

L'UTILISATION DE MICRO-ORDINATEURS DE TERRAIN POUR LA COLLECTE EN DEMOGRAPHIE

Bote 2

Benoît FERRY et Pierre CANTRELLE

ORSTOM

Unité de Recherche Population-Santé
75006 Paris, France

1 - INTRODUCTION

Depuis longtemps les recherches méthodologiques en démographie se sont essentiellement intéressées aux phases finales du processus d'enquête. Elles ont systématiquement privilégié l'analyse, y compris à partir de données peu fiables, et n'ont accordé que très peu d'intérêt à l'étape initiale, à savoir la collecte.

Elles ont souvent privilégié l'amélioration des erreurs de sondage plutôt que les erreurs d'observation, l'analyse poussée et l'introduction dans des modèles plutôt que la qualité des données.

Dépendant les techniques de collecte et d'exploitation se sont beaucoup développées dans les trente dernières années. Une littérature importante existe sur l'organisation des enquêtes, les théories de questionnaires, les principes de codification, les méthodes de saisie et de contrôle, de traitement informatique.

Dans le processus des enquêtes quantitatives la phase de préparation des données - contrôle en bureau, chiffrement, saisie, contrôles divers et dépouillement - peut s'avérer très longue, voire pour certaines enquêtes, décourager tout projet d'analyse des données collectées, autre qu'une analyse sommaire. Certaines enquêtes ne dépassent même pas le stade de liasse de questionnaires sans exploitation. Dans le cas contraire, ce sont au mieux plusieurs mois, plus généralement plusieurs années qui s'écoulent entre la collecte et la rédaction des résultats.

A titre d'exemple, pour les 42 enquêtes nationales de l'Enquête Mondiale Fécondité menées dans les pays en développement, le temps moyen écoulé entre la fin du terrain et la fin de la saisie a été de 6,6 mois, pour arriver à des données contrôlées et propres de 7,8 mois supplémentaires et enfin aux premières tabulations, de nouveau 6,2 mois; soit au total 20,7 mois en moyenne entre la fin du terrain et les premières tabulations. (THOMPSON 1984).

C'est pourquoi des efforts ont été faits pour rapprocher au maximum l'exploitation de la collecte sur le terrain avec les outils de saisie contrôlée sur micro-ordinateurs de type PC.

Depuis l'apparition des micro-ordinateurs portatifs (lap-top) de nouvelles directions ont été prises pour effectuer la saisie sur le terrain-même. Plus récemment certains logiciels ont été

09 NOV. 1989

1.2.15

ORSTOM Fonds Documentaire

N° : 27.095 ep1

Cote : B M

P160

développés pour une passation directe des questionnaires sur le terrain avec les micros portatifs.

Le moment semble donc venu de faire le point sur les progrès actuels et potentiels que procure la micro-informatique.

2 - LES FAIBLESSES ET LES POINTS PIVOTS DES OPERATIONS DE COLLECTE

La collecte des données sur le terrain et leur préparation pour les traitements statistiques constituent souvent la faiblesse des enquêtes. Ceci est dû parfois à l'organisation de l'opération, mais plus généralement à la sous-estimation de la difficulté des tâches à mener.

La première étape après le terrain est constituée par le contrôle des questionnaires en bureau. Le personnel a des difficultés à corriger les erreurs, à rétablir les cohérences à partir des seuls questionnaires; de plus ce contrôle est très loin d'être exhaustif. Enfin certaines erreurs sont impossibles à corriger telles que erreurs de filtre, omissions de question, pertes de pages du questionnaire.

A la codification d'autres types de problèmes peuvent intervenir en particulier les erreurs de chiffrement ou les omissions de codes.

A la saisie apparaissent les problèmes liés à l'environnement informatique de l'enquête. Il y a d'abord des problèmes dus à l'accès au matériel. Les ordinateurs requièrent la présence d'un personnel souvent spécialisé pour assurer la programmation, le fonctionnement des logiciels et des machines.

Après la saisie vient la partie de contrôle des données. Une liste des erreurs est produite au premier passage des données à travers les contrôles de structure. Une équipe de correcteurs formés sur le questionnaire et ses contrôles vont corriger les erreurs jusqu'à ce que l'on atteigne le pourcentage maximum d'erreur fixé dans le cahier des charges de l'enquête. Il arrive parfois que ce pourcentage d'erreur soit si important que tout traitement des données de l'enquête doit être abandonné. Cette phase de contrôle est répétée pour chaque type de contrôle (contrôle de structure, contrôle des valeurs minimales et maximales autorisées pour chaque variable, contrôle de cohérence entre les variables) jusqu'à la tolérance d'erreur. La résolution des incohérences finales est toujours longue et difficile. Chacune de ces parties est souvent prise en charge par un programme spécifique car il n'existe que très peu de programmes généraux de contrôle adaptés.

Globalement, ce qui pose le plus de difficultés à la réussite des opérations de collecte touche à la logistique de l'opération, aux problèmes budgétaires et de calendrier, à la supervision des enquêteurs sur le terrain et à la préparation des données pour traitement informatique. Peut-on y remédier ?

Une opération de collecte doit viser la rapidité pour rapprocher une demande formulée des résultats obtenus, la qualité et la précision des données, enfin les coûts les plus faibles possibles.

Dans ce but on devrait s'efforcer de n'avoir qu'une seule équipe de la conception aux analyses finales. Le responsable garderait ainsi la maîtrise de l'ensemble des phases. Il devrait en particulier préparer l'exploitation des la conception de l'enquête. Ceci suppose d'avoir les outils qui permettent cette maîtrise, ainsi que la possibilité d'effectuer rapidement l'ensemble d'une opération. La micro-informatique permet de répondre à ce souhait.

