

98,9% des femmes et 99,6% des couples. Celles de naissance des enfants sont un peu moins exactes à 93,8% pour les hommes, 97,9% pour les femmes et 98,3% pour les couples. Pour les migrations, les dates exactes à plus ou moins un mois près sont beaucoup moins fréquentes : 61,8% pour les hommes, 65,2% pour les femmes et 67,3% pour le couple. Si l'on se contente d'une précision à plus ou moins un an près, ces pourcentages montent à 88,4% pour les hommes, 90,3% pour les femmes et 91,8% pour le couple. En dépit de ces datations erronées, les erreurs ne modifient pas ou changent peu l'ordre logique dans lequel les événements se produisent. Cet ordre est correctement remémoré, comme le montrent de nombreuses analyses biographiques faites sur ces données⁽³⁾.

Ainsi les quotients instantanés de migration après le mariage ne sont pas discernables entre eux, selon que l'on travaille sur les réponses de l'homme, de la femme, des conjoints réunis ou des registres de population belges. Si l'on fait intervenir dans une analyse semi-paramétrique diverses caractéristiques de ces séjours (date de début de séjour par rapport à celle du mariage, nombre d'enfants en début de séjour, statut d'occupation du logement), les mêmes effets apparaissent comme significatifs, quelle que soit la source utilisée. Enfin un modèle encore plus complexe analysant les liens entre première migration et première naissance après mariage en fonction de diverses caractéristiques de la femme, montre un effet identique de tous les paramètres sauf un, correspondant à la déclaration de l'homme sur l'âge de sa femme au mariage (avant 20 ans). On voit à nouveau apparaître ici la moins bonne qualité des déclarations des hommes isolés. On peut ainsi conclure que, dans la majorité des cas, les erreurs de datation jouent le rôle d'un bruit de fond d'où l'on peut extraire une information cohérente quelle que soit la source utilisée. La mémoire semble donc fiable, là où l'analyse l'exige.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

Une relecture effectuée immédiatement après la collecte par les Directions régionales de l'Insee a permis aux enquêteurs, le plus souvent par simple appel téléphonique, de redresser certaines lacunes ou contradictions : cette étape a contribué pour beaucoup à la qualité des données car un contrôle tardif ne permet plus que le seul constat des erreurs.

L'organisation du traitement des questionnaires à l'Ined évitait une détérioration de l'information : la même personne enchaînait la codification, la saisie, l'examen des diagnostics fournis par un programme de 200 tests de cohérence et les corrections des fichiers jusqu'à apurement complet des erreurs « redressables » sans invention d'information. Les tests repèrent, par exemple, le divorce injustifiable d'un célibataire, le chevauchement de périodes successives, les périodes non décrites, etc. La collecte s'étant achevée au moment

⁽³⁾ Pour plus de détails voir Daniel Courgeau (1991).

où l'Insee expérimentait la nouvelle nomenclature des catégories sociales (CS), on a jugé de circonstance d'opérer une double codification des professions selon la nomenclature des catégories socioprofessionnelles (CSP) de 1954 et des professions et catégories socioprofessionnelles (PCS) de 1982. Enfin, un programme de codification automatique des communes a été mis au point, compte tenu des 150 000 mentions dans les questionnaires. L'indication sur écran des numéros de département et des toutes premières lettres de l'intitulé de la commune suffisant à son identification et à sa codification.

4.2. Fichiers de base de l'enquête

Il existe une base SAS de l'enquête qui est constituée d'un enregistrement par individu. Une partie des questions n'y figure cependant pas.

4.3. Choix de la saisie de la variable temps

La saisie des mois et années déclarées dans le questionnaire a conduit à une élaboration originale. Le traitement des dates mal déclarées a retenu une formule de correction sans falsification des dates déficientes : le codeur substituait à une date absente ou incohérente la date centrale des intervalles disponibles en en portant la mention sur le code du mois de l'événement. Les valeurs 21 à 32 indiquaient l'omission du mois, 41 à 52 celle de l'année, 61 à 72 signifiaient que la date originelle était incompatible avec le reste du questionnaire.

4.4. Choix de l'unité de temps pour l'analyse

Si l'on a parfois utilisé le mois, la plupart des analyses n'ont tenu compte que de l'année de l'événement.

4.5. Utilisation des méthodes d'analyse des biographies

Les logiciels suivants ont été utilisés :

- ROOT.RAT et EVACOV.FOR (écrits par Éva Lelièvre);
- RATE (écrit par Nancy Tuma);
- SAS;
- TDA (écrit par Goetz Rohwer);
- STATA (routines reprenant les programmes ROOT.RAT et EVACOV.FOR écrites par Philippe Bocquier).

4.6. Autres méthodes d'analyse

Tris simples ou multiples, régressions, modèles logit, modèles multi-niveaux.

4.7. Les utilisateurs des données

- Brigitte Baccaïni (thèse, articles);
- Marie-Ange Cambois (article);

des revenus sur lesquels il avait fait silence, voire une résidence secondaire dont il n'avait pas parlé.

Mais le questionnement au téléphone ne permet pas seulement de corriger des réponses écrites et d'obtenir une réponse à des questions non répondues ; il permet aussi de poser, sur le ton et au rythme de la conversation, *des questions qui ne figurent pas au questionnaire*, qui l'auraient allongé au-delà du raisonnable, et risquaient fort d'être mal acceptées, venant de leur Caisse de retraite⁽¹⁷⁾. Ainsi par exemple nous avons pu apprendre, des sujets joints au téléphone, des informations *sur leurs enfants* : le métier de ces enfants (et la part des chômeurs...), les raisons et la date de la venue en province de ces enfants (beaucoup de Parisiens de cette génération, migrants ou non, ont au moins un enfant en province), si les enfants des Parisiens retirés en province qui vivent dans la même région qu'eux y sont venus avant ou après eux, si des enfants ont l'intention d'y venir, où habitent les petits-enfants adultes, etc. Ces informations supplémentaires (elles portent sur *les enfants*, mais aussi sur *les relations sociales*, sur *les vacances* – où, avec qui ? –, sur *les animaux domestiques*) apportent beaucoup à la compréhension des façons de vivre.

