forme de deux fichiers de composition différente. D'abord celui de l'exploitation exhaustive légère, puis celui, plus riche, du sondage au quart. Deux fichiers de référence ont alors été constitués, dont le dessin d'enregistrement est différent :

*Pour l'Alsace,
deux fichiers
de référence*

- un **premier fichier** (noté **A**) composé en fonction de la disponibilité des informations de l'exploitation exhaustive. Ce fichier prend en compte l'ensemble des individus au recensement de 1990.

- un **second fichier** (noté B) prend en compte un individu sur quatre au recensement de 1990 (sondage au 1/4).

Différents essais de recodification ont été effectués pour disposer de fichiers de plus faible taille, donc plus maniables. Deux ont survécu :

- un fichier A' , de petite taille, adapté à l'analyse des migrations alternantes (création d'un indicateur de migrations alternantes et contraction des modalités pour certains codes).
- un fichier B', pour l'étude des migrations résidentielles interdépartementales (suppression des codes communes).

Un troisième fichier (C) a été constitué qui permet de regrouper en un seul les données provenant à la fois de l'exploitation au quart et exhaustive du recensement de 1990. Il a été testé sur le Bassin Potassique mais il est reproductible à l'identique pour n'importe quelle zone. Le principe est de recréer un nouveau fichier au sein duquel les caractéristiques de tous les individus du fichier de référence sont recodifiés par rapport à leur position relative à la zone d'étude, en l'occurrence le bassin potassique. Ce procédé permet d'obtenir immédiatement non seulement des résultats pour la zone d'étude, mais encore de les comparer avec ceux d'un territoire avec lequel la confrontation a un sens (en l'occurrence, l'Alsace ou le Haut-Rhin).

Un fichier unique, intégrant l'exhaustif et le quart adapté à une zone d'étude

Tous ces fichiers ont en commun d'être de structure très simple, dans un format classique pour être interrogés facilement par les utilisateurs de l'informatique.

Mais la simplicité initiale qui paraît nécessaire à une bonne maîtrise des fichiers n'exclut pas une progressive complexification. Dans des phases ultérieures, on peut envisager la confection d'un fichier de référence qui comprendrait l'ensemble des informations élémentaires contenues dans tous les fichiers informatisés des recensements. On peut aussi imaginer des traitements spécifiques de l'information de base qui donneraient davantage de cohérence à des variables dont les modalités n'ont pas toujours le même sens sur toute la période.

Des adaptations possibles

La souplesse

Un écueil à éviter était celui de la lourdeur. Des fichiers trop volumineux sont difficilement

exploitables, malgré les progrès apparus dans le stockage et le traitement des données informatiques. Aussi, il fallait choisir entre différentes options tout en assurant une grande souplesse dans l'exploitation des données.

*Un petit
nombre de
variables...*

Les données de base sont rassemblées dans de gros fichiers archivés. Si les trois derniers recensements sont relativement accessibles et facilement exploitables par les agents de l'INSEE habitués à manipuler les langages informatiques des gros systèmes, les deux précédents, ceux de 1962 et de 1968 présentent l'inconvénient d'être structurés dans un format qui n'a plus guère cours, de sorte qu'ils ne sont pas très faciles à exploiter. Aussi, même si, un agent qui connaît un tant soit peu l'informatique peut espérer extraire des cinq fichiers l'information dont il a besoin, il reste que l'opération est très longue. Il n'est donc pas rationnel que chacun aille rechercher un peu toujours les mêmes données. En ce sens créer un fichier unique lisible par le plus grand nombre ne devrait pas soutenir la moindre objection tant est évident le gain de temps qu'il fait réaliser à l'ensemble des utilisateurs.

Le traitement des données à des fins d'étude se fait de préférence en aval, sur de plus petits fichiers. On peut distinguer deux situations différentes, avec bien entendu toute la gamme des cas intermédiaires. Soit il s'agit d'un chargé d'étude de l'INSEE travaillant directement sur les gros systèmes et alors celui-ci est en mesure de réaliser l'exploitation en fonction de ses besoins propres en recodifiant le fichier de façon ad hoc. Soit l'utilisateur du fichier est extérieur à l'INSEE et il doit pouvoir travailler sur un fichier de taille modeste, par exemple sur disquette en vue d'un traitement sur micro-ordinateur.