3 - L'AVENEMENT DE LA MICRO-INFORMATIQUE

Les applications de la micro-informatique ont porté tout d'abord sur l'analyse et les modèles; on se souvient du vibrant plaidoyer de Nathan KEYFITZ en séance plénière au Congrès de Florence pour le micro-ordinateur, outil de base du démographe. Diverses comparaisons ont été faites par KELLER (1986) et STRONG (1987)

Bien vite on s'est intéressé aux possibilités des micro-ordinateurs pour la saisie des données. Dans un premier temps ils ont permis de répliquer le travail des postes de saisie des données chiffrées sur bordereaux, puis rapidement sont apparus des logiciels plus spécifiques à la saisie de questionnaires. Une évaluation en a été présentée par DISKIN (1986).

Plus récemment des logiciels assez complets de saisie contrôlée, de corrections et de traitements statistiques ont été développés et adaptés aux enquêtes démographiques complexes; notamment le logiciel ISSA de Westinghouse (CUSHING 1987) utilisé pour toutes les enquêtes complexes DHS sur micro-ordinateurs, ou encore ISPC de l'US Bureau of Census (DISKIN 1987).

D'autres types de logiciels sont couramment utilisés pour la saisie et la préparation des données. Ils sont centrés sur des bases de données et procurent des écrans de saisie. Les logiciels de base de données sont en effet peu chers, puissants et facilement utilisables. On peut mentionner DBASE III et ses semblables dont certains logiciels spécifiques ont été développés pour les enquêtes.

Le micro-ordinateur permet un travail en temps réel, une organisation des données vraiment structurées relationnelles, la convivialité par l'accès par menu et les menus d'aide; Pour le personnel qui entre les données, l'avantage principal est qu'il retrouve à l'écran les pages du questionnaire où il peut chiffrer tout ou partie des zones préparées.

Cependant il est toujours assez difficile de paramétrer ces logiciels et cela nécessite souvent l'intervention d'un programmeur. L'introduction des sauts de question et les contrôles de cohérence sont toujours complexes. Les procédures de corrections demeurent lourdes.

4 - DECENTRALISATION DE LA SAISIE DES DONNEES

La tendance a été de rapprocher les micro-ordinateurs du lieu de collecte pour accélérer la préparation et le traitement des données et, au besoin, corriger les informations en retournant sur le terrain. Ce type d'approche de la collecte est à la base d'une série d'enquêtes menée par la Banque

Mondiale dans lesquelles le micro-ordinateur de type PC a été choisi comme pivot. (WORLD BANK 1986). Ils ont été placés près du lieu d'enquête. Le questionnaire comprenait environ 750 variables dans une structure complexe. Un gros logiciel a été développé pour adaptation immédiate à l'opération et les données ont pu être ainsi traitées dans un délai de trois mois.

L'organisation informatique a été déterminante dans le déroulement de l'enquête. Le système de saisie était proche du terrain et consistait en cinq ordinateurs. Les machines permettaient d'entrer et corriger les données pendant la suite du terrain. Les ordinateurs n'ont pas été un élément additionnel à une enquête classique mais se sont révélés comme les pivots de l'enquête autour desquels tout s'organisait.

Un des problèmes importants était la gestion des questionnaires : stockage, administration de ceux-ci et saisie sur les machines. Les ordinateurs personnels devaient aider à briser les goulots d'étranglement du codage, de la saisie, des corrections des données et du problème de l'accès à un grand site informatique.

Toutes les informations sont précodées sur le questionnaire de manière à ne faire saisir aux opérateurs que des données de type numérique. Au cours de la saisie, le label correspondant à la valeur saisie s'affiche à l'écran car le dictionnaire des variables a été préprogrammé antérieurement. La saisie est contrôlée par des vérifications des valeurs limites des variables. Une fois la saisie terminée des contrôles de cohérence entre variables sont mis en œuvre par exemple âge et date de naissance, puis des contrôles sur les sauts conditionnels sont effectués. Enfin des contrôles entre différentes parties du questionnaire complètent cette phase de vérification.

Après ce contrôle des données un listing des erreurs, omissions, incohérences est imprimé et les questionnaires qui ont posé problème sont réadministrés la semaine suivante pour corriger les erreurs sur le terrain. De cette manière les équipes ne quittent le terrain qu'après avoir corrigé les erreurs, au moins celles de la première partie du questionnaire.

6 - MISE A JOUR D'UNE BASE DE DONNEE A PARTIR D'UNE ENQUETE

D'autres progrès ont été réalisés à partir des logiciels de bases de données qui permettent d'organiser un suivi de population en mettant à jour la base de données par des enquêtes de terrain.

Une expérience de ce genre est menée actuellement au Bangladesh par l'International Center for Diarrhoeal Disease Research. (MOZUMDER 1986). Un programme spécifique de saisie autour d'une base de données a été écrit. Les enquêteurs travaillent sur le terrain avec un registre imprimé puis saisissent ensuite eux-mêmes en bureau leurs données collectées sur la base de données informatisée. Cette saisie comprend une phase de contrôle où de nombreux contrôles croisés entre variables sont mis en œuvre. Ces contrôles génèrent des messages d'erreur avec diagnostic et les données peuvent être corrigées en retournant sur le terrain. Ce processus minimise les erreurs de la collecte démographique. La phase d'édition contribue à une meilleure formation des enquêteurs en leur montrant

les interactions entre phénomènes démographiques et les sensibilise aux erreurs logiques.

