

## **CAPI ET BLAISE POUR L'ENQUÊTE EMPLOI EN FRANCE**

*Jean-Luc Heller*

*INSEE - Département des Projets, France*

En mars 1993, CAPI a assuré la moitié de la collecte de l'enquête annuelle sur l'Emploi (Labour Force Survey). 630 enquêteurs répartis sur tout le territoire (y compris dans l'île de l'océan indien de La Réunion) se sont rendus dans 45000 logements pour interroger ainsi plus de 70000 personnes.

### **1. Une opération qui vient de loin**

En France, pour l'INSEE, des essais se sont déroulés entre 1987 et 1989 sous le terme de "saisie portable". Les premières "Propositions d'utilisation de matériels de saisie portable pour des enquêtes auprès des ménages" ont ainsi été émises en 1987. Une expérience était lancée sur l'enquête quadrimestrielle de conjoncture auprès des ménages. L'autre expérimentation importante s'est déroulée en janvier-février 1989 à Nantes et Marseille avec une enquête pilote sur l'Emploi. Elle traçait déjà les grandes lignes d'une architecture possible: portables de collecte type Toshiba 1000, PC hôte dans les régions, transmissions, messagerie, utilisation de BLAISE.

Tous les problèmes techniques n'étaient pas résolus mais l'impression était qu'ils étaient en passe de l'être à l'horizon de quelques années. Les expériences dans les autres pays incitaient à poursuivre dans cette voie. Restait à créer les conditions d'un véritable passage en production avec ce que cela comportait de moyens à dégager dans la préparation, de convivialité et de fiabilité des outils à mettre en oeuvre.

### **2. L'enquête Emploi, pourquoi, comment?**

L'enquête Emploi, par son ampleur et sa permanence, les contraintes de délai de sortie des résultats justifiait un tel investissement. La rénovation

intervenue pour l'enquête de 1990 anticipait déjà ce choix en mettant en place un système intégré faisant largement appel au chiffrage automatique et à la reprise des anomalies sur un poste de travail informatisé. A contrario, les études avaient soulevé la difficulté qu'il y a à informatiser la collecte d'une enquête sans projet d'ensemble de rénovation des traitements.

Compte tenu des risques et de l'ampleur de l'opération, le choix a été fait d'une implantation progressive. En avril 1991, un nouveau test, en parallèle avec l'enquête Emploi de mars, a touché deux régions. A la même époque, tout le réseau des enquêteurs faisait l'objet d'une enquête préalable sur son attitude vis à vis de CAPI. La bonne acceptation du procédé par les ménages et les enquêteurs a été ainsi confirmée. Pour l'enquête de mars 1992, CAPI s'est limité à sept régions, pour la partie entrante de l'échantillon, soit un neuvième environ de l'enquête. La comparabilité d'une année sur l'autre des sous-échantillons était ainsi préservée. L'extension nationale en mars 1993 à près de la moitié de l'enquête, s'est faite aussi en fonction du schéma de renouvellement de l'échantillon.

### **3. Un triple système à mettre en oeuvre**

CAPI ne peut se limiter à l'usage d'un micro portable pour réaliser des interviews avec un protocole de questionnement assisté : transmissions, suivi de gestion, interface avec l'existant sont partie intégrantes du système et représentent une bonne partie de la charge de conception. Il a fallu définir, tester et mettre en place un triple système, ce qui a représenté deux ans et demi de travail pour les quatre personnes de l'équipe de projet.

Le système de questionnement détermine le poste de travail de l'enquêteur : sur un micro-ordinateur portable, celui-ci dispose de la liste des unités à interroger et des programmes de questionnement assurant questions et contrôle des réponses saisies. Ce poste intègre aussi des fonctions d'auto-formation par EAO et de communication pour l'envoi des données et messages.