*...pour le
plus grand
nombre
d'utilisateurs*

C'est ce dernier cas de figure qui est privilégié du fait qu'il donne au plus grand nombre la possibilité de traiter des données qui, autrement, seraient inaccessibles soit parce que l'environnement technique ne se prête pas à la fourniture des données, soit parce que les fichiers n'ont pas été constitués en vue de leur diffusion.

Il y aura donc une très grande souplesse d'utilisation des données si on peut, par des programmes standards d'extraction et de recodification sélectionner les seules données utiles à l'étude et les fournir sous forme de fichiers disponibles à la vente.

La transparence

La tentation est toujours grande pour un chargé d'études de fabriquer son propre fichier, sans

prendre le temps de le documenter, de telle sorte que la personne qui l'a mis au point est la seule à savoir l'exploiter utilement, dans l'hypothèse favorable où elle a toujours dans la tête le mode d'emploi. Cette pratique courante a souvent été critiquée et c'est d'ailleurs pourquoi bien des travaux intéressants n'ont pas survécu à leur concepteur. Aussi pour donner à la construction du fichier historique une chance de pérennisation, mais aussi pour assurer sa crédibilité, les différentes phases du travail ont été décrites avec l'idée qu'il devait pouvoir être reproduit à l'identique en cas de défaillance. Cette documentation doit être mise à jour au fur et à mesure de l'avance des travaux.

*Une
documentation
"mise à jour"*

La transparence doit aussi permettre d'assurer de rendre le système évolutif. D'abord parce que si la présentation technique n'est pas trop ardue, il sera possible à des esprits curieux de la critiquer pour améliorer le système dans une phase ultérieure. Ensuite parce que l'on doit pouvoir greffer sur le fichier actuel des extensions.

En effet la création du fichier historique nécessite essentiellement l'appel des fichiers de base pour l'extraction et la recodification d'une sélection de données. Une fois au point cette procédure, il n'est pas difficile de reconstruire un nouveau fichier en tous points identique aux précédents, à ceci près que les imperfections repérées dans la première phase auront été éventuellement corrigées et surtout que de nouveaux codes auront été créés (tels des indicateurs de migrations). Evolutif, le système doit évidemment pouvoir prendre en compte au fur et à mesure les critiques et suggestions des utilisateurs.

*Un système
évolutif*

* * *

**

3 - A QUI S'ADRESSE LE FICHIER HISTORIQUE DES RECENSEMENTS ?

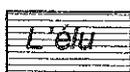
Instrument destiné à la présentation de données synthétiques sur l'évolution de la population et de l'emploi, le fichier historique concerne potentiellement un grand public. Comme révélateur de la dynamique spatiale, il doit être utile à tous ceux qui ont en charge l'animation et le développement d'une parcelle de territoire.

*Un instrument
de dialogue*

En effet, l'aménagement concerté du territoire nécessite que soit mise à la disposition des partenaires une information chiffrée claire et reconnue comme objective. Cette information qui gît dans les fichiers de l'INSEE et des autres producteurs d'information statistique permet de nourrir le débat, de structurer le dialogue, d'argumenter la prise de décision. Aujourd'hui encore, quoi qu'on en ait, ce sont encore les recensements de la population qui fournissent le matériau de base de la statistique locale. La mise à disposition en un même fichier des principales informations qu'ils ont fournies depuis trente ans est un moyen non seulement de les restituer, mais encore et surtout une possibilité de les valoriser, dans la mesure où la dimension temporelle ajoutée à la dimension spatiale multiplie la richesse contenue dans les fichiers.

*Un
enrichissement
mutuel*

En outre si l'information statistique, une fois mise en forme et à la portée de tous, peut être une aide à la connaissance du milieu local, inversement, la connaissance du milieu local peut elle aussi constituer une aide à la statistique.