D'autres opérations sur cette base de données sont possibles :

- Impression d'état ou de rapport partiel, calcul de taux démographiques.
- Interrogation de la base sur une sélection de ménages, de villages, d'une aire géographique déterminée ou sur la valeur d'une ou plusieurs variables (des opérateurs logiques et de comparaison permettent alors de relier les différentes variables entre elles)
- Extraction d'un sous-fichier pour effectuer un travail spécifique

6 - LES ENQUETES A PASSAGES REPETES GEREES PAR UN MICRO-ORDINATEUR

Les bases de données sur micro ont permis de gérer des enquêtes à passages répétés avec des intervalles relativement courts entre passages. Les questionnaires individuels peuvent être personnalisés en fonction des caractéristiques du sujet au passage précédent.

Cette approche est expérimentée par une équipe de l'ORSTOM dans une étude pluridisciplinaire sur la zone de Niakhar au Sénégal depuis mars 1983. (GARENNE 1984)

Le principe d'un bon questionnaire est que la question exacte soit clairement imprimée. Ceci est simple à réaliser dans le cas d'une enquête à un seul passage mais ce n'est pas le cas lorsque l'enquête est une étude suivie dans une région rurale d'un pays en développement. Les questions à poser dans une enquête démographique dépendent du sexe, de l'âge et dans un suivi du statut au passage précédent. Les questions qui doivent être intégrées dans le questionnaire doivent tenir compte de ces contraintes. Les micro-ordinateurs par contre permettent la préparation des questionnaires pour le passage suivant, personnalisé en fonction des réponses du passage précédent.

Dans l'enquête du Sénégal le micro-ordinateur imprime le statut de la personne au passage précédent ainsi que les questions à poser à la prochaine visite. A côté du nom de l'individu, de son numéro, son sexe, sa date de naissance, son lien de parenté au sein du ménage, son groupe ethnique et sa religion, l'ordinateur imprime les questions sous forme simplifiée. Il y a une ligne par question et il peut y avoir plusieurs lignes pour le même individu, l'enquêteur devant répondre à chaque question.

L'avantage de ce système est qu'il fournit dans le questionnaire à administrer des informations saisies aux passages précédents et qu'il permet donc de confirmer ces données et de saisir de nouvelles informations cohérentes avec les données précédentes.

Ce système est très efficace pour obtenir des données de qualité dans les enquêtes suivies, la seule contrainte étant de bien gérer dans l'espace les questionnaires individuels, chaque questionnaire correspondant à un individu précis.

L'enquête a montré l'efficacité de ce questionnaire informatise sur l'enregistrement des décès précoces. Deux visites rapprochées sur le terrain furent faites avec des méthodes différentes. Un passage avec des questions classiques sur les événements survenus depuis le dernier passage et un autre passage avec des questions exactes sur le dernier-né, possibles grâce au questionnaire informatisé. La deuxième méthode avec les questionnaires personnalisés par l'informatique a permis un gain considérable dans la qualité des données, ici de fortes sous-estimations dues aux omissions furent évitées.

7 - ENQUETES SUR DES MICRO-ORDINATEURS PORTATIFS SUR LE TERRAIN

Le développement actuel de la micro-informatique avec l'apparition d'ordinateurs de plus en plus puissants et d'encombrement réduit permet une meilleure utilisation de ceux-ci encore plus en amont dans le processus de l'enquête.

En 1983 la révolution du microprocesseur faisait rêver Griffith FEENEY de futurs systèmes d'interrogation ou de conduite d'enquêtes informatisées (FEENEY 1983). L'apparition des premiers micro-ordinateurs portatifs a permis à d'autres de réaliser ses souhaits.

Les micro-ordinateurs sont devenus portables, légers, fiables et ergonomiques. Ils peuvent supporter le programme d'une ou plusieurs enquêtes et stocker de manière sûre les données de plusieurs journées de terrain. Ces machines ont aussi des possibilités de transmission et doivent avoir un certain degré de compatibilité avec les machines qui traiteront les données : code des données ASCII ou EBCDIC, systèmes d'exploitation identiques, langage commun ou système de liaison.

Depuis quelques années sont apparues ce type de machine d'abord sous système d'exploitation CP/M, puis plus récemment MS-DOS.

On a vu émerger certaines applications qui, à partir d'un logiciel écrit spécialement, généralement en Basic, permettent de saisir directement des informations sur une machine de terrain. Ces premières applications très dédiées ont concerné principalement la collecte d'informations simples.

Il s'agissait souvent en fait d'effectuer la codification-saisie sur le terrain ou d'apporter une aide à l'enquêteur sur les nomenclatures. Citons les expériences de certains instituts de statistiques européens, France, Suède et Pays-Bas notamment. Les opérations menées jusqu'à présent ne l'ont été qu'à titre expérimental, ponctuel et à partir de logiciels spécifiques à chaque application. Elles n'ont fait l'objet que de quelques publications descriptives.

Ainsi l'objectif du Netherlands Central Bureau Of Statistics était de tester la faisabilité d'enquêtes sur ces machines. (BEMELMANS-SPOK 1985, CBS 1987). Il désirait savoir si cette technologie était bien reçue par les enquêteurs et les enquêtés. Un des buts était aussi d'estimer les possibilités et l'efficacité de la programmation d'enquêtes sur ces machines portatives.