Le système de gestion est appuyé sur le poste de travail des services régionaux qui supervisent la collecte. Il inclut des fonctions de préparation

et de suivi de la collecte. Il assure aussi les interfaces sur site central avec l'amont, pour importer les bases de sondage et les données antérieures, et avec l'aval, pour effectuer transferts et mise à disposition des données du terrain pour les traitements complexes ultérieurs. En l'occurrence, l'implantation progressive, à l'identique avec les questionnaires papier, qui imposait de maintenir en parallèle deux systèmes, s'est révélé une contrainte forte.

Le système de transmission assure la communication entre les différents acteurs (enquêteurs, services régionaux), pour les transferts des données saisies par les enquêteurs, l'information sur le déroulement des opérations, l'envoi des programmes et messages.

#### **4. Les choix informatiques: architecture, matériel, logiciel**

L'architecture s'articule autour d'un centre national informatique de l'INSEE, à Aix en Provence, centre collecteur des enquêtes. Les données de collecte ainsi que les programmes et messages sont centralisés physiquement sur gros système, chaque service régional en conservant la maîtrise logique par l'intermédiaire de son poste de travail. Pour cela, les services gestionnaires des enquêtes en région sont dotés d'un poste PC de bureau, avec des fonctions d'impression locale et de communication avec l'ordinateur central par le réseau INSEE.

Sur le terrain, il faut un matériel suffisamment léger, fiable et autonome qui offre en même temps une bonne qualité d'écran et une capacité de traitement acceptable. Le choix actuel est celui du T1800 de Toshiba. Un petit boîtier modem externe permet à l'enquêteur de transmettre les données à partir de son domicile par sa ligne téléphonique et un accès au réseau X25 (TRANSPAC 2400).

Les éléments principaux du dispositif logiciel font appel à BLAISE, et à ARBITER (produit commercial d'origine américaine) pour les transferts de données, programmes et messages entre les acteurs du système. De nombreux

développements complémentaires ont été écrits en Pascal, Cobol et procédures sur micro et site central pour les fonctions de gestion, transfert et mise en forme des données.

## **5. Une construction autour de BLAISE.**

BLAISE a été utilisé pour les programmes de questionnement en profitant de la souplesse des options offertes. Les textes des messages ont été traduits et adaptés. Les valeurs des touches de fonctions ont été modifiées pour offrir une ergonomie plus cohérente. La codification a été mise en oeuvre pour les indications géographiques (pays, département de résidence ou d'origine) ou de spécialisation des formations. La lecture d'informations externes sert à l'insertion de l'identification de l'enquêteur dans les fichiers à partir des paramètres de personnalisation du micro nécessaires à l'établissement de la connexion. ABACUS fournit des états d'avancement. En revanche l'utilisation de la fonction NOTEPAD n'a pas donné satisfaction dans les versions de BLAISE utilisées (2.33 puis 2.39).

Autour de BLAISE de nombreux développements complémentaires ont été écrits: menu général de lancement, procédures de lancement des programmes et conversion des données, et surtout programmes de liens entre les bases BLAISE et programmes de sélection et de marquage des données à transmettre. Une meilleure connaissance de notre part des possibilités de BLAISE (en particulier de MANIPULA) aurait sans doute permis de répondre plus légèrement aux fonctionnalités souhaitées.

Ces programmes, qui représentent une charge importante, ont été rendus nécessaires par le choix d'offrir une organisation identique à celle des questionnaires papiers qui continuaient à être utilisés en parallèle: par exemple offrir le choix de la personne du ménage à interroger indépendamment de l'ordre de saisie de la liste des personnes du ménage; ou donner la possibilité à l'enquêteur d'une visualisation de l'ensemble de ses questionnaires avec indication de leur état (Status): à faire, incomplet, bon à transmettre, transmis.