4. TRAITEMENT DE L'INFORMATION

Au préalable, il faut insister sur le fait que les données se présentent sous la forme de documents matériels : relevés à partir de documents administratifs, enquêtes auto-administrées, annotées et complétées au téléphone par les membres de l'équipe, enquêtes chez l'enquêté saisies sur papier ou bandes magnétiques. Seule *une partie* de l'information a été informatisée : la description de base pour la totalité des individus du panel, et les 3 enquêtes décennales. Sur des thèmes particuliers, pour des sous-populations spécifiques et sur des échantillons beaucoup plus restreints, on peut disposer d'une information beaucoup plus riche. De plus, à partir de l'ensemble des documents disponibles, on a réalisé des *résumés-analyses* pour près de 70 % de nos enquêtés. Ils ont été rédigés en rassemblant tout ce que l'on sait d'eux sur divers thèmes, et permet de comprendre leur situation, leur façon de vivre.

4.1. Traitement du questionnaire, principes et difficultés de codification

S'il fallait ne citer qu'un point, ce serait l'opiniâtreté avec laquelle nous avons travaillé à coder une catégorie sociale (CS individuelle et CS du ménage). Elle devait décrire au mieux une situation parfois complexe et fluctuante à l'échelle d'une vie professionnelle entière. Nous avons choisi de caractériser *la CS de la maturité*. Nous avons porté une attention particulière à la distinction entre employés peu qualifiés et plus qualifiés, et à la limite inférieure de la catégorie des cadres. Nous avons de la sorte abouti à une classification

(17) Nous avions fait de même pour le premier panel, lors des 2 grandes enquêtes de 1985-1986 et 1995-1996, dans lesquelles la majorité des sujets avaient eux aussi été interrogés par téléphone.

en 9 postes fondée sur type de métier et qualification⁽¹⁸⁾. Nous avons aussi constitué une variable agrégée en 4 postes, qui est, cette fois, une échelle. Le premier niveau regroupe personnels de service, employés et ouvriers non qualifiés (la mobilité à l'intérieur de cette catégorie a été fréquente au cours de la vie de travail), le second regroupe employés et ouvriers qualifiés.

4.2. Fichiers de base des suivis

Nous avons évoqué plus haut la complexité du corpus des données (cf. 1.3). Nous avons rassemblé autour d'un fichier principal une vingtaine de fichiers correspondant aux différentes sources documentaires (3 enquêtes principales, actes de naissance et de décès...) et aux différentes unités d'observation (enquêtés, résidences, états matrimoniaux, enfants...).

Les histoires matrimoniale, professionnelle et résidentielle sont saisies dans autant de fichiers différents. L'histoire résidentielle est répartie en 2 fichiers : un fichier « mort » correspondant à l'histoire rétrospective antérieure à l'enquête de 1985 et un fichier « vivant » correspondant à l'histoire post-retraite et mis à jour annuellement. Il faut ajouter qu'un certain nombre de dates remarquables dans la vie de l'enquêté se trouvent dans d'autres fichiers : arrivée à Paris, acquisition d'une résidence secondaire, début de cohabitation avec de vieux parents, etc.

4.3. Choix de la saisie de la variable temps

Jusqu'à présent les dates ont été saisies avec une précision à l'année, exception faite de la naissance connue au jour près. Compte tenu de la durée des séjours en institution, qui peuvent être très courts, les dates d'entrée en institution et de décès sont saisies au jour près.

4.4. Choix de l'unité de temps pour l'analyse

Soit nous utilisons des durées annuelles (définies par différences de dates annuelles), soit nous employons des couples avant/après un événement choisi. Pour l'âge au décès ou les durées de séjours finaux en institution, nous pouvons travailler au jour près.

4.5. Utilisation des méthodes d'analyse des biographies

Nous n'avons pas eu recours à de telles méthodes. Ce n'est pas une position de principe et nous comptons le faire. Mais nous savons que notre corpus n'est pas achevé : un cinquième de nos enquêtés sont encore vivants et nous n'avons pas épuisé les possibilités de mises à jour de nos données.

4.6. Autres méthodes d'analyse

Nous avons fait, bien sûr, de très nombreuses tabulations, et quelques analyses factorielles. Nous avons très souvent constitué des listes d'enquêtés selon des critères choisis afin de retourner aux documents matériels, de les

Les principales difficultés ont porté sur le recensement des logements. Parfois, l'enquêté oubliait le premier logement puis le mentionnait ensuite. La date de début du parcours logement était parfois fautive en raison du choix entre la date des 25 ans, le mariage, ou l'indépendance résidentielle.

Des problèmes également dans la partie ménage où l'on reconstituait la liste de toutes les personnes avec qui l'enquêté a vécu au moins un an. Ces erreurs ont pu, dans l'ensemble, être rectifiées à la lecture du questionnaire. En revanche, on n'a aucune idée des personnes oubliées consciemment ou inconsciemment (comme les filles au pair, apprenti...)

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

• Circuit du questionnaire

Une première phase, qui s'est avérée essentielle, compte tenu de la difficulté du questionnaire, a consisté à relire attentivement tous les questionnaires et éventuellement à recontacter les enquêtés en cas d'incohérence ou de non-réponse. Cette phase de relecture s'est effectuée à l'Insee en collaboration avec les chercheurs de l'Ined.

La codification a eu lieu à l'Ined ainsi que la saisie par des vacataires étudiants, encadrés par le service des enquêtes.

• Difficultés de la codification

La principale difficulté était de déterminer à quel moment on commençait le début de l'histoire résidentielle. Malgré des instructions précises : date d'émancipation, date de mariage ou à défaut date des 25 ans de l'enquêté (il fallait prendre la plus ancienne des trois), des erreurs ont été commises. En fait, un certain nombre d'enquêteurs oubliaient, dans un premier temps, le premier logement, lorsque l'enquêté, âgé de plus de 25 ans vivait chez ses parents, et rectifiaient ensuite en fin de la partie logement. Les numéros de logement ne correspondaient plus. Le questionnaire avait été conçu de façon trop rigide et ne permettait pas d'intercaler une période résidentielle oubliée dans un premier temps.

La deuxième difficulté a été de bien noter le numéro de logement dans lequel un événement (naissance des enfants, séparation du couple...) est arrivé. Souvent, les enquêteurs confondaient date de départ du logement X avec date d'arrivée dans le logement X+1. La relecture a permis de rectifier ces erreurs.

Autre difficulté : la codification du nombre de pièces. Malgré des instructions détaillées, il reste des imprécisions en particulier pour les chambres de bonne, les agrandissements du logement. Un individu pouvait très bien avoir habité un logement de trois pièces pendant une dizaine d'années, puis un logement de cinq pièces après avoir agrandi son logement en achetant le logement de deux pièces contigu. Dans ce cas, on a considéré cet agrandissement comme un déménagement dans le même immeuble.