L'Élu local, qu'il soit maire ou conseiller général, est la "cible" privilégiée du fichier historique dans la mesure où l'information proposée se veut être, à terme, extrêmement sobre, claire et bien présentée. Tableaux et graphiques doivent avoir immédiatement un sens. Il doit être possible de caractériser la zone par des indicateurs comparatifs appropriés : une commune avec l'ensemble du bassin d'emploi ou du département, et le département par rapport à la région, la ZEAT à l'ensemble de la France, toujours dans la perspective historique pour mettre en évidence les évolutions, qu'elles soient parallèles ou non. Avec l'aide de Saphir, il doit être possible d'exprimer en quelques chiffres les spécificités du morceau de territoire dont l'Élu a la charge.

*Présenter
les spécificités
de la zone*

Quand il s'agit d'une commune fondue dans une agglomération comme le sont celles des banlieues des villes, l'intérêt est peut-être plus grand encore. En effet l'entité formée par la commune est physiquement une abstraction puisque les limites sur le terrain ne sont pas visibles, contrairement à ce qui se passe dans les zones rurales où, le plus souvent, le contenu de la commune, du moins son agglomération, s'inscrit clairement sur le terrain. Les limites du ban communal traversent parfois un grand ensemble ou l'espace foncier d'une entreprise. Or les mouvements migratoires sont considérables à travers des frontières tout à fait artificielles et les personnes qui les franchissent n'ont pas le sentiment de pénétrer dans une autre commune puisqu'il y a continuité du bâti.

*Des limites
administratives
aux limites
statistiques*

C'est précisément là où la réalité est complexe que le besoin est grand de disposer d'informations sur des zones qui, pour n'avoir pas de réalité physique, en ont une, administrative. Le maire se doit de connaître la structure de la population dont il a la charge, qui l'a élu et devrait quelques années plus tard le réélire. Il doit par ailleurs se préoccuper de suivre l'évolution de l'emploi ne serait-ce que parce le produit de la taxe professionnelle versée par les entreprises implantées dans sa commune lui apporte une part importante des recettes.

D'autres considérations justifient également la fourniture d'indicateurs statistiques. Par exemple pour examiner après coup les conséquences des choix qui ont été fait dans le développement (évaluation des politiques). Si quelques données de recensement ne sauraient apporter seules l'évaluation nécessaire, ne serait-ce qu'à cause de leur caractère épisodique, elles sont néanmoins précieuses.

Malgré l'utilité des chiffres du recensement, on constate que, trop souvent, les maires qui ont pourtant joué un rôle essentiel dans la collecte des données, ne connaissent pas les possibilités offertes par des traitements appropriés pour leur commune.

*Le diagnostic,
l'évaluation*

L'aménageur

C'est l'élu qui fait le choix mais c'est le technicien (l'aménageur du territoire, l'agent de développement local) qui prépare les dossiers et fait des propositions. Les besoins d'information sont alors beaucoup plus étendus que pour l'élu dont ce n'est pas la fonction de s'investir dans des investigations statistiques, mais en principe les agents sont formés un tant soit peu à l'analyse statistique des données administratives et

disposent généralement d'outils adaptés au calcul et à la représentation graphique. Ce public est un client privilégié pour le fichier historique dans la mesure où la fourniture immédiate des données de base (de structure et d'évolution) lui permet un gain de temps considérable. Le but sera atteint si Saphir met de l'ordre dans les données de base, permettant ainsi de faire ressortir les caractères spécifiques qui méritent seuls une investigation ultérieure. Cette analyse se poursuit par l'interrogation des fichiers du recensement sur les variables qui ne figurent pas dans le fichier historique (logement, ménage) ou d'autres fichiers administratifs (des directions de l'Équipement ou du Travail par exemple). Mais s'il est naturel que l'élu et le technicien disposent des mêmes cadres d'analyse simple sur lesquels on s'accorde pour fixer les grandes lignes de l'évolution passée, c'est la fonction du technicien de chercher à aller plus loin dans l'investigation.

L'agent de développement local, un utilisateur privilégié

Le fichier historique permet notamment de suivre sur trente ans les flux migratoires. Le réseau très complexe des mouvements quotidiens et alternants qui se forment au voisinage des villes et évoluent dans le temps peut être assez simplement schématisé grâce au fichier historique, ce qui lui permet d'être un instrument utile à la définition de périmètres urbains pertinents nécessaires à toute élaboration de schémas directeurs. Le partenariat avec l'Agence d'urbanisme de Strasbourg (ADEUS) a sans doute beaucoup aidé à la connaissance du périmètre urbain et, en contrepartie, il a contribué à affiner la démarche d'analyse associée au fichier historique.