Fondamentalement on s'en calé sur l'organisation existante de l'enquête Emploi. L'enquêteur reçoit une liste de logements à repérer décrits à partir des données du Recensement général. Il met à jour les caractéristiques du logement ou enrichit la base. C'est un premier questionnaire BLAISE qui utilise en information externe les données de la base de sondage. Une base des logements à interroger effectivement est produite à partir de ce premier repérage. Ils font l'objet d'une deuxième questionnaire BLAISE sur les caractéristiques du logement et du ménage répondant. Au terme de ce questionnaire, la base des personnes adultes est constituée pour passer le troisième questionnaire BLAISE, le questionnaire individuel. Il a fallu gérer dynamiquement (hors BLAISE) les liens entre les trois bases.

Les contraintes de place mémoire à l'exécution ont aussi pesé. La taille du questionnaire individuel a imposé d'en faire un programme BLAISE séparé; une table de la taille maximum possible d'un ménage était trop coûteuse en mémoire.

Des modules spécifiques ont été nécessaires pour déterminer les données bonnes à envoyer: estimées comme telles par l'enquêteur, complètes et cohérentes entre les différents niveaux (logements où tous les adultes ont correctement répondu à un questionnaire individuel, logements vacants, vides, refus, etc.). Après envoi, un nouvel ensemble de modules met à jour les bases pour éviter une sélection lors de l'envoi suivant. Si l'exécution des programmes spécifiques (PASCAL et COBOL) est relativement rapide, les nombreuses conversions BLAISE - ASCII sont très pénalisantes.

## **6. L'utilisation particulière des "carnets de tournée" (BLOCK APPOINTMENT)**

Tout ceci repose sur un usage particulier des carnets de tournée de BLAISE.

Dans le cas où une interview doit être reportée, BLAISE prévoit l'utilisation d'un carnet de tournée où l'enquêteur note les renseignements sur un rendez-

## *CAPI et BLAISE pour l'enquête Emploi en France*

vous donné par le ménage, date, heure et adresse par exemple. Lors de sa prochaine tournée, l'enquêteur consulte ce carnet et une fois l'interview réalisée, la ligne disparaît du carnet de tournée. Telle est l'utilisation normale d'un carnet de tournée prévue dans BLAISE.

Le principe du carnet de tournée peut être détourné de sa vocation originale pour devenir un moyen de donner à l'enquêteur une liste d'unités à interroger, dans laquelle il sélectionne celle qu'il désire enquêter. On peut ainsi intégrer des indications d'adresse et de repérage du logement, et une fois l'interview réalisée, certaines variables du carnet de tournée peuvent être modifiées en fonction du contenu du questionnaire (utilisation du paragraphe CHECK du BLOCK APPOINTMENT). A la fin de l'interview, on revient au carnet de tournée.

Lors du chargement du micro de l'enquêteur, on peut ainsi lui donner une liste d'unités à enquêter, obtenue en convertissant en BLAISE le fichier ASCII \*.A02. Ce fichier \*.A02 comprend les indications qu'on veut voir figurer à l'écran pour guider l'enquêteur dans sa tournée, informations décrites dans le BLOCK APPOINTMENT, elles correspondent aux paragraphes QUEST et ROUTE de ce BLOCK.

Ci-dessous deux exemples de carnets de tournée utilisés pour l'enquête Emploi.

\* carnet de tournée des logements

IMLOC	COM	ADRESSE	SITU_LOG	E
00101	PAR	18RUE BERLIOZ	PORTAIL FER FORGE	A
00201	PAR	20RUE BERLIOZ	PORTE JAUNE	B
00301	PAR	22RUE BERLIOZ	AU FOND DE LA COUR	I
00401	PAR	24RUE BERLIOZ	BAT ESC ETG RCPTE DROITE	_
00402	PAR	24RUE BERLIOZ	BAT ESC ETG RCPTE GAUCHE	_
00403	PAR	24RUE BERLIOZ	BAT ESC ETG 01PTE FOND	_