D'une manière générale, on a sous-estimé la mobilité du parc lui-même. Sans changer de logement, l'enquêté pouvait vivre dans un appartement dont le nombre de pièces s'était modifié ; son statut d'occupation pouvait également changer. Les cas les plus fréquents étaient le passage du statut d'hébergé par la famille à celui de locataire ou propriétaire (les parents décèdent et l'enfant reprend le bail à son nom ou encore hérite du logement, le passage de locataire d'un logement loi de 1948 à propriétaire par exemple).

4.2. Fichiers de base de l'enquête

Plusieurs fichiers existent :

- un fichier contenant les principales variables démographiques et professionnelles ;
- un fichier logement avec toutes les données de chaque période résidentielle ;
- un fichier corésidents ;
- un fichier sur les différents membres de la famille (lieu de résidence, statut d'occupation...);
- un fichier sur les autres logements, les résidences secondaires, les projets de retraite et le logement idéal.

4.3. Choix de la saisie de la variable temps

La variable temps a été saisie à partir du mois et de l'année.

4.4. Choix de l'unité de temps pour l'analyse

C'est l'année qui a été le plus souvent retenue. Les mois ont été pris en compte dans le calcul de durée ou d'âge.

4.5. Utilisation des méthodes d'analyse des biographies

ROOT.SAS (mis au point par Éva Lelièvre)

Autres Logiciels utilisés pour l'exploitation :

- logiciels de saisie de l'Ined : SID ;
- logiciels de traitement statistique : SAS, SPAD, SPADT.

4.6. Autres méthodes d'analyse

Analyses sur les trajectoires familiales, résidentielles et géographiques.

4.7. Les utilisateurs des données

Des chercheurs essentiellement :

- Catherine Bonvalet, Ined ;
- Dominique Maison, Ined, Université Paris X ;
- Paul White, Professeur à l'Université de Sheffield, Grande-Bretagne ;

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

Les premières données tirées de l'enquête ont été exploitées par l'Office central de statistique en collaboration étroite avec l'Institut de statistique et de démographie, à l'École d'économie de Varsovie. L'ensemble du procédé de codification et de contrôle a été exécuté par l'Office central de statistique. Afin de vérifier la cohérence des données du questionnaire, nous avons mis au point un programme spécial qui enregistre et contrôle la cohérence des informations. Le programme de base se composait de 200 sous-programmes chargés de s'assurer de la qualité des données.

Dans un deuxième temps, l'Institut de statistique et de démographie recevait les données de l'Office central de statistique. Au sein de l'Institut ont été élaborés de nouveaux programmes destinés à vérifier les données et s'appuyant sur des options spécifiques pour évaluer la qualité des données par rapport au procédé des échantillons et à l'étude de la cohérence des structures et répartitions empiriques en général.

4.2. Fichiers de base de l'enquête

L'enquête rétrospective polonaise de 1988 a un manuel d'utilisateur en version polonaise et anglaise. Les données sont disponibles aux formats TXT, DBF, SPSS et ZOO.ARCHIVE pour le programme TDA. Les données sont organisées suivant les dix sections du questionnaire.

4.3. Choix de la saisie de la variable temps

Les variables temps n'ont pas été tout de suite entrées dans la base de données. Les bases de données comprennent les dates des événements, par exemple : la date du premier mariage – le mois et l'année – et la date de la première dissolution. L'utilisateur des bases de données peut créer des variables temps, dans le cas présent la durée du mariage.

4.4. Choix de l'unité de temps pour l'analyse

Le temps, selon l'analyse, est envisagé comme :

- le temps calendaire ou l'âge,
- la durée du processus depuis son origine ou la durée de séjour dans l'état étudié.

4.5. Utilisation des méthodes d'analyse des biographies

De nombreuses méthodes de pointe d'analyse biographique ont été utilisées au cours du travail de recherche sur les données de l'enquête. Cela n'a posé aucun problème, au contraire nous pensons avoir contribué au développement de ces méthodes.

reconstruire et aussi plus indiscret. À la suite d'une première collecte, on a décidé de l'insérer à la fin de l'entretien, mais cette décision n'a pas amélioré la crédibilité.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

Dans le questionnaire, il n'y avait pas de questions ouvertes et cela a simplifié nettement la codification. Quelques problèmes se sont posés pour l'identification de la réponse à certaines questions. Par exemple, le type de localité (il n'a pas toujours été facile de délimiter le centre historique par rapport au reste de la ville, ou celui-ci de la zone suburbaine), le statut professionnel (la distinction entre employé et ouvrier est particulièrement floue), le type de logement (comment par exemple, classer une maison de campagne rénovée qui ressemble à une petite villa indépendante).

Pour le reste, les codes étaient déjà spécifiés dans la case de réponse du questionnaire. La saisie a été effectuée en « couple » – l'enquêteur et un de mes collaborateurs – et à ce stade, cela a permis de réduire les erreurs à un niveau négligeable.

4.2. Fichiers de base de l'enquête

Le fichier de base, à partir duquel des fichiers allégés à format fixe ne contenant que des informations circonstanciées sont extraits lors des différents traitements, est un fichier formé d'un premier enregistrement consacré aux informations générales sur le ménage enquêté et aux données sur ses perspectives de logement. Il s'agit de toutes les informations contenues dans les volets 1 et 5 intégrées sous la forme de trois chiffres qui indiquent respectivement le nombre de membres du ménage, le nombre d'activités professionnelles enregistrées et celui des logements occupés. Pour chacune de ces séries d'informations, un enregistrement suit et donc : un pour chaque personne qui fait ou a fait partie du ménage, un pour chaque activité professionnelle exercée durant son appartenance au ménage, un pour chaque logement occupé. Chacun de ces quatre types d'enregistrement a son propre format fixe, mais le nombre d'enregistrements relatif à chaque ménage enquêté reste, bien sûr, variable.