Les expérimentateurs étaient conscients des avantages que l'on peut tirer de cette technologie. Les enquêtes devaient

en particulier montrer l'efficacité de la saisie avec contrôle des erreurs et le cheminement guidé dans le questionnaire suivant les valeurs des réponses. Face à ces avantages, une évaluation des coûts d'achat des machines, des coûts et des difficultés de programmation, de la qualité du travail de l'enquêteur et de sa relation avec l'enquêté était nécessaire.

Plusieurs enquêtes expérimentales ont été menées. Les matériels choisis étaient des micro-ordinateurs NEC et la programmation des enquêtes fut faite en Basic. Deux enquêtes servirent de support à ces tests: une enquête de relevé de prix et une enquête de consommation relativement simple. Un protocole de comparaison statistique entre le questionnaire papier et l'enquête avec l'ordinateur fut établi. Les machines furent très bien reçues par les interviewés car aucun refus n'a été justifié par la technologie employée. Les enquêteurs ont marqué une nette préférence pour l'enquête menée avec les machines car celles-ci contrôlaient les réponses et les guidaient dans le questionnaire en gérant les sauts de questions.

Ces expériences ont démontré pour les hollandais la faisabilité d'enquêtes avec des micro-ordinateurs portatifs. Les avantages sont multiples (disparition du papier, suppression du codage de la saisie et de la vérification, transmissions rapides, temps réduit, meilleure qualité des données...) mais l'application de cette technologie, proche de la saisie classique, est limitée aux enquêtes simples. Au moment de ce test, les capacités mémoire des machines, les difficultés et les coûts de programmation observés pour une enquête simple ne permettaient d'envisager l'utilisation de cette technologie avec un programme dédié pour des enquêtes complexes.

Diverses autres institutions ont mené ou mènent actuellement des enquêtes simples programmées généralement en Basic sur des micro-ordinateurs portatifs.

D'autres utilisations des micro-ordinateurs portatifs consistent à transférer simplement un logiciel de PC de bureau sur le terrain. L'opération similaire à celle du bureau fait défiler pour l'enquêteur des écrans et à saisir des codes. C'est le cas de bon nombre de logiciels de traitement de données mentionnés plus haut, dont on déporte le module d'entrée de données sur les micro-portatifs de terrain.

On peut mentionner en particulier l'expérimentation du logiciel ISSA sur le terrain au Guatemala (OCHOA 1988) qui reproduit le questionnaire papier sur l'écran du micro portatif avec la possibilité pour l'enquêteur de coder au fur et à mesure et d'être contrôlé sommairement au niveau des champs.

De même la Banque Mondiale dans une expérimentation sur l'enquête niveau de vie au Ghana a transporté son module de saisie sur des micro-portatifs.

D'autres utilisent sur des portatifs des logiciels centrés sur les bases de données du type DBASE III et ses semblables qui permettent de coder des informations dans des écrans en mettant à jour une base de donnée.

Pour ces approches, on retrouve les mêmes limites que sur le matériel de saisie en bureau. La situation vient cependant sensiblement se compliquer du fait que l'on est sur le terrain

et que les enquêteurs ne disposent ni des mêmes facilités d'environnement ni d'outils spécifiques à l'interview.

Parmi les problèmes rencontrés sur le terrain avec ces logiciels on peut mentionner la nécessité de coder, des sauts de question souvent mal mis en oeuvre, la difficulté de remonter en arrière dans le questionnaire, de se reprendre en cas d'erreur ou d'incohérences combinées.

De plus ces logiciels nécessitent généralement une préparation importante par un spécialiste pour leur mise en oeuvre et leur ajustement à une enquête spécifique.

Dans ces approches les contrôles de cohérence ne peuvent pas être introduits facilement au niveau des enquêteurs. Ils sont généralement entrepris après le terrain en bureau sur une autre machine avec impression d'un listing d'erreur et retour aux questionnaires ou au terrain quand cela est possible.

8 - LES PREMIERES APPROCHES D'INTEGRATION INFORMATIQUE DE LA COLLECTE ET LEURS CONTRAINTES

Lors du Symposium de clôture de l'Enquête Mondiale Fécondité en Avril 1984, les chercheurs associés au projet, tirant les leçons des opérations effectuées et s'appuyant sur les nouvelles possibilités de l'informatique, ont été les premiers à aborder le sujet de façon intégrée. Ils ont montré, dans un bref document (THOMPSON 84), comment l'informatique pourrait, sous réserve de la disponibilité de logiciels correspondants, améliorer considérablement l'efficacité, la rapidité et la qualité des tâches à chacune des phases des enquêtes et finalement aboutir à de nouvelles stratégies en matière de collecte de données.

Il ne s'agirait plus de collecter les données sur un questionnaire papier, mais sur un micro-ordinateur de terrain portable. La qualité de la collecte peut être sensiblement améliorée en incluant des contrôles de cohérence au moment de l'interview sur ce micro-ordinateur.

On conçoit facilement les nombreux avantages que l'on peut tirer d'enquêtes menées avec des micro-ordinateurs portatifs:

- Suppression du codage, de la saisie et parfois de la vérification des questionnaires si on contrôle la saisie sur la machine.
- Contrôle immédiat des erreurs sur le terrain.
- L'enquêteur est guidé dans son cheminement dans le questionnaire.
- Limitation des erreurs de transcription car il n'y a plus qu'une seule saisie des données.
- Rapidité des transmissions et de la communication des résultats et donc de leur exploitation.
- Transmission de l'intégralité des données sans perte.
- Stockage des données brutes pour exploitation ultérieure.