*CAPI et BLAISE pour l'enquête Emploi en France*

00501 PAR 18BRUE BERLIOZ VILLA NEUVE \_

\* carnet de tournée des personnes adultes

NUMERO	IMLOC	NO	PRENOM	SAN_N	REMARQUE	ETAT
2147483634	00201	01	PIERRE	1	1930	_____
2147483634	00201	02	JOELLE	2	1936	MANQUE ADRESSE EMPLOYEUR INC
2147483634	00201	03	VALENTINE	2	1906	BON
2147483635	00101	01	DANIEL	1	1951	AIX
2147483635	00101	02	ROSA	2	1938	AIX
2147483635	00101	03	DAVID	1	1966	AIX
2147483635	00101	04	SARAH	2	1977	AIX

La mise à jour du carnet de tournée en fonction du contenu du questionnaire qui vient d'être traité est très utile. Exemples: détermination et affichage de l' "état du questionnaire" (dans les exemples ci-dessus: bon, incomplet, non traité ou envoyé) permettant à l'enquêteur d'avoir une vision synthétique de l'avancement de son travail, zone de remarques sur les questions manquantes ou à compléter.

Cette mise à jour est faite par le CHECK du BLOCK APPOINTMENT au moment de quitter le questionnaire. Par exemple, pour mettre à jour la question ETAT du questionnaire comme zone du carnet de tournée, le paragraphe CHECK du BLOCK APPOINTMENT comporte une clause du type :

```
IF (condition sur Qst) THEN COMPUTE ETAT:='BON'ELSE COMPUTE ETAT:='INC'ENDIF;
```

"Qst" est une question posée dans le routage du questionnaire principal ou bien une question cachée du questionnaire principal calculée dans le CHECK

principal à partir d'une (ou de plusieurs) question figurant dans le routage principal avant la dernière question posée du questionnaire.

Cette utilisation spécifique nécessite de lancer le contrôle syntaxique avec une version modifiée du programme CAPI.P23 ou CAPI.P24.

Tout ceci a contribué à donner aux enquêteurs une bonne impression de facilité et de clarté dans la mise en oeuvre de CAPI, malgré leurs craintes initiales très fortes.

## **7. Formation et assistance: deux "investissements" stratégiques**

Le postulat du projet était de s'adresser au réseau d'enquêteurs actuel, sans recrutement ou sélection particulière. Compte tenu de l'hétérogénéité de la formation initiale, de l'absence très générale d'expérience informatique, et des craintes ressenties vis à vis de l'usage de cet outil, la formation était un enjeu crucial.

Un plan de formation ambitieux a permis d'assurer à chaque enquêteur quatre jours de formation au procédé CAPI, complétés par deux jours sur l'enquête proprement dite. Les formateurs, constitués en "binômes" issus des services d'enquêtes et des services informatiques régionaux, ont eux-mêmes suivi une formation de onze jours.

A l'issue de la formation, les enquêteurs sont dispersés sur tout le territoire, parfois à plusieurs centaines de kilomètres des Directions Régionales. Ils ont à leur disposition une application informatique complète mais ne peuvent avoir recours à l'assistance de proximité qui joue un grand rôle dans l'informatisation d'un travail de bureau. L'assistance à distance, organisée soigneusement sous forme de permanences, a associé les responsables régionaux des enquêtes et les informaticiens. L'assistance intervient dès la période de formation et prend une importance particulière lors de la phase de collecte dont les délais sont stricts. Il a souvent été fait appel à un relais national pour des problèmes plus complexes.



## **8. Des délais réduits malgré un temps de questionnement plus rigide**

La saisie directe à la source et la transmission automatisée suppriment les différents goulots d'étranglement de la méthode traditionnelle sur papier: constitution de "paquets", ruptures dans la chaîne de traitement entre les différents ateliers concernés. Par rapport à l'expérience de 1992, les améliorations apportées au matériel et à l'utilisation des logiciels ont permis d'obtenir rapidité et fiabilité des transmissions, assurées ainsi par réseau téléphonique dans 95% des cas.