4.3. Choix de la saisie de la variable temps

Je n'ai aucun regret d'avoir décidé de me contenter du niveau de précision de l'année, même si cela augmente les situations d'apparente contemporanéité des événements. Du reste, comme l'a déjà démontré J. Duchêne⁽¹⁾, quand on sollicite la mémoire à propos d'événements lointains, il est souvent

⁽¹⁾ J. Duchêne, 1985, Un test de fiabilité des enquêtes rétrospectives « Biographie familiale, professionnelle et migratoire », Chaire Quetelet : *Migrations internes*, Louvain la Neuve.

les points de vue, après les échecs du Parti majoritaire et avec les espérances de l'opposition. Il est vrai qu'en travaillant la qualité d'un questionnaire on peut le rendre attrayant, mais comment penser que les Roumains ont pu prendre plaisir à raconter leur vie dans cette période de contestations et d'incertitudes alors que les valeurs les plus sûres étaient remises en question? Pourtant, nous avons espéré que l'objectivation de leur passé allait apparaître pour beaucoup d'enquêtés comme dédramatisante et nous restons convaincue des effets libérateurs de la communication, en général.

L'interview a été considérée par les enquêtés comme trop longue et exigeant trop de précisions au niveau des questions concernant le temps et l'espace. La reconstitution du calendrier de vie requiert un effort de concentration important et l'exercice auquel est soumise la mémoire devient parfois périlleux.

En ce qui concerne la technique de recueil de l'information biographique, la lecture singulière que j'ai faite d'une bonne partie des questionnaires a mis en lumière les tentatives des enquêteurs et des enquêtés pour interpréter les questions et se les approprier, soit parce que les catégories proposées se sont révélées insuffisantes pour rendre compte des particularités des itinéraires, soit parce que les nuances que ces personnes ont souhaité apporter, c'est-à-dire leur version des événements de leur propre vie, n'ont pu s'inscrire dans le cadre pré-établi des questionnaires. Les précisions que certaines personnes ont demandé à l'enquêteur de noter débordent parfois le cadre du questionnaire, témoignant du souci des enquêtés de rendre compte de la complexité de leur histoire. Les pages de garde des questionnaires portent aussi trace de notations saisies «à chaud» par l'enquêteur. L'ensemble de ces observations a beaucoup enrichi le questionnaire. Bien que le questionnaire standardisé semble permettre une plus grande objectivation des faits biographiques, j'ai remarqué une élaboration subjective de la présentation par l'individu de son parcours, à travers la sélection des événements proposés et le type de précisions apportées.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

Pour assurer l'homogénéité de la codification nous avons préparé un document «Instructions de codification». Il permet parfois de rectifier dans le bon sens les erreurs de l'enquêteur, ou tout simplement donne des règles de classement aussi claires que possible. Mais comme les instructions de codification ne peuvent envisager tous les cas, ces situations et les consignes afférentes ont été enregistrées au fur et à mesure qu'elles sont apparues au chiffrage de l'enquête.

Je me suis chargée de la relecture de l'ensemble des questionnaires. Le chiffrage de l'enquête et sa saisie ont été assurés par les salariés de l'IMAS-Bucarest.

De façon simplifiée, le questionnaire «3B.R» comprend quatre unités statistiques : l'enquêté, ses enfants, ses périodes d'emploi, ses résidences principales.

Cette structure est reproduite par le programme de saisie : chaque questionnaire comprend autant d'enregistrements informatiques que d'enfants, de périodes d'emploi et de résidences principales.

Le codeur va substituer, à une année absente ou incohérente, l'année centrale de l'intervalle des possibles en en portant mention sur le code de l'année de l'événement. On a retenu cette formule de correction des dates déficientes, utilisée par l'Ined, car cet artifice de recodage est d'une grande efficacité informatique.

4.2. Fichiers de base de l'enquête

Rappelons que nous n'avons recueilli que les résidences ou les emplois occupés durant au moins six mois. Pour la cohérence des biographies, une succession de périodes de moins de six mois, totalisant cette durée, sera décrite comme une période unique désignée par «résidences variables», «types d'activité variables», «lieux d'emploi variables». Cette consigne a cependant pour inconvénient d'éliminer la majorité des périodes de maladie et de chômage, et peut donner lieu à une variation par suite du mode de calcul.

Pour la saisie des questionnaires, nous avons utilisé le programme SID (Saisie interactive des données). À l'Ined, ce programme ne réalise que les contrôles de champ (chaque variable ne prend que les valeurs admises). La liste des codes admis est fermée. Le programme de saisie SID de notre enquête «3B.R», tel que nous l'avons élaboré à l'Ined contient 27 blocs ; pour le bloc «enfants» nous avons prévu 15 occurrences et pour les blocs «emplois et logements», 30 occurrences. La saisie a été réalisée par le Centre d'informatique de la Commission nationale de statistique de la Roumanie, à la suite d'un accord passé avec l'Ined.

Les *questions ouvertes* seront traitées à part, par la méthode sociologique de l'analyse du contenu. On a trié un certain nombre de questionnaires, les mieux remplis, en espérant qu'ils illustrent de façon concrète une situation statistiquement fréquente. Les questions ouvertes ont été soumises à l'analyse de contenu, les catégories de l'analyse étant élaborées et établies sur la base de la lecture et de l'analyse de ces réponses. Nous avons assuré la relecture des questionnaires et l'analyse de contenu.

Les fichiers de base de l'enquête sont : la base d'analyse SAS et la base d'analyse par TDA.

4.3. Choix de la saisie de la variable temps

La saisie de la variable temps se fait par année.

4.4. Choix de l'unité de temps pour l'analyse

Dans l'analyse, nous avons traité le temps par des variables de durée. La particularité des données de durées est qu'elles peuvent s'interpréter facilement comme résultant d'un processus stochastique sous-jacent. Ce processus

les rapports avec les hommes, dans les relations avec les autres femmes (perception du rôle des femmes, vie associative). Il s'agit d'analyser les causes et les spécificités des migrations féminines, d'étudier les différentes stratégies mises en œuvre par les femmes pour s'insérer en milieu urbain, d'appréhender comment elles se représentent leur rôle dans la société de manière plus générale. La migration féminine se traduit-elle par l'affirmation de l'individu et la conscience d'une amélioration de son statut ?

L'enquête complémentaire qualitative a été conduite au moyen d'entretiens individuels auprès d'un sous-échantillon de l'enquête biographique. Cinquante entretiens ont été réalisés à Dakar auprès de 39 femmes et 12 hommes. Les personnes choisies pour l'entrevue correspondaient au profil souhaité à partir des critères, définis au préalable, concernant le statut matrimonial, l'activité, le profil migratoire. Il s'agit d'entretiens semi-directifs conduits en langue locale avec l'aide d'un interprète. L'ensemble des interviews a été enregistré et par la suite retranscrit en français pour en permettre l'analyse.

