LE CODAGE AUTOMATIQUE D’UN CARNET DE DEPENSES EST-IL PLUS COMPLEXE QUE CELUI D’UN CARNET D’ACTIVITES?

F. DESCHAMPS(*) et S. DESTANDAU(**) et F. DUMONTIER(***)

(*) INSEE, Unité Méthodes Statistiques
(**) INSEE, Division conditions de vie des ménages

Introduction:

Dans cet article, nous vous présentons les préparatifs à la codification automatique avec Sicore1 de deux enquêtes auprès des ménages : Emploi du temps (EdT2) et Budget de Famille (Bdf3) et leurs résultats.


Nous vous montrons qu’en dépit de fortes similitudes sur le dispositif des deux enquêtes, les travaux préparatoires et les bilans de codification (finaux pour EdT et provisoires pour Bdf) sont bien différents du fait du type de variable à coder, de la nomenclature associée, et du dispositif de recueil de l’information.

---

1) Pour plus de renseignements sur Sicore et ses connaissances, on se reportera à l’annexe nº1
2) Pour plus de renseignements sur l’enquête Emploi du Temps, on se reportera à l’annexe nº2
3) Pour plus de renseignements sur l’enquête Budget de Famille, on se reportera à l’annexe nº5
4) L’enquête Bdf est encore sur le terrain à la date de ces journées.
1. La variable activité et son support de collecte : le carnet

1.1 Le carnet d’emploi du temps journalier

Le carnet journalier de l’enquête Emploi du Temps permet de décrire toutes les activités d’un même jour de certains individus du ménage interrogé. Celui de l’enquête de 1986 était découpé en 288 lignes représentant les tranches de 5 mn d’une journée (de 0h à 24h) ; un individu kish et son conjoint éventuel devaient remplir un carnet chacun.


1.2 Une variable homogène et structurée

L’enquêteur insiste auprès des ménages pour qu’ils décrivent leurs activités de façon claire, par une phrase simple, dans leur langage courant (exemple : je conduis les enfants à l’école) et sur le fait qu’il fallait que l’individu ne décrive qu’une activité à la fois. Nous avons obtenu des phrases structurées du type “sujet + verbe + (compléments)”.

Or, très souvent, plusieurs activités pouvaient être réalisées simultanément. On demandait alors aux enquêtes de n’en noter que deux. La plus importante des deux était alors appelée “activité principale” et inscrite sur une ligne du carnet avec sa durée. La deuxième dite “activité secondaire” était aussi inscrite sur la même ligne du carnet que la première, mais dans une autre colonne - moins large. La durée, par construction du carnet, correspondait à celle de l’activité principale associée.

Chaque plage horaire de 10 mn tient sur deux pages face à face du carnet (cf. annexe n°3). La page de droite est réservée aux intitulés des activités primaires et secondaires, la page gauche, aux variables annexes. Sur la page de droite, on a délibérément offert deux fois plus de place pour les intitulés des activités primaires que pour les secondaires, de façon à laisser aux individus assez de place pour décrire correctement les premières. En revanche, on acceptait pour les activités secondaires une description plus sommaire (comme « radio », « conversation »).
Extrait de la page gauche d’un carnet de l’enquête EdT 1998-99

<table>
<thead>
<tr>
<th>Marquez vos différentes occupations de la journée en indiquant les heures de début et de fin de chaque occupation à l’aide d’accolades dans la colonne de gauche</th>
<th>Faites-vous autre chose en même temps ? (lecture, conversation, radio, TV...)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7h00</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>je dors</td>
</tr>
<tr>
<td>20</td>
<td>je fais ma toilette</td>
</tr>
<tr>
<td>30</td>
<td>je m’habille</td>
</tr>
<tr>
<td>40</td>
<td>je prépare le petit déjeuner</td>
</tr>
<tr>
<td>50</td>
<td>je déjeune</td>
</tr>
<tr>
<td>8h00</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>je fais la vaisselle</td>
</tr>
<tr>
<td>20</td>
<td>je range la cuisine</td>
</tr>
<tr>
<td>30</td>
<td>je fais le ménage</td>
</tr>
<tr>
<td>40</td>
<td>je vais au travail avec ma femme et un voisin</td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>9h00</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

Cette présentation était la même lors des enquêtes précédentes.
2. La nomenclature initiale de la variable activité

La nomenclature de 1986 contenait 199 postes de base à trois chiffres, qui pouvaient être regroupés en huit grands postes correspondants au premier chiffre :

<table>
<thead>
<tr>
<th>Premier chiffre</th>
<th>Titulé du poste</th>
<th>Exemples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Besoins physiologiques</td>
<td>Sommeil, soins personnels (se laver, s'habiller...) soins médicaux, repas</td>
</tr>
<tr>
<td>2</td>
<td>Temps de travail professionnel et temps de formation</td>
<td>Travail professionnel, formation professionnelle, études (étudiants, lycéens), autres formations</td>
</tr>
<tr>
<td>3</td>
<td>Travaux domestiques</td>
<td>Cuisine, ménage, soins du linge, courses, services administratifs, bricolage, jardinage, divers...</td>
</tr>
<tr>
<td>4</td>
<td>Soins aux personnes</td>
<td>S'occuper des enfants, jouer avec, soins matériels ou médicaux aux adultes</td>
</tr>
<tr>
<td>5</td>
<td>Sociabilité</td>
<td>Réceptions et sorties, conversation, téléphone, courrier, religion, participation civique et entraide</td>
</tr>
<tr>
<td>6</td>
<td>Loisirs</td>
<td>Sport, promenade, chasse - pêche, médias, spectacles, passe-temps et jeux</td>
</tr>
<tr>
<td>8</td>
<td>Trajets</td>
<td>Trajets domicile-travail, trajets liés aux enfants, autres trajets</td>
</tr>
<tr>
<td>9</td>
<td>Remplissage du carnet Insee</td>
<td></td>
</tr>
</tbody>
</table>
3. Pourquoi utiliser Sicore\(^5\) pour coder l'activité?

Pour l’enquête précédente de 1986, le codage de l’activité figurant dans les carnets journaliers était effectué par une seule équipe de codeurs du CNE de Toulouse et pesait très lourd en heures manuelles de codage (45 000 heures environ).

Le CNE de Toulouse n’existait plus, et les DR ne disposant pas d’une telle ressource de moyens manuels pour que l’enquête puisse avoir lieu sous la même forme qu’en 1986, il fallait trouver un moyen de réduire ces charges. Nous avons alors étudié la possibilité d’utiliser Sicore. Cela constituait une première : l’activité quotidienne était une nouvelle variable jamais codée et, de plus, les activités des anciennes enquêtes n’avaient jamais été saisies sur support informatique.

Toutefois, une chose était sûre, nous aurions moins de cas à traiter manuellement que lors de l’enquête précédente. Mais nous ne savions pas, au tout début des préparatifs relatifs à l’utilisation de Sicore, si le codage avec Sicore serait d’aussi bonne qualité qu’un codage manuel qui aurait été effectué de la même manière qu’en 1986.

4. La constitution des connaissances

4.1 Le premier fichier d’Apprentissage Brut (Fab) des activités

Dans la documentation de l’enquête de 1986, pour donner un aperçu de la diversité des libellés et de la richesse de cette nomenclature, figurait une liste de 3 918 de libellés différents, issue de 900 carnets de l’enquête de 1986, chaque libellé étant associé à un poste de la nomenclature. Après la saisie de cette liste de libellés et de codes, un premier fichier d’apprentissage était né. Il comportait 3 918 lignes.

Afin de tester ce Fab, la technique utilisée par Pascal Rivière\(^6\) a été :
- d’extraire un échantillon du Fab
- d’effectuer l’apprentissage sur le reste du Fab
- de coder cet échantillon.

---

\(^5\) Pour plus de renseignements sur Sicore et ses connaissances, on se reportera à l’annexe n°1

\(^6\) À l’époque, en mars 1995, Pascal Rivière était C.P.S. du projet Sicore, alors en phase d’achèvement. À cette occasion, il a formé Jean-Louis Pan Ke Shon, expert de la variable activité.
Dix-neuf échantillons ont été ainsi extraits et codés de cette façon. En moyenne, le taux moyen de réussite de ces 19 codages a été de 40.4% (avec un minimum de 35% et un maximum de 46%) dont 32.4% codés simples et 8% codés multiples.

La qualité moyenne brute, pourcentage de “bien codés” parmi les libellés codés automatiquement a été de 50.5%. L’expression de “bien codés” a ici un sens précis, dans la mesure où chaque libellé de l’échantillon est issu du Fab et par conséquent accompagné d’un code, que l’on appellera code de référence. Un libellé “bien codé” est un libellé dont le contenu obtenu automatiquement est égal au code de référence.

Mais, le fait qu’il existe des codes multiples pour un même libellé va conduire de façon mécanique à coder des libellés dans un code faux (i.e. différent du fichier d’apprentissage) et conduire à une très nette et artificielle diminution de la qualité du chiffrage automatique. En étudiant les codés multiples, Pascal Rivière a évalué à 10.9%, cette proportion de mal codés mécaniquement. Pour calculer la véritable qualité, il fallait tenir compte de cet effet pour redresser la qualité moyenne brute. Il a donc évalué une qualité redressée de 76%.

4.2 L’amélioration standard des connaissances pour la variable activité

La mise en œuvre de la boucle Sicore conduit à ce que l’expert variable effectue toujours le même genre d’opérations sur les connaissances relatives à la variable :

- L’ajout d’expressions originales au sein du Fab, issues de la saisie de carnets soit de l’enquête EdT précédente de 1986 soit des tests ayant eu lieu sur le terrain.

- La déclinaison des libellés déjà présent dans le Fab.
Les déclinaisons verbales au sens le plus strict en font partie.
    Par exemple, ayant “JARDINAGE”, on peut rajouter : “J’AI JARDINE”,
    “NOUS JARDINONS”, “JARDINANT”, ...
On peut également imaginer des variantes plus larges des libellés.
   Par exemple, ”REVEIL” peut amener à “MAMAN ME REVEILLE”.

- Le choix des caractères blancs ou vides et des mots vides.
    Par exemple, supposons que toutes les déclinaisons du verbe être et tous les articles et prépositions sont éliminés à la normalisation (ce sont des mots vides). Si, de plus “MAISON” est un autre mot vide car il indique un lieu plus qu’une action, alors le libellé “JE SUIS A LA MAISON” devient un

---

7 En l’occurrence Jean-Louis Pan Ké Shon
libellé blanc, alors qu'il pourrait être codé en repos, ou en réflexion. Les mots vides doivent alors être revus.

- Le choix, l'intégration, la mise en ordre et l'élimination de synonymes inutiles. Le choix des synonymes se fait grâce à l'étude des fréquences des libellés soit mal codés par Sicore, soit non codés. Leur intégration dépend de la nomenclature et doit respecter un ordre à tester pour une utilisation optimale.

Par exemple, supposons que dans la liste des synonymes, figurent dans l'ordre toutes les déclinaisons de "DORS", puis toutes celles de "SOMNOLE"; ensuite, dans la liste des synonymes "SOMNOLE" sera associé à "DORS", alors, quelle que soit la forme verbale du verbe somnoler, celle-ci sera codée comme "DORS". Certains synonymes peuvent avoir un effet non envisagé initialement qui aboutit à leur suppression.

Ainsi, voici quelques caractéristiques respectives du premier et du dernier environnement de codage de la variable activité :

<table>
<thead>
<tr>
<th>Environnement</th>
<th>Premier</th>
<th>Dernier⁹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre de lignes du Fab</td>
<td>3 918</td>
<td>13 400</td>
</tr>
<tr>
<td>Nombre de synonymes</td>
<td>0</td>
<td>2 388</td>
</tr>
</tbody>
</table>

4.3 La mesure de la qualité du codage automatique

Etant donnés les petits volumes de libellés lors des 3 tests de cette enquête, il a été possible de demander un codage manuel aux personnes effectuant la saisie des libellés d'activité. Cela a permis d'avoir un code de référence que Sicore pouvait comparer au code qu'il attribuait, et donc d'évaluer une certaine qualité de codage automatique. De plus, l'expert variable a analysé les rejets et les codés multiples de la codification automatique pour améliorer l'environnement Sicore ainsi que le prévoit le dispositif de la boucle Sicore.

⁹ L'environnement qui a servi au recodage complet des libellés de l'enquête à la fin de celle-ci.
Les résultats de ces tests sont les suivants :

<table>
<thead>
<tr>
<th>Date</th>
<th>Nombre de libellés</th>
<th>Environnement SICORE</th>
<th>Codés Sicore</th>
<th>Qualité(^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>test méthodologique</td>
<td>3 253</td>
<td>Fab = 9 500 lignes 1 000 synonymes</td>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>juin – juillet 1996</td>
<td></td>
<td>introduction des mots joker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(^{er}) test terrain</td>
<td>4 270</td>
<td>Fab = 10 500 lignes 1 650 synonymes</td>
<td>83%</td>
<td>93.5%</td>
</tr>
<tr>
<td>juin 1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2(^{ème}) test terrain</td>
<td>4 220</td>
<td>1(^{er}) codage même environnement que test précédent</td>
<td>78.5%</td>
<td>80.5%</td>
</tr>
<tr>
<td>septembre 1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2(^{ème}) test terrain</td>
<td>4 220</td>
<td>2 (^{ème}) codage environnement amélioré</td>
<td>81.6%</td>
<td>96.1%</td>
</tr>
</tbody>
</table>

Juste après le deuxième test terrain, la codification automatique par Sicore donnait plus de 80% de libellés codés avec une excellente qualité.

4.4 L’utilisation de variables annexes

Le codage du premier Fab sur lui-même a permis de détecter des libellés qui menaient à plusieurs codes. (1/5 des libellés - cf. annexe 4).

Ceci a confirmé, ce qui était prévisible, au vu de la structure de la nomenclature, qu’il fallait introduire des variables annexes dans le carnet qui donnent des règles de codage pour Sicore.

*Par exemple, "BOIS CAFE" pouvait être codé 146 (prendre le café ou le thé à domicile), 156 (prendre le café ou le thé sur son lieu de son travail), 163 (prendre le café, le thé chez des amis, voisins, parents lors de visites, réceptions) ou 166 (prendre le café, le thé à domicile, avec des amis, voisins, lors de réceptions, visites).*

En 1986, certaines variables annexes étaient relevées en clair sur le carnet dans deux colonnes séparées en face de chaque activité : "en présence de qui ?" et le "lieu de l’occupation". Ces deux variables annexes se sont donc imposées d’elles-mêmes. Le nombre de modalités a été fixé à 5 pour la variable lieu ("chez soi", "lieu de travail", "à l’extérieur", "traitement autre domicile-travail" et "autre trajet") et à 4

---

\(^6\) La qualité calculée ici est la proportion de codes parmi les codés automatiques, qui sont égaux aux codes manuels que les DR avaient codés manuellement avant la saisie.

\(^11\) En septembre 1997, nous avons fait deux codages, l’un en gardant, sans l’enrichir, l’environnement de juin 1997, et l’autre en l’enrichissant au vu des résultats du codage de juin 1997. Nous avions alors obtenu ce que nous espérions, c’est à dire un taux final de codage dépassant 80% avec une excellente qualité (96%).
pour la variable présence ("seule", "autre personne du ménage", "amis, voisins, parenté, collègue" et "autre personne").

En plus de ces variables, d’autres variables annexes ont pu être définies sur les carnets de 1986, et reprises en 1998. En particulier, la variable de but figurait aussi sur le fichier de 1986 (mais pas sur le carnet) ; cette variable n’était pas directement remplie par les individus en 1986, elle était chiffrée par l’enquêteur en cas d’ambiguïté à partir des intitulés des activités. Pour l’enquête de 1998-1999, le nombre de modalités a été fixé à 4 : "personnel ou pour son ménage", "professionnel", "pour un autre ménage" et "associatif".

Ainsi, à partir de ces variables annexes, une cinquantaine de règles de codage ont été initialisées. Elles étaient au nombre de 99 à la fin de l’enquête.

**Exemple de règle de codage sur un exemple simple**

*Libellé écrit sur carnet*

"JE PREPARE UNE PIZZA POUR LES ENFANTS".

1ère étape : normalisation

"PREPARE PIZZA ENFANTS"

2ème étape : synonymisation

"PREPARE ALIMENTS ENFANTS"

3ème étape : application des règles avec variable de but

- si but personnel : 311
- si but professionnel : 211
- si but pour un autre ménage : 543
- si but associatif : 542
- si but non renseigné : 311

(arbitrairement choisi, car le plus fréquent)

12 A savoir l’heure de début d’activité, l’heure de fin d’activité, la durée en minutes de l’activité et le statut professionnel de l’individu : salarié actif occupé, agriculteur actif occupé, autre indépendant actif occupé, étudiant ou élève ou autre (chômeur, retraité, personne au foyer ...).
4.5 L’utilisation de mots joker

L’étude des libellés en rejet et des libellés codés par Sicore vers différents codes (codés multiples) a permis de déceler un besoin d’utilisation de mots joker.

Par exemple, des prénoms apparaissent fréquemment dans les compléments d’activité : “dors sommeil” est suffisant pour classer l’activité, les informations supplémentaires risquant de parasiter le codage. Ainsi “Dors sommeil Pascal” ne serait pas codé. En effet, la recherche Sicore est faite sur les quatre premiers mots, et dans ce cas, Sicore ne reconnaîtrait pas «Pascal». La syntaxe de l’introduction du mot-joker est “Dors sommeil §§”, signifiant que quelle que soit la valeur des deux derniers mots, il ne faut tenir compte que des deux premiers.

Comme on peut l’imaginer, ce procédé radical nécessite une introduction circonscrpte des jokers. Nous nous sommes limités aux libellés présentant le moins de risque, c’est-à-dire ceux possédant au minimum 2 mots pour qu’il n’y ait pas de contrôle de redondance sur un mot joker. Eventuellement, dans les cas rares où un seul mot était signifiant et sans ambiguïté, un mot joker doit être rajouté.

5. Les conséquences de l’utilisation de Sicore

5.1 Des modifications dans la présentation du carnet

Dès le premier test sur le terrain (en juin 96), les trois variables annexes citées au paragraphe 4.3 (“en présence de qui?”, le “lieu de l’occupation” et “votre activité est dans un but …”) ont été présentées sur la page droite du carnet en face de chaque activité principale (cf. annexe n°3). Bien que ces variables ne servent pas systématiquement pour la codification de toutes les activités, on a demandé qu’elles figurent sur toutes les lignes du carnet, ce qui a été effectivement bien compris puisque sur le fichier final, on constate que le pourcentage de variables non renseignées est inférieur à 1%.

5.2 La modification de la nomenclature

La nomenclature utilisée lors des tests et au début de l’enquête a été réduite de 199 postes à 103.
En effet, de passer d’un codage manuel à un codage automatisé nous a obligés à établir des règles écrites, précises, qui ne laissent place à aucune ambiguïté et nous a amenés à agréger certains postes.

Premier exemple « manger »
Certains se limitent à ce seul descriptif, d’autres fournissent plus de détail par exemple « je mange un casse-croûte ». Peut-être les premiers ont-ils aussi mangé un casse-croûte, mais on ne le sait pas. Par contre on sait que dans les deux cas, il y a eu l’action de « manger ». Cela explique que dans la nomenclature de 1998, dans le doute, nous n’avons pas essayé de garder la nature du repas (casse-croûte, collation, café, sandwich...). Par contre, nous avons gardé, le lieu du repas et avec qui était pris ce repas, informations qui pouvaient être connues grâce aux variables annexes. Les postes repas sont ainsi passés de 20 à 4.

De même « dormir »,
le poste " sommeil " est devenu un poste global agrégeant les siestes, et le repos.

De même le poste « transport »
nous n’avons plus distingué les différents modes de transport (à pied, en voiture, ...), réduisant le nombre de postes " trajets de 15 à 2. Il aurait fallu ajouter une variable annexe supplémentaire en cas de transport ce qui nous semblait alourdir le carnet inutilement. En effet, sur les carnets de 1986, on a constaté que le mode de transport qui devait figurer en clair, avait souvent été oublié. D’ailleurs, peu d’exploitations de 1986 ont été faites à ce niveau de détail.

Par contre d’autres décisions ont été prises indépendamment du fait qu’on utilise une codification automatique ou non.

Par exemple,
nous avons supprimé les codes qui étaient apparus en 1986 associés à des activités effectuées par moins de 100 individus.

Avec ce nouvel environnement, 80% des 3 253 libellés issus du test méthodologique de juin 1996 ont été codés.

Lors de l’enquête, par contre, nous avons ajouté des codes nouveaux. Ceci a été rendu possible grâce à l’utilisation des variables annexes et à la fin de l’enquête, la nomenclature est passée de 103 à 139 postes.

En effet, grâce aux variables de but, nous avons pu dissocier certains postes ; en particulier les postes concernant les travaux ménagers ou les aides, si l’activité était faite « pour un autre ménage » ou « dans un but associatif ».

INSEE Méthodes 17
6. Au bilan

6.1 Une réussite dans la quantité et la qualité du codage

Nous avions décidé, au moins pour les premières vagues, étant donné le caractère saisonnier de la variable, qu’une étude approfondie des rejets serait faite pour améliorer l’environnement Sicore. Pour cela, à chaque vague, l’expert variable, à partir de la liste des biens codés, triée par libellés et codes, vérifiait le bon codage. De plus, à partir des rejets triés par libellés, il pouvait comprendre pourquoi le libellé était mal codé et agir sur l’environnement Sicore pour qu’il soit codé lors des vagues selon le dispositif décrit ci-dessus.

Les libellés d’activités de l’enquête se répartissaient comme suit :

<table>
<thead>
<tr>
<th>Activité</th>
<th>Principale</th>
<th>Secondaire</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre de libellés</td>
<td>316 097</td>
<td>113 581</td>
<td>429 678</td>
</tr>
<tr>
<td>Pourcentage</td>
<td>73,6%</td>
<td>26,4%</td>
<td>100%</td>
</tr>
</tbody>
</table>

L’ensemble de l’enquête comportait 429 678 libellés à coder, la plupart étant des activités principales.

Les caractéristiques de l’environnement utilisé lors du recodage final sont les suivantes :

- FAB : 13 400
- Synonymes : 2 388
- Règles logiques : 99
- Nomenclature : 139
- Variables annexes : 9

À l’issue d’un recodage complet des 8 vagues grâce à cet environnement, le taux de codage automatique de l’ensemble des libellés de l’enquête a atteint plus de 90% avec une qualité excellente dépassant les 95%.

<table>
<thead>
<tr>
<th>Activité</th>
<th>Principale</th>
<th>Secondaire</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taux de codage automatique</td>
<td>89,9%</td>
<td>98,2%</td>
<td>92,2%</td>
</tr>
</tbody>
</table>

L’activité secondaire a été très facile à coder et très bien codée. En effet, elle se résume à quelques postes ("conversation", "radio", "télévision"), souvent notés tels quels dans le carnet et qui à eux trois représentent plus de 90% de l’occurrence des activités secondaires.
Le gain d'efficacité entre la première vague et la dernière vague n'est que de 2 points environ, passant de 90,4% à 92,2%. Ce gain a surtout été fait par des enrichissements à l'issue des 3 premières vagues.

**Pour mieux comprendre le gain d'efficacité entre la première et la dernière vague:**

*Lors du dernier test de septembre 1997, nous avions atteint une efficacité de 81,6% mais il ne s'agissait que des activités principales qui sont plus difficiles à coder que les secondaires. Il ne faut pas comparer ce 81,6% avec l'efficacité globale de la 1ère vague de 90,4% (ensemble des activités primaires et secondaires).*

*On peut cependant estimer l'efficacité globale du dernier test à 86%, en tenant compte de la part respective des activités primaires et secondaires dans le volume total des activités à coder.*

*Le gain d'efficacité obtenu par l'enrichissement des bases de connaissances fait par l'expert variable, après le dernier test, et servant au codage de la 1ère vague pourrait ainsi être estimé environ à 4 points (différence entre 90,4% et 86%).*

### 6.2 Une réussite en gain de temps manuel

On constate que le temps passé pour traiter ces rejets d'activité est de 1 694 heures pour environ 40 000 rejets au total. Si l'on ajoute le temps de saisie des libellés (environ 6 000 heures), on constate un gain de 22 000 heures de temps manuel par rapport aux 30 000 heures de 1986. Ceci dépasse de loin les espérances les plus optimistes du début d'enquête.


Le gain de cette méthode ne s'arrête pas là. En effet, c'est la première fois que nous avons à disposition un fichier de libellés d'activités quotidiennes de cette importance sur lequel nous pouvons travailler (analyse textuelle...), afin d'harmoniser les comparaisons à l'avenir, et effectuer toutes sortes de recodifications possibles.

Un gain certain en temps manuel, certes, mais il faut tenir compte du temps de création, gestion, formation à l'environnement Sicore qu'on peut estimer globalement à environ 6 mois de travail à temps complet par un contrôleur de l'INSEE.
6.3 Des échanges facilités par une organisation centralisée

Pour l’enquête, une organisation entièrement centralisée a été mise en place\(^{13}\). Tous les documents de l’enquête étaient traités par le GSAS de la DR des Pays de la Loire. Une petite équipe était spécialisée dans la reprise des rejets de codage sur la base du volontariat.

Dans le cas de reprises jugées délicates, et/ou dans le cas où les renseignements affichés à l’écran de reprise n’étaient pas suffisants pour coder l’activité, chaque codeur pouvait, grâce au traitement par lots, revenir au dossier papier complet pour mieux situer l’activité dans l’ensemble de la journée ou consulter des variables contenues dans l’ensemble du dossier. Il en référerait au responsable de l’équipe qui tranchait.

Des réunions fréquentes avaient lieu au sein de l’équipe pour mettre en commun les problèmes rencontrés qui étaient alors transmis à l’expert variable par le responsable. Cette mise en commun des problèmes rencontrés a deux avantages. Tout d’abord, elle permet de rendre la codification manuelle assez homogène. D’autre part, elle facilite le travail de l’expert variable qui n’a affaire qu’à un seul interlocuteur.

6.4 L’impact négatif sur les comparaisons temporelles

Le succès de la codification automatique de la variable activité sur cette enquête doit être nuancé car ce codage réussi dépend des règles établies et de leurs variables annexes choisies. Cet ensemble de choix rend plus difficiles les comparaisons avec les enquêtes précédentes codées manuellement.

En effet, d’une part les libellés des activités de l’enquête de 1986 et des précédentes n’ont pas été saisis, ce qui rend leur codage automatique par Sicore impossible ; et d’autre part, dans les éditions précédentes de l’enquête EdT, les variables annexes n’existaient pas sous la même forme que celle de l’édition de 1998-1999.

6.5 Le choix non neutre des variables annexes et des règles

Le codage automatique a l’avantage d’être homogène sur l’ensemble des libellés de l’enquête. En revanche, en 1986, les variables annexes étaient notées “en clair” sur

\(^{13}\) Pour plus de renseignements sur l’enquête Emploi du Temps, on se reportera à l’annexe n°2
le carnet quand elles l'étaient. Et c'était le codeur qui interprétait l'ensemble des informations du carnet, et codait les activités ; ce qui aboutissait à des codes hétérogènes. Par ailleurs, nous avons pu constater, sur certains carnets de 1986, que le codage manuel de l'époque n'était pas exempt d'erreurs.

Par contre, en cas de mauvaise écriture des règles, l'erreur de codage automatique entraînée sera systématique, alors qu'en 1986, une erreur dépendait des codeurs manuels et était plus diluée dans l'ensemble des activités. Cependant, si lors de l'exploitation de l'enquête de 1998, on se rend compte d'une telle erreur, on peut la corriger systématiquement en corrigeant la règle, ce qui n'est pas le cas pour 1986.

L'enquête de 1998-1999 est actuellement en cours d'exploitation et nous nous rendons compte de l'impact du remplissage des variables annexes et de leur choix dans les règles qui leur sont associées dans la codification automatique.

*Par exemple, nous avons regardé le temps de travail des assistantes maternelles dont la profession consiste à garder des enfants chez elles*14. Lorsqu'elles faisaient leur ménage, ou leurs courses, elles ont eu plutôt tendance à mettre "but personnel", ce qui est tout à fait normal, et à cause de cette variable annexe de but qui joue prioritairement dans nos règles, cette durée n'a pas été comptée en temps quotidien de travail, mais en temps domestique.

*Le temps de travail, professionnel est calculé par ailleurs dans l'enquête soit par le semainier de travail, soit par une question globale (quel est votre temps habituel de travail ?). Au vu des résultats, on s'est rendu compte que le temps de travail moyen des assistantes maternelles calculé par le carnet tout au long de la journée, est nettement inférieur aux deux autres temps de travail. En effet, quand on leur demande globalement combien de temps elles travaillent, elles on plutôt tendance à donner le temps où elles sont responsables de l'enfant, et non celui où elles s'occupent réellement de lui.*

*Si on avait eu à disposition, lors de la codification automatique, une CS plus fine, permettant d'isoler cette profession, on aurait pu coder différemment le travail professionnel des assistantes maternelles en construisant des règles adaptées sachant que lorsque l'enfant était avec elles, c'était un temps professionnel et non un temps domestique et rendre la variable de présence prioritaire par rapport à celle de but. Nous aurions alors perdu le détail de leurs activités quotidiennes, qui auraient été considérées alors comme des activités secondaires. Mais où est la vérité ?*

14 La CS actuellement dans le fichier n'isole pas les assistantes maternelles. C'est un travail de déchiffrement de l'activité professionnelle en clair qui a permis à un chercheur exploitant l'enquête de les repérer.

6.6 Le souhait du responsable de l’enquête EdT

Le souhait du responsable d’enquête serait que l’expert Sicore qui a un poste permanent, fasse vivre et grandir cet environnement Sicore pour permettre, pourquoi pas, des enquêtes Emploi du Temps plus fréquentes ou codifier des libellés d’activités quotidiennes dans d’autres enquêtes. En effet, l’enquête Emploi du Temps actuellement n’a lieu que tous les 10 ans, et il serait dommage que cette base de connaissance devienne caduque faute d’un expert variable, qui en général, n’est nommé que pour la durée d’une enquête.
II - SICORE et les produits de dépense dans l’enquête Budget de Famille 2000-2001

1. La variable produit et son support de collecte : le carnet

1.1 Le carnet de comptes

Le carnet de l’enquête Bdf court sur 14 jours et non sur un jour. Tous les individus de plus de 14 ans du ménage interrogé doivent inscrire toutes les dépenses qu’ils effectuent au cours de ces 2 semaines de collecte dans ce carnet. L’enquêteur se déplace 3 fois chez le ménage, en particulier pour expliquer l’objectif et l’utilisation du carnet, vérifier le bon remplissage sur la période de collecte et le ramasser.

1.2 Une variable hétérogène et non structurée

Mais revenons sur la définition d’une dépense dans l’enquête Budget de Famille. Elle se décompose en 4 entités : une quantité, un produit, un montant et un lieu d’achat (exemple : 1 - place de théâtre - 250F - FNAC). Elle est aussi associée à un individu d’un ménage appartenant à l’échantillon initial, et elle est repérée dans le temps par un jour parmi les 14 d’une vague donnée (cf. annexe n°6).

Cet individu peut avoir effectué cet achat de 2 manières différentes qui seront consignées dans le carnet. Ou bien, il a réalisé un certain nombre d’achats (plus de 10 articles) dans un seul lieu et possède un ticket de caisse comportant toutes les informations sur ses dépenses ; il devra donc coller son ticket sur la partie gauche du carnet. Ou bien, s’il a effectué une dépense isolée, il devra l’inscrire sur la partie droite du carnet dans le “tableau de dépenses”.

- **Lorsque la dépense est isolée**, l’individu titulaire du carnet doit remplir les 4 champs figurant dans le tableau (partie droite du carnet) :

<table>
<thead>
<tr>
<th>Quantité/unité</th>
<th>Nature de la dépense (produit)</th>
<th>Montant de la dépense</th>
<th>Type de magasin (lieu d’achat)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 litres</td>
<td>lait</td>
<td>18,00</td>
<td>épicerie</td>
</tr>
<tr>
<td>1</td>
<td>place de théâtre</td>
<td>250,00</td>
<td>FNAC</td>
</tr>
</tbody>
</table>

- **Dans le cas où il colle son ticket de caisse** (achats dans un supermarché), il doit vérifier que son ticket comporte sensiblement les mêmes informations que celles du tableau de dépenses (non compris la quantité et/ou l’unité qui
sont des informations non obligatoires). Si cela n’est pas le cas, il doit ajouter l’information manquante afin d’identifier la dépense aussi bien que s’il l’avait notée directement dans le tableau de dépenses.

Le plus souvent, les tickets de caisse ne comportent pas des intitulés de produits aussi facilement identifiables que ceux figurant sur la partie “tableau de dépenses”. Par conséquent, il doit identifier le produit de manière claire en complétant de manière manuscrite chacun des produits incompréhensibles : c’est ce que l’on appelle le complément de la dépense.

Exemples :
- Si le libellé du ticket de caisse est “1 pack Yoplait 20F”, le consommateur devra tout d’abord inscrire le magasin s’il n’apparaît pas sur le ticket. Il devra d’autre part compléter le produit Yoplait par “yaourts” Les informations sur le produit qu’il note constituent des compléments de dépense. Lors de la saisie des carnets, elles seront saisies isolément de l’intitulé du produit tel qu’il est inscrit sur le ticket.
- Dans l’exemple de ticket de caisse ci-dessous, figurent à droite les compléments de dépense censés identifier les produits achetés.

<table>
<thead>
<tr>
<th>INTERMARCHÉ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GINI 10x20CL 6 1/1.95</td>
<td>18.70</td>
</tr>
<tr>
<td>VOLVIC</td>
<td>11.70</td>
</tr>
<tr>
<td>OUIL.AUTO MAG</td>
<td>169.00</td>
</tr>
<tr>
<td>JOUETS LOISIRS</td>
<td>25.80</td>
</tr>
<tr>
<td>MIEL ACACIA</td>
<td>15.30</td>
</tr>
<tr>
<td>COQUIGRAIN 65/70x6</td>
<td>6.40</td>
</tr>
<tr>
<td>BATONNETS COTON</td>
<td>3.45</td>
</tr>
<tr>
<td>RICORE</td>
<td>20.30</td>
</tr>
<tr>
<td>FEMME</td>
<td>49.90</td>
</tr>
<tr>
<td>GENIE GEL</td>
<td>11.80</td>
</tr>
<tr>
<td>MAJESTY BOEUF</td>
<td>2.10</td>
</tr>
<tr>
<td>BTE PYRENE.</td>
<td>21.95</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>365.35</strong></td>
</tr>
</tbody>
</table>

⇒ 4 housses siège voiture
⇒ 6 oeufs
⇒ 1 slip femme
⇒ 1 boite alimentation chat

Il est clair que le produit intitulé “FEMME” est difficilement identifiable s’il n’est pas associé à un complément d’information.
Ainsi, les champs de saisie pour la **partie tickets de caisse** sont les suivants :

<table>
<thead>
<tr>
<th>Nature de la dépense</th>
<th>Complément de la dépense</th>
<th>Montant de la dépense</th>
<th>Type de magasin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 pack Yoplait</td>
<td>Yaourts</td>
<td>20,00</td>
<td>Supermarché</td>
</tr>
</tbody>
</table>

Tandis que ceux de la partie du **tableau de dépenses** sont les suivants :

<table>
<thead>
<tr>
<th>Quantité/unité</th>
<th>Nature de la dépense</th>
<th>Montant de la dépense</th>
<th>Type de magasin</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 packs</td>
<td>lait</td>
<td>18,00</td>
<td>Épicerie</td>
</tr>
</tbody>
</table>

Ainsi, les informations collectées pour un produit donné sont **hétérogènes** car elles proviennent de 2 supports différents : ticket de caisse ou tableau de dépenses.

D’autre part, il suffit d’examiner plusieurs tickets de caisse provenant de commerces différents pour s’apercevoir que l’information pour un même produit est **non** structurée contrairement à la variable activité. Les produits issus des tickets de caisse peuvent être des marques de produit, des chiffres, des abréviations et/ou des codes. Nous verrons dans le paragraphe 6 les différents problèmes rencontrés face à un libellé aussi hétérogène.

### 1.3 Les caractéristiques de la variable et leurs évolutions dans le temps

La saisie des carnets de l’enquête BdF précédente a permis de montrer que déjà en 1994-1995, les dépenses des ménages inscrites dans ces carnets provenaient majoritairement de tickets de caisse (à 56% environ). D’autre part, le complément de dépense était présent pour un tiers des dépenses provenant des tickets de caisse. Enfin, le nombre de dépenses par carnet représente un volume considérable. En novembre 1998, lors du premier test terrain, la tendance s’était accentuée.

<table>
<thead>
<tr>
<th>Fichier</th>
<th>Carnets 1994-95</th>
<th>Test1(^{15}) nov. 1998</th>
<th>Test2 juin 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre de dépenses par carnet</td>
<td>57</td>
<td>65</td>
<td>89</td>
</tr>
<tr>
<td>Pourcentage de dépenses provenant de tickets de caisse</td>
<td>55,75%</td>
<td>60,9%</td>
<td>63,01%</td>
</tr>
<tr>
<td>Pourcentage de dépenses issues de tickets de caisse comportant un complément</td>
<td>34,66%</td>
<td>37,22%</td>
<td>32,11%</td>
</tr>
</tbody>
</table>

\(^{15}\) Le premier test a eu lieu en novembre 1998 dans les DR de Lille, Lyon et Nancy. Le deuxième test s’est déroulé en juin 1999 dans les DR de Lyon, Orléans et Rennes.

**INSEE Méthodes**
Ces caractéristiques ont sans doute des conséquences sur les taux de codification. La présence d'un complément d'information pour un produit provenant d'un ticket de caisse aura un rôle important autant du point de vue de la codification automatique que celui de la reprise manuelle des non-codés automatiquement.

2 La nomenclature de la variable produits

2.1 La nomenclature initiale de l'enquête de 1994-95

La nomenclature officielle en 1994-95 se divisait en 916 postes, écrits sur 5 positions, emboîtés et divisés en 9 grandes fonctions :
Cependant, toutes les fonctions ne contenaient pas le même nombre de postes : l'alimentation était très nettement majoritaire sur les autres : les postes de la fonction "Produits alimentaires, boissons et tabac" représentaient 29% de l'ensemble des postes.

2.2 Le changement de nomenclature en janvier 1999

En 1997, Eurostat avait recommandé, dans des soucis de comparaisons internationales, une nomenclature de produits pour l'enquête BDF, la COICOP-HBS. Celle-ci comprend 13 grandes fonctions, s'écrit sur 5 positions et comprend des postes emboîtés comme ceux de la nomenclature de 1994-1995. Néanmoins, afin de garder une continuité de nomenclature avec les précédentes éditions de BDF, nous avons rajouté une position et sommes arrivées à 6 positions. Les 13 fonctions de la nomenclature européenne sont celles-ci :

---

16 cf. bibliographie
17 Classification Of Individual COnsumption by Purpose - Household Budget Survey

INSEE Méthodes 26
<table>
<thead>
<tr>
<th>Numéro de la fonction</th>
<th>Intitulé de la fonction</th>
<th>Exemples de produit</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Produits alimentaires et boissons non alcoolisées</td>
<td>Pain, yaourts, jus d’orange</td>
</tr>
<tr>
<td>02</td>
<td>Boissons alcoolisées, tabac et stupéfiants</td>
<td>Cigarettes, champagne</td>
</tr>
<tr>
<td>03</td>
<td>Articles d’habillement et chaussures</td>
<td>Pantalon pour homme, jupe pour fille, soutien-gorge, tissu, cravate, espadrilles pour garçon, pressing pour les vêtements</td>
</tr>
<tr>
<td>04</td>
<td>Logement, eau, électricité, gaz et autres combustibles</td>
<td>Loyer, GDF, charbon, papier peint</td>
</tr>
<tr>
<td>05</td>
<td>Ameublement, équipement ménager en entretien courant de la maison</td>
<td>Lit, sonnette, cafetière électrique, voilage, assiette, piles, savon, pressing pour le linge d’ameublement, baby-sitter</td>
</tr>
<tr>
<td>06</td>
<td>Santé</td>
<td>Sirop contre la toux, sparadrap, lunettes, médecin, frais d’hospitalisation</td>
</tr>
<tr>
<td>07</td>
<td>Transports</td>
<td>Voiture, vélo, pneus, vidange, huile moteur, pêage, billet de train, carte orange</td>
</tr>
<tr>
<td>08</td>
<td>Communications</td>
<td>Facture de téléphone, timbres</td>
</tr>
<tr>
<td>09</td>
<td>Loisirs et culture</td>
<td>Télévision, disques, sac de couchage, planche à voile, monopoly, aquarium, bulbe de tulipe, flipper, cinéma, bowling, vacances, journaux, agrafes, stylo</td>
</tr>
<tr>
<td>10</td>
<td>Enseignement</td>
<td>Frais de scolarité</td>
</tr>
<tr>
<td>11</td>
<td>Hôtels cafés et restaurants</td>
<td>Hôtel, repas, café au comptoir</td>
</tr>
<tr>
<td>12</td>
<td>Autres biens et services</td>
<td>Coton, nourrice agréée, coiffeur, déodorant, bouillotte, réveil, cartable, assurances</td>
</tr>
<tr>
<td>13</td>
<td>Postes hors champ de la comptabilité nationale</td>
<td>Impôts, toiture, cadeaux, remboursement de prêts, permis de chasse</td>
</tr>
</tbody>
</table>

Remarque : les exemples de produits en gras de ce tableau représentent des produits ayant changé de fonction par rapport à la nomenclature de 1994-95.

Les exemples dans la dernière colonne de ce tableau prouvent que le changement par rapport à la nomenclature n’était pas mineur.
3. Pourquoi utiliser Sicore pour coder le produit ?


L’examen du codage manuel de 1994-1995 nous poussait aussi dans cette direction afin de rendre plus homogène le codage.

*Par exemple, en 1994-95, le « savon » pouvait être codé manuellement dans la fonction 4 Equipement du logement ou dans la fonction 8 Autres biens et Services s’il se présentait sous la forme de gel douche ou pas.*

Pour l’édition 2000-2001 de l’enquête, et pour la 1ère fois depuis que l’enquête Bdf existe, l’outil de codification automatique Sicore est utilisé pour coder, entre autres 18, les produits achetés et inscrits dans les carnets.

4. La constitution des connaissances

4.1 Le premier Fichier d’apprentissage brut des produits

Pour utiliser Sicore, il faut constituer en particulier un fichier d’apprentissage (Fab). Là encore, sur l’enquête Bdf tout comme sur l’enquête EdT, aucun fichier n’existait. En effet, les dépenses des carnets de l’enquête Bdf précédente n’avaient pas été saisies.


18 D’autres variables sont en outre codées par Sicore : tout d’abord, la CSP au sein du tronc commun ; ensuite le type de magasin dans le carnet. Cette dernière information pose nettement moins de problèmes de codification que le produit de la dépense ; nous n’en parlerons donc pas ici pour nous focaliser sur le produit.
A l’été 1998, l’expert variable n’étant pas nommé, c’est l’expert Sicore et le responsable d’enquête qui ont constitué les bases de connaissances de la variable produit dans leurs premières versions. A ces libellés de postes de nomenclature, ont été rajoutés des exemples similaires à ceux de la colonne du tableau de la nomenclature COICOP-HBS (cf. paragraphe II - 2.2) et des libellés de marques de produit issus d’un site de commerce électronique. A ce stade, le Fab comptait 3 179 lignes.

4.2 L’amélioration standard des connaissances pour la variable produit

Pour la variable produit, la boucle Sicore a permis l’enrichissement des connaissances suivant plusieurs axes à partir de trois sources d’informations : les carnets de l’enquête de 1994-1995 et ceux des deux tests terrain :

- Le choix des caractères blancs ou vides

- Le choix, l’intégration, la mise en ordre et l’élimination de synonymes
  On dénombre deux types de synonymes. Tout d’abord, on trouve les synonymes “classiques”.
  
  Par exemple,
  
  "BOCAL" = "CONSERVE",
  "RIDEAU" = "VOILAGE".

  Ensuite, viennent les synonymes créant de nouvelles expressions et aboutissant à une réduction du nombre de mots au sein du libellé normalisé.
  
  Par exemple,
  
  "HARICOT VERT" = "HARVER",
  "FROMAGE BLANC" = "FROMBLAN",
  "SOUS VETEMENT" = "SOUSVET".

  Ces expressions ne correspondent pas forcément aux abréviations rencontrées sur les tickets de caisse.

- Le choix des mots vides

  Par exemple, toutes les quantités sont éliminées lors de la phase de normalisation : "GRAMME", "PACK", "LITRE", "LOT" ...


20 En l’occurrence, Frédérique Deschamps
Certains mots sont éliminés mais cette élimination a entraîné d’autres modifications des connaissances.

Par exemple, la définition de “PAIRE” en mot vide entraîne un codage automatique similaire des libellés “UNE PAIRE DE CHAUSSONS” en tant que pantoufles et “CHAUSSE” en tant que pâtisserie. Si on veut continuer à éliminer le mot “PAIRE”, il faut ajouter un synonyme au bon endroit dans la liste des synonymes : "PAIRE CHAUSSON" = "PAIRCHAUSSON".

- Les paramètres de l’apprentissage
Ces premiers codages prenaient en compte 4 mots de 12 caractères chacun après normalisation.

- L’extension du nombre de lignes au sein du Fab constitue l’essentiel du travail de constitution des connaissances.

Sicore a été lancé sur les différents fichiers disponibles, soit issus de la saisie de carnets de 1994-1995, soit issus de la saisie des carnets des deux tests.

- Les fichiers de libellés de produits associés à un code de référence (manuel) ont été traités de trois manières différentes selon le code automatique fourni éventuellement par Sicore.

1) Pour les codés conformes, nous avons éventuellement tiré un échantillon, puis nous avons étudié tous les libellés concernés. Dans les cas où le code automatique était erroné, nous avons corrigé les connaissances en conséquence.

2) Nous avons intégrés les codés non conformes dans le FAB après arbitrage sur le code.

3) Pour les libellés non codés, c’est-à-dire sans code automatique valide fourni par Sicore (les erreurs de redondance plus les échecs de codage), étant donné le nombre de postes de la nomenclature de 1994-1995, nous avons appliqué une méthode de traitement par mot-clef. Cette méthode consiste, pour un poste précis de la nomenclature, à rechercher tous les libellés dont un des mots est en relation avec ce poste.

Pour plus de renseignements sur Sicore et ses connaissances, on se reportera à l’annexe n°1.
Par exemple, pour le poste " saucisses fraîches, cuites ou fumées "; les mots-clés cherchés dans ce fichier des non codés sont :

Cette méthode permet de traiter rapidement une grande partie des libellés non codés se référant au poste choisi et évite la manipulation fastidieuse de la grande quantité de papier que représente la nomenclature. Elle présente deux inconvénients : d’une part, le nombre de postes de la nomenclature étant important, cette tâche est répétitive et nécessite de l’imagination ; d’autre part, elle a pour conséquence la récupération de libellés comportant un des mots-clés, mais sans aucun rapport avec le poste choisi.

Sur l'exemple précédent, nous avons récupéré le libellé suivant :
" TICKET DE BUS POUR STRASBOURG " ou " LIVRE SUR TOULOUSE ".
Nous les avons alors codés sous le code ticket de bus ou livre.

- Pour les fichiers de libellés de produits sans code de référence (pas de code automatique, ni de code manuel de 1994-95), le traitement est plus long, car il n'y a pas de référence à la nomenclature autre que celle issue de Sicore.

① Pour les codés automatiquement par Sicore, nous avons vérifié que le code fourni par Sicore était juste. Le cas échéant, nous corrigions les bases de connaissances.

② Pour les non codés automatiquement par Sicore, deux méthodes ont été employées.
Au début, un codage manuel de ces différents produits a été réalisé et cette liste de produits associés à un code était simplifiée et intégrée au Fab. Puis, lassées du fait du volume de la nomenclature, la fréquence d’apparition des libellés de produits a été utilisée pour traiter ces libellés. Le Fab a été complété par ces libellés de produits, codés manuellement par nos soins, non existants jusqu’alors dans le Fab et dont la fréquence d’apparition était supérieure à 1.
4.3 L’utilisation optimale des informations auxiliaires

Des tests ont été menés sur les carnets saisis du test 1 de novembre 1998 afin de mesurer l'impact de la présence d'informations sur le codage automatique : le complément de dépense et la quantité.

Nous avons testé si ces informations étaient mieux placées en début ou en fin de libellés et si elles étaient nécessaires. Les résultats des multiples codages sont les suivants :

<table>
<thead>
<tr>
<th>Type de dépenses</th>
<th>Total</th>
<th>Codés</th>
<th>Redondants</th>
<th>Non codés</th>
</tr>
</thead>
<tbody>
<tr>
<td>avec le complément de dépense après le libellé de produit et la quantité avant</td>
<td>18 998</td>
<td>8 729 (45,95%)</td>
<td>4 736 (24,94%)</td>
<td>5 530 (29,11%)</td>
</tr>
<tr>
<td>avec le complément de dépense après le libellé de produit et sans la quantité</td>
<td>18 998</td>
<td>8 886 (46,77%)</td>
<td>4 711 (24,80%)</td>
<td>5 398 (28,41%)</td>
</tr>
<tr>
<td>avec le complément de dépense et la quantité avant le libellé de produit</td>
<td>18 998</td>
<td>8 241 (43,38%)</td>
<td>4 814 (25,35%)</td>
<td>5 940 (31,27%)</td>
</tr>
<tr>
<td>avec le complément de dépense avant le libellé de produit et sans la quantité</td>
<td>18 998</td>
<td>8 398 (44,20%)</td>
<td>4 789 (25,21%)</td>
<td>5 808 (30,57%)</td>
</tr>
<tr>
<td>avec le libellé seul, sans complément ni quantité</td>
<td>18 998</td>
<td>9 394 (49,45%)</td>
<td>4 458 (23,47%)</td>
<td>5 143 (27,07%)</td>
</tr>
</tbody>
</table>

De ces tests de codage, il ressort que, si on intègre le complément de dépense, il vaut mieux qu'il soit placé après le libellé de la dépense. Quant à la quantité, le codage est meilleur lorsque le libellé ne contient pas de quantité.

4.4 Un nombre de mots plus adapté

Lors de l'étude réalisée sur les carnets saisis du test 1 de novembre 1998 (cf. paragraphe II - 4.3), nous avons aussi remarqué que les meilleurs taux de codage automatique étaient obtenus en ne prenant ni le complément de dépense ni la quantité (cf. tableau précédent). Ce résultat était surprenant : en général, des précisions permettent de mieux coder.

L'explication de ce problème vient du nombre de mots pris en compte dans les paramètres de codage Sicore : 4 dans ces codages. En effet, le complément d'information ne sert que quand le libellé du ticket de caisse est très court (sur 1 mot par exemple) ce qui est rare compte tenu des informations parasites qui font souvent
partie du libellé provenant d’un ticket de caisse, même si certaines sont éliminées par la normalisation.

Pour tenir compte du complément de dépense, nous avons alors réalisé des tests sur le nombre de mots du libellé normalisé comprenant le complément à prendre en compte. Jusqu’ici, nous prenions en compte un libellé de 4 mots, ce qui était insuffisant. Les résultats de ces tests montrent que considérer un libellé à 6 mots permet de coder plus de produits que prendre en compte 4 mots d’un libellé. En revanche, il ne sert à rien, a priori, de prendre en compte 8 mots.

<table>
<thead>
<tr>
<th>Nombre de mots du libellé normalisé</th>
<th>Codés</th>
<th>Redondants</th>
<th>Echecs</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 mots</td>
<td>8 886 (46,77%)</td>
<td>4 711 (24,80%)</td>
<td>5 398 (28,41%)</td>
</tr>
<tr>
<td>6 mots</td>
<td>11 155 (58,72%)</td>
<td>2 524 (13,29%)</td>
<td>5 316 (27,958%)</td>
</tr>
<tr>
<td>8 mots</td>
<td>11 155 (58,72%)</td>
<td>2 524 (13,29%)</td>
<td>5 316 (27,958%)</td>
</tr>
</tbody>
</table>

⇒ Au bilan, les codages prennent 6 mots dans le libellé issu de la normalisation.

### 4.5 La non-utilisation des variables annexes

Comme pour la variable activité, il a été question un temps de présenter une partie du carnet de dépenses (le tableau de dépenses) avec des variables annexes du type: destination de la dépense (pour l’individu, pour le ménage, pour la famille, pour d’autres personnes), lieu de dépense (grande surface, petite surface, autre cas).

Mais, compte tenu du comportement actuel des ménages à grouper de plus en plus leurs achats (par exemple sur le test 2 de juin 1999, le pourcentage de produits provenant de tickets de caisse s’élève à 63%), le problème des tickets de caisse demeurait.

Enfin, le fait d’avoir le lieu de dépense en variable précédée (en quelques modalités) comme le but de l’activité dans l’enquête EdT, limitait les possibilités d’exploitation et de collaboration avec les autres services divisions de l’INSEE : la division des prix à la consommation et les divisions Services et Commerces s’intéressent aux lieux d’achats. Ainsi, nous avons abandonné cette piste, mais nous gardons en tête l’idée des règles de codage (cf. paragraphe II - 8.2).
5. Les conséquences de l'utilisation de Sicore

5.1 Les modifications dans la présentation du carnet

La présentation des carnets s’est vue simplifiée par rapport à la mise en page des carnets de 1995 car nous avons tout d’abord essayé de faciliter le codage automatique des dépenses\(^{22}\). Ainsi, dans le carnet, le ménage n’inscrit que ses dépenses quotidiennes.

Ensuite, un seul type de carnet par rapport à l’enquête précédente a été utilisé : toutes les personnes du ménage ont un carnet du même type à remplir\(^{23}\).

En novembre 1998, le test 1 avait aussi pour objectif de tester la présentation des carnets. En effet, les carnets utilisés étaient de deux types. Les premiers comportaient une colonne “quantité” isolée de la colonne “nature de la dépense” sur les pages concernant les dépenses isolées (tableau de dépense), tandis que les seconds fusionnaient ces deux informations. Au bilan, nous avons conservé un tableau de dépenses avec la colonne “quantité/unité” isolée.

Il a été décidé aussi de faciliter le plus possible le travail de remplissage du carnet par l’individu en lui laissant de la place pour inscrire ses dépenses sur le tableau de dépenses ou compléter l’information fournie par les tickets de caisse : le format des carnets a donc été changé par rapport à celui adopté en 1994-95 ; on ne peut plus réellement parler de carnet de dépenses, c’est davantage un cahier de dépenses.

\[
\begin{array}{l}
\text{Ainsi le tableau de dépenses du carnet comprend 4 champs :} \\
\quad \text{quantité et unité,} \\
\quad \text{nature de la dépense,} \\
\quad \text{montant de la dépense,} \\
\quad \text{type de magasin.}
\end{array}
\]

Et, lors de la formation des gestionnaires d’enquête qui forment eux-mêmes les enquêteurs, l’attention s’est portée plus particulièrement sur le complément de dépense à ajouter dans le cas de produits provenant de tickets de caisse.

\(^{22}\) Deux types de tableaux ont été transférés sur CAPI ; ils concernaient l’autoconsommation et les repas pris à l’extérieur du domicile du titulaire du carnet et de ses enfants de moins de 14 ans (éventuels).

5.2 L’identification difficile de certains produits

Sicore ne distingue pas les majuscules des minuscules et ne tient pas compte de la ponctuation. Ceci a pour conséquence que certains produits ne peuvent pas être codés de manière certaine ; des choix ont donc dû être faits.

*Par exemple, "PATE" peut être codé vers plusieurs codes en 01 (pâtes Panzani, pâti en terrine ou pâte feuilletée) et vers le 09 (Pâté, pour chiens). On peut citer aussi "HACHE" qui peut être codé dans la fonction 09 (horticulture) ou 01 (viande).*

D’autre part, le fait que l’on n’utilise pas de variable annexe pour coder automatiquement les produits entraîne une identification plus délicate de certains produits.

*Par exemple, le terme "COTON" peut correspondre à plusieurs produits : le coton à tricoter (03), le coton à démaquiller (12). Autre exemple, "CREME" peut être fraîche (01), Nivea (12). "CARTOUCHE" fait référence aux cigarettes (02), à l’encre et à la chasse (09). De même, "HUILE" peut être alimentaire (01) ou pour la voiture (07).*

En revanche, un codeur aboutirait à arbitrer ces cas facilement à partir des informations auxiliaires (comme le lieu de dépense ou la quantité) disponible en clair pour lui. Des traitements spécifiques sont donc envisagés mais uniquement à la fin du traitement des carnets. À ce moment là, une règle de codage sera utilisée ; elle affinera le codage du produit à partir du magasin.

*Par exemple, "COTON" associé au lieu de dépense "PHARMACIE" pourra être codé à la bonne place en 12.*
5.3 L’adaptation à la nomenclature

Il a été arbitré des codes pour des libellés ambigus pour la codification automatique par Sicore.

Par exemple, le libellé “ PATES ” sera codé comme les coquillettes Panzani et le libellé “ PATE ” comme les rillettes.

D’autre part, pour les libellés généraux, ne permettant pas de déterminer les 6 chiffres du code, le caractère “ * ” a été inséré dans des codes. Ces cas nécessitent absolument la présence du complément de dépense pour être codés au plus juste. Sans ce dernier, on reste à un niveau général agrégé²⁴.

Par exemple, “ BOUCHERIE ” ne permet pas de déterminer quelle viande a été achetée ; on lui a donc attribué le code 0112**. De la même manière, a été créé le code à 6 étoiles “ ***** ” afin de coder des libellés ne contenant aucune information précise, comme “ TICKET DE CAISSE ”, “ LIQUIDE ”, ...

Enfin, il arrive que certains libellés comprennent plusieurs produits de nature différente, et même appartenant à des fonctions différentes. Dans ce cas de multiplicité de produits, le plus grand nombre de chiffres en commun pour les codes est gardé et complété par des étoiles.

Par exemple, “ FLEURS DENTIFRICE ” sera codé également en *****. Mais, “ OEUFS YAOURTS ” est codé en 0114**.

Lors de la future exploitation des fichiers, il faudra traiter ces cas.

5.4 Le dédoubllement du Fab des produits

L’examen attentif des produits issus des tickets de caisse met en évidence des libellés ayant une forme différente des libellés issus du tableau de dépense (cf : II - 6).

L’expert variable a alors décidé de dédoubler le Fab produits afin de tenir compte des spécificités des libellés des tickets de caisse :
- elle a supprimé du Fab des produits issus des tickets de caisse, les libellés de produits qui avaient très peu de chance d’être achetés dans des grandes surfaces
- elle a alors concentré ses efforts sur les produits alimentaires qui sont les produits les plus fréquents sur les tickets de caisse.

²⁴ Cf. le paragraphe 6.2
5.5 Une organisation des traitements des carnets complètement éclatée

Compte tenu du volume des carnets à saisir\(^{25}\), tous les ateliers de saisie en exercice en mai 2000 ont dû travailler sur les carnets de l’enquête BdF 2000-2001. Étant au nombre de 13 contre 18 DR de collecte, des regroupements de DR ont été effectués selon la taille des GSAS.

Pour la reprise des carnets de dépenses, le volume de rejets à traiter\(^{26}\) et l’organisation de la saisie ont eu des conséquences non négligeables sur l’organisation de la reprise qui est en fait, tout comme la saisie, bien différente de celle choisie pour EdT.

Nous avons pu maintenir le même nombre de sites de reprise : 13 DR réalisent ce travail grâce à un outil de reprise\(^{27}\). En revanche, le travail de codage manuel peut être fait par des équipes différentes de celles qui effectuent la collecte (les DEM) ou la saisie (GSAS).

Cette organisation ne facilite
- ni l’homogénéisation des codes manuels attribués aux produits rejetés par Sicore.
- ni le travail de l’expert variable qui doit jongler avec les 13 interlocuteurs différents au minimum\(^{28}\).

---

\(^{25}\) Cf. annexe n°5

\(^{26}\) Ce volume de rejets à traiter a été calculé à partir d’une prévision de taux de codage relativement faible mais réaliste à l’époque (cf. annexe n°5)

\(^{27}\) Cet outil de reprise a été développé par Anne-Marie Duval au CNI de Lille (cf. annexe n°5)

\(^{28}\) En effet, dans certaines DR, le travail de reprise peut être effectué à la fois par la DEM et par le GSAS ce qui multiplie les interlocuteurs.
6. Les conséquences des caractéristiques des dépenses recueillies

L’examen des libellés provenant à la fois des tableaux de dépenses et des tickets de caisse nous a permis de déceler des problèmes.

6.1 L’orthographe des libellés et les abréviations

- Des fautes d’orthographe sont fréquentes dans la partie tableau de dépenses et gênent la codification automatique quand elles n’aboutissent pas à un échec du codage.
  
  Par exemple, les termes “YAOURT”, “BEEFSTEACK” ont de nombreuses variantes.

- Les tickets de caisse décrivent souvent les produits grâce à des abréviations et des codes internes au magasin.
  
  Par exemple, nous avons déniché les libellés suivants assortis de leur complément de dépense :
  
  “P ELEA 1.2 ECR” LAIT
  “2X250ML SH2/1 SEC” SHAMPOOING

De plus, “COQUI” peut amener aux produits suivants : des coquillettes ou une coquille de poisson ; et “BOUCHE” peut renvoyer à un article de boucherie ou à une bouchée à la reine.

Les abréviations de ”YAOURT”, ”BEEFSTEACK”, ”POMMES DE TERRE” sont aussi nombreuses.

► La majorité des fautes d’orthographe et des abréviations, en tous cas les recensées, ont été rajoutées sous forme de synonymes. Les codes internes ont eux été intégrés dans la liste des mots vides relative aux connaissances.

6.2 La dilution de l’information

Pour certains libellés, l’information principale qui permettrait un codage automatique facilement est diluée par des informations parasites, comme la couleur, la marque, la quantité, le type de produit, ... Ces problèmes sont particulièrement vrais pour les libellés issus des tickets de caisse. En effet, les tickets de caisse présentent souvent l’inconvénient de décrire le produit en une multitude de mots à cause de la présence du complément de dépense et de la quantité et souvent de la marque.

Par exemple,
- nous avons trouvé les libellés suivants assortis de leur éventuel complément de dépense :

```
" TENNIS REVERS CUP'S "
" GRESSINS ALICE "
" MARTI RIZ 1/2 LG ETOVE "
" PAIN 24 TRS FIN "
" 1 ER PRIX1/2 ECRIE "
```

- nous avons aussi trouvé les libellés suivants sans complément de dépense :

```
" P JEAN GARN BOUCHEE REINE "
" BA FRUIT SOLEIL 6X1 YAOUR "
" FL 750ML PAI VAISELLE "
```

⇒ L’augmentation du nombre de mots du libellé normalisé pris en compte lors du codage a permis de régler en partie ce problème.

6.3 L’absence d’information pertinente

Au contraire du cas précédent, l’information principale qui permettrait un codage automatique facilement n’apparaît pas dans le libellé faute d’un complément de dépense. Une fois de plus, ces problèmes sont particulièrement vrais pour les libellés issus des tickets de caisse.

Par exemple, les libellés suivants sont trouvés :
```
" NON ALIMENTAIRE "
" FR ET LEGU "
" FRIT M "
```

⇒ L’acceptation de codes agrémentés de " * " a permis de régler ce problème.

6.4 Les non-dépenses présentes dans les tickets de caisse

Les tickets de caisse présentent des lignes qui ne correspondent pas à des dépenses stricto sensu : les annulations de produit précédemment saisi, les remises et ristournes, les consignes, les bons de réduction, les bons d’achat, ...

Les libellés suivants ont été trouvés :
```
AUBERGINE
ANNULATION ARTICLE
BON REDUCT.
REMISE CARTE PASS FRAISES 250G
```

⇒ Un code particulier a été attribué à ces différents libellés : 999999. Lors de l’exploitation, nous saurons ainsi qu’il faudra soit retrancher le montant associé, soit l’éliminer purement et simplement.
7. Au Bilan

7.1 Les résultats de codage au cours de la préparation des connaissances

① Chronologiquement, la préparation des connaissances a commencé par l'exploitation des carnets de 1994-1995. A cette époque, Sicore prenait en compte les 4 premiers mots du libellé normalisé. Avant l'utilisation des libellés du test 1, les résultats de codage de Sicore sur un fichier de référence29 étaient les suivants :

<table>
<thead>
<tr>
<th>Résultats du codage automatique du fichier de référence</th>
<th>Nombre de lignes</th>
<th>Pourcentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codés</td>
<td>20 395</td>
<td>53,21%</td>
</tr>
<tr>
<td>Non codés</td>
<td>17 931</td>
<td>46,79%</td>
</tr>
<tr>
<td>Total</td>
<td>38 326</td>
<td>100%</td>
</tr>
</tbody>
</table>

② Avec la même version des connaissances qui a fourni ces résultats, voici les résultats de codage des produits du test 1 de novembre 1998 :

<table>
<thead>
<tr>
<th>Résultats du codage automatique des produits de novembre 1998</th>
<th>Nombre de lignes</th>
<th>Pourcentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codés</td>
<td>8 767</td>
<td>46,15%</td>
</tr>
<tr>
<td>Non codés</td>
<td>10 231</td>
<td>53,85%</td>
</tr>
<tr>
<td>Total</td>
<td>18 998</td>
<td>100%</td>
</tr>
</tbody>
</table>

Ces résultats de codage sont moins bons que ceux obtenus précédemment sur le fichier de référence des produits de 1994-95. Cela s'explique par les nouveaux produits apparus depuis cette époque et l'augmentation du pourcentage des libellés provenant des tickets de caisse.

③ Par la suite, les libellés du test 1 ayant servi à l'enrichissement des bases de connaissances et à prendre la décision d'utiliser 6 mots dans le libellé normalisé, nous avons recodé le fichier de référence. Ce genre d'opérations nous permet de mesurer les progrès engendrés par l'exploitation des libellés du test 1. Les résultats de codage de ce fichier ont donc été tout à fait encourageants ; nous avons gagné plus de 10 points dans le codage (53,21% contre 66,33%) :

---

29 Ce fichier de 38 326 lignes était issu des carnets de 1994-95. Il n'a pas été utilisé pour enrichir les connaissances de Sicore. Il nous a servi de référence.
<table>
<thead>
<tr>
<th>Résultats du codage automatique du fichier de référence</th>
<th>Nombre de lignes</th>
<th>Pourcentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codés</td>
<td>25 421</td>
<td>66,33%</td>
</tr>
<tr>
<td>Non codés</td>
<td>12 905</td>
<td>33,67%</td>
</tr>
<tr>
<td>Total</td>
<td>38 326</td>
<td>100%</td>
</tr>
</tbody>
</table>

Cette même version des connaissances a servi aussi à coder les produits du test 2 de juin 1999.

<table>
<thead>
<tr>
<th>Résultats du codage automatique des produits de juin 1999</th>
<th>Nombre de lignes</th>
<th>Pourcentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codés</td>
<td>9 118</td>
<td>60,24%</td>
</tr>
<tr>
<td>Non codés</td>
<td>6 017</td>
<td>39,76%</td>
</tr>
<tr>
<td>Total</td>
<td>15 135</td>
<td>100%</td>
</tr>
</tbody>
</table>

Même si les taux de codage étaient inférieurs à ceux obtenus sur le fichier de référence (66,33% contre 60,24%), ces résultats ont été plutôt satisfaits compte tenu du fait que le dernier enrichissement du Fab a été réalisé à partir de produits de novembre 1998, c’est à dire de certains produits saisonniers spécifiques à l’automne. Or, le test 2 s’étant déroulé au printemps, d’autres produits spécifiques ont pu apparaître sans avoir pu être codés.

Un sous-fichier du test 230 a permis également de fournir les premiers résultats de codage après la décision de dédoubler le Fab des produits. Ainsi, à cette époque, nous avons pu constater l’écart de codage égal à 25 points entre produits issus de tickets de caisse et produits issus des tableaux de dépense.

<table>
<thead>
<tr>
<th>Résultat du codage automatique des produits de juin 1999</th>
<th>produits des tickets de caisse</th>
<th>produits des tableaux</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nombre de lignes</td>
<td>%</td>
</tr>
<tr>
<td>Codés</td>
<td>2 896</td>
<td>51,96%</td>
</tr>
<tr>
<td>Non codés</td>
<td>2 678</td>
<td>48,04%</td>
</tr>
<tr>
<td>Total</td>
<td>5 574</td>
<td>100%</td>
</tr>
</tbody>
</table>

---

30 Ce sous fichier est constitué des produits recueillis dans la DR de Rhône Alpes lors du test 2 de juin 1999. Deux autres DR ont participé à ce test.
7.2 - Le premier environnement utilisé pour l’enquête en juin 2000

L’enquête a démarré sur le terrain en mai 2000. En juin 2000, les premiers codages ont été effectués. Les caractéristiques générales de ce premier environnement sont les suivantes :

<table>
<thead>
<tr>
<th>Environnement</th>
<th>TABLEAU DE DEPENSES</th>
<th>TICKET DE CAISSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre de mots</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Nombre de lettres par mots</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>FAB :</td>
<td>37 394 lignes</td>
<td>31 299 lignes</td>
</tr>
<tr>
<td>Nombre de codes comportant au moins une *</td>
<td>462</td>
<td>80</td>
</tr>
<tr>
<td>Nombre de synonymes</td>
<td>1 193</td>
<td>1 193</td>
</tr>
</tbody>
</table>

Voici la répartition des libellés de produits par grande fonction de la nomenclature :

<table>
<thead>
<tr>
<th>Environnement</th>
<th>TABLEAU DE DEPENSES</th>
<th>TICKET de CAISSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fonction</td>
<td>nombre de libellés</td>
<td>nombre de libellés</td>
</tr>
<tr>
<td>01</td>
<td>22 696 (60,69%)</td>
<td>20 998 (67,09%)</td>
</tr>
<tr>
<td>02</td>
<td>1 398</td>
<td>1 187</td>
</tr>
<tr>
<td>03</td>
<td>1 321</td>
<td>1 204</td>
</tr>
<tr>
<td>04</td>
<td>418</td>
<td>274</td>
</tr>
<tr>
<td>05</td>
<td>2 908</td>
<td>2 585</td>
</tr>
<tr>
<td>06</td>
<td>491</td>
<td>146</td>
</tr>
<tr>
<td>07</td>
<td>641</td>
<td>206</td>
</tr>
<tr>
<td>08</td>
<td>129</td>
<td>16</td>
</tr>
<tr>
<td>09</td>
<td>3 551</td>
<td>2 638</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>576</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>2 253</td>
<td>1 786</td>
</tr>
<tr>
<td>13</td>
<td>573</td>
<td>147</td>
</tr>
<tr>
<td>99</td>
<td>37</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>37 394</td>
<td>31 299</td>
</tr>
</tbody>
</table>

Il est logique de voir que le nombre de libellés dans le Fab produits issus des tickets de caisse dans le domaine de l’enseignement est nul et que la part de l’alimentaire (fonction 01) dans ce Fab est la plus importante (67,09%).
7.3 **Les premières vagues de l’enquête**

7.3.1 **Les caractéristiques des libellés recueillis au début de l’enquête**

Par rapport aux caractéristiques des carnets du test 2 de juin 1999, le pourcentage de produits des vagues 1 et 2 provenant des tickets de caisse a légèrement diminué : 63,01% contre 61,13 % en 2000 ce qui peut être une bonne chose pour le taux de codage automatique puisque la faiblesse du codage demeure sur les produits des tickets de caisse.