Définir des périmètres urbains

Le monde de l'éducation

Hormis la fourniture de données de référence nécessaires à une bonne connaissance de l'espace géographique facilitant les travaux de recherche aussi bien en géographie qu'en sociologie, le fichier historique pourrait peut-être trouver une application spécifique dans l'enseignement. En effet, les micro-ordinateurs ont investi les salles de classes et le traitement des données est accessible à tous.

En l'occurrence, on pourrait imaginer comment des enfants s'initieraient à la connaissance de leur environnement géographique immédiat par la possibilité qu'ils auraient de traiter les données relatives aux migrations de leur propre commune (canton) par rapport aux localités voisines. Ainsi pourraient être mis en évidence des phénomènes qui ont pris tant d'importance, tels que la polarisation vers les communes-pôles au détriment des campagnes (voire l'effet inverse dans

Apprendre à comprendre son environnement géographique

certains cas). A l'aide de logiciels graphiques et cartographiques, on pourrait représenter non seulement des flux, mais des structures pour différencier les différentes catégories de migrants. En l'occurrence, c'est l'aspect connaissance du milieu local qui est privilégié, mais bien entendu la même démarche pourrait être appliquée sur un territoire plus vaste, l'ensemble des départements et régions. Le processus de désertification de certaines régions et la dynamique démographique qui concerne d'autres zones pourraient être mises en cartes et présentées en fondu enchaîné. Les croisements des variables étant multiples (presque sans limites), il serait alors possible de représenter divers aspects des problèmes étudiés, ce qui ferait du fichier un instrument pédagogique.

Le grand public

Dans la mesure où les journalistes s'intéressent aux études réalisées à partir du fichier historique, le grand public peut être lui aussi concerné dès lors que les articles rédigés sur la base du fichier historique sont repris par les médias. On pense évidemment aux aspects liés à la localisation des activités qui intéressent le grand public. Mais en dehors des problèmes économiques et d'aménagement du territoire, on pense aussi à certains domaines auxquels la population est naturellement intéressée. Ainsi en Alsace, il y a en permanence des interrogations sur la capacité de la région à faire venir, et à retenir, les cadres supérieurs dont son économie a besoin. Il y a là une image de la région souvent perçue comme peu attractive. Dans le même ordre d'idée, on découvre en Alsace des singularités dans les taux d'activité féminins et dans leur évolution.

*Aujourd'hui
l'informatique,
à quand
l'information
pour tous ?*

Autre exemple :

On découvre qu'il existe dans la région un certain déficit de la population féminine jeune. Localement, en particulier dans certaines zones du nord et du sud de la région, le déficit atteint parfois 20% à 30% pour les générations nées autour de 1950. Une analyse appropriée permet de savoir si ce sont surtout des femmes qui partent ou des hommes qui arrivent, d'où et avec quelles caractéristiques. Tout cela simplement, en perspective historique.

* *

**

4 - LES LIMITES

Le fichier historique doit apporter beaucoup non pas tant par la nouveauté des informations -toutes étaient virtuellement disponibles puisque contenues dans les fichiers de base- que par le mode d'obtention et de présentation des résultats, ainsi que leur mise en perspective. Toute information, peut devenir, à terme, quasiment disponible instantanément. Pour séduisants que puissent être les résultats obtenus, ils n'en sont pas moins parfois très fragiles. En effet, comme dans toute exploitation statistique il existe des limites dans la qualité et dans l'interprétation des résultats qu'il ne faut pas oublier pour une bonne analyse.

L'incertitude des échantillons

La première difficulté provient de ce que les fichiers de référence correspondent à des exploitations par sondage. Cette limitation est incontournable dans le cas où l'INSEE n'a pas réalisé de saisie exhaustive des bulletins du recensement comme en 1968. Dans d'autres cas on disposait à la fois de données reposant sur l'exploitation exhaustive et par sondage. L'incertitude inhérente à l'utilisation d'échantillons peut être grande, notamment quand il s'agit du recensement de 1962 puisque le taux de sondage était de 1 sur 20.