3.4. Évaluation de la collecte

Comme nous l'avons mentionné plus haut, un des points faibles de la collecte pour l'enquête de Dakar est le fort taux de déperdition (refus, déguisés ou non, mobilité de certains travailleurs, etc.) au niveau de la collecte biographique. Cette déperdition semble pouvoir être réduite par un enchaînement plus rapide entre l'enquête ménage et l'enquête biographique.

Malgré de nombreux contrôles sur le terrain, des corrections de cohérence de dates ont été faites au bureau après fusion des trois biographies, sans trop de problèmes cependant, grâce aux informations relevées sur la fiche AGEVEN. D'ailleurs, cette fiche semble avoir bien servi : elle constituait un support très utile pour le remplissage des questionnaires. De plus, les enquêtés semblent avoir apprécié cette manière de résumer leur vie, et demandaient souvent qu'on leur en donne une copie comme souvenir, et probablement aussi pour leur éviter de refaire le même travail de reconstitution pour d'autres occasions.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

L'essentiel de la codification s'est fait directement sur le terrain, à l'exception des professions, dont on notait de façon détaillée la description. À l'issue du terrain, une nomenclature exhaustive des professions recueillies a été établie. Des regroupements ont été opérés en s'inspirant de la Classification internationale type des professions (CITP-88), tout en attachant une attention particulière aux activités informelles.

4.2. Fichiers de base de l'enquête

Pour chaque période, les dates de début et de fin ont été saisies. Cela dit, étant donné qu'on ne prenait pas en compte les périodes de moins de six

les hommes et pour certaines femmes travaillant dans le commerce. Les équipes ont ainsi été obligées de demeurer sur le terrain parfois jusqu'après 23 heures. La période d'enquête chevauchait celle des élections législative et présidentielle. Contrairement à nos attentes, ces événements n'ont aucunement perturbé le déroulement de l'enquête. Les efforts faits en ce qui a trait à la sensibilisation (communiqué à la radio et prise de contact avec les populations) ne sont certainement pas étrangers à ce résultat. Les problèmes les plus importants se sont posés plus particulièrement pendant la période de carême religieux du mois de mars qui coïncidait également avec le moment de l'année où les températures sont les plus élevées. Il était parfois difficile de demander à des enquêtés qui n'ont ni mangé ni bu depuis plus de douze heures de prendre une heure pour répondre à un questionnaire. Ce problème a également affecté le rendement de certains enquêteurs qui respectaient les interdits liés à la pratique du Ramadan.

Les problèmes de mémoire n'ont pas été particulièrement importants. En fait, la fiche AGEVEN, qui résume en quelque sorte la vie de l'individu, s'est avérée être un outil particulièrement performant pour la collecte de données biographiques auprès des populations du Mali. La précision mensuelle de certains événements n'a, par ailleurs, pas pu être donnée pour de nombreux événements et ce, malgré l'utilisation du calendrier historique et du calendrier des saisons. Un calendrier des fêtes musulmanes aurait probablement été plus utile et devrait être utilisé à l'avenir.

4. TRAITEMENT de L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

Le précodage de la majorité des questions a accéléré le travail. En ce qui a trait aux lieux géographiques, aux professions et secteurs d'activité, le principe a été retenu de transcrire les réponses et de faire la codification *a posteriori*. Une liste de codes géographiques a ensuite été élaborée permettant notamment de distinguer les quartiers de Bamako. La nomenclature des professions, produite lors du Recensement malien de 1987, a été utilisée pour la codification des professions. Trois enquêteurs formés furent détachés pour porter les codes sur les questionnaires avant de les acheminer vers la saisie.

4.2. Fichiers de base de l'enquête

Les programmes de tirage aléatoire et de mise en forme des listes nominatives des individus ont été élaborés à partir des logiciels Basic et D-Base. Les programmes de saisie et d'apurement des fichiers ont été élaborés à l'aide de la version 2.26 du logiciel ISSA (Internat System for Survey Analysis) développé par l'Institute for Resource Development/Westinghouse. Avec le même outil, toutes les applications de transformation de fichiers ont été développées. Cette opération a duré jusqu'en août 1992. Les fichiers de données ont ensuite été transformés pour être traités à partir du logiciel STATA.

et allégés a, une fois de plus, été prouvée. De même, l'enregistrement chronologique des différents événements, connus par l'individu, s'est révélé efficace pour retracer son histoire de vie. Cependant des difficultés ont été rencontrées dans la collecte des informations concernant l'emploi, et dans le report des dates de la fiche AGEVEN au questionnaire ménage.

• *Cohérence de la biographie et problèmes de mémoire*

Un bon nombre de problèmes de cohérence ont été rencontrés dans les données collectées, notamment en ce qui concerne la date de naissance et la date de début de la première résidence, ainsi que la continuité des événements résidentiels et professionnels. Les problèmes de mémoire ont été à la base de nombreuses omissions et imprécisions dans le recueil des dates de certains événements, notamment familiaux (décès du père ou de la mère, mois de naissances, etc.).

• *Difficultés rencontrées*

La durée des interviews, qui duraient en moyenne 45 minutes par personne éligible, a provoqué dans certains cas des énervements et des abandons de la part des enquêtés. Il y a également le niveau intellectuel et la situation économique de l'individu, qui jouaient un rôle important sur sa plus ou moins bonne réceptivité vis-à-vis du recueil de son histoire de vie, étant entendu qu'un chômeur, par exemple, était moins enclin à se plier à cette opération, considérée comme une perte de temps et sans bénéfice pour lui. Les personnes âgées et de faible niveau intellectuel ne se sentaient pas concernées par cette étude, la trouvant trop « scientifique », n'apportant pas de réponses concrètes à leurs préoccupations de l'heure : trouver un travail décent, se nourrir, se loger et se soigner.

Il y a également l'implication personnelle des enquêtés, qui étaient très diserts en ce qui concerne leurs problèmes quotidiens et les « méfaits » des autorités publiques, mais répugnaient à répondre aux questions précises du questionnaire biographique.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

À la fin de chaque journée de collecte, les enquêteurs remettaient aux contrôleurs tous les questionnaires, ménages et biographiques, qui avaient été intégralement remplis. Les contrôleurs devaient s'assurer, à leur niveau, que tous les questionnaires qui leur étaient remis avaient été correctement remplis, en mettant un accent particulier sur la datation des événements. Ils devaient vérifier si les dates des événements figurant sur la fiche AGEVEN avaient été reportées sans erreur sur le questionnaire biographique. En cas d'erreur, les contrôleurs effectuaient les corrections nécessaires. En cas de problème grave

(information manquante ou erronée), le questionnaire était renvoyé sur le terrain, à charge pour l'enquêteur de recueillir la bonne information.