La programmation en langage structuré

Les questionnaires complexes comme les questionnaires démographiques sont hiérarchisés : un niveau général d'identification du ménage, puis un niveau individuel ou

un groupe de questions est posé autant de fois qu'il y a d'individus puis d'autres groupes de questions à poser à chaque individu ou à tout le ménage. Le Basic ne peut manipuler que très difficilement de telles structures car il n'est pas structuré. Par contre les langages structurés modernes comme le Pascal, le C, le Modula 2, l'Ada et d'autres encore incitent et encouragent la modularité et la structuration des programmes correspondant à des structures de données complexes.

On pourrait donc envisager de programmer spécifiquement des opérations de collecte dans des langages structurés. Cependant un certain nombre d'inconvénients subsistent:

- Ils n'ont pas tous l'universalité
- Chaque langage structuré requiert la présence d'un compilateur spécifique sur la machine de terrain
- Les langages structurés nécessitent un apprentissage ou une certaine expertise en informatique de la part des programmeurs. Les modifications du programme de l'enquête ne peuvent se faire que par l'équipe de développement.

Mais un certain nombre d'avantages apparaissent :

- Leur structure modulaire permet une évolution facile des programmes. On peut envisager de réutiliser des modules du programme d'une enquête pour en concevoir une autre.
- Ces langages sont le plus souvent compilés, donc il n'y a pas de danger de détérioration des programmes de plus, la plupart des erreurs sont détectées en compilation donc avant le terrain.
- Le danger de perte de données est beaucoup plus faible.
- Le code compilé est homogène et transparent pour l'utilisateur et n'a pas besoin d'interpréteur résident sur l'ordinateur portatif.
- L'exécution de programmes compilés est infiniment plus rapide que celle de programmes interprétés donc les temps de réponse sont plus courts.

Un langage de programmation de questionnaire : BLAISE

Cette conception de la collecte informatisée n'élimine pas le problème de la programmation. Les langages structurés ont cependant inspiré les concepteurs d'enquête de manière beaucoup plus fructueuse. On peut en effet comparer la structure d'un questionnaire à celle d'un langage de programmation moderne.

Ainsi les chercheurs du Netherlands Central Bureau of Statistics ont conçu un langage de conception de questionnaire, le Blaise (DENTENEER 1986, BETHLEHEM 1987, CBS 1987)

Ce langage, très inspiré du langage Pascal, décrit les questions, les réponses possibles, les cheminements dans le questionnaire, les contrôles des valeurs maximales et minimales et les contrôles d'incohérence :

- Chaque question est décrite par le label de la question, le type de la réponse (alphanumérique, numérique, modalité) et les réponses possibles dans un bloc ressemblant fort à une procédure Pascal (bloc Quest).
- Le cheminement dans le questionnaire et les sauts de questions sont décrits dans d'autres blocs (blocs Route).
- Les contrôles de cohérence entre variables sont aussi regroupés dans des blocs (blocs Check).

- Les questions peuvent de même être regroupées dans des blocs suivant les thèmes du questionnaire. Ces groupes de questions contiennent la description de celles-ci, les sauts de questions et les contrôles de cohérence. Le questionnaire peut ainsi être fortement structuré pour correspondre à une enquête complexe. Les concepts de blocs et de tableaux de questions empruntés au Pascal permettent de concevoir des enquêtes hiérarchisées comme les questionnaires ménages.

Le système Blaise est composé de plusieurs groupes de programmes. Le programme décrivant le questionnaire est édité avec un logiciel de traitement de texte classique puis ce texte passe dans un logiciel d'analyse syntaxique et est compilé pour produire un programme exécutable.

- Le logiciel d'analyse syntaxique Blaise lit le questionnaire et vérifie la syntaxe. Si la syntaxe est correcte il produit un programme Pascal. Il y a un logiciel d'analyse syntaxique spécifique pour chaque machine virtuelle.

- Le compilateur Blaise génère un programme exécutable en Pascal. Le programme Pascal produit par le logiciel d'analyse Blaise est combiné avec plusieurs procédures standard Blaise. Ensuite le compilateur Turbo-Pascal est utilisé pour générer l'exécutable. Il y a différents groupes de procédures Blaise pour chaque machine virtuelle.

- Les interfaces : Les données produites sont propres, directement utilisables et peuvent être converties dans des formats compatibles avec des logiciels de traitement statistiques (SPSS par exemple) ou des bases de données (DbaseIII).

Ici la programmation dédiée est abandonnée au profit d'un langage spécifique aux questionnaires. On travaille en terme de génération et un questionnaire écrit en Blaise peut être la source de plusieurs produits finaux différents : enquête papier, téléphone... De plus le processus de l'enquête est conçu comme un enchaînement logique d'opérations générées autour du questionnaire.

Ce langage Blaise étant très proche du langage de programmation Pascal comporte les avantages et les inconvénients de ce dernier.

Bien que conçu dans une approche globale de l'informatisation des enquêtes le système Blaise souffre de l'ésotérisme de son langage de définition de questionnaire qui ramène aux travers de la programmation spécifique d'une enquête.

9 - LE SYSTEME INTEGRE DE COLLECTE

Les acquis décrits précédemment ont finalement abouti à une approche nouvelle, celle du nouveau logiciel SIC, développé initialement par l'ORSTOM :

Il s'appuie sur un savoir faire en matière d'enquête et non à partir des grilles de saisie adaptées.