Ainsi un questionnaire peut être entièrement validé et traité en 48 heures. La comparaison des indicateurs d'avancement du traitement de l'enquête entre 1992 (sur papier pour 8/9èmes) et 1993 (où CAPI représente près de la moitié) confirme l'amélioration. En fin de troisième semaine, 33% des questionnaires individuels sont intégrés dans le fichier d'enquête et 16% validés, contre respectivement 8% et 2% en 1992.

Le rythme global de collecte n'a pas été affecté par l'introduction de CAPI, bien que les enquêteurs aient fait état d'un temps de questionnement parfois plus long qu'avec le papier. Les mesures directes effectuées sur les micro-ordinateurs laissent penser que l'augmentation reste cependant faible et devrait encore se réduire, voire tendre à disparaître au fur et à mesure de la généralisation du procédé. Par ailleurs l'ensemble du travail hors interview se trouve allégé. La quasi totalité du réseau considère CAPI comme un progrès et souhaite son extension rapide à d'autres enquêtes.

## **9. Une qualité accrue allégeant la charge de correction**

Outre l'économie de la saisie, CAPI permet un allègement en moyens manuels de la phase consacrée à la reprise pour correction des questionnaires. Par rapport au traitement papier, on peut estimer autour de 35 000 heures (saisie et reprise) le gain annuel apporté par un traitement entièrement CAPI.

Par construction, BLAISE supprime les erreurs d'identification, les erreurs

de routage (partie ou question manquante ou remplie à tort), ainsi que les erreurs de modalité des réponses. Les contrôles de cohérence possibles entre les réponses ont été limités. Il ne faut pas perturber l'interview par le règlement de litiges trop complexes. Et il est parfois bon de s'en tenir à l'enregistrement d'une situation, même avec ses aspects contradictoires ou incohérents, quitte à reprendre plus tard les anomalies.

La mesure du gain apporté par CAPI est encore perturbée par des phénomènes liés à l'apprentissage et à d'inévitables mises au point. On peut avancer le taux de questionnaires "bons pour la première fois" qui résulte du niveau de qualité requis par paramétrage lors des contrôles et des chiffrements automatiques. Il faut le ramener à un indice pour apprécier la variation relative cet indicateur. Ainsi en 1992 la partie de l'échantillon entrant enquêtée par CAPI a atteint un indice de 126 (questionnaires bons pour la première fois) par rapport au 100 de la partie hors CAPI. En 1993, cet indice s'établit à 132. Autre critère de qualité, le taux de réponse à l'enquête ne semble pas être affecté même si les relances semblent pâtir du nouveau procédé.

En première analyse les biais par rapport au papier ne semblent pas sensibles mais l'enquête de 1993 doit permettre de tester d'éventuelles divergences entre papier et CAPI à partir de l'étude statistique d'environ 5000 logements de la région parisienne en deux échantillons comparables.

## **10. En conclusion**

Le procédé est très favorablement accueilli par les enquêteurs et bien accepté par les ménages, contrairement à certaines craintes initiales. Les délais sont effectivement réduits par la suppression des manipulations et de la saisie des questionnaires papier et la collecte de données de meilleure qualité qui allège la charge de correction. Les investissements réalisés centralement (travail de l'équipe de projet et des autres unités sollicitées, acquisition du matériel, formation de formateurs) ou localement (formation, préparation, assistance) se traduisent dès maintenant par une économie substantielle de moyens de saisie et de traitement. Le bilan, positif, conduit à souhaiter l'extension de CAPI. A travers l'expérience acquise par le traitement de

l'enquête Emploi l'infrastructure générale est en place et la formation du dispositif de collecte largement acquise. Reste à "accueillir" les utilisateurs suivants : enquête trimestrielle Emploi, enquêtes complémentaires, enquêtes conditions de vie, loyers et charges, enquêtes régionales. Ce qui, en l'absence de procédures très générales et paramétrables, ne peut ce faire encore qu'au prix d'un investissement élevé en conception informatique.