Une fois le questionnaire validé par les contrôleurs, il était remis aux superviseurs qui se chargeaient de vérifier la cohérence des réponses fournies et la séquence des dates et des événements; après cette opération, les questionnaires étaient rangés selon leur grappe d'appartenance, dans des boîtes d'archivage. La codification a commencé à l'issue de la collecte. Certains enquêteurs ont été retenus pour cette opération qui a duré plus que prévu (un peu plus de trois semaines). Les agents de codification ont procédé dans un premier temps au dépouillement systématique de toutes les questions ouvertes : ethnies, profession, diplôme obtenu, les questions relatives à la perception de la crise, etc. Cette phase a également permis de compléter les questionnaires biographiques à l'aide des fiches AGEVEN, et de corriger un bon nombre des incohérences des informations. Le dépouillement des questions ouvertes a relevé des réponses non appropriées aux questions posées. C'est le cas entre autres des mesures prises par les autorités pour sortir le Cameroun de la crise, les raisons de se retrouver au foyer et de n'avoir pas cherché un emploi, etc.

Une fois ce classement effectué, les agents de codification, sous le contrôle des superviseurs, procédaient au dépouillement systématique de toutes les questions ouvertes (noms des départements, profession, branche d'activité, ethnies, etc.). Au terme de ce dépouillement, les superviseurs ont élaboré une grille de codification, grille à partir de laquelle les agents de codification ont procédé à l'inscription des codes sur toutes les questions ouvertes. Des problèmes subsistent dans la nomenclature des activités et des professions.

Dans l'ensemble, la plupart des problèmes rencontrés dans la codification ont été résolus. Toutefois en ce qui concerne la variable ethnies, il était difficile de rattacher certaines ethnies déclarées aux grands groupes de la nomenclature existante. Des codes spéciaux ont été adoptés pour ces ethnies. Il reviendra à chaque chercheur, au moment des analyses, de procéder au regroupement nécessaire.

Parallèlement au travail de codification, un informaticien bien rodé en matière de saisie des données d'enquête a été recruté pour la conception du masque de saisie à l'aide du logiciel ISSA. Il aurait dû être associé à l'élaboration du questionnaire, notamment pour le problème de saut de questions. En effet la difficulté rencontrée dans l'élaboration du masque de saisie est l'absence, dans le questionnaire, des variables autorisant le saut. L'informaticien était obligé de mettre des 9 à la place des vides qu'aurait engendrés le saut des variables. Cette situation est liée aux contraintes du logiciel ISSA.

Ainsi, une fois le travail de codification achevé, toutes les conditions étaient réunies pour passer à la saisie des données collectées.

Ce travail a été effectué par des agents de saisie recrutés par l'Iforid et formés, pour la saisie, par l'informaticien ayant conçu le masque de saisie. Les questionnaires étaient saisis par grappe, et la saisie commençait par l'entrée des informations concernant le ménage, suivies des biographies des personnes éligibles dans ce ménage.

a tiré parti de cette aptitude du volet biographique, en changeant la séquence des parties du questionnaire pour commencer par ce volet : dans les cas où le choix de la personne qui devait être interrogée pour la biographie était évident, cette enquêtrice commençait l'entrevue par la section biographique. Ainsi, elle pouvait établir, dès le départ, une bonne communication avec les personnes interrogées et profiter de toutes les informations de la biographie, considérées à juste raison comme plus fiables, pour vérifier les réponses aux autres parties du questionnaire.

Il convient de souligner que le questionnaire, tel qu'il a été appliqué à Bogota, autorise une bonne évaluation de l'information collectée, du fait que les contrôles de cohérence possibles sont nombreux et que, pour les personnes ayant répondu au volet biographique, il est possible de comparer les données consignées dans la matrice biographique avec toutes les données des sections précédentes, ce qui permet un contrôle très fouillé de l'information collectée. De plus, la réalisation d'entretiens approfondis auprès d'un sous-échantillon de l'échantillon du premier passage de l'enquête statistique (de même que le second passage, actuellement en cours) a constitué en soi une opportunité très riche d'évaluation de l'enquête : l'approche qualitative, développée par l'anthropologue de notre équipe de recherche, a mis en évidence la bonne qualité globale de l'enquête statistique.

Même si l'appréciation générale du volet biographique est largement positive, on ne peut pour autant passer sous silence les problèmes de *mémoire*. Une « évidence » mérite toutefois d'être rappelée : l'univers enquêté à Bogota est une population jeune, d'autant plus que bon nombre de nos zones d'étude dans la capitale colombienne correspondent à des quartiers récents occupés par de jeunes adultes. De plus, l'obligation qui est faite aux habitants de Bogota de fournir des extraits authentifiés de leurs documents, pour toute démarche auprès de l'administration, contribue à une bonne mémorisation des dates. Enfin, le peuplement de Bogota se nourrit traditionnellement d'une migration relativement directe, comportant peu d'étapes intermédiaires. L'évaluation très positive du questionnaire doit donc être replacée dans ce contexte, nettement favorable à une collecte biographique centrée sur des questions de mobilité spatiale.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

L'étape de critique et de codification a débuté dès la fin de la collecte. Afin de tirer parti de leur bonne connaissance du questionnaire, nous avons fait effectuer ce travail par les deux personnes ayant supervisé la collecte, chacune d'elles vérifiant et codifiant les questionnaires renseignés par les enquêtrices n'ayant pas travaillé sous sa responsabilité.

La majeure partie du questionnaire étant précodée, les questions ouvertes, nécessitant donc une codification, sont peu nombreuses. Elles concernent : l'âge, les lieux (pays, département, municipalité, secteur de Bogota), les professions,

les branches d'activité, et la fréquence de séjour dans les autres logements. Les codes sont notés dans les cases grises prévues à cet effet dans le questionnaire.

L'information recueillie dans le calendrier de la section sur les *Systèmes de résidence* est codée sous la forme d'une variable synthétique dont les modalités décrivent les principaux types de rythme de séjour dans un logement : séjour continu, périodicité hebdomadaire de 1-2 jours par semaine, périodicité hebdomadaire de 3 jours ou plus par semaine, périodicité mensuelle de moins d'une semaine par mois, etc. Le nombre de jours passés dans le logement en question ayant été noté par l'enquêtrice dans les cases correspondant à la question 3, il ne fait pas l'objet de codification au sens propre du terme ; il fait par contre l'objet d'une vérification, et d'une éventuelle correction, en fonction de l'information contenue dans le calendrier.