<table>
<thead>
<tr>
<th></th>
<th>Nombre de carnets</th>
<th>Nombre de produits</th>
<th>Provenance du produit</th>
<th>Nombre de produits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vague 1</td>
<td>2 002</td>
<td>137 740</td>
<td>Ticket</td>
<td>83 981 (60,97%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Manuscrit</td>
<td>53 759 (39,03%)</td>
</tr>
<tr>
<td>Vague 2</td>
<td>1 785</td>
<td>129 309</td>
<td>Ticket</td>
<td>79 322 (61,34%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Manuscrit</td>
<td>49 987 (38,66%)</td>
</tr>
<tr>
<td>Total</td>
<td>3 787</td>
<td>267 129</td>
<td>Ticket</td>
<td>163 303 (61,13%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Manuscrit</td>
<td>103 746 (38,87%)</td>
</tr>
</tbody>
</table>

D’autre part, comme lors des formations des gestionnaires d’enquête, nous avions insisté sur l’importance d’un complément de dépense dans le cas d’un produit issu des tickets de caisse, nous avons voulu connaître ce nombre de produits issus des tickets de caisse qui comportaient un complément de dépense. Dans les carnets du test 2, le complément existait dans seulement 32,11% des cas. Et nous avons été agréablement surpris avec un taux au-dessus de 40 % pour les deux premières vagues de l’enquête.

<table>
<thead>
<tr>
<th></th>
<th>Nombre de produits issus des tickets de caisse</th>
<th>Existence de complément de dépense</th>
</tr>
</thead>
<tbody>
<tr>
<td>vague 1</td>
<td>83 981</td>
<td>38 379 (45,70%)</td>
</tr>
<tr>
<td>vague 2</td>
<td>79 322</td>
<td>33 111 (41,74%)</td>
</tr>
<tr>
<td>Total</td>
<td>163 303</td>
<td>71 490 (43,78%)</td>
</tr>
</tbody>
</table>

7.3.2 **L’enrichissement des connaissances en cours d’enquête**

Comme pour l’enquête EdT, l’expert variable enrichit les Fab produits au fur et à mesure des vagues afin de bien améliorer le taux de codage automatique et par

---

31 La vague 1 couvrait la période du 9 mai au 18 juin 2000 tandis que la vague 2 s’étalait du 19 juin au 30 juillet 2000 en métropole.
conséquent de diminuer le volume des produits à coder manuellement dans les DR assurant la reprise. Elle vérifie aussi le codage automatique sur un échantillon de libellés.

L’enrichissement des fichiers de référence porte sur :
- l’aspect quantitatif : les libellés non reconnus (échecs) ou reconnus partiellement (redondants) sont isolés puis triés par fréquence d’apparition. Les libellés ou type de libellés les plus fréquents sont traités en priorité, soit en ajout simple, soit en synonymie puis ajoutés.
- l’aspect qualitatif : les libellés codés automatiquement sont vérifiés, "individuellement" s’il s’agit de libellés à haute fréquence d’apparition, par tirage d’échantillon parmi l’ensemble des codés et ponctuellement par poste de la nomenclature.

Par ailleurs, des contrôles sont opérés par l’expert variable lorsque les DR signalent des anomalies dans la codification automatique.

Après tous ces travaux, au 30 octobre 2000, quatre versions des environnements produits tickets de caisse ont été livrés et ont permis d’avoir un taux de codification automatique en augmentation.

En effet, sur la vague 1, le taux de codage automatique est de 69,22%, quel que soit l’environnement et la provenance de la dépense alors que le test 2 de juin 1999 était codé seulement à 63,26%. Sur la vague 2, ce taux passe au-dessus de 70% : 71,67%.

<table>
<thead>
<tr>
<th>Résultat du codage automatique PRODUITS</th>
<th>Nombre de produits VAGUE 1</th>
<th>% VAGUE 1</th>
<th>Nombre de produits VAGUE 2</th>
<th>% VAGUE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codés</td>
<td>95 345</td>
<td>69,22%</td>
<td>92 674</td>
<td>71,67%</td>
</tr>
<tr>
<td>Non codés</td>
<td>42 395</td>
<td>30,78%</td>
<td>36 635</td>
<td>28,33%</td>
</tr>
<tr>
<td>Total</td>
<td>137 740</td>
<td>100,00%</td>
<td>129 309</td>
<td>100,00%</td>
</tr>
</tbody>
</table>

7.3.3 Les produits des tickets de caisse de plus en plus codés

La mise à disposition de nouveaux environnements portant sur les libellés des produits issus des tickets de caisse sur les premières vagues de l’enquête est tout à fait payante. En effet, ce sont sur les libellés de tickets de caisse qu’ont porté tous les efforts de l’expert variable. On passe ainsi de 61,58% sur la vague 1 à 65,58% sur la vague 2 c’est-à-dire un gain de 4 points.

32 Pour plus de renseignements sur l’enquête Bdf, on se reportera à l’annexe n°5

INSEE Méthodes
<table>
<thead>
<tr>
<th></th>
<th>produits des tickets de caisse</th>
<th>produits isolés</th>
</tr>
</thead>
<tbody>
<tr>
<td>Résultat du codage automatique</td>
<td>Nombre de produits</td>
<td>%</td>
</tr>
<tr>
<td>Codés</td>
<td>51 715</td>
<td>61,58%</td>
</tr>
<tr>
<td>Redondants</td>
<td>10 044</td>
<td>11,96%</td>
</tr>
<tr>
<td>Non codés</td>
<td>22 222</td>
<td>26,46%</td>
</tr>
<tr>
<td>Total</td>
<td>83 981</td>
<td>100 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>produits des tickets de caisse</th>
<th>produits isolés</th>
</tr>
</thead>
<tbody>
<tr>
<td>Résultat du codage automatique</td>
<td>Nombre de produits</td>
<td>%</td>
</tr>
<tr>
<td>Codés</td>
<td>52 016</td>
<td>65,58%</td>
</tr>
<tr>
<td>Redondants</td>
<td>9 204</td>
<td>11,60%</td>
</tr>
<tr>
<td>Non codés</td>
<td>18 102</td>
<td>22,82%</td>
</tr>
<tr>
<td>Total</td>
<td>79 322</td>
<td>100 %</td>
</tr>
</tbody>
</table>

Nous sommes plutôt satisfaits des résultats obtenus sur les deux premières vagues de l’enquête par rapport au taux de codification des tests mais le travail d’enrichissement du Fab produits est quasi quotidien et par conséquent fastidieux.

7.3.4 La découverte de nouveaux problèmes liés à la collecte des tickets de caisse

Nous nous sommes aperçues en examinant les carnets des vagues 1 et 2 d’un phénomène qui gênait (voire même empêchait) le codage automatique des produits et que nous n’avions pas du tout rencontré lors des tests. Sur la partie ticket de caisse des carnets, nous avions demandé que soit ajouté un complément de dépense pertinent pour coder plus facilement le produit. Or compte tenu du fait que sur la partie tableau de dépense, la colonne “quantité / unité ” est clairement séparée du reste, nous récupérerons assez fréquemment une quantité dans le complément de dépense du ticket de caisse.

Nous avons donc essayé de rappeler aux gestionnaires d’enquête qui eux-mêmes transmettent cette information aux enquêteurs sur le terrain que le complément de dépense ne devait pas contenir uniquement la quantité et unité du produit. L’information ajoutée devait être “pertinente”.

Note : dans les Directions Régionales où c’est la même équipe qui forme les enquêteurs et qui réalise la reprise des non codés, le phénomène est moins fréquent. En effet, elle comprend mieux l’idée de pertinence pour un complément de dépenses.
8. Les prévisions

8.1 Pour les vagues suivantes

Des facteurs de variation dus à la variable produit peuvent laisser présumer des disparités fortes de codage sur l’ensemble de l’enquête.

Tout d’abord, le marché de la grande consommation est tel que des nouveaux produits apparaissent chaque jour, inscrits le plus souvent dans les tickets de caisse. De ce point de vue, il semble que le travail de suivi des bases de connaissances devra se poursuivre tout au long de l’enquête.

Ensuite, comme nous l’avons déjà dit, la variable a un fort caractère saisonnier. Or, les deux tests ont porté sur de courtes périodes ne couvrant pas l’année entière. Certes, nous pouvons espérer que les carnets saisis de 1994-1995 ont couvert une bonne partie de l’année, mais ils étaient peu nombreux et commencent à vieillir un peu. En particulier, les périodes comme celles des fêtes de fin d’année et des vacances d’hiver représentent des périodes de consommation spécifiques non rencontrées ou presque jusqu’ici.

Enfin, les tests ont eu lieu sur une petite partie du territoire métropolitain33. On peut s’attendre à ce que des produits régionaux spécifiques, comme ceux de la Corse ou d’Alsace-Lorraine, ne soient pas pris en compte dès maintenant : du travail reste à fournir pour ces cas-là.

8.2 A la fin de l’enquête

A la fin de l’enquête, une règle de codage sera utilisée ; elle affinera le codage du produit à partir du magasin.

*Par exemple, un café acheté dans une épicerie sous forme de paquet et un café payé au comptoir sont actuellement codés automatiquement avec le même code. Or, la nomenclature des produits (COICOP-HBS) donne 2 codes différents.*

De plus, nous effectuerons aussi un recodage complet avec la dernière version des bases de connaissance de tous les produits afin d’harmoniser les codes des produits. Nous traiterons, aussi, les codes 999999 correspondant aux non dépenses (cf. paragraphe II - 6.4).

33 L’enquête a aussi lieu dans les DOM avec les mêmes Fab complétés par quelques produits plus locaux.
Conclusion : le mot de l’ex-expert Sicore

Ce papier présente la mise en place d’un dispositif d’enquête sur deux enquêtes assez lourdes. En effet, utiliser la codification automatique au sein d’une enquête est loin d’être neutre, même si le gain (essentiellement en heures manuelles) est réel. Nous avons donc décrit ici les grandes étapes de la mise en place de la codification automatique sur deux variables nouvelles : les activités quotidiennes et les produits de dépense, ainsi que les principaux problèmes concrètes rencontrés.

En ce qui concerne les activités quotidiennes, une bonne partie du travail préparatoire à l’enquête sur les bases de connaissances a porté sur la construction des règles de codage. En effet, pour l’enquête Emploi du Temps, la nomenclature et la structure du recueil des données (le carnet) se prêtait bien à l’utilisation des règles de codage. Evidemment, le travail a aussi consisté à constituer et enrichir le fichier d’apprentissage. Toutefois, le taux de codification automatique a dès le début été bon : la maintenance des bases de connaissances en cours d’enquête a été relativement légère34. Au final, le taux de codification automatique a atteint 92,2% et on a constaté une excellente qualité du codage final.

En revanche, pour les produits, le travail préparatoire à l’enquête BdF pour construire les bases de connaissances a été relativement colossal et fastidieux, essentiellement parce que, pour une même variable, nous traitions une information hétérogène : les produits isolés transcrits manuellement par l’enquête et les produits issus des tickets de caisse collés sur le carnet. Et, nous avons vu que l’information recueillie via les tickets de caisse était toute différente et nécessitait un traitement à part. Ainsi, deux bases de connaissances sont nécessaires pour coder une seule et même variable. Tout cela a entraîné (en plus de la vacance du poste d’expert variable pendant un laps de temps non négligeable) une préparation insuffisante avant le démarrage de l’enquête. Finalement, les taux de codification restent bas (70%) et le travail sur les bases de connaissances toujours d’actualité, étant donné que l’enquête est en cours.

Outre la présentation de méthodes pouvant servir à la mise en place de la codification automatique d’une nouvelle variable par Sicore, ce papier montre les difficultés d’une telle opération : le résultat de la codification ne dépend pas uniquement de la structure de l’enquête sur laquelle elle est mise en place ; elle dépend également de la variable intrinsèque qui est codée (nomenclature, structure de l’information recueillie, qualité du recueil de l’information).

34 Dans la mesure où le travail de maintenance de bases de connaissances relatives à une variable via Sicore peut être léger.
Loin de prouver que l'utilisation de Sicro repose sur quelques principes généraux lors de l'application, nous avons montré que nous ne pouvons pas préconiser des conseils universels à appliquer lors de la construction et de l'amélioration des bases de connaissances. En réalité, les seules recommandations que l'on pourrait donner seraient d'être tenace et de rester patient ...
Bibliographie

1 - Sur la codification automatique en général


2 - Sur Sicore en particulier


RIVIERE P., SICORE, un outil et une méthode pour le chiffrement automatique à l’INSEE, Courrier des Statistiques, n°74, août 1995

SCHUHL P., SICORE, the INSEE Automatic Coding System, Proceedings de l’ARC, Annual Research Conference, Census Bureau, Washington, mars 1996

DESCHAMPS F., RIVIERE P., Codage automatique (1) : SICORE, le Recensement de la Population 1999, Préparation (I), INSEE-Méthodes N°79-80, pp. 267-299

3 - Sur l’enquête Emploi du Temps


Note à l’attention de M Maurin - P.Rivi ère - N°057/C530/PR/PR du 08/03/1995
Note à l’attention de M Maurin “Codage automatique des libellés d’emploi du
temps au moyen de Sicore - 3ème test-résumé” - F Dumontier et J L Pan Ké Shon -
N°2/F340 du 21/1/1997

Note à l’attention de D Guillermot “4ème test Codage Sicore-EdT” - F Dumontier
et J L Pan Ké Shon - N°17/F340 du 5/2/1998

4 - Sur l’enquête Budget de Famille

Présentation dans le cadre de la réunion du réseau de concepteurs d’enquêtes - CAE
- UMS - 26 février 1999 - “Enquête Budget de Famille 2000 : la préparation de la
codification par SICORE” - F. Deschamps et S. Destandau

Présentation dans le cadre de la réunion du réseau de concepteurs d’enquêtes - CAE
- UMS - 25 juin 1999 - “La qualité dans la future enquête Budget de Famille” - N.
Cérani, S. Destandau et H. Fréchou

Les enquêtes sur le budget des ménages dans l’Union européenne - Méthodologie et
recommandations pour l’harmonisation - 1997 - Eurostat -
Annexe N°1 : Sicore et ses connaissances...

Sicore est un logiciel de codification automatique créé par et pour l’INSEE dans les années 1990.

L’objet de ce papier n’est pas de présenter Sicore dans sa globalité, mais de comparer la manière dont il a été utilisé lors de la préparation de l’enquête Emploi du Temps 1998-1999 et lors de celle de Budget de Famille 2000-200135.

D’ailleurs, pour plus d’approfondissement sur la logique et l’architecture interne de Sicore, le lecteur pourra se référer aux articles cités en bibliographie.

1. Le codage automatique : de quoi s’agit-il ?

1.1 Coder

Assez simplement, le codage est l’action de transformer un texte en un code. Dans le contexte qui nous intéresse ici, il s’agit donc de convertir un produit de dépense consigné par l’enquêté vers le code de la nomenclature correspondant.

En théorie, tous les libellés devraient être codés. En pratique, cela n’est pas toujours le cas, et ceci pour plusieurs raisons. Tout d’abord, le libellé peut ne pas correspondre à ce à quoi on s’attend.

Par exemple, "CADRE" ne suffit pas à cerner une PCS et "ALPES" ne permet pas d’attribuer un code département.

Ensuite, le libellé peut ne pas contenir les informations suffisantes.

Dans le Val-de-Marne, deux communes commencent par "VILLENEUVE" : "VILLENEUVE-SAINT-GEORGES" et "VILLENEUVE LE ROI". Ainsi, le seul libellé "VILLENEUVE" ne nous permettra pas de coder, même si nous savons que la commune en question est dans le Val-de-Marne.

Enfin, la nomenclature peut comporter des flous que chacun résoudra à sa manière, même si le résultat final du codage est non homogène. Nous verrons des exemples de ce cas plus tard.

En un mot, coder n’est pas facile, surtout lorsque l’on sait que tous les libellés doivent avoir un code à la fin.

35 Pour cette enquête, nous nous intéresserons quasi uniquement à la variable "produit de la dépense" ; mais d’autres variables y sont codées : d’une part la CS contenue dans le tronc commun et d’autre part la variable "lieu de la dépense" qui, plus simple que la variable "produit de la dépense" ne nous intéressera pas dans ce papier.
1.2 Coder automatiquement

Chiffrer automatiquement, c’est faire coder le libellé par une machine. En tout état de cause, ce choix de la codification automatique de variables au sein d’une chaîne d’enquête n’est pas neutre.

En effet, ce choix induit des changements majeurs dans le calendrier de l’enquête, dans la présentation des supports de ces variables et dans son déroulement. Il faut d’abord prévoir en amont du codage automatique une saisie de qualité aussi bonne que possible, ce qui est relativement coûteux. Ensuite, comme le codage automatique n’est jamais efficace à 100%, il convient de mettre en place un dispositif permettant de reprendre les libellés non codés par la machine à la main (grâce à ce que l’on appelle un “outil de reprise” élaboré spécialement) ; ici encore cette opération est coûteuse manuellement et informatiquement, même si le volume de ces libellés est moindre que si l’ensemble des libellés était à coder manuellement.

Evidemment, les difficultés de codage que nous avons évoquées plus haut se retrouvent ici, et de manière plus forte encore : si la personne qui paramètre la codification automatique a du mal à arbitrer des codes, la machine ne pourra pas faire mieux.

2. Sicore : comment fonctionne-t-il ?

2.1 Principe général

Le principe de codage de Sicore est, en théorie, assez simple : en entrée, il reçoit les libellés de la saisie et en sortie, il renvoie le résultat de son codage (qu’il y ait un code ou pas).

Pour cela, Sicore agit en deux temps : d’abord, il normalise le libellé de la saisie de manière à le simplifier et l’homogénéiser à ce qu’il connaît (ou croit connaître). Ensuite, il tente de reconnaître ce libellé simplifié ; il s’agit là de l’étape de codage à proprement parler.

Toutes ces références dont on parle ici sont ce que l’on appelle en langage Sicore les bases de connaissances. De ce fait, il y a une totale séparation entre le programme et les connaissances relatives à une variable : les programmes sont toujours les mêmes quelle que soit la variable à coder. Mais cela implique évidemment que lorsque les connaissances d’une variable n’existent pas, il faut les écrire depuis le début.
2.2 La phase de normalisation

La première étape est dite de *normalisation*. Elle comporte plusieurs actions successives qui utilisent une partie des bases de connaissances de chaque variable. Ces actions successives sont :

- L’élimination des caractères blancs, c’est à dire des caractères qui seront remplacés par un espace.

  *Exemples : l’apostrophe, la virgule, le point d’interrogation, ...*

- L’élimination des caractères vides, c’est à dire des caractères qui seront éliminés.

  Pratiquement, seul le point est déclaré pour traiter les abréviations.

- L’élimination des mots vides, c’est à dire des mots qui seront éliminés.

  À la différence de ce qui précède, cette rubrique est bien spécifique à la variable traitée.

  *Exemples pour "produit de dépense" : "L", "PROMOTION", "LOT", ...*

- L’utilisation des synonymes, c’est à dire des mots (ou groupes de mots) qui seront remplacés par d’autres mots (ou groupes de mots). Cette action a particulièrement un rôle d’entonnoir, dans le sens où elle permet, en plus de son action simple de synonymisation ("AGNEAU" = "MOUTON" *par exemple*), de traiter des éventuelles fautes d’orthographe et/ou de saisie ("AGNAU" = "AGNEAU"). Il convient alors d’être très attentif dans l’ordre dans lequel les synonymes sont rentrés dans Sicore.

- Le calibrage, c’est à dire donner au libellé obtenu à l’issue des étapes précédentes une longueur fixée, en limitant le nombre de mots et la taille des mots (*exemple : 5 mots de 10 caractères*). Lorsque les mots sont trop courts ou trop peu nombreux, on complète par des blancs.

2.3 Le fichier d’apprentissage

Lors de l’étape de codage, nous avons dit que Sicore essayait de rapprocher le libellé à coder de quelque chose qu’il reconnaît ou croit reconnaître. En disant cela, nous avons fait directement référence au fichier d’apprentissage qui est sans doute la partie des bases de connaissances la plus importante et la plus intuitive. Il s’agit d’un fichier contenant sur chaque ligne un libellé et le code associé à ce libellé.

Dans un premier temps, en fait avant la normalisation, le fichier d’apprentissage est dit *brut* : c’est celui sur lequel les améliorations sont effectuées, pour des soucis de meilleure lisibilité.
Dans le cas des pays et nationalités, le Fab a cette allure :

<table>
<thead>
<tr>
<th>Code</th>
<th>Libellé</th>
</tr>
</thead>
<tbody>
<tr>
<td>99101</td>
<td>COPENHAGUE</td>
</tr>
<tr>
<td>99101</td>
<td>ARHU1</td>
</tr>
<tr>
<td>99101</td>
<td>DANEMARK</td>
</tr>
<tr>
<td>99101</td>
<td>DANOIS</td>
</tr>
<tr>
<td>99438</td>
<td>DOMINIQUE</td>
</tr>
<tr>
<td>99438</td>
<td>DOMINIQUE</td>
</tr>
<tr>
<td>99438</td>
<td>ROSEAU</td>
</tr>
<tr>
<td>99408</td>
<td>REPDOMINICA1</td>
</tr>
<tr>
<td>99408</td>
<td>REP DOMINICAINE</td>
</tr>
<tr>
<td>99408</td>
<td>REPUBDOMINIC</td>
</tr>
<tr>
<td>99408</td>
<td>REPUBLIQUE DOMINICAINE</td>
</tr>
<tr>
<td>99408</td>
<td>SAINT DOMINGUE</td>
</tr>
</tbody>
</table>

Le code est sur l’ensemble des 13 premiers caractères, le libellé correspondant suit. On se rend compte ici que plusieurs libellés correspondent au même code ; en revanche, il n’est pas envisageable qu’un même libellé soit associé à plusieurs codes différents, même si cela peut exister dans les faits. Par exemple, une personne se déclarant de nationalité “ BASQUE ” peut tout à fait être dans les faits, espagnole ou française.

Enfin, le Fab peut comporter des mots joker. Il s’agit de mots dont la totalité des caractères peuvent être égaux à n’importe quel caractère (lettre ou chiffre). Cette technique, certes risquée, permet de traiter, par exemple, tous les libellés commençant par un certain groupe de mots. Par exemple, tous les libellés de profession commençant par “ DIRECTEUR D’ECOLE ” peuvent être traités, même si l’enquête a rajouté le nom de l’école dans laquelle il est directeur (“ DIRECTEUR DE L’ECOLE VICTOR HUGO ”) ou la commune36 (“ DIRECTEUR DE L’ECOLE COMMUNALE DE MONNETIER ”).

En pratique, et spécifiquement sur les variables qui nous intéressent ici, le Fab est la partie des bases de connaissances la plus travaillée et celle qui s’enrichit le plus au cours du temps.

### 2.4 L’utilisation éventuelle d’information supplémentaire

Or, il arrive que dans certains cas, pour certaines variables, on ne puisse coder directement un libellé. Sans même parler du cas de la PCS, on ne peut pas toujours

---

36 Ceci n’est pas tout à fait exact : les directeurs d’école primaire ou maternelle, publique ou privée sont en 4214, à l’exception des directeurs d’école à classe unique classés en 4211. Toutefois, cette subtilité de déclaration est difficilement exigible de l’enquêteur et de l’enquêté.
coder une commune à partir du libellé si on n’a pas le département dans lequel est située la commune. Dans de tels cas, on utilise de l’information supplémentaire sous forme de variables annexes que l’on traite dans des règles logiques. Le fichier des règles logiques est une nouvelle partie des bases de connaissances.

Dans le cas de la commune, on regarde toujours le département déclaré. Pour la PCS, 14 variables peuvent être nécessaires au codage. Dans les faits, on ne prend en compte que celles nécessaires au regard du libellé.

2.5 Les paramètres d’apprentissage : la mise en œuvre de la qualité du codage

A ce stade, on pourrait être tenté de croire que toutes les connaissances nécessaires au codage de la variable sont décrites. Ceci n’est pas tout à fait exact. Pour pouvoir coder le plus rapidement et avec la meilleure qualité possible, on définit ce que l’on appelle des paramètres d’apprentissage. Lorsque Sicore va apprendre toutes les connaissances écrites jusqu’ici, il va découper les libellés du fichier d’apprentissage, après normalisation, en groupe de deux lettres : les bigrammes.

De manière à vérifier la qualité du code attribué par Sicore, on définit des bigrammes de redondance, c’est-à-dire des bigrammes que l’on souhaite vérifier quoi qu’il arrive. Une fois le code attribué, Sicore va scrupuleusement vérifier tous ces bigrammes de redondance entre le libellé à coder et le libellé qu’il a de plus proche dans les connaissances qu’il a apprises.

Il suffit alors qu’un seul bigramme du libellé à coder soit différent du même bigramme du libellé le plus proche dans les connaissances apprises pour que l’on décrète qu’il y a erreur de redondance. Dans ce cas, même si Sicore renvoie un code, on considère le libellé comme non codé. En pratique, les 2 ou 3 premiers bigrammes des (2 ou 3) premiers mots sont souvent pris comme bigrammes de redondance ; cela dépend du fichier des connaissances de chacune des variables.

Prenons un exemple concret sur la variable “produit de la dépense”.
Initialement, les trois premiers bigrammes des deux premiers mots du libellé normalisé sont redondants. Cela donne, pour 4 mots de 12 lettres (ou 6 bigrammes) chacun :

Soit maintenant le libellé brut suivant “une bouteille d’assouplissant” qui fait partie du Fab. Normalisé, ce libellé devient “BOUTEILLE ASSOPLISSANT”. Imaginons maintenant que Sicore doive coder le libellé brut suivant “une

37 A titre d’illustration, il faut savoir qu’il existe trois “MARSEILLE” en France
bouteille de soupline". Normalisé, ce libellé devient "BOUTEILLE, SOUPLINE". Pour un codeur manuel, ce libellé ne posera aucun problème. Pour Sicore, il ne pourra être codé car, si le premier mot "BOUTEILLE" ne posera pas de problème, Sicore ne reconnaîtra pas le deuxième : dans le contexte décrit précédemment, ce libellé ne passera pas le contrôle de redondance38.

De plus, d'autres paramètres d'apprentissage permettent à l'utilisateur d'agir sur la rapidité de codage et sur d'autres points de la phase de codage. Ces paramètres ne seront pas développés ici car ils ne nous intéressent pas particulièrement. De plus, l'ensemble du paramétrage n'est pas immédiat à mettre au point, et se fait par tâtonnement.

Tous ces paramètres d'apprentissage permettent à l'utilisateur de jouer sur un des grands paradoxes du codage automatique : le dilemme efficacité – fiabilité ; autrement dit, plus on code, moins on code juste. Il est facile de coder à 100% : il suffit de renvoyer l'ensemble des libellés sur un seul et unique code, via les mots joker. Arbitrer ce dilemme est beaucoup plus délicat et dépend très généralement des moyens alloués, notamment en reprise des libellés non codés.

2.6 La phase d'apprentissage

Une fois que toutes ces connaissances sont écrites et consignées, nous pouvons les faire apprendre, au sens littéral du terme, par Sicore lors de l'apprentissage. Lors de cette phase, Sicore prend en entrée les libellés normalisés du fichier d'apprentissage, les découpe en bigrammes et construit un arbre de questionnement selon les paramètres d'apprentissage définis par l'utilisateur. Nous ne nous attarderons pas ici sur la construction de cet arbre, même si, en théorie, beaucoup de choses sont à expliquer. Le lecteur espérant plus d'explications se reportera aux ouvrages cités en bibliographie.

38 En pratique, les choses sont plus complexes du fait de la multiplicité des lignes dans le fab, y compris celles commençant par le mot "bouteille" : jus d'orange, eau, alcool, …
2.7 Le codage à proprement parler

Une fois ces connaissances apprises, le codage à proprement parler peut commencer. Sicore prend alors un des libellés qu'on lui donne à coder, le normalise, le découpe en bigrammes et parcourt ses connaissances jusqu'à éventuellement trouver un libellé qui s'en rapproche. Trois cas généraux peuvent se présenter :

1° S’il n’en trouve pas, on parle d’échec de codage, Sicore passe au libellé suivant.

2° Si Sicore trouve un code, il effectue les contrôles de redondance que l’utilisateur lui a donné l’ordre de faire. Dès qu’un contrôle échoue, Sicore indique qu’il y a eu un problème de redondance, et donne le code qui aurait été trouvé s’il n’y avait pas eu de problème 39.

3° Si tous les contrôles sont satisfaisants, Sicore renvoie le “bon” code, en tous les cas, un code qui est pris comme étant bon. Les études de la qualité de codage détermineront éventuellement si certains codes sont justes ou pas.
En cas d’utilisation de variables annexes, le principe est le même : Sicore, après avoir traité le libellé suivant le précédent schéma parcourt les règles logiques de Sicore pour traiter l’information supplémentaire remplie par l’enquêté.

3. Une organisation spécifique autour de Sicore est indispensable

Il est assez naturel d’entrapercevoir le fait que choisir d’insérer Sicore au sein d’un traitement d’enquête est loin d’être neutre. C’est la raison pour laquelle, dès le déroulement du projet Sicore, il est apparu la nécessité d’imposer des règles d’organisation en cas d’utilisation de Sicore.

Cet aspect organisationnel revêt essentiellement deux composantes 40 : la gestion (élaboration éventuelle et mises à jour) des bases de connaissances et la définition des acteurs autour de Sicore avec la distribution des rôles.

39 Malgré tout, dans de tels cas, le libellé est considéré comme non codé.

40 En fait, un troisième élément est très important : l’insertion de Sicore au sein de la chaîne de traitement de l’enquête ; mais cet aspect ne nous intéresse pas ici, nous ne le traiterons pas.
3.1 La gestion des bases de connaissances
Le premier aspect met en place un processus cyclique appelé boucle Sicore dans deux cas : lorsque la variable est utilisée de manière permanente (comme la PCS-CS) ou lorsque la période d'enquête est suffisamment longue (comme pour les deux enquêtes qui nous intéressent ici).

En effet, pour qu'une application de codage automatique demeure efficace à long terme, il faut qu'elle soit vivante : le langage évolue, de nouvelles expressions apparaissent, et les nomenclatures changent également (apparition de nouveaux métiers, fusion de communes, ...). Si ce n'est pas le cas, les gestionnaires de la reprise qui traitent les libellés non-codés automatiquement récupèrent à chaque fois les mêmes mots, les mêmes termes, ce qui entraîne une démotivation de leur part ; et un surcoût non négligeable pour l'enquête du fait que des libellés pourraient être traités automatiquement.

Ensuite, il faut s'assurer que les connaissances de la variable ne contiennent pas d'erreurs. En effet, malgré toute l'attention portée au travail d'élaboration des connaissances et toutes les précautions que l'on peut prendre, des erreurs peuvent être mises en évidence au sein des bases de connaissances.

De plus, pour les deux variables qui nous intéressent ici (« activité quotidienne » et « produit de dépense »), nous verrons plus loin que les conditions de codage étaient initialement tout à fait similaires.

Le principe de cette boucle Sicore est assez simple : améliorer les bases de connaissances existantes pour coder plus et/ou mieux41, en s'appuyant sur des résultats de codage précédents : on doit s'assurer que le codage précédent n'est pas trop mauvais, et faire en sorte que le nombre de non codés diminue progressivement.

Toutes ces consignes assez naturelles à imaginer restent assez vagues. C'est là toute la difficulté de la mise à jour des bases de connaissances : il n'existe pas de règle miraculeuse qui amélioreraient systématiquement le codage. On peut se contenter de fournir des pistes méthodologiques : tirer des échantillons de libellés codés et extrapoler la qualité résultante, essayer de traiter les principales fautes d'orthographe et/ou de saisie en rajoutant des synonymes, traiter en priorité les libellés non codés qui sont les plus fréquents après normalisation, ...

Cette boucle Sicore amène donc des changements des bases de connaissances. Ces changements sont possibles à plusieurs niveaux : synonymes, fichier d'apprentissage, paramètres d'apprentissage... Ils sont également possibles dans les deux sens, dans le sens d'une extension de ces bases, comme dans celui d'un rétrécissement : en effet, en cas de qualité jugée trop médiocre, on peut se rendre compte, par exemple, qu'un synonyme a un double sens non envisagé initialement et décider de l'éliminer.

41 Cela dépend de l'option qui est prise sur l'arbitrage efficacité – fiabilité.
A la fin de cette phase d'amélioration, des procédures de vérification sont mises en place. Les plus couramment utilisées sont au nombre de deux : le codage du Fab sur lui-même qui doit arriver sur un codage total sans le moindre problème et le codage d'un précédent fichier d'enquête qui ne doit pas voir l'efficacité et la qualité changer du tout au tout par rapport au précédent codage.

3.2 La structure des acteurs gravitant autour de Sicore

La deuxième composante essentielle de l'aspect organisationnel lié à Sicore concerne les acteurs fondamentaux de la codification automatique. En théorie, ils sont au nombre de cinq.

Le système Sicore s'organise autour d'un expert Sicore. Celui-ci est le point d'entrée, le relais, pour tout statisticien qui voudrait utiliser le codage automatique dans son application. Il centralise les bases de connaissances, gère le réseau des experts de variables et les groupes de travail correspondants, prend en compte les demandes sur le codage automatique et conseille les utilisateurs, organise les formations, spécifie les évolutions de l'outil, gère la communication interne et externe. L'expert Sicore est en lien étroit avec l'informaticien Sicore qui est chargé de la maintenance évolutive de l'outil, de l'intégration de Sicore dans les chaînes de traitement d'enquête, et, parallèlement, de la documentation informatique de Sicore.

Viennent ensuite les utilisateurs de Sicore. D'abord, il y a l'expert variable (un expert pour chaque variable codée par Sicore) est la cheville ouvrière du système. C'est lui qui a la lourde tâche de construire et de mettre à jour la base de connaissances qui l'intéresse. L'informaticien de l'enquête doit écrire la chaîne de traitement informatique, et donc y insérer les traitements induits par l'utilisation de Sicore, ceci en coordination avec l'informaticien Sicore. Enfin, le statisticien d'enquête est la personne à l'origine de tout ; c'est lui qui passe la commande du codage automatique et supervise toute son enquête, en particulier le codage automatique. C'est par exemple cette personne qui peut arbitrer entre efficacité et fiabilité.

4. Les différentes utilisations de Sicore antérieures à EdT et à BdF

L'une des particularités de Sicore est son aspect transversal : il est utilisable pour toutes les enquêtes ménages pour coder des variables diverses. Avant les deux enquêtes dont il est question ici, Sicore codait plusieurs variables. On comptait alors deux variables simples, sans variable annexe : le département, et le code pays/nationalité.
Ensuite, nous avons le cas un peu plus compliqué de la commune : il faut absolument que le département soit renseigné et on compte en plus une variable annexe : la date à laquelle on veut que le code se réfère. Si aucune date n’est renseignée, le codage donne le code connu le plus récent : de ce point de vue, la date est une variable annexe non nécessaire.

Enfin, la PCS représente la variable la plus lourde à gérer : 14 variables annexes, de natures très diverses et pour certaines déjà issues de codification précédente, en plus du libellé, si possible assorti d’une indication de grade pour les employés de la fonction publique.

Toutes ces variables évoquées précédemment, avant le lancement de l’enquête EdT, avaient la particularité d’avoir déjà leur base de connaissances et étaient utilisées pour des enquêtes diverses : toutes celles utilisant le tronc commun des enquêtes ménages, l’Etat-civil, des enquêtes régionales de déplacement, ...

Nous allons voir qu’avec les deux enquêtes EdT et BEdF, les choses sont un peu différentes : rien n’existait avant la phase de préparation des deux enquêtes.
Annexe n°2 - L’enquête Emploi du Temps


L’objectif principal de ces enquêtes est de savoir quelles sont les occupations quotidiennes des individus des ménages ordinaires et leur durée.

Les supports de l’enquête de 1998-1999 sont tous papier :
- la fiche-adresse
- les questionnaires
- les carnets journaliers

Le carnet journalier auquel nous nous intéressons ici permet de décrire toutes les activités d’un même jour (24 heures) de chaque individu de 15 ans et plus du ménage interrogé.
- les semainiers

L’enquêteur se déplace deux fois chez le ménage. La première fois, il remplit le questionnaire ménage, explique le remplissage du carnet (en faisant un exercice avec le ménage), fixe le jour de remplissage du carnet, administre les questionnaires individus avec les personnes (de 15 ans et plus), présentes, et laisse au ménage les carnets à remplir et un “semainier” de travail pour les actifs occupés. Il revient une seconde fois chez le ménage, pour ramasser les carnets, les contrôler avec le ménage et pour administrer les questionnaires individus qui ne l’avaient pas été lors de la 1ère visite.

1. L’organisation du traitement des documents

Tout le travail de saisie, d’apurement, de lancement du codage automatique et de reprise de codage de tous les documents a été confié au GSAS de la DR des Pays de la Loire.

Chaque personne de cet atelier pouvait revenir au dossier papier complet. Des réunions fréquentes avaient lieu au sein de l’équipe pour mettre en commun les problèmes rencontrés qui étaient alors transmis, soit au responsable d’enquête, soit à l’expert variable, par le responsable.

Les dossiers papier au niveau ménage étaient groupés par lots.
2. L’outil de reprise

Une fois les carnets saisis et apurés, le lancement de la codification automatique était réalisé par une personne de l’équipe qui traitait alors les reprises. Pour ce faire, un outil a été mis au point par l’informaticien de l’enquête (Isabelle Rebourg du CNIA).

Lors des reprises effectuées par les personnes du GSAS grâce à cet outil, apparaissaient non seulement le libellé à coder, ses 9 variables annexes, mais aussi les intitulés des activités précédentes et suivantes, pour mieux situer l’activité dans la journée.

Grâce à un menu spécifique de cet outil, l’expert variable avait la possibilité de voir l’avancement du codage et de demander des listes d’activités codées et non codées triées selon plusieurs critères.

L’ensemble de l’enquête comportait 429 678 libellés à coder, la plupart étant des activités principales (316 097) et le reste des activités secondaires (environ 27% de l’ensemble des activités).
ANNEXE N°3 – Exemple de remplissage du carnet journalier de l’Enquête Emploi du Temps

Vous comprendrez facilement comment procéder en regardant les exemples et en appliquant les consignes suivantes :
- Notez vos occupations régulièrement et plusieurs fois dans la journée afin d'éviter les oubliés.
- Notez vos occupations que vous délimitez par une accolade aux heures de début et de fin de chaque occupation.
- Il doit y avoir une utilisation de toutes les heures de la journée; on peut très bien réfléchir, se reposer, parler, ou attendre quelque chose pendant une partie de la journée, et il faut le marquer.

<table>
<thead>
<tr>
<th>Marquez vos différentes occupations de la journée en indiquant les heures de début et de fin de chaque occupation à l'aide d' accolades dans la colonne de gauche</th>
<th>Faites-vous autre chose en même temps ? (lecture, conversation, radio, TV...)</th>
<th>Lieu ou trajet une réponse possible</th>
<th>en présence de qui 3 réponses possibles</th>
<th>Votre activité est dans un but une réponse possible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chez soi</td>
<td>Lieu de travail</td>
<td>A l'extérieur</td>
<td>Trajet domestique travail</td>
<td>Autres trajets</td>
</tr>
<tr>
<td>7h00</td>
<td>je dors</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>10</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>20</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>30</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>40</td>
<td>je fais ma toilette radio</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>50</td>
<td>je m'habille radio</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>8h00</td>
<td>je prépare le petit déjeuner conversation</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>10</td>
<td>je déjeune TV</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>20</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>30</td>
<td>je fais la vaisselle TV</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>40</td>
<td>je range la cuisine TV</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>50</td>
<td>je fais le ménage</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>9h00</td>
<td>je vais au travail avec ma femme et un voisin conversation</td>
<td>0 1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Date</td>
<td>Lignes à coder</td>
<td>Environnement Score</td>
<td>Efficacité activité principale</td>
<td>Efficacité activité secondaire</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>mars 1995</td>
<td>construction FAB</td>
<td>199 codes d'activité FAB= 3918 lignes</td>
<td>100% codés</td>
<td></td>
</tr>
<tr>
<td></td>
<td>recherche codes multiples</td>
<td>Pas de règles</td>
<td>82% codés simples</td>
<td></td>
</tr>
<tr>
<td></td>
<td>codage FAB sur FAB</td>
<td>Pas de synonymes</td>
<td>18% multiples (1/5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test 1</td>
<td>6 big sur les 2 1er mots</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19 échantillons de 200 lignes test de codage</td>
<td>19 FAB complément</td>
<td>40.4% codés (en moy)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32.4% codés simple</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.0% codés multiples</td>
<td></td>
</tr>
<tr>
<td>été 95</td>
<td>Test 2 1er essai</td>
<td>199 codes d'activité FAB= 8500 libellés</td>
<td>67% codés</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2500 lignes issues 84 carnet</td>
<td>650 synonymes</td>
<td>dont 13% multiples</td>
<td></td>
</tr>
<tr>
<td></td>
<td>saisie de enquête 1986</td>
<td>6 big sur 1er et 2ème mot</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50 règles</td>
<td>70% codés</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 big sur 3ème mot</td>
<td>dont 3% multiples</td>
<td></td>
</tr>
</tbody>
</table>
### Test de redondance sur le terrain

<table>
<thead>
<tr>
<th></th>
<th>Test 3</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>juin 96</td>
<td>3253 libellés issus test 2 DR</td>
<td>100 codes d'activité</td>
<td>80% codés</td>
<td>90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>variables annexes</td>
<td>FAB = 9500 libellés</td>
<td>introduction de jokers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1060 synonymes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Test 4</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>juin 97</td>
<td>4270 libellés issus test 2 DR</td>
<td>amélioration</td>
<td>83%</td>
<td>93.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAB = 10500 lignes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1650 synonymes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Test 5</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sept 97</td>
<td>4220 libellés issus de deux DR</td>
<td>1er codage</td>
<td>78.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>même environnement que test 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2ème codage</td>
<td>environnement amélioré</td>
<td>81.6%</td>
<td>98%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>activités principales et secondaires</td>
<td>96.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>86%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Enquête de l'15/2/98 au 15/2/99</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vague 1</td>
</tr>
<tr>
<td></td>
<td>55 410 libellés à coder</td>
</tr>
<tr>
<td></td>
<td>77 règles</td>
</tr>
<tr>
<td>du</td>
<td></td>
</tr>
<tr>
<td>15/2/99</td>
<td></td>
</tr>
<tr>
<td>au</td>
<td></td>
</tr>
<tr>
<td>29/3/98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recodification 8 vagues</td>
</tr>
<tr>
<td>avril 98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>99 règles</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annexe no5 : L’enquête Budget de Famille

L’enquête Budget de Famille ressemble étrangement à sa petite sœur l’enquête Emploi du Temps. En effet, BdF est une enquête ancienne (depuis 1965 sous ce nom) et réalisée tous les 5 ans par l’INSEE. Elle comporte 8 vagues de 6 semaines réparties sur un an. La dernière en date s’étale de mai 2000 à mai 2001 pour la version métropolitaine. La taille de son échantillon est supérieure à celle d’EdT : 20 000 logements métropolitains.

Son objectif principal consiste à mesurer le plus précisément possible les dépenses, les consommations et les ressources des ménages français.

Pour cela, elle repose sur 2 types de support :
- des questionnaires posés au ménage sous CAPI
- des carnets de dépenses papier.

Avant de réaliser l’enquête en grandeur nature, deux tests ont eu lieu sur le terrain grâce à des enquêteurs de 5 Directions Régionales différentes en novembre 1998 et juin 1999.

1. L’organisation du traitement des carnets papier

Compte tenu du nombre de FA de l’enquête BdF supérieur à celui de l’enquête EdT(20 000 contre 12 000), du nombre de dépenses par ménage (par rapport au nombre d’activités dans un carnet journalier), un choix d’organisation bien différent de celle adoptée pour l’enquête EdT en ce qui concerne le traitement des carnets a été nécessaire.

1.1 La saisie des carnets de dépense

Sur l’ensemble des vagues de l’enquête, le volume à traiter a été estimé ainsi :

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre de FA de l’échantillon initial métropole + DOM</td>
<td>23 000</td>
</tr>
<tr>
<td>Nombre de ménages répondants métropole + DOM</td>
<td>16 000</td>
</tr>
<tr>
<td>Nombre moyen estimé de dépenses par ménage</td>
<td>134</td>
</tr>
<tr>
<td>Nombre estimé de dépenses à saisir</td>
<td>2 144 000 dépenses</td>
</tr>
</tbody>
</table>

42 Initialement nous n’avions pas prévu d’échantillon de réserve métropolitain. Cet échantillon comporte 2000 FA en plus des 18 000 FA standards pour la métropole

INSEE Méthodes
Compte tenu de ce volume, tous les ateliers de saisie en exercice en mai 2000 ont dû saisir les dépenses des carnets. Etant au nombre de 13 contre 18 DR de collecte, des regroupements de DR ont été effectués selon la taille des GSAS.

*Par exemple, le GSAS de Toulouse a pris en charge la saisie des carnets de la DR du Languedoc Roussillon et de ceux de l’Auvergne en plus de ceux de leur DR.*

### 1.2 La reprise des carnets de dépenses

Pour la reprise des carnets de dépenses, le volume à traiter et l’organisation de la saisie a eu des conséquences non négligeables sur l’organisation de la reprise qui est en fait elle aussi bien différente de celle choisie pour EdT.

*une estimation de volume à reprendre démoralisante*

<table>
<thead>
<tr>
<th>Taux estimé de reprise des produits en 2000 (4 vagues)</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre estimé de produits à reprendre en 2000</td>
<td>857 600 produits à reprendre</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taux estimé de reprise des produits en 2001 (4 vagues)</th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre estimé de produits à reprendre</td>
<td>428 800 produits à reprendre</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taux estimé de reprise des magasins</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre estimé de magasins à reprendre</td>
<td>112 000 magasins à reprendre</td>
</tr>
</tbody>
</table>

*l’organisation de la saisie, une contrainte pour la reprise*

Compte tenu de la multitude des sites de saisie, 2 options ont été prises :

- d’une part, nous sommes parties du principe qu’il fallait déplacer le moins possible les carnets aprè leur saisie de manière à ce que les personnes effectuant la reprise puissent les consulter.

Les 13 DR effectuant la saisie ont donc été chargées aussi de la reprise, l’organisation au sein de la DR étant laissée entièrement libre. En réalité, pour la majorité des 13 DR (8), la reprise des carnets s’est effectuée au sein des DEM. Dans 4 DR, le travail de codage manuel est partagé entre la DEM et le GSAS et enfin dans seulement 1 DR (Nantes, là où les carnets d’EdT ont été entièrement traités) toute la reprise a été confiée au GSAS.

- d’autre part, la codification automatique n’est pas lancée par les différentes équipes assurant derrière la reprise des rejets comme pour EdT mais par l’informaticien de l’enquête au niveau central (CNI de Lille). Il assure ensuite la mise à disposition des dépenses non codées automatiquement auprès de ces 17 équipes ((GSAS+DEM)*4 + 8 + 1) dans les 13 DR de reprise afin qu’elles les codent manuellement grâce à un outil informatique spécifique à BdF.
2. L’outil de reprise

En plus de constituer le Fab produits, tout le travail de constitution des connaissances a permis de choisir une présentation des carnets, mais aussi de spécifier l’outil de reprise qui est utilisé par les 17 équipes en DR durant plus d’un an, pour coder manuellement des dépenses non codées automatiquement.

- En effet, d’une part, il nous a paru intéressant de conserver l’information du type de codage (codé, redondant, échec) c’est à dire l’écho de Sicore dans l’affichage des dépenses. Ainsi face à un produit (ou un magasin) redondant, le codeur peut s’aider du code proposé par Sicore pour coder manuellement.
- D’autre part, l’outil n’affiche pas seulement le produit rejeté (c’est-à-dire redondant ou en échec) mais bien l’ensemble des informations relatives à ce produit c’est à dire
  - s’il provient du tableau de dépenses :
    * la quantité et unité quand elle existe,
    * le montant de la dépense,
    * le type de magasin.
  - s’il provient d’un ticket de caisse :
    * le montant de la dépense,
    * le type de magasin.

Bien entendu, toutes les autres informations relatives à l’identification du ménage, de l’individu et de la date de l’enquête peuvent apparaître à la demande du codeur afin qu’il évite de revenir au carnet papier.

Le codeur peut donc être confronté au cas où le produit est à coder mais le magasin est codé automatiquement. Dans ce cas, le libellé du magasin peut l’aider à coder le produit.
Exemple : “1 - COUPE - 220 FRANCS - COIFFEUR”
ANNEXE n°6 – Exemple de remplissage du carnet de dépenses de l’Enquête Budget de Famille

Nième jour de collecte
Jour de la semaine : Mercredi
Date : 18-11-1998

TICKETS DE CAISSE

<table>
<thead>
<tr>
<th>Article</th>
<th>Quantité</th>
<th>Unité</th>
<th>Nature de la dépense</th>
<th>Montant en francs</th>
<th>lieu de dépense</th>
</tr>
</thead>
<tbody>
<tr>
<td>GINIM 10x20 CL 6</td>
<td>18.70</td>
<td>XXXXX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOLVIC</td>
<td>11.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUTIL, AUTO MAG</td>
<td>169.00</td>
<td>→</td>
<td>housses siège voiture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOUET LOISIRS</td>
<td>25.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIEL ACACIA</td>
<td>15.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COQUIGRAIN 65g 70x6</td>
<td>6.40</td>
<td>→</td>
<td>œufs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAJESTY BOEUF</td>
<td>2.10</td>
<td>→</td>
<td>alimentation chat conserve</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BANGA 4x20 CL</td>
<td>6.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEMME</td>
<td>49.90</td>
<td>→</td>
<td>slip femme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RICORE 250G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BATONNETS</td>
<td>3.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COTON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENIE GEL</td>
<td>11.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HTE PYRENE</td>
<td>21.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

→ cadeau

TABLEAU DES DÉPENSES

<table>
<thead>
<tr>
<th>Quantité</th>
<th>Nature de la dépense</th>
<th>Montant en francs</th>
<th>lieu de dépense</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>facture EDF</td>
<td>782.20</td>
<td>prélèvement</td>
</tr>
<tr>
<td>2 billets</td>
<td>train</td>
<td>1250</td>
<td>SNCF</td>
</tr>
<tr>
<td>1</td>
<td>magazine</td>
<td>10.50</td>
<td>presse</td>
</tr>
<tr>
<td>1</td>
<td>pull over cadeau</td>
<td>250</td>
<td>Monoprix</td>
</tr>
<tr>
<td>3 places</td>
<td>cinéma</td>
<td>150</td>
<td>Gaumont</td>
</tr>
</tbody>
</table>

69
Annexe n° 7 : Glossaire

BdF : enquête Budget de Famille

CAPI : Collecte Assistée par Informatique

CNE : Centre National d’Exploitation dans une DR

CNI : Centre National Informatique
(Aix - Lille - Orléans - Nantes - Paris)

COICOP-HBS : nom de la nomenclature européenne (Eurostat) des enquêtes Budget de Famille

DEM : Division Enquête Ménage (SES) dans une DR

DG : Direction Générale

DR : Direction Régionale

EdT : enquête Emploi du Temps

FA : Fiche-Adresse

Fab : Fichier d’apprentissage brut

GSAS : Atelier de saisie (SAR) dans une DR

PCS : Catégorie socio-professionnelle à 4 chiffres

CS : Catégorie socio-professionnelle à 2 chiffres
CODER LA PROFESSION :
NOUVELLES PROCEDURES,
VIEUX PROBLEMES

A. CHENU (*) et F. GUGLIELMETTI (**)  
(*) CREST  
(**) INSEE - Unité "Méthodes Statistiques"

La profession des individus et leur situation par rapport au marché de l’emploi présentent de multiples facettes (métier, classification professionnelle, taille de l’établissement...). Il n’est donc pas étonnant que le codage statistique de la profession requière un questionnement lui-même multiple et prenne des formes différentes selon qu’il concerne des salariés ou des non-salariés, les salariés d’une petite entreprise ou ceux de la fonction publique, etc.

Le nombre et l’ordre des questions posées en vue du codage de la profession ou de la catégorie socioprofessionnelle varient d’un dispositif statistique à un autre. La diversité des dispositifs est grande. À son niveau à deux chiffres, la catégorie socioprofessionnelle peut être renseignée aussi bien dans des fichiers de DADS (déclarations annuelles de données sociales) que dans des données d’état civil, dans des recensements de la population que dans diverses enquêtes Insee auprès des ménages ou des employeurs, dans des sondages d’opinion à caractère politique ou commercial aussi bien que dans des documents administratifs de toute sorte. Ni le diplôme, ni le revenu, ni aucun autre indicateur de statut social ne présentent une telle «portabilité», ne circulent aussi facilement d’une sphère de description du monde social à une autre.
CS et PCS

Entrée en vigueur en 1982, la nomenclature des professions et catégories socioprofessionnelles (PCS) comporte, dans son niveau le plus détaillé (code à quatre chiffres), 455 positions élémentaires, correspondant à des ensembles homogènes de professions. Le nombre de positions passe à 553 dans la version rénovée (PCS 2000).

Les deux premiers chiffres du code définissent la catégorie socioprofessionnelle (CS). Détailée en 42 postes, elle peut être utilisée sous une forme regroupée en 24 postes pour les publications courantes. Le niveau le plus agrégé (premier chiffre du code) correspond aux grands groupes socioprofessionnels (8 postes).

Dans le langage à première vue commun des catégories socioprofessionnelles, les mêmes mots ne recouvrent pas toujours les mêmes objets. Les spécialistes de la mortalité différentielle savent par exemple qu’il y a bien plus d’employés au sens de l’état civil qu’au sens du recensement. Au sein d’un même dispositif, des changements apparemment minimes dans l’ordre des questions posées ou dans les procédures de traitement de l’information créent parfois de fortes discontinuités dans les séries. On sait qu’à une question sur la classification au sens des conventions collectives, la modalité « employé » est plus souvent choisie si elle vient avant l’item « cadre » que si elle lui fait suite, parce que dans le premier cas elle est prise comme synonyme de salarié, etc.

On se propose de présenter ici quelques observations permettant de caractériser la robustesse du codage socioprofessionnel tel qu’il est pratiqué au recensement et à l’enquête Emploi. Ces deux dispositifs vont connaître dans un proche avenir de substantielles réformes qui auront d’inévitables conséquences sur les chaînes de collecte et de traitement de la profession. On vise aussi à apporter une petite contribution aux vastes interrogations que suscite le devenir des nomenclatures sociales et professionnelles, dans un contexte marqué par des changements dans trois domaines interdépendants : demande sociale (administrations, presse, entreprises, milieux scientifiques...) en matière de statistiques sur les professions, l’emploi, la stratification sociale, technologie des enquêtes statistiques, intégration statistique internationale.
Un contexte changeant

Demande sociale : les « professions et catégories socioprofessionnelles » sont-elles toujours intéressantes ? Sait-on ce qu'on mesure lorsqu'on renseigne la CS ? La profession a-t-elle la même valeur aujourd'hui et hier comme indicateur de position sociale ?

Technologie des enquêtes statistiques : avec l'extension de la collecte assistée par ordinateur et la facilité croissante du traitement automatique des libellés en clair recueillis en réponse à des questions ouvertes, quelles modifications peut-on (ou doit-on), dans tel ou tel dispositif, apporter à la chaîne de collecte et de traitement qui aboutit au codage de la CS et/ou de la PCS ?

Intégration statistique internationale : le particularisme des catégories socioprofessionnelles à la française les vouez-il à une disparition à terme ? Quels argumentsjustifieraient le maintien de leurs spécificités, ou leur accès au statut de modèle, dans le cadre d'une Europe statistique tendant à promouvoir successivement une standardisation des nomenclatures sectorielles, professionnelles, sociales ?

On se limite ici à l'examen partiel du deuxième de ces trois volets.

1. Les catégories socioprofessionnelles à l'épreuve de la réitération


La réflexion amorcée ici s'appuie sur deux rapprochements. D'abord celui du codage de la PCS, pour les mêmes personnes interrogées deux fois, au recensement (codage semi-automatique selon la procédure Colibri) et à l'enquête Emploi de 1982 (codage

---

1 Cf. l'article de même titre de D. Merliié (1990).
2 Colibri : CODage en Lignes des Bulletins du Recensement Individuel
« manuel » à l'aide d'une documentation principalement constituée d'un guide et
d'un index alphabétique sur papier) [Chenu 1997]. Ensuite celui du codage de la
PCS et des variables connexes au test du recensement de 1997 à partir des mêmes
bulletins traités de manière semi-automatique d'une part, et selon une procédure de
contrôle principalement « manuelle » d'autre part3 [Guglielmetti 2000].

Au rapprochement de 1982, les divergences entre les deux chiffrements de la PCS
résultent tantôt des différences dans la nature des déclarations que les personnes ont
fournies dans le cadre du recensement et de l'enquête Emploi (« flou de
déclaration »), tantôt de celles dans le codage effectué par l'Insee (« flou de
classement »), tantôt d'un cumul des deux. Elles peuvent être particulièrement
fréquentes parce que 1982 est l'année où l'on met en oeuvre pour la première fois
une toute nouvelle nomenclature socioprofessionnelle [Desrosières et Thévenot
1989]. Par ailleurs les deux procédures ne sont pas de même nature : le questionnaire
de l'enquête Emploi est administré par un enquêteur, celui du recensement est
déposé puis récupéré par un agent recenseur ; le recensement « général » de la
population est une opération qui sollicite avant tout le civisme des répondants
[Kramarz 1991], tandis que le contexte d'une enquête focalisée sur l'emploi invite à
détailer la description des situations de travail ou de chômage.

3 Un des objectifs du test du recensement est d'évaluer et d'améliorer la chaîne de
codage de la PCS. La mise en oeuvre de l'application Sicore-PCS (Sicore : Système
Informatique de CODage par REconnaissance) permet de coder environ sept
bulletins sur dix. Les bulletins restants font l'objet d'une « reprise » manuelle (avec
la documentation papier), reprise elle-même informatiquement assistée par Sicore.
Tous les bulletins codés automatiquement font l'objet d'un codage de contrôle
manuel. Les autres sont codés une seconde fois en reprise. L'analyse des écarts entre
les deux codages conduit à des modifications de la base de règles Sicore et de
l'organisation de la reprise manuelle.

3 Les différences résultent de mécanismes complexes qui mériteraient une analyse
approfondie. A chaque divergence, il faut répondre à toute une série de questions :
existe-t-il une règle de codage précise ? Si oui, est-elle la même dans les tables du
codage automatique et dans les instructions papier ? Si elle est unique, pourquoi
n'est-elle pas appliquée dans un des cas : non connue, jugée erronée ? S'il n'existe
pas de consigne précise, quels critères implicites ont été mis en oeuvre dans les deux
codages ? Résultent-ils de la formation, de consignes générales locales, de l'initiative
individuelle ?
Au test de 1997, les différences résulent uniquement de variations dans les procédures de codage à partir d’un même bulletin⁴.

Globalement, parmi les actifs occupés au sens du recensement, le taux de divergence dans le codage de la PCS à quatre chiffres est de 43 % au rapprochement de 1982, et 29 % à celui de 1997.

**Taux de divergence entre les deux codages socioprofessionnels**

- **au recensement et à l’enquête Emploi de 1982**

<table>
<thead>
<tr>
<th>Groupe socioprofessionnel au recensement de 1982</th>
<th>Classement à l’enquête Emploi de 1982</th>
<th>score D⁴*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(pₓ) groupe différent</td>
<td>(pₓ) même groupe</td>
</tr>
<tr>
<td></td>
<td>au recensement de 1982</td>
<td>catégorie différente</td>
</tr>
<tr>
<td>1. agriculteurs</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>2. indépendants</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>3. cadres</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>4. professions intermédiaires</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>5. employés</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>6. ouvriers</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td>Ensemble</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

⁴ Les différences résultent de mécanismes complexes qui mériteraient une analyse approfondie. A chaque divergence, il faut répondre à toute une série de questions : existe-t-il une règle de codage précise ? Si oui, est-elle la même dans les tables du codage automatique et dans les instructions papier ? Si elle est unique, pourquoi n’est-elle pas appliquée dans un des cas : non connue, jugée erronée ? S’il n’existe pas de consigne précise, quels critères implicites ont été mis en œuvre dans les deux codages ? Résultent-ils de la formation, de consignes générales locales, de l’initiative individuelle ?

*INSEE Méthodes*
au traitement semi-automatique et au contrôle manuel du test de recensement de 1997

<table>
<thead>
<tr>
<th>Groupe socioprofessionnel</th>
<th>Classement au codage de contrôle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(P) groupe différent</td>
</tr>
<tr>
<td>1. agriculteurs</td>
<td>3</td>
</tr>
<tr>
<td>2. indépendants</td>
<td>8</td>
</tr>
<tr>
<td>3. cadres</td>
<td>11</td>
</tr>
<tr>
<td>4. professions intermédiaires</td>
<td>12</td>
</tr>
<tr>
<td>5. employés</td>
<td>8</td>
</tr>
<tr>
<td>6. ouvriers</td>
<td>5</td>
</tr>
<tr>
<td>Ensemble</td>
<td>8</td>
</tr>
</tbody>
</table>

*Voir l’encadré.


- Le premier, qui comprend les agriculteurs et les ouvriers, se définit par la sûreté de l’appartenance au groupement à un chiffre et l’incertitude des divisions internes (notamment au niveau des catégories à deux chiffres).

- Le deuxième, composé des employés et des indépendants, se prête à une bonne description en gros comme en détail (le niveau qui « marche » le mieux est celui des catégories à deux chiffres).

- Le troisième, avec les cadres et les professions intermédiaires, se caractérise par beaucoup de flottement au niveau le plus agrégé et des taux de divergence moyens (ou médiocres dans le cas des cadres en 1997) dans les classements de détail.

Un score global de « descriptibilité » statistique peut mesurer l’aptitude moyenne des doubles processus de codage à opérer des classements qui convergent à la fois en gros et en détail (voir l’encadré). En 1997 comme en 1982, il atteint ses valeurs les plus élevées chez les employés, les indépendants, les agriculteurs. Mais cet indicateur gomme la configuration différenciée des écarts de codage qui, très schématiquement, oppose les agriculteurs et les ouvriers, vastes ensembles aux divisions internes peu robustes, et les autres regroupements, dont les contours
globaux sont moins nets mais dont la structuration interne en professions ou en catégories socioprofessionnelles à deux chiffres est d'une meilleure lisibilité.

Un score de descriptibilité

Le croisement du codage de la PCS pour les mêmes personnes caractérisées à deux reprises permet de situer les catégories au long d'une échelle de « descriptibilité » à quatre positions.

Lorsque même le repérage le plus grossier, celui parmi l'un des six groupes socioprofessionnels, diffère aux deux sources (rapprochement enquête Emploi-recensement de 1982) ou au terme des deux procédures de codage des mêmes bulletins (test du recensement de 1997), la descriptibilité des personnes concernées est au niveau zéro. On se situe au niveau 1 lorsque la catégorie à deux chiffres diffère mais que le premier chiffre est le même, au niveau 2 lorsque la profession à quatre chiffres diffère mais que la catégorie est la même dans les deux dispositifs, et enfin au niveau 3, le plus élevé possible, lorsque les deux codages à quatre chiffres coïncident.

La série des quatre pourcentages \( p_0 \) à \( p_3 \) décrivant la composition d'une population d'effectif \( n \) au regard de cette échelle constitue le profil de descriptibilité de cette population dans la nomenclature des PCS. Sous une hypothèse rustique d'équidistance entre les barreaux successifs de l'échelle, on peut définir un score de descriptibilité socioprofessionnelle \( D \), égal à la moyenne des scores individuels :

\[
D = \frac{(p_1 + 2p_2 + 3p_3)}{3}
\]

\( D \) prend la valeur minimale 0 si aucun individu n'est classé dans le même groupe socioprofessionnel dans les deux dispositifs, et la valeur maximale 100 si tous les codages coïncident au niveau le plus détaillé de la nomenclature.

Le fait que les scores de descriptibilité soient médiocres dans le cas des cadres et membres des professions intellectuelles supérieures est surprenant. C'est au sein de cette catégorie que les emplois sont le plus fortement institutionnalisés. Les uns le sont sur le mode des corps de la fonction publique, d'autres se rapprochent des professions au sens anglo-saxon du terme (formation de niveau supérieur contrôlée par la profession elle-même, monopole d'exercice garanti légalement, déontologie spécifique). Dans tous les cas, les contours des professions ont en principe une grande visibilité sociale. La nomenclature est-elle trop détaillée ? Les procédures de questionnement et de codage sont-elles inadaptées ? Toujours est-il que la forte variabilité des codages requiert ici des explications particulières.

Les profils de descriptibilité des employés et des cadres varient beaucoup selon le sexe : la catégorisation des hommes comme cadres et des femmes comme employées, situations professionnelles assez typiques, est robuste, alors que celle des hommes
employés et des femmes cadres est plutôt incertaine (chiffres non reproduits ici; pour 1982, voir Chenu 1997).

L'observation des « échanges » entre les groupes, respectivement au recensement et à l’enquête Emploi, confirme la robustesse des phénomènes décrits ci-dessus.

Classement des actifs par groupe socioprofessionnel

- au recensement et à l’enquête Emploi de 1982

<table>
<thead>
<tr>
<th></th>
<th>Recensement</th>
<th>Enquête Emploi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>agriculteurs</td>
<td>indépendants</td>
</tr>
<tr>
<td>agriculteurs</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>indépendants</td>
<td>0</td>
<td>74</td>
</tr>
<tr>
<td>cadres</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>prof. intermédiaires</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>employés</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ouvriers</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ensemble</td>
<td>66</td>
<td>80</td>
</tr>
</tbody>
</table>

- au traitement semi-automatique et au contrôle manuel du test de recensement de 1997

<table>
<thead>
<tr>
<th></th>
<th>Codage semi-automatique</th>
<th>Codage de contrôle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>agriculteurs</td>
<td>indépendants</td>
</tr>
<tr>
<td>agriculteurs</td>
<td>66</td>
<td>0</td>
</tr>
<tr>
<td>indépendants</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>cadres</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>prof. intermédiaires</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>employés</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ouvriers</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ensemble</td>
<td>66</td>
<td>82</td>
</tr>
</tbody>
</table>

Dans les deux cas, ce sont les mêmes groupes qui échangent entre eux : cadres et professions intermédiaires, professions intermédiaires et employés, employés et ouvriers. Le rapprochement avec la mesure de la qualité au recensement de 1982 confirme cette très grande permanence des flottements entre les groupes [Guglielmetti 2000, p.33].

d'ouvriers) au sens du recensement sont classés dans les professions intermédiaires à l'enquête Emploi. Le groupe des professions intermédiaires est aussi celui qui, au test du recensement de 1997, donne lieu aux reprises les plus fréquentes (24% de libellés non reconnus). Le questionnement approfondi de l'enquête Emploi repère un plus grand nombre de situations « intermédiaires », celui du recensement réserve une plus large place aux descriptions les plus « typifiées », il durcit les contrastes.

Le schéma ci-après vaut pour le recensement comme pour l'enquête Emploi :

**Attractions entre catégories socioprofessionnelles du fait du double codage**

![Diagram](image)

Lecture :
- I : Indépendants
- C : Cadres
- PI : Professions intermédiaires
- E : Employés
- O : Ouvriers

L'épaisseur du trait est proportionnelle au volume des échanges.
La surface des cercles est proportionnelle aux effectifs.
2. Les libellés de profession en clair

La question ouverte sur la profession était, jusqu’au recensement de 1962, l’unique question (outre le secteur d’activité économique) posée en vue du codage du métier et de la catégorie socioprofessionnelle, et elle est toujours la première question posée dans la partie emploi du bulletin du recensement et dans les enquêtes auprès des ménages.