Les différents modules de ce logiciel correspondant d'une part aux tâches du responsable d'une enquête qui va concevoir et mettre au point son questionnaire, le tester et analyser ses données dès que le terrain sera terminé, et d'autre part au

travail d'enquêteurs sur le terrain qui ne s'occuperont que de l'interview.

Le logiciel SIC a été conçu de telle sorte que du personnel non informaticien puisse utiliser l'ensemble des fonctionnalités en rendant transparent les opérations techniques et les ordres strictements informatiques.

Il est écrit de façon modulaire en vue d'améliorations et de fonctionnalités complémentaires.

Il vise les enquêtes par interviews, tous types de questionnaires, y compris les plus complexes, mais peut s'appliquer également à des saisies de données et de documents sur le terrain; il assume également les enquêtes à passages répétés et les suivis de population, tant dans les pays développés que les pays en développement.

Les fonctions du logiciel SIC

Le logiciel SIC est une chaîne de logiciels intégrés permettant d'assurer trois opérations :

- La génération d'un questionnaire ou d'un formulaire sur le micro-ordinateur de type PC grâce à un éditeur interactif
- Le questionnement guidé et contrôlé sur le terrain.
- Après téléchargement des données provenant du terrain, décompactage des données codées et propre dans un format lisible par le plupart des logiciels de traitements statistiques.

Le module générateur

L'idée globale est de donner des outils au concepteur pour concevoir l'ensemble des éléments d'un questionnaire de façon éventuellement désordonnée en contrôlant systématiquement la pertinence et la syntaxe des informations rentrées. Pour cela il dispose d'un ensemble de menus et fenêtres qui lui permet d'effectuer les grandes opérations de structure de son questionnaire : définition d'une question, d'un sous-questionnaire. Correspondant à ces actions des sous-fenêtres spécifiques aux paramètres de définition des questions et des sous-questionnaires peuvent être remplies partiellement ou totalement selon la façon dont travaille le concepteur. A chacun de ces champs sont associés un certain nombre de contrôles qui assurent la qualité des informations rentrées.

Générant progressivement son questionnaire il est à tout moment possible de consulter les informations déjà rentrées par le système de fenêtrage multiple, mais également de modifier tout ou partie de ce questionnaire par l'accès aux fenêtres de paramètres et la possibilité d'insérer ou de supprimer des questions ou des sous-questionnaires. Un listing rassemblant sous une forme claire ces paramètres peut également être généré à tout moment.

A partir de là, il est possible d'utiliser le module interview soit sur la machine centrale pour que le concepteur puisse simuler le travail de terrain, soit, en téléchargeant ce fichier de paramètres, sur une ou plusieurs machines de terrain.

Ce module permet également de prévoir des enquêtes à plusieurs passages ou les réponses d'un passage précédent peuvent être reaffichées et servir de filtre ou d'élément de contrôle de cohérence pour des questions du nouveau passage.

SIC permet un nombre théoriquement illimité de questions et de sous-questionnaires simples ou multiples; Il offre la possibilité de 6 niveaux de hiérarchie.

Le module Interview

C'est l'outil de l'enquêteur. Il est résident sur la machine de terrain et fonctionne en relation avec l'ensemble des paramètres décrivant le questionnaire. Il fournit également un certain nombre d'outils spécifiques au travail de l'enquêteur et inhibe toutes les autres fonctions informatiques du micro de terrain. Il permet le stockage dans un format compact des données collectées et la possibilité de réinterroger ou de mener de nouveaux questionnaires sur les mêmes individus; c'est le cas en particulier des enquêtes à passage répétés. Les enquêteurs peuvent avoir une autonomie de longue durée, de plus d'une semaine, sur le terrain; et si l'on préfère possibilité de décharger les données à tout moment.

L'utilisation de ce module est très simple et facile d'accès, en regard de la faible qualification de ses utilisateurs. Seules les questions pertinentes apparaissent successivement en clair à l'écran, et l'enquêteur consigne les réponses de façon totalement explicite; il est de plus informé en permanence par des messages éventuels des instructions, des incohérences et de ce qu'il peut faire.

Le module concentration et traitement des résultats

Les données collectées sur une ou plusieurs machines de terrain peuvent être remontées sur la machine centrale où elles sont successivement rassemblées dans un seul fichier puis décompactées afin d'obtenir un fichier dans un format normalisé (ASCII). Parallèlement on récupère un fichier dictionnaire de variables issu de la génération. Ces deux fichiers permettent alors de s'interfacer immédiatement et facilement avec les logiciels classiques de traitement statistique des enquêtes (SAS, SPSS, DBASE ...).

Le logiciel SIC, écrit en Turbo-Pascal et en Turbo-C, fonctionne sous DOS 2.1 et plus, sur micro-ordinateurs PC et compatibles (XT, AT ou 386 avec ou sans disque dur) pour la machine centrale, et sur portatif (MS-DOS) compatible PC pour la machine de terrain (TOSHIBA T1000, EPSON PX16, GAID...).

SIC a également fonctionné sur des machines CPM (HUSKY, EPSON PX4...); mais le renforcement des fonctionnalités ne permet pas pour l'avenir de conserver cette possibilité.

Il est à présent développé et commercialisé sous le nom de QUESTOR par la Société AGERIS, sous licence ORSTOM.

10 - APPLICATIONS DU SIC

Plusieurs séries d'expérimentations (LEAHY 1986) ont d'abord été menées dans les centres d'état-civil de Dakar en collaboration avec la Direction de la Statistique du Sénégal, pour la saisie des registres.