Pour ce qui concerne le volet *biographique*, le travail de codification consiste à :

- faire figurer dans la colonne Date toutes les dates correspondant à des événements répertoriés dans les différentes biographies du volet ;
- entourer dans la colonne Étape de la biographie 2 (Lieu de résidence) chacune des croix correspondant à une migration ;
- noter les codes de lieux (pays, département, commune, quartier de Bogota) ;
- dans les biographies 3 (relation de parenté) et 4 (statut d'occupation du logement) : entourer les codes des cases correspondant à la première année d'une étape ;
- dans les biographies 5 (éducation) et 6 (activité) : entourer les codes des cases correspondant à la première année et à la dernière année d'une étape, après avoir noté les codes d'activité et de branche d'activité ; si l'étape est encore en cours au moment de l'enquête, noter 97 au bas de la colonne correspondante ;
- dans les colonnes de la biographie 7 (famille et corésidence) : entourer les codes correspondant aux événements ; si l'étape est encore en cours au moment de l'enquête, noter 97 au bas de la colonne correspondante.

Cinq semaines de travail ont été nécessaires à ces deux personnes pour mener à bien l'étape de critique et codification, la majeure partie de leur temps étant consacrée à la critique ; cette étape a fait l'objet d'un contrôle quotidien par les chercheurs⁽²¹⁾.

4.2. Fichiers de base de l'enquête

La *structuration des données* a cherché à satisfaire deux objectifs : d'une part, permettre la saisie directe des données à partir du questionnaire d'enquête,

(21) Pour l'enquête dans les villes du Casanare, quatre personnes, ayant toutes participé à la collecte, ont réalisé la codification des 2 000 questionnaires ; la durée moyenne de critique/codification d'un questionnaire a été plus longue qu'à Bogota, du fait de la plus grande complexité des situations et des trajectoires rencontrées dans ces villes.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

• *Enquête statistique principale et auprès des sans-logis*

Pendant les opérations de collecte, une première série de vérifications des questionnaires a été effectuée au jour le jour sur place : première vérification par les superviseurs, deuxième par la chercheuse responsable du projet assistée d'une allocataire. Une fois la collecte achevée, l'analyse critique des questionnaires et leur codification ont été réalisées par une équipe de cinq membres, comprenant les deux superviseurs et trois des enquêteurs ayant participé à la collecte, donc tous déjà familiarisés avec les données du questionnaire. Un enquêteur/superviseur codifiait les questionnaires qu'il n'avait pas lui-même remplis ou vérifiés au moment de la collecte. Chaque questionnaire codé était ensuite vérifié par un autre membre de l'équipe. Le responsable du projet a effectué ensuite une deuxième vérification de l'ensemble des questionnaires, en ciblant plus particulièrement sur certaines questions. L'étape de critique et codification a débuté un mois et demi après la fin de l'enquête statistique principale, et a duré trois mois (de mi-juin 1995 à mi-septembre 1995).

Une grande partie des questions étant déjà précodées au moment de la collecte, la majorité du temps, au cours de l'étape d'analyse critique et de codification, a été consacrée à la vérification de la cohérence de l'information collectée. Les rares questions ouvertes, nécessitant une codification ultérieure, concernaient les variables de lieu (pays, État, district, ville dans l'aire métropolitaine, localité dans l'agglomération de Delhi), la caste, les langues, les professions, la fréquence du séjour dans les autres logements.

Les listes des codes pour les castes et les langues ont été établies sur la base de la liste exhaustive des cas rencontrés au cours de l'enquête. Les codes des localités dans l'agglomération de Delhi correspondent aux coordonnées du lieu, selon un atlas de Delhi utilisé pour la codification. Pour les professions nous avons repris la nomenclature à trois degrés utilisée dans les recensements et enquêtes nationales officielles.

La variable « fréquence du séjour », dans la section « Système de résidence », est une variable synthétique; codée à partir des informations reportées graphiquement dans le calendrier prévu à cet effet ; ses modalités décrivent les principaux rythmes de séjour dans un logement.

Dans la section « Étapes migratoires », les informations du tableau reconstituant de manière chronologique la totalité des étapes résidentielles n'ont pas été codées telles quelles, mais ont servi à la création de nouvelles variables au moment de la codification. Pour l'enquête principale, deux variables supplémentaires ont été créées : le nombre d'étapes migratoires avant la dernière arrivée à Delhi, et le nombre de logements différents occupés dans Delhi depuis la dernière arrivée dans la capitale. Pour l'enquête auprès des sans-logis une troisième variable a été ajoutée : le nombre total de logements (y compris

abris de nuit et aires de couchage à l'extérieur) occupés dans Delhi avant le séjour en cours dans la capitale. Cette dernière variable permettra d'identifier les personnes qui ont vécu dans Delhi à différentes périodes et entre-temps dans d'autres villes ou villages en dehors de la capitale.

La saisie, de même que la correction des données et leur organisation en fichiers exportables sous SPSS, a été sous-traitée à un centre de calcul. Les informaticiens du centre ont travaillé à partir des manuels d'instructions détaillées pour chaque étape de la préparation des données. La saisie des données a été effectuée directement à partir des questionnaires codés, à l'aide d'un programme CLIPPER qui comprenait des contrôles des valeurs possibles sur toutes les variables du questionnaire. Une double saisie a été effectuée, suivie d'un contrôle de comparaison pour détecter et corriger les erreurs éventuelles (à l'aide de programmes écrits en DBase et en FoxPro). Une série de contrôles de cohérence entre variables d'un même chapitre ou de chapitres différents du questionnaire a été également appliquée (à l'aide de programmes écrits en DBase et en FoxPro). À partir du listing des erreurs, chaque cas a été examiné et corrigé en retournant aux questionnaires : cette étape du travail a été effectuée par une équipe restreinte de 3 enquêteurs/superviseurs, suivie directement par le responsable du projet.

• *Deuxième passage à Noida*

Au cours de la collecte les questionnaires ont été vérifiés au jour le jour par l'allocataire de recherche en charge de cette enquête. Pour la codification du volet biographique, qui constitue la partie spécifique du questionnaire de cette enquête par rapport à celui de l'enquête statistique principale, ont été suivis les principes adoptés pour la partie biographique des enquêtes sur les mobilités spatiales à Bogota et dans trois villes du Casanare, et dont les questionnaires ont servi de modèle pour la collecte biographique à Noida⁽⁷⁾.