Si, au recensement, on se contentait de la déclaration de la profession en clair pour coder dans la nomenclature des PCS, on obtiendrait 28% de codes « erronés », dont 10% au niveau du groupe socioprofessionnel (source : test du recensement de 1997). Ce résultat est susceptible d’une interprétation contradictoire : le questionnement sur les « variables annexes » (statut, position professionnelle, fonction, activité économique de l’établissement employeur...), qui sont sollicitées à des degrés divers (le statut l’est presque toujours) est d’une grande importance pour le codage de la profession. Mais inversement, sept fois sur dix, les questions complémentaires défient in fine pour rien, ne fournissant qu’une information redondante. Le codage de la profession s’appuie avant tout sur un intitulé en clair.

On a vu que près d’une personne sur deux interrogée sur sa profession respectivement au recensement et à l’enquête emploi est classée à ces deux sources sous deux rubriques différentes. Cette divergence peut résulter pour partie de différences dans la façon dont est posée la question ouverte sur la profession, ainsi que de différences dans la nature des variables annexes et dans la façon dont elles sont sollicitées aux deux sources. Examinons dans un premier temps la forme de la question ouverte, et les caractéristiques des réponses en clair qui lui sont données.
Question ouverte sur la profession, au recensement et à l’enquête Emploi

<table>
<thead>
<tr>
<th>Recensement de mars 1982</th>
<th>Enquête Emploi de juin 1982</th>
</tr>
</thead>
<tbody>
<tr>
<td>(pour la profession antérieure)</td>
<td></td>
</tr>
<tr>
<td>(14) AVEZ-VOUS DEJA TRAVAILLÉ ?</td>
<td>A1 a - Profession principale. Intitulé précis de profession.</td>
</tr>
<tr>
<td>OUI → Quelle était votre profession principale ?</td>
<td></td>
</tr>
<tr>
<td>NON</td>
<td></td>
</tr>
<tr>
<td>(pour la profession actuelle)</td>
<td></td>
</tr>
<tr>
<td>(15) PROFESSION EXERCÉE ACTUELLEMENT</td>
<td></td>
</tr>
<tr>
<td>Soyez précis (Ex : électricien d'entretien de robot, comptable d'assurances, technicien chimiste, etc.)</td>
<td></td>
</tr>
<tr>
<td>Si vous êtes agent de la Fonction Publique, de l'État ou des collectivités (y compris HLM, hôpitaux publics), précisez votre grade (corps, catégorie...)</td>
<td>(suivent deux questions complémentaires pour les ouvriers ou exploitants agricoles et le personnel domestique).*</td>
</tr>
</tbody>
</table>

* L'Instruction aux enquêteurs indique :

Intitulé précis, mais concis, sans mots inutiles. Si des précisions viennent spontanément, marquez-les en A.1c et non à la suite en A.1a. Cet intitulé servira pour réaliser une codification automatique de la profession. Il ne doit pas comporter d'abréviations, ni de signes tels que parenthèses, guillemets, tirets, etc., mais peut comporter des sigles connus : SNCF, EDF, PTT, ou encore PL ou VL pour les chauffeurs. Pour les deux questions A.1a et A.1b, vous devez bannir toute indication de lieu (à la ville de ...) et en général les mentions d'entreprise, sauf si elles sont vraiment importantes (RATP, SNCF, Michelin...)

et, pour le grade :

... Il ne faut pas réécrire le grade s'il a été déjà mentionné comme profession ... Les sigles usuels sont permis, PEGC, AAP...

Ainsi les deux protocoles sont largement similaires. La différence majeure est que la question de l'enquête Emploi porte sur la profession principale, et que cette restriction est absente au recensement. Corrélativement, on rencontre au recensement de longs libellés, décrivant des activités multiples, difficiles à reconnaître et à coder. Au contraire, à l'enquête Emploi, les enquêteurs professionnels guident les répondants, sélectionnent et standardisent l'information qui leur paraît utile, et
enregistrent au total des libellés sensiblement plus courts, en moyenne, que ceux qui figurent sur les bulletins de recensement.

Cependant la liste des libellés les plus fréquents est presque la même aux deux sources. Les libellés en clair n’étant pas disponibles pour 1982, on compare, à un niveau agrégé, les professions en clair au test du recensement de 1997 (successivement pour la profession antérieure, la profession actuelle (libellé et grade, et libellé seul) et la profession principale (libellé seul) à l’enquête Emploi de 1998. La taille des échantillons est limitée à 70 000 répondants (taille de l’échantillon de l’enquête Emploi).

Ces 70 000 répondants utilisent :

- au recensement, pour déclarer leur profession antérieure, 17 700 libellés différents, dont 13 900 occurrences uniques ;

- au recensement, pour déclarer leur profession actuelle, 20 400 libellés différents (23 400 quand on ajoute le grade dans la fonction publique), dont 16 200 occurrences uniques ;

- à l’enquête Emploi, pour déclarer leur profession principale, 15 100 libellés différents, dont 10 100 occurrences uniques.

**Morphologie des libellés de profession en clair au recensement et à l’enquête Emploi**

<table>
<thead>
<tr>
<th></th>
<th>Recensement profession antérieure</th>
<th>Recensement profession actuelle (libellé + grade)</th>
<th>Recensement profession actuelle (libellé seul)</th>
<th>Enquête Emploi profession principale (libellé seul)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nombre de libellés différents</td>
<td>17700</td>
<td>23400</td>
<td>20400</td>
<td>15100</td>
</tr>
<tr>
<td>fréquence moyenne</td>
<td>4,0</td>
<td>3,0</td>
<td>3,4</td>
<td>4,6</td>
</tr>
<tr>
<td>Hapax (occurrences uniques)</td>
<td>13900</td>
<td>19100</td>
<td>16200</td>
<td>10100</td>
</tr>
<tr>
<td>Occurrences multiples:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nombre de libellés</td>
<td>3800</td>
<td>4300</td>
<td>4200</td>
<td>5000</td>
</tr>
<tr>
<td>nombre de répondants</td>
<td>56100</td>
<td>50900</td>
<td>53800</td>
<td>59900</td>
</tr>
<tr>
<td>fréquence moyenne</td>
<td>14,6</td>
<td>11,8</td>
<td>12,8</td>
<td>12,1</td>
</tr>
</tbody>
</table>

Source : test recensement 1997, enquête Emploi 1998

Quand un retraité ou un chômeur déclare son ancienne profession principale, le recul qu’il a par rapport à sa situation d’emploi le conduit à fournir une déclaration plus concise, plus simple, plus vague peut-être, que s’il était au travail, d’où un taux

---

3 Si l’enquête Emploi laisse une ligne à l’enquêteur pour décrire la profession, le recensement en propose deux au recensé.
moindre d’occurrences uniques et des fréquences plus élevées pour les libellés courants.

Les résultats pour l’enquête Emploi révèlent bien l’ampleur de la normalisation des libellés opérée par les enquêteurs : les hapax (libellés n’apparaissant qu’une fois) sont plus rares, mais en moyenne les descriptions sont plus concises qu’au recensement.

**Longueur des libellés au recensement et à l’enquête Emploi (pour 1000 répondants)**

<table>
<thead>
<tr>
<th>Nombre de mots</th>
<th>Recensement profession antérieure</th>
<th>Recensement profession actuelle (libellé + grade)</th>
<th>Recensement profession actuelle (libellé seul)</th>
<th>Enquête Emploi profession principale (libellé seul)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>126</td>
<td>229</td>
<td>270</td>
<td>343</td>
</tr>
<tr>
<td>2</td>
<td>743</td>
<td>484</td>
<td>505</td>
<td>542</td>
</tr>
<tr>
<td>3</td>
<td>111</td>
<td>179</td>
<td>167</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>71</td>
<td>32</td>
<td>11</td>
</tr>
<tr>
<td>5 ou plus</td>
<td>7</td>
<td>37</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td><strong>Longueur moyenne (en nombre de mots)</strong></td>
<td>2,0</td>
<td>2,2</td>
<td>2,1</td>
<td>1,8</td>
</tr>
<tr>
<td>- Hapax</td>
<td>2,6</td>
<td>3,2</td>
<td>2,9</td>
<td>2,5</td>
</tr>
<tr>
<td>- Occurrences multiples</td>
<td>1,9</td>
<td>1,9</td>
<td>1,8</td>
<td>1,7</td>
</tr>
</tbody>
</table>

*Source : test recensement 1997, enquête Emploi 1998*

La taille modale est toujours de deux. Mais les professions déclarées en deux mots représentent les trois quarts des déclarations de professions antérieures, et seulement la moitié des professions actuelles. C’est là que l’écart entre la taille des libellés rares et celle des libellés courants est la plus faible. On peut voir là aussi l’effet du temps sur la « stabilisation » de la déclaration : plus le temps passe, plus la déclaration se fige autour d’un libellé « décanté ».

Les distributions des libellés de profession actuelle au recensement sont les plus étalées : il y a deux fois plus de libellés d’au moins trois mots qu’à l’enquête emploi. Les réponses fournies par les personnes recensées sont moins standardisées que celles saisies par les enquêteurs, et comportent une plus grande part de redondance avec les informations provenant des questions ultérieures.
Neuf des dix libellés les plus fréquents sont communs aux deux listes. Parmi les 50 libellés fréquemment utilisés respectivement au recensement et à l’enquête emploi, on en retrouve 41 dans les deux sources. Ces 50 libellés sont ceux déclarés par 30 % des répondants à l’enquête Emploi, mais par 18 % au recensement. Ainsi les « secrétaires » (sans autre indication) représentent 2,6 % des répondants à l’enquête Emploi et 1,7 % au recensement.
Les 10 libellés les plus fréquents

<table>
<thead>
<tr>
<th>Libellé</th>
<th>N</th>
<th>/1000</th>
<th>Rang</th>
<th>EE</th>
<th>Libellé</th>
<th>N</th>
<th>/1000</th>
<th>Rang</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>secrétaire</td>
<td>1179</td>
<td>16,8</td>
<td>1</td>
<td>1</td>
<td>secrétaire</td>
<td>1789</td>
<td>25,6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>assistante maternelle</td>
<td>746</td>
<td>10,7</td>
<td>2</td>
<td>4</td>
<td>comptable</td>
<td>860</td>
<td>12,3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>femme de ménage</td>
<td>573</td>
<td>8,2</td>
<td>3</td>
<td>3</td>
<td>femme de ménage</td>
<td>845</td>
<td>12,1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>comptable</td>
<td>543</td>
<td>7,8</td>
<td>4</td>
<td>2</td>
<td>assistante maternelle</td>
<td>833</td>
<td>11,9</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>aide soignante</td>
<td>488</td>
<td>7,0</td>
<td>5</td>
<td>7</td>
<td>infirmière</td>
<td>769</td>
<td>11,0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>infirmière</td>
<td>371</td>
<td>5,3</td>
<td>6</td>
<td>5</td>
<td>agriculteur</td>
<td>758</td>
<td>10,8</td>
<td>6</td>
<td>44</td>
</tr>
<tr>
<td>agent</td>
<td>350</td>
<td>5,0</td>
<td>7</td>
<td>8</td>
<td>aide soignante</td>
<td>694</td>
<td>9,9</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>employée de bureau</td>
<td>323</td>
<td>4,6</td>
<td>8</td>
<td>16</td>
<td>agent d'entretien</td>
<td>687</td>
<td>9,8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>vendeuse</td>
<td>323</td>
<td>4,6</td>
<td>9</td>
<td>10</td>
<td>maçonnier</td>
<td>632</td>
<td>9,0</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>maçonnier</td>
<td>315</td>
<td>4,5</td>
<td>10</td>
<td>9</td>
<td>vendeuse</td>
<td>617</td>
<td>8,8</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

Source : test recensement 1997, enquête Emploi 1998


Il est donc nécessaire, pour aller plus avant dans l’explication de la variabilité des formes de déclaration et des résultats du codage, d’examiner le rôle des variables connexes. Deux questions principales permettent le codage de la catégorie socioprofessionnelle à deux chiffres, l’une a trait au statut (non salarié ou salarié, salarié de l’État, d’une entreprise, d’un particulier…), l’autre, posée aux seuls salariés, porte sur la position dans l’emploi (ouvrier non qualifié / ouvrier qualifié / employé / technicien / agent de maîtrise / ingénieur ou cadre). D’autres questions entrent surtout en jeu dans le chiffrage de la profession détaillée. On s’intressera ici à la position dans l’emploi, dont le rôle est tout à fait stratégique. Les nombreux réaménagements qu’ont subis les questions correspondantes témoignent de difficultés qui méritent une attention particulière. Sans l’information sur la classification dans l’emploi, 13,5% des salariés au test du recensement...

INSEE Méthodes

85
de 1997 - environ un bulletin sur sept - ne seraient pas codés dans la PCS qui leur est affectée sur la base de l'ensemble des renseignements collectés.

3. La position dans l'emploi, ou le statisticien embarrassé

Au recensement de 1954, la question ouverte sur la « profession principale au moment du recensement » était surmontée d'un long paragraphe de consignes invitant notamment les ouvriers à indiquer leur niveau de qualification, les fonctionnaires leur grade, etc. Parmi les personnes concernées, beaucoup omirent de donner ces renseignements ; la CSP fut chiffrée par imputation, tous les soudeurs salariés, par exemple, étant considérés comme des ouvriers qualifiés.

À partir de 1962, diverses questions fermées font suite à la question ouverte sur la profession. Une ou deux questions permettent d'abord de renseigner le statut ; en 1962, 1968 et 1975, on demande ensuite aux salariés de répondre à l'une des trois questions suivantes :

a - Si vous êtes ouvrier, précisez la qualification de votre emploi actuel (manœuvre /OS /OQ-OHQ)

b - Si vous êtes agent de l'État, d'une collectivité locale ou d'un service public (EDF, SNCF, etc.) ou militaire de carrière, précisez votre grade (question ouverte)

c - Si vous êtes dans un autre cas, précisez votre position hiérarchique (question ouverte).

En 1975, une question supplémentaire permet de ventiler les non-salariés en fonction du nombre de personnes qu'ils emploient. En 1982, elle est intégrée dans la question sur le statut.

En 1990, la question sur la position dans l'emploi connaît des modifications substantielles et comporte des listes d'une longueur accrue : « Indiquez la position professionnelle de votre emploi actuel » : manoeuvre ou OS / OQ-OHQ / agent de maîtrise dirigeant des ouvriers... / agent de maîtrise dirigeant des techniciens / technicien, dessinateur, VRP... / instituteur, assistant(e) social(e) ... / ingénieur ou cadre / professeur et personnel de catégorie A de la fonction publique / employé de bureau ... / autre cas. La question sur la fonction propose à tous les salariés une liste modifiée. Une personne qui a répondu « employé de commerce » à la question ouverte sur la profession coche « employé de commerce » à la question sur la position professionnelle et « commerce » à la question sur la fonction. Le taux de non réponses est relativement élevé.

En 1999, on retourne à la formulation de 1982 : « Indiquez la catégorie professionnelle de votre emploi ». La liste est de nouveau modifiée : les deux premiers postes (ouvriers) sont inchangés, viennent ensuite « agent de service, aide soignant(e), employé de maison », « employé de commerce, employé de bureau, personnel administratif de catégorie C ou D de la fonction publique ». La rubrique « autre cas » disparaît, laissant dans l’embarras les policiers ou les artistes. La liste des fonctions ne subit que des retouches mineures. La redondance entre les diverses questions n’est pas diminuée.


Des tests en cours pour le recensement rénové ou pour l’enquête Emploi marquent un éventuel retour au type de questionnement des années soixante, différencié selon que le répondant est ou non salarié de l’État. Ils portent aussi sur différents ordres des modalités : si la modalité « employé » vient tôt dans une liste, elle peut être comprise comme « salarié » ; une modalité « direction générale » est moins souvent choisie si elle vient après « cadre » que si elle vient avant, etc.

L’évolution d’ensemble, notamment depuis 1982, peut être caractérisée par un allongement des listes de positions dans l’emploi, et une plus grande redondance entre les questions. Un positionnement en termes de grilles des conventions collectives était suggéré par la liste ouvrier / employé / technicien / agent de maîtrise / ingénieur ou cadre proposée en 1982. Les salariés de l’État n’ont retrouvés que leur compte, mais ils pouvaient cocher « autre cas » et indiquer leur grade à la question suivante. Les listes ultérieures, plus copieuses et plus hétérogènes, comportent des termes qui peuvent être interprétés comme des noms de professions - voire, la culture statistique se diffusant dans le monde profane, comme des catégories socioprofessionnelles - plutôt qu’à des positions dans l’emploi ou à des classifications professionnelles proprement dites [Fermanian et Lantin, 1998]. Mais il est vrai que la seule référence aux grands postes constitutifs des grilles Parodi-
Croizat, vieilles de plus d’un demi siècle, pose problème dans un contexte où le langage des conventions de branche et d’entreprise évolue et où l’emploi de termes tels que « ouvrier », « employé », « agent de maîtrise » dans les libellés de profession est en recul [Burnod et Chenu, à paraître].

La difficulté à identifier la nature de l’information recueillie à la question sur la position dans l’emploi est patente dans le cas des techniciens. Un « technicien d’atelier » au sens de la convention collective de la métallurgie est en principe classé dans la catégorie socioprofessionnelle des ouvriers qualifiés de type industriel, et non dans celle des techniciens. Un technicien au sens des grilles d’emplois de la Sécurité sociale est un employé administratif d’entreprise au sens des CS. Une personne dont le libellé de profession est « technicien en organisation du travail » est codée dans une catégorie socioprofessionnelle de technicien, employé ou ouvrier selon la position dans l’emploi qu’elle déclare, et en employé administratif par défaut. Un « technicien en électroménager », quelle que soit la classification professionnelle qu’il déclare, est toujours codé en ouvrier qualifié. Et la formule « technicien de surface » est un libellé de profession souvent donné comme exemple d’école - en fait très rare - et conduisant à un codage non dans la catégorie socioprofessionnelle de technicien, mais dans celle d’ouvrier non qualifié de type artisanal.

La multiplication de cas témoignant de la polysémie du mot « technicien » ne met pas nécessairement en question la pertinence d’une rubrique « technicien » au sein d’une nomenclature socioprofessionnelle. Mais elle complique le travail du statisticien. À la question sur leur position professionnelle, certaines personnes peuvent répondre en des termes de jargon d’entreprise (le technicien de la sécurité sociale qui se déclare technicien), d’autre peuvent traduire ce jargon dans un langage interprofessionnel (le technicien de la sécurité sociale qui se déclare employé). Dans beaucoup de cas, les réponses aux autres questions (profession en clair, secteur d’activité...) assurent un codage socioprofessionnel satisfaisant. Mais l’ambiguïté est parfois irréductible.
Position professionnelle déclarée et codage socioprofessionnel

Test du recensement de 1997, codage semi-automatique. Champ : salariés

<table>
<thead>
<tr>
<th>Position professionnelle déclarée</th>
<th>Groupe socioprofessionnel codé</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>indépendants</td>
<td>cadres</td>
</tr>
<tr>
<td>cadres</td>
<td>7</td>
<td>136</td>
</tr>
<tr>
<td>professions intermédiaires</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>employés</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ouvriers</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Ensemble</td>
<td>10</td>
<td>144</td>
</tr>
</tbody>
</table>

| Décalés (%)**                     | //            | 6      | 20           | 16       | 14       | //    |

* Parmi les personnes ayant déclaré une position professionnelle donnée, part de celles qui sont classées par l’Insee dans une catégorie socioprofessionnelle dont l’intitulé est en décalage par rapport à la position professionnelle déclarée.

** Parmi les personnes classées par l’Insee dans une catégorie socioprofessionnelle donnée, part de celles qui se sont situées à une position professionnelle décalée par rapport à cette CS.

Dans 17 % des cas, il n’y a pas de correspondance simple entre la position professionnelle déclarée et l’appartenance à un groupe socioprofessionnel de la nomenclature. Ainsi, 20 % des « professions intermédiaires » au sens des catégories socioprofessionnelles sont des personnes qui ne se classent pas spontanément parmi les agents de maîtrise, les techniciens ou le cadre B de la fonction publique. 23 % des personnes qui se déclarent « ingénieur, cadre d’entreprise » ne sont pas classées parmi les cadres.

Certains des décalages entre position dans l’emploi et catégorie socioprofessionnelle relèvent du cas du « technicien de surface » qui déclare une position de technicien et que l’Insee range sans hésitation parmi les ouvriers non qualifiés. D’autres résultent de ce que certaines personnes interrogées ont du mal à comprendre sur quoi porte la question relative à la position professionnelle. Les variations dans la formulation de la question témoignent de ce que les statisticiens ont eux aussi du mal à cerner l’objet (« position dans l’emploi », ou « position professionnelle », ou « catégorie dans l’emploi », etc.). Ces difficultés sont fonction du caractère éminemment historique, daté et localisé, des termes dans lesquels employeurs et salariés conviennent de nommer les emplois et de déterminer les salaires. Il y a vingt ou trente ans, ces termes étaient souvent, dans la grande industrie, ceux des grilles Parodi-Croizat. Ils se sont toujours mal appliqués au monde de la boutique. La multiplication des conventions collectives faisant référence à des critères classants, celle des accords d’entreprise, et le recul du taux de syndicalisation des salariés s’accompagnent d’une diversification des appellations d’emploi. Une analyse plus approfondie des formes nouvelles de déclaration d’emploi et de leurs combinaisons avec les variables connexes est nécessaire à l’amélioration des chaînes de traitement statistique déterminant la position dans une nomenclature telle que la PCS.
4. Le libellé en clair, ressource majeure

Le libellé de profession en clair permet presque toujours de chiffrer une PCS à quatre chiffres, et, sept fois sur dix, la prise en compte des autres renseignements (position dans l'emploi, fonction, secteur d'activité...) n'entraîne aucune modification du codage. Ce libellé en clair est donc une ressource statistique majeure, dont des logiciels tels que Sicore permettent de tirer parti à assez peu de frais. La similitude des listes de libellés de profession en clair les plus fréquents au recensement et à l'enquête sur l'emploi de 1997-1998 indique que cette ressource est assez peu sensible au contexte de collecte, dépôt-retrait au recensement, questionnement en face à face à l'enquête Emploi.

Une des difficultés qui caractérise la collecte aux récents recensements de la population est liée aux redondances entre les différentes questions nécessaires au codage de la PCS, et aux variations dans l'interprétation de la question sur la position dans l'emploi. Dans des enquêtes avec questionnement en face à face assisté par ordinateur, l'usage de questions filtrés devrait permettre à l'avenir de réduire la redondance dans le questionnement. Une plus grande priorité donnée dans les règles de codage au traitement du libellé de profession en clair permettrait de limiter les risques de divergence entre la signification du codage dans des dispositifs par dépôt-retrait et dans des enquêtes en face à face.
Références bibliographiques


L’INTERNET :
UNE NOUVELLE VOIE POUR COLLECTER
LES ENQUÊTES DE BRANCHE EN FRANCE.

J-M. BEGUIN et V. DEROIN
SESSI

Dans le courant de l’année 2000, le SESSI (Service des études et des statistiques industrielles du Secrétariat d’État français de l’industrie) a progressivement proposé à toutes les entreprises (3018 en septembre) de l’échantillon des enquêtes mensuelles de branche qu’il gère directement, de substituer à la réponse habituelle par courrier, une collecte par l’internet sur le site "sessi.fr". De nombreux échanges par voie télématicque (EDI)1 existent déjà, quelquefois depuis plusieurs années, entre l’administration d’État et les entreprises2. Mais c’est la première fois en France que l’internet est utilisé par l’administration pour collecter les réponses d’entreprises à une enquête statistique.

Toutes ces enquêtes mensuelles sont à la base de l’élaboration de l’indice de la production industrielle, réalisée par l’INSEE. Pour le SESSI, il s’agit d’une enquête par sondage, dont l’échantillon représente en l’an 2000 un peu plus de 3.000 entreprises ; le taux de réponse final atteint 90% (en nombre d’entreprises répondantes). Les questionnaires imprimés par le SESSI et envoyés aux entreprises, par courrier, sont personnalisés, en ce sens qu’ils ne contiennent que les intitulés des produits fabriqués par l’entreprise et pour lesquels elle a été retenue dans l’échantillon. Mais les réponses des mois précédents ne sont pas rappelées et il n’y a pas de retour aux entreprises enquêtées leur donnant les résultats de l’enquête ; ceux-ci ne figurent que dans les publications du SESSI, voire de l’INSEE au niveau de l’indice de la production industrielle. Les données collectées sont peu nombreuses (une dizaine en moyenne) et simples : données de production ou de facturation par produit, dans la majorité des cas.

1 Échanges de données informatisés ou Electronic Data Interchange.
2 Par exemple, la collecte électronique des déclarations annuelles de résultats des entreprises est pratiquée par l’administration fiscale depuis une dizaine d’années.
1. Les objectifs poursuivis

Les objectifs ont été fixés très tôt et n’ont pratiquement pas varié pendant le développement du projet. Le projet s’inscrit d’abord dans un contexte général de simplification administrative et d’ouverture des services de l’État aux nouvelles technologies de l’information.

Le projet est présenté comme une expérimentation car il paraît illusoire d’espérer au début des taux de réponse très importants. À terme, on vise cependant des gains en termes de qualité (fiabilité et délais) et de coût, ainsi qu’une amélioration des relations avec les entreprises grâce aux retours d’informations fournis par le serveur.

Les deux premiers objectifs sont liés, bien évidemment, à la disparition de la phase de saisie des questionnaires, actuellement réalisée dans les équipes de gestion du Sessi ou sous-traitée dans le cas d’un nombre important de questionnaires : en effet, disparaît alors cette étape intermédiaire, source de coûts et d’erreurs potentielles.

Il est surtout prévu dès le départ de redonner à l’entreprise les résultats qu’elle a fournis lors des trois derniers mois. Elle sera donc en mesure de vérifier que les chiffres qu’elle a fournis sont exacts. Dans le fonctionnement actuel, il n’y a cependant pas de contrôle de validité, déclenché en temps réel, qui demanderait à l’entreprise de confirmer des variations qui pourraient nous paraître anormales. Ceci sera peut-être introduit dans la deuxième version du site.

Mais l’originalité principale réside dans la volonté d’offrir des informations, en retour, aux entreprises : informations publiques et générales (donc même accessibles sans mot de passe) concernant les objectifs de l’enquête, le cadre juridique, les nomenclatures utilisées, les liens avec d’autres sites internet du ministère,… et surtout informations personnalisées, sous forme graphique, permettant à l’entreprise de se positionner dans l’ensemble de la production française. On espère surtout ainsi transformer une relation purement régaliennne (l’État collecte les informations dont il a besoin) en une relation plus contractuelle (l’État fournit des informations en retour) dans laquelle le mode d’échange via l’internet permettra une forme de dialogue élargi avec les entreprises, plus commode que le courrier habituel ou le téléphone.

On ne négligera pas un dernier objectif de management interne : le fait de travailler sur un projet résolument moderne et novateur est une source de motivation non négligeable, non seulement pour les acteurs du projet, mais même au-delà pour l’ensemble des gestionnaires de l’enquête.
2. La chronologie du projet

Des premières réflexions à la mise en œuvre complète, il se sera écoulé environ trente mois ; le développement véritable n'a pris toutefois qu'environ six ou sept mois. Le calendrier final s'est en effet accéléré au 2ème semestre 99 dans le contexte suivant : les Organisations Professionnelles Agréées (OPA) qui réalisent environ la moitié des enquêtes de l'industrie, sont confrontées à des problèmes de moyens et abandonnent progressivement ces enquêtes. Début 2000, la Fédération des Industries Électriques, Électroniques et de la Communication (FIEEC) a ainsi « rendu » au SESSI l'enquête dont elle assurait jusque là la réalisation auprès de 300 entreprises environ. Or, elle proposait déjà, depuis 5 ans, une collecte par voie télématicque que près des deux tiers des entreprises utilisaient ; elle avait basculé cette collecte sur l'internet en septembre 1999 mais une petite centaine d'entreprises seulement avaient suivi le basculement sur l'internet. Le SESSI, en reprenant l'enquête, se devait d'être en mesure de proposer un service par Internet à ces entreprises dès janvier 2000.

Au total, les principales étapes ont été les suivantes :

**Mars 1996** : proposition du Sessi de mettre en place une collecte électronique de données sur les enquêtes de branche ; 3 techniques analysées : minitel (conclusion : peu ergonomique), internet (peu développé en entreprises), disquettes (choix peu moderne pouvant vite devenir dépassé).

**Août 1997** : le nouveau Premier ministre annonce son « programme d'action gouvernementale pour la société de l'information » (PAGSI) et en fait une orientation forte du gouvernement ; le SESSI relance ses réflexions dès octobre 1997, mais cette fois ci avec une optique internet.

**Octobre 1997 - mai 1998** : réflexions en interne ; création d'une structure de projet ; intervention courte d'un consultant pour préciser des choix techniques ; étude préalable ; choix d'externaliser le serveur auprès d'un fournisseur de services internet (FSI).


**Juillet 1999** : la FIEEC nous demande de reprendre à partir du mois de janvier 2000 les enquêtes qu'elle gérait jusque là en partie sur internet. Le projet devra être prêt pour cette date.

Février 2000 : ouverture du service pour les 85 entreprises de la FIEEC déjà habituées à ce support de communication pour la collecte du mois de janvier 2000 (sans cryptage ni nom de domaine).

Février à juin 2000 : achat du certificat et implémentation du nom de domaine


Septembre 2000 : ouverture du site pour les 2400 entreprises restantes (pour la collecte de juillet et août).

3. Les principaux choix du projet (stratégiques ou techniques)

La vie d’un projet est émaillée de choix et de décisions implicites ou explicites. On retrace ici ceux qui paraissent porteurs d’enseignements.

Réaliser \textit{un site simple, de fonctionnement très robuste, minimisant les temps de réponse}. Techniquement cela signifie que le site, développé par des informaticiens du SESSI, n’est pratiquement constitué que de pages HTML préconstruites ; il n’y a pas de composition "dynamique" des pages. L’entreprise donne son numéro SIREN et les pages personnalisées correspondantes s’affichent sur le PC de l’entreprise\(^4\). Nous nous félicitons plutôt de ce choix qui nous a permis de tenir les délais imposés par la reprise de l’enquête de la FIEEC, de nous familiariser avec les techniques internet, sans gérer trop de nouveautés à la fois, et de n’avoir rencontré aucun problème de fonctionnement lorsque nous sommes passés en production.

\textit{Travailler par étapes, avec des maquettes successives, validées par le groupe de projet} : il nous a paru important d’associer les gestionnaires et les entreprises à l’ergonomie du site. Avant le démarrage en production, quelques entreprises de la FIEEC ont accepté de participer à des tests.

\(^3\) En application des consignes gouvernementales, le nom du site aurait dû comporter le nom du ministère et le suffixe « gouv » (pour gouvernement). On demandait donc une exception aux règles de nomenclature.

\(^4\) En simplifiant un peu les choses, pour 3000 entreprises, le serveur contient donc quelques 12000 pages HTML, pré-composées pour chaque mois de collecte à raison de 4 pages par entreprise : 3 pages rappellent les 3 mois précédents et la 4\(^e\) sert à collecter le mois courant.
Séparer les pages publiques et les pages de résultats individuels : on a vu que cette double offre était un des objectifs du projet ; nous nous étions même posé la question de les mettre ou non sur le même serveur. Pour l'instant, les pages publiques ne semblent guère consultées, mais le site est encore jeune. Viennent aussi d'être ajoutées des pages d'informations liées à une enquête thématique.

Mettre l'accent sur la sécurité : celle-ci est double : pour accéder aux pages protégées, l'entreprise doit fournir un mot de passe de 8 caractères et toutes les données sont protégées par un cryptage dit de type SSL s'appuyant sur une clef de 40 bits (données circulant sur le net cryptées lors de leur envoi, puis décryptées à l'arrivée sur notre serveur). Cette clef est fournie par l'achat d'un certificat qui authentifie également le site auprès des entreprises répondantes. En fin de session, l'entreprise est informée de ce que ses données sont bien parvenues au serveur. Le dispositif nous donne satisfaction pour l'instant.

Réaliser une montée en charge progressive du serveur, et proposer une assistance téléphonique (hot-line) : la volonté d'assurer une continuité de service avec la collecte assurée par la FIEEC jusqu'en décembre 1999 nous a conduits à être prêts à ouvrir le site dès le début 2000 pour les 85 entreprises qui répondaient par internet à la FIEEC. A cette date, ni le nom de domaine, ni le certificat n'avaient pu être obtenus. Nous avons malgré tout ouvert le site, auquel les entreprises accédaient directement par son adresse IP et sans cryptage des transmissions (ce qui correspondait d'ailleurs aux conditions dans lesquelles les entreprises communiquaient avec la FIEEC jusque là). Le dispositif d'assistance comportait une personne susceptible d'aider les entreprises : celles-ci peuvent avoir des problèmes de paramétrage des navigateurs, des réseaux, de mots de passe, etc... Il est impératif de pouvoir les aider. Nous avons ouvert ensuite auprès de 600 entreprises au mois de juin et, devant le faible taux d'utilisation (5%) qui ne saturait pas notre hot-line, étendu aux 2400 dernières, début septembre.

Informer les entreprises de l'ouverture du service, uniquement par courrier : nous nous sommes contentés pour l'instant d'envoyer deux lettres, l'une au correspondant qui répond habituellement aux enquêtes ; l'autre au Président Directeur Général (PDG) de l'entreprise, insistant sur le volet "retour d'informations" qui donne la position de l'entreprise vis-à-vis de l'ensemble de la production française, ce qui nous paraissait spécifiquement intéressant pour un dirigeant ; il est à noter que même si

---

5 Au départ. Elle peut ensuite le modifier.
6 Il s'agit d'une clef asymétrique. Au delà de 40 bits, il faut des autorisations gouvernementales très spécifiques que nous n'avons pas demandées. Avec la prochaine loi sur la Société de l'Information qui précise les conditions de mise en œuvre de la signature électronique, on peut espérer pouvoir rallonger la longueur de la clef.
7 Choisies au hasard ; il n'y a pas eu de critère de sélection sur la taille ou le secteur ; il s'agissait juste d'une montée en charge. Au départ, on imaginait des vagues successives de 600 entreprises.
l'entreprise a répondu par courrier et non par l'internet, elle peut consulter le site et voir les graphiques la concernant ainsi que les chiffres globaux de la branche. Lors de la mise en place pour les entreprises de la FIEEC, nous n'avions envoyé le mot de passe qu'au PDG, voulant ainsi qu'il "autorise" formellement le correspondant à répondre par l'internet. Ce fut une erreur ; la plupart n'ont pas compris qu'il fallait qu'ils transmettent le mot de passe à nos correspondants habituels et ceux-ci n'arrivaient pas à l'obtenir. Lors des 2 vagues suivantes, nous avons décidé de transmettre directement les mots de passe à nos correspondants dans les entreprises, ce qui n'a généré aucun problème. L'information est complétée par un rappel sur les questionnaires de la possibilité de répondre par Internet et/ou de consulter des résultats de l'ensemble du secteur d'activités.

Rendre aussi transparente que possible, pour les gestionnaires de l'enquête au SESSI, cette possibilité offerte aux entreprises : nous avons pu rendre le questionnaire internet quasiment identique au questionnaire papier. Les gestionnaires sont informés par e-mail dès qu'une entreprise a répondu et ils peuvent consulter la réponse (l'imprimer au besoin), immédiatement. Les réponses parvenues par l'internet sont intégrées dans l'application de gestion de l'enquête exactement comme si elles résulteraient de la saisie des questionnaires papier. Ces mises à jour sont réalisées toutes les nuits par un programme qui explore le serveur, repère les questionnaires remplis par les entreprises dans la journée et les charge dans l'application de gestion. Pour un test, cette façon de procéder minimise à l'évidence les risques. Pour des développements ultérieurs, il conviendra probablement de tirer davantage partie de la spécificité d'internet et de ne pas se contenter de "reproduire" les questionnaires papier.

4. Les coûts

Sur le plan matériel, il a fallu acquérir un serveur physique (100 kF), un serveur web de Microsoft (inclus dans Visual interdev), (5kF), un certificat (1kF/an) acheté auprès de la société Thawte, et un logiciel de contrôle d'accès par mots de passe (de la société SecuriSite) (environ 20kF). On ne compte pas ici les compilateurs des langages utilisés, ni les dispositifs anti-intrusion (fire-walls) que le ministère avait déjà acquis par ailleurs. Le coût d'investissement est donc inférieur à 140kF (environ 20 000$ US).

8 Il faut bien sûr dans ce cas qu'elle attende que son courrier nous soit parvenu et que ses données aient été saisies.
9 Ceci peut paraître tout à fait évident, mais ne l'est pas compte tenu des problèmes de sécurité : pour éviter que des personnes de l'extérieur du ministère et de l'intérieur ne puissent accéder en même temps, au même serveur, il a fallu le dupliquer et réaliser un site miroir.
10 Il s'agit d'une filiale de Verisign.
Les coûts de développements, y compris le chargement des données collectées dans l'application de gestion existante, ont été de l'ordre de 90 jours x homme, auxquels il faut ajouter le temps consacré par l'encadrement, la rédaction d'un manuel, la formation des gestionnaires. On retiendra le chiffre total de 100 j/h.

Les coûts de maintenance informatique sont très faibles. Les coûts de production, en revanche, le sont un peu moins : d'un côté, la nécessité de disposer d'une assistance téléphonique mobilise de 5 à 10 % (selon que l'on se trouve en début ou en cours d'enquête) d'un cadre à temps plein ; de l'autre, le fait que la collecte par internet se superpose à une application existante mobilise un opérateur environ 20 minutes tous les jours.

5. Les résultats et les réactions des entreprises

Le taux de réponse spontanée initial, hors entreprises de la FIEEC, (5,5%) est resté plutôt inférieur à ce à quoi nous nous attendions.

Cependant, depuis l'ouverture à la totalité des entreprises, le taux des répondantes a atteint 8,6% (hors FIEEC). Le fait de rappeler sur le questionnaire papier l'existence du site devrait permettre de gagner régulièrement de nouvelles entreprises.

"L'erreur" de communication du mot de passe vis-à-vis des 85 entreprises qui répondaient à la FIEEC par l'internet nous en a fait "perdre" quelques-unes : en septembre, nous n'avions récupéré que 75 d'entre elles. De même, sur les 174 lettres envoyées uniquement au Président Directeur Général par défaut de connaissance du correspondant mensuel, seules 2 réponses par internet nous sont parvenues.

À la mi-octobre, 312 entreprises (y.e. Fieec) avaient utilisé Internet au moins une fois pour retourner leur réponse ; cependant le maximum de réponses obtenues pour un même mois est de 284. Cette instabilité de réponse s'explique par le fait que 10 entreprises anciennement Fieec n'ont pas utilisé le site sessi.fr pour répondre (cf. paragraphe ci-dessus) et aussi du fait que plusieurs entreprises restent non répondantes pendant 2 ou 3 mois puis envoient le même jour leurs 2 ou 3 questionnaires.

Au total, pour 284 entreprises ayant choisi à la mi-octobre de répondre via l'internet, nous avons reçu environ 110 coups de téléphone ; on notera d'ailleurs que les entreprises nous contactent encore beaucoup plus volontiers par téléphone que par e-mail ; mais c'est probablement logique pour du dépannage.

En début de chacune des 3 phases d'ouverture, les demandes des entreprises ont porté essentiellement sur les mots de passe, les problèmes de sécurité (ou associés au dispositif SSL) et de paramétrage des navigateurs.
En cours d'enquête, les autres appels concernent des demandes d'aide à la navigation suite à des erreurs de saisie de Siren, de formats de valeurs, de visualisation des données.

Nous notons très peu de réactions sur la convivialité du site ou le contenu des pages statistiques.

Les conclusions, à la mi-octobre, de l'analyse des 312 Siren (y.c. fieec) ayant répondu via Internet sont les suivantes:

1) la ventilation selon la taille de l'entreprise montre une utilisation croissante de l'internet en fonction de la taille : 21 % de nos répondantes par internet sont des entreprises de 20 à 249 salariés, contre 54 % pour les entreprises de 1000 salariés et plus.
2) la ventilation selon le secteur d'activités de l'entreprise montre une prépondérance des industries de pointe, électriques, électroniques, de communication : leur taux de réponse avoisine les 30% (rappelons que les entreprises Fieec sont incluses dans l'analyse de ces 312 Siren). La chimie, le verre et les moules et modèles répondent aussi bien par ce moyen (environ 10%) ; tous les autres secteurs sont dans la moyenne du taux de réponse, l'industrie du meuble reste en retrait avec seulement 3% de répondantes.

<table>
<thead>
<tr>
<th>Secteurs</th>
<th>Taux de réponse internet</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 - Industrie textile</td>
<td>7 %</td>
</tr>
<tr>
<td>20 - Travail du bois</td>
<td>6 %</td>
</tr>
<tr>
<td>22 - Edition, imprimerie, reproduction</td>
<td>7 %</td>
</tr>
<tr>
<td>23 - Cokéfaction, raffinage, industries nucléaires</td>
<td>8 %</td>
</tr>
<tr>
<td>24 - Industrie chimique</td>
<td>11 %</td>
</tr>
<tr>
<td>25 - Industrie du caoutchouc et des plastiques</td>
<td>5 %</td>
</tr>
<tr>
<td>26 - Fabrication d'autres produits minéraux non métalliques</td>
<td>10 %</td>
</tr>
<tr>
<td>27 - Métallurgie</td>
<td>5 %</td>
</tr>
<tr>
<td>28 - Travail des métaux</td>
<td>6 %</td>
</tr>
<tr>
<td>29 - Fabrication de machines et équipements</td>
<td>9 %</td>
</tr>
<tr>
<td>30 - Fabrication de machines de bureau et matériel informatique</td>
<td>29 %</td>
</tr>
<tr>
<td>31 - Fabrication de machines et appareils électriques</td>
<td>28 %</td>
</tr>
<tr>
<td>32 - Fabrication d'équipements de radio, de télévision, communication</td>
<td>32 %</td>
</tr>
<tr>
<td>33 - Fabrication d'instruments médicaux, de précision, d'optique,</td>
<td>33 %</td>
</tr>
<tr>
<td>d'horlogerie</td>
<td></td>
</tr>
<tr>
<td>36 - Fabrication de meubles ; industries diverses</td>
<td>2 %</td>
</tr>
</tbody>
</table>

3) la ventilation selon la région montre une prépondérance des réponses provenant d'Ile-de-France et du Nord-Est : 16 % des répondantes sont implantées en Ile-de-France, 15 % en Alsace, 13 % en Franche Comté, 11 % en Lorraine et 10 % en Nord pas-de-Calais. La région Rhône-Alpes est aussi bien représentée avec 9 % des répondantes. Cf. carte page suivante.

En croisant ces 3 critères "taille-secteur-région", on note un effet taille par rapport à la région Ile-de-France: le pourcentage important des répondantes en Ile-de-France s'explique par le fait que c'est aussi là que sont implantées les plus grosses entreprises. Sans mener une étude plus approfondie, on ne détecte donc pas a priori d'effet géographique propre.
De même en ce qui concerne le secteur d'activité : les secteurs 30 et 32 (fabrication de matériels informatiques, d'équipements de radio, télévision et de communication) sont ceux qui concentrent les entreprises de plus grande taille.

Enfin, on note encore que la région Alsace, seconde région après l'Ile-de-France en termes de réponses internet, est aussi celle où sont le plus implantées les entreprises des secteurs 29 et 31 (fabrication de machines et équipements, et matériels électriques).

La répartition régionales des entreprises répondant par Internet

Résultats de la nouvelle campagne 2001, au 15/05/2001 :

L'enquête mensuelle 2001 concerne un échantillon de 3865 entreprises.

La totalité de ces entreprises a été informée de la possibilité de répondre par internet dès le lancement de l'enquête en février.
En 2001, le taux de réponse par internet est de 12 % sur la période de janvier ; il a atteint 16,9 % pour la période de mars 2001.

On observe donc une augmentation sensible du nombre et du pourcentage d'entreprises qui préfèrent utiliser ce moyen de réponse, à relier très probablement à l'accroissement des accès à internet dans les entreprises.

Une analyse des 611 entreprises répondantes (au 15/05/01) confirme la ventilation déjà établie sur les répondantes de 2000.

En termes de région, l'Ile-de-France et l'Alsace restent prépondérantes (respectivement 22 % et 24 %).

Les secteurs d'activités des industries de pointe, électriques, électroniques et de communication sont toujours les mieux représentés (Naf 30 = 57 % ; Naf 31 = 30 % ; Naf 32 = 44 %).

Les entreprises de plus de 1000 salariés ont répondu par internet pour 33 % d'entre elles, contre 13 % pour les entreprises de 20 à 249 salariés.

6. Les principaux enseignements

Une fois la décision prise et les objectifs fixés, la réalisation d'un site de ce type n'est finalement ni très coûteuse, ni très longue si l'on veut bien exclure la phase de maturation du projet pendant laquelle il n'y a pas vraiment eu de travail réalisé, et les délais induits par la réorganisation du ministère. Ce fut, en tout cas dans le cas présent, un projet assez modeste, réalisé par une équipe qui a acquis une compétence internet pour l'occasion. Les coûts sont faibles et les délais ont été tenus, au moins pour prendre le relais de la FIEEC.

Les difficultés principales sont venues du côté administratif (obtention du nom du domaine et du certificat) où nous ne les attendions pas mais où nous avons pu fort heureusement être aidés dans le contexte du ministère, même si l'inclusion du site dans l'infrastructure ministérielle a été en elle-même source de complexité.

Si c'était à refaire, nous reprendrions très probablement les mêmes choix techniques et organisationnels, à l'exception de la communication sur le mot de passe.

En revanche, nous sommes plutôt désagréablement surpris par le faible taux de réponse spontanée. Nous savons qu'environ 70% des entreprises industrielles françaises disposent d'un accès à l'internet ; mais ceci ne signifie pas que nos

---

11 Après l'ouverture par la FIEEC pour la collecte de janvier, nous pensions ouvrir réellement le service pour la collecte des données relatives au mois de février, alors que nous n'avons pu le faire que sur maiw.
correspondants habituels en disposent ; il semble au contraire que l'accès à l'internet soit géré dans les entreprises de façon assez restrictive. Nous ne connaissons pas non plus leur adresse de courrier électronique, s'ils en ont une, pour les relancer chaque mois. Aucun dispositif d'enquête ne s'est préoccupé jusqu'à présent de collecter et conserver ce type d'information, plus difficile à gérer car attaché à la personne et pas seulement à l'entreprise ou à la fonction dans l'entreprise, comme peuvent l'être des numéros de fax ou de téléphone. Il ne suffit certainement pas, en tout cas, de proposer une collecte via l'internet, pour susciter la réponse des entreprises.

7. L'avenir

À court terme, et en premier lieu, il nous faut mieux faire connaître le site, par des annonces sur les questionnaires papier, des informations auprès des syndicats professionnels et l'enregistrement dans nos fichiers des adresses électroniques de nos correspondants que l'on demandera désormais systématiquement dans nos enquêtes. Nous étudierons aussi la possibilité d'interventions directes orales auprès des industriels (salons, réunions professionnelles).

Une enquête qualité sera réalisée en automne 2001, pour mesurer le degré de satisfaction des entreprises qui utilisent le service (différencier les pages de questionnaires des pages de retour d'informations) et les raisons de la non utilisation des autres et mettre en place un plan d'actions pour inciter les entreprises à utiliser l'internet.

À moyen terme (1 à 2 ans), le site qui se veut comme on l'a vu expérimental, sera refait à la suite de la refonte générale de l'application qui permet d'exploiter l'enquête. Même si les objectifs n'en sont pas encore complètement fixés, on peut songer aux directions suivantes :

- intégrer les 2 applications de collecte papier et internet dans l'application globale de gestion de l'enquête ;

- changer le principe de fonctionnement du serveur (générer les pages dynamiquement au lieu de charger des pages statiques) ;

- rajouter des contrôles de qualité interactifs ;

- améliorer les retours d'information : ceci fait d'ailleurs l'objet d'un sous-projet, financé par Eurostat, qui doit être mené dès 2001. Nous souhaiterions en particulier pouvoir différencier la personne qui remplit le questionnaire de celle qui recevra les retours d'information ;

- développer l'interactivité avec les entreprises : nous espérons proposer par exemple aux grosses entreprises de pouvoir répondre en nous envoyant directement des fichiers issus de leur système d'information, sans passer par une phase de saisie à l'écran.
- améliorer la sécurité dans le cadre de la mise en place par le ministère de l'économie d'une infrastructure de gestion de clés publiques qui permettront aux entreprises de s'authentifier.

Nous tâcherons enfin de tirer les leçons complètes de cette expérimentation pour étendre le principe de collecte par l'internet à d'autres enquêtes. D'ores et déjà, la documentation de la prochaine enquête thématique du SESSI, qui porte sur les échanges internationaux intra-groupes, est accessible sur le site. Il sera intéressant de voir si cette possibilité est exploitée par les entreprises ou non.
Annexe

La ventilation des 312 répondantes internet selon la taille de l'entreprise :

<table>
<thead>
<tr>
<th>Tranche d'effectif</th>
<th>Nombre d'entreprises EMB</th>
<th>Nombre de répondantes Internet</th>
<th>Taux des répondantes Internet</th>
</tr>
</thead>
<tbody>
<tr>
<td>02 (hors tranche, non D.O.)</td>
<td>72</td>
<td>9</td>
<td>12 %</td>
</tr>
<tr>
<td>30 (20 à 49 sal)</td>
<td>499</td>
<td>26</td>
<td>5 %</td>
</tr>
<tr>
<td>40 (50 à 99 sal)</td>
<td>635</td>
<td>42</td>
<td>7 %</td>
</tr>
<tr>
<td>50 (100 à 249 sal)</td>
<td>880</td>
<td>82</td>
<td>9 %</td>
</tr>
<tr>
<td>60 (250 à 499 sal)</td>
<td>415</td>
<td>44</td>
<td>11 %</td>
</tr>
<tr>
<td>70 (500 à 999 sal)</td>
<td>293</td>
<td>50</td>
<td>17 %</td>
</tr>
<tr>
<td>80 (1000 à 1999 sal)</td>
<td>134</td>
<td>31</td>
<td>23 %</td>
</tr>
<tr>
<td>90 (2000 et + sal)</td>
<td>86</td>
<td>27</td>
<td>31 %</td>
</tr>
</tbody>
</table>

Source : les entreprises répondantes au 16/10/2000
La ventilation des 312 répondantes internet selon la région d'implantation de l'entreprise :

<table>
<thead>
<tr>
<th>Région</th>
<th>Nombre d'entreprises EMB</th>
<th>Nombre d'entreprises internet</th>
<th>Taux réponse internet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile de France</td>
<td>781</td>
<td>125</td>
<td>16 %</td>
</tr>
<tr>
<td>Alsace</td>
<td>107</td>
<td>17</td>
<td>15 %</td>
</tr>
<tr>
<td>Franche Comté</td>
<td>88</td>
<td>12</td>
<td>13 %</td>
</tr>
<tr>
<td>Lorraine</td>
<td>109</td>
<td>13</td>
<td>11 %</td>
</tr>
<tr>
<td>Nord Pas-de-Calais</td>
<td>147</td>
<td>15</td>
<td>10 %</td>
</tr>
<tr>
<td>Rhône-Alpes</td>
<td>475</td>
<td>43</td>
<td>9 %</td>
</tr>
<tr>
<td>Centre</td>
<td>135</td>
<td>12</td>
<td>8 %</td>
</tr>
<tr>
<td>Haute-Normandie</td>
<td>91</td>
<td>8</td>
<td>8 %</td>
</tr>
<tr>
<td>Midi-Pyrénées</td>
<td>84</td>
<td>7</td>
<td>8 %</td>
</tr>
<tr>
<td>Limousin</td>
<td>25</td>
<td>2</td>
<td>8 %</td>
</tr>
<tr>
<td>PACA</td>
<td>66</td>
<td>5</td>
<td>7 %</td>
</tr>
<tr>
<td>Picardie</td>
<td>120</td>
<td>9</td>
<td>7 %</td>
</tr>
<tr>
<td>Poitou-Charente</td>
<td>54</td>
<td>4</td>
<td>7 %</td>
</tr>
<tr>
<td>Pays de Loire</td>
<td>179</td>
<td>12</td>
<td>6 %</td>
</tr>
<tr>
<td>Aquitaine</td>
<td>96</td>
<td>6</td>
<td>6 %</td>
</tr>
<tr>
<td>Champagne-Ardenne</td>
<td>84</td>
<td>5</td>
<td>5 %</td>
</tr>
<tr>
<td>Bretagne</td>
<td>85</td>
<td>5</td>
<td>5 %</td>
</tr>
<tr>
<td>Languedoc-Roussillon</td>
<td>35</td>
<td>2</td>
<td>5 %</td>
</tr>
<tr>
<td>Bourgogne</td>
<td>108</td>
<td>6</td>
<td>5 %</td>
</tr>
<tr>
<td>Auvergne</td>
<td>85</td>
<td>3</td>
<td>3 %</td>
</tr>
<tr>
<td>Basse-Normandie</td>
<td>58</td>
<td>1</td>
<td>1 %</td>
</tr>
</tbody>
</table>

*Source : les entreprises répondantes au 16/10/2000*
LE TELEPHONE MOBILE
DANS LES ENQUETES PAR SONDAGE

G. ROY et A. VANHEUVERZWYN

MEDIAMETRIE - Direction Recherche et Méthodes

1. Introduction

1.1. Le développement des enquêtes par téléphone

L'usage du téléphone dans les enquêtes par sondage s'est sans cesse affirmé depuis l'avènement des enquêtes d'opinion et ce dans des domaines très diversifiés. En effet, le téléphone est désormais souvent préféré au face à face et ce pour de multiples raisons (simplicité du contrôle, homogénéité de la passation du questionnaire, évacuation des difficultés d'accès aux personnes ou foyers enquêtés, coût moindre et taille d'échantillon plus importante). Même s'il est peut-être plus facile de refuser de répondre lors d'une enquête par téléphone que dans le cadre du face à face, le côté « impersonnel » et neutre du téléphone permet d'aborder plus facilement des sujets sensibles et de toucher des populations a priori plus réticentes, comme par exemple les personnes âgées.

L'existence d'un annuaire utilisable comme une base de sondage a également favorisé l'utilisation du téléphone dans les enquêtes par sondage. En France comme à l'étranger, les opérateurs de téléphonie filaire publient les annuaires des abonnés (Les pages blanches pour France Télécom) comprenant l'ensemble des numéros attribués et la commune de localisation de la ligne téléphonique. Cette dernière information est très utile notamment dans le cadre de la stratification sur des critères géographiques comme la région ou la taille de la commune.

De plus, Médiamétrie a mis au point, en 1990, une méthode de génération aléatoire de numéros permettant d'atteindre des ménages en liste rouge, c'est-à-dire dont le numéro n'apparaît pas dans l'annuaire. Lorsque l'appel d'un numéro demeure infructueux après quatre rappels consécutifs, la méthode consiste à incrémenter ce numéro de un. Parmi les numéros générés par ce procédé, certains sont attribués et d'autres ne le sont pas. Une partie des numéros attribués apparaissent dans l'annuaire, les autres sont sur liste rouge [2].

Mais l'utilisation du téléphone filaire comme mode d'administration d'enquêtes par sondage sous-entend que les foyers équipés de téléphone filaire sont représentatifs de l'ensemble de la population des unités enquêtées (foyers ou individus), c'est-à-
dire, d’une part, que le taux d’équipement en téléphone filaire est élevé et, d’autre part, que les foyers non-équipés de téléphone filaire n’ont pas un comportement atypique au regard des variables d’intérêt. Si ces hypothèses sont encore vérifiées aujourd’hui, il est légitime de se demander si l’arrivée du téléphone mobile ne pourrait pas nuancer cette situation.

1.2. Évolution du taux d’équipement en téléphone filaire

Alors qu’en 1997, 95,8% des ménages de France métropolitaine étaient équipés d’un téléphone filaire\(^1\), ils sont 91,5% au deuxième trimestre 2000 selon l’enquête 24000 Multimédia. On s’éloigne donc progressivement de la configuration qui avait permis le développement des enquêtes par téléphone : 8,5% des ménages sont aujourd’hui exclus des enquêtes par téléphone filaire. Près des trois quarts de ces ménages sont équipés d’au moins un téléphone mobile.

Tableau 1 : Taux d’équipement des ménages français en téléphone filaire

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9,3</td>
<td>14,9</td>
<td>30,2</td>
<td>51,0</td>
<td>90,0</td>
<td>93,7</td>
<td>95,8</td>
<td>94,3</td>
<td>92,6</td>
<td>91,5</td>
</tr>
</tbody>
</table>

* : Source INSEE.
** : Source enquête 24000 Multimédia de Médiamétrie.
*** : Source enquête 24000 Multimédia de Médiamétrie. 2\(^{ème}\) trimestre 2000.

Selon l’enquête 24000 Multimédia de Médiamétrie, 6,6% des individus de 15 ans et plus n’étaient pas équipés de téléphone filaire au 2\(^{ème}\) trimestre 2000, parmi lesquels 5,2% disposaient d’un téléphone mobile, exclusivement donc. Ainsi, même si le taux de pénétration du téléphone filaire est en légère régression et que le taux de pénétration du téléphone mobile augmente, de plus en plus de personnes sont joignables par téléphone - mobile ou filaire : au 2\(^{ème}\) trimestre 2000, 98,6% des individus de 15 ans et plus étaient joignables par téléphone ; la substitution au téléphone filaire semble donc en marche.

On sait par ailleurs que la structure socio-démographique des exclusifs du mobile est très différente de celle des équipés de téléphone filaire, notamment sur les critères d’âge et de catégorie socio-professionnelle : ils sont plus jeunes, plutôt CSP- et font davantage partie de foyers ne comportant qu’une personne. Dans le cadre des enquêtes par téléphone, la part des individus non accessibles par téléphone filaire augmente, posant la question de la pertinence des échantillons.

L’enquête 75000 de Médiamétrie, étude de référence du marché radiophonique national, est réalisée par téléphone filaire. La percée du téléphone mobile a conduit

\(^1\) Source Médiamétrie.
Médiamétrie à s’investir dans une veille adaptée en menant régulièrement des enquêtes par téléphone mobile destinées à évaluer la faisabilité d’enquêtes par téléphone mobile, à apprécier l’acceptation des personnes jointes et à mesurer une éventuelle différence de comportement, en particulier d’audience de la radio, entre la population touchée par la 75000 et celle, que l’on peut supposer atypique, des exclusifs du mobile.

2. Méthodologie d’une enquête par téléphone mobile

2.1. Une enquête par téléphone mobile ?

Si l’on est fondé à parler de substitution fonctionnelle entre le téléphone mobile et le téléphone filaire, les deux modes de communication sont-ils pour autant comparables en tous points ?

Le téléphone mobile est, naturellement, attaché à un individu plutôt qu’à un foyer ; ce peut être un avantage pour les enquêtes individuelles même si, souvent, celles-ci n’échappent pas à une description du foyer (quant à son chef, sa composition, son équipement…).

Lorsqu’on appelle un numéro de téléphone filaire, on sait où se trouve géographiquement la personne appelée : dans son logement, sur sa commune de résidence… Un numéro de téléphone filaire est intrinsèquement riche d’enseignements, jusque et y compris (potentiellement) l’adresse qui lui correspond et, par voie de conséquence, le type de quartier, d’habitat, etc. En l’absence d’annuaire, un numéro de téléphone mobile ne contient aucune information de ce type et ne peut donc pas s’accommoder d’un plan de sondage à registre géographique, sauf à retenir des quotas géographiques. Encore faut-il en préciser la nature (domicile, lieu d’enquête, …).

Une personne qui répond à un numéro de téléphone filaire est dans son environnement privé, disponible pour répondre au téléphone, éventuellement disponible pour répondre à une enquête et y consacrer le temps raisonnablement nécessaire. Une personne qui répond sur un numéro de téléphone mobile n’est pas systématiquement dans un environnement aussi favorable ; elle peut être professionnellement occupée, privativement indisponible, avoir une activité incompatible (comme par exemple être au volant de sa voiture ; attention à la responsabilité de l’appelant en cas d’accident). La prise de rendez-vous est possible (sur rappel de l’appelant à heure convenue) mais cela ne porte pas toujours ses fruits.

2 Ceci est cependant à nuancer car un annuaire peut ne pas enregistrer les déménagements

INSEE Méthodes 111
Il n'est pas rare qu'un téléphone mobile ne soit principalement utilisé que pour appeler ; il est le reste du temps sur messagerie et l'appelant peut alors laisser un message demandant à l'appelé de rappeler ; cette voie est a priori moins portée que la prise de rendez-vous.

Enfin, il reste une inconnue sur d'éventuels problèmes techniques. En effet, les réseaux de téléphonie mobile ne couvrent pas strictement l'ensemble du territoire et il se peut que certains appels n'aboutissent pas ou que les communications soient de mauvaise qualité, voire coupés.

2.2. Le problème de la base de sondage

Dans le cadre des enquêtes par téléphone filaire, l'annuaire (pages blanches) de France Télécom constitue une base de sondage présentant de nombreux avantages pratiques : il couvre l'ensemble du territoire et est régulièrement mis à jour.

Dans le cadre d'enquêtes par téléphone mobile, on ne dispose pas d'une telle base de sondage. Les seules informations qui nous sont communiquées sont les plages - préfixes de quatre chiffres - octroyées par l'ART3 aux différents opérateurs de téléphonie mobile. La constitution d'échantillons de numéros de téléphone mobile passe donc par un générateur de nombres aléatoires parmi les préfixes de type 06.XX qui ont effectivement été attribués aux opérateurs.

De plus, contrairement à France Télécom, les préfixes attribués par les opérateurs de téléphonie mobile n'ont aucune signification sur le plan géographique. Enfin, on ne dispose d'aucune information suffisamment précise pour être considérée comme donnée de cadrage sur la population des équipés de téléphone mobile.

Globalement donc : pas de base de sondage, aucune information géographique ni de structure, la seule méthodologie envisageable ici est donc le sondage aléatoire simple dans un ensemble de numéros dont tous ne sont pas nécessairement attribués par les opérateurs.

2.3. L'enquête par téléphone mobile du printemps 2000

La veille évoquée plus haut se décline en une série d'enquêtes de rythme annuel. Une opération pilote a été menée à l'été 1998, destinée à recueillir des premiers éléments sur la faisabilité des enquêtes par téléphone. Cette opération ayant été plutôt concluante, notamment sur le taux d'acceptation, une première enquête « grandeur nature » a été menée au printemps 2000. Cette enquête a ainsi été réalisée sur une période de quatre semaines, avec environ 60 interviews par jour du lundi au

3 Autorité de Régulation des Télécommunications
vendredi entre 17h30 et 21h30. Au total, 1215 interviews ont été réalisées, parmi lesquelles on s’est astreint à toucher 50% d’individus possédant également un téléphone filaire (les « mixtes ») et 50% d’exclusifs du mobile : on a finalement 607 mixtes et 608 exclusifs du mobile. Cette stratification permet de sur-représenter notablement la population des exclusifs du mobile\(^4\); c’est en effet celle qu’il est impossible d’atteindre dans l’enquête 75000 et donc celle qui nous intéresse particulièrement.

Étant donné le coût des appels sur téléphone mobile, on ne peut se permettre de poser le même questionnaire que celui de l’enquête 75000, qui, dans son ensemble, dure près d’une demi-heure. On se limite donc aux questions sur l’audience radio le jour même entre 5h du matin et 17h30. A ce bloc de questions s’ajoutent celles d’ordre socio-démographique, ainsi que celles relatives à l’utilisation du téléphone mobile et à la résiliation ou l’intention de résilier le téléphone filaire. Le questionnaire dure finalement 11 minutes en moyenne.

Pour des raisons de sécurité, a été insérée une question préliminaire permettant de déterminer si l’individu interrogé est au volant d’un véhicule. Si c’est le cas, l’enquêteur interrompt l’entretien en proposant un rendez-vous. Enfin, pour chaque numéro tiré, on tente d’établir le contact jusqu’à cinq fois, ce qui n’exclut pas que l’interview se fasse en plusieurs fois ; au-delà de ces cinq essais, on abandonne le numéro.

3. Y a-t-il complémentarité des modes d’enquête ?

Pour répondre à cette question, il convient de comparer l’enquête réalisée par téléphone mobile et l’enquête 75000 sur deux niveaux :

- le recueil de l’information,
- les résultats.

3.1. Le recueil de l’information

Pour réaliser 1215 interviews d’individus de 15 ans et plus (608 exclusifs du mobile et 607 mixtes), 13906 numéros ont été nécessaires.

\(^4\) Selon l’enquête 24000 Multimédia de Médiamétrie, au 2\textsuperscript{ème} trimestre 2000, 10,4% des individus de 15 ans ou plus équipés de téléphone mobile ne possédaient pas de téléphone filaire.
Tableau 2 : Résultat du dernier appel dans l’enquête par téléphone mobile (Avril 2000)

<table>
<thead>
<tr>
<th>Interviews réalisées</th>
<th>1215 (8,7%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refus de continuer</td>
<td>1810 (13,0%)</td>
</tr>
<tr>
<td>Sans réponse</td>
<td>4867 (35,0%)</td>
</tr>
<tr>
<td>dont</td>
<td></td>
</tr>
<tr>
<td>Messagerie</td>
<td>2867 (20,6%)</td>
</tr>
<tr>
<td>Occupé</td>
<td>1393 (10,0%)</td>
</tr>
<tr>
<td>Sonnerie « dans le vide »</td>
<td>607 (4,4%)</td>
</tr>
<tr>
<td>Hors quota</td>
<td>1457 (10,5%)</td>
</tr>
<tr>
<td>Hors étude</td>
<td>4557 (32,8%)</td>
</tr>
<tr>
<td>dont</td>
<td></td>
</tr>
<tr>
<td>Numéro non attribué (message opérateur)</td>
<td>4347 (31,3%)</td>
</tr>
<tr>
<td>Hors champ</td>
<td>210 (1,5%)</td>
</tr>
<tr>
<td>Total</td>
<td>13906 (100%)</td>
</tr>
</tbody>
</table>

Le quota de 50% d’exclusifs du mobile et 50% de mixtes mobile et filaire a été respecté quotidiennement. Chaque jour d’enquête, on avait donc un certain nombre d’individus qui ne correspondaient plus aux quotas. Pour la plupart, des rendez-vous ont été fixés à un autre jour : certains ont permis la réalisation d’une interview et 1457 - les hors quota -, qui par ailleurs avaient accepté de répondre, n’ont pas été rappelés.

Dans la catégorie « hors étude », sont inclus les numéros non attribués et les numéros hors champ. Les numéros hors champ correspondent notamment à des individus de moins de 15 ans et constituent 1,5% de l’ensemble des numéros. Pour 31,3% des 13609 numéros utilisés, un message de l’opérateur indiquait la non-attribution du numéro appelé. On peut ainsi chiffrer le coût induit par l’absence de base de sondage.

Certains numéros, qui apparaissaient occupés ou sonnant dans le vide au premier appel, se sont avérés non attribués lors des rappels. On ne peut donc savoir
exactement quelle est la proportion de numéros effectivement non attribués parmi les 13906 utilisés.

Le « rendement » des appels peut être évalué sur deux registres : un registre économique d’une part, et un registre technique d’autre part.

✓ Rendement économique

1215 interviews ont été réalisées pour 13906 numéros utilisés, soit un taux de réalisation de 8,7%. En moyenne sur l’ensemble des 13906 numéros utilisés, 4,7 appels ont été réalisés par numéro.

✓ Rendement technique

On constate 1810 refus de continuer et 1215 interviews réalisées. Sur la base des numéros utiles, c’est-à-dire en excluant les hors quota et hors étude, le taux de refus est ainsi de 20,4% et le taux de succès de 13,7%. Pour les numéros ayant abouti à une interview complète (1215 numéros), il a fallu entre 1 et 30 (r)appels pour mener à bien l’interview : le contact a bien été établi dans les cinq premiers appels, mais des prises de rendez-vous ou des coupures pendant l’entretien ont parfois entraîné l’obligation de rappeler et par conséquent la multiplication des appels. La répartition du nombre d’appels nécessaires pour mener à bien l’interview est la suivante :

Tableau 3 : Nombre d’appels nécessaires par interview complète

<table>
<thead>
<tr>
<th>Nombre d’appels</th>
<th>Répartition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 appel</td>
<td>27,1%</td>
</tr>
<tr>
<td>2 appels</td>
<td>25,3%</td>
</tr>
<tr>
<td>3 appels</td>
<td>14,9%</td>
</tr>
<tr>
<td>4 appels</td>
<td>9,5%</td>
</tr>
<tr>
<td>Entre 5 et 9 appels</td>
<td>17,4%</td>
</tr>
<tr>
<td>10 appels et plus</td>
<td>5,8%</td>
</tr>
</tbody>
</table>
Le nombre moyen d’appels, pour les numéros ayant abouti à une interview complète, est 3,4. Pour 70% de ces 1215 interviews, le questionnaire a été rempli lors d’un seul et même appel, qui n’était pas forcément le premier, et pour les 30% restant, il a été coupé et repris dans d’autres appels. Enfin, 43% de ces interviews ont été réalisées suite à une prise de rendez-vous.

Dans l’enquête 75000, la durée moyenne du questionnaire est de 26 minutes et on impose quotidiennement des quotas sur les critères croisés sexe par âge et sexe par activité. Même si les résultats des contacts dans l’enquête par téléphone mobile ne sont donc pas comparables avec ceux de l’enquête 75000+, il est intéressant de faire le parallèle entre les deux.
Tableau 4 : Résultat du dernier appel dans l’enquête 75000 (Novembre-Décembre 1999)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviews réalisées</td>
<td>15203 (12,2%)</td>
</tr>
<tr>
<td>Refus de continuer</td>
<td>45274 (36,3%)</td>
</tr>
<tr>
<td>Sans réponse</td>
<td>49923 (40,0%)</td>
</tr>
<tr>
<td>Hors quota</td>
<td>2183 (1,7%)</td>
</tr>
<tr>
<td>Hors étude(^5)</td>
<td>12213 (9,8%)</td>
</tr>
<tr>
<td>Total</td>
<td>124796 (100%)</td>
</tr>
</tbody>
</table>

Dans l’enquête 75000, réalisée par téléphone filaire, on n’a que 9,8% d’appels hors étude, et ce grâce à l’existence d’un annuaire téléphonique. On remarque que le taux d’appels hors quota (1,7%) est très faible. Un élément d’explication est que, contrairement au téléphone mobile, le filaire est un équipement du foyer : un appel permet a priori de choisir une personne parmi celles présentes dans le foyer, et de remplir ainsi plus facilement les quotas.