Ensuite dans une enquête simple au Sénégal avec le personnel enquêteur de l'ORSTOM dans la banlieue de Dakar. Des interviews ont été menés dans des concessions, d'une part à partir d'un questionnaire papier et d'autre part sur les micro-portatifs à des fins de comparaisons. Plus récemment en Guinée à Conakry, plusieurs tests ont été menés avec le personnel du Service de la Statistique.

Plusieurs enquêtes en vraie grandeur ont été également entreprises dans la région parisienne; en particulier à la Cité des Sciences et de l'Industrie de La Villette auprès des usagers de la médiathèque. Pour cette opération un questionnaire d'une quarantaine de questions a été généré un matin; la formation des enquêtrices a eu lieu en début d'après-midi; le questionnaire a ensuite été modifié à la suite de leurs remarques, et l'enquête a commencé le même jour en fin d'après-midi. L'enquête a porté sur 309 questionnaires; déchargement des données le soir du dernier jour, et sortie des tableaux le lendemain.

L'INSEE a testé et évalué lui-même (INSEE 1988) ce logiciel dans un sous-échantillon de l'enquête de conjoncture auprès des ménages avec 10 enquêteurs munis chacun d'un micro-portatif. Cette enquête ménage comportait environ 235 questions dans une hiérarchie complexe. 250 ménages ont été interrogés.

D'autres opérations de plus grande envergure sont actuellement en préparation. Mentionnons en particulier la réplique de l'enquête niveau de vie auprès des ménages (LSMS) en Côte d'Ivoire dans un certain nombre de grappes, en collaboration avec la Banque Mondiale, pour la préparation des méthodologies des enquêtes ménage dans le cadre du programme Dimension Sociale de l'Ajustement.

En Guinée est entreprise au cours de l'été 1988 une étude de faisabilité d'observation suivie, dont l'outil central est le SIC, en vue de son application ultérieure éventuelle au niveau national dans le cadre du Système Guinéen de Données sur la Population.

11 - INTERET D'UN LOGICIEL SPECIFIQUE A LA COLLECTE SUR LE TERRAIN

Certaines questions préalables semblaient encore en suspens (CUSHING 1987) quant à l'exécution d'enquête sur micro-ordinateurs portatifs sur le terrain.

En particulier l'utilisation de PC portatifs était-elle possible avec du personnel non formé en informatique, à plus forte raison dans les PED? L'expérience a montré sa facilité d'utilisation par des enquêteurs, sans aucune formation en informatique, notamment les enquêtes au Sénégal et en Guinée. Pour l'enquêteur, c'est simplement une machine d'enquête; et elle a pour avantage de faciliter son travail et de valoriser sa fonction.

Par ailleurs l'introduction d'une machine dans une enquête allait-elle troubler la relation enquêteur-enquêté ? Les expériences déjà mentionnées ont montré qu'il n'en était rien. Mais dans les pays où l'informatique est reconnue et peut susciter une défiance, en particulier dans les enquêtes à domicile ? Les enquêtes en France viennent confirmer les expériences étrangères, notamment en Hollande, en dissipant cette crainte.

Des développements entrepris pour réaliser le Système Intègre de Collecte et des différentes enquêtes menées tant en France qu'en Afrique, se dégagent un bilan favorable à l'utilisation de logiciels spécifiques à la collecte de données qui intègrent création du questionnaire, passation sur le terrain et préparation des données.

Simplicité

- Plus besoin de programmer, il suffit d'une approche logique de l'enquête.
- Supprime toute la chaîne de gestion des documents.
- L'enquêteur est complètement guidé dans sa saisie; il a moins de questions à se poser; il est guidé pas à pas sur les questions pertinentes; les sauts de questions ou de groupes de questions se font automatiquement.
- Possibilité à tout moment de remonter dans le questionnaire pour consulter ou modifier.
- Les questionnaires papier disparaissent, la codification, la saisie et le nettoyage des données également; les micro portatifs assurent ces tâches sur le terrain.

Rapidité

- Réduction des délais par suppression du temps de saisie des bordereaux, de retour aux documents de base pour correction éventuelle après contrôle.
- La mise au point des tabulations peut se faire dès avant l'opération, être testée, et exécutée dès le retour des premiers enregistrements.
- On obtient sur le terrain même des données contrôlées et propres, exploitables dans les heures qui suivent la fin du terrain.
- C'est le plus court chemin entre la conception de l'étude et ses résultats et donc une nouvelle dynamique pour les décideurs.

Efficacité et qualité

- La maîtrise, sans aide d'informaticiens, par une seule équipe de l'opération dans son ensemble.
- Amélioration de la qualité par le contrôle direct à la saisie, et le fait de n'avoir plus d'étape de transcription et de contrôle intermédiaire, source d'erreurs. Dès le terrain les contrôles de cohérence peuvent être complexes, l'enquêteur possède les outils de reprise.
- Possibilité de simuler le travail de l'enquêteur, mais aussi de superviser son travail par des outils efficaces.
- Interface automatique avec plusieurs logiciels de traitement statistique et de base de données.
- Par la possibilité de traitement immédiat, il est possible de fournir des résultats sur le terrain même (feed-back).

Formation

- Aucune qualification informatique n'est nécessaire.
- Le concepteur apprend en une journée à générer un questionnaire.
- L'enquêteur apprend en quelques heures à passer ses questionnaires avec le micro portable.

Le coût total d'une opération est finalement moindre :

- en moins, suppression des coûts de programmation, de codification, de saisie, de correction des données, de retours éventuels sur le terrain.
- en plus, le matériel, dont les prix tendent à la baisse, est amorti rapidement.