4.2. Fichiers de base de l'enquête

• *Enquête statistique principale*

Les données de l'enquête statistique principale sont organisées sous forme de 3 fichiers composés d'enregistrements de taille fixe, avec un identifiant ménage commun aux trois fichiers :

- 1 fichier au niveau Ménage (variables d'identification et caractéristiques du logement) ;
- 1 fichier au niveau Individu pour les membres du ménage (variables démographiques et socio-économiques et variables sur les mobilités quotidiennes et résidentielles) ;
- 1 fichier au niveau Individu pour les membres de la famille ne faisant pas partie du ménage enquêté (dernière partie du questionnaire).

⁽⁷⁾ Cf. Dureau et Florez, ce volume.

Ce problème de mémoire a été particulièrement important dans la collecte des dates précises des événements. Pour résoudre d'avance le problème de concomitance de certains événements, on demandait la date précise de chaque événement, en année et en mois. Cependant cette exigence de précision a été fort difficile à satisfaire pour de nombreux enquêtés. De plus, en mettant l'accent sur la date précise des événements et non sur leur séquence, on a perdu de l'information. Plus que le mois précis de l'événement, c'est la séquence des différents événements qui est nécessaire pour connaître la relation causale ; or cette séquence est sûrement plus facile à mémoriser que des dates précises. Cette mauvaise (ou insuffisante) récupération de la séquence des différents événements est la principale faiblesse du questionnaire, qui a donc été revu sur cet aspect.

Une autre difficulté de la collecte, directement liée à la conception du questionnaire, réside dans la juxtaposition de deux biographies (annuelle et mensuelle). L'information collectée dans la seconde semble de moins bonne qualité, par lassitude des enquêtés, et aussi parce que la différence avec la première histoire de vie n'a pas toujours été bien perçue, par les enquêtés ou même par les enquêtrices. Dans beaucoup de cas, la biographie des cinq dernières années n'est apparue que comme une répétition de la première, alors qu'il s'agissait de rechercher les événements de moindre durée, les situations de résidence ou de travail irrégulières, instables ou intermittentes. Cette confusion découle, assurément, de la faiblesse de la formation des enquêtrices. De plus, la juxtaposition des deux types de biographies affecte l'ensemble du questionnaire qui apparaît particulièrement long, n'incitant pas les enquêtrices à consacrer beaucoup de temps à chaque question.

En conclusion et à propos des deux histoires de vie (annuelle et mensuelle), la faiblesse du nombre d'événements courts et multiples que connaissent de petites sous-populations durant une même année, nous a incités à supprimer le questionnaire mensuel dans la nouvelle version du questionnaire (national). La multiplicité de résidence dans une même année de vie représente moins de 0,5 % des cas ! C'est donc négligeable, même à Tijuana.

4. TRAITEMENT DE L'INFORMATION

4.1. Traitement du questionnaire, principes et difficultés de codification

Le questionnaire ménage prévoit la codification de l'information au moment de la collecte.

Pour le questionnaire biographique, après l'entretien, le questionnaire est relu une première fois par la responsable de la supervision (appartenant à l'équipe de chercheurs) en présence de l'enquêtrice. En cas de contradiction dans l'information collectée ou de manque de clarté ou d'insuffisance, l'enquêtrice doit retourner au logement sélectionné pour corriger ou compléter l'information. Le questionnaire est ensuite revu par la responsable de supervision.

La codification a été faite par des employées de l'Inegi ayant une expérience en codification d'enquêtes (de fait, deux des enquêtrices). La supervision

de la codification a été réalisée par la responsable de la supervision du travail de terrain de l'équipe des chercheurs.

Le questionnaire biographique comprend deux formes de codification. La première pour les questions ouvertes, très peu nombreuses, la majorité des questions étant précodifiées. Les questions ouvertes concernent les lieux de résidence (État, municipale, localité), les emplois et les branches d'activité. La question sur les situations spéciales a été étudiée séparément et ne figure pas dans la base de données.

La seconde codification a consisté à remplir toutes les années vides, sans événement, par exemple les premières années avant la scolarisation, pour le thème « Scolarité », les années sans travail, pour « Activité économique », etc. Pendant l'enquête, les seules années codifiées ont été celles où un événement s'était produit et aussi celles qui marquaient un relevé de durée, lorsque le questionnaire, par un trait vertical, marquait le maintien d'un état donné. Après l'enquête, toutes les années sans événement ont été codées par un zéro (0).

Plutôt que de coder les mois des événements, c'est seulement l'ordre de séquence qui a été retenu lorsque deux types d'événements différents avaient lieu la même année (mariage et nouvel emploi, déménagement et naissance d'un enfant,...). De nouvelles variables ont donc été créées, appelées ORDEN, qui apparaissent pour chaque thème de question/champ d'événement : une variable d'ordre pour les lieux de résidence, une pour l'activité économique, une pour chaque conjoint et une pour chaque enfant. En l'absence d'événements concomitants la même année, toutes ces variables sont codées 0.

C'est lors de la codification que certains points ont été précisés. Il n'était pas toujours clair dans l'énoncé du questionnaire si, pour le mariage par exemple, l'événement était seulement daté (codifié l'année du mariage) ou aussi associé à une durée (toutes les années pendant lesquelles la personne reste mariée). Pour la future enquête nationale, il est prévu des codes différents pour les événements datés et ceux associés à une durée (par exemple, U (Union) pour l'année du mariage et EU pour les années de vie maritale).

La codification des emplois a dû être modifiée. Certains changements de situation auraient été éliminés, par exemple un changement d'emploi dans la même catégorie, selon la méthode de codification prévue. Il a fallu insérer un nombre supplémentaire au code des emplois pour décrire l'ordre des emplois dans la vie professionnelle.

En ce qui concerne l'histoire scolaire, un nouveau code a été prévu pour les années qui suivent les périodes de scolarité : le premier chiffre définit l'état « ne plus suivre d'études » et le second définit le niveau d'études obtenu.

4.2. Fichiers de base de l'enquête

La saisie de l'information du questionnaire de ménage a été réalisée avec Data Entry de SPSS, alors que pour la saisie du questionnaire biographique, on a utilisé le programme EXCEL étant donné le faible nombre d'histoires de vie à saisir. Cependant, la saisie des histoires de vie a été très laborieuse à superviser. Il est évident qu'il faut un programme spécial de saisie du questionnaire