*L'imprécision
inhérente
aux
sondages*

Quand elles existent, les données issues de l'exploitation exhaustive peuvent être de qualité moindre que celles des sondages. Les contrôles de différentes phases d'exploitation étaient peut-être moins rigoureux. Mais l'affirmation qui tendrait à montrer qu'une information par sondage bien collectée est meilleure qu'une information exhaustive victime de biais doit être nuancée. Le choix des fichiers de sondage résulte aussi de la nécessité de ne pas charger trop de données pour ne pas alourdir les exploitations.

Abondance de données nuit parfois

En fait une extension intéressante consisterait à enrichir le fichier de référence par la prise en compte, pour un même recensement, des données issues des différents modes de collecte. On disposerait ainsi parfois de deux données pour caractériser, à un même moment, la même variable. Cet avantage pourrait être aussi un inconvénient. Un utilisateur non averti pourrait en être troublé.

Pourtant, de fait, c'est déjà ce qui se passe. Nombre de nos clients sont troublés face aux différences qui apparaissent parfois entre des chiffres qui se rapportent au même contenu.

*Un
exemple
sur
Altkirch*

Soit Altkirch, qui a été choisie car elle est la première des communes du Haut-Rhin sur la liste de celles qui ont plus de 2 000 habitants.

Le fascicule jaune donne des séries historiques pour les trois derniers recensements, le fascicule vert présente seulement les deux derniers recensements.

Nous lisons :

	1982	1990	Variation 90/82	
			absolue	en %
Population totale				
fascicule jaune (1/1)	5 278	5 090	- 188	- 3,6
fascicule vert (1/4)	5 488	5 081	- 407	- 7,4
Chômeurs				
fascicule jaune (1/1)	127	105	- 22	- 17,3
fascicule vert (1/4)	136	68	- 68	- 50,0

Selon les chiffres choisis, on pourra donc tirer des conclusions différentes puisque les taux d'évolution varient du simple au double quand il s'agit de la population totale, du simple au triple quand il s'agit du chômage.

Les données du fascicule jaune correspondent au dépouillement exhaustif, ceux du fascicule vert proviennent du sondage au quart.

Les différences qui apparaissent ne sont certes pas incompatibles avec les formules donnant l'incertitude des résultats puisque des données au quart sont forcément imprécises quand il s'agit de population de faible taille.

Ce qu'il faut seulement dire, c'est que la publication de chiffres aussi différents est une gêne évidente pour nos clients.

*Une gêne
évidente pour
établir un
diagnostic
de zone*

Ne pourrait-on pas envisager d'harmoniser les résultats afin de ne donner, au moins pour les variables essentielles, qu'un seul chiffre par recensement? Il y aurait sans doute une procédure à inventer qui limiterait les conséquences fâcheuses de telles divergences dont on imagine facilement qu'elles peuvent être utilisées dans le débat local.

Bien sûr un fichier historique qui contiendrait les données de base des différentes phases de dépouillement serait certes très lourd, mais assorti d'un mode d'emploi, il permettrait sans aucun doute un traitement approprié des résultats.

L'avantage d'un fichier historique qui contiendrait les deux chiffres serait d'offrir à l'utilisateur le choix de la bonne donnée, non pas en fonction de l'opportunité politique (encore que l'on ne saurait faire grief à un élu du choix des armes pour justifier son point de vue), mais en fonction de la plausibilité quand deux chiffres sont contradictoires. En effet les chiffres doivent être replacés dans le contexte historique.

Des faiblesses dans la collecte

Les recensements ne sont pas exempts d'anomalies de collecte. Ce peut être dans telle ville des défaillances d'agents recenseurs, voire même localement des fraudes. Ces accidents vont créer parfois des ruptures dans les séries sans gravité quand il s'agit de faire des études sur les structures, mais assurément plus gênantes quand il s'agit de travaux sur les évolutions établies d'un recensement à l'autre. De même on sait que des biais se sont produits au moment de la collecte, par exemple sur le mode de chiffrage des femmes actives dans l'agriculture ou sur le code de l'activité économique et de la commune de travail du salarié.