✓ Rendement économique

15203 interviews ont été réalisées pour 124796 numéros utilisés, soit un taux de réalisation de 12,2%.

✓ Rendement technique

Sur la base des numéros utiles, le taux de refus est évalué à 41% et le taux de succès à 13,8%. Ce taux de refus s’explique en grande partie par la durée du questionnaire.

On constate donc que, malgré un rendement économique plus faible, l’enquête réalisée par téléphone mobile a un rendement technique comparable, voire même meilleur, que l’enquête 75000. En effet, le taux de succès est sensiblement le même et le taux de refus plus faible. Certes, les deux enquêtes ne sont pas du tout

\(^5\) Sont compris dans le poste « Hors étude » les numéros professionnels, les fax, les numéros non attribués et les hors champ de l’enquête.
comparables, mais, au vu de l'expérience acquise, il semble que l'utilisation du téléphone mobile comme mode d'enquête n'ait pas reçu d'opposition particulière de la part des enquêtés. Il devrait être pertinent d'apporter un complément qualitatif à ce constat.

3.2. Les modes d'enquête

Dans près de 65% des cas, l'enquête se trouvait à son domicile au moment où l'interview a été finalisée, 7% étaient sur leur lieu de travail, 8,1% dans la rue et 7,2% chez des amis. Cette répartition est assez homogène entre les exclusifs du mobile et les mixtes mobile et filaire.

Afin de comparer les deux modes d'enquête, on analyse les résultats des questions communes, à savoir les questions relatives à l'audience de la radio, sur la sous-population des mixtes mobile et filaire, celle que l'on touche dans les deux enquêtes.

Pour cela, on compare les résultats d'audience cumulée\(^6\) entre 5h et 17h30, ventilés sur les principaux critères socio-démographiques. Bien que les résultats d'audience cumulée de l'enquête par téléphone mobile semblent plus faibles que ceux issus de la 75000, les écarts sont peu élevés. Afin de mettre en exergue les éventuels écarts, des tests de Student ont été effectués : en ce qui concerne le Total Radio, aucun écart n'est significatif au seuil de 95%. Pour ce qui est des stations, à peine plus de 13% des écarts sont significatifs, mais les écarts ne sont pas significatifs sur l'âge qui est le critère le plus discriminant dans l'explication du comportement d'audience de la radio.

Globalement, on peut donc conclure à la cohérence des résultats entre l'enquête réalisée par téléphone filaire et celle réalisée par téléphone mobile, pour la sous-population des mixtes mobile et filaire.

4. Comparaison entre exclusifs du mobile et mixtes

La comparabilité des résultats des deux enquêtes nous porte à penser qu'il serait envisageable d'utiliser conjointement les deux modes de recueil dans une même enquête. On pourrait ainsi toucher 98,6% des individus de 15 ans et plus\(^7\) et assurer la représentativité de l'échantillon.

Mais l'ajout d'un sur-échantillon d'exclusifs du mobile ne serait pas sans conséquence sur le coût de l'enquête et devrait, pour être financé, recueillir l'assentiment du Comité Radio de Médiamétrie. Au-delà de la volonté d'avoir un

\(^6\) L'audience cumulée d'une station S sur la tranche horaire T est estimée par la proportion d'individus ayant écouté la station S pendant la tranche T.

\(^7\) Selon l'enquête 24000 Multimédia de Médiamétrie, 1,4% des individus de 15 ans et plus ne sont pas équipés de téléphone (ni filaire, ni mobile).
échantillon aussi proche que possible de la population de référence, il convient de déterminer si le comportement des exclusifs du mobile diffère de celui des équipés filaire au regard des variables d’intérêt, i.e. s’il justifie l’enquête sur les exclusifs du mobile.

4.1. Les exclusifs du mobile ont-ils un comportement radio spécifique ?

Il s’agit ici d’évaluer la relation entre la non-possession d’un téléphone filaire et le comportement d’audience de la radio. Pour cela, on compare les résultats d’audience entre les deux sous-échantillons (exclusifs du mobile et mixtes mobile et filaire) de l’enquête par téléphone mobile avec ventilation par critères socio-démographiques.

Pour ce qui est du Total Radio, près de 45% des écarts sont significatifs au seuil de 95%, avec une moindre consommation du média Radio chez les exclusifs du mobile.

Tableau 5 : Principaux écarts significatifs entre exclusifs du mobile et mixtes mobile et filaire

<table>
<thead>
<tr>
<th>Station</th>
<th>Groupe</th>
<th>Audience cumulée de 5h à 17h30 (du 27/03/00 au 21/04/00)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Exclusifs du mobile</td>
</tr>
<tr>
<td>Total Radio</td>
<td>15-24 ans</td>
<td>70,0</td>
</tr>
<tr>
<td></td>
<td>25-34 ans</td>
<td>68,9</td>
</tr>
<tr>
<td></td>
<td>35-49 ans</td>
<td>5,6</td>
</tr>
</tbody>
</table>

Sur le Total Radio tous critères confondus, on observe un écart moyen de près de 10% sur l’audience cumulée de 5h à 17h30. Les exclusifs du mobile écoutent donc moins la radio que les équipés filaire, cet écart étant localisé notamment sur les stations généralistes.

4.2. Quel impact au global ?

Même si les exclusifs du mobile écoutent moins la radio que les mixtes mobile et filaire, leur poids dans l’univers des 15 ans et plus n’est que de 5,2%. Seul un impact réel de leur insertion sur les résultats d’audience justifierait un sur-échantillon coûteux d’exclusifs du mobile.
Pour évaluer cet impact, nous avons fusionné les données provenant de l’enquête 75000 et les exclusifs du mobile issus de l’enquête par téléphone mobile, pour créer ainsi un échantillon de simulation. Chaque individu a un poids de sondage. Afin d’évaluer l’impact réel des exclusifs du mobile sur les résultats d’audience, il convient d’effectuer un calage permettant de remettre les exclusifs du mobile et les équipés filaire à leurs poids réels respectifs.

Cette simulation montre que l’impact global, sur l’audience cumulée comme sur la durée d’écoute par individu, est minime : les écarts relatifs, en valeur absolue, n’excèdent pas 4% pour l’audience cumulée et atteignent 5,4% pour la durée d’écoute par individu. Sur la cible privilégiée des 15-24 ans, les écarts relatifs sont un peu plus important. Enfin, on analyse l’audience quart d’heure par quart d’heure entre 5h et 17h30 : la quasi superposition des courbes (cf. graphe 2) conduit à conclure que la prise en compte des exclusifs du mobile ne modifie pas, aujourd’hui, sensiblement les choses.

Graphe 2 : Audience quart d’heure par quart d’heure entre 5h et 17h30

4.3. Recherche du seuil critique

Dans le même esprit que précédemment, toujours en réalisant des simulations, on se propose de déterminer le taux d’exclusifs du mobile à partir duquel les écarts en terme d’audience cumulée entre 5h et 17h30 deviendraient significatifs.
Notre démarche consiste en fait à modifier les marges du critère de calage fixées, dans la simulation du paragraphe précédent, à 6% d’exclusifs du mobile et 94% d’équipés filaire. À chaque étape, on incrèmente de 1% la part des exclusifs du mobile et donc leur poids dans l’échantillon recomposé, et on mesure les écarts d’audience cumulée.

Ce n’est qu’à partir de 19% d’exclusifs du mobile, sur l’ensemble des 15 ans et plus équipés de téléphone, que l’écart devient significatif à 95% pour le Total Radio, c’est-à-dire que les intervalles de confiance des deux résultats d’audience cumulée sont disjoints. À 25% d’exclusifs du mobile, l’écart devient significatif pour quelques stations.

5. Conclusion

Ce système de veille a permis de mettre en évidence que, aujourd’hui, les exclusifs du mobile, encore en faible proportion, ne remettent pas en cause les résultats d’audience de la radio issus de l’enquête 75000, bien que leur comportement d’audience soit différent de celui des équipés filaire. Ce constat est valable pour l’audience de la radio, mais n’est pas généralisable : il dépend des variables d’intérêt.

Le développement d’internet pourrait également porter atteinte à la représentativité des enquêtes par téléphone filaire. En effet, lorsqu’il est connecté à Internet par ligne de téléphone filaire, le foyer n’est pas joignable sur cette ligne. De plus, certains foyers consacrent une ligne téléphonique exclusivement à Internet, ce qui introduit de nouveaux biais dans la base de sondage, à savoir les pages blanches de France Télécom.
Bibliographie


VERS UNE MESURE FIABLE DE L'AMPLEUR DES USAGES DE DROGUES ILLICITES EN FRANCE

F. BECK

Observatoire Français des Drogues et Toxicomanies

Dans son ouvrage *Pourquoi les hommes usent-ils de stupéfiants?* Léon Tolstoï offrait en 1890 une vision quelque peu catastrophiste du monde des drogues : « Continuellement et partout on rencontre des gens qui pourtant aiment leurs enfants et sont prêts à faire pour eux de multiples sacrifices, mais qui néanmoins dépensent en eau de vie, vin, bière, haschisch ou opium, et surtout en tabac, l'argent qui pourrait nourrir ces enfants et même les sauver de la misère. » L'usage des adverbes, sensés interpeller le lecteur, est également révélateur d'un certain flou dans l'estimation que Léon Tolstoï fait de l'ampleur du phénomène.

Sans remonter si loin, il suffit de demander à des étudiants de première année de sociologie combien de jeunes parmi les 15-19 ans ont déjà pris du cannabis, pour qu'ils répondent volontiers 80 ou 90 %, puisque dans leur perception, presque tous les jeunes fument du cannabis au moins occasionnellement. Or les chiffres des enquêtes déclaratives donnent plutôt 40 %. Enfin, les médias, principaux vecteurs de l'information sur l'ampleur des usages de drogues, n'assurent pas toujours cette transmission sans égarements. Ainsi, l'annonce en juillet 2000, dans la presse espagnole puis française, de chiffres ahurissants concernant la consommation de

---


2 De Quincey, lui-même usager d'opium, s'était livré en 1822 à un exercice similaire dans ses *Confessions d'un mangeur d'opium anglais*. « [Les mangeurs d'opium] forment en vérité une catégorie fort nombreuse. J'en ai acquis la conviction, voici quelques années, en faisant le compté de ceux qui, dans une petite fraction restreinte de la société anglaise (celle des hommes que distinguent leurs talents ou qui jouissent d'un rang éminent) m'étaient connus directement ou indirectement. [...] Or, étant donné qu'une seule fraction, relativement si limitée, offrait pareils cas par dizaines (et cela au su d'un seul observateur), on en pouvait inférer naturellement que l'ensemble de la population anglaise en fournirait proportionnellement autant. » Pour lui, il était inéluctable, sur les seules bases d'observations ethnocentrees, que les opiomanes proliféraient dans l'Angleterre du XIXème siècle.

INSEE Méthodes
cocaïne en Espagne attire également l'attention : plus de 40 % des jeunes en auraient déjà pris. Il faut savoir que les chiffres courants oscillent entre 1 % et 5 %. Il s'agissait en fait d'une étude ethnographique sur les jeunes de milieux urbains adeptes de la vie nocturne et recrutés dans des clubs ou des after. Manipuler du chiffre sur les usages de drogues prête donc parfois à confusion et l'enjeu va ici être de savoir dans quelle mesure les techniques classiques d'enquête sont valides pour décrire ce phénomène. Par exemple, l'enquête décennale de santé de 1991 (INSEE, SESI, CREDES), menée en population générale, porte entre autres sur les consommations d'alcool, de tabac et de médicaments psychotropes mais pas sur les drogues illicites. Ne faut-il pas, dès lors que l'observation porte sur un objet rare et potentiellement caché, s'imposer d'aller à la rencontre de populations beaucoup plus difficiles à atteindre, qu'elles soient marginalisées ou protégées par leur statut social ? Ces individus échappant aux outils standard d'observation peuvent être les sans domicile, certains usagers de cocaïne bien insérés socialement, les non francophones... L'enquête en population générale doit-elle s'imposer d'inclure ces populations difficiles à atteindre mais parfois particulièrement concernées par la consommation de psychotropes ?

Les objectifs d'une telle mesure sont multiples : il va s'agir à la fois de connaître les prévalences (terme emprunté au langage médical désignant le nombre de cas d'une maladie sur l'ensemble de la population pour une période donnée) des usages de substances licites et illicites au sein de la population, mais aussi leurs évolutions dans le temps et de pouvoir effectuer des comparaisons internationales. Dans le même ordre d'idée, des informations sur l'incidence et le début des usages sont d'utiles indicateurs pour décrire le phénomène. D'autres recherches, telles que celles axées sur la mise en évidence des facteurs associés aux consommations de substances psychoactives et la mesure de leur lien avec les usages, apparaissent également essentielles à quantifier.

Mais il convient de définir avec précision l'objet de l'étude. Les drogues réunissent une grande variété de produits dont la liste évolue dans le temps. La diversité des substances en jeu dans ce domaine alourdit les protocoles envisagés pour décrire la situation. La pureté (au sens chimique) variable des produits, les dénominations multiples, les effets dépendant des dosages, les divers modes d'administration et les « usages durs de drogues douces » (Aquatias et al., 1997) rendent d'autant plus incertain ce cadre. A titre d'exemples, l'acide lysergique (LSD), souvent associé au support qu'est le buvard, peut aussi être consommé sous forme de micropointes, les amphétamines peuvent s'injécter et l'héroïne se fumer.

Le champ d'observation des drogues en France a été clairement étendu aux substances licites à la fin des années 90 (Roques, 1998). Il apparaît désormais inconcevable d'envisager les consommations de drogues en s'en tenant aux

3 Pour plus de détails, voir la revue Pénombre n°24, janvier 2001.
4 Nombre de nouveaux cas sur une période donnée, rapporté à la population totale.
substances figurant sur la liste des stupéfiants telle qu'elle est présentée dans le Code de la Santé Publique. La liste des substances peut dès lors se résumer ainsi :

- Les produits licites ou dont l'usage est réglementé (alcool, tabac, médicaments psychotropes…)
- Les produits illicites (cannabis, cocaïne, héroïne, LSD, champignons hallucinogènes…)
- Les produits licites détournés de leur usage (produits à inhaler, les médicaments pris hors traitement…)
- Les nouvelles drogues (kétamine, GHB, DMT…)\(^5\)
- Les produits inscrits sur la liste des substances et méthodes dopantes (Dorosz, 2001)

Malgré un certain « trouble des frontières » ( Ehrenberg, 1998), des limites peuvent ainsi être tracées en termes de produits. Mais l’observateur est vite confronté à un continuum d’usages qu’il est parfois difficile de réduire à quelques agrégats nécessaires à l’analyse statistique : quel type d’usager compter et dans quelles conditions de validité ? Entre la consommation de drogues au sens large et leur usage problématique, la limite est parfois située au niveau de la dépendance, mais une telle distinction reste discutable. Une sorte d’échelle des comportements peut toutefois être esquissée à partir de l’intensité de l’usage et de ses conséquences, allant de l’expérimentation à la pathologie profonde en passant par l’usage occasionnel, récréatif, répété ou régulier, contrôlé ou pas, l’excès, le désir compulsif, l’abus, la dépendance et la marginalisation, la polyconsommation achevant de complexifier le tableau. Elargissant encore le champ, des équipes de recherche allemande, néerlandaise et maltaise ont mené une observation conjointe des pratiques maniaques du jeu et des formes plus usuelles d’appétence\(^6\) dans leurs enquêtes statistiques. Cette concomitance, présente dans des ouvrages de référence sur les addictions\(^7\) (Lowinson \textit{et al.}, 1997), a d’autant plus d’intérêt que, comme le soulignent Valleur et Bucher (1997), les conduites addictives donnent fréquemment lieu à un regroupement : de nombreux joueurs invétérés sont également des usagers compulsifs de drogues.


\(^5\) La kétamine et le GHB (acide gamma hydroxybutyrique) sont des anesthésiants vétérinaire ou à usage hospitalier qui, en fonction du dosage, peuvent avoir de forts effets hallucinogènes. Le DMT (Diméthyltryptamine) est un hallucinogène à la durée d’action courte mais très puissant.

\(^6\) Tendance qui pousse un individu à assouvir ses besoins, ses instincts.

\(^7\) Le jeu pathologique s’y retrouve aux côtés des troubles de l’appétit, de l’adhésion aux sectes et de la dépendance au sexe.
clairement sur la nécessité d’un tel outil en France, à l’instar des États-Unis ou de la plupart des pays européens.

1. **Le choix fondamental de la base de sondage et du mode d’échantillonnage**

La France ne dispose pas d’un registre national de la population, à l’inverse d’autres pays européens (Suède, Finlande, Pays-Bas, Allemagne…). Seuls le recensement de la population et le fichier des abonnés du téléphone peuvent être utilisés comme base de sondage nationale, mais les enquêteurs de l’INSEE sont seuls autorisés à se servir de la base issue du recensement. Il existe aussi d’autres bases telles que les fichiers EDF, les listes électorales, les listes des contributions directes… mais elles sont toutes insuffisantes pour représenter correctement la population générale. Le recours à une base aréolaire, c’est à dire définie sur des critères géographiques, permettrait une investigation plus précise des zones géographiques typiquement concernées par le phénomène de consommation de substances psychoactives (cités, quartiers sensibles, quartiers “branchés”…), mais cette solution impose la mise en œuvre d’un coûteux dispositif d’encadrement des enquêteurs. L’enquête mise en place au Portugal en 2001 repose néanmoins sur une telle base (Balsa, 2001).⁸

Le choix du mode d’échantillonnage appliqué à la base de sondage est également primordial. Entre enquêtes par quotas et sondages probabilistes de toutes sortes s’opposent différentes écoles. La méthode des quotas, souvent utilisée par les instituts de sondage qui en maîtrisent bien l’application, constitue une méthode empirique peu onéreuse, rapide et relativement bien adaptée aux petits échantillons. Dans la pratique, la variance des estimateurs fournis par ce type de sondage est estimée “plutôt faible” tant que la taille de l’échantillon reste petite (moins de 2000 individus). L’intervalle de confiance obtenu par estimation sur la méthode des quotas est même majoré par l’intervalle équivalent dans le cas d’un sondage stratifié à allocation proportionnelle, calculé en assimilant à une strate chaque croisement de critère sur lequel il faut respecter un quota (Deville, 1992).

En revanche, cette méthode ne permet pas d’effectuer de tests statistiques en toute rigueur. Une majorité de chercheurs et de méthodologues préconisent l’utilisation des sondages aléatoires, en particulier des sondages stratifiés. La méthode des quotas présente deux autres désavantages : celui de laisser échapper les individus les plus difficiles à joindre (il faut parfois plus de dix tentatives avant de pouvoir joindre un ménage, ce qui autorise une enquête aléatoire, alors qu’avec la méthode des quotas,

---

⁸ Après avoir stratifié sur les sept grandes régions portugaises, une sélection aléatoire de municipalités sera opérée. Les chercheurs de l’université de Lisbonne effectuent ensuite un tirage aléatoire proportionnel de 1 500 sections à l’intérieur desquelles sont sélectionnés des foyers. A l’intérieur du foyer, le choix de l’individu à interviewer est également aléatoire, pour obtenir in fine environ 15 000 enquêtés.
un ménage injoignable au premier essai est abandonné au profit du suivant sur la liste) et celui d’empêcher tout contrôle de la non-réponse.

Un autre défaut des sondages par quotas est le biais de sélection induit par la liberté qui est laissée à l’enquêteur de sélectionner les individus qui vont faire partie de l’échantillon, surtout dans les enquêtes en face à face. À titre d’exemple, lors des tests préalables au lancement de l’enquête sur l’Analyse du Comportement Sexual des Français (ACSF), les chercheurs avaient insisté sur l’importance d’arriver à rentrer en contact avec des individus “ à risque ”. Les enquêteurs avaient pris cette consigne tellement à cœur que l’échantillon de ce test s’est retrouvé constitué d’une part exceptionnellement élevée d’homosexuels, catégorie que l’équipe de recherche craignait au départ de voir sous-représentée. Un biais du même ordre est peut-être apparu dans les enquêtes réalisées par l’institut IFOP pour le Comité Français d’Education pour la Santé (CFES) en 1990, 91 et 92, en face à face et sur la base d’un échantillon constitué par quotas. Ces enquêtes, dont le but était de mesurer l’impact d’une campagne télévisée de prévention des drogues, comportaient également des questions de consommation pour le cannabis et quelques autres produits illicites. Or, alors que tous les spécialistes s’accordent à constater une banalisation et une augmentation de la consommation et de l’expérimentation de cannabis en France au cours des années 90, ces enquêtes fournissent pour les [18-44 ans] des chiffres nettement plus élevés que ceux des enquêtes menées les années suivantes : environ 32 % en 1990 et 27 % en 1991 et 1992, alors que les sondages plus récents et réalisés suivant des méthodes diverses montrent une croissance de 18 % en 1992 (Beck, 1998) à 31 % en 1999 (Beck et Peretti-Watel, 2000). S’il n’est pas possible de mesurer à quel point le contexte, le mode de collecte ou d’autres facteurs méthodologiques ont pu influencer les déclarations, la cause la plus probable reste le biais d’échantillonnage lié à la méthode des quotas qui a conduit les enquêteurs à contacter une population "anormalement concernée" en comparaison de celle qu’un échantillonnage aléatoire aurait pu fournir.

La nécessaire étape de formation des enquêteurs revêt donc une importance cruciale dans le cadre des enquêtes réalisées selon la méthode aléatoire, et de tels exemples sont utiles pour leur montrer que la bonne volonté est moins payante que le respect le plus strict du hasard. Ils peuvent ainsi être incités à enquêter à des heures différentes de la journée et à renouveler de nombreuses fois la tentative de contact en cas d’échec ou, dans le cadre du face à face, à respecter l’ensemble du territoire

9 Parmi lesquelles on trouve le téléphone en aléatoire ou par quotas et le face à face par quotas.
10 Précisons que l’objectif de la plupart de ces enquêtes, en particulier les post-tests réalisés par l’IFOP en 1990, 91 et 92, n’était pas d’offrir une mesure du niveau d’usage de cannabis, mais que cette question était plutôt envisagée comme une variable transversale.
qui leur est attribué au départ même s’il est moins coûteux de rester dans un périmètre limité et parfois plus tentant d’éviter les quartiers « sensibles ».

2. Améliorer la qualité de la passation

2.1. Optimiser le taux de réponse à l’enquête

Le statut obligatoire des enquêtes figure parmi les méthodes les plus efficaces pour diminuer les refus de réponse. Son incidence sur le nombre et la qualité des réponses a déjà été montrée (Berthier et Dupont, 1996), mais, pour d’évidentes raisons éthiques, il n’est pas question d’y recourir sur un sujet sensible tel que celui de l’usage des drogues. Le recours à une lettre-avis est en revanche courant et a déjà fait la preuve de son efficacité (Spira et Bajos, 1993 ; INSEE, 1996).

Puisqu’il est difficile de l’éviter, il apparaît primordial de connaître les raisons du refus (est-il lié au thème, au fait que l’enquêté refuse d’ouvrir sa porte à un inconnu ou de passer du temps au téléphone... ?) car il est vraisemblable que cette population ait des caractéristiques particulières. Il est ainsi envisageable de proposer à l’individu refusant de répondre à un entretien un questionnaire auto-administré qu’il pourrait renvoyer par la poste, en lui donnant également un numéro vert auquel il pourra se renseigner s’il a des problèmes pour remplir le questionnaire et se faire confirmer le contexte d’anonymat et de confidentialité. De récents travaux méthodologiques menés aux Pays-Bas par le Centrum voor Drugsonderzoek (CEDRO) de l’Université d’Amsterdam dans le cadre des enquêtes sur les usages de drogues ont par ailleurs montré que la rémunération des enquêtés (de l’ordre de 10 Euros) diminuait sensiblement les refus, cette diminution étant d’autant plus nette que la somme augmentait, même modestement.

Aborder un sujet sensible sans aucun contexte peut occasionner un nombre considérable de refus, liés tant à la gêne qu’à l’incompréhension. En effet, il apparaît nécessaire d’intégrer la consommation de drogues dans une problématique plus générale et donc moins stigmatisante (pouvant être la santé, des sujets de consommation plus généraux, les modes de vie ou les conduites déviantes). Plusieurs pays (Grèce, Suisse, Danemark, Autriche, Belgique, Québec...) posent des questions sur les usages de substances psychoactives dans les enquêtes de santé. En France, en 1992, 1993 et 1995 auprès des adultes et en 1997 auprès des jeunes, les baromètres santé du CFES ont abordé le thème des drogues illicites, même si celui-ci était relativement sommaire, dans le cadre d’un questionnement multithématique (Baudier et al., 1994 ; Baudier et al., 1995 ; Baudier et Arènes, 1997 ; Baudier, Janvin et Arènes, 1998). Le fait de se situer dans une telle perspective est un bon vecteur pour une enquête car c’est un sujet de préoccupation pour de nombreux individus (fréquemment classé second derrière le chômage dans les sondages d’opinion). Néanmoins, ce choix génère implicitement un taux de refus lié au contexte qui va parfois être jugé trop intime, sensible ou même inintéressant pour des gens qui s’estiment en bonne santé.
2.2. Assurer la qualité et la fiabilité des réponses

Le vocabulaire employé peut être très différent d'une population à l'autre et le grand nombre de substances en jeu achève de compliquer les choses. L'emploi de l'argot ("keps", "tarpe", "beuh"...) peut, dans certains cas, s'avérer inefficace mais la nature même de ce langage fait qu'il n'est pas accessible à tous. Dans le cas de la cocaïne et du crack, le fréquent amalgame fait entre les deux produits, notamment dans le contexte de la rue, incitera à l'emploi d'autres dénominations ("caillou", "galette", "freebase", "rocher") pour faire la distinction. Dans le même ordre d'idées, le terme "speed" peut désigner soit l'ensemble des stimulants synthétiques, soit simplement les amphétamines. Il faudrait théoriquement s'assurer que les enquêtés donnent tous le même sens aux mots clés, mais dans ce cas, l'essentiel n'est pas forcément que l'ensemble de l'échantillon comprenne une question, mais plutôt que ceux qui sont directement concernés la saisissent parfaitement. Une enquête sur les représentations des Français en matière de drogues (EROPP), menée par l'OFDT en 1999, montre ainsi que certains consommateurs de cannabis répondent "non" à la question "Avez-vous déjà pris du cannabis", mais par l'affirmative à la question "Quelles sont les autres drogues que vous avez déjà prises?", en précisant "un beuze", "le bedo" ou "du zetla" (Beck et Peretti-Watel, 2000). Dès lors le concepteur se retrouve-t-il contraint de construire un questionnaire étoffé sur un sujet précis qui, en terme d'usage, ne va concerner qu'une petite minorité de la population interrogée, du moins pour les substances autres que l'alcool, le cannabis, le tabac et les médicaments psychotropes.

Parmi les solutions envisageables, le recours à des filtres s'avère très efficace pour les comportements rares, de même que la conception d'un lexique pour les enquêteurs qui, notamment en face à face, leur permet de proposer des alternatives aux interviewés dubitatifs. Les rapports d'activité des associations de soutien aux usagers, les relevés d'entretien des ethnologues, voire les rapports de police, s'avèrent de précieux outils pour repérer certaines dénominations spécifiques d'une substance donnée.

La clarté de présentation du questionnaire, de l'enchâinement des questions, la simplicité du vocabulaire employé et la lisibilité des filtres s'avèrent des critères essentiels. Il convient de contextualiser les questions afin de limiter l'équivocité et les difficultés de mémorisation telles que, par exemple, l'effet de "télescopage des dates", qui amène le répondant à inclure un événement dans la période récente alors que celui-ci en fait plus ancien (Auriat, 1996). Pour éviter que l'ordre des modalités proposées n'influence les réponses, les systèmes "Computer Assisted Personnail / Telephone Interview" (CAPI et CATI) permettent des rotations aléatoires des modalités de réponse. Dans le même ordre d'idée, il est illusoire de proposer un trop grand nombre de modalités (et elles ne doivent être trop complexes) lors d'un entretien téléphonique, sous peine de focaliser l'attention du répondant sur quelques unes d'entre elles (la première qui lui semble correspondre à sa réponse ou les dernières entendues). À ce titre, l'étude des débriefings avec les enquêteurs et des rapports techniques des instituts de sondage est instructive. De
nombreuses informations sur les écueil à éviter y figurent : lors de l’enquête EROPP 1999, les questions qui ont posé le plus de problèmes ne concernaient ni les opinions sur les drogues et les toxicomanies, ni les consommations de produits psychoactifs mais les questions portant sur les usages des proches et sur le positionnement politique. Enfin, l’effet-enquêteur doit pouvoir être mesuré, notamment en fonction des caractéristiques socio-démographiques de celui-ci (Berthier, Deville et Néros, 1996).

2.3. Créer un contexte de confiance pour limiter la dissimulation

Les faibles prévalences d’usage de cannabis au cours de la vie relevées dans les enquêtes en population générale pour les tranches d’âge les plus élevées sèment le doute sur la validité des chiffres, même si une étude allemande a pu montrer la robustesse de telles valeurs (Kraus, Bauernfeind et Herbst, 1998). Par ailleurs, concernant l’alcool, il a été estimé que les enquêtes françaises de consommation correspondaient à seulement 60% de l’alcool effectivement distribué dans l’hexagone (Got et Weill, 1997). Une telle approximation n’est évidemment pas envisageable sur un produit illicite, mais il est très probable qu’une sous-déclaration existe également, du fait même de son caractère délictueux. Les tests chimiques (urine, cheveux, sueur), contraignant techniquement (les traces d’héroïne et de cocaïne ne restent que quelques jours dans les urines) et très coûteux, ne sont pas réalisables en population générale. Ils ont néanmoins déjà été utilisés et placés en regard des résultats des déclarations par questionnaire aux Etats-Unis (Hancock et al., 1991) et même en France, dans une enquête du Service de Santé des Armées de 1996, après accord des sujets, permettant ainsi de repérer à la fois la sur et la sous-déclaration (Louboutin-Croc et al., 1997). Les résultats étaient édifiants, les deux méthodes aboutissant à des prévalences du cannabis au cours de la vie de l’ordre de 40% pour les jeunes hommes mais les déclarations et les tests ne concordaient pas : 20% d’individus positifs au test mais ne déclarant pas d’usage dans le questionnaire, 20% dans la situation symétrique et environ 20% positifs et déclarant un usage dans le questionnaire... Il faut préciser que le contexte, les « trois jours », avec l’enjeu de telle ou telle affectation ou d’une réforme, pouvait biaiser considérablement les réponses12.

Pour inciter les répondants à donner des informations fiables, il semble incontournable de bien expliquer le contexte de l’enquête pour les motiver et les mettre en confiance sur la confidentialité en formant les enquêteurs à être persuasifs et en adoptant des techniques respectant explicitement la confidentialité aux yeux des enquêtés (sceller le questionnaire sous enveloppe et aller avec lui poster la lettre dans une boîte, laisser l’enquête répondre seul sur un ordinateur portable...).

12 Dans l’esprit des appelés, déclarer un usage limitait les possibilités d’affectation mais pouvait permettre d’échapper au Service national.
Il est aussi envisageable de placer quelques questions de contrôle discrètes permettant de juger partiellement la sous-estimation. À titre d'exemple, lors de l'enquête ACSF, au sein d'un questionnaire très important (plus de 400 questions) se trouvaient deux questions sur la masturbation féminine posées à toutes les interrogées alors qu'un filtre aurait dû "épargner" la seconde question à celles ayant répondu par la négative à la première. Ce test futurist a permis de constater une très nette disparité pour les femmes qui n'étaient que 34 % à reconnaître cette pratique alors que 51 % d'entre elles prétendaient parvenir toujours, plutôt facilement ou plutôt difficilement à l'orgasme par la masturbation (Béjin, 1993).

Il peut ainsi être imaginé de mettre en place un contrôle de la validité du témoignage par des questions similaires sur la consommation de cannabis par exemple. Dans les premières questions, serait demandé : "Avez-vous déjà fumé du cannabis ?" et, nettement plus loin dans le questionnaire, parmi des questions relatives à l'ivresse, "Lorsque vous fumez un joint, ressentez-vous une modification de votre état de conscience ?". Ce type de dispositif pourrait aussi être très efficace concernant les consommations de médicaments détournés de leur usage, tant pour la dissimulation volontaire que pour l'omission, à condition toutefois d'en user avec parcimonie pour éviter les interruptions de questionnaire. Il s'agirait donc ensuite de croiser ces différentes questions afin de déceler les incohérences au niveau individuel. Ces incohérences interrogent autant la qualité de certaines réponses fournies dans ce type d'enquête que les normes sociales qui en régulent l'expression. En effet, ce phénomène peut également être le reflet des diverses contraintes normatives qui s'exercent autour de la consommation de substances psychoactives.

La sous-estimation est très redoutée lorsque l'interrogation concerne un comportement répréhensible, en particulier lorsqu'il est illicite. Elle ne doit toutefois pas masquer le risque de surestimation (volonté d'impressionner l'enquêteur par des pratiques exceptionnelles, défi, crainte de passer pour un "anormal" lorsque les pratiques sont aussi courantes que l'expérimentation d'alcool, de tabac ou de cannabis...). L'utilisation de leurs (dummys drugs), tel que le désormais célèbre "mop" peut être envisagée pour détecter les individus qui prétendent connaître, phénomène assez courant (Bless et al., 1997 ; Beck et Peretti-Watel, 2000), ou même consommer (nettement plus rare) ces produits imaginaires (Ramsay et Spiller, 1997 ; Beck, Legleye et Peretti-Watel, 2001). Il faut néanmoins utiliser ces techniques avec prudence, certaines personnes pouvant avoir pris des drogues dont elles ne connaissent pas le nom, ou croire reconnaître le nom d'une substance qu'elles ont effectivement consommée.

L'impossibilité de pouvoir s'isoler, notamment dans les enquêtes par téléphone, pouvant entraîner certaines dissimulations (lorsqu'un adolescent se trouve en compagnie d'un de ses parents par exemple), il est important d'une part de fournir des modalités de réponse n'indiquant en rien le contenu de la question posée (oui, pas d'accord...) et d'autre part d'être en mesure de proposer de rappeler à un meilleur moment (Janvrin, Gautier et Arènes, 1997). Dans le même esprit, il peut être utile, dans le cas d'un entretien en face à face, de présenter à l'interviewé un carton sur lequel sont écrits les termes délicats pour éviter d'avoir à les prononcer.

*INSEE Méthodes*
(Lagrange, Lhomond et al., 1997). Dans le cas particulier de l’interrogation d’un individu sur son entourage (enquête “proxy”), envisageable sur un sujet sensible, s’observe parfois une tendance, relevée au cours de l’enquête ACSF, à se prétendre entouré d’individus « dans la norme » (Richard-Zappella, 1996).

Ces facteurs de perturbations peuvent être explorés grâce au recueil de l’avis de l’enquêteur à propos de la sincérité du répondant. Même si l’enquêteur peut avoir envie de déclarer que les réponses sont sincères comme gage de la qualité de son travail, cet indicateur permet néanmoins de prendre un recul profitable par rapport aux réponses fournies. Il ne faut cependant pas nourrir trop d’illusions quant à la diminution de la sous-déclaration et à la bonne volonté de certains des répondants parmi les plus concernés, la consommation de substances psychoactives s’inscrivant couramment dans un processus de rejet et/ou de dépassement du système. “Nous sommes des empêcheurs de tourner en rond puisque nous tournons en spirale” écrit Borel (1995) à propos des usagers d’ecstasy dans les fêtes techno, illustrant joliment ce caractère insaisissable.

Sur ce point comme sur l’ensemble de ceux concernant la qualité de la passation, la présence des équipes de recherche sur les terrains (pour la formation des enquêteurs dans tous les types d’enquêtes et sur le plateau lors des enquêtes téléphoniques) et le recours à des modes de collecte qui garantissent sans équivoque la confidentialité sont des gages de réussite primordiaux.

3. Description du dispositif mis en place en France par l’OFDT

3.1. Quelles enquêtes ?

Devant l’entrelacs des contraintes et la diversité des populations à étudier, notamment en termes d’âge, l’OFDT a mis en place un système permettant une couverture globale du champ tout en pointant sur les populations les plus concernées par les phénomènes d’usage. Il permet d’observer précisément les comportements aux âges d’entrée dans la consommation.

Le choix du mode de collecte s’est avéré primordial. Il a été guidé par des travaux antérieurs menés en France sur l’analyse des comportements sexuels. Le test réalisé en 1991 pour déterminer le mode le plus approprié à ce sujet sensible, entre le téléphone et le face à face avec une partie du questionnaire en auto-administré, concluait que chaque méthode avait ses avantages et que les taux de refus étaient similaires dans les deux cas (ACSF, 1992). Les chercheurs avaient noté une plus grande facilité à répondre, une meilleure cohérence et un contrôle de terrain plus efficace pour le téléphone, mais aussi une tendance à donner des réponses plus proches de la norme qu’en face à face. En conséquence, le surcoût financier occasionné par une enquête en face à face n’était pas apparu nécessaire. Une des questions du test portait directement sur la consommation de drogues au cours de la
vie (tous types de drogues illicites confondus), et les résultats n'étaient pas significativement différents d'un mode de collecte à l'autre. Plus récemment, une analyse secondaire a montré que les adultes répondaient plus facilement par téléphone qu'en face à face à cette question (Guilbert et al., 1999).

A l'étranger, les efforts méthodologiques consentis pour assurer une bonne qualité d'enquête sur les sujets sensibles en général et sur la toxicomanie en particulier sont parfois remarquables. Le cas du British Crime Survey dont le protocole a été modifié en 1994 pour passer à l'utilisation (coûteuse) d'ordinateurs portables en auto-administré (CASI13) en est un exemple notable.

Des tests menés par le National Institute on Drug Abuse (NIDA) aux États-Unis en 1991 ont montré que le recueil en face à face avec auto-administré est préférable à celui effectué par téléphone, mais ces résultats tiennent surtout à une couverture téléphonique parcellaire et excluant trop de ménages supposés être concernés par la consommation de drogues (Turner, Lessler, Devore, 1992). Dans une étude auprès d'usagers d'héroïne ayant recours aux soins, Bale (1979) montre que la voie postale offre des résultats d'une qualité supérieure ou égale à celle de l'entretien. Dans l'ensemble, les travaux méthodologiques américains sur les enquêtes en population générale se prononcent plutôt en faveur des questionnaires auto-administrés (Aquilino, 1994 ; Rodgers et al., 1999).

Sous l'égide de l'Observatoire Européen des Drogues et des Toxicomanies (OEDT), un projet méthodologique coordonné par l'Université d'Amsterdam a consisté en la répétition du même questionnaire avec, dans chacun des quatre pays engagés (Pays-Bas, Suède, Royaume Uni et Grèce), trois échantillons de 800 personnes par mode de collecte, afin de déterminer lequel se révélait le mieux adapté. Le questionnaire avait été construit par un groupe de travail européen réfléchissant à l'amélioration de la comparabilité internationale des enquêtes. Ces différents travaux concluaient à des résultats assez similaires entre face à face, téléphone et auto-administré en population adulte (de Winter, 2000).

L'OFDT aurait dû participer au projet européen, mais le partenariat mis en place avec l'Unité Méthodologie Statistique de l'INSEE (profiter du passage des enquêteurs de l'Enquête Permanente de "Conditions de Vie des ménages" pour laisser au répondant un autoquestionnaire) a finalement été rejeté par le Comité de Direction de l'INSEE pour des raisons déontologiques et statistiques14. Les autres solutions, moins fiables, envisagées avec des instituts de sondage privés ont également été rejetées.

13 CASI : Computer Assisted Self Interview
14 Les raisons invoquées étaient principalement la réticence à engager l'image de l'INSEE sur des résultats portant sur un sujet sensible, ainsi que le fait que parmi les enquêteurs de l'INSEE figurent des fonctionnaires et que ces derniers ont obligation de dénonciation de tout comportement illégal porté à leur connaissance. Cette obligation était d'ailleurs en contradiction avec les lois instituant le secret statistique.
Pour les adultes, c'est donc le système CATI qui a été choisi, notamment pour des raisons de coût et de facilité à encadrer le terrain et avec la contrainte d'interroger l'ensemble des ménages y compris ceux inscrits sur listes rouge ou orange (Riandey et Firdion, 1993 ; Beck et al., 2000).

Le dispositif s'articule ainsi sur le Baromètre Santé 2000 du CFES, enquête téléphonique qui aborde les différents comportements et attitudes de santé des Français, en particulier en matière de consommation de substances psychoactives. La durée moyenne d'un entretien est d'environ 35 minutes.

Les baromètres santé existent depuis 1992 et ont notamment permis de mettre en évidence la banalisation de l'usage du cannabis au cours des années 90. L'OFTDT a participé à la mise en place de l'exercice 2000, entraînant l'augmentation du nombre de questions relatives aux drogues et de la taille de l'échantillon. Au départ, l'idée était de suréchantillonner les jeunes adultes (dans la mesure où ils constituent la tranche d'âge la plus concernée\(^\text{15}\)) mais la multiplication des partenaires aux intérêts divers et des financements correspondant ont permis un échantillon de plus de 13,500 individus de 12 à 75 ans.

Le protocole du Baromètre Santé 2000 présente de nombreux avantages car la procédure est pérenne et robuste. Afin d'interroger également les ménages sur liste rouge, les numéros ont été tirés dans l'annuaire, le dernier chiffre de chaque numéro a été incrémenté (+ 1) et l'annuaire inversé a été ensuite utilisé pour récupérer l'adresse des ménages sur liste blanche pour l'envoi d'une lettre-avis. Avec cette méthode, les individus issus de ménages inscrits sur liste rouge représentent environ un cinquième de l'échantillon, à qui l'envoi de la lettre-avis était proposé après la prise de contact téléphonique. Au maximum, 12 appels ont été tentés avant d'abandonner un numéro. Une fois le ménage contacté, la sélection de l'individu se faisait par la méthode "anniversaire"\(^\text{16}\). L'anonymisation du fichier est garantie par l'effacement du numéro de téléphone au terme de l'entretien.

La réalisation, quelques mois plus tard, entre mars et juillet 2000 de l'Enquête sur les Violences Envers les Femmes en France (ENVEFF\(^\text{17}\)) menée dans des conditions méthodologiques très proches, mais uniquement auprès de 7,000 femmes de 20 à 59 ans (Jaspard et al., 2001), offre des perspectives d'exploration de l'effet du contexte (santé et comportements liés à la santé d'une part et violences subies d'autre part)

\(^{15}\) Ce type d'échantillonnage a notamment été mis en œuvre en 1997 lors d'une enquête sur les habitants d'Amsterdam (Abraham et al., 1998). Le suréchantillonnage est parfois aussi pratiqué sur des critères autrement plus contestables, tels que les "ethnies", aux Etats-Unis (NHSDA, 1996) et au Royaume-Uni (Fitzgerald et Hales, 1996) notamment.

\(^{16}\) Sélection du prochain membre du ménage qui fêtera son anniversaire.

\(^{17}\) Enquête commanditée par le Service des Droits des femmes et le Secrétariat d'Etat aux Droits des femmes, en partenariat avec l'ANRS, la CNAF, le FAS, l'IHESI, l'OFTDT, les Conseils régionaux d'Ile de France et de PACA et la mission de recherche Droit et Justice.
dans la mesure où les mêmes questions d’usage des substances psychoactives ont été posées dans les deux enquêtes, plutôt en fin de questionnaire. Ces deux études sont d’ailleurs peut-être les deux dernières grandes enquêtes téléphoniques n’incluant pas les téléphones portables dans la base de sondage, dans la mesure où le taux de couverture du téléphone fixe, maximal à la fin des années 90 (96 %), est au début de l’année 2001 retombé aux alentours de 90 % (Roy et Vanheuverzyn, 2000). Au-delà des adaptations méthodologiques rendues nécessaires dans le déroulement des enquêtes par l’introduction des téléphones portables, l’affectation d’un appareil à une personne plus qu’à un ménage (encore que cette idée reste incertaine) remet en question le calcul des probabilités d’inclusion d’un individu dans l’échantillon.


La première a eu lieu de mars à mai 1999 dans une trentaine de pays européens sur la base d’un questionnaire commun centré sur l’usage des drogues. Il s’agit de l’enquête en milieu scolaire “European School Survey on Alcohol and Other Drugs ” (ESPAD) mise en place par l’INSERM, l’OFDT et le MENRT. Elle offre de réelles perspectives, tant du point de vue du suivi de l’évolution en France que du point de vue de la comparabilité au niveau européen. Si le projet européen concerne en priorité les 15-16 ans, il a semblé opportun de profiter du protocole mis en place pour interroger les 14-19 ans à moindre coût, l’échantillon total atteignant environ 12.000 élèves qui ont été tirés au sort par classe (deux classes sondées entièrement dans chacun des 300 établissements sélectionnés, de la quatrième à la terminale) parmi tous les types d’établissements de l’enseignement public mais aussi du privé. En classe, en présence d’une infirmière ou d’un médecin scolaire, ils ont répondu individuellement à un autoquestionnaire d’une durée d’une heure. Cette enquête fait suite à d’autres investigations en milieu scolaire qui comportaient des questions sur les drogues mais dont le thème principal était la santé (Choquet et Ledoux, 1994), les modes de vie des lycéens parisiens (De Peretti et Leselbaum, 1999) et les conduites déviantes (Ballion, 1999). Reconduite tous les quatre ans à la même époque, elle permettra des comparaisons dans le temps.

Les principaux biais des enquêtes en milieu scolaire sont, du point de vue de la base de sondage, les non-scolarisés et les absentéistes et l’effet de groupe dû au fait que tous les élèves d’une même classe sont interrogés. Du point de vue du mode de collecte, il faut envisager un probable effet de groupe consécutif à la proximité des camarades de classe pendant le remplissage du questionnaire, susceptibles d’exercer une sorte de contrôle social informel des pairs.

Pour remédier au moins en partie à ces biais, l’OFDT a mis en place une Enquête annuelle sur la Santé et les Consommations lors de l’Appel et de Préparation à la Défense (ESCAPAD). Elle a lieu au cours de la Journée d’Appel et de Préparation à la Défense qui remplace le service militaire depuis octobre 1998 et concerne aussi

INSEE Méthodes

135
les filles depuis avril 2000. Tous les appelés présents un jour donné (environ 14.000) remplissent un autoquestionnaire pendant 25 minutes, le nombre de questions posées étant nettement inférieur à celui des deux autres enquêtes. Ils ont entre 17 et 19 ans, cette tranche d’âge très restreinte offrant une forte puissance statistique par âge exact.


Le système d’observation des usages de substances psychoactives à l’adolescence s’est donc récemment vu enrichi d’une enquête complète, sur plusieurs âges, en milieu scolaire, et d’une enquête au questionnaire plus léger mais dont l’étroitesse de la tranche d’âge observée et la taille d’échantillon offrent une grande robustesse, et dont la périodicité annuelle permettra à terme un suivi de tendances.

Au final, ce dispositif d’enquêtes profite de certaines circonstances favorables (milieu scolaire, journée d’appel de préparation à la défense) pour mettre en œuvre auprès des plus jeunes des modes de collecte privilégiés. Il reste néanmoins précis pour les jeunes adultes [18-44 ans] grâce à des tailles d’échantillon importantes.

### 3.2. Quelques limites du dispositif

Concernant les adultes, si l’entretien téléphonique semble bien adapté à la population des ménages “ordinaires”, ce mode de collecte peut s’avérer défaillant pour certaines populations particulières qui, dans le cadre de la consommation de drogues, sont porteuses d’une information complémentaire. Cela peut être le cas pour des raisons techniques (ménage non-abonné, individu absent aux heures d’appel courantes) mais aussi pour des raisons culturelles (problèmes de langue, méfiance, réticence à dévoiler des comportements intimes dans un tel contexte de distance). Ces populations réunissent entre autres des personnes résidant dans les “villes” des zones périurbaines.


A priori, même s’il ne réglerait pas tous les problèmes, le mode le mieux adapté serait le face à face administré par une personne jeune (20-30 ans), connue dans la cité, capable d’expliquer les objectifs de l’enquête. Il apparaît assez clairement que l’investigation menée dans ces quartiers doit être préparée de façon très attentive et tenir notamment compte des réseaux relationnels spécifiques qui s’y tissent.
En règle générale, les enquêtes quantitatives sont peu appropriées à l’approche compréhensive car la diversité des usages peut être rapidement gommée par les agrégats statistiques. Par exemple, les individus qui ont consommé plus de dix fois du cannabis au cours de l’année vont être agrégés en une seule catégorie statistique même s’ils renvoient à une multiplicité de profils. Cette limite était encore plus criante au début des années 90 quand les prévalences étaient plus faibles et que les individus qui avaient consommé au moins dix fois du cannabis au cours de leur vie étaient désignés “ usagers réguliers ” (Choquet et Ledoux, 1994).

Le suivi des tendances récentes et des nouvelles drogues est un travail de longue haleine qui se satisfait mieux d’investigations de terrain, participatives, culturelles. Pour répondre à ces deux questions, les approches ethnographiques, notamment en réseau, centrées sur des milieux d’usagers (OFDT, 2000 ; Fontaine et al., 2001) ou sur des groupes particuliers (Aquatias, 1998) se révèlent bien meilleures que les enquêtes quantitatives. Sur un produit tel que la kétamine par exemple, l’enquête ESCAPAD 2000 ne compte que 9 individus en ayant déjà pris sur 14.000 interrogés, parce que son usage est certes rare, mais sans doute aussi parce qu’il ne peut être déclaré que spontanément, en réponse à une question ouverte18 (Beck, Legleye et Peretti-Watel, 2000).

Il convient ici de rappeler que d’autres modes d’échantillonnage, comme les “ boules de neige19 ”, par ailleurs extrêmement efficaces pour décrire un certain nombre de phénomènes liés à l’usage des drogues et atteindre des populations dites cachées (Ingold, 1992), ne peuvent fournir d’indicateurs fiables par rapport au reste de la population. En effet, calculer un âge moyen sur une population construite par boule de neige au sein d’une sphère prédéterminée constituerait une méthode contestable dans la mesure où les probabilités d’inclusion sont incalculables et où les différents groupes d’âge ne se mélangent pas forcément.

3.3. Une alternative : le comptage des usagers à problèmes

Le repérage des usagers à problèmes est aussi particulièrement difficile en raison du manque de consensus sur les définitions (peut-on avoir de grosses consommations qui ne soient pas jugées problématiques ?) La réflexion s’oriente actuellement vers la notion de conduite “ atypique ”, qui peut, dans le cas du cannabis, être définie par le fait de consommer souvent le matin ou souvent en solitaire, mais elle reste pour l’instant à l’état d’ébauche. De plus, les questions susceptibles d’être mobilisées pour la définition des usagers à problèmes, souvent stigmatisantes, nécessitent un entretien approfondi où le contexte de consommation doit être abordé dans le détail. C’est le cas, pour l’alcool, avec le test clinique DETA20 qui repère les usagers

---

18 Cette question portait sur les “ Autres drogues ” consommées.
19 Permettent d’atteindre des membres d’un groupe grâce au réseau relationnel, de proche en proche.
20 Diminuer, Entourage, Trop, Alcool. Il s’agit d’une traduction du test américain CAGE.
d'alcool "à risque". Sont considérés comme tels ceux qui répondent par l'affirmative à au moins deux de ces questions :

- Avez-vous déjà ressenti le besoin de diminuer votre consommation de boissons alcoolisées ?

- Votre entourage vous a-t-il déjà fait des remarques au sujet de votre consommation ?

- Avez-vous déjà eu l'impression que vous buviez trop ?

- Avez-vous déjà eu besoin d'alcool dès le matin pour vous sentir en forme ?

Si cet indicateur est utile dans les enquêtes en population générale, notamment pour étudier l'évolution du nombre de positifs au test dans le temps (Baudier et Arènes, 1997), sa validité a récemment été remise en question dans sa version américaine (Bisson, Nadeau et Demers, 1999).

Pour l'estimation du nombre des usagers à problèmes d'opiacés (hééroïne et dérivés de l'opium) et de cocaïne, des méthodes spécifiques à l'observation des populations difficiles à joindre, telles que capture/recapture, sont privilégiées. Si cette méthode a été conçue pour des zones géographiques restreintes et clairement délimitées, des chiffres nationaux peuvent être envisagés en extrapolant plusieurs estimations locales de prévalence.

L'objectif d'une telle méthode est d'estimer la taille d'une population à partir de données issues de différentes sources. Conçue au 19ème siècle pour estimer des tailles de populations animales (Seber, 1986 ; 1992), notamment les poissons (Pollock, 1991), cette technique a été utilisée en démographie pour valider des données issues de registres d'Etat civil (Sekar et Deming, 1949) puis à partir des années 70 en épidémiologie. Au cours des années 90, elle a également aidé dans les comptages des sans-abri (Burt, 1992). Dans le cas de la toxicomanie, les sources peuvent être la police, les urgences hospitalières, centres d'accueil pour usagers de drogues...

La méthode est simple : pour estimer une population totale de N usagers à problèmes d'opiacés, il faut disposer d'au moins deux sources d'observation S1 et S2. Dans ce qui suit, l'indice 1 indique la présence dans la source, l'indice 2 signifie l'absence dans cette source et le premier indice concerne la source 1, le deuxième la source 2.

Dans la source S1, il y a N1 cas, qui sont les n12 cas uniquement retrouvés dans la source S1, et les n11 cas présents dans les deux sources (doublons). De même, dans la source S2, se trouvent N2 = n11 + n21 cas. Enfin, la population N comprend des individus qui n'ont été recensés par aucune des deux sources, au nombre de n22. On a donc N = N1 + N2 + n22.

Pour deux sources de recueil de l'information, et sous l'hypothèse d'indépendance des sources, on a : n22 = (n12 - n21)/n11
Il est ainsi possible d’estimer la population totale de la manière suivante :

\[ N = n_{11} + n_{21} + n_{12} + \frac{(n_{12} \cdot n_{21})}{n_{11}} \]

Cette estimation n’est valable que sous plusieurs conditions :

- La définition d’un usager doit être rigoureusement la même d’une source à l’autre, alors que dans la réalité, elle peut être différente selon les sources. De plus, là encore, le problème de la frontière se pose : les usagers consommant surtout des produits de substitution aux opiacés, tels que la méthadone ou le Subutex®, doivent-ils être inclus ?

- La période d’observation et la zone géographique doivent être précisément limitées. Il ne doit y avoir ni arrivée, ni départ d’usagers pendant le recueil (autrement dit, la population doit être stable), aussi est-il essentiel de réaliser un recueil simultané dans chaque source et de limiter la durée de l’enquête.

- Tous les doublons, le plus souvent identifiés par méthode semi-manuelle\(^\text{21}\), doivent être contrôlés. S’ils ne sont pas tous identifiés, le dénominateur \((n_{11})\) est sous-estimé et \(N\) est donc sur-estimé, et inversement.

- Il doit y avoir indépendance des sources : la présence d’un individu dans une source ne doit pas modifier sa probabilité de présence dans l’autre source. Lorsqu’il existe une dépendance positive entre deux sources, \(N\) est sous-estimé et réciproquement, si elle est négative il y a sur-estimation de \(N\).

- La probabilité de capture doit être la même pour chaque individu.

Dans le cas de deux sources, l’indépendance des sources n’est vérifiable qu’à l’appréciation sur le terrain, mais si trois sources sont disponibles, il est possible de réaliser des estimations à partir de modèles log-linéaires (Cormack, 1989 ; Wickens, 1993). Ceux-ci permettent de prendre en compte les interactions entre sources et l’hétérogénéité de la population étudiée. Ainsi, il n’est plus nécessaire de disposer de sources indépendantes entre elles.

Une telle méthode a été initiée en France en 1995 et mise en place sur l’agglomération toulousaine (Bello, 1997) puis sur cinq sites (Lens, Lille, Marseille, Nice et Toulouse) en 1999 (Chevallier et al., 2001), l’OFDT envisageant de réaliser des estimations nationales à partir de cette étude multicentrique.

\(^{21}\) Pour des questions de respect de l’anonymat, la recherche des doublons s’est faite, en France, sur la première lettre du nom, les trois premières lettres du prénom et la date de naissance. Deux fiches pouvaient dès lors présenter soient une stricte identité, soit de fortes similitudes pouvant conduire à un nouvel examen en détail afin de se prononcer avec certitude.
Dans le cas des produits licites, tels que l'alcool, le tabac ou les médicaments psychotropes, il est toujours possible de partir de l'observation des données de vente comme le propose par une allégorie Jünger (1970) : « On peut se faire une idée de cette consommation massive de drogues dans les usines à produits pharmaceutiques, devant les centrifugeuses, dont les cachets jaillissent en une succession rapide. Ils s'unissent en rivières multicolores qui, à leur tour, se ramifient jusque dans les villages et les foyers les plus lointains. », mais chacun sait à quel point il est difficile d'estimer la production de produits illicites...

D'autres approches, telles que le recours aux statistiques institutionnelles (interpellations, demandes de soins) sont possibles. Des estimations nationales de prévalence sont parfois fournies sur ces bases, mais elles mesurent souvent l'activité des services plus que l'ampleur véritable du phénomène. Elles reposent la plupart du temps sur des méthodes de calcul fustes et parfois tautologiques qui consistent à déterminer, par exemple, un taux de recours au soin (sur des bases qui ne peuvent être absolument rigoureuses) puis à appliquer ce taux à la population repérable des usagers de drogues en traitement. La solution retenue est souvent de proposer une fourchette de prévalences à partir des résultats de différentes stratégies (Costes, 1988; 1990). La confrontation des données administratives et des déclarations issues d'enquêtes en population générale est souvent édifiante. Ainsi Jaspard et al. (2001) mettent-ils en regard les 3.350 viols déclarés par les femmes majeures sur l'année 1998 à la police ou à la gendarmerie et l'estimation issue de leur enquête (entre 32.000 et 64.000 femmes de 20 à 59 ans auraient été victimes de viols au cours de l'année 1999).

4. Conclusion

Les enquêtes en population générale se révèlent particulièrement efficaces pour quantifier les usages occasionnels ou réguliers des produits dont la consommation est banalisée. Dès que les produits ou les comportements d'usages deviennent rares, les tailles d'échantillon et les différents biais liés aux bases de sondage classiques compromettent la justesse d'une observation quantitative. Ceci est d'autant plus vrai

---

22 Une vision caricaturale de ces méthodes est donnée par un texte d’Anslinger et Cooper (1937) qui montre à quel point le choix de l'instrument de mesure lui-même et le niveau d'investissement consenti déterminent mécaniquement le niveau estimé : « La menace de la dépendance à la marijuana est relativement nouvelle en Amérique. En 1931, son dossier au Bureau des Stupéfiants américain ne faisait pas plus de cinq centimètres d'épaisseur, alors qu'aujourd'hui les rapports envahissent d'innombrables rayons de bibliothèque. » Harry Anslinger, créateur du système de lutte contre la drogue aux États-Unis, a dirigé le Bureau des Stupéfiants de 1931 à 1962. Précisons qu'il souhaitait par ce texte ériger définitivement la marijuana au rang de drogue et dramatiser la situation en matière de drogues aux États-Unis, afin de promouvoir la politique de lutte contre ces produits dans la foulée de la prohibition de l'alcool.
que la déclaration de tels comportements peut être vécue comme compromettante par l’enquêté.

L’avenir est sans doute dans la diversification des modes de collecte pour une même enquête, telle qu’elle est pratiquée aux Pays-Bas dans l’enquête menée en 2000 par le CEDRO par exemple, où les enquêtés reçoivent une lettre contenant un argumentaire et un questionnaire (papier ou disquette) en plusieurs langues (néerlandais, arabe et turque), ainsi qu’une proposition de répondre via Internet (solution retenue par 6% des répondants). Elle pourrait inciter les équipes de recherche françaises à proposer des alternatives à l’enquête téléphonique auprès des adultes en cas de refus ou même dès l’envoi de la lettre-avis.

A tous les niveaux, le caractère illicite de l’usage de drogues hypothèque les chances de recueillir une information réellement fiable, d’autant qu’elle contribue à faire de la réponse de l’enquêté un geste politique, dans la mesure où il sait que sa réponse alimentera le débat public. L’analyse d’une question ouverte posée à la fin de l’enquête ESCAPAD a également permis de voir à quel point les jeunes interprètent aussi l’enquête comme le signe d’une prise en compte des questions relatives aux drogues par les pouvoirs publics (Beck, Legleye et Peretti-Watel, 2000). Il convient ainsi de toujours rester prudent dans les analyses car si l’usage d’une substance illicite était dépénalisé en France et que, par la suite, le niveau d’usage de cette substance révélé par les enquêtes augmentait, ce pourrait être dû en partie à une plus grande aisance dans la déclaration ou à un désir de recomposer sa biographie.

Les enquêtes en population générale sont utiles car tous les travaux sur la toxicomanie nécessitent des données de cadrage quantitatives. Par exemple, la connaissance de la proportion d’usagers dans une population particulière n’a de sens que mise en regard de celle estimée sur l’ensemble des individus ayant le même profil socio-démographique dans la population générale. Ainsi, dans quelle mesure (au sens de l’instrument de mesure dont il s’agit de disposer) faut-il s’inquiéter de la proportion de consommateurs du cannabis parmi les jeunes adultes conducteurs ayant eu un accident de la circulation, dès lors qu’elle est similaire à celle de l’ensemble des jeunes en âge de conduire ? Si interdire le volant aux individus sous influence cannabique ou plus généralement psychoactive est une mesure légitime puisqu’il est possible de mesurer le risque encouru, sanctionner la présence de cannabis dans les urines l’est nettement moins dès lors que les traces de cette substance peuvent s’y avérer rémanentes pendant plus d’un mois.

Mais au-delà de ces considérations politiques, les données quantitatives servent également à mieux guider les attitudes adoptées par les acteurs de terrain : une infirmière scolaire ou un travailleur social changeront sans doute le regard qu’ils portent sur un adolescent surpris en train de fumer un joint s’ils ont en tête que plus de la moitié des jeunes se sont déjà, un jour ou l’autre, retrouvés dans la même situation.

Malgré les écarts qui jalonnent la construction d’un dispositif d’observation en population générale et les limites inhérentes à ce type d’approche, il ne faut pas
négliger l'utilité des données quantitatives issues d'un tel recueil. Au final, on peut retenir que l'essentiel est d'essayer d'améliorer l'existant et ne pas espérer répondre à un trop grand nombre de questions grâce aux enquêtes en population générale. Bien observer dès lors à privilégier le suivi des tendances et à accepter les biais sans jamais les oublier, en considérant qu'à biais constant une évolution peut être, le cas échéant, jugée significative.

**Bibliographie**


PANEL D’INDIVIDUS VERSUS PANEL DE LOGEMENTS OU : QUE PEUT-ON DIRE DE LA QUALITÉ DU PANEL EUROPÉEN ?

P. BREUIL-GENIER, N. LEGENDRE et H. VALDELIEVRE

Insee, Division Revenus et Patrimoine

Introduction : panel d’individus versus panel de logements

Le panel européen mis en place à la demande d’Eurostat en 1994 est un panel d’individus : toutes les personnes considérées comme répondantes lors de la première vague (automne 1994) ont été de nouveau sollicitées pour les vagues annuelles suivantes à condition de ne pas avoir été non-répondantes pendant deux vagues de suite¹. La collecte de la vague 7 s’est déroulée à l’automne 2000, mais seules les vagues 1 à 4 (1994 à 1997) seront exploitées ici. Les « individus panels » sont suivis même s’ils déménagent, sauf s’ils sortent du champ de l’enquête (en déménageant en institution ou à l’étranger) ou bien s’ils décèdent (dans ces deux cas, on qualifiera ces individus de « hors champ » dans la suite).

Un panel d’individus diffère donc dans son principe d’un panel de logements. En effet, dans des panels de logements - comme les enquêtes Emploi ou les Enquêtes Permanentes Conditions de Vie (PCV) - une partie des logements enquêtés une année sont de nouveau enquêtés l’année suivante (un tiers dans les enquêtes Emploi, la moitié dans les enquêtes PCV). Dans le cas de l’enquête Emploi, il n’est donc possible de reconstituer un historique des trajectoires d’activité sur trois ans que pour les personnes qui n’ont pas déménagé pendant trois ans. Or on peut penser que les personnes n’ayant pas déménagé ne sont pas représentatives de la population française. C’est par exemple le constat que fait Magnac (1997)² qui exploite trois années de l’enquête Emploi pour étudier l’influence des stages sur l’insertion professionnelle des jeunes de 18 à 29 ans : ne retenir pour l’étude que les jeunes qui ont été enquêtés trois années de suite conduit à travailler sur un échantillon de personnes plus jeunes, venant de familles où les parents ont un niveau d’éducation plus faible mais divorcent moins³. Le biais potentiel lié à l’utilisation d’un sous-

¹ A chaque vague d’enquête sont également interrogées toutes les personnes qui appartiennent au ménage d’un individu panel.
³ Magnac (1997) note par ailleurs que ce biais de sélection ne serait pas un problème pour l’analyse économétrique si les caractéristiques inobservables qui expliquent la sélection n’influençaient pas le
échantillon de personnes n’ayant jamais déménagé est par ailleurs d’autant plus susceptible d’être important que les populations étudiées sont mobiles ou dans des situations précaires. Ainsi, dans l’étude de Lagarenne et Legendre (2000), 20 % des travailleurs pauvres présents dans l’enquête Revenus Fiscaux de 1996 (basée sur le tiers médian de l’enquête Emploi de 1997) ne répondent pas à l’enquête Emploi 1998. Cette proportion monte à 32 % pour les travailleurs pauvres qui ont connu l’année précédente un parcours incluant des passages vers l’inactivité, alors qu’elle n’est que de 9 % pour les personnes qui ont exercé une activité d’indépendant pendant les douze derniers mois.

Si se restreindre aux personnes n’ayant jamais déménagé pour réaliser des études longitudinales est susceptible de générer des biais, ce que l’on illustrera dans le cas du panel européen (Ib), le recours à un panel d’individus ne résout pas en pratique tous les problèmes. Ainsi, il est difficile de suivre les individus qui déménagent (Ia), et ceux que l’on retrouve ne sont pas parfaitement représentatifs de l’ensemble (Ib). Toutefois, l’utilisation d’une pondération adaptée pour l’exploitation longitudinale d’un panel de logements peut limiter ce dernier biais (Ic). Si la pondération peut également permettre de réduire le biais lié à l’absence de suivis en cas de déménagement dans le cadre d’un panel de logement, elle conduit à multiplier très fortement le poids de certains individus (Id).

Au final, le recours à un panel d’individus reste sans doute de toute façon préférable, dès lors que le phénomène étudié n’est pas indépendant de l’existence d’un déménagement, ou que le nombre de vagues d’enquêtes à exploiter est important ou la population à étudier est mobile. Mais quel est le prix à payer pour disposer d’un panel d’individus ? Financièrement d’abord, le coût moyen par questionnaire rempli est plus élevé (I1a), d’où des échantillons souvent plus réduits. La précision des phénomènes que l’on peut étudier à partir du panel reste-t-elle alors suffisante, ou, en d’autres termes, celui-ci continue-t-il à être suffisamment représentatif de la population de l’année courante au fur et à mesure que le nombre de vagues augmente (I1c) ? On verra que sur l’exemple des taux de transition entre emploi, chômage et inactivité, l’éventuelle perte de représentativité de la vague 4 du panel par rapport à l’enquête Emploi semble limitée.

---

phénomène étudié (en l’occurrence l’insertion) mais précise que cela n’est probablement pas le cas (par exemple, si un jeune qui trouve un emploi stable quitte ses parents, il n’est plus observé).


5 Nous ne discutons pas dans cette étude de l’alternative consistant à n’effectuer qu’une enquête ponctuelle comportant un important calendrier rétrospectif. Ce type d’enquête a l’avantage de ne pas poser de problème d’attrition, mais a l’inconvénient d’être sensible aux problèmes de mémoire (ce qui interdit sans doute les questions sur certains thèmes, comme les revenus).
1. Peut-on (et faut-il ?) suivre les individus qui déménagent ?

1.1 Plus de 8 individus sur 10 sont encore répondants après un déménagement

Parmi les individus panels de 17 ans et plus (encadré 1) qui remplissent un questionnaire individuel lors d’une vague, environ 90 % en remplissent encore un la vague suivante (tableau 2). Bien que le taux de réponse soit plus faible parmi ceux qui ont déménagé (annexe), il reste élevé (entre 75 et 84 % selon les vagues). Au total, 82 % des personnes interrogées en première vague ont également répondu aux deux suivantes. Ce taux de réponse sur trois vagues sur les données françaises est proche de ceux observés en Belgique, en Grèce, en Allemagne ou au Luxembourg. Seuls le Portugal et l’Italie arrivent à des taux de réponse nettement supérieurs (90 %), tandis qu’en Grande-Bretagne et en Irlande les taux de réponse sont plus faibles (65 %)\(^6\).

Si l’on avait renoncé à suivre les individus qui déménagent, la part d’individus de la première vague pour lesquels on aurait eu des réponses pour les trois vagues suivantes aurait été de 58 % (et non de 73 %, tableau 1), soit une diminution supplémentaire de 20 % de la taille de l’échantillon. Et cette perte d’échantillon s’accentue avec le nombre de vagues\(^7\).