12 - CONCLUSIONS

On a vu la rapide progression réalisée au cours des dernières années grâce à la micro-informatique aboutissant au matériel de saisie portable et aux logiciels de collecte intégrée.

Les matériels continuent à évoluer rapidement sur un plan technique. Parallèlement les logiciels sont appelés à de nouveaux développements.

Les progrès, se traduisent par une amélioration de l'équation qualité-rapidité-coût, vont transformer la collecte des données. Les rapports entre utilisateurs et producteurs vont créer une nouvelle dynamique pour les décideurs. Des stratégies de collecte plus pertinentes vont pouvoir être mises en place.

Sans être esclave de l'outil, les démographes vont se libérer d'une série de contraintes qui jusqu'à présent pesait sur la collecte des données, et par conséquent la demande, à plus forte raison dans les pays où les données manquent le plus, comme en Afrique.

REFERENCES

- BENELMANS-SPORK M.E.J., SIKKEL D. (1985). Data collection with hand-held computers. Department of Statistical Methods. Netherlands Central Bureau of Statistics, Voorburg.
- BETHLEHEM J.G., DENTENEER D., HUNDERPODL A.J. KELLER W.J. (1987). The Blaise system for computer-assisted survey processing. Department of Statistical Methods. Netherlands Central Bureau of Statistics, Voorburg.
- C.B.S. (1987). Automation in Survey Processing. CBS Select N°4. Netherlands Central Bureau of Statistics. 144 p. Voorburg.
- CUSHING J., CANTOR D., STURGIS R. (1986). Microcomputer use with large scale surveys in developing countries. Paper presented at the Population Association of America Annual Meeting, San Francisco.
- CUSHING J. (1987). Processing the Demographic and Health Surveys : The Micro Approach. Communication présentée à la Conférence de l'IIIS à Tokyo.
- DENTENEER D. et KELLER W.J. (1986) Blaise, a language for questionnaire specification. Compstat - Rome.
- DISKIN B. (1986) An evaluation of micro-computer data-entry software package for use in developing countries for census and survey processing. US Bureau of Census.
- DISKIN B. (1987) Integrated microcomputer processing system. United States Bureau of Census, Washington.
- FEENEY G. (1983). A microprocessor revolution in data collection ?. Asian and Pacific Census Forum - East-West Population Institute - East-West Center - Voi 10 N°1, Honolulu.
- FERRY B. et BERGES J.C. (1985). Système intégré de collecte - Description technique et phases du projet. DRSTOM UR Population-Santé. Note d'information n°34, Paris.

- FERRY B., CANTRELLE P., BERGES J.C. COUILLET F. (1986). La saisie et le traitement d'informations statistiques en milieu sévère à partir de micro-ordinateurs portatifs. Séminaire CIE-INSEEM-ORSTOM-INEL. Estimation de la mortalité du jeune enfant (0-5 ans) pour guider les actions de santé dans les pays en développement. INSEEM Vol 145, pp 277-302.
- GARENNE M. (1984). Les concepts de l'analyse longitudinale et ses implications pour la collecte des données : exemple de l'utilisation de questionnaires informatisés pour améliorer l'enregistrement des décès précoces au Sénégal en milieu rural (Niakhar). Communication au séminaire de l'Institut du Sanej sur les enquêtes EMIS. Bamako.
- INSEE. (1988). Test de saisie portable. Enquête de Conjoncture auprès des ménages. Rapport d'évaluation par J.P. BOCQUET. INSEE-Direction Régionale de Rouen. 13p + 12p + annexes.
- KELLER W.J. (1986). Statistical Software for Personal Computer. COMPSTAT-86. pp 332-337.
- LYBERG L. (1985). Plans for computer-assisted data collection at statistics Sweden. Statistical Research Unit. Statistics Sweden. Stockholm.
- MOZUMDER KORSHED ALAM A.B.M., PHILLIPS J.F., LEON D. HOSSAIN M.B. (1986). The sample registration system, a microcomputer system for monitoring demographic dynamics and health and family planning service operations in rural Bangladesh. Paper for the Population Association of America Annual Meeting, San Francisco.
- OCHOA L.H., ORTUZAR J., WILLARD D. (1988). Procesamiento interactivo de encuestas complejas con microcomputadores portátiles : un experimento en Guatemala. Paper presented at the Seminar CELADE-IUSSP on Collection and Data Processing of Demographic Data in Latin America. May 23-27. Santiago.
- STRONG M.A. (1987). Software for Demographic Research. Population Index. 53(2). pp 183-199.
- THOMPSON R. (1984). The potential of new technology for future survey data processing. World Fertility Survey 1972-1984. Symposium. Methodology Session 3 Paper 2. London.
- WESTLAKE A., CROFT T., THOMPSON R. (1984). Realising the Potential of New Technology for Processing Sample Surveys in developing countries. World Fertility Survey - WFS/tech 2328, London.
- WORLD BANK (ed) (1986). What PCs add to surveys. The World Bank - Research News - Vol 6 N°3, Washington.

SUMMARY

Considering the key points of data collection and editing in demography, the authors look at the recent improvements given by micro-computers on these topics. They show the use of micros closer to the field-work for data-entry, the use of lap-top computers on field with specific programs and data bases, generally replicating desk tasks on field. New integrated researches, specially focused on data collection and editing gave new software packages specific for data collection with lap-tops. The authors illustrate this with the Integrated System of Data Collection, a new software, based on a radically different approach, which has been already used successfully with lap-tops on field. With micros, demographers gained already new tools for analysis, they are now getting real tools for data collection.