*Des erreurs
d'observation*

Le difficile problème des nomenclatures

Enfin les changements de nomenclature sont aussi à l'origine de ruptures fâcheuses. En ne retenant que trois postes pour désigner l'activité économique on ne risque certes pas de faire de trop graves erreurs d'interprétation. Il n'en est pas de même pour les codes de catégories socio-professionnelles dont les modalités ont été profondément remaniées entre 1975 et 1982. La variation brute du nombre des ouvriers n'a assurément pas de sens sur cette période. En revanche on peut admettre que des indicateurs qui comparent le pourcentage d'ouvriers d'un bassin d'emploi au même pourcentage régional gardent leur sens, surtout dans une perspective à long terme. De même les bilans migratoires peuvent probablement être calculés sans dommage par CS.

*Des
ruptures
dans
les séries*

Aux difficultés inhérentes aux nomenclatures s'ajoutent celles relatives aux découpages géographiques. Des communes ont changé de département, certaines ont changé

de contours, d'autres ont fusionné et parfois l'union n'est qu'éphémère au grand dam des statisticiens. Tous ces handicaps ne sont sans doute pas insurmontables même s'il est irréaliste de prétendre les dépasser dans de brefs délais.

L'ordre caché

L'expérience appuyée sur l'observation d'une multitude de séries provenant de l'exploitation du fichier historique rassure sur la capacité des recensements à fournir une information intéressante, même pour des zones géographiques de faible taille. Sauf quand il s'agit de zones très petites (il n'est pas possible de fixer un seuil pour la taille d'une zone au-dessous de laquelle le fichier historique serait inexploitable) pour quelques cas exceptionnels, les résultats obtenus semblent tout à fait satisfaisants.

*Faire
parler
les
chiffres...*

S'agissant de la première exploitation nationale faite au niveau des départements, les principales anomalies décelées concernent :

- la Corse, région pour laquelle le suivi historique des données est impossible,
- les navettes en 1962 : par exemple, les Bretons qui ont déclaré la région parisienne comme lieu de travail sont beaucoup trop nombreux. De même les actifs pour lesquels un lieu de travail à l'étranger a été déclaré sont beaucoup trop nombreux. Cette anomalie est d'autant plus frappante qu'elle touche souvent des départements éloignés de la frontière.

*...au besoin
en leur
donnant une
correction
appropriée...*

Dans le premier cas, il vaut mieux faire taire les chiffres, dans le second cas, une correction doit suffire.

De toutes façons il en est des recensements comme des autres sources statistiques. Leurs insuffisances ne doivent pas les condamner. S'agissant de la géométrie, on dit parfois que tout l'art est de savoir raisonner sur des figures fausses. Par analogie, tout l'art du statisticien n'est-il pas de dégager des idées justes à partir de chiffres qui ne le sont pas, ou ne le sont pas toujours? Il existe, par delà les inexactitudes, une cohérence dans les chiffres qu'il faut savoir déceler.

*...sinon
les faire
taire*

S'il est souhaitable, pour ne pas déboussoler l'utilisateur, de conserver la possibilité de se raccrocher aux chiffres de base tels qu'ils sont publiés dans les fascicules des recensements, il n'en demeure pas moins qu'un objectif fascinant serait de créer des séries temporelles lissées, corrigées des principales anomalies repérées, dans le but de fournir des représentations adéquates des phénomènes sur lesquels on poursuit l'investigation. Mais la réalité elle-même n'est-elle pas sans aspérités ?

*Comment lisser
les séries
sans gommer
la réalité ?*

L'ordre, dit Süssmilch, se dissimule dans les cas particuliers et dans le détail, mais la concordance apparaîtra avec plus d'évidence quand, avec le temps, on aura rassemblé des listes plus nombreuses et plus importantes, "car l'ordre ne peut être révélé que par les grands nombres : les inexactitudes des petits chiffres disparaissent d'autant plus que les nombres que l'on compare sont plus grands". Pour n'en donner qu'un exemple, à recenser une maison une par une, on trouverait tantôt rien que des filles, tantôt rien que des garçons, mais, à partir du moment où l'observation peut jouer sur de grandes masses et de longues périodes, l'ordre caché derrière l'irrégularité apparente des phénomènes ne tarde pas à se manifester.

Jacqueline HECHT, Pour une histoire de la Statistique tome I - ECONOMICA