*Tableau 1 - Incidence du déménagement sur le pourcentage d’individus réinterrogés*

<table>
<thead>
<tr>
<th></th>
<th>Absence de suivi en cas de déménagement</th>
<th>Suivi en cas de déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vague 2</td>
<td>82</td>
<td>88</td>
</tr>
<tr>
<td>Vagues 2 et 3</td>
<td>71</td>
<td>82</td>
</tr>
<tr>
<td>Vagues 2, 3 et 4</td>
<td>58</td>
<td>73</td>
</tr>
</tbody>
</table>

*Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee*

*Champ : individus panels de 17 ans ou plus ayant rempli un questionnaire individuel à la vague 1 – Données non pondérées*

*Lecture : parmi les individus interrogés à la vague 1, 73 % répondent aux trois vagues suivantes, mais seuls 58 % répondent aux trois vagues suivantes et ne déménagent pas entre les vagues 1 et 4*


### Tableau 2 - Taux de déménagement et taux de réponse en fonction de l’existence d’un déménagement

<table>
<thead>
<tr>
<th></th>
<th>v1-&gt;v2</th>
<th>v2-&gt;v3</th>
<th>v3-&gt;v4</th>
<th>v1-&gt;v4</th>
</tr>
</thead>
<tbody>
<tr>
<td>répartition</td>
<td>91</td>
<td>90</td>
<td>90</td>
<td>89</td>
</tr>
<tr>
<td>taux de réponse</td>
<td>90</td>
<td>94</td>
<td>89</td>
<td>91</td>
</tr>
<tr>
<td>n'ont pas déménagé</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ont déménagé</td>
<td>8</td>
<td>79</td>
<td>9</td>
<td>84</td>
</tr>
<tr>
<td>dont :</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ont déménagé avec leur ménage</td>
<td>6</td>
<td>81</td>
<td>7</td>
<td>86</td>
</tr>
<tr>
<td>dont :</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ont déménagé en se séparant de leur ménage</td>
<td>2</td>
<td>73</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>hors champ (dont décès)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>total /moyenne</td>
<td>100</td>
<td>88</td>
<td>100</td>
<td>93</td>
</tr>
<tr>
<td>effectif</td>
<td>14334</td>
<td>12677</td>
<td>12992</td>
<td>12018</td>
</tr>
</tbody>
</table>

**Source** : panel européen, vagues 1 à 4 (1994 à 1997), Insee

**Champ** : individus panels de 17 ans ou plus ayant rempli un questionnaire individuel à la première des deux vagues considérées - Données non pondérées

**Lecture** : parmi les individus de 17 ans ou plus ayant rempli un questionnaire individuel en vague 1, 91 % n’ont pas déménagé en V2, et parmi eux, 90 % ont répondu au questionnaire V2.
Dans plus de la moitié des cas, les non-réponses faisant suite à un déménagement seraient dues au fait que le nouveau ménage n’a pas été retrouvé (soit l’enquêteur n’a pas obtenu la nouvelle adresse, soit le ménage n’a pas été retrouvé à cette adresse). Lorsque la personne est retrouvée à sa nouvelle adresse, sa probabilité de réponse semble comparable à celle des enquêtés n’ayant pas déménagé (tableau 3).

Au sein de ceux qui déménagent, le taux de non-réponse varie encore en fonction du « type » de déménagement. Il est en effet plus important chez les personnes qui quittent leur ménage (enfants qui partent, conjoints qui se séparent) que chez celles qui déménagent avec l’ensemble de leur ménage. Les taux de non-réponse sont particulièrement élevés (plus d’un tiers) en cas de séparation de conjoints ou de départ d’une personne autre qu’un conjoint ou un enfant. En effet, c’est dans ces cas que retrouver l’adresse de la personne qui est partie est le plus difficile (tableau 3).

**Tableau 3 - Fréquence et causes de la non-réponse en fonction du type de déménagement**

<table>
<thead>
<tr>
<th></th>
<th>répartition</th>
<th>taux de non-réponse</th>
<th>dont : non réponse liée à une adresse non retrouvée</th>
<th>dont : non réponse liée à un refus, une absence de contact ou une raison inconnue</th>
</tr>
</thead>
<tbody>
<tr>
<td>déménagement sans éclatement</td>
<td>76</td>
<td>18</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>déménagement avec éclatement</td>
<td>24</td>
<td>26</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>séparation du conjoint</td>
<td>19</td>
<td>36</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>départ d’un enfant</td>
<td>73</td>
<td>22</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>départ d’une autre personne</td>
<td>8</td>
<td>44</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>total déménagement</td>
<td>100</td>
<td>20</td>
<td>11</td>
<td>9</td>
</tr>
</tbody>
</table>

*Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee*

*Champ : Individus panels de 17 ans et plus ayant rempli un questionnaire individuel à la vague précédente et qui ont déménagé entre les deux vagues tout en restant dans le champ de l’enquête. Moyennes non pondérées sur les vagues 2, 3 et 4.*
1.2 Ce sont les jeunes et les moins insérés sur le marché du travail qui déménagent le plus

Parmi les individus de 17 ans et plus répondants en vague 1 qui sont restés dans le champ de l’enquête lors des trois vagues suivantes, 22 % ont déménagé entre les vagues 1 et 4, parmi lesquels 68 % sont répondants les 3 vagues suivantes. Ne pas réinterroger ces individus après leur déménagement pose deux problèmes : en premier lieu, cela réduit fortement la taille de l’échantillon exploitable pour des études longitudinales - donc la précision des estimations - mais surtout, cela risque d’introduire un biais dans les résultats, si les individus que l’on ne réinterroge pas sont différents de ceux que l’on réinterroge (et en particulier si ces différences portent sur des caractéristiques inobservables). Or, comme nous allons le voir, les individus qui sont les plus susceptibles de déménager sont les jeunes et les personnes les moins insérées sur le marché du travail, c’est-à-dire deux populations pour lesquelles l’intérêt d’études longitudinales est particulièrement important. Mais nous verrons aussi que ce sont également ces populations que l’on a le plus difficulté à suivre en cas de déménagement.

Près d’un jeune de 17-25 sur deux déménage en trois ans

La probabilité de déménagement varie fortement en fonction de l’âge. Elle est maximale pour les 21-25 ans, qui sont 55 % à déménager au cours des trois années étudiées, et elle reste très élevée pour les 26-30 ans (43 %) et les 17-20 ans (36 %). Après 26 ans, le taux de déménagement en trois ans décroît avec l’âge, jusqu’à un taux de 7 % chez les personnes de plus de 61 ans (tableau 4).

Tableau 4a : Ventilation des individus panel selon l'existence d'un déménagement et la réponse aux vagues 2 à 4, en fonction des caractéristiques des individus en début de période.

<table>
<thead>
<tr>
<th>Caractéristiques individuelles</th>
<th>Ensemble</th>
<th>Pas de déménagement</th>
<th>Déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ensemble</td>
<td>Répondants V2, V3 et V4</td>
<td>Autres</td>
</tr>
<tr>
<td>Niveau d'études atteint</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aucun diplôme</td>
<td>100</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>CAP, BEP</td>
<td>100</td>
<td>74</td>
<td>27</td>
</tr>
<tr>
<td>BEPC</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Baccalauréat</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Baccalauréat plus 2 années</td>
<td>100</td>
<td>79</td>
<td>21</td>
</tr>
<tr>
<td>Supérieur à bac plus 2</td>
<td>100</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Total des individus</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actif occupé</td>
<td>100</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Chômeur</td>
<td>100</td>
<td>64</td>
<td>36</td>
</tr>
<tr>
<td>Etudiant</td>
<td>100</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>Retraité</td>
<td>100</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>Autre inactif</td>
<td>100</td>
<td>75</td>
<td>26</td>
</tr>
<tr>
<td>Total des individus</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Âge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-20 ans</td>
<td>100</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>21-25 ans</td>
<td>100</td>
<td>69</td>
<td>31</td>
</tr>
<tr>
<td>26-30 ans</td>
<td>100</td>
<td>76</td>
<td>24</td>
</tr>
<tr>
<td>31-40 ans</td>
<td>100</td>
<td>79</td>
<td>21</td>
</tr>
<tr>
<td>41-50 ans</td>
<td>100</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>51-60 ans</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Plus de 61 ans</td>
<td>100</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>Total des individus</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee
Tableau 4a (suite) : Ventilation des individus panels selon l'existence d'un déménagement et la réponse aux vagues 2 à 4, en fonction des caractéristiques des individus en début de période.

<table>
<thead>
<tr>
<th>Caractéristiques individuelles</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Non retrouvés</th>
<th>Autres non réponses</th>
<th>Part</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statut des actifs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salarié</td>
<td>100</td>
<td>78</td>
<td>22</td>
<td>75</td>
<td>60</td>
<td>15</td>
<td>25</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td>88</td>
</tr>
<tr>
<td>Indépendant</td>
<td>100</td>
<td>76</td>
<td>24</td>
<td>84</td>
<td>65</td>
<td>19</td>
<td>16</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Total des actifs</td>
<td>100</td>
<td>78</td>
<td>22</td>
<td>76</td>
<td>61</td>
<td>15</td>
<td>24</td>
<td>17</td>
<td>3</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>PCS des actifs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculteur</td>
<td>100</td>
<td>82</td>
<td>18</td>
<td>90</td>
<td>74</td>
<td>16</td>
<td>11</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Artisan</td>
<td>100</td>
<td>71</td>
<td>29</td>
<td>80</td>
<td>60</td>
<td>20</td>
<td>20</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Cadre supérieur</td>
<td>100</td>
<td>80</td>
<td>20</td>
<td>73</td>
<td>60</td>
<td>13</td>
<td>27</td>
<td>20</td>
<td>2</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Profession intermédiaire</td>
<td>100</td>
<td>82</td>
<td>18</td>
<td>74</td>
<td>62</td>
<td>12</td>
<td>27</td>
<td>21</td>
<td>2</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Employé</td>
<td>100</td>
<td>77</td>
<td>23</td>
<td>75</td>
<td>59</td>
<td>16</td>
<td>25</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>Ouvrier</td>
<td>100</td>
<td>76</td>
<td>24</td>
<td>76</td>
<td>61</td>
<td>16</td>
<td>24</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>Total des actifs</td>
<td>100</td>
<td>78</td>
<td>22</td>
<td>76</td>
<td>61</td>
<td>15</td>
<td>24</td>
<td>17</td>
<td>3</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Type de contrat de travail des salariés</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDD</td>
<td>100</td>
<td>73</td>
<td>27</td>
<td>61</td>
<td>47</td>
<td>14</td>
<td>39</td>
<td>26</td>
<td>6</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>CDI</td>
<td>100</td>
<td>79</td>
<td>21</td>
<td>76</td>
<td>62</td>
<td>15</td>
<td>24</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td>90</td>
</tr>
<tr>
<td>Total des salariés</td>
<td>100</td>
<td>78</td>
<td>22</td>
<td>74</td>
<td>60</td>
<td>14</td>
<td>26</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td>100</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), INSEE
Champ : individus panel de 17 ans ou plus en 1994 ayant rempli un questionnaire individuel en première vague. Individus hors champ en vagues 2, 3 ou 4 exclus.
Caractéristiques de 1994. Données pondérées (pondérations de la vague 1)
### Tableau 4b : Ventilation des individus panels selon l'existence d'un déménagement et la réponse aux vagues 2 à 4, en fonction des caractéristiques de leur ménage en début de période.

<table>
<thead>
<tr>
<th>Caractéristiques ménage</th>
<th>Ensemble</th>
<th>Pas de déménagement</th>
<th>Déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ensemble</td>
<td>Répondants V2, V3 et V4</td>
<td>Autres</td>
</tr>
<tr>
<td>TYPE DE MÉNAGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personne seule</td>
<td>100</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>Couple sans enfant</td>
<td>100</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>Couple avec enfant</td>
<td>100</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Famille monoparentale</td>
<td>100</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>Autre type de ménage</td>
<td>100</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>TOTAL DES MÉNAGES</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>ZEAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile de France</td>
<td>100</td>
<td>71</td>
<td>30</td>
</tr>
<tr>
<td>Bassin parisien</td>
<td>100</td>
<td>76</td>
<td>24</td>
</tr>
<tr>
<td>Nord</td>
<td>100</td>
<td>74</td>
<td>26</td>
</tr>
<tr>
<td>Est</td>
<td>100</td>
<td>76</td>
<td>24</td>
</tr>
<tr>
<td>Ouest</td>
<td>100</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Sud-ouest</td>
<td>100</td>
<td>77</td>
<td>23</td>
</tr>
<tr>
<td>Centre-est</td>
<td>100</td>
<td>77</td>
<td>24</td>
</tr>
<tr>
<td>Méditerranée</td>
<td>100</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>TOTAL DES MÉNAGES</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>STRATE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commune rurale</td>
<td>100</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Unité urbaine &lt; 20000 hab</td>
<td>100</td>
<td>76</td>
<td>24</td>
</tr>
<tr>
<td>20000&lt;= u.u. &lt;=100000 hab</td>
<td>100</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>u.u. de plus de 100000 hab</td>
<td>100</td>
<td>100</td>
<td>26</td>
</tr>
<tr>
<td>TOTAL DES MÉNAGES</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee

**Tableau 4b (suite) : Ventilation des individus panels selon l’existence d’un déménagement et la réponse aux vagues 2 à 4, en fonction des caractéristiques de leur ménage en début de période.**

<table>
<thead>
<tr>
<th>Caractéristiques ménage</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Pas de déménagement</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Déménagement</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Non retrouvés</th>
<th>Autres non réponses</th>
<th>Part</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statut d’occupation du logement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locataire (yc, sous locataire et logé gratuitement)</td>
<td>100</td>
<td>73</td>
<td>28</td>
<td>61</td>
<td>46</td>
<td>15</td>
<td></td>
<td>39</td>
<td>27</td>
<td>6</td>
<td>6</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Propriétaire</td>
<td>100</td>
<td>76</td>
<td>24</td>
<td>89</td>
<td>69</td>
<td>20</td>
<td></td>
<td>11</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Total des ménages</td>
<td>100</td>
<td>75</td>
<td>25</td>
<td>78</td>
<td>60</td>
<td>18</td>
<td></td>
<td>22</td>
<td>15</td>
<td>3</td>
<td>4</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td><strong>Propriétaire (pour locataires)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Famille</td>
<td>100</td>
<td>75</td>
<td>25</td>
<td>77</td>
<td>58</td>
<td>18</td>
<td></td>
<td>23</td>
<td>17</td>
<td>3</td>
<td>4</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Propriétaire privé</td>
<td>100</td>
<td>72</td>
<td>28</td>
<td>50</td>
<td>38</td>
<td>12</td>
<td></td>
<td>50</td>
<td>34</td>
<td>9</td>
<td>7</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>HLM ou organisme public</td>
<td>100</td>
<td>73</td>
<td>27</td>
<td>70</td>
<td>52</td>
<td>18</td>
<td></td>
<td>30</td>
<td>21</td>
<td>4</td>
<td>5</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Total des ménages de locataires</td>
<td>100</td>
<td>72</td>
<td>28</td>
<td>62</td>
<td>46</td>
<td>16</td>
<td></td>
<td>38</td>
<td>26</td>
<td>6</td>
<td>6</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td><strong>Revenu mensuel / uc (quartiles)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revenu&lt;4400 F</td>
<td>100</td>
<td>70</td>
<td>30</td>
<td>76</td>
<td>55</td>
<td>22</td>
<td></td>
<td>24</td>
<td>15</td>
<td>5</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4400 F&lt;revenu&lt;6500 F</td>
<td>100</td>
<td>74</td>
<td>26</td>
<td>80</td>
<td>61</td>
<td>19</td>
<td></td>
<td>21</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>6500 F&lt;revenu&lt;9330 F</td>
<td>100</td>
<td>79</td>
<td>21</td>
<td>78</td>
<td>64</td>
<td>14</td>
<td></td>
<td>22</td>
<td>15</td>
<td>3</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>9330 F&lt;revenu</td>
<td>100</td>
<td>76</td>
<td>24</td>
<td>78</td>
<td>61</td>
<td>17</td>
<td></td>
<td>22</td>
<td>16</td>
<td>3</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Total des ménages</td>
<td>100</td>
<td>75</td>
<td>25</td>
<td>78</td>
<td>60</td>
<td>18</td>
<td></td>
<td>22</td>
<td>15</td>
<td>3</td>
<td>4</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

*Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee*


<table>
<thead>
<tr>
<th>Trajectoire d’emploi entre nov. 93 et oct.94 (12 mois)</th>
<th>Ensemble</th>
<th>Pas de déménagement</th>
<th>Déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ensemble</td>
<td>Répondants V2, V3 et V4</td>
<td>Autres</td>
</tr>
<tr>
<td>Emploi tous les mois</td>
<td>100</td>
<td>79</td>
<td>21</td>
</tr>
<tr>
<td>tous les mois à temps complet</td>
<td>100</td>
<td>79</td>
<td>21</td>
</tr>
<tr>
<td>au moins un mois en temps partiel</td>
<td>100</td>
<td>82</td>
<td>18</td>
</tr>
<tr>
<td>au moins un mois en indépendant</td>
<td>100</td>
<td>76</td>
<td>25</td>
</tr>
<tr>
<td>Emploi et non emploi</td>
<td>100</td>
<td>69</td>
<td>31</td>
</tr>
<tr>
<td>emploi puis chômage</td>
<td>100</td>
<td>62</td>
<td>38</td>
</tr>
<tr>
<td>chômage puis emploi</td>
<td>100</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>alternances d’emploi et de chômage</td>
<td>100</td>
<td>68</td>
<td>32</td>
</tr>
<tr>
<td>activité et inactivité (+ chômage)</td>
<td>100</td>
<td>72</td>
<td>29</td>
</tr>
<tr>
<td>Jamais d’emploi</td>
<td>100</td>
<td>73</td>
<td>28</td>
</tr>
<tr>
<td>Ensemble (âge actif)</td>
<td>100</td>
<td>76</td>
<td>24</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997). Insee
Champ : individus panel de 17 ans à 60 ans répondants en 1994. Individus hors champ en vagues 2, 3 ou 4 exclus. Données pondérées (pondérations de la vague 1)
Bien que les 17-25 ans aient un fort taux de non-réponse en cas de déménagement (environ un tiers de non-répondants), suivre ceux qui déménagent permet d’obtenir des réponses sur quatre vagues pour 70 % de cette classe d’âge, alors que n’interroger que ceux qui n’ont pas déménagé ne permettrait d’obtenir cette information longitudinale que pour 48 % des 17-20 ans et 33 % des 21-25 ans (tableau 4), soit, en moyenne sur ces tranches d’âges, un échantillon deux fois plus faible qu’en cas de suivi lors de déménagements. L’intérêt (tout du moins en termes de taille d’échantillon) de suivre les individus qui déménagent reste important au sein de la classe d’âge immédiatement supérieure (26-30 ans), puisque, sur 76 % d’individus répondant aux quatre vagues, seuls 44 % n’ont pas déménagé. En revanche, au-delà de 30 ans, la proportion d’enquêtés qui pourraient être interrogés avec succès quatre vagues de suite en l’absence de suivi en cas de déménagement dépasse 60 %.

Mener des analyses de comportement sur la seule population des jeunes n’ayant pas déménagé implique de généraliser à l’ensemble des jeunes des comportements observés sur la moitié ou un tiers d’entre eux. Cela n’est bien sûr problématique que si les comportements des jeunes qui déménagent diffèrent de ceux qui ne déménagent pas. Or, pour les jeunes en particulier, on peut craindre l’existence de liens directs et indirects entre changements sur le marché du travail et changements de domicile (encadré 3). En effet, chez les moins de 25 ans, le déménagement correspond près de quatre fois sur dix au départ du foyer parental (tableau 5). Et les trajectoires des jeunes quittant leurs parents diffèrent sans doute fortement de celles des jeunes restant chez leurs parents, notamment en matière d’accès à l’emploi. Plusieurs études récentes ont abordé les liens entre les différentes étapes des jeunes vers l’autonomie. Ainsi, la moitié des jeunes des générations 1968 à 1971 ont quitté leurs parents avant 21,1 ans. Un quart les a quittés avant 19,2 ans, et un quart après 23,8 ans. L’âge médian de l’accès au premier emploi (21,2 ans) est très proche de l’âge médian de premier départ de chez les parents. L’accès à un emploi stable et la mise en couple sont eux plus tardifs (respectivement 23,3 ans et 23,4). Et l’âge d’accès au premier logement indépendant (non payé par les parents ou mis gratuitement à disposition par la famille) est intermédiaire (22,5 ans). Le nombre d’années médian entre premier emploi de plus de 6 mois et accès à un logement autonome est proche de zéro) (Galland, 2001).9

Tableau 5 - Type de déménagement chez les 25 ans et moins

<table>
<thead>
<tr>
<th>déménagement sans éclatement</th>
<th>répartition</th>
</tr>
</thead>
<tbody>
<tr>
<td>déménagement avec éclatement</td>
<td>45</td>
</tr>
<tr>
<td>séparation du conjoint</td>
<td>3</td>
</tr>
<tr>
<td>départ d'un enfant</td>
<td>39</td>
</tr>
<tr>
<td>départ d'une autre personne</td>
<td>3</td>
</tr>
<tr>
<td>total déménagement</td>
<td>100</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee
Champ : Individus panels de 17 à 25 ans ayant rempli un questionnaire individuel à la vague précédente et qui ont déménagé entre les deux vagues. Moyennes non pondérées sur les vagues 2 à 4.

Pour essayer de mettre en évidence des liens directs entre changements de logement et de situation sur le marché du travail, une première manière de faire consiste à comparer la situation d’emploi des jeunes qui déménagent avant et après leur déménagement. En fait, les déménagements s’accompagnant d’une modification quasi-simultanée de situation sur le marché de l’emploi restent peu fréquents (encadré 3). Ainsi, seuls 5 % des déménagements des 17-20 ans correspondent à une transition des études à l’emploi dans les mois entourant le déménagement. Dans 64 % des cas, les 17-20 ans sont en études deux mois avant comme deux mois après le déménagement, et dans 14 % ils sont en emploi avant comme après. Chez les 21-25 ans, la situation la plus fréquente est l’emploi avant et après le déménagement (deux tiers des déménagements). Dans cette tranche d’âge, seuls 7 % des déménagements s’accompagnent d’accès à l’emploi.

Ces résultats suggèrent que la plupart des déménagements ne seraient pas liés à l’accès à l’emploi, mais pourraient par exemple correspondre à des déménagements avec la famille (ce que montrait déjà l’analyse par type de déménagement). Toutefois, ces résultats ne tiennent compte que des liens *directs et quasi simultanés* entre accès à l’emploi et déménagement. Or, un étudiant peut commencer à travailler tout en résidant chez ses parents, et ne les quitter que 6 mois après, une fois son autonomie financière assurée. Lorsqu’il déménagera, il sera considéré ici comme passant de l’emploi à l’emploi, alors que l’on pourrait considérer que son déménagement comme son accès à l’emploi s’inscrivent dans un processus d’accès à l’autonomie (lien direct mais non simultané entre accès à l’emploi et déménagement) (cf. dossier « jeunes : l’âge des indépendances » *Economie et Statistique* 337-338, 2001). Et par ailleurs, ces résultats ne tiennent pas compte de liens *indirects* entre emploi et logement (encadré 3).

Pour tenir compte des liens indirects ou à plus long terme entre accès à l’emploi et déménagement, on peut comparer les taux d’emploi à trois ans (i.e. en septembre 1997) des jeunes initialement en études en septembre 1994 (tableau 6). La proportion de jeunes en emploi est beaucoup plus élevée chez ceux qui ont déménagé que chez les autres. L’erreur commise sur le taux d’accès à l’emploi en n’étudiant que les jeunes n’ayant pas déménagé serait de 6 points pour les 17-20 ans.
(19 % au lieu de 25 %), et de 19 points pour les 21-25 ans (35 % au lieu de 54 %). 
Et, selon la tranche d’âge considérée, ceux qui ont déménagé représentent la moitié 
or plus des deux tiers de ceux qui ont un emploi ! Ainsi, étudier les trajectoires 
d’accès à l’emploi des jeunes sur trois ou quatre ans sans disposer d’information sur 
ceux qui ont déménagé conduit au minimum à très fortement sous-estimer les taux 
d’accès à l’emploi, et, au pire, à des estimations de comportements dont on n’est pas 
assuré qu’elles puissent être généralisées à l’ensemble de la population.

Tableau 6 - Taux d’emploi à trois ans (septembre 1997) en fonction de la situation 
en septembre 1994 et de l’existence d’un déménagement

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>17-20 ans en étude</td>
<td>25</td>
<td>19 [après redressement 19]</td>
<td>36</td>
<td>49</td>
</tr>
<tr>
<td>21-25 ans en étude</td>
<td>54</td>
<td>34 [après redressement 35]</td>
<td>72</td>
<td>71</td>
</tr>
<tr>
<td>Chômeurs (tous âges)</td>
<td>49</td>
<td>44 [après redressement 47]</td>
<td>60</td>
<td>42</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee

Champ : individus panel de 17 ans ou plus ayant rempli un questionnaire individuel aux vagues 1 et 4. Résultats pondérés à l’aide d’une pondération longitudinale redressant de la non-réponse en vague 4, sauf le résultat « après redressement » qui est, lui, pondéré à l’aide d’une pondération redressant également de l’absence de suivi en cas de déménagement (cf. infra).

Lecture : Parmi les 17-20 ans en études en septembre 1994, 25 % occupent un emploi en septembre 1997. Cette proportion est de 19 % parmi ceux qui n’ont pas déménagé, et de 36 % parmi ceux qui ont déménagé, si bien que ces derniers représentent 49 % de ceux qui ont un emploi en septembre 1997.

Plus d’un chômeur sur trois déménage en trois ans, et 26 % d’entre eux ne sont pas retrouvés

Les personnes qui semblent les moins insérées sont également particulièrement mobiles, et souvent assez difficiles à ré-interroger en cas de déménagement. Comme dans le cas des jeunes, le lien entre déménagement et changement de situation sur le marché du travail peut être direct (par exemple, déménagement vers le lieu où l’on a trouvé un emploi) ou indirect (les moins insérés sur le marché de l’emploi sont aussi ceux dont le logement est le moins stable).
Ainsi, parmi les chômeurs d’octobre 1994, 37 % vont déménager dans les trois ans (contre 22 % pour l’ensemble de la population de 17 ans ou plus). À peine plus de la moitié de ces chômeurs ayant déménagé seront réinterrogés. En effet, les chômeurs constituent la catégorie dont il est le plus difficile de retrouver l’adresse en cas de déménagement (celle-ci n’est pas retrouvée une fois sur quatre). Les enquêtes d’âge actif (17-60 ans en première vague) qui déménagent ont des trajectoires d’activité initiales plus instables ou moins insérées que les enquêtés qui ne déménagent pas. Ainsi, le taux de déménagement en 3 ans (i.e. entre les vagues 1 et 4) est de 39 % parmi les personnes ayant connu l’emploi et le non-emploi lors de la première année, contre 26 % dans la population d’âge actif dans son ensemble.

Au sein des salariés actifs occupés, 39 % de ceux qui sont en CDD en début de période vont déménager dans les trois ans, contre 24 % pour les actifs employés en CDI. De manière générale, le taux de déménagement est plus élevé chez les salariés (25 %) que chez les indépendants (16 %). On retrouve ce résultat lorsque l’on analyse les taux de déménagement par catégorie socio-professionnelle : les agriculteurs sont beaucoup plus sédentaires que les autres actifs (11% seulement déménagent), les artisans également mais dans une moindre mesure (20%), tandis que les salariés ont des taux de déménagement plutôt supérieurs à la moyenne, variant de 24 % pour les ouvriers à 27 % pour les cadres.

Même si les écarts de taux de déménagement et de taux de réponse en fonction de la situation de la personne sur le marché de l’emploi sont moins marqués que les écarts en fonction de l’âge, ils restent forts10. De ce fait, les risques de biais liés à l’absence de suivi des personnes qui déménagent, moins importants que ceux liés à l’âge, restent notables. Ainsi, parmi les chômeurs de septembre 1994, 44 % de ceux qui n’ont pas déménagé ont un emploi trois ans plus tard, contre 49 % de l’ensemble. Si l’on cherche à estimer le taux d’emploi à trois ans de l’ensemble des chômeurs à partir des seuls chômeurs n’ayant pas déménagé (en les repondérant pour les rendre représentatifs de l’ensemble), on aboutit à 47 %, soit une valeur peu différente de la valeur estimée sur l’ensemble des chômeurs (49 %) (tableau 6). Mais il n’en reste pas moins que 42 % des chômeurs qui ont un emploi en fin de période ont déménagé pendant les trois années étudiées.

10 Des premiers résultats de régressions logistiques confirment que l’âge est effectivement un facteur explicatif de premier ordre, mais que, à âge donné, la situation sur le marché de l’emploi pourrait avoir un impact moins important que le laissaient suggérer les statistiques descriptives : ainsi, si les personnes les moins insérées sur le marché de l’emploi déménagent plus, c’est avant tout parce elles appartiennent aux âges où l’on franchit les différentes étapes menant à l’indépendance.
Ainsi, les résultats ci-dessus suggèrent

1. que les statistiques descriptives que l’on peut tirer d’échantillons de personnes n’ayant pas déménagé peuvent ne pas être trop éloignées des statistiques que l’on obtient sur des panels d’individus, si l’une des trois conditions suivantes est satisfaite :
   - soit les comportements de ceux qui ont déménagé et des autres sont proches,
   - soit les taux de déménagement pour la population concernée ne sont pas trop importants (population peu mobile, ou période d’observation courte),
   - soit le redressement par post-stratification permet de rapprocher ces deux estimations, ce qui n’est possible que si les différences entre ceux qui déménagent et les autres s’expliquent bien en fonction de caractéristiques observables prises en compte dans le calcul de la pondération ;

2. que, même si, grâce à l’usage d’une pondération adaptée, les statistiques descriptives sont peu faussées, se limiter aux personnes n’ayant pas déménagé réduit fortement l’échantillon d’étude quand les taux de déménagement sont élevés, et réduit encore plus fortement l’échantillon des personnes qui connaissent l’événement étudié si celui-ci est plus fréquent parmi ceux qui déménagent. Cela peut rendre délicate la généralisation à l’ensemble de la population de comportements observés sur ceux qui ne déménagent pas, et ce, d’autant plus que ceux qui déménagent diffèrent des autres, en particulier si ces différences portent sur des caractéristiques non observables (et donc non susceptibles d’être redressées). Cette étude traite peu des problèmes posés par les biais liés à des caractéristiques inobservables. Toutefois, si on considère une variable future comme inobservable (ou comme révélatrice d’une variable inobservable à la date courante), l’exemple des taux d’emploi à trois ans met en évidence des différences importantes parmi ceux qui déménagent et les autres.

Une autre manière d’apprécier l’impact potentiel de variables difficilement observables est d’étudier le taux de déménagement futur en réponse à des questions d’opinion. Un déménagement futur est d’autant plus probable que la personne est peu satisfaite de son logement : le taux de déménagement dans les

---

11 La période à laquelle il est fait référence ici est la période pendant laquelle la personne ne « doit » pas déménager. Cette période correspond donc à l’intervalle de temps entre la première et la dernière vague d’enquête, et non à la longueur du calendrier d’activité exploité.
12 Malheureusement, dans cet exemple, seul l’âge est contrôlé, et non les autres caractéristiques observables. D’autres analyses plus complètes des biais introduits par des caractéristiques non observables peuvent être envisagées, mais n’ont pas été menées ici faute de temps. Par exemple, on pourrait comparer dans une équation de salaire les résidus obtenus pour les individus qui vont déménager et pour les autres.

166

INSEE Méthodes
trois ans dépasse 50 % pour ceux qui ne sont « pas satisfaits du tout » de leur logement en début de période, et décroît jusqu’à 16 % pour les personnes qui sont « très satisfaits » de leur logement. Une analyse rapide « toutes choses égales par ailleurs » confirme que cet effet reste très significatif à âge, sexe, occupation et catégorie sociale donnés. Les autres questions d’opinion ont un effet plus modéré sur la probabilité de déménagement, et cet effet est souvent plus corrélé aux déterminants socio-économiques usuels (donc, peut être plus facilement corrigé). Ainsi, un tiers des personnes se déclarant « pas satisfaits du tout de leur travail ou situation principale » vont déménager dans les trois ans, contre environ un cinquième de celles qui s’en déclarent « assez satisfaites », « satisfaites » ou « très satisfaites ». Toutefois, à âge et situation sur le marché de l’emploi donnés, la satisfaction vis-à-vis de l’emploi n’apparaît plus liée de manière significative au déménagement. Il en est de même, pour les actifs occupés, de la satisfaction vis-à-vis de la distance à leur lieu de travail.

Le déménagement est plus fréquent aux bas niveaux de vie, dans les familles monoparentales et chez les locataires.

Compte tenu de l’importance du taux de déménagement et des taux de réponse pour déterminer les risques de biais et de diminution de la taille d’échantillon, il est important d’avoir une idée des variations de ces indicateurs en fonction des caractéristiques des personnes et de leurs ménages (tableaux 4). L’objet ici n’est pas de fournir un modèle explicatif de la probabilité de déménagement en fonction de caractéristiques individuelles. En effet, le déménagement dépend vraisemblablement largement d’événements familiaux (départ de chez les parents, mise en couple, arrivée d’un enfant, divorce…) qui ne sont pas évoqués dans ce texte. Le but est plutôt de fournir des indications sur le type de populations que l’on suit moins bien dans des panels de logements, afin de permettre au lecteur souhaitant réaliser une étude sur ce type de données d’apprécier les risques de biais pour les populations qui l’intéressent. Nous nous limitons donc ici au commentaire de statistiques descriptives (s’appuyant sur les caractéristiques des personnes et des ménages en début de période, c’est-à-dire en 1994).

Ce sont les individus appartenant à des ménages dont le revenu mensuel du ménage par unité de consommation se situe dans le 1er quartile (soit moins de 4400 F par mois) qui ont la plus forte probabilité de déménager dans les trois ans (24 %). Il semblerait que la probabilité de déménagement, ainsi que la difficulté à retrouver les personnes, soient encore plus importantes pour le premier décile de revenu. Les locataires sont également plus mobiles, et ils le sont d’autant plus que le propriétaire est un propriétaire privé (50 % de taux de déménagement). A l’opposé quand le propriétaire est un membre de la famille le taux de déménagement est nettement plus faible (23 %). Enfin, ce taux est de 30 % lorsque le propriétaire est un organisme public.

Les personnes appartenant à des familles monoparentales sont celles qui ont la probabilité la plus forte de déménager, soit 32 %, contre 26 % pour les personnes seules célibataires. Ces déménagements correspondent dans environ la moitié des cas
au départ d’un enfant, et dans l’autre moitié, à un déménagement de l’ensemble de la famille monoparentale. A l’inverse, ce sont les couples sans enfant qui sont les plus stables (taux de déménagement de 19 %).

C’est dans les unités urbaines de plus de 100 000 habitants ainsi que dans l’unité urbaine de Paris que l’on observe la plus forte proportion de déménagements (respectivement 27 % et 25 %). C’est également sur Paris que la part des non retrouvés est la plus forte (ils représentent un quart des cas de déménagement). A l’opposé, les communes rurales sont celles qui connaissent le moins de déménagements (14 %). Si on regarde les individus ayant déménagé suivant la Zone d’Etudes et d’Aménagement du Territoire (ZEAT) de résidence initiale, la Méditerranée, l’Ile de France puis la région Rhône-Alpes ont des taux de déménagement supérieur aux autres régions. C’est également dans les ZEAT à plus fort taux de déménagement qu’il est le plus difficile de retrouver les personnes qui déménagent (un quart d’entre elles ne sont pas retrouvées en Île-de-France ou en Méditerranée, contre un septième environ dans l’ensemble de la France).

En revanche, la nationalité ainsi que le sexe de l’individu n’ont pas d’effet significatif sur la probabilité de déménager.

1.3. Dans quelle mesure la pondération permet-elle de compenser la non-représentativité des individus suivis ?

Si suivre les personnes en cas de déménagement limite les risques de biais, cela ne résout pas tous les problèmes de « représentativité » de la population suivie par rapport à la population dans le champ. En effet, les personnes qui répondent à une vague donnée ne sont pas « représentatives » des personnes interrogées à la vague précédente, car l’attrition d’une vague sur l’autre est sélective : parmi les personnes répondant à une vague, les personnes les moins bien insérées sont plus susceptibles d’être non-répondantes à la vague suivante. Ainsi, 79 % des personnes interrogées lors de la première vague qui ont été en emploi toute l’année précédant la collecte répondent également aux trois vagues suivantes, tandis qu’elles ne sont que 62 % parmi celles qui ont connu l’emploi puis le chômage l’année précédant la première collecte (tableau 7c). Dans cette partie, nous nous intéressons à l’importance des biais liés à cette sélectivité de l’attrition, et examinons dans quelle mesure l’utilisation d’une pondération (longitudinale) adaptée permet de les réduire.

13 Notons que le calage (longitudinal) dans un panel se distingue d’un calage transversal. En gros, le principe du calage longitudinal est de rendre les répondants de la vague n+i représentatifs des répondants de la vague n. Ce calage peut donc s’appuyer sur les informations recueillies dans la vague n, et éventuellement sur des informations extérieures (structure de la population en n+i). A l’inverse, pour une enquête transversale, le calage ne peut s’appuyer que sur les quelques informations contenues dans la base de sondage (par exemple, le type de logement) et sur des informations extérieures (par exemple, la distribution par âge et sexe de la population à l’enquête Emploi). Les possibilités de redressement sont donc plus limitées dans une opération transversale.
Sont particulièrement sur-représentés parmi les personnes interrogées en première vague qui ne seront pas réinterrogées aux trois vagues suivantes (tableaux 7) :

- les sans-diplômes (29 % de ceux qui ne seront pas réinterrogés, contre 25 % de ceux qui seront réinterrogés),

- les chômeurs (11 % contre 6 %), et, dans une moindre mesure, les retraités (23 % contre 21 %),

- les 21-25 ans (12 % contre 9 %) et les plus de 61 ans (26 % contre 22 %),

- parmi les actifs, les artisans (10 % contre 7 %), ouvriers (27 % contre 24 %) et les employés (31 % contre 29 %),

- parmi les salariés, ceux qui sont en Contrat à Durée Déterminée (13 % contre 10 %),

- les personnes seules (16 % contre 13 %) et les couples sans enfant (28 % contre 26 %),

- les habitants d’Île de France (23 % contre 18 %),

- les locataires (42 % contre 38 %),

- les 25 % les plus pauvres (30 % contre 23 %),

- les personnes ayant alterné entre emploi et non-emploi au cours des douze derniers mois (17 % contre 12 %), et celles n’ayant pas eu d’emploi (33 % contre 27 %).

Le biais sur les statistiques descriptive (non pondérées) dû à l’attrition dépend, avant pondération, de deux facteurs : de la part des personnes réinterrogées et de l’importance des différences entre les personnes réinterrogées et non-réinterrogées. A condition d’essayer de suivre les personnes qui déménagent et de s’intéresser à des populations pas trop spécifiques, les personnes non réinterrogées restent en général minoritaires (elles représentent 24 % de l’ensemble des enquêtés d’âge actif dans le cas d’un suivi sur 4 vagues). De plus, les différences (observables) entre personnes réinterrogées et personnes non réinterrogées restent encore modérées dans le cas d’un suivi sur quatre ans (cf. ci dessous). Au total, ces différences sont suffisamment peu marquées en population générale pour que la sélectivité de l’attrition ne se traduise pas par des biais importants sur les statistiques descriptive. Ainsi, sur l’exemple de la trajectoire sur 12 mois avant la première collecte des personnes d’âge actif, 61 % de celles qui seront réinterrogées ont été en emploi tous les mois, contre 51 % de celles que ne seront pas réinterrogées. Comme les secondes représentent 24 % du champ et les premières 76 %, la proportion de personnes en
emploi sur douze mois sur l’ensemble de la population est de 58 %\textsuperscript{14}, valeur peu éloignée du 61 % mesuré sur les seules personnes réinterrogées.

Mais bien sûr, le biais lié à l’attrition pourrait devenir important dans le cadre d’une étude centrée sur une sous-population mal suivie (par exemple les chômeurs) pour des variables dont la moyenne serait très différente sur l’échantillon suivi et l’échantillon non suivi. Par ailleurs, en général, les caractéristiques associées à une plus forte probabilité de déménagement sont également celles qui se traduisent par une moindre probabilité de réinterrogation en l’absence de déménagement. L’absence de suivi en cas de déménagement accentue donc les biais inhérents à l’utilisation d’un panel.

Le recours à une pondération adaptée (pondération longitudinale, encadré 3) permet de corriger en partie les biais liés à l’attrition, tout du moins pour la part de ces biais qui s’explique en fonction de caractéristiques observées dans les vagues d’enquêtes précédentes\textsuperscript{15}. Ainsi, l’estimation de la proportion de personnes toujours en emploi l’année précédant la première vague effectuée à partir des répondants aux quatre vagues (60 %) se rapproche de la valeur obtenue sur les répondants de première vague (61 %) (tableau 8). Ce résultat n’est pas mécanique, dans la mesure où la procédure de redressement longitudinal n’utilise pas d’information sur la situation des individus par rapport à l’emploi pour le calage.

\textsuperscript{14} soit 76\%*61\%+24\%*51\%.

\textsuperscript{15} ou éventuellement uniquement dans la vague courante si l’on effectue un calage sur des marges connues à l’aide d’autres enquêtes.
**Tableau 7 : Représentation de différents sous-échantillons**


<table>
<thead>
<tr>
<th>Caractéristiques individuelles</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Non retrouvés</th>
<th>Autres non réponses</th>
<th>Parti des non-retrouvés dans les non-réponses</th>
<th>Taux de déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Niveau d'études atteint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aucun diplôme</td>
<td>26</td>
<td>25</td>
<td>29</td>
<td>30</td>
<td>29</td>
<td>35</td>
<td>11</td>
<td>10</td>
<td>14</td>
<td>15</td>
<td>48</td>
<td>9</td>
</tr>
<tr>
<td>CAP, BEP</td>
<td>20</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td>27</td>
<td>21</td>
<td>55</td>
<td>22</td>
</tr>
<tr>
<td>BEPC</td>
<td>26</td>
<td>26</td>
<td>25</td>
<td>26</td>
<td>26</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>25</td>
<td>51</td>
<td>23</td>
</tr>
<tr>
<td>Baccalauréat</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>10</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td>Baccalauréat plus 2 années</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>18</td>
<td>19</td>
<td>15</td>
<td>17</td>
<td>45</td>
<td>34</td>
</tr>
<tr>
<td>Supérieur à bac plus 2</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>16</td>
<td>18</td>
<td>12</td>
<td>12</td>
<td>47</td>
<td>36</td>
</tr>
<tr>
<td>Total des individus</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>49</td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actif occupé</td>
<td>51</td>
<td>53</td>
<td>45</td>
<td>49</td>
<td>51</td>
<td>43</td>
<td>56</td>
<td>59</td>
<td>48</td>
<td>52</td>
<td>47</td>
<td>34</td>
</tr>
<tr>
<td>Chômeur</td>
<td>8</td>
<td>6</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>9</td>
<td>13</td>
<td>11</td>
<td>21</td>
<td>12</td>
<td>62</td>
<td>37</td>
</tr>
<tr>
<td>Étudiant</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>50</td>
<td>41</td>
</tr>
<tr>
<td>Retraité</td>
<td>22</td>
<td>21</td>
<td>23</td>
<td>26</td>
<td>25</td>
<td>29</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>11</td>
<td>32</td>
<td>7</td>
</tr>
<tr>
<td>Autre inactif</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>Total des individus</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>49</td>
</tr>
<tr>
<td><strong>Âge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-20 ans</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>12</td>
<td>11</td>
<td>13</td>
<td>14</td>
<td>47</td>
<td>36</td>
</tr>
<tr>
<td>21-25 ans</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>23</td>
<td>23</td>
<td>26</td>
<td>22</td>
<td>53</td>
<td>55</td>
</tr>
<tr>
<td>26-30 ans</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>19</td>
<td>21</td>
<td>19</td>
<td>13</td>
<td>59</td>
<td>43</td>
</tr>
<tr>
<td>31-40 ans</td>
<td>19</td>
<td>20</td>
<td>16</td>
<td>19</td>
<td>20</td>
<td>14</td>
<td>22</td>
<td>22</td>
<td>21</td>
<td>20</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>41-50 ans</td>
<td>19</td>
<td>20</td>
<td>17</td>
<td>21</td>
<td>22</td>
<td>19</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>45</td>
<td>11</td>
</tr>
<tr>
<td>51-60 ans</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>15</td>
<td>14</td>
<td>15</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>38</td>
<td>12</td>
</tr>
<tr>
<td>Plus de 61 ans</td>
<td>23</td>
<td>22</td>
<td>26</td>
<td>27</td>
<td>26</td>
<td>32</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>12</td>
<td>32</td>
<td>7</td>
</tr>
<tr>
<td>Total des individus</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>49</td>
</tr>
</tbody>
</table>
### Tableau 7a (suite) : en fonction de caractéristiques observables des personnes en 1994.

<table>
<thead>
<tr>
<th>Caractéristiques individuelles</th>
<th>Ensemble</th>
<th>Enseignement</th>
<th>Enseignement</th>
<th>Enseignement</th>
<th>Enseignement</th>
<th>Enseignement</th>
<th>Enseignement</th>
<th>Enseignement</th>
<th>Enseignement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statut des actifs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salaire</td>
<td>88</td>
<td>86</td>
<td>87</td>
<td>85</td>
<td>92</td>
<td>89</td>
<td>90</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Indépendant</td>
<td>13</td>
<td>12</td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Total des actifs</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>PCS des actifs</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Agriculteur</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Cadre supérieur</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>14</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Profession intermédiaire</td>
<td>23</td>
<td>23</td>
<td>18</td>
<td>23</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Ouvrier</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>31</td>
<td>31</td>
<td>30</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Total des salariés</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Type de contrat de travail des salariés</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDD</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>CDI</td>
<td>90</td>
<td>90</td>
<td>87</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>84</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>Total des salariés</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Source : Panel européen, vagues 1 à 4 (1994-1997), INSEE.

<table>
<thead>
<tr>
<th>Caractéristiques des ménages</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Autres</th>
<th>Ensemble</th>
<th>Répondants V2, V3 et V4</th>
<th>Non retrouvés</th>
<th>Autres non-réponses</th>
<th>Part des non-retrouvés dans les non-réponses</th>
<th>Taux de déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type de ménage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personne seule</td>
<td>14</td>
<td>13</td>
<td>16</td>
<td>13</td>
<td>13</td>
<td>15</td>
<td>16</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>19</td>
<td>45</td>
</tr>
<tr>
<td>Couple sans enfant</td>
<td>26</td>
<td>26</td>
<td>28</td>
<td>28</td>
<td>27</td>
<td>31</td>
<td>23</td>
<td>23</td>
<td>25</td>
<td>19</td>
<td>19</td>
<td>55</td>
</tr>
<tr>
<td>Couple avec enfant</td>
<td>42</td>
<td>44</td>
<td>37</td>
<td>42</td>
<td>44</td>
<td>36</td>
<td>43</td>
<td>46</td>
<td>38</td>
<td>40</td>
<td>47</td>
<td>23</td>
</tr>
<tr>
<td>Famille monoparentale</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>41</td>
<td>32</td>
</tr>
<tr>
<td>Autre type de ménage</td>
<td>14</td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>15</td>
<td>13</td>
<td>12</td>
<td>16</td>
<td>14</td>
<td>51</td>
<td>20</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>ZEAT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Île de France</td>
<td>19</td>
<td>18</td>
<td>23</td>
<td>19</td>
<td>18</td>
<td>22</td>
<td>22</td>
<td>20</td>
<td>32</td>
<td>20</td>
<td>60</td>
<td>25</td>
</tr>
<tr>
<td>Bassin parisien</td>
<td>17</td>
<td>17</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td>14</td>
<td>48</td>
<td>26</td>
</tr>
<tr>
<td>Nord</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Est</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>42</td>
<td>20</td>
</tr>
<tr>
<td>Ouest</td>
<td>13</td>
<td>14</td>
<td>12</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>Sud-ouest</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>5</td>
<td>11</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Centre-est</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>7</td>
<td>14</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>Méditerranée</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>15</td>
<td>12</td>
<td>24</td>
<td>15</td>
<td>60</td>
<td>26</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Strate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commune rurale</td>
<td>27</td>
<td>28</td>
<td>23</td>
<td>29</td>
<td>30</td>
<td>26</td>
<td>17</td>
<td>18</td>
<td>14</td>
<td>18</td>
<td>42</td>
<td>14</td>
</tr>
<tr>
<td>Unité urbaine &lt; 20000 hab</td>
<td>16</td>
<td>17</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>13</td>
<td>18</td>
<td>40</td>
<td>22</td>
</tr>
<tr>
<td>20000&lt;=u.u.&lt;=100000 hab</td>
<td>13</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>12</td>
<td>15</td>
<td>14</td>
<td>15</td>
<td>10</td>
<td>17</td>
<td>35</td>
<td>24</td>
</tr>
<tr>
<td>u.u. de plus de 100000 hab</td>
<td>27</td>
<td>27</td>
<td>28</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>33</td>
<td>33</td>
<td>35</td>
<td>30</td>
<td>53</td>
<td>27</td>
</tr>
<tr>
<td>Commune de l'unité urbaine de Paris</td>
<td>17</td>
<td>16</td>
<td>19</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>19</td>
<td>17</td>
<td>28</td>
<td>17</td>
<td>61</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Caractéristiques des ménages</td>
<td>Ensemble</td>
<td>Répondants V2, V3 et V4</td>
<td>Autres</td>
<td>Ensemble</td>
<td>Répondants V2, V3 et V4</td>
<td>Autres</td>
<td>Ensemble</td>
<td>Répondants V2, V3 et V4</td>
<td>Non retrouvés</td>
<td>Autres non-réponses</td>
<td>Part des non-retrouvés dans les non-réponses</td>
<td>Taux de déménagement</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>--------------------------</td>
<td>-------</td>
<td>----------</td>
<td>--------------------------</td>
<td>-------</td>
<td>----------</td>
<td>--------------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Statut d’occupation du logement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locataire (yc, sous locataire et logé gratuitement)</td>
<td>39</td>
<td>38</td>
<td>42</td>
<td>30</td>
<td>30</td>
<td>33</td>
<td>69</td>
<td>70</td>
<td>70</td>
<td>64</td>
<td>51</td>
<td>39</td>
</tr>
<tr>
<td>Propriétaire</td>
<td>61</td>
<td>63</td>
<td>58</td>
<td>70</td>
<td>71</td>
<td>67</td>
<td>31</td>
<td>30</td>
<td>30</td>
<td>36</td>
<td>44</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>49</td>
</tr>
<tr>
<td>Propriétaire (pour locataires)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Famille</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>17</td>
<td>17</td>
<td>16</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>10</td>
<td>39</td>
<td>23</td>
</tr>
<tr>
<td>Propriétaire privé</td>
<td>46</td>
<td>46</td>
<td>47</td>
<td>37</td>
<td>38</td>
<td>37</td>
<td>60</td>
<td>60</td>
<td>67</td>
<td>55</td>
<td>56</td>
<td>50</td>
</tr>
<tr>
<td>HLM ou organisme public</td>
<td>41</td>
<td>41</td>
<td>40</td>
<td>46</td>
<td>46</td>
<td>47</td>
<td>32</td>
<td>32</td>
<td>27</td>
<td>36</td>
<td>43</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>49</td>
</tr>
<tr>
<td>Revenu mensuel / uc (quartiles)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revenu&lt;4400 F</td>
<td>25</td>
<td>23</td>
<td>30</td>
<td>24</td>
<td>23</td>
<td>30</td>
<td>27</td>
<td>25</td>
<td>37</td>
<td>24</td>
<td>59</td>
<td>24</td>
</tr>
<tr>
<td>4400 F&lt;revenu&lt;6500 F</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>26</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>25</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>6500 F&lt;revenu&lt;9330 F</td>
<td>25</td>
<td>27</td>
<td>21</td>
<td>25</td>
<td>27</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td>22</td>
<td>24</td>
<td>46</td>
<td>22</td>
</tr>
<tr>
<td>9330 F&lt;revenu</td>
<td>25</td>
<td>26</td>
<td>24</td>
<td>25</td>
<td>25</td>
<td>24</td>
<td>25</td>
<td>27</td>
<td>18</td>
<td>27</td>
<td>39</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>49</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee
Tableau 7c : en fonction des trajectoires d'emploi sur les douze derniers mois avant la première collecte.

<table>
<thead>
<tr>
<th>Trajectoire d'emploi entre nov. 93 et oct. 94 (12 mois)</th>
<th>Ensemble</th>
<th>Pas de déménagement</th>
<th>Déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ensemble</td>
<td>Répondeurs V2, V3 et V4</td>
<td>Autres</td>
</tr>
<tr>
<td>Emploi tous les mois</td>
<td>58</td>
<td>61</td>
<td>51</td>
</tr>
<tr>
<td>tous les mois à temps complet</td>
<td>44</td>
<td>46</td>
<td>38</td>
</tr>
<tr>
<td>au moins un mois en temps partiel</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>au moins un mois en indépendant</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Emploi et non emploi</td>
<td>13</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>emploi puis chômage</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>chômage puis emploi</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>alternance d'emploi et de chômage</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>activité et inactivité (+ chômage)</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Jamais d'emploi</td>
<td>29</td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td>Ensemble (âge actif)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee

Champ : individus panel de 17 ans à 60 ans répondants en 1994. Individus hors champ en vagues 2, 3 ou 4 exclus. Données pondérées (pondérations de la vague 1)
Tableau 8 : Biais lié à la sélectivité de l’attrition et effet correcteur du redressement — exemple de la proportion d’enquêtés ayant toujours été en emploi et de la proportion d’enquêtés ayant connu l’emploi et le non-emploi.

8a : Proportion d’enquêtés ayant toujours été en emploi entre novembre 93 et octobre 94.

<table>
<thead>
<tr>
<th>Echantillon</th>
<th>Type de pondération</th>
<th>Proportion / Biais</th>
<th>Ensemble</th>
<th>Pas de déménagement</th>
<th>Déménagement entre V1 et V4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Répondants de 1994</td>
<td>Transversale 1994</td>
<td>Proportion</td>
<td>58,4%</td>
<td>61,4%</td>
<td>49,5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biais</td>
<td>2,4%</td>
<td>1,8%</td>
<td></td>
</tr>
<tr>
<td>Répondants de 1994 à 1997</td>
<td>Transversale 1994</td>
<td>Proportion</td>
<td>60,8%</td>
<td>63,2%</td>
<td>52,9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biais</td>
<td>1,8%</td>
<td>1,6%</td>
<td>1,5%</td>
</tr>
<tr>
<td>Longitudinale</td>
<td>Proportion</td>
<td>59,5%</td>
<td>63,0%</td>
<td>51,0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biais</td>
<td>1,5%</td>
<td>1,6%</td>
<td>1,5%</td>
<td></td>
</tr>
<tr>
<td>Répondants de 1994 et 1997</td>
<td>Longitudinale</td>
<td>Proportion</td>
<td>59,3%</td>
<td>62,1%</td>
<td>50,8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biais</td>
<td>0,9%</td>
<td>0,7%</td>
<td>1,3%</td>
</tr>
<tr>
<td>Répondants de 1994 et 1997 n’ayant pas déménagé</td>
<td>Longitudinale</td>
<td>Proportion</td>
<td>60,5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biais</td>
<td>2,1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee
Champ : individus panel de 17 ans à 60 ans répondants en 1994. Individus ayant rempli un questionnaire individuel en vague 1 et qui ne sont pas hors champ aux vagues 2, 3 et 4. Données pondérées.
8b : Proportion d'enquêtés ayant connu l'emploi et le non-emploi entre novembre 93 et octobre 94.

<table>
<thead>
<tr>
<th>Echantillon</th>
<th>Type de pondération</th>
<th>Proportion / Biais</th>
<th>Ensemble</th>
<th>Pas de déménagement</th>
<th>Déménagement entre V1 et V4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Répondants de 1994</td>
<td>Transversale 1994</td>
<td>Proportion</td>
<td>13,0%</td>
<td>10,7%</td>
<td>19,4%</td>
</tr>
<tr>
<td>Répondants de 1994 à 1997</td>
<td>Transversale 1994</td>
<td>Proportion</td>
<td>11,7%</td>
<td>9,9%</td>
<td>17,8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biais</td>
<td>-1,3%</td>
<td>-0,8%</td>
<td>-1,6%</td>
</tr>
<tr>
<td>Longitudinale</td>
<td>Proportion</td>
<td>12,2%</td>
<td>10,0%</td>
<td>18,7%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biais</td>
<td>-0,8%</td>
<td>-0,7%</td>
<td>-0,7%</td>
<td></td>
</tr>
<tr>
<td>Répondants de 1994 et 1997</td>
<td>Longitudinale</td>
<td>Proportion</td>
<td>12,4%</td>
<td>10,3%</td>
<td>18,7%</td>
</tr>
<tr>
<td></td>
<td>Biais</td>
<td>-0,6%</td>
<td>-0,4%</td>
<td>-0,7%</td>
<td></td>
</tr>
<tr>
<td>Répondants de 1994 et 1997</td>
<td>Longitudinale</td>
<td>Proportion</td>
<td>11,4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n'ayant pas déménagé</td>
<td></td>
<td>Biais</td>
<td>-1,6%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee
Champ : individus panel de 17 ans à 60 ans répondants en 1994. Individus ayant rempli un questionnaire individuel en vague et qui ne sont pas hors champ aux vagues 2, 3 et 4. Données pondérées.
1.4 La pondération permettrait-elle de compenser l’absence de suivi des individus qui déménagent ?

Sur quatre vagues, les biais liés à l’attrition semblent modérés, et peuvent être réduits grâce à une pondération adaptée. Bien que les biais liés à l’absence de suivi en cas de déménagement soient, eux, plus importants, on peut se demander s’il ne serait pas également possible de les corriger par re-pondération. C’est l’exercice auquel nous nous livrons maintenant. Pour ce, les individus qui ont déménagé entre les vagues 1 à 4 sont considérés comme non-répondants, ce qui porte le taux de non-réponse à 38 % (non-réponse « au sein du même logement »), contre 22 % si ceux qui déménagent ne sont pas considérés comme non répondants. Considérer ceux qui ont déménagé comme non-répondants divise donc le taux de réponse par 1,2. Les poids des « répondants au sein du même logement » de la vague 4 doivent donc en moyenne être multipliés par 1,2 par rapport aux poids obtenus lorsque les personnes qui déménagent sont suivies, et par 1,6 par rapport aux poids de première vague.

En fait, les taux de « non-réponse au sein du même logement » ne sont pas homogènes dans la population. Selon les caractéristiques des personnes, ils s’échelonnent entre 15 et 82 %. Parmi les 55 catégories homogènes vis-à-vis de la non-réponse au sein du même logement qui ont été définies (encadré 3), quatre ont un taux de non-réponse au sein du même logement égal ou supérieur à 80 %, et 20 ont un taux de non-réponse qui dépasse 50 %. Trois des quatre catégories dont le taux de non-réponse est le plus élevé correspondent à des individus de 25 ans à 35 ans. Comme les taux de non-réponse au sein du même logement sont très dispersés, le facteur correctif appliqué aux pondérations est très variable d’une catégorie à l’autre : les poids initiaux sont multipliés au moins par 1,2 et au plus par 5,6 (par 1,09 au moins et 2,2 lorsque ceux qui déménagent sont suivis). Les poids obtenus sont donc beaucoup plus dispersés que lorsque ceux qui déménagent sont suivis (tableau 9) : ils varient dans un rapport de 1 à 22 (contre un rapport de 1 à 9 en cas de suivi lors des déménagements). Un seul individu peut alors fortement peser dans les résultats, ce qui peut poser des problèmes s’il est atypique (ou si les informations le concernant sont affectées d’erreurs de mesure).

---

16 Taux de non-réponse calculés sur les enquêtes de première vague de 17 ans et plus qui ne sont pas hors champ lors des vagues suivantes.

INSEE Méthodes
Tableau 9 : Caractéristiques des poids longitudinaux en vague 4 des individus panel

<table>
<thead>
<tr>
<th></th>
<th>Ponderation valable pour l'ensemble des répondants</th>
<th>Ponderation valable pour les répondants n'ayant pas déménagé</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moyenne</td>
<td>3 961</td>
<td>4 945</td>
</tr>
<tr>
<td>Ecart type</td>
<td>1 326</td>
<td>2 243</td>
</tr>
<tr>
<td>D1</td>
<td>2 824</td>
<td>3 152</td>
</tr>
<tr>
<td>Q1</td>
<td>3 176</td>
<td>3 564</td>
</tr>
<tr>
<td>Médiane</td>
<td>3 693</td>
<td>4 259</td>
</tr>
<tr>
<td>Q3</td>
<td>4 402</td>
<td>5 526</td>
</tr>
<tr>
<td>D9</td>
<td>5 226</td>
<td>7 363</td>
</tr>
<tr>
<td>C95</td>
<td>5 931</td>
<td>9 020</td>
</tr>
<tr>
<td>C99</td>
<td>9 892</td>
<td>13 348</td>
</tr>
<tr>
<td>Minimum</td>
<td>2 195</td>
<td>2 265</td>
</tr>
<tr>
<td>Maximum</td>
<td>19 485</td>
<td>49 200</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee
Champ : individus panel de 17 ans et plus en 1997 ayant rempli un questionnaire individuel en vague 4.
Lecture : dans le cas où les individus sont suivis lorsqu'ils déménagent, un individu panel remplissant un questionnaire individuel en vague 4 représente en moyenne 3 961 personnes dans la réalité. En l'absence de suivi en cas de déménagement, un individu panel « répondant au sein du même logement » en vague 4 représente 4 945 personnes dans la réalité, soit 1,2 fois plus.

C'est pour les 21-30 ans que l'absence de suivi en cas de déménagement aurait l'impact le plus fort sur les poids. En effet, dans ce groupe d'âge, l'absence de suivi en cas de déménagement conduirait en moyenne à presque doubler les poids par rapport aux poids calculés en cas de suivi. Pour les 17-20 ans, l'incidence du suivi sur les poids resterait assez importante (poids 1,5 fois plus élevés que les poids calculés en cas de suivi). Elle serait en revanche relativement faible pour les plus de 30 ans (multiplication par 1,1 à 1,3).

S'il est en théorie possible de redresser les données pour corriger de l'absence de suivi des personnes qui déménagent (ainsi que de l'attrition), cela ne permet évidemment pas de compenser les pertes d'effectifs dans certaines catégories : les poids calculés sont très dispersés, et peuvent être assez élevés. Par ailleurs, une telle pondération, basée sur le redressement suivant les caractéristiques socio-démographiques usuelles, ne redresse qu'imparfaitement les biais. Ainsi, le taux de personnes qui ont été continûment en emploi l'année précédant la première vague est estimé après redressement à 60,5 % (à partir de l'échantillon des répondants au sein du ménage), soit une valeur plus proche de la valeur non-redressée (60,8 %) que de la vraie valeur (58,4 %) (tableau 8). L'estimation de la proportion de personnes qui ont connu l'emploi et le non-emploi souffre des mêmes défauts (estimation après redressement de 11,4 %, plus proche de la valeur non redressée (11,7 %) que de la vraie valeur (13 %)). Le redressement, s'il permet de corriger des biais s'expliquant bien en fonction des caractéristiques observables, et notamment les « effets de structure », atteint ses limites lorsqu'il s'agit de tenir compte de biais qui ne sont pas fortement corrélés aux caractéristiques socio-économiques usuelles.

INSEE Méthodes 179
2. Quel est le prix à payer pour suivre les personnes qui déménagent ?

2.1 Un coût par interview réalisé plus élevé

Un panel d'individus est plus coûteux qu'un panel de logements. En effet, en premier lieu, un effort supplémentaire doit être fait pour suivre les individus, et éventuellement pour assurer le transfert de leur dossier d'enquête à une direction régionale de l’Insee et / ou à un autre enquêteur. L’interrogation des personnes retrouvées peut également s’avérer plus coûteuse. C’est le cas par exemple si elles ont déménagé dans une zone hors échantillon maître de l’Insee (d’où des coûts de transports plus élevés pour les enquêteurs ainsi que des pertes de temps et des coûts de repérage des nouvelles adresses). Par ailleurs, il est plus fréquent que les personnes qui déménagent ne soient pas enquêtées (soit parce qu’elles ne sont pas retrouvées, soit parce qu’elles sont en fait hors champ, décédées...), d’où un coût moyen par interview réalisé plus élevé. Enfin, les panels d’individus imposent en général une charge plus lourde aux enquêteurs, car ceux-ci sont en moyenne suivis plus longtemps que dans un panel de logements (parce que le nombre de vagues est plus élevé ou que les individus ne sortent pas de l’enquête par déménagement). Cette charge implique des efforts particuliers pour s’assurer la coopération des enquêtés (cadres, courriers, diffusion de résultats...). Elle pourrait également se traduire par des refus supplémentaires (entraînant des coûts supérieurs par interview réalisée), sauf si le fait que le même enquêteur retrouve le même enquêté d’une année sur l’autre facilite la coopération de ce dernier ou si le fait que certaines variables (état civil) ne soient pas redemandées à chaque fois permet d’alléger notablement la charge d’enquête (Caron et Ravalet, 200017).

A budget donné, l’échantillon d’un panel d’individus sera donc a priori de taille plus réduite que celui d’un panel de logements. 7 344 ménages ont répondu à la première vague du panel européen, et des questionnaires individuels ont été remplis pour environ 14 000 individus de 17 ans et plus. Est-il alors possible d’estimer valablement sur la base d’un échantillon de cette taille la fréquence d’événements relativement rares, comme les transitions sur le marché de l’emploi ? C’est la question à laquelle nous nous intéressons maintenant.

2.2 Construction des calendriers d’activité à partir du panel, et qualité des données

Toutefois, avant de pouvoir comparer les résultats tirés du calendrier d’activité du panel et ceux que l’on observe sur les enquêtes Emploi (panel de logements de taille plus importante), il convient de traiter certains problèmes spécifiques au calendrier du panel, comme les incohérences que l’on observe sur les recouvrements de calendrier entre les différentes vagues ou les cas où plusieurs occupations peuvent être déclarées pour un même mois (encadré 4).

Si l’impact de ces problèmes est faible si l’on s’intéresse à une situation un mois donné sur le marché du travail, ils peuvent être plus importants si l’on s’intéresse aux transitions. En effet, les transitions restent peu fréquentes un mois donné : à peine plus de 1 % par mois (cf. ci-dessous), ce qui représente dans le panel moins de 200 observations par mois. Ainsi, un problème de déclaration, de définition ou de qualité des données qui conduirait à générer 20 transitions supplémentaires par mois augmenterait de 10 % la probabilité de transition.

Compte tenu de l’importance potentielle de ces problèmes pour l’analyse des transitions, il nous a semblé important de commencer notre travail par une évaluation de la qualité des données, et éventuellement une correction de ces dernières. L’impact des corrections menées dépend des variables étudiées. Ainsi :

- L’apurement a conduit à modifier moins de 1 % des mois renseignés, mais a surtout conduit à renseigner environ 3,5 % des mois en plus (notamment les mois pour lesquels seule une activité « non principale » (activité secondaire, maladie, chômage partiel, formation...) était renseignée, ou les mois correspondant à des trous de calendrier). Au total, 10 % des individus sont concernés.

- Le choix des règles de définition de la situation principale du mois introduit un léger biais en faveur de l’activité, puisqu’une transition vers le chômage inférieure à un mois n’est pas prise en compte. Si l’on avait choisi de privilégier le chômage sur l’emploi puis sur l’inactivité, seuls 0,3 % des mois auraient été modifiés, mais 5,4 % des enquêtés de 17 ans et plus auraient été concernés. Seul un mois aurait été modifié pour la moitié de ces derniers (2,8 %). Pour 2,3 % des individus (soit 373 individus), ce choix aurait entraîné une modification du nombre de transitions (celui-ci aurait varié d’une unité dans 0,8 % des cas, de deux unités dans 1,1 % des cas). La séquence des transitions (suite des activités exercées) est également modifiée dans ce cas. Toutefois, cela n’implique une modification de la trajectoire sur 48 mois (selon la typologie du tableau 10) que dans 0,8 % des cas (soit 80 individus parmi ceux qui sont présents sur les quatre vagues) : si l’on privilégie le chômage sur l’emploi, plus de la moitié de ces individus ne sont plus continûment en emploi sur 48 mois.

INSEE Méthodes 181
Tableau 10 : Incidence de la règle de définition de l'activité principale et de l'apurement sur la définition des trajectoires sur 48 mois

<table>
<thead>
<tr>
<th>Règle de priorité pour déterminer l'activité principale :</th>
<th>Emploi domine chômage qui domine inactivité OUI</th>
<th>Chômage domine emploi qui domine inactivité</th>
<th>Emploi domine chômage qui domine inactivité NON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prise en compte des corrections ponctuelles :</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emploi tous les mois</td>
<td>52,8</td>
<td>52,3</td>
<td>57</td>
</tr>
<tr>
<td>tous les mois à temps complet</td>
<td>38,6</td>
<td>38,2</td>
<td>42</td>
</tr>
<tr>
<td>au moins un mois en temps partiel</td>
<td>7,2</td>
<td>7,1</td>
<td>8</td>
</tr>
<tr>
<td>au moins un mois en indépendant</td>
<td>7,1</td>
<td>7,0</td>
<td>8</td>
</tr>
<tr>
<td>Emploi et non emploi</td>
<td>30,5</td>
<td>30,8</td>
<td>31</td>
</tr>
<tr>
<td>emploi puis chômage</td>
<td>1,8</td>
<td>2,0</td>
<td>2</td>
</tr>
<tr>
<td>chômage puis emploi</td>
<td>1,1</td>
<td>1,2</td>
<td>1</td>
</tr>
<tr>
<td>alternances d'emploi et de chômage</td>
<td>10,1</td>
<td>10,2</td>
<td>11</td>
</tr>
<tr>
<td>activité et inactivité (+ chômage)</td>
<td>17,5</td>
<td>17,3</td>
<td>17</td>
</tr>
<tr>
<td>Jamais d'emploi</td>
<td>16,6</td>
<td>16,9</td>
<td>11</td>
</tr>
<tr>
<td>Trajectoire à trous</td>
<td>0,0</td>
<td>0,0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee
Au total, l’absence de hiérarchisation entre les différentes occupations possibles (les situations d’emploi étant placées sur le même plan que la maladie, la formation, ou le chômage partiel) comme la variété des choix offerts - mais pas toujours compris - se sont traduits par des calendriers parfois difficilement exploitables sans travail supplémentaire, car ne permettant pas de déterminer facilement si, un mois donné, la personne a été active occupée, chômeuse ou inactive. Ces problèmes devraient en partie disparaître à partir de la vague 5, grâce à la généralisation de la Collecte Assistée Par Informatique (qui réduira par ailleurs les erreurs de saisie). En effet, dans le questionnaire informatique, l’enquêté doit décrire les différentes périodes consécutives qu’il a connues sur le marché de l’emploi. Ces différentes périodes ne portent que sur les principales situations proposées sur les questionnaires papier (et excluent notamment la maladie, les autres formations...). Ce nouveau mode d’interrogation assure qu’aucun mois ne puisse être vide (toute période doit débuter à la fin de la précédente) et permet de déterminer sans ambiguïté quand l’individu a été actif occupé, chômeur ou inactif. Au sein de chacune de ces périodes, l’enquêté doit ensuite préciser le nombre de jours de maladie, de chômage partiel ou de formation qu’il a connus, ce qui permet de récupérer une information proche de celle recueillie avec les questionnaires papier. Enfin, lors de l’enquête, une incohérence entre les déclarations d’une vague et celles de la vague précédente peut être repérée informatiquement dès la collecte, ce qui donne à l’enquêteur la possibilité de la faire corriger à l’enquêté. Avec le passage à CAPI et l’expérience de plus en plus grande des enquêteurs, la qualité des calendriers d’activité du panel devrait donc notablement s’améliorer à partir de la vague 5.

18 Il permet par ailleurs de connaître exactement la durée en jours de chaque période et d’observer toutes les transitions, même s’il y en a plusieurs dans le mois.
2.3 L’estimation des taux de transition avec un échantillon plus faible reste-t-elle satisfaisante (comparaison des taux de transition entre enquêtes emploi et panel européen) ?

Pour « valider » les transitions observées sur le panel, on a souhaité les rapprocher de celles observées sur les calendriers d’activité de l’enquête Emploi. Cette validation n’est bien sûr que partielle. En effet, les calendriers de l’enquête Emploi sont, comme ceux du panel, déclaratifs (la codification d’une activité repose largement sur la perception de la signification des différentes rubriques par l’enquêté). Toutefois, elle permet de s’assurer que la taille limitée de l’échantillon du panel, les incohérences observées lors des recouvrements ou encore les différents choix méthodologiques qui ont été faits (traitements des mois à plusieurs occupations, des occupations « non principales ») ne conduisent pas à des résultats très différents de ceux que l’on tirerait de l’analyse des calendriers des Enquêtes Emploi19.

Pour tenir compte du fait que les modalités des calendriers d’activité sont différentes dans l’enquête Emploi et dans le panel - et que, plus les rubriques sont détaillées, plus le « bruit » est important (encadré 4) - on a choisi de se limiter à trois catégories dans la suite de ce travail : emploi, chômage et inactivité.

Les transitions les plus fréquentes sont les transitions entre emploi et chômage (0,4 % des individus un mois donné, soit une cinquantaine de cas par mois20) (tableau 11). En moyenne sur l’ensemble de la période étudiée (mars 1993 à septembre 1997), les taux de transition observés sur les enquêtes Emploi et le panel européen sont proches, sauf pour les transitions entre emploi et inactivité, pour lesquelles les taux de transitions sont supérieurs de 0,1 % dans le panel.

19 En effet, dans l’enquête Emploi, seule une occupation peut être déclarée un mois donné, et le nombre de types d’occupations est plus restreint. Enfin, nous n’exploitons ici les calendriers d’activité année par année, si bien qu’aucun problème de recouvrement de calendrier ne se pose.

20 Ces probabilités de transition sont non conditionnelles à la situation de l’individu le mois précédent. Autrement dit, on rapporte le nombre de transitions emploi-chômage observées au nombre d’individus présents, et non au nombre d’individus en emploi.
### Tableau 11 : Comparaison des taux de transition sur le marché du travail entre le panel et les enquêtes Emploi 1993-1997

<table>
<thead>
<tr>
<th>Enquête</th>
<th>Emploi</th>
<th>Chômage</th>
<th>Inactivité à</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paneurropien</td>
<td>1993 à 1997</td>
<td>0,46</td>
<td>0,10</td>
<td>0,20</td>
</tr>
<tr>
<td>Pan neurropien</td>
<td>1993 à 1996</td>
<td>0,42</td>
<td>0,20</td>
<td>0,21</td>
</tr>
<tr>
<td>Pan neurropien</td>
<td>1993 à 1997</td>
<td>0,42</td>
<td>0,20</td>
<td>0,21</td>
</tr>
<tr>
<td>Pan neurropien</td>
<td>1993 à 1997</td>
<td>0,42</td>
<td>0,20</td>
<td>0,21</td>
</tr>
</tbody>
</table>


Chômage : individus de 17 ans ou plus lors de l’enquête.

Lecture : En moyenne sur la durée d’observation du panel (janvier 93 à septembre 97), la probabilité d’observer une transition emploi/chômage un mois donné pour un individu est de 0,42 %. Sur la période correspondant aux calendriers de l’enquête Emploi, elle est de 0,43 %.
Graphiques 1 et 2 - Comparaison des données des trois premières vagues du panel européen et des Enquêtes Emploi

1. Transitions depuis l'emploi

2. Transitions depuis le chômage

Champ : individus de 17 ans ou plus lors de la vague d'enquête (pondérés de manière à être représentatifs de la population qui avait 17 ans ou plus)
Toutefois, si les transitions entre emploi et inactivité sont plus fréquentes sur le panel européen que sur les enquêtes Emploi, cet effet est surtout marqué sur la période correspondant à la première vague du panel, à cause probablement d’un problème de questionnement (encadré 4). Les différences, plus mineures, qui subsistent au-delà de la première vague pourraient être dues au fait que l’importance des calendriers d’emploi est différente dans les deux enquêtes. Alors que le calendrier d’activité, très complet, est au cœur du panel, il est en fin de questionnaire de l’Enquête Emploi (et ne porte que sur l’activité principale)21. On retrouverait alors un phénomène analogue à ce que l’on observe sur les revenus par exemple : plus les questionnements sont détaillés, plus ils permettent de récupérer de l’information22. Ici, l’accent mis sur l’importance du calendrier dans le panel pourrait conduire à recueillir un nombre de transition supérieur dans cette enquête. La différence entre le panel et les enquêtes Emploi pourrait également être en partie liée aux thèmes qu’elles traitent. En effet, on note que si les taux de transition entre emploi et inactivité sont supérieurs dans le panel, les taux de transition entre emploi et chômage sont eux inférieurs à ceux de l’enquête Emploi : sensibilisés dans cette dernière enquête par les questions portant sur le chômage, les enquêtés pourraient-ils se déclarer plus facilement au chômage qu’en inactivité ?

Par ailleurs, l’écart entre le taux de transition dans le panel et dans l’enquête Emploi dépend fortement du mois calendrier, probablement en raison des effets de mémoire : les transitions apparaissent plus fréquentes dans l’enquête Emploi que dans le panel entre novembre et mars (mois de collecte de l’enquête Emploi), tandis qu’elles sont plus importantes entre mars et octobre dans le panel (graphique 4). Fait exception à cette règle le mois d’août, avec un nombre supérieur de transitions dans l’enquête Emploi (pour une raison inexpliquée). Enfin, les valeurs observées lors des recouvrements de calendrier dans le panel sont particulièrement élevées, et il est difficile de faire la part exacte sur ces deux mois de l’effet de mémoire et des incohérences lors des recouvrements entre vagues du panel (encadré 4).

21 On peut se demander si, notamment, le panel permettrait de mieux saisir les alternances entre emploi ponctuel et inactivité des personnes principalement inactives. Ceci sera à examiner sur les données.

INSEE Méthodes 187
Graphique 4 - Comparaison des taux de transition mensuels entre emploi, chômage et inactivité dans le panel et dans les enquêtes Emploi : effet du mois

Source : Insee, panel européen, vagues 1994 à 1997; enquêtes Emploi 94 à 98
Champ : individus de 17 ans ou plus lors de l’enquête
Lecture : Sur la période 1994-1996, les taux de transition entre emploi, chômage et inactivité observés sur le panel sont en moyenne supérieurs de 0,4 % à ceux observés sur les Enquêtes Emploi. A cet écart moyen s’ajoutent des écarts supplémentaires selon le mois considéré : -0,25 % en janvier, -0,15 % en février...

Enfin, le taux de transition dans le panel diminue de plus en plus (relativement au taux de transition dans l’enquête Emploi) au fur et à mesure que le nombre de vagues augmente (tableau 12 et graphiques 1 et 2). Cette baisse de la fréquence des transitions dans le panel peut être interprétée comme le signe d’une perte de représentativité du panel au fur et à mesure que le nombre de vagues augmente. Cette baisse reste toutefois encore modérée en vague 4 : l’écart absolu entre taux de transition est de 0,1 % (pour un taux de transition total de 1,4 % environ). Il est très légèrement plus important si on estime les taux de transition dans le panel à partir des seuls individus répondants lors des trois vagues. Il faudrait pouvoir disposer d’un nombre supérieur de vagues pour voir dans quelle mesure la baisse de représentativité pourrait être gênante 23.

23 L’étude devrait alors sans doute se baser plutôt sur des probabilités de transition conditionnelles.
Tableaux 12 - Différence entre les taux de transition sur le marché du travail estimés sur le panel et les enquêtes Emploi 1993-1997

12a. Ensemble des individus de 17 ans et plus

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>emploi à chômage</td>
<td>-0.05</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.04</td>
</tr>
<tr>
<td>emploi à inactivité</td>
<td>0.11</td>
<td>0.11</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>total emploi vers non emploi</td>
<td>0.06</td>
<td>0.11</td>
<td>0.04</td>
<td>-0.05</td>
<td>-0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>chômage à emploi</td>
<td>-0.06</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>inactivité à emploi</td>
<td>0.10</td>
<td>0.10</td>
<td>0.05</td>
<td>0.03</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>total non-emploi vers emploi</td>
<td>0.04</td>
<td>0.11</td>
<td>0.08</td>
<td>-0.01</td>
<td>-0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>chômage à inactivité</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>inactivité à chômage</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>total</td>
<td>0.10</td>
<td>0.27</td>
<td>0.15</td>
<td>-0.04</td>
<td>-0.10</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Champ : individus de 17 ans ou plus lors de l’enquête

12b. Individus de 17 ans et plus présents lors des quatre vagues du panel

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>emploi à chômage</td>
<td>-0.06</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.08</td>
<td>-0.09</td>
<td>-0.05</td>
</tr>
<tr>
<td>emploi à inactivité</td>
<td>0.12</td>
<td>0.11</td>
<td>0.06</td>
<td>0.04</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>total emploi vers non emploi</td>
<td>0.06</td>
<td>0.09</td>
<td>0.03</td>
<td>-0.04</td>
<td>-0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>chômage à emploi</td>
<td>-0.07</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>inactivité à emploi</td>
<td>0.11</td>
<td>0.10</td>
<td>0.06</td>
<td>0.03</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>total non-emploi vers emploi</td>
<td>0.03</td>
<td>0.10</td>
<td>0.07</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>chômage à inactivité</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>inactivité à chômage</td>
<td>-0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>total</td>
<td>0.09</td>
<td>0.25</td>
<td>0.13</td>
<td>-0.01</td>
<td>-0.12</td>
<td>0.08</td>
</tr>
</tbody>
</table>


Champ : individus de 17 ans ou plus lors de l’enquête

Lecture : En moyenne de mars 93 à septembre 97, la probabilité d’observer une transition emploi/chômage un mois donné est inférieure de 0.04 points dans le panel si on l’estime à partir de l’ensemble des individus, et de 0.05 si on l’estime à partir des seuls individus présents lors des trois vagues.

Au final, la validation interne des calendriers d’activité du panel comme la comparaison avec les enquêtes Emploi indiquent que les transitions entre emploi et chômage que l’on définit à partir du panel sont de bonne qualité, et en particulier que le risque de transitions fictives (liées à la complexité du calendrier) est limité. Et si l’attrition biais sans doute légèrement les taux de transition, ce biais reste en vague 4 plus limité que celui lié à l’effet de mémoire.
Conclusion

Les enquêtes répétées visent souvent plusieurs objectifs : le statisticien souhaite à la fois disposer de données en coupe fiables, pouvoir mesurer avec précision les évolutions entre deux vagues d'enquêtes successives, et disposer de données individuelles longitudinales. Malheureusement, ces objectifs sont difficilement compatibles (Caron et Ravalet, 2000) :

- La reconstitution de micro-données longitudinales portant sur une période relatively longue (4 ans ici) n'est possible que grâce à un panel d'individus. En effet, bien que le taux de déménagement soit faible chaque année (de l'ordre de 10 %), le taux de déménagement cumulé sur plusieurs années devient lui beaucoup plus élevé, et il s'ajoute de plus à l'attrition : ne pas suivre les personnes qui déménagent revient à se priver d'une part de plus en plus importante de l'échantillon, et nuit fortement à la représentativité de ce dernier. Les statistiques descriptives peuvent en être fortement modifiées (exemple des taux d'emploi à trois ans des jeunes étudiants), non seulement parce que le déménagement peut être directement lié au phénomène étudié (les jeunes qui trouvent un emploi quittent leurs parents), mais surtout, parce qu'il peut également y être lié plus indirectement en raison de l'hétérogénéité individuelle (les personnes dont la situation sur le marché du travail est la plus précaire sont également celles dont la situation de logement est la plus instable, notamment parce qu'il s'agit fréquemment de jeunes en phase d'accès à l'indépendance).

- Pour pouvoir estimer les évolutions d'une grandeur avec précision, un panel d'individus semble également préférable, dans la mesure où il permet de s'assurer que les variations observées d'une période à l'autre ne sont pas (ou peu) liées à des changements d'échantillon.

- A l'inverse, la qualité des estimations en coupe est meilleure dans des enquêtes transversales ou des panels de logements, qui ne sont pas sensibles à l'attrition.

Si le panel de logements semble un bon compromis pour disposer à la fois d'estimations en coupe de qualité et de mesures d'évolutions précises (enquête Emploi, enquêtes Permanentes Conditions de Vie), il ne permet pas de suivre les personnes sur longue période. Ainsi, choisir entre un panel d'individus et un panel de logements revient dans une large mesure à arbitrer entre représentativité en coupe et possibilité de suivi sur longue période. Ce travail montre qu'il est possible d'arbitrer

---

24 Il faut toutefois signaler que le recours à un panel de logements, donc à une enquête répétée, peut avoir une incidence sur la qualité de l'information transversale. En effet, les ménages peuvent réagir différemment lors de passages ultérieurs de l'enquêteur (soit qu'ils comprennent mieux l'enquête, soit au contraire qu'ils essaient de limiter le temps de questionnement). Et l'attitude de ce dernier peut également varier (passation plus rapide des questions sur les points qui n'ont pas changé...). Ainsi, il semblerait que les trois tiers de l'enquête Emploi, même redressés indépendamment, ne conduisent pas à des résultats transversaux identiques.
en faveur d’un panel d’individus (sur 4 ans) sans devoir sacrifier de manière importante la qualité des estimations en coupe. En effet, la sélectivité de l’attrition sur 4 ans est suffisamment modérée pour que la structure de l’échantillon de répondants aux quatre vagues reste assez proche de la structure des répondants de départ. Par ailleurs, inclure des individus non-panels dans l’échantillon transversal permet de redresser une partie de ces biais. Et l’usage de pondérations longitudinales permet également de réduire les écarts observés entre estimations menées sur un échantillon cylindré et échantillon complet.\textsuperscript{25}

Au total, les taux de transition entre emploi, chômage et inactivité estimés à partir du panel restent proches de ceux de l’enquête Emploi. Mais bien sûr, il serait intéressant de refaire ce type d’étude avec un panel de plus longue durée...

\textsuperscript{25} Tout du moins, ceux de ces écarts qui sont liés à des caractéristiques observables.
Encadré 1 - Champ de l'étude

Dans cette étude, l'unité statistique retenue sera l'individu panel de 17 ans et plus :
- En premier lieu, l'étude sera basée sur les individus et non sur les ménages. En effet, il est délicat de raisonner au niveau du ménage dès lors que l'on souhaite effectuer des analyses longitudinales, car les contours d'un ménage ne sont pas stables dans le temps (par exemple en cas de décès, divorce, décohabitation, mariage, naissance...). Notamment, un ménage peut être « éclaté » si seuls certains de ses membres déménagent.

- Le choix d'une limite d'âge inférieure à 17 ans est à la fois dicté par les données (les questionnaires individuels du panel, qui contiennent notamment les calendriers d'activité, ne sont posés qu'aux 17 ans et plus) et par les thèmes qui vont servir d'exemple (principalement, l'analyse des trajectoires d'emploi). Pour l'analyse des trajectoires sur le marché de l'emploi, une limite d'âge supérieure sera également imposée (60 ans en première vague).

- Enfin, dans la première partie, la plupart des résultats portent sur les seuls individus ayant rempli un questionnaire individuel en première vague. Parmi ces 14 334 individus, 10 409 (73 %) ont également répondu aux trois vagues suivantes. En revanche, dans la seconde partie, les taux de transition mensuels entre emploi, chômage et activité sont - sauf mention du contraire - estimés sur l'ensemble des individus de 17 ans et plus à la vague considérée (16 007 personnes), qu'ils aient ou non été présents depuis la première enquête (il suffit qu'ils soient présents lors des mois étudiés).
Encadré 2 - Liens entre déménagement et modification de situation sur le marché de l'emploi

Compte tenu de l'importance des thématiques liées à la reprise ou la perte d'emploi (en particulier dans les études portant sur des panels), il est intéressant de voir dans quelle mesure déménagement et modification de la situation sur le marché de l'emploi peuvent être liés. Un tel lien peut être direct (par exemple, une personne déménage pour aller exercer un emploi qu'elle a trouvé loin de son domicile, ou elle déménage à la suite d'une perte d'emploi vers un logement moins cher) ou indirect (les personnes les moins insérées sur le marché de l'emploi auraient une résidence moins stable).

Des liens directs ou des changements simultanés peu importants...

Intéressons-nous dans un premier temps aux liens « directs » entre changements sur le marché de l'emploi et déménagement. Une première manière de le voir si de tels liens existent est de comparer le statut d'occupation des personnes qui déménaient avant et après leur déménagement (tableau E.1). Que ce soit chez les jeunes ou sur l'ensemble de la population, les changements de situations coïncidant avec un déménagement semblent rares. Plus précisément, dans 87% des cas, les personnes qui sont en emploi, au chômage, en études ou en inactivité (autre que les études) deux mois avant un déménagement sont encore dans la même situation deux mois après. Et il n'est pas possible de savoir à travers le calendrier d'activité si ces personnes ont changé d'emploi (ou d'établissement d'études).

Ainsi, si l'on prend l'exemple des 26-50 ans, ceux-ci sont 9 fois sur 10 dans la même situation deux mois avant et deux mois après le déménagement (tableau E.1). Dans 4% des cas, les déménagements les concernant s'accompagnent d'un retour à l'emploi, et dans 5% des cas, d'une perte d'emploi. Ce dernier chiffre est un peu plus élevé pour les femmes (6%) que pour les hommes (4%). L'étude de Dormont et Dufour-Kippelen (2000) sur les jeunes peu diplômés confirme qu'il y a des interactions significatives entre durées d'accès au CDI et durées de cohabitation avec

Les analyses de cet encadré ne sont menées que pour les seules personnes qui répondent, pour lesquelles il est possible d'exploiter la date de déménagement, le calendrier d'activité et le motif de déménagement. Plus précisément, l'échantillon est restreint aux personnes ayant répondu aux quatre vagues.

Comparer les situations deux mois avant et deux mois après permet de s'affranchir d'éventuels « bruits » lors du mois du déménagement (absence d'activité durant ce mois pour permettre le déménagement). Une variante « 3 mois avant - 3 mois après » a été effectuée, et donne des résultats analogues.

les parents, mais que ces interactions restent modestes en comparaison avec l’effet d’autres variables explicatives.

**Tableau E.1 – Situations sur le marché de l’emploi deux mois avant et deux mois après le déménagement**

<table>
<thead>
<tr>
<th>de 17 à 20 ans</th>
<th>ensemble</th>
<th>hommes</th>
<th>femmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>études -&gt; études</td>
<td>64</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>emploi -&gt; emploi</td>
<td>14</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>sans emploi (hors études) -&gt; sans emploi</td>
<td>12</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>dont initialement au chômage</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dont initialement inactif (hors études)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>études -&gt; emploi</td>
<td>5</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>autres situations (emploi-&gt;pas emploi, sans emploi-&gt;emploi)</td>
<td>6</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>dont initialement inactif (hors études)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ensemble</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>de 21 à 25 ans</th>
<th>ensemble</th>
<th>hommes</th>
<th>femmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>emploi -&gt; emploi</td>
<td>67</td>
<td>70</td>
<td>61</td>
</tr>
<tr>
<td>études -&gt; études</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>études -&gt; emploi</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>emploi -&gt; sans emploi</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>sans emploi (hors études) -&gt; emploi</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>sans emploi (hors études) -&gt; sans emploi</td>
<td>12</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>dont initialement au chômage</td>
<td>8</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>dont initialement inactif (hors études)</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>ensemble</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>de 26 à 50 ans</th>
<th>ensemble</th>
<th>hommes</th>
<th>femmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>emploi -&gt; emploi</td>
<td>74</td>
<td>86</td>
<td>64</td>
</tr>
<tr>
<td>sans emploi (hors études) -&gt; emploi</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>emploi -&gt; sans emploi</td>
<td>5</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>sans emploi (hors études) -&gt; sans emploi</td>
<td>16</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>dont initialement au chômage</td>
<td>7</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>dont initialement inactif (hors études)</td>
<td>9</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>autre situations (études -&gt; études)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ensemble</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>de 51 à 60 ans</th>
<th>ensemble</th>
<th>hommes</th>
<th>femmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>inactivité -&gt; inactivité</td>
<td>41</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>emploi -&gt; emploi</td>
<td>33</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>emploi -&gt; sans emploi</td>
<td>14</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>sans emploi (hors études) -&gt; sans emploi</td>
<td>11</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>dont initialement au chômage</td>
<td>1</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>dont initialement inactif (hors études)</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ensemble</td>
<td>110</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>plus de 61 ans</th>
<th>ensemble</th>
<th>hommes</th>
<th>femmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>inactivité -&gt; inactivité</td>
<td>99</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>autres situations (emploi-&gt;emploi, emploi-&gt; sans emploi)</td>
<td>1</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>ensemble</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee*


194  

*INSEE Méthodes*
Une deuxième façon d’étudier les liens directs entre déménagement et changement de situation sur le marché de l’emploi est de demander aux personnes qui déménagent la raison de leur déménagement. Seuls 6 % des déménagements sont liés à une reprise d’emploi (tableau E.2). 15 % supplémentaires sont justifiés par une « autre raison liée à l’emploi », comme un rapprochement du lieu de travail. Et les 80 % restants relèvent d’autres motivations (raisons liées au logement...).

**Tableau E.2 - Principal motif du déménagement entre deux vagues successives en fonction de la trajectoire entre ces deux vagues**

<table>
<thead>
<tr>
<th></th>
<th>vous ou une autre personne du ménage avez trouvé un emploi ici</th>
<th>pour une autre raison liée à l’emploi (rapprochement du lieu de travail)</th>
<th>pour des raisons liées au logement</th>
<th>pour d’autres raisons</th>
<th>ensemble des principaux motifs exprimés</th>
</tr>
</thead>
<tbody>
<tr>
<td>toujours en emploi</td>
<td>4</td>
<td>15</td>
<td>53</td>
<td>28</td>
<td>100</td>
</tr>
<tr>
<td>alternance emploi / non emploi</td>
<td>13</td>
<td>19</td>
<td>30</td>
<td>39</td>
<td>100</td>
</tr>
<tr>
<td>jamais en emploi</td>
<td>3</td>
<td>12</td>
<td>35</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>ensemble</td>
<td>6</td>
<td>15</td>
<td>44</td>
<td>36</td>
<td>100</td>
</tr>
</tbody>
</table>

*Source : panel européen, vagues 1 à 4 (1994 à 1997), Insee*

*Champ : individus panel de 17 à 60 ans ayant rempli un questionnaire individuel en 1994 et aux deux vagues consécutives étudiées. Pourcentages sur les seules réponses exprimées (non-réponse d’environ 5 %). Moyennes sur les principaux motifs de déménagements exprimés dans les vagues 2, 3 et 4*

Ces résultats, comme les précédents, invitent à relativiser l’importance des liens directs entre déménagement et situation sur le marché de l’emploi : dans la plupart des cas, le changement de logement ne serait pas concomitant à une modification de la situation sur le marché de l’emploi.

*... mais des comportements fortement différenciés entre ceux qui déménagent et les autres.*

Les analyses ci-dessus ne portent que sur des changements quasi-simultanés dans les domaines du logement et de l’emploi. Or, le déménagement peut être assez éloigné dans le temps de la reprise d’emploi, par exemple dans le cas d’une personne qui retrouve un emploi et ne décide de déménager que lorsqu’elle est sûre que son nouvel emploi sera durable. Par ailleurs, ces analyses ne tiennent pas (ou peu) compte des liens indirects entre emploi et logement liés à l’hétérogénéité individuelle (les personnes les moins bien insérées sur le marché de l’emploi sont sans doute plus susceptibles d’avoir une situation de logement instable).

Pour juger de l’ensemble de ces effets, il est possible de comparer les situations vis-à-vis de l’emploi des personnes qui déménagent et des autres. Ainsi, le taux d’emploi à trois ans de jeunes élèves ou étudiants est beaucoup plus élevé parmi ceux qui ont déménagé (cf. texte). Plus généralement, les trajectoires sur le marché de l’emploi de ceux qui ont déménagé sont beaucoup moins stables que celles de ceux qui n’ont pas changé de logement : 38 % de ceux qui ont déménagé entre l’automne 1994 et l’automne 1997 ont connu à la fois l’emploi et l’absence d’emploi durant cette
période, contre 20 % de ceux qui n’ont pas changé de logement. Autrement dit, 39 % de ceux qui ont connu à la fois l’emploi et le non-emploi ont déménagé. Le lien indirect entre emploi et logement joue sans doute un rôle important dans ces résultats, car les taux de déménagement entre 1994 et 1997 semblent varier à peu près autant en fonction de la trajectoire entre 1994 et 1997 (tableau E.3) qu’en fonction de la trajectoire antérieure (tableau 7c).

**Tableau E.3 - Trajectoire entre novembre 1994 et octobre 1997 (vagues 2 à 4) selon l’existence d’un déménagement entre les vagues 1 et 4**

<table>
<thead>
<tr>
<th>Trajectoire entre nov. 94 et oct 97 (36 mois)</th>
<th>Ensemble</th>
<th>Pas de déménagement</th>
<th>Déménagement entre v1 et v4</th>
<th>Taux de déménagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emploi tous les mois</td>
<td>54%</td>
<td>57%</td>
<td>46%</td>
<td>22%</td>
</tr>
<tr>
<td>tous les mois à temps</td>
<td>39%</td>
<td>40%</td>
<td>37%</td>
<td>24%</td>
</tr>
<tr>
<td>au moins un mois en temps</td>
<td>8%</td>
<td>8%</td>
<td>6%</td>
<td>20%</td>
</tr>
<tr>
<td>au moins un mois en</td>
<td>7%</td>
<td>9%</td>
<td>3%</td>
<td>11%</td>
</tr>
<tr>
<td>Emploi et non emploi</td>
<td>25%</td>
<td>20%</td>
<td>38%</td>
<td>39%</td>
</tr>
<tr>
<td>emploi puis</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>23%</td>
</tr>
<tr>
<td>chômage puis</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>44%</td>
</tr>
<tr>
<td>alternances d’emploi et de</td>
<td>8%</td>
<td>7%</td>
<td>14%</td>
<td>42%</td>
</tr>
<tr>
<td>activité et inactivité (+)</td>
<td>13%</td>
<td>10%</td>
<td>20%</td>
<td>40%</td>
</tr>
<tr>
<td>Jamais d’emploi</td>
<td>21%</td>
<td>23%</td>
<td>15%</td>
<td>19%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>26%</td>
</tr>
</tbody>
</table>

*Source : panel européen, vagues 1 à 4 (1994 à 1997). Insee*

Au total, se restreindre à un échantillon de personnes n’ayant pas déménagé pour étudier les trajectoires sur le marché de l’emploi conduit à se limiter à un échantillon de personnes dont les trajectoires d’emploi sont plus stables ou plus tournées vers l’emploi. Cela ne s’explique pas tant par le fait qu’elles n’ont pas connu de déménagement associé à un changement d’emploi, que par le fait que leurs caractéristiques diffèrent (et que l’effet de certaines de ces caractéristiques, inobservables, ne peut être corrigé par redressement). Les statistiques descriptives tirées d’échantillons de personnes n’ayant pas déménagé sont donc biaisées vers une plus grande stabilité ou une meilleure insertion sur le marché du travail. Mais surtout, une étude des changements sur le marché de l’emploi menée sur les seules personnes n’ayant pas déménagé conduit à ne pas tenir compte d’une part très importante des personnes qui ont connu l’événement étudié.
Encadré 3 Les poids longitudinaux des individus panel adultes dans le panel européen

Le principe général de calcul des pondérations longitudinales

Dans le panel, le poids d'un individu change d'une vague à l'autre. En effet, à chaque vague, les individus panel répondants doivent être représentatifs de la population de l'année, et leurs poids doivent donc être recalés pour tenir compte de l'attrition d'une vague sur l'autre. Ainsi si sur 100 individus répondants à la vague n (supposés identiques), seuls 50 sont répondants à la vague n+1, le poids (longitudinal) moyen de chaque répondant de n+1 sera le double de son poids en n. En fait, les individus ne sont pas identiques, si bien que le coefficient multiplicatif appliqué à chacun dépend de ses caractéristiques. Plus précisément, la méthode de calcul des poids est la suivante. Un modèle logit dichotomique permet dans un premier temps de repérer les variables qui permettent de mieux prédire la non-réponse. Sur la base de ces résultats, les individus sont regroupés en fonction de leurs caractéristiques pour définir des catégories homogènes par rapport à la non-réponse (chaque catégorie devant être d'effectif suffisant). Enfin, on calcule pour chaque catégorie le taux de non-réponse d'une vague à la suivante, ou plus généralement, à la vague n+i. Les pondérations de la vague n+i (notées BASE_{n+i}) s'obtiennent alors à partir des pondérations de la vague n par la formule : BASE_{n+i} = BASE_n / (1 - taux de non-réponse).

Les facteurs explicatifs de la non-réponse d'une vague sur l'autre.

Les taux de non-réponse varient en fonction des caractéristiques des enquêtés. Dans une optique de calcul de pondérations longitudinales, il est possible d'étudier les variations des taux de réponse à la vague n+i (sachant que la personne est répondante en vague 1 ou en vague n) en fonction de toutes les informations recueillies aux vagues précédentes. L'analyse des taux de non-réponse est donc plus riche que dans une enquête transversale, puisque l'on n'est pas obligé de se restreindre à l'analyse de l'impact des caractéristiques décrites dans la base de sondage.

Sur l'ensemble de la population, le principal facteur explicatif de la non-réponse d'une vague sur l'autre est l'appartenance à un ménage dont certains membres sont partis entre les deux vagues considérées. Viennent ensuite : la zone urbaine de résidence (l'agglomération parisienne a un taux de non-réponse plus élevé) et le type de contrat de travail de la personne de référence (les taux de réponse sont plus élevés quand la personne de référence est en CDI qu'en CDD), puis sa catégorie socio-professionnelle (ce sont les ménages agriculteurs qui répondent le plus, et les ménages d'ouvriers et d'artisans ou commerçants qui répondent le moins) et le niveau des charges de logement (des charges trop élevées entraînent une probabilité de non-réponse accrue à la vague suivante). Selon les vagues et les modèles, d'autres
variables peuvent avoir une influence, comme le type de ménage (les couples avec enfants ont des taux de non-réponse d’une vague sur l’autre plus faibles) ou le diplôme de la personne de référence (l’attrition est plus faible pour les diplômés de l’enseignement supérieur)\textsuperscript{29}.

Dans le cas où les personnes qui déménagent ne sont pas suivies, sont considérées comme « non-répondantes au sein du même logement » à la fois les personnes qui ne répondent plus, mais également toutes celles qui ont déménagé. Les facteurs explicatifs de cette non-réponse au sein du même logement ne sont plus les mêmes que précédemment : notamment, le fait d’appartenir à un ménage dont certains membres sont partis n’a plus d’influence, alors que la variable d’âge devient très fortement significative. Finalement, parmi les 6 variables retenues pour expliquer la non-réponse au sein du même logement, la plus discriminante est l’âge de la personne. Viennent ensuite le statut d’occupation du logement et la catégorie socio-professionnelle de la personne de référence, et enfin le type de contrat de travail de la personne de référence, la zone urbaine de résidence et le niveau des charges de logement. Par regroupement de classes d’individus obtenues par croisement de ces modalités, 55 catégories homogènes vis-à-vis de la non-réponse au sein du même logement sont définies.

\textsuperscript{29} Cf. document de travail à venir sur la pondération de la vague 4 du panel.

198

\textit{INSEE Méthodes}
Encadré 4 : qualité du calendrier d’activité du panel européen

Des calendriers particulièrement complexes et détaillés

Dans sa version française, le calendrier d’activité du panel européen est particulièrement complexe et détaillé. Il permet de connaître mois par mois depuis janvier 1993, les situations professionnelles des individus âgés de 17 ans et plus l’année de l’enquête. 17 situations différentes sont proposées : 9 correspondent au croisement entre le type d’emploi (Contrat à Durée Indéterminée, Contrat à Durée Déterminée ou indépendant) et la durée du travail (temps complet, temps partiel de plus ou de moins de 15 heures par semaine), 4 correspondent aux principales formes d’inactivité (retraite ou préretraite, études initiales, autre) ou au service national. Enfin, les 4 dernières modalités concernent l’existence d’activités secondaires, de formations (non initiales), de chômage partiel et enfin, d’absences pour cause de maladie. Dans le questionnaire français du panel européen\(^{30}\), deux situations peuvent être recueillies pour un mois donné.


Qualité des calendriers d’activité du panel au sein de chaque vague..

Dans un premier temps, des contrôles de qualité et de cohérence ont été menés séparément sur les fichiers d’activité correspondant aux trois premières vagues du panel. Plus de 500 lignes des calendriers d’activité\(^{31}\) ont été corrigées sur la première vague (soit 2,6 % des lignes), près de 300 en deuxième vague (1,4 %) et plus de 150 en troisième vague (0,8 %). La recherche d’erreurs ayant été faite sensiblement de la même manière sur les trois vagues, il est intéressant de constater à travers ces chiffres que la qualité du panel s’est améliorée avec le temps (ce que l’on peut, au moins pour partie, imputer à la formation des enquêteurs).

Voici une liste non exhaustive des problèmes détectés sur les lignes des calendriers d’activité :

\(^{30}\) Ce n’est pas le cas dans la version européenne du questionnaire.

\(^{31}\) Il existe une ligne d’activité par type d’activité (occupation) déclarée au cours d’une vague par individu.
Les lignes d’activités pour lesquelles aucun mois n’était rempli ont été renseignées dans la mesure du possible à l’aide des informations des calendriers de revenu (ou des questionnaires individuels). Cela concerne moins de 0,5 % des lignes en V1 (dont la moitié a été corrigée), et cette proportion diminue nettement en V2 et en V3.

Les individus ne déclarant pour une vague donnée que des mois de maladie se sont vu ajouter une seconde ligne d’activité (voire une troisième), en fonction de leurs déclarations au questionnaire individuel (occupation, dates de fin d’activité...) et aux autres questionnaires (calendrier d’activité des autres vagues, fiches revenus...). En effet, toutes les lignes ‘maladie’ ne concernaient pas des travailleurs en absence temporaire (congés maladie), mais elles étaient parfois remplies par des retraités ou d’autres inactifs. Ces corrections portent sur 150 personnes en V1, mais sur moins d’une quinzaine en V2 et moins de 10 en V3.

De nombreuses erreurs de saisie sur les codes activité ont également été détectées et corrigées, comme la confusion à la saisie entre les codes ‘11’ et ‘17’ ou ‘10’ et ‘16’.

De manière générale, l’activité « chômage partiel » est apparue peu fiable. Alors que cette modalité visait à récupérer de l’information sur les individus auxquelles leur entreprise imposait une activité réduite (du fait d’une perte d’activité de nature économique), certains enquêtés se sont déclarés en chômage partiel car à temps partiel et souhaitant travailler plus, ou encore car ils avaient quitté leur emploi au milieu de mois et avaient donc été partiellement chômeurs un mois donné. Enfin, certains chômeurs ont sans doute été classés en chômage partiel parce que la ligne d’activité « chômage partiel » figurait sur la première page du questionnaire (et non la ligne chômage, reléguée sur la suivante).

Une analyse systématique a été menée pour détecter les actifs occupés qui auraient connu un ou deux mois d’inactivité pendant l’été. En effet, certaines personnes se sont déclarées inactives pendant leurs congés (parfois non sans raison, comme pour les assistances maternelles ou les surveillants dans les lycées). Lorsqu’il n’y avait pas de déclaration de changement d’activité, on a alors considéré que la personne était en fait occupée. Au total, les déclarations d’une dizaine d’actifs ont été corrigées en vague 1. Les ruptures de calendrier lors des mois d’été étaient en revanche beaucoup plus fréquentes pour les étudiants (qui ne se déclaraient plus en études lors des mois d’été, et étaient donc classés parmi les autres inactifs). Les corrections pour les étudiants ont été effectuées au cas par cas.

Comme l’indiquent certains résultats ci-dessus, la qualité de l’information dans le panel semble s’être améliorée avec le nombre de vagues. Certains problèmes n’apparaissent en fait qu’en vague 1, du fait d’un questionnement légèrement différent (ou d’enquêteurs moins rompus aux subtilités de l’enquête). Ainsi, une
comparaison entre les taux de transition entre emploi et inactivité dans le panel et les enquêtes Emploi indique que ces taux sont surestimés dans le panel pendant les 21 premiers mois (correspondant au calendrier d’activité de la vague 1) (graphique E.1). En effet, en première vague, la ligne « autre inactif » n’existait pas dans les calendriers d’activité. Cette ligne devait être reconstituée en négatif (absence d’autre activité), avec les risques que cela comporte.

**Graphique E.1 - Transitions entre emploi et inactivité : comparaison des taux de transition du panel et de l’enquête Emploi**

![Graphique E.1](image)

Source : Insee, panel européen, vagues 1994 à 1997; enquêtes Emploi 94 à 98  
Champ : Individus de 17 ans ou plus lors de l’enquête  
Lecture : Sur la période mars 1993 (mois 3)- septembre 1994 (mois 21), les taux de transition entre emploi et inactivité observés sur le panel sont en moyenne supérieurs de 0,1% à ceux observés sur les Enquêtes Emploi.

... et incohérences détectées lors des recouvrements entre vagues

On détecte également des incohérences sur les calendriers d’activité lorsque ceux-ci se recouvrent, comme au mois d’octobre 1994 entre les deux premières vagues, ou aux mois d’octobre à décembre 1995 (resp. 1996) entre les vagues 2 et 3 (resp. 3 et 4) (schéma E.1). Chaque mois, de 1 à 4% d’individus changent d’occupation « principale » d’après les trois premières vagues du panel (graphique E.2). Bien sûr, ces changements sont de nature très différente. Pour certains, il s’agit simplement d’un passage du temps partiel au temps plein, et pour d’autres, de la perte d’emploi ou du passage à la retraite. Mais plus les changements sont observés à un niveau fin, plus les risques d’erreurs (transitions fictives ou mal datées) sont importants :

- En vague 1 par exemple, la première ligne du calendrier d’activité (correspondant au CDI temps plein) a souvent été renseignée pour des individus qui se déclaraient en CDD ou à temps partiel lorsqu’on les interrogeait sur leur situation du moment dans le questionnaire individuel. Heureusement, la qualité du remplissage des calendriers d’activité s’est améliorée avec le nombre de vagues.
- Certains changements « mineurs » sont moins bien repérés par ces calendriers (par exemple, les transitions temps partiel / temps plein ou inactivité / chômage)...

Ces phénomènes expliquent les forts taux de transition observés entre les vagues 1 et 2 (mois de septembre à novembre 1994). Alors que les courbes correspondant aux différentes années sont proches sur les autres mois, la période automnale se caractérise par des écarts plus importants (graphique E.2). On constate notamment que les taux de transition observés entre septembre et novembre 1994 (passage du calendrier vague 1 au calendrier vague 2) sont plus importants que ceux observés à la même époque en 1993 (relevés sur le calendrier de la seule première vague). L'accroissement du taux de transition en automne se réduit lors des deuxième et troisième vagues (puis lors des troisième et quatrième vagues) (graphique E.2), ce qui renforce l'idée que la qualité du panel s'améliore notablement avec le nombre de vagues\(^\text{32}\).

**Schéma I - Les calendriers d’activité des quatre premières vagues du panel : mois sur lesquels portent les enquêtes et périodes de recouvrement**

\[\text{Act} \quad \text{Recouv}\]

\[\begin{array}{cccc}
oct. à déc. & sept. à déc. & sept. à déc. & sept. à déc. \\
\end{array}\]

\[\begin{align*}
\text{act} & : \\
\text{jan. 93 à oct. 94} & : \\
\text{oct. 94 à sep/déc. 95} & : \\
\text{oct. 95 à sep/déc. 96} & : \\
\text{oct. 96 à sep/déc. 97} & : \\
\text{oct. 94} & : \\
\text{oct. à déc. 95} & : \\
\text{oct. à déc. 96} & : \\
\end{align*}\]

---

\(32\) Si l'on regroupe les Contrats à Durée Indéterminée (CDI) (resp. les contrats à durée déterminée (CDD), les activités indépendantes) de plus de 15 heures par semaine avec les CDI (resp. les CDD, les activités indépendantes) de moins de 15 heures par semaine (et compte tenu des apports prélablement effectués), seules 179 lignes activités n'ont pas la même valeur pour le mois d'octobre 1994 en vagues 1 et 2, et 148 pour les mois d'octobre à décembre en vagues 2 et 3. Ces problèmes de recouvrement ont été traités de manière à assurer la plus grande continuité dans les calendriers. Une attention particulière a été portée aux individus ne déclarant pas d'activité lors du mois d'enquête.
Graphique E.2 - Taux mensuels de changements d'occupation principale d'après le panel européen en fonction du mois et de l'année (nomenclatu re du calendrier d'activité en 19 postes).

Champ : individus de 17 ans ou plus lors de la vague d'enquête (pondérés de manière à être représentatifs de la population qui avait 17 ans ou plus)

Résumer toutes les situations déclarées pour un mois donné en une occupation principale

Dans le calendrier d'activité du panel, plusieurs occupations différentes peuvent être déclarées un même mois. Ainsi, un individu peut être le même mois en CDD temps plein et en maladie (occupations simultanées), ou bien en CDD temps plein et au chômage s'il a été licencié au cours du mois (occupations consécutives). Le traitement des cas d'occupations multiples peut bien sûr avoir une influence sur les transitions que l'on repère. Pour cette étude, les différents types d'occupations présentés un mois donné ont été synthétisés en une occupation dite principale. Pour ce, l'activité a été privilégiée sur le chômage, lui-même privilégié sur l'inactivité. Ce choix implique que de courts passages par le chômage (inférieurs à un mois) ne sont pas pris en compte. Implicitement, on ne s'intéresse donc qu'aux périodes de chômage ayant au moins une durée de l'ordre du mois. Au terme de cette étape, les occupations obtenues pour certains mois restaient difficiles à classer entre activité, inactivité et chômage. C'était le cas par exemple lorsque l'occupation principale du mois était une activité secondaire (1) ou une formation complémentaire (qui peut avoir lieu dans le cadre d'un emploi ou non). C'était également le cas pour des occupations telles la maladie ou le chômage partiel, qui - compte tenu des déclarations observées - ne semblaient pas uniquement concerner des actifs. Un peu plus de 1 % des mois d'activité étaient concernés. Le choix qui a été fait est de

INSEE Méthodes
prolonger les occupations déclarées le mois précédant et le mois suivant la maladie, le chômage partiel, la formation complémentaire ou l’activité secondaire. Lors de ce traitement, entre un cinquième et un quart des cas de chômage partiel ou de maladie ont ainsi été classées en chômage ou inactivité, ainsi que la moitié des cas de formations complémentaires, et les deux tiers des activités secondaires. Au total, 7 % des calendriers individuels ont été modifiés à cette étape.
Annexe : taux de déménagement et d'éclatement de ménages.

Une sous-estimation négligeable des taux de déménagement

Dans le panel européen, plusieurs variables contiennent de l'information sur l'existence d'un déménagement ou d'un « éclatement de ménage » (situations dans lesquels certains membres présents dans le ménage à l’enquête précédente le quittent). Toutefois, ces variables ne sont pas toutes renseignées systématiquement. Certaines ne le sont notamment que lorsque que le ménage accepte de répondre au questionnaire. Pour la présente étude, toute l’information disponible a été utilisée pour repérer les cas de déménagement ou d’éclatement de ménage, mais les taux de déménagement et d’éclatement restent malgré cela encore légèrement sous-estimés. En effet, si on peut supposer que l’on arrive à détecter presque tous les cas où tous les membres du ménage quittent le logement, en revanche, quand le ménage est non-répondant (et que tous ses membres n’ont pas déménagé), il n’est pas possible de savoir si certains membres du ménage l’ont tout de même quitté. On est amené à supposer qu’aucun membre du ménage n’est parti, ce qui conduit à des taux de déménagement (ou d’éclatement de ménage) légèrement sous-estimés. Ce biais est toutefois négligeable. En effet, la non-réponse est de l’ordre de 10 %, et le taux d’éclatement de ménage est de quelques pourcents (et est sans doute plus faible au sein des ménages qui déménagent), soit un biais inférieur à 0,5 %.33

Réconcilier des informations multiples mais souvent incohérentes

Déménagement du ménage

Lorsque le ménage est non-répondant, on utilise les variables TRESFA, remplies par l’enquêteur, qui décrivent les résultats des essais de contact de ce dernier avec le ménage, ainsi que la variable TCHAD (qui décrit l’existence d’un changement d’adresse depuis la vague précédente - téléphonique34 ou en vis-à-vis - et est donc plus difficilement exploitables). Lorsque le ménage est répondant, on peut de plus utiliser la question qui lui est directement posée de savoir s’il a déménagé (LOGID). Les codes des communes de résidence lors des différentes vagues d’enquêtes sont aussi utilisés pour trancher les cas litigieux. Toutefois, toutes ces informations ne sont pas toujours cohérentes entre elles (cf. notamment les travaux de L. Gobillon).

33 En revanche, le fait que l’on ne puisse pas savoir si une personne a quitté le ménage si ce dernier ne répond pas interdit de calculer de manière valable des taux de réponse pour les individus restant dans le ménage, selon que d’autres membres du ménage sont partis ou non (on aboutirait par exemple en vague 1 à un taux de réponse en vague 2 anormalement élevé de 96 % pour les individus qui n’ont pas déménagé à l’inverse d’autres membres de leur ménage : ce taux élevé est simplement dû au fait qu’on ne sait en général quand un individu a quitté le ménage que si un autre membre au moins du ménage accepte de répondre).
34 Des vagues téléphoniques ont été intercalées entre les premières enquêtes en vis-à-vis réalisées chaque année à l’automne. Elles avaient pour principal objet de repérer les principales modifications affectant le ménage, mais n’ont pas été exploitées.

INSEE Méthodes 205
Notamment, dans environ la moitié des cas, le code de la commune change alors que le ménage déclare ne pas avoir changé de logement. Il s’agit souvent de problème de codage d’arrondissement. Dans tous les cas, l’information donnée par le ménage a été privilégiée, en particulier quand elle s’accompagne de réponses à des questions sur la date et les motifs du déménagement.

Déménagement de l’individu

Pour les ménages répondants, les personnes ayant quitté leur ménage sont identifiées avec la variable TNOC (mouvement des personnes), et bien sûr par comparaison des présents dans le ménage aux différentes vagues ou examen du numéro de ménage.

Au total, les informations relatives aux mouvements des individus ou des ménages ou aux difficultés rencontrées pour interroger le ménage ou certains de ses membres sont multiples, souvent incohérentes ou incomplètes. La détermination des personnes ayant déménagé repose donc sur des règles de décision, dont la modification pourrait faire légèrement varier les statistiques sur la non-réponse et ses causes.
LA MESURE DES COMPÉTENCES :
LES LOGIQUES CONTRADICTOIRES DES
ÉVALUATIONS INTERNATIONALES

X. D'HAUTFOEUILLE (*) , F. MURAT(**) et T. ROCHER(***)

(*) INSEE, Unité "Méthodes Statistiques"
(**) INSEE, Division "Etudes Sociales"
(*** ) Ministère de l'Education Nationale, DPD

Le besoin d'évaluer le fonctionnement du système éducatif sur des bases objectives est devenu particulièrement fort, à la fois chez les acteurs de ce système (décideurs politiques, professeurs, etc.) et chez les « consommateurs » (les élèves et leurs parents). Depuis 1992, l'OCDE publie chaque année pour répondre à cette demande une série d'indicateurs sur l'école1. Cette publication propose des informations sur les moyens mis en œuvre (part du PIB consacrée à l'éducation, formation des enseignants, taille des classes...), ainsi que des chiffres relatifs aux résultats obtenus : taux d'accès en fin d'études secondaires, efficacité des diplômes sur le marché du travail mais aussi des indicateurs concernant les compétences des populations scolaires ou ayant fini leurs études. La place grandissante de ces derniers indicateurs incite à s'interroger sur leur valeur.

Dans un premier temps, on présentera les objectifs des évaluations de compétences, c'est-à-dire comment elles permettent d'éclairer le débat sur le fonctionnement du système éducatif. On s'intéressera plus particulièrement au cas des enquêtes internationales. Nous donnerons ensuite un aperçu des théories statistiques inhérentes à ces domaines. Les deux dernières parties seront consacrées aux problèmes concernant l'élaboration de la mesure et la comparabilité des résultats d'un pays à l'autre. Ces difficultés affectent les niveaux moyens aussi bien que les inégalités au sein de chaque population.

1 Cette publication annuelle a pour titre « Regards sur l'Education ».

INSEE Méthodes 207
1. Evaluation des compétences : définition et utilité

1.1. Evaluation individuelle et évaluation statistique

A moins d'avoir effectué sa scolarité en marge du système scolaire habituel, il est probable que le lecteur aura déjà répondu, plus ou moins directement, à plusieurs centaines de tests d'évaluation de ses compétences. On peut donner quelques exemples : la dictée de CE2, le contrôle de math de 3ème, un bac « blanc », le vrai bac, un concours de la fonction publique, les tests psychotechniques des « 3 jours », la note administrative pour les attachés INSEE... Toutes ces formes d'évaluations ont des objectifs bien différents mais au moins un point commun : elles ont une implication essentiellement individuelle. C'est l'individu interrogé qui a intérêt à réussir l'évaluation, pour obtenir un diplôme ou une place sur le marché du travail ou quelques mois d'avancement dans son échelle de salaires.

Les évaluations statistiques dont nous allons parler ont une visée bien distincte : c'est le système éducatif dans son ensemble qui est évalué au travers d'un échantillon d'individus. Ceux-ci n'ont généralement rien à gagner. Il s'agit d'étudier le niveau de la population dans son ensemble (c'est-à-dire en moyenne ou en considérant plus finement la distribution des résultats, en particulier pour les groupes les plus en difficulté). Ce changement de perspective explique bien des particularités des évaluations statistiques de compétences.

De plus, ces évaluations sont souvent accompagnées d'un questionnement statistique plus « classique ». En même temps qu'on évalue un élève, on va recueillir une somme d'informations relatives à ses caractéristiques sociodémographiques, ses opinions sur l'école, celles de ses parents, les opinions et pratiques pédagogiques de ses professeurs, les caractéristiques de l'établissement qui l'accueille... Dans le cas où l'échantillon est suivi sur plusieurs années, on s'intéresse aussi à son parcours scolaire, aux différents établissements qu'il a fréquentés. On le voit, il existe une quantité formidable de données qui pourront servir dans certains cas à expliquer les

\[2\] On verra d'ailleurs que cela peut avoir un certain effet sur leur motivation.

\[3\] Il ne faut cependant pas établir une frontière trop rigide entre les deux types d'évaluations. Certaines peuvent remplir, de façon plus ou moins satisfaite, les deux objectifs. L'obtention du baccalauréat n'est pas qu'un événement heureux pour le candidat : un taux élevé de réussite est aussi un motif de satisfaction pour l'école. De même les évaluations nationales en CE2 et 6ème ont des objectifs prioritairement pédagogiques (il s'agit d'aider chaque professeur à repérer les difficultés rencontrées par les élèves) mais sont aussi utilisées pour donner une image de l'état des compétences à ces deux niveaux.
compétences de l’individu et dans d’autres cas seront elles-mêmes éclairées par la
connaissance de ces compétences.

La constitution d’une évaluation de compétence fait appel à un processus un peu
différent de celui d’une enquête ordinaire :

- Tout d’abord, un groupe de travail composé de professeurs et de pédagogues
  constitue un protocole d’évaluation à partir d’un tableau de compétences. Il est
  alors important de bien définir la ou les compétences que l’on veut évaluer (cf.
  1.4). Un soin particulier est apporté à ce que chaque compétence jugée
  importante soit évaluée par un ou plusieurs exercices. Un autre aspect qui fait
  différer sensiblement ces protocoles des évaluations ordinaires est la recherche
  d’une standardisation la plus poussée possible. Tout cahier d’exercices doit
  pouvoir être corrigé de la même façon par n’importe quel correcteur (qu’on
  suppose tout au moins spécialiste du domaine évalué). Cela explique peut-être
  pourquoi les protocoles comportent souvent des QCM et pourquoi certains
  domaines sont rarement concernés (expression écrite et a fortiori orale).

- Vient ensuite une phase plus habituelle de recueil de l’information, passant par
  la constitution d’un échantillon (en utilisant souvent le degré « école » ou «
  classe ») l’envoi des questionnaires, leur passation, la correction (souvent par les
  professeurs des élèves interrogés), le retour des questionnaires, leur saisie
  (parfois par lecture optique) et les premiers traitements statistiques
  indispensables (« nettoyage » des fichiers, repondération…). Les enquêtes à
  domicile présentent, quant à elles, des difficultés particulières.

- C’est alors la phase la plus spécifique de ce type d’enquête : l’élaboration
  d’indicateurs de résultats. On a recours à des modélisations assez complexes,
  que l’on présentera dans les parties suivantes.

- Enfin, une fois les indicateurs de résultats construits, il ne reste plus qu’à les
  étudier dans des directions variées (voir chapitre 1.2.)

Cette présentation sommaire de la façon dont on élabore une évaluation de
compétences met en évidence la diversité des acteurs. Parmi les plus importants, on
peut citer : les pédagogues à l’origine des protocoles, les psychométriciens élaborant
la mesure des compétences et les statisticiens étudiant les résultats.

INSEE Méthodes 209
1.2. Utilité des évaluations de compétences

L’analyse de ces évaluations peut se faire sous des angles très divers. Nous allons présenter différentes voies possibles en nous attardant davantage sur celles qui seront abordées dans la suite.

1) **Analyses pédagogiques** : même si nous n’insisterons pas sur cet aspect, il faut se souvenir que les résultats publiés ont peu de sens si l’on ne revient pas au protocole d’évaluation. On ne peut se contenter d’une moyenne en affirmant qu’on mesure la « lecture » ou les « mathématiques ». L’analyse par sous-domaines, voire par exercices doit être menée, généralement par les personnes à l’origine des questionnaires, qui pourront à la lumière de leurs attentes permettre de dire si les résultats sont satisfaisants ou non. Ce type d’études peut conduire les décideurs à changer les programmes, pour que soit mis l’accent sur un domaine qui est apparu peu maîtrisé.

2) **Analyses statistiques** : on secontente ici d’un indicateur très synthétique de résultats pour étudier les liens qu’il entretient avec d’autres variables. Cette analyse peut se faire à deux niveaux :

   a) **Niveau individuel** : on étudie de façon plus ou moins fine, dans le cadre d’un échantillon donné, les écarts de résultats entre individus. Cela peut consister à calculer les différences de scores moyens entre plusieurs groupes (écarts entre filles et garçons, entre ouvriers et cadres...). Pour aller plus loin, on a souvent recours aux techniques économétriques, pour isoler les effets propres de chaque caractéristique. D’autres pistes d’études à l’échelle individuelle peuvent être envisagées : lien entre compétence et parcours scolaire, mesure de la progression en procédant à deux évaluations, au début et en fin d’année, études docimologiques (comparaison des résultats à une évaluation standardisée et des notes données par le professeur)... Les évaluations de compétence permettent d’éclairer d’une façon originale le débat sur le fonctionnement du système éducatif.

Cependant, ce débat s’est déjà beaucoup développé sans avoir recours aux comparaisons de compétences. On étudie souvent les inégalités devant l’école à l’aide des différences de parcours scolaires (obtention de tel ou tel diplôme, redoublement, etc.). La netteté des résultats généralement dégagés conduit à se demander ce qu’apporte l’usage des comparaisons de compétences.

En fait, la plupart du temps, on peut regretter que les indicateurs de parcours scolaire donnent une vision trop schématique de la réussite ou de l’échec : la population ne se ramènera pas à deux groupes, les « bons » et les « mauvais », ceux qui sont en retard et les autres. De même, dans une perspective diachronique, on évoque souvent la dévaluation des diplômes ce qui remet en cause l’usage de cette donnée comme mesure absolue de la
réussite scolaire. Enfin, la question se pose de façon encore plus cruciale quand on s’intéresse aux élèves des écoles primaires car pour mettre en évidence des inégalités sociales, encore faut-il mesurer des écarts ! Or, au début de l’école primaire, il n’y a pas de diplôme, pas encore de retard significatif et pourtant il serait audacieux d’affirmer qu’il n’existe encore aucun écart entre enfants. C’est là sans doute la grande utilité des évaluations de compétences que de permettre d’étudier les écarts sociaux alors qu’ils ne sont pas encore matérialisés en divergences de parcours scolaires. Il va sans dire que ces considérations ne remettent pas en cause les résultats obtenus jusqu’à présent en utilisant l’accès à tel diplôme ou telle formation comme indicateur de réussite scolaire. L’ampleur des écarts qu’ils permettent de mettre en évidence, malgré leur caractère un peu schématique, justifie leur usage. De plus, ces variables ont une signification claire et concrète (on peut savoir parfaitement qui a son bac ou non), ce qui, on le verra, n’est pas forcément le cas des mesures psychométriques.

b) **Niveau agrégé**: on ne retient alors qu’une valeur très synthétique calculée sur l’ensemble de l’échantillon (la moyenne, l’écart-type ou l’écart entre deux groupes caractéristiques) et on compare cette valeur à celle que l’on trouve sur un autre population à une évaluation identique. On peut ainsi envisager différents types de comparaisons :

i) **Comparaisons temporelles** : on compare l’indicateur avec la valeur obtenue autrefois, pour déterminer si la situation s’est améliorée ou pas.

ii) **Comparaisons internationales** : on compare l’indicateur avec ce que l’on trouve dans d’autres pays pour situer le système éducatif français et chercher des pistes auprès des pays les meilleurs. Comme dans le cas des comparaisons temporelles d’ailleurs, les analyses peuvent se ramener à deux grandes questions :

(1) *Étude des moyennes*. Le niveau est-il bon ? la moyenne est-elle élevée ?

(2) *Étude des écarts-type*. Les inégalités sont-elles importantes ? L’écart-type est-il élevé ou l’écart entre enfants de cadres et enfants d’ouvriers est-il élevé ?

Ce sont ces deux questions et surtout la première qui vont nous préoccuper ici, dans le cadre particulier des enquêtes internationales.
1.3. Les enquêtes internationales

Les récentes enquêtes internationales à grande échelle concernent généralement les mathématiques, les sciences ou la lecture. Le processus d’élaboration des questionnaires et de traitement des données est sensiblement différent de ce qui se fait en France (par exemple à la Direction de la Programmation et du Développement) sur deux points importants : d’une part, les concepteurs des évaluations ont souvent recours à la technique des cahiers tournants (l’ensemble des items conçus, généralement assez large, est découpé en parties non disjointes d’une quarantaine de questions ; chaque individu ne passera qu’un seul de ces sous-ensembles) ; d’autre part, les responsables de ces enquêtes font souvent appel à des techniques spécifiques de traitement des données. Enfin, le souci d’établir des comparaisons équitables impose une expertise poussée des procédures d’échantillonnage et de calcul des résultats.

Dans la suite, nous allons tirer la plupart de nos exemples des deux dernières enquêtes effectuées par les organismes internationaux. Il convient donc de les présenter rapidement.

1.3.1. TIMSS (Third International Mathematics and Science Study)

Plus de quarante pays ont participé à cette enquête sur les mathématiques et les sciences, organisée par l’IEA (International Association for the Evaluation of Educational Achievement) pendant l’année scolaire 1994/1995. Trois populations différentes étaient visées : la population 1 regroupant les élèves de 9 ans (le plus souvent en 5ème et 4ème années de scolarité), la population 2 les élèves de 13 ans (en 7ème et 8ème années) et la population 3 les élèves de fin d’études secondaires.

Un test différent était administré à chacune des trois populations. L’évaluation portait sur plusieurs sous-domaines comme la géométrie ou les probabilités pour les mathématiques ou bien la physique ou la chimie pour les sciences. Cependant, les concepteurs n’ont conservé que deux échelles : une pour les mathématiques et l’autre pour les sciences, afin de publier un classement général pour chacun de ces deux domaines (même si des résultats par sous-domaines, voire par items, sont aussi présentés).

1.3.2. IALS (International Adult Literacy Survey)

Cette enquête a eu lieu la première fois en 1994. Elle concernait les individus âgés de 16 à 65 ans de huit pays. En 1997, on comptait au total environ vingt-cinq pays de l’OCDE ayant participé aux différentes vagues de l’enquête. L’Office national de statistique canadien (Statistics Canada) et l’institut privé américain ETS (Educationnal Testing Service) ont mené cette étude en collaboration avec l’OCDE.
Le test, inspiré de l’enquête américaine NALS (National Adult Literacy Survey), visait à mesurer les capacités de lecture et de compréhension de documents de la vie courante. Partant du concept de « littératie », l’évaluation concernait trois domaines : la compréhension de textes suivis, la compréhension de textes schématiques et la compréhension de textes quantitatifs. Pour chacun de ces domaines, les concepteurs ont défini cinq niveaux de compétences. En général, les publications donnent pour chaque domaine la répartition des personnes interrogées, selon ces niveaux. Il n’y a pas d’analyse pédagogique plus détaillée.

Il est important de souligner qu’il s’agissait là, contrairement à TIMSS, d’une enquête portant sur une population d’adultes interrogés à leur domicile. Ce public, ayant pour une part quitté le système scolaire depuis longtemps, se montre sans doute moins réceptif à un questionnement sur leurs compétences. De plus, il va de soi que l’on ne cherche pas à mesurer des compétences scolaires, mais alors, que veut-on évaluer ?

1.4. Les compétences en question

L’évaluation statistique de compétences amène à s’interroger de façon marquée sur ce que l’on veut mesurer. Les intitulés des enquêtes font référence aux mathématiques, aux sciences ou à la littératie. Les responsables se proposent de donner une image des compétences d’une population dans ces domaines. Ceci amène plusieurs questions. On peut tout d’abord se demander si la compétence en mathématiques ou en lecture existe ou si au contraire on ne mesure que des performances à un test (performances que l’on juge plus ou moins représentatives des tâches que l’individu peut être amené à remplir). Et même si l’on accepte, au vu de l’analyse statistique, l’existence de facteurs latents, rien n’assure qu’on puisse les réduire à un seul. Dans le cas des mathématiques, on verra que l’on peut distinguer compétences en algèbre et en géométrie. La question est de savoir jusqu’à quel niveau de détail on doit aller et s’il faut supposer une dimension différente pour chaque processus mental (il y aurait une dimension pour l’addition, une pour la multiplication, une pour le repérage d’un mot, une pour le repérage d’une lettre, etc.). Cette préoccupation est extrêmement importante dans le domaine scolaire, où du fait de l’étendue des programmes, on pose souvent des exercices sur des sujets très variés (elle est beaucoup moins forte dans le cas des tests de QI, où les items recourent à des raisonnements très proches voire quasi identiques).

On peut aussi s’interroger sur le caractère universel de la dimension mesurée. La lecture en français et la lecture en japonais relèvent-elles de la même compétence ? Même une discipline comme les « sciences » recouvre des définitions et des contenus d’enseignement très différents d’un pays à l’autre.

Ce problème est encore plus délicat lorsqu’on travaille sur des populations adultes (cas de l’enquête IALS). On peut s’intéresser aux compétences qui sont utiles sur le
marché du travail mais on se heurte à une nouvelle difficulté : est-ce qu’un jeune homme ou une personne mûre ont besoin des mêmes compétences pour trouver un emploi ? Est-ce que ce sont les mêmes compétences qui facilitent l’insertion professionnelle pour un homme ou une femme, en France et au Chili ?

Un dernier point mérite d’être évoqué concernant les difficultés de mesure. Il peut paraître plus facile de repérer les individus qui ne maîtrisent pas les compétences de base pour effectuer convenablement les actes de leur vie quotidienne que de distinguer chez les personnes n’ayant pas de difficultés, celles qui sont très compétentes de celles qui le sont un peu moins. En particulier, on peut penser que la hiérarchie obtenue sera très sensible au domaine évalué. Pour reprendre la problématique du paragraphe précédent, une fois que l’on a vu que tel individu avait les bases en lecture et en calcul pour s’insérer convenablement dans la société, faut-il voir s’il est Shakespeare ou cordonnier ?

Ce problème concernant la dimensionnalité du domaine d’intérêt est fondamental car la constitution du questionnaire et les traitements statistiques à effectuer en dépendent. Cependant, le discuter en détail conduirait à aborder les théories psychologiques, sociologiques et philosophiques, autant que techniques. Nous nous placerons donc par la suite dans le cadre dominant des évaluations internationales, qui suppose l’unidimensionnalité du domaine évalué, en évoquant parfois les quelques alternatives qui sont proposées.
2. Eléments de psychométrie

Avant de commenter les résultats issus des évaluations internationales, il paraît souhaitable de présenter leurs fondements théoriques. On insistera plus particulièrement sur les Modèles de Réponse à l’Item (MRI), dont l’emploi est largement répandu dans les organismes internationaux.

2.1. Présentation rapide et historique

La tentation de mettre l’âme humaine en chiffre est très ancienne. Au siècle dernier, on a cherché dans les caractéristiques du cerveau, en particulier son volume, l’origine de l’intelligence, avec souvent l’intention plus ou moins explicite de justifier les inégalités sociales, raciales et sexuelles. Dans un autre état d’esprit, les premiers tests d’intelligence élaborés par Binet étaient destinés à repérer les enfants en difficultés pour leur apporter l’aide nécessaire. Ils sont toutefois assez rapidement devenus des outils au service d’une vision plus conservatrice, voire eugéniste, de la société. Pour cette raison et à cause d’un certain manque de fondement du point de vue de la théorie psychologique, ils ont été assez contestés avant de retrouver un certain crédit.

Du point de vue technique, l’étude des réponses aux tests d’intelligence a donné lieu à une théorie assez développée, partant le plus souvent d’une approche probabiliste. Les outils d’analyses de données ont ainsi vu le jour, en grande partie pour alimenter le débat sur la structure factorielle de l’intelligence (un seul facteur, appelé g, ou plusieurs comme celui d’intelligence verbale, etc.). Dans le domaine des évaluations scolaires, la dernière décennie a vu le développement rapide des Modèles de Réponse à L’Item (MRI) très souvent employés dans le cadre des comparaisons internationales.

Le petit exposé qui va suivre ne prétend pas être un cours complet et structuré de psychométrie mais donne quelques éléments indispensables à la compréhension des résultats publiés dans le domaine de l’évaluation des compétences.

2.2. La théorie de la mesure

2.2.1. Les différents « niveaux » de mesure

Les manuels de psychométrie commencent généralement par des considérations assez détaillées sur la nature de la mesure. Les psychométriciens ne se contentent pas de la distinction entre variables qualitatives et variables quantitatives mais ont élaboré une théorie de la mesure beaucoup plus formalisée.

On considérera comme mesure un ensemble de symboles, généralement chiffrés, représentant des faits empiriques. Les différents types de mesure doivent respecter quelques propriétés de base (définition d’une relation d’identité symétrique et
transitive) et se distinguent entre elles en fonction des propriétés supplémentaires qu’elles possèdent. On les classe ainsi en différents « niveaux », que l’on peut aussi définir en donnant les transformations applicables sans perte d’information. Ces notions permettent ensuite de définir les statistiques appropriées. Quatre échelles principales sont généralement présentées :

- l’échelle **nominale**, où la mesure est un simple ensemble d’étiquettes. Toute bijection de l’ensemble d’étiquettes dans un autre donne une mesure équivalente (en revanche, une transformation agrégeant certains cas donne une mesure non équivalente). Exemples : la nomenclature de professions INSEE.

- l’échelle **ordinale**, qui impose l’existence d’une relation d’ordre. On obtient une mesure équivalente si la mesure finale respecte une relation d’ordre et que deux objets classés dans un certain ordre pour la première mesure le sont pour la seconde (la transformation doit donc être monotone stricte). Exemples : les classes d’âge, les niveaux scolaires, les classements sportifs...

- l’échelle **d’intervalles**, où l’on arrive à une notion plus proche de ce que l’on entend par mesure, c’est-à-dire une donnée chiffrée sur laquelle on peut effectuer des opérations arithmétiques. La propriété fondamentale de ce niveau de mesure est que les écarts entre les valeurs ont une signification. Un écart de 10 unités a la même importance si le niveau est 20, -350 ou 2000 : il est 2 fois plus grand qu’un écart de 5. Ce rapport entre les écarts doit se retrouver dans toutes les mesures équivalentes ce qui implique que l’on n’obtient une mesure équivalente que par une transformation affine sur les nombres. Exemples : les températures °C ou °F, la position sur un axe de coordonnées, la plupart des variables psychologiques...

- l’échelle de **rapport**, où non seulement les écarts, mais aussi les rapports, doivent être comparables. Cela impose l’existence d’un 0 absolu. Seule la multiplication par une constante permet de passer du mètre à l’inch, de la livre au kilo, de la seconde à l’année, de l’euro au dollar.

### 2.2.2. Le choix des indicateurs

Les précédentes distinctions ne sont pas simplement d’ordre théorique mais ont des implications pratiques fortes : il est indispensable de s’interroger sur la nature de la mesure quand on souhaite construire des indicateurs de résultats.

Il y a plusieurs façons de voir si la mesure sur laquelle on travaille autorise ou non une statistique. On peut par exemple s’intéresser à la façon dont la statistique est obtenue et l’interdire si elle utilise des opérations non autorisées : par exemple, en toute rigueur, on ne peut pas calculer de moyenne sur des données ordinales puisque l’addition n’est pas possible ; on ne peut pas non plus calculer de rapport interdéciles sur des échelles d’intervalles.
Une autre méthode consiste à déterminer si l’indicateur donne les mêmes résultats sur des échelles équivalentes, obtenues par les transformations possibles. Nous allons prendre deux exemples : celui de l’écart-type et du coefficient de variation. Ces indicateurs sont-ils valides sur les échelles d’intervalles et sur les échelles de rapport ?

On passe d’une échelle d’intervalles $E_{int}$ de mesure $M_{int}$ à une échelle d’intervalles équivalente $E_{int'}$ de mesure $M_{int'}$ par une transformation linéaire $M_{int'}=a\times M_{int}+b$ (avec $a$ différent de 0, de préférence strictement positif). Pour une échelle de rapport $E_{rap}$ de mesure $M_{rap}$, la transformation est $M_{rap'}=a\times M_{rap}$ (avec $a$ différent de 0).

Voyons si l’interprétation de deux écarts-type est affectée par le changement de mesure. Si deux populations ont sur l’échelle $E_{int}$ des écarts-type $s_1$ et $s_2$ tels que $s_1=K \times s_2$ alors sur l’échelle $E_{int'}$, $s_1'=a \times s_1$, $s_2'=a \times s_2$ et l’on retrouve $s_1'=K \times s_2'$. On voit immédiatement que les transformations permises pour les échelles de rapports respectent aussi cette propriété. L’ordre des dispersions pour l’une des mesures est le même pour l’autre.

En ce qui concerne le coefficient de variation commençons par les échelles de rapports. Si deux populations ont sur l’échelle $E_{rap}$ des coefficients de variation $CV_1=s_1/m_1$ et $CV_2=s_2/m_2$, en passant à l’échelle équivalente $E_{rap'}$, on obtient $CV_1'=s_1'/m_1'(a \times s_1)/(a \times m_1)=s_1/m_1=CV_1$. Le coefficient de variation n’est pas affecté par un changement d’échelle (c’est ce qui explique sa popularité). Les interprétations sont donc équivalentes d’une échelle à l’autre.

En revanche si l’on travaille sur des échelles d’intervalles, la transformation est plus compliquée et donne comme équivalent $CV_1'=s_1'/m_1'=(a \times s_1)/(a \times m_1+b)$ et $CV_2'=s_2'/m_2'=(a \times s_2)/(a \times m_2+b)$. On ne voit pas clairement l’équivalence des deux échelles. Plutôt que d’effectuer un travail d’analyse sur la fonction $f(x,y)=(ax+y)/(ax+b)$, nous allons prendre un petit exemple. Soient deux séries de température en °F qui donnent pour l’une : une moyenne de 50°F et un écart-type de 5°F et pour l’autre une moyenne de 80°F et un écart-type de 10°F. Les coefficients de variation sont 0,1 et 0,125 : la deuxième série paraît pour cet indicateur plus dispersée. Si nous convertirions toutes ces températures en °C (°C=5/9 °F-160/9) les moyennes deviennent respectivement 10 °C et 26,67 °C, les écarts-type 2,78 °C et 5,56 °C et les coefficients de variations 2,78 et 2,08 : c’est la première série qui paraît alors la plus dispersée ! Il paraît donc difficile de donner un jugement fiable à partir de cet indicateur. On note en revanche que l’examen des écarts-type sont concordants. Cela ne signifie pas que cet indicateur soit vraiment pertinent mais il respecte au moins la clause de base de tout jugement scientifique, la non-contradiction.

INSEE Méthodes 217
2.3. Indicateurs classiques

Les indicateurs les plus simples que l'on puisse construire à partir d'une épreuve composée de N questions proposées à un échantillon de K individus sont les suivants :

- On mesure le niveau de l'individu par un score, la proportion de bonnes réponses, c'est-à-dire le nombre de bonnes réponses divisé par le nombre total de questions.

- On mesure la difficulté d'un item par la proportion d'individus qui le réussissent.

- Pour les items, on définit généralement une autre caractéristique : sa discrimination. Il s'agit de la corrélation, calculée sur l'ensemble des individus, entre la réussite à l'item et le score calculé sur la totalité des items (ou très souvent en enlevant l'item en question). Plusieurs indicateurs peuvent être proposés. Le plus simple mais pas forcément le meilleur est tout simplement le coefficient de corrélation linéaire entre le score et l'indicateur de réussite à l'item. La discrimination indique dans quelle mesure la réussite à tel item permet de prédire un score élevé. Dans l'absolu, les items les plus discriminants sont les meilleurs (car ils reflètent bien la dimension évaluée par l'ensemble de l'épreuve).

Il convient d'ajouter à cela quelques notions permettant de juger de la qualité globale d'une épreuve. On a généralement recours à deux types d'indicateurs :

- Indicateurs de fidélité : le calcul d'un score revient donc dans le cas le plus simple à faire la somme des indicatrices de réussite aux items. Comme dans tous les cas où l'on additionne des données, il faut vérifier qu'aucune carotte ne s'est glissée parmi les choux. En d'autres termes, la sommation n'est pertinente que si tous les items mesurent la même chose. Ceci se teste de plusieurs façons qui reviennent toutes à étudier les corrélations entre items. Si les corrélations entre items sont élevées, quand réussir à l'un accroît significativement les chances de réussir à un autre, on considère que l'épreuve est fidèle ou qu'elle a une bonne consistance interne. On verra plus loin les méthodes utilisées pour évaluer cette fidélité quand on cherchera à évaluer l'unidimensionnalité de l'épreuve.

- Indicateurs de validité : pour s'assurer que l'on a une bonne épreuve, une autre méthode consiste à comparer les résultats qu'elle donne soit avec une donnée connue et censée être liée à la compétence (par exemple, le redoublement dans le cas d'une évaluation scolaire) soit avec une autre épreuve faisant autorité. Les indicateurs de validité donnent donc une idée de la valeur prédictive de l'épreuve.
Tous ces indicateurs ont le mérite de la simplicité mais ils ont quelques inconvénients. Tout d’abord, ils ont les défauts de tous les pourcentages, c’est-à-dire que leur « comportement sur les bords » devient particulier. Une population ayant un score moyen élevé, proche de 100 % (ou faible et proche de 0) risque de présenter des écarts entre individus assez faibles par rapport à une population dont le taux de réussite est proche de 50 %. Or ce résultat sera certainement inversé si l’on change d’épreuve, en en donnant une plus difficile, pour faire en sorte que le taux de réussite de la première population s’approche de 50 %, alors que celui de la deuxième tende vers 0. Pour reprendre les termes de la théorie de la mesure que nous avons développée, il est à craindre que les scores en % ne suivent ni une échelle de rapport, ni une échelle d’intervalle, quand on s’écarte trop de 50 %

Les indicateurs posent d’autres problèmes quand on veut les comparer entre eux. Il n’est pas possible de dire que telle population est moins bonne que telle autre, parce qu’elle a un score moyen inférieur. Il faut que chacune des populations ait été soumise une évaluation identique. Or dans bien des cas, c’est une contrainte très forte.

---

4 Ce résultat ne semblera pas singulier à ceux qui ont suivi dans la Revue Française de Sociologie le débat sur la comparaison des taux d’accès. Si le taux d’accès des enfants d’ouvriers passe de 1 % à 10 % tandis que celui des enfants de cadres passe de 39 % à 51 %, peut-on dire que les inégalités ont diminué ? Doit-on faire la différence des taux, leur rapport, étudier les taux de variation ?

Là aussi, on voit que toutes ces questions amènent à se demander quel est le niveau de mesure du taux d’accès. Est-ce une mesure d’intervalle ou de rapport ? En fait, les chercheurs semblent considérer qu’elle n’est véritablement ni l’une ni l’autre mais simplement ordinale. Son échelle naturelle serait logistique : il faudrait donc appliquer une transformation du type log(p/(1-p)) pour aboutir à une échelle d’intervalle. Cependant, on s’accorde aussi à dire que si les pourcentages restent dans des valeurs moyennes (entre 25 et 75 %) on peut les considérer comme relevant directement d’une mesure d’intervalle. La conclusion pratique que l’on peut tirer de l’examen des articles abordant ce problème est que les indicateurs logistiques sont sans doute les plus justes mais qu’ils doivent surtout servir à vérifier les commentaires portant sur des indicateurs plus simples. La lisibilité ne doit pas être sacrificée à un trop grand souci de rigueur. De plus, même si l’indicateur logistique conduit à dire qu’un écart de taux d’accès au baccalauréat de 92 % à 99 % est statistiquement énorme, en pratique, c’est une inégalité relativement peu apparente, même si cela peut suggérer que le problème des inégalités n’est pas résolu et risque de réapparaître à un niveau plus sélectif du système éducatif.
3. Les modèles de réponse à l’item

3.1. Présentation

Les modèles de réponse à l’item (MRI ou IRT en anglais) proposent d’expliquer la réussite à un item en fonction de la compétence de l’individu et de l’item lui-même. La probabilité pour un individu j de répondre à la question i sera notée :

$$\Pr(x_{ij} = 1) = F(\theta_j, \beta_i)$$

où $x_{ij}$ est la réponse de j à la question i, $\theta_j$ la compétence de l’individu j, $\beta_i$ le vecteur caractéristique de l’item i et $F$ la fonction de lien$^5$.

De tels modèles présentent l’avantage de séparer les concepts, puisque compétences individuelles et caractéristiques des items sont définies de façon totalement indépendante$^6$. On verra dans la partie 3.2. les hypothèses sous-jacentes à ce modèle.

On remarque une certaine analogie avec les modèles économétriques de panels : les différentes questions posées à un individu peuvent être assimilées à des prises d’informations successives. La compétence $\theta_j$ correspondrait à « l’effet fixe » ou effet individuel dans ces modèles.

3.1.1. Le modèle de Rasch

Il existe plusieurs MRI suivant le nombre de paramètres que l’on utilise pour décrire l’item. Historiquement, le premier est le modèle à un paramètre, ou modèle de Rasch (1960) :

$$\Pr(x_{ij} = 1|\theta_j, b_i) = \frac{1}{1 + e^{(b_i-\theta_j)}}$$

où $b_i$ est la difficulté de l’item i.

---

$^5$ On se limitera ici à la fonction logistique mais d’autres choix sont possibles.

$^6$ Sur ce point, les psychométriciens tombent parfois dans l’ambiguïté : certains transforment l’hypothèse (i.e. la relation proposée par le modèle et ses implications sur les propriétés requises par les données) en fait imposé par la nature du modèle. Par exemple, certains auteurs affirment que les paramètres de difficulté peuvent être calibrés sur n’importe quel échantillon puisqu’ils ne dépendent pas du groupe d’individus sur lesquels ils sont calculés. Cet avantage existe mais il est théorique : il faut avant tout qu’il y ait adéquation entre le modèle et les données.
Une même échelle latente s’applique donc aux individus et aux questions, ce qui fait tout l’intérêt du modèle. \( (b_i - \theta_j) \) représente la différence de niveaux entre les deux : si \( (b_i - \theta_j) < 0 \), la probabilité de réussite sera supérieure à 0.5, inférieure sinon.

Ce modèle fait l’hypothèse que la dépendance entre la probabilité de réussir un item et la compétence des individus peut être représentée comme sur le graphique 1. Les courbes des 3 items pris en exemple sont toutes parallèles. La difficultés de l’item correspond au point d’inflexion de la courbe. On voit que c’est aussi la valeur de la compétence des individus qui ont autant de chance de réussir l’item que de le rater.

**Graphique 1 : modèle à un paramètre (ou modèle de Rasch)**

![Graphique 1](image)

**3.1.2. Le modèle à deux paramètres**

Pour affiner le modèle, on peut ajouter une caractéristique à l’item :

\[
Pr(x_{ij} = 1 | \theta_i, a_i, b_i) = \frac{1}{1 + e^{-Dna_i(\theta_i - b_i)}}
\]

où \( a_i \) est le coefficient de discrimination de la question i et D un facteur d’échelonnement permettant le passage à la fonction de lien ogive normale (constante égale à 1,7).

Si \( a_i \) est grand, \( (\theta - b) \) sera « amplifié » : une personne ayant une compétence plus forte que la difficulté de la question réussira très probablement la question tandis qu’une personne de compétence moindre n’aura que très peu de chances d’y répondre avec succès. A l’extrême inverse, si ce coefficient est nul, la probabilité de réussite sera constante quelle que soit la compétence de l’individu. C’est pourquoi \( a_i \) est appelé coefficient de discrimination. Le graphique 2 présente les courbes caractéristiques théoriques de quelques items. On voit que contrairement au cas précédent, elles ne sont plus parallèles et peuvent se croiser.
3.1.3. Le modèle à trois paramètres

Certains items peuvent être réussis au hasard, comme les questions à choix multiples. L’ajout d’un facteur de chance $c_i$ peut dans ce cas s’avérer pertinent. Le modèle devient alors

$$\Pr(x_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-D_{a_i}(\theta_j, -b_i)}}$$

Ainsi, on considère que la probabilité de réussir l’item est supérieure à $c_i$, quelle que soit la compétence de l’individu. Si par exemple la réponse à la question est oui ou non, on pourra poser $c_i=0.5$ puisque même si on ne connaît absolument pas la réponse, on a une chance sur deux de répondre correctement. Mais l’ajout de ce paramètre n’est pas toujours si simple car si la question posée est une question « piège », il se peut très bien que l’on observe des taux de réussite inférieurs à 50 %, en contradiction avec le modèle.
3.2. Hypothèses des modèles

En théorie, les modèles de réponse à l’item présentent des avantages très intéressants pour les constructeurs de tests ou d’évaluations. Cependant, ils reposent sur des hypothèses fortes, qu’il est nécessaire de vérifier avant toute utilisation.

3.2.1. L’hypothèse d’unidimensionnalité

Les performances des individus aux items ne sont mesurées que par un seul paramètre. On postule donc que la réussite aux items est conditionnée par un seul et même facteur. Dans la pratique, cette hypothèse n’est évidemment pas totalement remplie vu la complexité des processus mentaux permettant la réussite aux items. De plus, il n’existe pas de méthode parfaite pour valider cette hypothèse car la dimensionnalité des items est un concept difficilement mesurable.

Deux approches ont cours. La première vise à mesurer la cohérence globale : on aura alors recours aux analyses factorielles ou aux α de Cronbach. La deuxième analyse le fonctionnement différentiel des items.

Les analyses factorielles

En général, on considère qu’une évaluation est unidimensionnelle si l’ensemble des items s’ajuste convenablement au premier facteur d’une analyse factorielle. De manière pratique, on compare l’inertie expliquée par le premier facteur à celle expliquée par le second. On considère en général que l’hypothèse est vérifiée lorsque le rapport des deux est supérieur à 4.

Le coefficient α de Cronbach

Ce coefficient est employé pour évaluer la consistance interne de l’épreuve. Dans le cas où tous les items sont dichotomiques, il est défini de la manière suivante:

\[ \alpha = \frac{N}{N - 1} \left( 1 - \sum_{i=1}^{N} p_i (1 - p_i) \right) / \sigma^2 \]

où \( p_i \) est la proportion de réussite à l’item i et \( \sigma^2 \) est la variance de l’épreuve (c’est-à-dire la variance des proportions de bonnes réponses obtenues par individu).

---

7 Cela ne signifie pas que l’évaluation entière repose sur une seule échelle. Dans IALS, par exemple, 3 axes de compétence sont définis. On ne s’intéressera donc ici qu’aux items mesurant la même dimension.
8 De manière plus générale, le α est défini comme la moyenne des corrélations entre toutes les sous-épreuves que l’on peut construire à partir des N items.
A partir d'une valeur de 0.8, on considère que l'hypothèse est acceptable. Mais plusieurs mises en garde peuvent être faites (Cortina, 1993) : par exemple, le degré de corrélation entre les items ne mesure pas forcément l'homogénéité de l'épreuve, c'est une condition nécessaire mais pas suffisante. Ce coefficient dépend de l'écart-type et une forte valeur de α peut provoquer d'une forte disparité de niveaux entre les individus de l'échantillon et non pas de l'homogénéité des items. De plus, il augmente sensiblement avec le nombre d'items de l'épreuve.

**Le fonctionnement différentiel des items (FDI)**

- La statistique de Mantel-Haenszel.

On cherche ici à repérer les items dont la réussite ne seraient pas uniquement conditionnée par le score de l'individu. On teste ainsi l'influence, à compétence constante, des facteurs comme le sexe, l'âge, les diplômes sur les réponses à une question donnée. Pour cela, on peut utiliser la statistique de Mantel - Haenszel. On considère deux groupes (par exemple les hommes et les femmes), qu'on découpe en J classes de compétences équivalentes. Le tableau ci-dessous présente les effectifs de chaque catégorie :

<table>
<thead>
<tr>
<th>Niveau j</th>
<th>Groupe 1</th>
<th>Groupe 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A_j</td>
<td>B_j</td>
<td>n_{1j}</td>
</tr>
<tr>
<td></td>
<td>C_j</td>
<td>D_j</td>
<td>n_{2j}</td>
</tr>
<tr>
<td></td>
<td>m_{1j}</td>
<td>m_{0j}</td>
<td>T_j</td>
</tr>
</tbody>
</table>

La statistique de Mantel-Haenszel est donnée par :

\[ \text{MH} = \frac{\left( \sum_{j=1}^{J} \left( \frac{A_j - E(A_j)}{2} \right)^2 \right)}{\sum_{j=1}^{J} \text{Var}(A_j)} \]

où \( E(A_j) = \frac{n_{ij} m_{ij}}{T_j} \) et \( \text{Var}(A_j) = \frac{n_{ij} n_{2j} m_{ij} m_{2j}}{T_j^2 (T_j - 1)} \)

Sous l'hypothèse \( H_0 : A/B = C/D \) (la réussite à l'item est indépendante du groupe considéré), on a approximativement : \( \text{MH} \sim \chi^2 \)

224

*INSEE Méthodes*
• Le biais statistique.

Dans le même ordre d'idée, une autre façon de procéder est d'utiliser l'indice suivant, proposé par Lord :

\[ d_j = \frac{\Delta b_j}{\sqrt{S_{j,1}^2 + S_{j,2}^2}} \]

où \( \Delta b_j = b_{j,1} - b_{j,2} \) est la différence des estimations de la difficulté de l'item \( j \) pour les pays 1 et 2.

\( S_{j,k} \) est l'erreur type de l'estimation de \( b_{j,k} \)

\( d_j \) est appelé usuellement SIB (pour Statistical Index of Bias).

A priori, les deux populations ont des compétences moyennes différents. Afin de supprimer cet effet, on impose que la moyenne des \( b_j \) soit nulle pour chacun des deux groupes.

On teste l'hypothèse \( H_0 : d=0 \) contre \( H_1 : d \neq 0 \).

Sous \( H_0 \), \( d \sim N(0,1) \). Donc, pour un seuil de 5\%, si \(|d| > 1.96\), la question présente un fonctionnement différentiel significatif.

Notons que ces deux statistiques ont le défaut de dépendre de la taille de la population. Si le nombre d'individus est trop faible, \( H_0 \) est presque systématiquement acceptée. En d'autres termes, on risque d'affirmer que les items correspondent bien au modèle simplement parce que le nombre d'individus ne permet pas de mettre en évidence un écart statistique. Ce risque est particulièrement important quand on travaille sur des échantillons de taille modeste et en tenant compte de l'usage des cahiers tournants (chaque item n'est passé que par une partie de la population).

3.2.2. L'hypothèse d'indépendance locale

Cette hypothèse signifie qu'à un niveau de compétence donné, il n'existe pas de corrélation entre les réponses aux items. Elle rejoint le postulat d'unidimensionnalité dans la mesure où seule la compétence des individus explique la réussite aux items. Par exemple, une question intitulée comme ceci : « À partir des résultats de la question précédente... » violerait le principe d'indépendance locale. En pratique, pour vérifier cette hypothèse, on utilise le coefficient \( \phi^2 \).
Lorsque les items sont dichotomiques, il s’agit du coefficient de corrélation linéaire $R^2$ entre les variables indicatrices de deux items :

$$
\phi^2 = \frac{(P_{11}P_{00} - P_{10}P_{01})^2}{P_1(1 - P_1)P_2(1 - P_2)}
$$

où $p_{ij}$ est la proportion de réponses (i à l’item 1, j à l’item 2)

$p_1$ est la proportion de bonnes réponses à l’item 1

$p_2$ est la proportion de bonnes réponses à l’item 2

En pratique, on calcule la moyenne de tous les $\phi$ obtenus sur l’ensemble d’un questionnaire, pour un groupe d’individus de niveaux homogènes. Si les résultats sont proches de 0, l’hypothèse d’indépendance locale est validée.

### 3.3. Estimation

#### 3.3.1. Estimation des paramètres d’items

Les modèles MRI ont ceci de très particulier qu’ils s’appuient sur des variables explicatives inobservables a priori. En effet, même si les items sont élaborés par des spécialistes (professeurs, psychométriciens) suivant certains critères de difficulté (comme la lisibilité du document, le nombre de distracteurs, la quantité d’informations à relever...), les paramètres des items ne sont pas déterminés au départ mais estimés conjointement aux scores des individus.

On se place tout d’abord dans le cas où tout le monde a répondu aux mêmes items. Soit $X = \{x_{ij}\}_{1 \leq i \leq N, 1 \leq j \leq K}$ la matrice de réponses des K individus aux N questions. Sous l’hypothèse d’indépendance locale, la fonction de vraisemblance s’écrit $^9$

$$
L(X \mid \theta, \beta) = \prod_{j=1}^{K} \prod_{i=1}^{N} P_{ij}^{x_{ij}} (1 - P_{ij})^{1-x_{ij}}
$$

où $\theta$ est le vecteur de scores de l’ensemble des individus

$\beta$ est le vecteur des paramètres des items

$P_{ij} = P(x_{ij}=1 \mid \theta, \alpha)$

---

$^9$ L’autre hypothèse est bien sûr qu’à $\theta$ constant, $P_{ij}$ est constant quel que soit le pays d’origine.
Soit f la densité de θ considérée comme une variable aléatoire continue. On a donc :

\[ L(X, \theta | \beta) = \prod_{j=1}^{K} \prod_{i=1}^{N} P_{ij}^{x_{ij}} (1 - P_{ij})^{1-x_{ij}} f(\theta) \]

Et par conséquent :

\[ L(X | \beta) = \int \prod_{j=1}^{K} \prod_{i=1}^{N} P_{ij}^{x_{ij}} (1 - P_{ij})^{1-x_{ij}} f(\theta) d\theta \]

A ce stade, la solution la plus simple est de considérer f comme connue. On peut postuler, par exemple, que θ suit une loi normale centrée réduite. On estime alors directement β par maximum de vraisemblance.

3.3.2. Les cahiers tournants

Jusqu’à présent, on a considéré que tous les individus répondaient à un seul questionnaire. En réalité, dans les évaluations internationales, on utilise le principe des cahiers tournants. Ainsi, dans IALS par exemple, les 114 items de départs sont répartis en sept blocs et ces blocs sont regroupés par trois dans sept questionnaires : le premier questionnaire, ainsi, est formé des blocs 1, 2 et 4 et le deuxième des blocs 2, 3 et 5. On a donc la possibilité de faire passer un plus grand nombre d’items sans alourdir le temps de passation, et d’obtenir plus de résultats item par item.

Par ailleurs, l’influence d’une question inadquate est réduite, puisque, dans notre exemple, 3/7 de la population seulement y répond. En contrepartie, il faut s’assurer que les items des différents blocs sont sur une échelle commune, ne serait-ce que pour que les individus soient également sur une échelle commune. Par exemple, une simple proportion de bonnes réponses ne conviendrait pas pour de tels questionnaires car si l’un des blocs était plus difficile que les autres, les scores obtenus seraient inférieurs à la moyenne sans pour autant refléter une faiblesse générale de cette sous-population. Le modèle MRI s’adapte quant à lui relativement bien à cette contrainte dans la mesure où il tient déjà compte des caractéristiques des questions. Un seul problème subsiste en fait : l’indétérmination du modèle. En effet, en posant, pour le modèle à deux paramètres :

\[ a_i^* = a_i / k_2, \ b_i^* = k_1 + k_2 b_i \]
\[ \theta_j^* = k_1 + k_2 \theta_j \]

on obtient :

\[ P \left( x_{ij} = 1 \mid \theta_j, a_i, b_i \right) = P \left( x_{ij} = 1 \mid \theta_j^*, a_i^*, b_i^* \right) \]

Ainsi, deux estimations successives de a et b peuvent donner des résultats différents. Il existe plusieurs techniques pour relier les paramètres entre eux, suivant le type de méthode de cahiers tournants utilisés. Si au moins un groupe d’items est commun (cas de IALS), on peut par exemple effectuer une estimation simultanée de tous les items et standardiser ensuite soit les b_i, soit les θ_j obtenus.
3.3.3. Estimation des scores : la méthode des valeurs plausibles

Une fois déterminé $\beta$, on peut maximiser en $\theta$ la fonction de vraisemblance $L(X|\theta, \beta)$. Cette solution est la plus simple mais possède l’inconvénient de donner des résultats infinis si la personne a répondu juste ou faux à toutes les questions. On peut contourner cet obstacle en utilisant une estimation bayésienne. Si la moyenne des estimateurs est sans biais par rapport à la moyenne de $f$, l’écart-type des estimateurs est en général plus petit que celui de $f$, ce qui peut poser des problèmes de comparaison de populations.

Une autre solution souvent retenue est la méthode des valeurs plausibles. L’objectif est de fournir des estimations optimales à l’échelle des populations et non des individus. Cela permet également d’obtenir des résultats continus et non discrets (avec la méthode de maximum de vraisemblance, par exemple, on obtient un nombre fini de $\theta$ différents). Pour ce faire, on tire au hasard une valeur de $\theta$ dans sa distribution a posteriori (c’est-à-dire la distribution de $\theta$ étant donné les résultats aux items). On s’appuie sur la relation bayésienne suivante :

\[
P(\theta | x) = C \frac{p(x | \theta) p(\theta)}{(1)}
\]

où C est une constante.

Les paramètres des items étant déterminés, $P(x | \theta)$ est connu. Reste à fixer la distribution a priori. On considère par exemple : $\theta \sim N(0, 1)^{10}$. De par la complexité de la vraisemblance, on peut supposer également que la distribution a posteriori est normale, et on calcule son espérance et sa variance à partir de (1). On tire ensuite un $\theta$ dans cette distribution. Très clairement, l’estimateur obtenu n’est pas optimal pour les individus puisqu’on a une probabilité non nulle de tirer un $\theta$ très faible pour un individu ayant réussi entièrement le questionnaire, et inversement.

\[10\] On pourra toujours contester cette hypothèse de normalité. La symétrie des compétences, par exemple, semble hasardeuse : peut-on dire qu’il y a autant de Proust que d’illettrés ? Souvent, on préfère se limiter à des populations a priori plus homogènes (et, on l’espère, davantage « normales »). Cela revient à introduire des caractéristiques socio-démographiques dans l’équation (1). Ce raffinement complique singulièrement les procédures d’estimation (NCES, [98]).
3.4. Relation avec la proportion de bonnes réponses

La proportion de bonnes réponses constitue sans doute l’estimateur le plus naturel de réussite d’un individu à une évaluation. On va chercher à savoir en quoi la compétence estimée par le modèle de réponse à l’item est « supérieure » à cette statistique élémentaire. Pour comprendre le lien entre ces deux approches a priori divergentes, analysons le modèle suivant :

\[ P(\theta) = Pr(x_i = 1 | \theta) = \frac{1}{1 + e^{-\theta}} \]

Il s’agit en quelque sorte d’un modèle MRI à 0 paramètre puisqu’on ne prend pas en compte les caractéristiques de l’item (on considère que \( \forall i \ b_i = 0 \)). On détermine \( \theta \) à l’aide de la fonction de vraisemblance suivante :

\[ L(X | \theta) = \prod_{i=1}^{N} P(\theta)^{x_i=(1)} (1 - P(\theta))^{1-(x_i=1)} \]

\( \theta \) vérifie : [ \( \ln(X | \theta) \) ]'=0

C’est-à-dire :

\[ \sum_{i=1}^{N} (x_{ij} = 1) \frac{P'(\theta)}{P(\theta)} - [1 - (x_{ij} = 1)] \frac{P'(\theta)}{1 - P(\theta)} = 0 \]

Or \( P'(\theta) = P(\theta) (1 - P(\theta)) \) donc

\[ P(\theta) = \frac{1}{N} \sum_{i=1}^{N} (x_i = 1) \]

En considérant \( p \) la proportion de réponses correctes, on obtient ainsi :

\[ \theta = \ln \left( \frac{p}{1-p} \right) \]

Dans la suite on notera \( \ln \left( \frac{p}{1-p} \right) = p_{log} \)

Ainsi, à une transformation logistique près, on peut affirmer que la proportion de bonnes réponses est la compétence obtenue pour un modèle MRI à 0 paramètre.

La question initiale se pose donc dans les termes suivants : l’apport des caractéristiques d’item modifie-t-il sensiblement les compétences des individus ?

Prenons tout d’abord le cas du modèle à 1 paramètre.

Supposons qu’on ait calculé \( \theta \) par maximum de vraisemblance. \( \theta \) maximise donc \( \ln(L) \) où \( L \) est la vraisemblance du modèle. Par conséquent\n
\[ [\ln(L)]'(\theta) = 0 \]

Soit

\[ \sum_{i=1}^{N} (x_i = 1) \frac{P_i'(\theta)}{P_i(\theta)} - [1 - (x_i = 1)] \frac{P_i'(\theta)}{1 - P_i(\theta)} = 0 \]
Or pour un modèle à un paramètre : $P_i'(\theta) = P_i(\theta)(1 - P_i(\theta))$. On obtient donc

$$\frac{1}{N} \sum_{i=1}^{N} (x_i = 1) = \frac{1}{N} \sum_{i=1}^{N} P_i(\theta)$$

En d'autres termes, la moyenne des probabilité de réussite du modèle est égale à la proportion $p$ d'items réussis.

Si l'on remplace $P_i(\theta)$ par son expression, on obtient

$$p = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{1 + e^{b_i - \theta}}$$

Il existe donc une relation fonctionnelle entre $p$ et $\theta$, et donc également entre $p_{\log}$ et $\theta$. De plus, $p$ est une fonction continue strictement croissante de $\theta$. Cela signifie qu'à un $p$ donné on associe un unique $\theta$. Donc, pour une épreuve de $N$ items sans cahiers tournants, comme il n'y a que $N-1$ valeurs possibles de $p$ (on ne tient pas compte des extrêmes qui sont des cas particuliers), il n'y aura également que $N-1$ valeurs possibles de $\theta$. L'utilisation des valeurs plausibles dans le cas d'un modèle à 1 paramètre est donc particulièrement pertinente si l'on souhaite obtenir un score continu.

Par ailleurs, comme $e^{b_i - \theta} > 0$, on a, par convexité de $x \rightarrow \frac{1}{1 + x}$ :

$$p \geq \frac{1}{1 + e^{-\theta} E_1}$$

où $E_1 = \frac{1}{N} \sum_{i=1}^{N} e^{b_i}$.

et, en considérant $b_{\min}$ le minimum des $b_i$ :

$$\frac{1}{1 + e^{b_{\min} - \theta}} \geq \frac{1}{1 + e^{b_i - \theta}}$$

Donc :

$$\frac{1}{1 + e^{b_{\min} - \theta}} \geq p \geq \frac{1}{1 + e^{-\theta} E_1}$$

Comme $x \rightarrow \ln(\frac{x}{1-x})$ est une fonction croissante de $x$, on peut encadrer $p_{\log}$ :

$$\theta - b_{\min} \geq p_{\log} \geq \theta - \ln(E_1)$$

230

INSEE Méthodes
Ce résultat théorique se retrouve en pratique de manière encore plus spectaculaire. Sur les données françaises de IALS, on obtient ainsi $R^2=0.96$ entre $p_{oq}$ et $\theta$. De plus, on constate que les divergences sont essentiellement dues à l'imprécision des estimateurs de $\theta$ dans le cas où les individus ont répondu à très peu d'items. Lorsqu'on retire les personnes ayant répondu à moins de 20 items (l'épreuve en compte en moyenne 50), le $R^2$ atteint 0.99. Et si l'on prend en compte seulement les individus ayant passé entièrement un même cahier, on obtient des $R^2$ de l'ordre de 0.999.

Le modèle à 2 paramètres possède l'avantage d'accorder plus ou moins de poids aux items dans l'estimation du $\theta$. En théorie, les items qui fonctionnent mal interviendront peu tandis que ceux qui discriminent parfaitement la population auront une importance cruciale dans l'estimation du $\theta$. Cela sous-entend également que contrairement au modèle à 1 paramètre, plusieurs $\theta$ sont possibles pour un $p$ donné. En pratique, les corrélations entre $\theta$ et les proportions de bonnes réponses sont encore très élevées, puisqu'on obtient $R^2=0.95$ sur l'ensemble des individus. Si l'on calcule cet indicateur cahier par cahier, les $R^2$ avoisinent 0.99.

D'une manière analogue, on constate que l'estimateur intuitif de la difficulté des items, c'est-à-dire la proportion d'individus réussissant l'item, est très proche des $b_j$. Toujours pour les données françaises de IALS, on obtient $R^2=0.96$ avec un modèle à 1 paramètre.

Etant donné, d'une part, la complexité des estimations, et d'autre part, la similitude des résultats avec les estimateurs intuitifs, on peut donc s'interroger sur la pertinence de la mise en œuvre d'un modèle MRI. Un argument est souvent invoqué : l'avantage qu'il procure pour résoudre le problème des cahiers tournants. Cela n'est pas faux. Mais, dans une épreuve comme IALS où les items ne sont pas étudiés un par un, rien ne démontre que les cahiers tournants sont indispensables.
3.5. Les limites du modèle

3.5.1. La question des extrêmes

Comment peut-on comprendre le modèle de réponse à l’item ? Partons d’un modèle avec seuil :

$$\Pr(x_{ij}=1) = \begin{cases} 1 & \text{si } (\theta_j - b_i) > s \\ 0 & \text{sinon} \end{cases}$$

On peut maintenant supposer qu’à chaque question, la compétence « vraie » de l’individu est perturbée par un aléa imputable à la question, aux conditions extérieures... et ne peut s’observer directement. En posant donc $\theta_{i_{obs}} = \theta_j + \varepsilon_{ij}$, on obtient, si $G$ est la fonction de distribution de $-\varepsilon_{ij} :$

$$\Pr(x_{ij}=1)=\Pr( -\varepsilon_{ij}< (\theta_j - b_i) - s )=G((\theta_j - b_i) - s )$$

Si l’on suppose que les résidus sont gaussiens d’écart-type constant, on obtient un modèle MRI à un paramètre. Et si on suppose que l’écart-type du résidu varie en fonction des questions, on obtient un modèle MRI à deux paramètres. L’écart-type du résidu correspond en fait au coefficient de discrimination $a_i$ de la question : si celui-ci est très faible, l’écart-type du résidu est tellement important qu’il empêche de distinguer la compétence individuelle. Si $a_i$ est très fort, l’écart-type est quasi-nul et on retrouve pratiquement notre modèle de seuil initial.

Un point paraît discutable : la compétence « vraie » peut être modifiée à la fois négativement par des facteurs aléatoires de perturbation, ce qui semble logique - distraction par des éléments extérieurs, manque de motivation, lassitude face au questionnaire etc. - mais aussi positivement, ce qui est nettement moins évident. On peut à la rigueur imaginer que l’attrait d’un individu pour le domaine abordé par la question favorise ses chances de réussite. Que penser, en revanche, d’un élève de 6ème qui aurait une probabilité non nulle de répondre à une question posée à Polytechnique ? Cette situation n’est pas si caricaturale que cela puisque le modèle MRI stipule justement que toutes les questions se valent et que seule compte la probabilité de réussite : réussir une question simple avec une probabilité de 0.8 équivalait à réussir une question difficile avec une probabilité de 0.07, par exemple. Partant de là, les concepteur de IALS n’ont construit qu’un très faible nombre d’items élémentaires (4 ou 5 suivant le questionnaire), et ont donc évalué en grande partie la compétence des individus à partir d’items plus ardus. Il semble pourtant hasardeux de considérer comme illétrée une personne qui n’aurait réussi « que » 7% des questions les plus difficiles, puisqu’a priori un illétré serait totalement incapable de répondre à ces questions. Les limites du modèle sont très claires ici.
De plus, la façon de présenter les résultats par les responsables de l’OCDE est parfois malhonnête. En effet, ils définissent une population de niveau 1 d’un point de vue statistique : il s’agit des individus qui auraient, suivant le modèle, 80 % de chance de réussir les items de ce niveau et moins pour les items de difficulté supérieure. Cependant, ils la présentent également comme l’ensemble des individus « ayant un niveau de compétences très faible ; par exemple, la personne peut être incapable de déterminer correctement la dose d’un médicament à administrer à un enfant d’après le mode d’emploi indiqué sur l’emballage ». On note déjà un glissement sémantique puisque d’un côté ces individus réussiraient régulièrement les items de base (et seulement ceux-là) tandis que la deuxième formulation semble suggérer qu’ils ne sont même pas capables de résoudre les problèmes les plus simples. Cette dernière formulation a sans doute contribué à ce que l’on considère cette population comme illettrée. Or, selon le calcul de l’OCDE, 40 % des Français appartenaient à cette catégorie, chiffre largement supérieur aux estimations les plus pessimistes du taux d’illettrisme. D’ailleurs, on constate que 94 % des individus ont répondu correctement à la question donné en exemple dans la définition de l’OCDE (lecture d’une notice d’un médicament). Est-ce vraiment compatible avec les 40 % obtenus par ailleurs ?

3.5.2. La même justice pour tous ?

Rappelons qu’une des hypothèses du modèle est que les items fonctionnent identiquement dans tous les pays. C’est une des conséquences du postulat d’unidimensionnalité. La statistique de Mantel-Haenszel ou le SIB permet de tester cette hypothèse (voir paragraphe 3.2.1.). L’existence de nombreux fonctionnements différentiels viole le principe d’unidimensionnalité, et remet totalement en cause la possibilité de classements internationaux, qui constitue pourtant l’enjeu politique majeur de ce type d’enquête. Nous allons maintenant étudier ce problème dans un cas particulier.
4. Les biais d’items et l’étude du niveau moyen

4.1. France - Etats-Unis : le match était-il truqué ?

Nous allons procéder maintenant à une mini-enquête internationale, pour montrer dans les grandes lignes, les traitements généralement effectués. L’exemple que nous allons présenter ici est issu de l’enquête TIMSS menée en 1995 au niveau de la cinquième et de la quatrième. On se restreindra à l’étude de l’épreuve de mathématiques dans deux pays : la France et les Etats-Unis. Environ 6000 élèves ont répondu à l’épreuve en France et plus de 10 000 aux Etats-Unis. On a déterminé les scores des élèves et les paramètres des items grâce au logiciel CONQUEST, à partir d’un modèle de Rasch, qui a également été utilisé par les responsables de TIMSS pour établir les résultats publiés.

Le logiciel propose différentes façons d’attribuer un score à chaque élève. Nous avons retenu la méthode la plus simple, dans le cadre des MRI, celle utilisant l’estimation du maximum de vraisemblance, qui donne un score très proche du logit du taux de réussite, voire de ce taux lui-même. Le graphique 4 présente la distribution des scores aux Etats-Unis et en France. Les élèves français apparaissent sensiblement meilleurs que leurs camarades américains.

Graphique 4 : Répartition des scores aux Etats-Unis et en France

Nous allons surtout nous intéresser maintenant aux paramètres des items. Le logiciel donne pour chaque item sa difficulté et son « fit », accompagné de l’erreur de mesure
(voir tableau 1). Le « fit » mesure l'écart entre la discrimination de l'item et la discrimination moyenne : si trop de « fits » sont différents de 1, il est souhaitable d'utiliser un modèle à 2 paramètres. L'indétermination linéaire affectant les difficultés est résolue en fixant la moyenne à 0.

Tableau 1 : extrait des paramètres d'items calculés par CONQUEST

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>UNWGHTED FIT</th>
<th>WGHTED FIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>ESTIMATE</td>
<td>ERROR</td>
</tr>
<tr>
<td>1 RMA01</td>
<td>0.052</td>
<td>0.016</td>
</tr>
<tr>
<td>2 RMA02</td>
<td>-1.057</td>
<td>0.017</td>
</tr>
<tr>
<td>3 RMA03</td>
<td>0.060</td>
<td>0.016</td>
</tr>
<tr>
<td>4 RMA04</td>
<td>0.460</td>
<td>0.016</td>
</tr>
<tr>
<td>5 RMA05</td>
<td>-0.543</td>
<td>0.016</td>
</tr>
<tr>
<td>6 RMA06</td>
<td>-1.270</td>
<td>0.017</td>
</tr>
<tr>
<td>7 RMB07</td>
<td>-0.710</td>
<td>0.021</td>
</tr>
<tr>
<td>8 RMB08</td>
<td>-0.647</td>
<td>0.021</td>
</tr>
<tr>
<td>9 RMB09</td>
<td>-0.264</td>
<td>0.021</td>
</tr>
<tr>
<td>(...)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>151 RMU01B</td>
<td>1.566</td>
<td>0.027</td>
</tr>
<tr>
<td>152 RMU02A</td>
<td>3.267</td>
<td>0.033</td>
</tr>
<tr>
<td>153 RMU02B</td>
<td>2.651</td>
<td>0.031</td>
</tr>
<tr>
<td>154 RMV02</td>
<td>1.915</td>
<td>0.028</td>
</tr>
<tr>
<td>155 RMV03</td>
<td>0.282</td>
<td>0.026</td>
</tr>
<tr>
<td>156 RMV04</td>
<td>1.259*</td>
<td></td>
</tr>
</tbody>
</table>

An asterisk next to a parameter estimate indicates that it is constrained.
Separation Reliability = 1.000
Chi-square test of parameter equality = 271886.991,
df = 155, Sig Level = 0.000

Les erreurs de mesure données ici pourraient permettre de tester des « fonctionnements différenciels ». Il suffirait de faire tourner le modèle dans chaque pays et de comparer pour chaque item les deux paramètres de difficulté obtenus, en tenant compte de l'erreur de mesure. Des tests montreraient si les paramètres de difficultés sont significativement différents. En fait, le logiciel CONQUEST intègre cette possibilité et le tableau 2 donne pour chaque item, ce qu’il faut ajouter au paramètre de difficulté calculé sur les deux pays, pour obtenir celui spécifique à la France. Si ce coefficient est significativement différent de 0 (on utilise l'erreur de mesure pour le tester) on peut dire que l’item est plus difficile en France qu'aux Etats-Unis (si son signe est positif) ou moins difficile (si le signe est négatif).
Tableau 2 : extrait des paramètres concernant les fonctionnements différentiels

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>UNWGHTED FIT</th>
<th>WGHTED FIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>ESTIMATE</td>
<td>ERROR</td>
</tr>
<tr>
<td>1 RMA01</td>
<td>-0.186</td>
<td>0.016</td>
</tr>
<tr>
<td>2 RMA02</td>
<td>-0.092</td>
<td>0.017</td>
</tr>
<tr>
<td>3 RMA03</td>
<td>0.257</td>
<td>0.016</td>
</tr>
<tr>
<td>4 RMA04</td>
<td>0.062</td>
<td>0.016</td>
</tr>
<tr>
<td>5 RMA05</td>
<td>-0.413</td>
<td>0.016</td>
</tr>
<tr>
<td>6 RMA06</td>
<td>-0.039</td>
<td>0.017</td>
</tr>
<tr>
<td>7 RMB07</td>
<td>0.124</td>
<td>0.021</td>
</tr>
<tr>
<td>8 RMB08</td>
<td>0.265</td>
<td>0.021</td>
</tr>
</tbody>
</table>

An asterisk next to a parameter estimate indicates that it is constrained. Separation Reliability = 0.997, Chi-square test of parameter equality = 42360.449, df = 155, Sig Level = 0.000

On le voit, la plupart des items présentent dans cette analyse un fonctionnement différentiel. Pour éclairer cela, dans le cas de trois items (A05, B08 et U01B), il est sans doute intéressant d’utiliser la technique rudimentaire du Khi² en classant la population totale suivant les quartiles de score et en regardant pour chaque quartile, le taux de réussite par pays à l’item étudié (on neutralise ainsi le fait qu’une des populations a des moins bons résultats). Normalement, en l’absence de biais, les taux de réussite devraient être identiques (les élèves « médiocres » américains devraient avoir autant de chance de réussir l’item que les élèves « médiocres » français). On constate au contraire qu’ils sont assez différents (voir tableau 3). Pour le premier item A05, on note que l’écart entre la France et les États-Unis est considérable : on tourne autour de 20 % par quartile et l’écart est encore plus fort pour l’ensemble de la population, du fait d’un effet de structure (les jeunes français sont en moyenne meilleurs donc…). L’item U01B donne des résultats opposés : là, que ce soit par quartile ou sur l’ensemble de la population, c’est en faveur des jeunes américains que l’on observe un écart !

Manifestement la réussite à chacun de ces deux items doit faire appel à des éléments extérieurs à la pure dimension mathématique, sinon les conclusions sur le niveau des français et des américains risquent d’être contradictoires suivant qu’on prend l’un ou l’autre. L’item B08 est intéressant en ce qu’il montre un biais alors que l’analyse trop précipitée des taux de réussite l’aurait laissé passer. En effet, les taux de réussite sur l’ensemble de la population paraissent équivalents en France et aux États-Unis.
(67,5 %) mais par quartile, on note un léger mais significatif avantage pour les jeunes américains et seul l’effet de structure explique l’équivalence sur la totalité des élèves. En d’autres termes, trouver des réussites équivalentes alors que les populations ont des niveaux a priori différents est le signe d’un biais. En fait, cela conduit à utiliser une méthode simplifiée pour tester les biais : on compare l’écart entre France et États-Unis sur un item donné avec celui que l’on trouve sur l’ensemble des items.

Tableau 3 : Taux de réussite par pays et par quartiles de score à trois items.

<table>
<thead>
<tr>
<th></th>
<th>A05 France</th>
<th>A05 USA</th>
<th>B08 France</th>
<th>B08 USA</th>
<th>U01B France</th>
<th>U01B USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premier quartile</td>
<td>52,2%</td>
<td>33,7%</td>
<td>49,5%</td>
<td>50,4%</td>
<td>2,1%</td>
<td>9,4%</td>
</tr>
<tr>
<td>Deuxième quartile</td>
<td>71,2%</td>
<td>46,5%</td>
<td>63,9%</td>
<td>65,9%</td>
<td>7,5%</td>
<td>26,5%</td>
</tr>
<tr>
<td>Troisième quartile</td>
<td>83,6%</td>
<td>57,4%</td>
<td>71,7%</td>
<td>73,7%</td>
<td>14,3%</td>
<td>52,3%</td>
</tr>
<tr>
<td>Quatrième quartile</td>
<td>94,0%</td>
<td>73,0%</td>
<td>79,6%</td>
<td>82,9%</td>
<td>34,5%</td>
<td>75,0%</td>
</tr>
<tr>
<td>Ensemble</td>
<td>75,7%</td>
<td>52,6%</td>
<td>67,5%</td>
<td>67,6%</td>
<td>13,2%</td>
<td>41,9%</td>
</tr>
</tbody>
</table>

En reprenant les données du tableau 2, on note de plus que les items se voyaient attribuer respectivement un coefficient de fonctionnement différentiel de −0,413 ; 0,265 et 0,985, tous significativement différents de 0 ce qui confirme l’existence de biais.

Cependant, cette exposé paraîtra un peu sec si l’on continue à traiter les items comme de simples numéros. C’est pourquoi nous allons présenter quelques exemples selon le type de biais rencontré. Par souci de clarté, on se contentera de donner les taux de réussite car les biais présentés sont généralement suffisamment forts pour apparaître ainsi.

4.2. Explications des biais statistiques

Il est ainsi possible de mettre en évidence un assez grand nombre d’items biaisés, qui semblent inégalement difficiles selon le pays. Les raisons qui peuvent expliquer un tel phénomène sont nombreuses et nous allons en présenter quelques-unes.

4.2.1. Les problèmes de traduction

Les évaluations internationales sont proposées dans la langue de chaque pays participant, ce qui est une source de biais importante. Sans parler des erreurs qui peuvent se glisser, il est de nombreux exemples où une connaissance sommaire de l’anglais et du français permet de voir que même si les traductions sont exactes, les exercices ne sont pas de difficultés équivalentes.
Ces problèmes de traduction semblent avoir été particulièrement nets dans le cas de l’enquête IALS. Ainsi, des différences importantes entre les textes français et américains ont subsisté, que l’on peut regrouper en trois catégories:\footnote{cf. F. Guérin Pacé et A. Blum, in « L’illusion comparative », 1999.}:

- les erreurs de traduction.

- l’absence de répétition des termes en français

- imprécisions des termes français par rapport à l’anglais.

Les erreurs de traduction regroupe ici tous les faux-sens commis lors de la traduction. Ces faux-sens rendent la question plus difficile, voire erronée. Par exemple, l’intitulé d’une question était le suivant : « Sur la carte météo, entourez l’endroit où de fortes précipitations sont prévues dans la semaine. » Or les cartes météo donnaient seulement les prévisions pour les journées de vendredi à dimanche. Et l’on s’est aperçu que bon nombre de personnes essayait de déduire, à partir des cartes du week-end, le temps qu’il ferait la semaine suivante. En américain, la question ne prétrait pas à confusion puisqu’elle portait sur la météo du week-end. Conséquence immédiate : le taux de réussite en France est 26 points en dessous celui des États-Unis, alors que l’écart n’est que de 12 points sur l’ensemble des items. On trouvera dans l’article cité en note bien d’autres exemples de ce type.

Dans le cas de TIMSS, on repère quelques cas où sans qu’on puisse parler d’erreurs, les exercices ne sont pas équivalents en français et en anglais.

Ainsi les items J15 et P09 portaient sur des triangles homothétiques. Sans entrer dans le détail de ces deux questions, il est intéressant de noter que l’expression anglaise « the triangles are similaire » est sans doute moins parlante que sa traduction, tout à fait exacte d’ailleurs, « le premier triangle est un agrandissement de l’autre » ce qui peut expliquer le très net écart en faveur de la France observé sur ces deux questions (79 % contre 66 % de réussite pour J15, 58% contre 28 % pour P09).

Parfois, un petit changement qui peut paraître anodin bouleverse la difficulté de la question. Ainsi, 68 % des élèves américains peuvent trouver l’équivalent de 3 parmi cinq propositions, quand on donne comme bonne réponse xyxyx. Dans le questionnaire français, on a préféré remplacer la bonne réponse par une autre expression équivalente (y.y.y) avec des conséquences importantes : seuls 32 % des français reconnaissent l’égalité soit deux fois moins que les américains. A titre de comparaison, une question portant sur l’égalité m+m+m+m+m=4m est réussie par 60 % des élèves français contre 43 % des américains ! On peut donc penser que c’est l’usage d’une formule plus « scientifique » et moins familière qui a perturbé les jeunes français. Le problème aurait pu dans ce cas être évité mais dans bien d’autres, on est complètement désarmé quand les notations ou les formulations habituellement
utilisées dans les deux langues sont très différentes, sans être a priori de même difficulté : soit on prend dans chaque langue la bonne expression et les élèves où elle est plus complexe sont défavorisés ; soit on harmonise les deux expressions mais cela conduit à proposer dans un pays une expression qui n'est pas celle habituellement utilisée.

De façon générale, il est indispensable d'accorder une place importante à la traduction des questionnaires, sans hésiter à avoir recours à la technique de rétroatraction (Cette méthode consiste à traduire en retour vers la langue de départ, et à confronter le texte obtenu avec l'original).

4.2.2. Les biais culturels

On parle de biais culturel quand certaines catégories d'individus sont favorisées (ou parfois défavorisées, on va le voir) par leur connaissance a priori du support sur lequel portent les questions. Pour prendre un exemple caricatural, une question portant sur la Révolution française risque de favoriser les élèves français, tandis que ce seront les élèves américains tireront avantage d'une question liée à la guerre de Sécession. Il faut noter que la familiarité avec le support peut parfois causer des déboires et jouer dans le mauvais sens. Ainsi, dans IALS, une question consistait à identifier les comédies à partir des critiques de 4 films. Or, on observe en France que de nombreux enquêtés ont donné comme réponse un film dont la description ne l'apparante pourtant pas à une comédie. La seule explication est la présence dans ce film de l'acteur Michel Blanc, bien connu des Français pour ses rôles dans de nombreuses comédies, mais peu célèbre à l'étranger. Ici, la connaissance a priori des individus a joué en leur défaveur.

Les exercices de TIMSS, de forme assez scolaire, ne semblent pas tomber souvent sous le coup des problèmes culturels. On peut cependant évoquer un facteur qui a pu jouer (mais son influence semble faible) : l'usage systématique du système métrique. Ainsi, une question demandait quelle était l'unité la plus adaptée pour mesurer la masse d'un œuf : le centimètre, le millimètre, le gramme ou le kilogramme. Cet exercice est réussi à 88 % en France contre 68 % aux États-Unis. Il n'est pas impossible que la plus grande familiarité des élèves français avec ces unités ait joué en leur faveur.

4.2.3. Les différences de programmes

Dans le cas des évaluations scolaires, l'une des sources les plus importantes de biais (mais ici la connotation péjorative de ce terme va paraître particulièrement inadéquate) est l'influence des programmes des différents pays. Pour rester dans un cadre simple, si dans tel pays, l'accent est particulièrement mis sur un certain domaine de la discipline évaluée, on risque d'observer que tous les items relevant de ce domaine seront biaisés en faveur du pays, tandis que beaucoup des autres items seront biaisés dans l'autre sens.
Nous n'avons malheureusement pas la place ici de présenter, dans le cadre de notre comparaison France/États-Unis, les items où un tel type d'explications paraît pertinent, c'est pourquoi nous nous contenterons d'une présentation agrégée. En effet, il est apparu à l'examen que de nombreux items « favorisant » la France relevaient de la géométrie, tandis que les items d'algèbre étaient « plus difficiles » dans notre pays. Le tableau 4 permet de vérifier ce résultat en présentant les scores moyens (en %) pour les différents domaines des mathématiques.

**Tableau 4 : taux de réussite de la France et des États-Unis, par niveau, pour les différents domaines de mathématiques**

<table>
<thead>
<tr>
<th>Niveau</th>
<th>Pays</th>
<th>Total</th>
<th>Fraction et sens des nombres</th>
<th>Géométrie</th>
<th>Algèbre</th>
<th>Représentation des données, probabilités</th>
<th>Mesure</th>
<th>Proportionnalité</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>151</td>
<td>51</td>
<td>23</td>
<td>27</td>
<td>21</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>4ème</td>
<td>France</td>
<td>61</td>
<td>64</td>
<td>66</td>
<td>54</td>
<td>71</td>
<td>57</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>États-Unis</td>
<td>53</td>
<td>59</td>
<td>48</td>
<td>51</td>
<td>65</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>5ème</td>
<td>France</td>
<td>51</td>
<td>53</td>
<td>58</td>
<td>39</td>
<td>63</td>
<td>49</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>États-Unis</td>
<td>48</td>
<td>54</td>
<td>44</td>
<td>44</td>
<td>60</td>
<td>36</td>
<td>38</td>
</tr>
</tbody>
</table>

Considérons l'exemple de la cinquième. Sur les 151 items que comprend l'épreuve, la France se situe un peu au-dessus des États-Unis (3 points exactement ce qui, du fait de l'imprécision due à l'échantillonnage, n'est pas significatif) mais la situation est très différente suivant les domaines : en géométrie, les élèves français distancent nettement leurs camarades américains (58 % de réussite aux 23 items de ce type contre 44 %) alors qu'en algèbre ils sont sensiblement en retrait (39 % de réussite aux 21 items contre 44 % aux États-Unis). On pourrait encore affiner l'analyse : parmi les items du quatrième domaine, les français ont peu brillé dans les questions de probabilité. Par exemple, si on leur dit qu'un vase contient 9 jetons numérotés de 1 à 9, ils sont 61 % (contre 70 % aux États-Unis) à affirmer qu'on a 4 chances sur 9 de tirer un jeton pair (beaucoup optent pour 1 chance sur 2).

Il est à peine besoin de souligner l'intérêt de ce genre d'analyses. Si l'on souhaite que les évaluations internationales ne se résument pas à un simple tableau d'honneur mais servent à définir une politique éducative, l'examen des résultats domaine par domaine, voire item par item, par les spécialistes des disciplines évaluées, permettra d'infléchir la teneur des programmes, en montrant les points où il convient de faire un effort particulier.
4.2.4. Les relations différentes au questionnaire

Nous achèverons cette présentation rapide des causes possibles de biais d’item par des considérations sur le rapport au questionnaire. La forme générale des questions peut influencer les réponses, si elle paraît aux élèves trop déroutante. Mais surtout, nous avons pu constater que les réactions des élèves face aux efforts que nécessite la réponse, donnent lieu à des stratégies très différentes en France et aux États-Unis. Nous détaillons par la suite deux exemples pour mettre en évidence ce phénomène mais des analyses plus agrégées permettraient sans aucun doute de confirmer ces résultats.

Deux exercices portaient sur la manipulation des expressions du premier degré. Nous ne nous attarderons pas sur leur différence en terme de contenu pédagogique. Les différences de forme paraissent aussi importantes. L’item L16 demandait : « trouver x tel que 10x-15=5x+20 ». L’item N13, lui aussi sous forme de question ouverte, demandait « si x=2, combien vaut (7x+4)/(5x-4) ? ».

Tout d’abord, il est intéressant de noter une différence entre la France et les États-Unis concernant les non-réponses. Pour la question L16, 32 % des français n’ont rien répondu, contre seulement 18 % des américains. Sachant que les non-réponses sont souvent interprétées comme des réponses fausses, cela serait le signe que la question est particulièrement difficile pour les français, et si difficile qu’ils n’essaient même pas de répondre. Or elle est réussie à 32 % en France soit 8 points de plus qu’aux États-Unis ! Notons qu’on retrouverait sur d’autres items de l’épreuve (sous forme de questions ouvertes essentiellement) cette « contradiction » entre taux de réussite et taux de non-réponses. Il semble que les élèves français ne fassent pas de difficulté pour vérifier la concordance de deux propositions et cocher la case correspondant à la bonne réponse mais nombre d’entre eux rechignent à s’investir dans un problème qui semble demander plus d’effort et surtout, exigence jugée souvent inadmissible, où il faut écrire sa réponse ! On peut voir là une certaine démotivation face à une évaluation qui paraît sans enjeu. D’un point vue technique, cela amène à se poser des questions sur le traitement des non-réponses.

\[12\] Dans de nombreuses enquêtes, les non-réponses se voient attribuer une signification différentes selon leur emplacement. Quand on trouve à partir d’un certain exercice un nombre important de non-réponses qui se suivent jusqu’à la fin du questionnaire, on considère que, manque de temps ou de motivation, la personne s’est arrêtée là et que puisqu’elle n’a pas répondu aux dernières questions, on ne les fait pas entrer dans le calcul du score (les modèles MRI permettent de faire ceci en tenant compte de la difficulté des questions non faites). En revanche, toute non-réponse « interne », c’est-à-dire suivie par au moins une réponse, bonne ou non, est considérée comme une mauvaise réponse. Alain Blum et France Guérin-Pace ont montré dans l’article déjà cité qu’une telle façon de voir était dangereuse et qu’il y avait bien des cas où la personne interrogée « sautait » des exercices non parce qu’ils

INSEE Méthodes 241
Le cas de l’item N13 est différent. Les taux de non-réponses sont comparables (17 % en France contre 13 % aux États-Unis). Ce qui est intéressant ici, c’est de distinguer les bonnes réponses. Un peu moins de la moitié des élèves en France comme aux États-Unis réussissent l’item (il serait donc biaisé, d’après ce que nous avons vu, en faveur des américains) mais surtout les bonnes réponses ne sont pas les mêmes dans les deux pays. Aux États-Unis, 42 % répondent « 3 » et 6 % répondent « 18/6 ». En France, ils sont 33 % seulement à répondre « 3 » et 15 % à s’être contentés de donner la fraction sans la réduire complètement. On peut y voir là encore, plus que des écarts de compétences, un différentiel de motivation face au questionnement.

4.2.5. Que faire des items biaisés ?

Les sources de biais sont nombreuses mais elles ne sont pas de même nature. Les erreurs de traduction ou de présentation peuvent être corrigées. En revanche, la recherche d’équivalents dans toutes les langues ou les différences de motivation des individus sont beaucoup plus difficiles à prendre en compte. Enfin, les différences de réussite dues aux programmes scolaires ne semblent pas devoir être effacées, compte tenu de leur intérêt pédagogique.

En fait, face au biais, deux attitudes extrêmes sont possibles. Les intégristes de la mesure unidimensionnelle, généralement dévots des MRI, considèrent que les items biaisés détériore la qualité de la mesure, qu’il faut les éliminer quand par négligence, ils subsistent. En effet, d’un point de vue théorique, les items n’ont aucun intérêt pris individuellement, ils ne sont que des intermédiaires servant à l’élaboration de la dimension évaluée, unique et unidimensionnelle. On peut imaginer une attitude plus (trop ?) souple qui consisterait à construire une épreuve censée mesurer au mieux ce que l’on veut et tout ce que l’on veut, à analyser les résultats pour déterminer a posteriori si l’on ne pourrait pas en tirer un ou plusieurs facteurs synthétiques dominants. On pourrait alors reprocher à une telle approche d’être trop empirique. Il est en effet préférable d’avoir une théorie bien définie avant d’aborder la constitution d’un protocole et l’analyse des résultats. Mais, plutôt que de rejeter systématiquement les items biaisés, il est plus intéressant de profiter de leur éclairage sur les différences de réussite entre les pays.

lui semblaient hors de portée mais par manque d’intérêt. Ce que nous venons d’exposer rejoint cette analyse.

242 INSEE Méthodes
4.3. Généralisation aux comparaisons multiples

Il semble intéressant de tenter de généraliser les analyses menées ci-dessus au cas des comparaisons entre plus de deux pays. Comment dans ce cas-là mettre en évidence des biais ? Pour cela, on peut étudier les profils de réussite des pays, c'est-à-dire pour chaque pays la hiérarchie des items selon leur pourcentage de bonnes réponses. Théoriquement, si on respecte l'hypothèse d'unidimensionnalité, tous les pays devraient avoir le même profil de réussite (les items ayant la même difficulté dans chaque pays, ils doivent se retrouver dans le même ordre en terme de réussite). Si on observe des variations de profils de réussite d'un pays à l'autre, on peut supposer que ces variations sont liées à des différences culturelles et ne sont pas le simple fruit d'un effet aléatoire. Pour le vérifier, nous avons procédé à une classification hiérarchique des pays selon leurs profils de réussite aux items de mathématiques de l'enquête TIMSS pour la population 2. Cette méthode est reprise d'une analyse similaire menée par F. Guérin-Pace et A. Blum sur les données de IALS (1999). D'ailleurs, on aboutit ici au même constat : la proximité des profils de réussite des pays coïncide en général avec leur proximité géographique, culturelle ou linguistique. D'où la formation de groupes de pays à peu près cohérents selon ces critères (voir graphique 5).

Les quatre pays asiatiques (Singapour, la Corée du Sud, Hong-Kong et le Japon) forment un groupe homogène. Une grande classe rassemble les pays d'Europe occidentale (sauf l'Espagne et le Portugal) ainsi que les États-Unis, le Canada, l'Australie et la Nouvelle-Zélande. De ce groupe se dégagent plusieurs sous-groupes intéressants comme par exemple les pays d'Europe du Nord ou les deux groupes de pays anglophones. Les pays d'Europe de l'Est se retrouvent pratiquement tous dans le même grand groupe (à part la Bulgarie et la Roumanie). Le dernier regroupement est certainement le plus hétérogène : même si l'on observe des proximités géographiques ou culturelles (comme l'Espagne et le Portugal, par exemple), il rassemble des pays aussi divers que Chypre, l'Iran ou la Bulgarie.

Pour illustrer l'arbre de classification, nous avons représenté graphiquement les croisements des pourcentages de bonnes réponses de l'Angleterre et de l'Australie d'une part, et de l'Angleterre et du Japon d'autre part (graphiques 6 et 7).

---

13 Ce résultat est certain pour les modèles à 1 paramètre (cas de TIMSS) où, on l'a vu, le taux de réussite est équivalent au paramètre de difficulté donné par le modèle. La situation est plus complexe quand on utilise un modèle à 2 paramètres.

INSEE Méthodes 243
Graphique 5 : Arbre de classification des pays selon le profil de réussite
Graphiques 6 et 7 : comparaison des profils de réussite de l'Angleterre avec ceux de l'Australie et du Japon

Chaque point représente un item dont l'abscisse est le pourcentage de bonnes réponses observé en Angleterre et l'ordonnée le pourcentage de bonnes réponses au Japon ou en Australie. Il ressort assez clairement de ces représentations que l'échelle de réussite des items de l'Angleterre est très proche de celle de l'Australie mais assez éloignée de celle du Japon. D'ailleurs, la corrélation des rangs s'élève à 0.95 dans le premier cas contre seulement 0.63 dans le second.

4.4. Stabilité des classements des moyennes

A partir de la classification précédente, on a établi une partition des pays en cinq classes. Pour chacune de ces classes, on a déterminé les items les plus caractéristiques, c'est-à-dire les items particulièrement réussis ou échoués par ces populations.

Le classement des pays se modifie sensiblement selon qu'on le calcule sur les points faibles ou sur les points forts d'une classe particulière. Par exemple, nous avons classé les pays selon le taux de réussite moyen obtenu aux 20 items constituant les points faibles du groupe des pays asiatiques. On constate que, par rapport au classement établi sur l'ensemble des items, ces pays voient leurs positions considérablement diminuées. Ainsi Singapour passe de la 1ère à la 4ème place, le Japon de la 2ème à la 14ème, la Corée du Sud de la 3ème à la 19ème et Hong-Kong de la 4ème à la 24ème. Une autre pondération des items peut donc amener à un classement final très différent.

Ce dernier résultat conduit à s'interroger sur la robustesse des classements publiés par l'OCDE. Les imprécisions du processus d'élaboration de la mesure impliquent une certaine instabilité dans les palmarès des niveaux moyens. Il faut cependant reconnaître que la simulation du paragraphe précédent n'est pas complètement satisfaisante sur ce point. En effet, on s'est placé là dans le pire des cas pour les pays asiatiques, sur la sous-épreuve parmi les millions que l'on pourrait constituer à partir
des items passés qui donne un résultat très divergent du résultat publié. Il pourrait être intéressant de procéder à ce genre de simulation sans a priori, en constituant à partir d’un ensemble d’items, toutes les sous-épreuves possibles (voir pistes de recherche). Ces analyses ont cependant l’intérêt de montrer qu’on observe parfois de véritables bouleversements de classements et surtout, il est assez significatif que l’épreuve qui défavorise Honk-Kong soit aussi celle, à peu de choses près, qui défavorise le Japon ou Singapour.

Ceci nous amène à reprendre la question des biais d’items, en envisageant leur influence sur les classements des niveaux moyens des pays. Il est toujours possible de contester les résultats d’une évaluation en remettant en cause la composition des épreuves, en arguant que la dimension ne serait pas tout à fait la même ainsi si une équipe d’un pays particulier avait composé l’épreuve. Cependant, la composition des épreuves est, ou devrait être, le fruit d’un compromis entre les différents pays participants. Ainsi, on dispose au final d’une « norme internationale ». Il est alors important de montrer comment la définition nationale s’écarte de cette norme pour expliquer les résultats.

Un point plus inquiétant concerne la motivation des individus. Il est possible que les individus répondent systématiquement moins bien par rapport à leur niveau réel, à cause de leur manque d’intérêt par rapport à l’évaluation. Dans le cas de IALS, l’examen des taux de non-réponses est particulièrement éclairant sur ce point : 45 % des ménages français de l’échantillon initial ont refusé de répondre à l’enquête ; 12 % n’ont répondu qu’au questionnaire portant sur leurs caractéristiques sociodémographiques, sans répondre aux exercices ; 6 % ont commencé à y répondre mais se sont arrêtés avant la fin. En définitive, 37 % seulement de l’échantillon initial est allé au bout du questionnaire d’évaluation, en omettant parfois certaines questions. Cela pose trois questions : les individus qui ont refusé de répondre sont-ils identiques aux répondants ? La non-réponse à une question est-elle le signe de l’ « incompétence » du sujet ou de son désintérêt ? Est-on certain que ceux qui répondent l’ont fait avec assez de sérieux ? On peut penser que ces questions sont moins déterminantes dans le cadre scolaire, où les épreuves sont vite assimilées aux évaluations « à enjeu » habituelles. En revanche, quand, comme dans le cas de IALS, on cherche à mesurer les compétences d’individus ayant parfois quitté l’école depuis longtemps, on peut se demander s’ils accepteront de jouer le jeu et de consacrer des efforts de concentration souvent importants pour répondre aux questions.
5. Les inégalités de compétences

Il n’est sans doute jamais inutile de rappeler le caractère réducteur d’un jugement fondé sur le seul examen d’une moyenne. La plupart des études accordent une place tout aussi importante à la dispersion des résultats, qui donnent une idée des inégalités scolaires. On les caractérise souvent par les différences de taux d’accès à tel niveau de scolarisation entre enfants de cadres et enfants d’ouvrier, entre garçons et filles. On a vu ce que les évaluations de compétences permettaient d’apporter en plus par rapport à ces analyses. Il semble donc important de vérifier que la mesure des compétences se fait de façon suffisamment précise pour donner une idée des écarts individuels.

Les investigations que nous allons mener dans ce chapitre concernent la catégorie la plus « statistique » des indicateurs d’inégalités, ceux mesurant la dispersion brute (écart-type, écart interquartile). Nous essayerons de voir aussi dans quelle mesure on peut généraliser les résultats à des indicateurs plus « sociaux » (différences de scores entre enfants de cadres et enfants d’ouvriers par exemple)\(^{14}\).

5.1. Inégalités des compétences et écart-type des performances

Nous avons déjà évoqué les difficultés à interpréter des écarts entre individus. En effet, si on utilise la simple proportion de bonnes réponses, le niveau de l’épreuve détermine la dispersion : faible si l’épreuve est très facile ou très difficile (taux moyen proche de 0 % ou de 100 %), élevée si l’épreuve est de difficulté moyenne pour la population considérée (taux moyen entre 25 et 75 %).

Cela suggère qu’il est difficile d’atteindre les inégalités de compétences et que la dispersion des scores que l’on va observer risque d’être très dépendante de l’épreuve que l’on a utilisée. En fait, la dispersion des performances va dépendre autant de la « qualité » de l’épreuve (ou, pour être plus précis, de son degré d’adaptation au public visé), que de la « réelle » dispersion des compétences. Comme contre-exemple, on peut imaginer une épreuve dont les questions seraient si mal posées que les élèves répondraient au hasard : la dispersion serait faible non parce que les compétences sont proches mais parce que l’épreuve les a mal mesurées.

\(^{14}\) Cependant les indicateurs bruts de dispersions sont souvent utilisés dans le cadre des comparaisons internationales, en partie à cause des difficultés que l’on rencontre à comparer des écarts sociaux définis à partir de catégories rarement équivalentes d’un pays à l’autre.

INSEE Méthodes 247
L'objectif d'une évaluation statistique est de mettre en évidence des groupes en opposition bien tranchée, de distinguer les meilleurs des moins bons. Pour mieux comprendre cela, on peut songer à la situation du professeur dans sa classe : ce n'est pas parce que tous ses étudiants savent compter qu'un professeur d'université va leur mettre 10 sur 10. Inversement, un enseignant du primaire évitera d'interroger ses élèves sur la mécanique quantique. Le professeur adapte ses exigences afin de pouvoir distinguer les élèves selon leurs compétences. Il construit donc une épreuve de difficulté moyenne avec des questions faciles pour repérer les élèves en grande difficulté et des questions difficiles pour départager les meilleurs.

On met ainsi en évidence une apparente contradiction d'objectifs : si l'un des buts principaux d'une bonne politique éducative est de maintenir des écarts peu importants de compétences entre individus, celui des concepteurs d'évaluation est de chercher la meilleure épreuve possible, celle qui distingue le mieux les meilleurs des moins bons, les pays performants des moins performants. Ils visent donc la plus grande dispersion possible, afin de pouvoir plus facilement expliquer les écarts. En ce sens le fait de trouver un écart-type élevé peut être le fait d'une épreuve bien adaptée autant que d'une population variée.

Ce résultat doit rendre extrêmement prudent quand on compare deux populations n'ayant pas passé la même épreuve de mathématiques par exemple (ne parlons pas des comparaisons entre disciplines dont le sens est encore plus problématique). Le fait de trouver dans un cas un écart-type des performances plus élevé signifie-t-il que la population est plus dispersée ou que l'épreuve proposée est mieux adaptée ? Mais le problème peut aussi se poser quand on utilise une seule épreuve sur des publics différents : les écarts de dispersions ne relèvent-ils pas alors davantage du degré d'adaptation de chaque épreuve au public plutôt que de réels écarts de compétences ? C'est ce dernier cas que nous allons plus particulièrement étudier dans le cas des comparaisons internationales : l'usage d'un même protocole partout permet-il d'avoir une bonne image de la dispersion des compétences au sein de chaque pays ?

5.2. Dispersion des résultats et qualité de l'épreuve

On a vu que l'une des façons d'apprécier la qualité d'un test était l'examen de sa fidélité (voir chapitre 2.3). Rappelons que la fidélité mesure le degré de corrélation entre items, afin de vérifier qu'ils dépendent tous partiellement d'un même facteur. Il peut être intéressant de confronter cette notion, et les indicateurs qui en découlent, avec la mesure de la dispersion.
5.2.1. Des notions interdépendantes

Tout d’abord, on peut noter une dépendance mathématique entre la plupart des indicateurs de fidélité et les indicateurs de dispersion. Pour prendre le plus courant, le $\alpha$ de Cronbach (voir chapitre 3.2.), il apparaît qu’à structure d’épreuve donnée en terme de taux de réussite, plus l’écart-type est élevé, plus le $\alpha$ le sera. Et inversement. Ceci concerne l’écart-type du taux de réussite et ne se généralise pas forcément aux scores issus des MRI. Cependant, sur le plan empirique, il n’est pas difficile de mettre en évidence une telle relation : dans le cas de TIMSS, quand on compare l’écart-type du pays avec la mesure de la fidélité de l’épreuve dans ce pays, la corrélation linéaire est extrêmement élevée (elle est par exemple de 0.95 en sciences pour la septième année de scolarité sur 39 pays participants). Les corrélations sont du même ordre dans l’enquête PIAAC.

Une telle ampleur dans les corrélations est assez préoccupante. En effet, on peut se demander si l’un de ces indicateurs n’est pas « mensonger » et simple décalque de l’autre. Il est possible par quelques expériences de montrer que les torts sont partagés. Par exemple, quand on travaille sur des sous-groupes a priori peu dispersés, on note des $\alpha$ beaucoup plus faibles que pour l’ensemble de la population. Il apparaît ainsi que les $\alpha$ calculés pour des groupes définis selon le niveau de diplôme des parents sont tous inférieurs à ce que l’on note sur l’ensemble de la population. L’épreuve serait donc adaptée à toute la population sans être mieux adaptée à aucun sous-groupe qui la compose. En d’autres termes, discriminer entre eux des polytechniciens ou des illettrés est plus difficile que faire le même travail sur l’ensemble de la population. De plus, quand on propose une épreuve à une population a priori très dispersée (la population totale par exemple, avec ses niveaux d’instruction très variables) on risque de voir son épreuve créditée de coefficient de fidélité important, quelle que soit sa réelle qualité. La plupart des participants des JMS savent résoudre l’équation $3x+2=5$ et répondre à la question : « A quel temps est la phrase : « j’aurais mangé » ? ». Ces deux questions risquent de poser problème à beaucoup de personnes ayant quitté l’école assez vite, il y a longtemps. Par conséquent, la corrélation entre les réussites à ces deux questions sera sans doute significative. De là à dire qu’elles mesurent la même chose...

5.2.2. Des dispersions variables

La dispersion de la population a donc une incidence directe sur l’indice de fidélité : plus les écarts sont élevés, meilleurs seront les $\alpha$. Inversement, l’image que l’on aura des inégalités de compétences au sein d’un pays risque de dépendre de la qualité de l’épreuve. Ainsi, en travaillant sur les « meilleurs » items (c’est-à-dire les plus discriminants, les mieux corrélés avec le score global) on aboutit à des indicateurs de dispersions beaucoup plus élevés. Il est aussi possible de construire des épreuves discriminantes pour certains pays mais pas pour d’autres et de montrer la sensibilité des classements obtenus. Prenons un exemple. On va comparer les résultats de l’Australie et de la Grèce en mathématiques, en fin d’étude primaire. Quand on
travaille avec les scores MRI calculés au niveau international, on trouve respectivement des scores moyens de 530 et 453 et des écarts-type de 98 et 95 : si les élèves grecs ont de moins bons résultats ils ne paraissent ni plus ni moins dispersés que les élèves australiens. Nous allons maintenant raisonner sur deux sous-épreuves du cahier 3 construites de façon à maximiser l’écart. L’une contient les items qui sont très discriminants en Grèce sans l’être en Australie ; l’autre contient les items qui sont discriminants en Australie sans l’être en Grèce. Nous présentons ci-dessous (voir tableau 5) les scores moyens sous forme de pourcentages puis mis sous une forme logistique (on applique donc au score individuel la transformation log(p/(1-p)) ce qui donne des résultats très proches, on l’a vu, des scores issus d’un modèle de réponse à l’item).

### Tableau 5 : Dispersions en Grèce et en Australie selon l’épreuve

<table>
<thead>
<tr>
<th></th>
<th>Australie</th>
<th></th>
<th>Grèce</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moyenne</td>
<td>Écart-type</td>
<td>Moyenne</td>
<td>Écart-type</td>
</tr>
<tr>
<td>Score total (en %)</td>
<td>61,6 %</td>
<td>19,9 %</td>
<td>43,7 %</td>
<td>19,8 %</td>
</tr>
<tr>
<td>Score total (logit)</td>
<td>0,191</td>
<td>0,334</td>
<td>-0,101</td>
<td>0,319</td>
</tr>
<tr>
<td>Score « à la grecque » (en %)</td>
<td>69,4 %</td>
<td>19,5 %</td>
<td>52,7 %</td>
<td>22,8 %</td>
</tr>
<tr>
<td>Score « à la grecque » (logit)</td>
<td>0,327</td>
<td>0,345</td>
<td>0,048</td>
<td>0,379</td>
</tr>
<tr>
<td>Score « à l’austrалиenne » (en %)</td>
<td>49,0 %</td>
<td>24,6 %</td>
<td>29,0 %</td>
<td>18,7 %</td>
</tr>
<tr>
<td>Score « à l’austrалиenne » (logit)</td>
<td>-0,020</td>
<td>0,423</td>
<td>-0,354</td>
<td>0,336</td>
</tr>
</tbody>
</table>

L’analyse du score global (sous forme % ou logistique) confirme ce que l’on trouve avec le score MRI : la moyenne de la Grèce est inférieure et les écarts-type sont proches. Quand on travaille sur les deux sous-épreuves, les résultats diffèrent, pas tellement en terme de moyenne (l’écart entre les pays reste à peu près du même ordre) mais plutôt en terme de dispersion. Sur l’épreuve construite à partir d’items particulièrement discriminants en Grèce, ce pays apparaît légèrement plus dispersé, tandis que sur l’autre épreuve, les écarts sont plus importants en Australie.

Ces résultats peuvent être sans peine généralisés à d’autres couples de pays, à d’autres cahiers, à d’autres disciplines. Il convient cependant d’en saisir la portée. Comme dans l’étude des moyennes, on montre l’ampleur du problème sans l’expliquer et en se plaçant dans le pire des cas, sans dire si cette hypothèse est réaliste. En d’autres termes, nous sommes capables d’extraire des sous-épreuves qui ne fonctionnent pas bien à partir de l’existant. Cela ne prouve pas pour autant que les constructeurs de tests n’ont pas réussi à établir une épreuve globale qui soit valable pour tous, même si elle contient des sous-parties qui posent problème.

En effet, tout ce que nous développons ici n’a d’intérêt que si l’on peut montrer qu’il existe un risque qu’une épreuve donnée s’adapte plus ou moins bien selon les pays. Quels sont les facteurs qui peuvent influer sur la fidélité nationale d’une épreuve ? L’examen de cette question est encore peu avancé mais il apparaît vite un facteur
assez accessible et très déterminant sur le degré d'adaptation d'une épreuve : sa difficulté.

5.3. Moyenne et écart-type

5.3.1. Relation entre niveau moyen et dispersion

Les modèles psychométriques du type MRI ont beau être très élaborés, il est sans doute impossible de recruter les jeunes polytechniciens à l'aide d'une épreuve composée d'une centaine d'additions. De même si l'on propose une épreuve par QCM sur les équations du second degré à des élèves de CE2 et si le score varie entre 0 et 20 % de réussite, il est tout de même un peu abusive d'en conclure que les compétences en mathématiques à ce niveau sont faibles et peu dispersées. Exemples caricaturaux ? C'est à voir...

Les deux enquêtes que nous avons prises comme exemples conduisent parfois à des situations assez similaires. Dans TIMSS, on a cherché à mesurer la progression des élèves entre le milieu de l'école primaire et le milieu du secondaire, en les plaçant donc sur une même échelle. On avait même envisagé de prolonger la comparaison jusqu'à la fin du secondaire. Quant à IALS, le but de cette enquête était clairement d'évaluer les compétences en lecture de l'ensemble de la population, ce qui inclue en France les polytechniciens, les énarques et les personnes n'ayant pas eu le certificat d'étude. De plus, ces diversités des publics au sein de chaque pays s'additionnent quand on cherche à construire une épreuve valable pour tous les pays. La question se pose véritablement de savoir s'il est possible de construire une épreuve qui convienne à tous. Ne risque-t-on pas, en proposant une épreuve facile, de trouver une faible dispersion dans les pays performants, tandis que les pays les moins bons apparaîtraient comme particulièrement inégalitaires ?

Prenons deux exemples. Le premier concerne les résultats en mathématiques des élèves en huitième année d'étude (niveau quatrième) dans une trentaine de pays

15 On peut exprimer ce problème au niveau des items (ce qui selon les psychométriciens donne des résultats équivalents aux considérations sur les individus). Y a-t-il plus d'écarts en terme de difficulté entre la résolution d'une addition et d'une règle de 3, d'une part, ou entre une équation du second degré et la résolution d'un algorithme de maximisation de la vraisemblance d'autre part ? Il est probable qu'un traitement statistique nous conduirait à affirmer que le premier écart est plus grand (beaucoup d'individus réussissent l'addition sans réussir la règle de 3 alors que dans une population quelconque les deux autres semblent également difficiles). Une analyse pédagogique ne donnerait pas forcément le même résultat. Les deux perspectives apparaissent d'ailleurs également légitimes mais se distinguent suivant le public et le niveau de compétence visé.
(enquête TIMSS). Le graphique 8 présente la relation entre l’écart-type du score et la moyenne par pays. Une corrélation fortement positive apparaît ($r^2=63\%$) : meilleure est la moyenne, plus grandes sont les inégalités. Si nous étudions maintenant les résultats de la population adulte en compréhension de textes quantitatifs (données issues de IALS, graphique 9), on note au contraire une corrélation fortement négative entre l’écart interquartile et la moyenne ($r^2=74\%$) : meilleure est la moyenne, moins grandes sont les inégalités ! Voilà deux résultats fort différents ! Bien sûr les notions mesurées ne sont pas identiques. De plus, la population n’est pas la même. Cependant, une telle contradiction dans les relations est troublante. La question est de savoir si ce résultat provient de la nature des épreuves (difficile dans le premier cas, facile dans le second) ou s’il touche réellement au lien entre inégalités et niveaux moyens (problème dont on voit sans peine l’importance car on touche à la question du lien entre démocratisation et évolution du niveau moyen).

Graphiques 8 et 9 : relation entre niveau moyen et dispersion pour TIMSS et IALS

5.3.2. Approche empirique

Pour mieux comprendre le lien entre la difficulté de l’épreuve et sa fidélité, nous allons extraire d’une épreuve donnée (le cahier 3 de mathématiques de TIMSS sur la population 1) différentes sous-épreuves de difficulté variée. Notons que cette méthode ne souffre plus des défauts précédemment exprimés : on ne se place plus, a posteriori, dans le pire des cas mais on teste une hypothèse sur un élément (la difficulté) indépendant du facteur étudié (la dispersion).

A partir des 39 items du cahier que nous avons retenu, nous avons construit 3 sous-épreuves différentes de 10 items chacune (bien sûr le fait de travailler sur relativement peu d’items est un peu embarrassant mais l’objectif est ici de dégager des tendances plutôt que d’aboutir à un chiffrage précis du phénomène) en considérant trois niveaux de difficulté (difficile/moyen/facile). Le tableau 6 donne les caractéristiques des scores sur l’ensemble de la population.
Tableau 6 : Dispersion selon la difficulté de l’épreuve

<table>
<thead>
<tr>
<th></th>
<th>Score en %</th>
<th>Score en logit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moyenne</td>
<td>Écart-type</td>
</tr>
<tr>
<td>Ed=Epreuve difficile</td>
<td>31,6 %</td>
<td>24,4 %</td>
</tr>
<tr>
<td>Em=Epreuve moyenne</td>
<td>52,6 %</td>
<td>26,8 %</td>
</tr>
<tr>
<td>Ef=Epreuve facile</td>
<td>78,6 %</td>
<td>20,8 %</td>
</tr>
</tbody>
</table>

Ce tableau présente les scores sous forme de pourcentages mais dans la suite nous n’utiliserons plus que les scores sous forme logistique. On retrouve les mêmes résultats concernant les scores en pourcentage sauf qu’il s’ajoute une dépendance due aux bornes 0 et 1 qui rend les relations (généralement quadratiques) entre moyennes et écarts-type encore plus fortes. L’examen de ce tableau permet de voir que nos épreuves sont bien distinctes en terme de difficulté. La deuxième étape consiste à calculer les moyennes et écarts-type par pays et à étudier les corrélations entre les valeurs obtenues. Il est tout d’abord intéressant de constater que lorsque l’on confronte les classements de moyennes obtenus selon les épreuves, les corrélations sont assez élevées (rappelons que l’on travaille maintenant sur une population de 27 pays) : la moins bonne est de 0,88 (entre Ed et Ef). En particulier, les corrélations avec le score moyen MRI calculé et publié par l’IEA sont toujours supérieures à 0,95. Cela semble indiquer que pour effectuer un classement des pays selon leur niveau moyen, la difficulté de l’épreuve importe peu, les classements sont assez stables. En revanche les résultats sont beaucoup moins bons quand on compare les écarts-type obtenus sur les différentes épreuves. Par exemple, il n’y a pas de lien entre les écarts-type calculés par pays sur l’épreuve Ed et ceux calculés sur l’épreuve Ef. Les pays qui paraissent très dispersés pour la première épreuve ne le sont pas forcément pour l’autre. La mesure de la dispersion au sein de chaque pays semble donc plus sensible à l’épreuve utilisée et les classements obtenus peu fiables.

Notre hypothèse est que la difficulté des épreuves influe sur le public auquel l’épreuve s’adapte le mieux et confère donc la plus grande dispersion de résultats. Nous allons tester cette hypothèse en cherchant pour chaque épreuve la relation entre l’écart-type et la moyenne. Le tableau 7 présente le $R^2$ issu d’une régression linéaire simple de l’écart-type sur la moyenne puis le $R^2$ obtenu en ajoutant le carré de la moyenne (cet ajout permet de tenir compte du fait que la dépendance peut avoir la forme d’une cloche : pour une épreuve de difficulté moyenne en particulier, l’écart-type sera faible pour les pays les moins bons, élevé pour les pays moyens et faible pour les pays les meilleurs). Pour la relation linéaire, nous indiquons par une flèche, si la corrélation est positive ou négative.

INSEE Méthodes 253
Tableau 7 : Relation moyenne-écart-type selon la difficulté de l’épreuve

<table>
<thead>
<tr>
<th>Épreuve</th>
<th>Relation linéaire</th>
<th>Relation quadratique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficile</td>
<td>55,9 %</td>
<td>61,7 %</td>
</tr>
<tr>
<td>Moyenne</td>
<td>15,7 %</td>
<td>67,8 %</td>
</tr>
<tr>
<td>Facile</td>
<td>66,7 %</td>
<td>78,4 %</td>
</tr>
</tbody>
</table>

On observe des relations assez fortes entre moyennes et écarts-type : la dépendance est assez nettement linéaire et positive pour l’épreuve difficile (meilleur est un pays plus ses résultats sont dispersés). La dépendance est négative pour l’épreuve facile. La relation linéaire est non significative pour l’épreuve moyenne mais l’ajout d’un terme quadratique fait augmenter fortement la valeur de la prédiction, ce qui rend compte de la forme en cloche de la dépendance.

5.4. Écarts sociaux

Les problèmes que nous venons d’exposer concernant la mesure de la dispersion brute ont évidemment une influence sur la mesure des inégalités sociales. En effet, on aurait tendance à voir une causalité inverse (s’il y a de grands écarts entre groupes sociaux, il risque d’y avoir de forts écarts absolus) mais comme nous venons de voir que les écarts entre individus sont appréhendés de façon toujours un peu déformée, il est probable que cette déformation affecte la mesure des inégalités sociales. Ainsi, on peut craindre que les conclusions quant à savoir quels sont les pays où il y a le plus d’écarts suivant l’origine sociale, soient très sensibles aux caractéristiques de l’épreuve choisie. Reprenons l’exemple de IALS. Le premier des deux graphiques ci-dessous permet de voir que l’écart « inter-diplômes » (i.e. la différence de scores entre les personnes ayant achevé des études supérieures et celles n’ayant pas fini des études secondaires) est très lié à la mesure brute de dispersion que constitue l’écart interquartile. Le graphique 11, quant à lui, montre le lien entre le premier écart et le score moyen : la forme de la dépendance rappelle celle que l’on trouvait pour l’écart interquartile (cf graphique 9). On peut se demander si cette dépendance a un sens, si les pays qui ont le meilleur niveau moyen sont aussi ceux qui maintiennent des écarts faibles entre les groupes sociaux ou si cette corrélation n’est qu’un artefact statistique dû au fait que l’épreuve était assez facile. En effet, si l’on effectue le même genre d’analyse sur TIMSS, en mathématiques, en utilisant l’origine sociale des parents des élèves interrogés, on trouve une dépendance positive : les pays les meilleurs sont ceux où il y a le plus d’écart, résultat qui ressemble à ce que l’on trouve pour l’écart-type et qu’il nous semble possible d’expliquer par le fait que les exercices étaient plutôt difficiles.

254

INSEE Méthodes
Les remarques développées dans ce chapitre amènent donc à s'interroger sur la validité de tous les indicateurs cherchant à mesurer les écarts entre individus. Est-il possible de conclure que tel pays est plus inégalitaire que tel autre, sachant qu'un ensemble différent d'exercices aurait pu conduire à des constats différentes. Parmi les facteurs qui peuvent perturber les classements, on a vu que la difficulté de l'épreuve jouait un rôle important ce qui est un problème en soi. En effet, le lien entre le niveau moyen atteint par l'ensemble de la population et les inégalités entre individus est un point crucial du débat sur l'école. Pour dire les choses en termes plus politiques, est-il possible de démocratiser le système éducatif, en élévant le niveau ? On a vu que selon qu'on travaillait sur TIMSS ou sur IALS, on aboutissait sur ce point à des conclusions divergentes.

Enfin, nous n'avons pas évoqué la question des indicateurs relatifs qu'on utilise parfois. Dans l'exemple précédent, on peut en effet se demander si le rapport entre l'écart inter-diplôme et l'écart interquartile donne une image assez bonne de l'influence du niveau d'étude. Loin d'être une évidence, cette question est en cours d'exploration.
6. Quelques pistes de recherche

Il n’a pas été possible de présenter dans le cadre de ce colloque l’ensemble de nos recherches, ni de mener à bien tout ce que nous avons entrepris. Nous allons donc présenter succinctement les pistes qu’il nous semble intéressant d’explorer.

- **Confronter les résultats d’enquêtes entre eux.** Bien sûr il n’existe pas au niveau international d’enquêtes concurrentes qui chercheraient à mesurer au même moment, dans la même matière, au même niveau de la scolarité les compétences des individus. Cependant, des enquêtes plus ou moins comparables sont effectuées à des époques différentes ou à des niveaux différents. Il peut être intéressant de croiser le classement des différents pays pour voir s’ils sont stables (relativement stables, évidemment car, si l’enquête a un intérêt, c’est justement parce qu’on s’attend à ce qu’ils bougent). Dans le cas de TIMSS, par exemple, il est important de noter que deux niveaux successifs étaient ciblés par les enquêtes : CE2-CM1 pour la population 1 ; cinquième-quatrième pour la population 2. Le fait de trouver des classements plus stables quand on compare deux niveaux successifs peut s’interpréter à l’aide de cette proximité mais le fait que ces niveaux sont évalués par le même protocole doit jouer un rôle. Il est à noter que la comparaison des inégalités bruts (écarts-type des scores) montrent une grande instabilité des classements quand on compare deux niveaux « éloignés », alors qu’ils sont assez proches pour les comparaisons CE2-CM1 et cinquième-quatrième.

- **Découper une épreuve.** Une méthode pour étudier de façon empirique la sensibilité des indicateurs peut être de tirer au hasard dans un ensemble d’items, une sous-épreuve dont on va confronter les résultats avec ce que donne la totalité. En procédant ainsi un assez grand nombre de fois, en imposant de plus quelques contraintes pour obtenir des épreuves de caractéristiques assez variées (en particulier en terme de difficulté) on pourra se faire une idée de la stabilité des classements. Les premiers essais que nous avons effectués montrent une bonne stabilité des classements de pays par niveaux moyens (on est à chaque fois très proche du classement obtenu avec l’ensemble des items) alors que les classements de dispersion sont beaucoup moins constants (il n’est pas rare de construire ainsi des épreuves qui donnent un classement corrélé autour de .50 avec le classement sur l’ensemble des items). De même, il apparaît que conformément à ce que l’on a dit, le lien entre moyenne et écart-type dépend fortement du niveau de difficulté de l’épreuve.

- **Les cahiers tournants.** Dans le même ordre d’idée, il peut être intéressant d’utiliser ces sous-épreuves déjà définies que sont les cahiers tournants. En comparant les classements de pays obtenus sur chaque cahier indépendamment, on a une autre image, un peu moins négative, de leur robustesse. Les classements de moyennes sont corréles à plus de .95 entre eux. Encore une fois, les corrélations sont moins bonnes pour les indicateurs de dispersion.
Cependant, les premières analyses montrent que l’usage des modèles MRI permet d’augmenter un peu la stabilité des classements d’un cahier à l’autre.

- **Approfondir la connaissance des modèles MRI.** Il semble indispensable de chercher à mieux comprendre le fonctionnement de ces modèles, pour déterminer ce qu’ils peuvent apporter. La technique des valeurs plausibles permettra peut-être d’améliorer l’estimation des écarts individuels. De plus, la confrontation des modèles à 1, 2 et 3 paramètres s’avétera instructive. En début d’année, les responsables de TIMSS ont ainsi fourni de nouvelles estimations des scores de compétences calculées à partir d’un modèle à 3 paramètres au lieu de 1 jusqu’à présent. Ce modèle donne des résultats sensiblement différents. En particulier, il n’y a plus de corrélation positive entre la moyenne et l’écart-type en mathématiques, pour la population d’élèves de collège. Au contraire, la corrélation est significativement négative, comme pour les autres scores ! Une telle inversion dans un résultat fondamental doit amener à s’interroger sur l’effet de la modélisation sur un même ensemble de données.

**Conclusion**

Le tableau dressé ici peut paraître globalement négatif, du fait que notre démarche critique s’est attachée à montrer les points sur lesquels des progrès restent à faire. Il convient de nuancer notre propos. Tout d’abord, la plupart des remarques concernent en premier lieu le cas des comparaisons internationales et ne s’appliquent que partiellement aux évaluations faites dans un seul pays. D’autre part, tout n’est pas à rejeter dans les comparaisons internationales et sur bien des points, elles apportent un éclairage intéressant sur le fonctionnement des systèmes éducatifs. Cependant, nous avons voulu montrer qu’il fallait les utiliser avec d’extrêmes précautions.

De plus, il va nous être impossible de rester dans le domaine de la critique. En effet, l’INSEE, en collaboration avec la DPD, l’INED et l’INETOP, s’est vue charger de mener une enquête méthodologique sur la question la plus délicate : l’évaluation des compétences des adultes. Quand l’enquête IALS a commencé, l’expérience française sur le sujet était à peu près nulle. Il existe certes de nombreuses évaluations menées auprès d’adultes, dans le cadre de la formation professionnelle par exemple, mais aucune ne visait à aller « chez les gens » pour évaluer leurs compétences. Le pari de faire de cette expérience dans le cadre d’une enquête internationale était risqué. Effectivement, devant les difficultés rencontrées concernant la passation des épreuves et le traitement des réponses, il a semblé préférable d’ajourner la participation française. En définitive, l’INSEE a certainement un rôle important à jouer pour faire progresser la connaissance française sur la question, en menant une enquête méthodologique à la fois sur le terrain et dans les modèles.
Eléments de bibliographie

Publications de la DPD

- La Direction de la Programmation et du Développement donne régulièrement dans ses publications annuelles des indicateurs relevant de la mesure de compétence (voir Etat de l’Ecole, Géographie de l’Ecole, Repères et Références).

- On trouvera dans les autres publications (les Notes d’information, la revue Education et Formations, les dossiers d’Education et Formations) des informations plus détaillées sur des opérations d’évaluation de compétences ayant été mené au CP, au CE2, au CM2, en sixième, en 4ème et 5ème, en 3ème générale et technologique, en Terminale, ainsi que des études spécifiques sur les jeunes de 17 ans passant les tests de lecture de la journée d’Appel Préparation Défense, le recrutement de l’élite scolaire (Grandes Ecoles), une comparaison fondée sur les résultats d’un échantillon de 1920 au certificat d’étude, etc.

Ouvrages généraux sur le système éducatif.


Articles récents sur les inégalités devant l’école


Psychométrie


Publications internationales


Pour une vision critique des comparaisons internationales


L’ENQUÊTE « IMAGE DES SITUATIONS PROFESSIONNELLES ET SOCIALES »

Y. LEMEL(*) et L. RAINWATER(**)

(*) CREST
(**) Harvard University

1. Introduction

La position sociale d’une personne dépend fortement de son occupation professionnelle et il y a un grand consensus sur l’évaluation et l’appréciation des occupations professionnelles. C’est ce consensus qui permet la construction d’échelles dites de « prestige professionnel ».

Ceci étant, bien d’autres éléments influent sans doute sur l’évaluation d’une position sociale. Comment organise-t-on les différentes informations que l’on a sur quelqu’un - son sexe, information la plus immédiate à obtenir ; son âge ; sa profession ; son niveau de ressources ; son niveau d’éducation ; etc - pour se faire une idée de sa position dans la société ? L’enquête dont on va présenter la méthodologie traite de tous ces ingrédients et de la manière dont on les combine.

Cette enquête adopte une méthodologie originale. Chaque questionnaire est une pièce originale et doit être considérée comme un échantillon particulier de questions. Il y a donc en fait 3000 questionnaire tous différents car une randomization complète du questionnement a été introduite. Concrètement, chaque personne doit répondre à 27 questions. Chaque question présente une « situation », obtenue en combinant plus ou moins aléatoirement un, deux, trois ou plus des éléments suivants : diplôme, revenu, profession, âge, sexe d’un individu, de son conjoint éventuel. Dans chaque cas, les personnes doivent attribuer une note, traduisant leur jugement sur le statut social d’un titulaire d’une telle situation. Toutes les situations sont « réelles », chacune d’entre elles (ou presque) ayant été observée dans une enquête. Les 27 situations sont différentes d’une personne à l’autre et il n’y a pas deux questionnaires semblables.

On présente dans cette note la méthodologie de l’enquête, c’est-à-dire essentiellement le contenu du questionnaire et la manière dont l’enquête a été acceptée et les dossiers remplis.
2. Architecture et construction des questionnaires

Toutes les questions de l’enquête adoptent en fait la même forme : l’enquête doit donner une appréciation de la situation décrite dans le corps de la question en utilisant une même échelle qui est explicitement présentée comme pouvant dépasser 100 mais doit pouvoir s’inscrire dans un bac à 4 chiffres (plus grande valeur possible a priori donc 9999).

2.1. Architecture

Les deux premières questions (notées W1 et W2) traitent de la situation de la personne, de son travail si elle en a un. Les vingt questions suivantes (notées V1 à V20) traitent de situations « élémentaires » dont l’ensemble a été tiré au sort de manière à représenter équitablement situations a priori hautes, basses et faibles. Les deux questions suivantes (S1 et S2) traitent de situations un peu plus complexes. Les cinq dernières (notées SV1 à SV5) de situations encore plus complexes. Ici encore des tirages au sort ont été faits pour assurer une égalité de présentation des « bonnes » et des « mauvaises situations ». Physiquement, W1 et W2 occupent la première page du questionnaire, V1 à V20, S1 et S2 la seconde page, SV1 à SV5 les deux dernières pages.

Compte tenu du mode de construction des questionnaires, les questions sont interchangeables, statistiquement parlant au moins, à l’intérieur des 4 ensembles distingués ci-dessus. Ceci étant elles diffèrent aussi, à l’intérieur de chacun de ces ensembles, par la place qu’elles occupent dans le questionnaire.

2.2. La question de base

Le libellé des questions peut varier un petit peu d’un cas à l’autre pour s’adapter aux particularités de celles-ci mais, fondamentalement, le contenu en est toujours le même.

En voici la présentation générale :

« Le respect, le prestige, la considération, la notoriété dépendent de nombreux éléments : niveau d’éducation, métier (actuel ou passé), ce qu’on gagne, la manière dont on se comporte, le lieu où il réside, etc. Indiquez-nous comment, à votre avis, les français jugent en général tel ou tel cas que nous allons vous présenter ?

Considérez, par exemple, quelqu’un âgé de 25 ans, qui est instituteur. A votre avis, les français considèrent-ils cette personne comme deux fois mieux lotie que la moyenne auquel cas vous lui attribuez la note de 200 - ou, au contraire, comme moitié moins bien lotie que la moyenne - auquel cas vous lui attribuez la note de 50 ? »
C’est tout à fait volontairement que le libellé de la question utilise de nombreux termes plus ou moins synonymes. Des tests préparatoires montrent que certains d’entre eux, pris isolément, peuvent susciter réaction d’incompréhension chez certaines personnes. Leur association lève ces difficultés. C’est tout à fait volontairement aussi que la question renvoie au jugement moyen des Français dans le but d’éviter des idiosyncrasies personnelles vis-à-vis, par exemple, de telle profession qu’on ne pourrait, pour des raisons parfaitement contingentes, exercer soi-même. Au surplus, les recherches sur ces sujets montrent classiquement que les résultats sont beaucoup moins dépendants qu’on ne le croirait de la formulation particulière retenue.

La question est posée en différentiel. Le libellé suggère de donner un score de 200 si la situation est deux fois meilleure que la situation moyenne, (implicitly dans ce libellé) cotée à 100. Ce n’est donc pas l’écart qui compte mais le rapport. La gamme d’évaluation proposée n’est pas finie (de 0 à 20 par exemple, échelle typique de la notation scolaire en France) mais infinie, dans une logique de rapport d’échelle. Cette formulation est inspirée de considérations de psychométrie et de psychologie sociale. Les appréciations du type de celui requis ici reflètent souvent des échelles de jugement de type géométrique, non métrique, organisée autour d’un point d’ancrage (ici supposé à 100). La conséquence statistique serait que la distribution des réponses, si l’hypothèse psychométrique est exacte, doit être de type log-normale de moyenne 100. On verra plus loin ce qu’il en est.

2.3. Les « situations » évaluées

Chaque questionnaire propose à évaluation :

- 20 situations définies par la combinaison : d’un âge ; d’un sexe ; d’un niveau d’éducation ou d’un niveau de revenu ou d’une profession ;
- 2 situations définies par la combinaison : d’un âge ; d’un sexe ; d’une composition de foyer ; d’un revenu ;
- 5 situations définies par l’âge, le sexe, le niveau de diplôme, la profession d’une personne ; le niveau de diplôme et la profession de son conjoint ; la composition de leur foyer ; le détail de leurs ressources financières annuelles.

On a distingué 10 niveaux de diplôme différents (cf. Annexe 1). Les combinaisons avec le sexe et l’âge de leur titulaire sont tirées au sort à partir de celles effectivement observées dans l’enquête sur les budgets familiaux mais le tirage des questions pour un questionnaire donné a été fait de manière à ce que chacun d’entre eux comprenne des positions au dessus et au dessous de la moyenne.

Environ 400 différents montants de revenu individuel sont utilisés dans l’enquête. Comme dans le cas précédent, les combinaisons avec sexe et âge sont issues de celles réellement observées dans une enquête mais avec tirage au sort pour
constituer un questionnaire assurant de rencontrer toutes les situations. Les montants varient de 10 000 à 3 300 000 Francs.

Environ 493 libellés d’occupation différents ont été utilisés. Chacun d’entre eux est décliné au féminin ou au masculin suivant le genre du titulaire. Leur ensemble assure une représentativité de l’univers professionnel et des statuts d’emploi salariés ou indépendants au niveau de la PCS. La combinaison âge, sexe, occupation est fondée sur les données du recensement avec affectation par tirage au sort de l’âge (pour chaque sexe séparément) dans une plage acceptable pour chaque profession séparément (c’est-à-dire en excluant les déciles extrêmes de cette distribution). Pour un questionnaire donné, le tirage au sort des combinaisons assurait un tiers de situation moyenne, haute et basse. L’ordre des questions était complètement aléatoire.

Quant aux descriptions complètes des familles, elles étaient en fait dérivées directement des 7500 questionnaires de l’enquête Budget de Famille, tirés au sort de manière à assurer une représentation du haut et du bas de l’échelle des revenus.

3. L’acceptation de l’enquête


3.1. Comment identifier les refus ?

Les enquêteurs ont réalisé (en interview directe) l’enquête principale PCV. De cette enquête sont issus 5472 dossiers individuels de personnes ayant 15 ans et plus, appartenant au sous-échantillon « sortant » (donc 76 personnes ayant juste 15 ans). Le taux de refus apparent est donc de 47,5%. Il ne s’agit cependant pas tout à fait du taux « réel » pour diverses raisons.

D’abord les enquêteurs ont pu plus ou moins bien respecter les consignes données (certains éléments montrent des difficultés dans certains cas). On peut se demander, en particulier, si les enquêteurs ont bien donné des questionnaires à tous les individus de plus de 15 ans de l’échantillon « sortant ». Aucun élément d’information n’est disponible sur ce point. Par ailleurs, les questionnaires renvoyés sont plus ou moins bien remplis de sorte qu’on peut se demander, pour les plus mal renseignés, s’ils ne sont pas des refus déguisés.

On utilisera la base suivante : tous les individus de 15 ans ou plus appartenant à un ménage du sous-échantillon sortant ayant accepté l’enquête PCV (ces individus sont
au nombre de 5472), complétés des individus n’appartenant pas au champ précédent, ayant répondu à l’enquête PCV de Mai et ayant renvoyé le questionnaire auto-renseigné (ces individus sont au nombre de 29). Cette base contient au total 5500 individus (et non 5501 = 5472 + 26 + 3, comme on pourrait s’y attendre, car un individu, du sous-échantillon « sortant » n’ayant pas renvoyé de questionnaire, a dû être éliminé, son numéro d’identification s’avérant erroné et ne permettant pas de retrouver son ménage d’appartenance).


3.2. Qui n’a donc pas renvoyé de questionnaire ?

Le « taux de renvoi » est donc de 52,2 %. Il varie, bien sûr, suivant les caractéristiques sociales et démographiques des personnes interrogées.

Un très rapide balayage est le suivant :

- Ce taux est le même pour les hommes et les femmes ;
- Il aurait une légère tendance à croître avec l’âge : un peu au-dessus de 45 % pour les plus jeunes, autour de 55 % pour les plus âgés ;
- Les variations suivant le niveau d’études paraissent faibles, une fois tenu compte du fait que les personnes n’ayant jamais fait d’étude ou s’étant arrêté avant la fin du primaire ont des taux de renvoi vraiment faibles (25 %). On peut supposer que les personnes les plus éduquées ont un peu plus renvoyé : un peu au-dessus de 55 % pour le supérieur et le technique long ;
- Les variations suivant le niveau de revenu paraissent plus marquées. Elles affecteraient la forme d’une courbe en U, les taux de renvoi atteignant presque les 60 % dans les tranches intermédiaires et se situant plutôt au dessous de la moyenne dans les tranches extrêmes ;
- Il y a des variations fortes avec l’indicateur de réussite financière : le taux croît régulièrement avec celle-ci depuis les 40 % pour les moins riches à 60 % pour les plus riches. Il faudrait approfondir le point, la signification exacte de l’indicateur, qui traite d’un flux et non d’un stock, devant être regardé : s’agit-il d’un effet du niveau de vie ? de la satisfaction financière ? y aurait-il, plus généralement, des effets liés à la plus ou moins grande satisfaction (quelle qu’en soit la raison) des individus ?
- Les indépendants - ce qui n’est pas pour surprendre - ainsi que les « autres inactifs » - ont moins renvoyé que les autres : les taux de renvoi sont à 45 % et moins (41 % pour les agriculteurs). Parmi les salariés, on peut imaginer qu’il
y a décroissance du taux de renvoi avec l'augmentation du statut social : de 49 % pour les ouvriers à 55 % pour les cadres supérieurs ;

- Un effet de la présence d'enfants apparaît au travers des différences par type de ménage : les personnes appartenant à des couples avec enfants renvoient un peu moins fréquemment. Les personnes seules jeunes répondent plus que les personnes âgées. Au total, le taux oscille entre 50 et 55 %, avec l'exception des personnes seules jeunes (peu nombreuses) dont le taux de renvoi atteint les 60 %.

- Enfin, le taux de renvoi est un peu plus faible dans l'agglomération parisienne. Mais il reste néanmoins de 45 % dans Paris intra-muros, qui est traditionnellement la zone la plus difficile à enquêter.

Reste à voir comment se présentent les effets propres. Une analyse logistique de type polytomique ordonnée suggère que les variables démographiques - sexe et âge - n'ont guère d'effets propres. L'appartenance socioprofessionnelle est à la limite de la significativité. Revenus, diplôme et épargne paraissent très significatifs.

En résumé, les taux de renvoi varient donc pour l'essentiel dans une plage de valeur entre 45 et 55 % avec, occasionnellement, quelques valeurs nettement plus faibles (les personnes ayant peu ou pas étudié, les agriculteurs, les personnes très riches) et, un peu plus souvent, quelques valeurs plus fortes de 60 à 65 % (les personnes relativement âgées, ou ayant une faible épargne, les personnes jeunes vivant seules). Il y a très certainement des strates de population combinant plusieurs de ces dimensions pour lesquelles les taux de renvoi seront très faibles. Ce taux ne semble cependant pas varier autant qu'on l'aurait imaginer a priori.

3.3. La qualité du remplissage des questionnaires

Ainsi qu'on l'a indiqué, les questionnaires auto-renseignés sont plus ou moins bien renseignés. Pour examiner plus avant ce point, on a calculé le nombre de cases laissées en blanc parmi les questions posées dans les pages 2, 3 et 4 du questionnaire, laissant de côté la page 1, c'est-à-dire les appréciations portées sur la situation personnelle et le travail (l'appréciation sur le travail est structurellement non renseignée dans nombre de cas car il n'y pas d'activité professionnelle). Ce nombre de cases peut s'échelonner entre 0 et 27. La moyenne de cet indicateur sur les 2871 dossiers est de 2,3 (il y a donc en moyenne 2 cases manquantes par dossier) et la variance de 6,8. En fait, la distribution est nettement bi modale avec une énorme majorité de dossiers sans aucun manque (près de 80 %) et une petite minorité avec beaucoup de manques : cf. tableau 1.

Où faut-il mettre le seuil séparant les deux sous-population implicites de cette distribution bi modale ? Il y a trois pointes de « non renseignements » : 27 questions non renseignées (dossiers totalement vides donc), 22 et 20 questions non renseignées. Les 33 dossiers avec 22 questions non-renseignées se composent : de 14 dossiers où ne sont remplis que SV1 et SV5 (ce sont les 5 dernières questions traitant des vignettes de situations « complexes »), 16 dossiers où ne sont remplis
que V1 à V5 (ce sont les 5 premières questions des pages 2 à 4), 3 autres cas. Les 11 dossiers avec 20 questions renseignées se composent : de 1 dossier où sont seulement renseignées les questions V1 à V5 et V9 ; et de 19 dossiers semblables : y sont seulement renseignées des 7 dernières questions1. A l’autre extrémité, l’examen des dossiers avec 5 questions non-renseignées ne suggère pas d’organisation particulière.

En somme, il y a :

- les personnes qui n’ont rien renseigné (tous en jugeant bon de renvoyer le document) : 152 sont dans ce cas ;
- celles qui n’ont donné d’appréciations que sur des situations complexes (mais peut-être plus réalistes) : 24 (14+10) sont dans ce cas ;
- celles qui ont commencé et abandonné : 16 sont dans ce cas.

Ce sont, sans doute, ces dernières seulement qu’on doit considérer sans conteste aucun comme des abandons. Pour les autres, on peut débattre. Il y a parmi les 152 répondants dont les dossiers sont vides des personnes dont il conviendrait plutôt d’assimiler les absences de renseignements à une égalité de notations de toutes les situations2.

Cette population de personnes ayant renvoyé le questionnaire mais l’ayant mal rempli n’est pas quelconque. Il s’agit de personnes plutôt âgées (presque 50 % d’entre elles ont plus de 65 ans), de très faibles niveaux d’études, retraités (pour 43 %). Elles vivent plutôt à la campagne ou dans des petites villes (51 %). L’interprétation de leur comportement comme une forme d’acquiescement à l’enquête accompagnée d’un sentiment d’incapacité à répondre paraît assez plausible. On note, par ailleurs, que les couples avec enfants qui avaient plus tendance que d’autres à ne pas renvoyer les questionnaires les renvoient renseignés, ce qui confirme bien l’interprétation en terme de disponibilités suggérée ci-dessus.

Tout ceci dit, il faut bien trancher. On a considéré que le seuil était à placer à 19 et que les dossiers contenant au moins 20 items non-renseignés étaient des « pseudo-refus ». Dans la définition proposée le taux d’« acceptation » est donc de 48,4 %.

1 Outre SV1 et SV 5 donc, les deux questions précédant cette série dans le questionnaire. Ce sont les deux seules, en dehors de la série SV1 à SV5, à ne pas traiter de situations « élémentaires ».
2 Ainsi certains questionnaires complètement vierges sont accompagnés de lettre de motivation expliquant, par exemple, que nos sociétés sont obsédées par le classement mais que toutes les positions dans la société sont également estimables. On pourrait fort bien considérer que ces personnes ont donné la même cote à toutes les questions proposées à leur appréciation.
3.4. Bilan final.

Finalement, le taux de retour après relance est de 52 %. On peut certes juger le verre à moitié plein comme à moitié vide. Nous aurions plutôt tendance à la voir à moitié plein même si nous aurions anticipé un taux moyen de retour plus élevé mais des taux de non réponse plus élevés dans certains cas.

On a vu que le taux de renvoi avait plutôt tendance à croître avec l’avancement en âge des personnes. Quand on tient compte du taux de remplissage des documents, le phénomène à tendance à s’atténuer voire disparaître. On le voit bien d’ailleurs au fait que le chiseaux de la vraisemblance pour les critères d’âge, de type de ménage et aussi de catégorie socioprofessionnelle - trois caractéristiques qui incorporent des dimensions d’âge dans leur construction - est plus faible lorsqu’on le calcule sur le taux d’« acceptation » qu’il ne l’est lorsqu’il est calculé sur le taux de « renvoi » - 28,88 au lieu de 35,70 ; 31,64 au lieu de 62,86 ; et 41,74 au lieu de 51,46 respectivement3. Inversement, les personnes les moins éduquées avaient déjà tendance à moins renvoyer. Le phénomène est donc encore plus accentué quand on examine les taux d’acceptation. Conséquence : la log vraisemblance, le pouvoir explicatif restera stable. Quant aux conclusions sur les indicateurs de niveau de vie, elles restent inchangées.

En substance, les caractéristiques socio-démographiques des répondants sont similaires à celles des non-respondants et on ne constate pas de biais particuliers. Bien entendu, que les caractéristiques sociales ou démographiques de l’échantillon répondant soient les mêmes que celles de la base de sondage n’implique pas que tous ces répondants ne partagent pas un même trait de caractère susceptible d’avoir un impact sur les résultats et de les biaiser éventuellement. Mais c’est un point qu’on ne peut guère étudier.

4. L’incertitude des jugements élémentaires

Les informations élémentaires composant une situation dont on demande l’évaluation peuvent eux aussi avoir une influence sur la facilité voire la possibilité pour les notateurs de donner une réponse. Si, par exemple, cas d’école, il était exclu par la coutume ou la loi qu’une femme ait une certaine activité professionnelle, le pourcentage de non déclarations pour les situations décrivant des femmes actives ayant cette profession seraient bien plus élevé que celui pour les situations décrivant des hommes actifs. Un élément jugé incongru susciterait perplexités et le nombre de notateurs ne pouvant donner d’évaluation augmenterait. Plus largement, la plus ou moins grande proximité aux notateurs d’un élément particulier influencera le taux de réponse pour cet élément. Le taux de non évaluation renseigne donc de manière

3 Dans le cas de la catégorie socioprofessionnelle, on notera aussi que les catégories d’indépendants, qui avaient déjà tendance à moins renvoyer, sont plus fréquentes à ne pas remplir ce qu’elles renvoient.
indirecte sur les composantes considérées comme faisant légitimement partie de ce qui peut être pris en compte pour juger du statut social.

Dans cet esprit, pour 5 043 des 57 420 évaluations, soit 8,7 %, la réponse est NSP ou Non Déclaré. En fait, comme on l’a dit, une part importante de ces questions non renseignées proviennent de questionnaires très peu renseignés en général mais renvoyés néanmoins. Le taux de 8 % surestime donc certainement le taux « réel » de refus des questions prises individuellement et, en tout état de cause, ces questionnaires très peu renseignés n’apportent guère d’information sur le degré auquel une situation est jugée « évaluable ».

L’inspection des variations du taux suggère globalement que, s’il n’y a guère de réserves chez les notateurs pour ce qui est des caractéristiques de genre, il y a plus de flottements pour certaines tranches d’âge et qu’il y aurait des indices (légers) donnant à penser que la dimension économique du statut social est la plus facile à évaluer. On rappelle qu’il existe une différence de traitement dans l’enquête entre les composantes du statut que les sociologues qualifient d’« acquises » et celles qu’ils qualifient d’« astreintes ». Chaque question nécessite en effet d’évaluer une personne d’âge, de sexe et de profession/de niveau de revenu/de niveau d’éducation connus. Les deux premiers éléments sont donc toujours présents de sorte qu’on ne peut pas savoir ce que leur présence induit comme difficultés supplémentaires pour l’évaluation. On peut seulement dans leur cas juger de différences relatives à l’intérieur de chacune d’entre elles.

4.1. Dans l’ordre : revenu, occupation et éducation

Le taux de non-déclaration varie avec la dimension « achieved » considérée. Il est de 9,7 ; 8,8 et 8,0 % pour les items comprenant une indication de niveau d’éducation, d’occupation et de revenu respectivement.

On notera que le revenu n’apparaît pas plus particulièrement difficile à évaluer que ne le sont les autres éléments, ce serait plutôt le contraire. Le montant des ressources est souvent réputé (au moins dans le contexte Français) comme étant très difficile à connaître : ce serait indiscret que d’interroger quelqu’un sur son revenu. Il est peut-être effectivement indiscret d’interroger sur ce sujet en sorte que l’information sera moins disponible que d’autres dans la vie courante, mais elle n’en sera pas moins pertinente. Telle est la conclusion que l’on peut tirer des réactions aux questions posées.

La hiérarchie se retrouve identiquement (à quelques cas près) quel que soit l’âge ou le sexe de la personne dont on propose d’évaluer la situation et ne dépend donc guère d’un biais éventuel dans les combinaisons de sexe et d’âge conjointes. Des analyses logistiques du taux de non réponse en fonction de l’âge, du sexe et des trois dimensions revenu, occupation, éducation confirment ces résultats.
4.2. Des doutes sur les personnes les moins jeunes, guère sur le genre.

Le taux de non évaluation d’une situation est (en moyenne) complètement indépendant du fait qu’il s’agit de la situation d’une femme ou de celle d’un homme. La moyenne est de 8,8% dans les deux cas. Ce résultat est bien confirmé par les analyses logistiques des taux intégrant les autres éléments contribuant à définir la situation à évaluer. L’effet propre du genre est fortement non significatif.

Le taux de non-évaluation dépend par contre nettement de l’âge. Il y a d’abord une lente diminution avec celui-ci suivi d’une augmentation assez rapide : cf. graph.1. L’ajustement de fonctions polynomiales ou de « spline » confirme bien cette évolution et suggère que l’inflexion se situerait entre quarante et cinquante cinq ans. Elle paraît un peu plus marquée pour les hommes que pour les femmes. Ce sont donc les personnes âgées et, quel que soit leur sexe, qui soulèvent le plus de problèmes aux évaluateurs.

Le point n’est évidemment pas pour surprendre. La littérature sur les processus de vieillissement s’interroge sur la perte de statut (au sens structurel du terme) dont seraient victimes les personnes âgées. Ceci étant, on notera que c’est avant l’âge de retrait de la vie active que se produit l’inflexion observée. Par ailleurs, on aurait pu imaginer qu’un phénomène analogue se produirait pour les plus jeunes qui, dans la réalité, ne sont généralement pas encore rentrés dans la vie active aux âges ici proposés et pour lesquels les interrogations sont nombreuses sur leur statut exact. Or, il ne semble pas que l’on trouve à leur égard les doutes qui apparaissent pour les plus âgés.

Au total, la différence dans l’« évaluabilité » du sexe et de l’âge est assez nette.

5. Les modalités de réponse

On va maintenant examiner les modalités de réponse qui sont utilisées. L’étude se fera sur les questionnaires renvoyés, chiffrés, on le rappelle, indépendamment de la qualité de leur remplissage. Sauf mention du contraire la variable W2 sera éliminée de l’analyse car elle n’est renseignée que pour les personnes actives.

5.1. Les valeurs rencontrées

Les notes varient de -200 (2 cas) à 9999 (3 cas. Par « cas » on entend ici une occurrence parmi les valeurs renseignées pour les questions W1, V1 à V20, S1 et S2, SV1 et SV5 dans l’ensemble des 2871 questionnaires). La totalité de la plage offerte a donc été utilisée du côté des valeurs positives. Les valeurs négatives ou nulles

---

4 Tous les tests de significativité doivent être considérés avec prudence compte tenu de la complexité de l’enquête. Nous les évoquons à titre indicatif.
n’étaient en rien suggérées dans le questionnaire, tous les exemples offerts utilisant des valeurs positives ; elles apparaissent néanmoins « spontanément » dans 57 cas.

Ceci étant, toutes les valeurs possibles entre les deux extrémités de la plage n’apparaissent pas, loin de là. Elles sont, en théorie, au nombre de 10 000 et quelques. Or on n’observe que 195 valeurs différentes, la valeur 100 se taillant la part du lion (20 646 occurrences soit 28 % des cas). Et, parmi ces 195 valeurs, une cinquantaine seulement ont attiré plus de 10 choix et 30 plus de 100. Il est clair que les valeurs rondes sont beaucoup plus fréquentes que les autres, le hit-parade des trente réponses les plus citées ne comprenant que des valeurs se terminant pas 0 (ou 5 pour cinq d’entre elles) et les 5 valeurs les plus citées étant :

<table>
<thead>
<tr>
<th>Valeur</th>
<th>Fréquence</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>21 062 fois</td>
</tr>
<tr>
<td>50</td>
<td>12 195 fois</td>
</tr>
<tr>
<td>200</td>
<td>9 647 fois</td>
</tr>
<tr>
<td>150</td>
<td>7 254 fois</td>
</tr>
<tr>
<td>80</td>
<td>3 985 fois</td>
</tr>
</tbody>
</table>

De très grandes valeurs apparaissent comme on l’a vu, mais elles apparaissent très rarement. Les valeurs inférieures ou égales à 1000 représentent 99,8 % des cas, (environ 2 % de notes sont supérieures à 1000 donc) et les valeurs inférieures ou égales à 2000 représentent 99,9 % des cas.

En laissant de côté les valeurs supérieures à 1000, (moins de 2 % des cas donc), la distribution apparaîtrait plutôt tri modale avec une nette asymétrie. Un premier mode se situerait aux alentours de 100, un second aux alentours de 50, un troisième autour de 200 comme on le voit en lissant la distribution (graphique 2). Le second mode est cependant nettement moins important que le premier et si on accentue le lissage, il disparaît entièrement. Ceci étant, la distribution est très discontinue puisqu’on ne trouve sur cet intervalle [-200, 1000] que 164 valeurs différentes et que les 20 plus fréquentes représentent presque 15 % de la totalité des cas.

5.2. Le problème des valeurs extrêmes.

Comment faut-il considérer les valeurs extrêmes ? D’un point de vue de pure description statistique d’une distribution, il s’agit évidemment d’« outliers ». Les valeurs négatives extrêmes en font d’ailleurs aussi partie. On serait donc porté à les supprimer. Encore faudrait-il s’assurer que leur suppression n’aurait pas trop de conséquences sur les résultats et qu’on ne peut non plus les identifier « sociologiquement » soit qu’on puisse leur associer une signification particulière, soit qu’elles paraissent les réponses données par des individus d’un certain type.

Le tableau 2 donne des éléments sur la sensibilité des caractéristiques centrales de la distribution des notes à la suppression des valeurs extrêmes. Les situations extrêmes sont les suivantes :
• moyenne de 125,5 et variance de 212,7 lorsqu'on tronque fortement à gauche sans tronquer à droite ;
• moyenne de 117,5 et variance de 80,7 lorsqu'on tronque fortement à droite sans tronquer à gauche ;

les autres résultats se plaçant évidemment entre ces deux extrêmes. La moyenne est donc en fait assez peu sensible aux troncatures, mais la variance l’est. Il semble par ailleurs que les résultats sont plus sensibles aux troncatures à droite qu’aux troncatures à gauche, un seuil apparaissant entre les notes de 6 000 et 2 000. Variances et moyennes diminuent fortement entre la troncature à 6 000 et celle à 2 000 tandis que tronquer juste à 6 000 ou plus haut paraît assez neutre.

Y-a-t-il cependant une signification sociologique possible de ces valeurs extrêmes ?

Les valeurs extrêmes, négatives inférieures à zéro, sont le fait d’un très petit nombre d’individus, 5 en fait. Un individu a produit les 2 cas de notations à -200, un autre 3 des cas de notation à -100 et trois individus tous les autres cas de notation à -100 (avec respectivement 17, 15 et 18 cas chacun). Ces trois derniers notateurs appartiennent de plus à la même famille - les deux parents plus leur fille - et les notes qu’ils ont fournies varient entre -100 et 100.

L’examen maintenant des cas de valeurs extrêmes positives montre symétriquement qu’un très petit nombre de notateurs sont en cause : dix suffisent pour produirent les 66 cas de notes supérieures strictement à 2 000. Un notateur produit 20 cas (mais les notes varient et ne sont pas uniformes, variance de l’ordre de 2 500). Un autre dix (là non plus il n’y a pas uniformité des notes, la variance est de l’ordre de 4 000 avec une valeur maximale de 9999).

Ni dans un cas, ni dans un autre, les dossiers ne semblent « bidonnés ». On a le sentiment qu’il s’agit plus d’un décalage des points de référence que d’autre chose. Ceci étant, il faudrait examiner cela plus attentivement et voir comment on peut l’articuler avec des hypothèses sur la nature des stimuli que constitueraient les questions. Le graphique 3 permet de voir qu’il y a un lien indéniable entre les valeurs maximales et minimales attribuées par une personne une fois ces valeurs standardisées c’est-à-dire rapportées à la moyenne des notes que la personne a donné et divisé par l’écart-type de celles-ci.

5.3. Différences éventuelles entre questions

La tâche de noter toutes les situations proposées peut être longue et fastidieuse. On peut donc imaginer que l’ordre des questions ne sera pas sans conséquence sur les réponses qui seront faites.

De fait, l’examen des valeurs moyennes des réponses aux diverses questions (graphique 4) montre qu’il semble bien avoir des effets : 1/ les variables « complexes » (SV1 à SV5) sont mieux évaluées en moyenne. L’effet est particulièrement marqué lorsqu’on tient compte des valeurs extrêmes repérées ci-
dessus et a tendance à s'affaiblir lorsqu'on exclue ces valeurs extrêmes. Il n'en reste pas moins net ; 2/ on peut imaginer qu'il y a une légère croissance de la note des questions simples au fur et à mesure que l'on avance dans le déroulement du questionnaire ; 3/ deux questions, V10 et V17, semblent nettement sous-estimées par rapport aux autres. Par contre V2, V9 et V13, V18 et V19 seraient surestimées mais l'effet est moins certain.

Pour ce qui est du 1er point, des explications différentes, non forcément alternatives, peuvent s'imaginer. L'une est que les situations, plus complexes mais peut-être plus compréhensibles, sont de ce seul fait mieux considérées par référence aux situations plus simples mais abstraites qui précèdent. Une autre serait que ces situations, plus complexes, mêlent plus fréquemment des éléments qui se compensent les uns les autres tandis que les situations «simples» mettent mieux en évidence les cas extrêmes. Il y a enfin que ces questions sont les dernières, les plus sensibles donc à effet d'«accoutumance positive» pour lequel un test explicite serait nécessaire.

6. L'évaluation des situations. Quel indicateur construire ?

Le point de référence pour les évaluations est supposé être le même pour tous : de 100. Au vu de ce qui précède, on peut se demander si c'est bien le cas. Tout le monde devrait donner la même note moyenne, aux aléas d'échantillonnage des questions près, aux questions d'évaluation (ou au sous-ensemble de ces questions qui traitent des situations explicitement présentées par ailleurs dans le questionnaire comme «moyenne», «le travail d'une secrétaire, d'un plombier ou d'un employé dans une banque»). Cela n'est pas le cas et il y a une nette variabilité de l'évaluation de la situation moyenne. Par contre, la moyenne des évaluations données pour sa propre situation (1ère question du questionnaire : W1) diffère peu de la moyenne (notée MEAN) des évaluations que l'on donne sur le panel des autres évaluations que l'on doit faire :

\[ W1 = 118,35 \text{ tandis que MEAN} = 124,39. \]

Les ordres de grandeur sont tout à fait comparables et tous deux nettement supérieurs à 100.

Plus largement, on peut imaginer diverses possibilités en combinant les possibilités sur le point de référence d'une part, le type d'échelle utilisée de l'autre (parmi d'autres). Par exemple :

- échelle absolue, ancrage absolu : 100 (ou toute autre valeur, la même pour tout le monde) ;
- échelle absolue, ancrage : individuel ;
- échelle relative, ancrage absolu : 100 (ou toute autre valeur, la même pour tout le monde) ;
- échelle relative, ancrage individuel.

INSEE Méthodes
Que faut-il choisir ?

A une échelle « absolue » correspond l’indicateur non transformé, à l’échelle « relative » correspond un indicateur obtenu par transformé logarithmique. Clairement les distributions des transformées logarithmiques dans le cas de W1 sont beaucoup plus proches de lois normales (tableau 3). Et, parmi celles-ci, c’est l’hypothèse d’un ancrage plutôt individuel qu’identique pour tout le monde qui permet le plus d’approcher une loi normale, notamment par le fait que de rapporter à la moyenne individuelle semble « lisser » considérablement la distribution. A l’aune de ce critère donc, les principes d’évaluation retenus donc seraient bien relatifs, comme imaginé initialement, mais plutôt par référence à des évaluations personnelles de la situation standard qu’en référence à une évaluation a priori de 100.

Un autre critère d’évaluation est le degré auquel l’index utilisé permet de révéler des disparités de comportement. Le tableau 4 donne des éléments à cet égard. Le modèle utilisé pour construire ce tableau est une régression des moindres carrés ordinaires utilisant en additif toutes les variables explicatives caractérisant les répondants, chacune d’entre elles décomposée en dummies. On obtient les mêmes conclusions que précédemment. L’utilisation de transformées logarithmiques améliore substantiellement la capacité différenciatrice et l’utilisation d’une référence à la moyenne individuelle paraît un peu meilleure.

Dans la suite du travail, on a utilisé le log des notes rapportées à la moyenne des évaluations5. L’indicateur ne suit pas vraiment une loi normale (les tests statistiques rejettent très clairement cette possibilité) mais il s’agit néanmoins d’une loi unimodale relativement symétrique (graphique x). Les quantiles à 10 % sont respectivement de - 0,82 et + 0,47, ceux à 2,5 % de - 1,18 et + 0,71.

*  
*  

La méthodologie qu’on vient de décrire permet de diversifier considérablement la palette des situations que l’on fait évaluer. En particulier, elle permet l’étude d’éventuels effets, interactifs ou non, qu’il eût été difficile d’examiner sinon.

Elle présente par contre l’inconvénient d’introduire des effets éventuels de grappe de différents ordres : ceux qui résultent du plan de sondage à plusieurs niveaux ayant permis de constituer l’échantillon des ménages, ceux qui résultent du fait que

---

5 Prendre un tel indicateur à sa valeur faciale sans tenir compte, dans les estimations, du fait qu’il est le ratio de deux variables aléatoires liées puisque le numérateur intervient partiellement dans le calcul du dénominateur ne va pas sans problèmes. Il serait sans doute souhaitable de disposer d’un modèle de décision permettant de relier sans ambiguïtés les évaluations latentes de chaque individu avec ses déclarations.
plusieurs personnes d'un même ménage sont interrogées, ceux qui résultent du fait qu'un même individu évalue différentes réponses. On pourrait envisager, puisqu'on dispose en fait d'un échantillon de questions, de procéder à leur égard comme on procède souvent vis-à-vis des répondants, eux aussi issus d'un échantillon, et d'introduire une post-stratification et des taux de pondération a posteriori en fonction de la plus ou moins grande représentativité du résultat. Pour l'instant, on ne l'a pas fait. Un examen assez sommaire suggère que l'effet de grappe lié à l'interrogation des membres d'un même ménage est faible.
Annexe 1 - Les niveaux de diplômes

<table>
<thead>
<tr>
<th>Niveau</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pas de diplôme</td>
</tr>
<tr>
<td>2</td>
<td>Certificat d’études primaires</td>
</tr>
<tr>
<td>3</td>
<td>BEPC</td>
</tr>
<tr>
<td>4</td>
<td>CAP, BEP ou diplôme équivalent</td>
</tr>
<tr>
<td>5</td>
<td>Baccalauréat technique</td>
</tr>
<tr>
<td>6</td>
<td>Baccalauréat général</td>
</tr>
<tr>
<td>7</td>
<td>Diplôme technique après le baccalauréat</td>
</tr>
<tr>
<td>8</td>
<td>BTS, DUT ou diplôme équivalent</td>
</tr>
<tr>
<td>9</td>
<td>Diplôme du 1er cycle d’Université</td>
</tr>
<tr>
<td>10</td>
<td>Diplôme de 2ème et 3ème cycle d’Université</td>
</tr>
<tr>
<td>Nombre de cases non renseignées</td>
<td>Nombre</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>0</td>
<td>2273</td>
</tr>
<tr>
<td>1</td>
<td>132</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>27</td>
<td>152</td>
</tr>
</tbody>
</table>
### Tableau 2 - Moyenne et dispersion en fonction des troncatures (Tous dossiers renvoyés, W2 exclu)

<table>
<thead>
<tr>
<th>Exclusion des valeurs supérieures à :</th>
<th>Exclusion des valeurs inférieures ou égales à :</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aucune</td>
</tr>
<tr>
<td></td>
<td>124,28</td>
</tr>
<tr>
<td></td>
<td>(212,16)</td>
</tr>
<tr>
<td>9.500</td>
<td>123,88</td>
</tr>
<tr>
<td></td>
<td>(202,56)</td>
</tr>
<tr>
<td>9.000</td>
<td>123,63</td>
</tr>
<tr>
<td></td>
<td>(196,56)</td>
</tr>
<tr>
<td>8.500</td>
<td>122,54</td>
</tr>
<tr>
<td></td>
<td>(170,52)</td>
</tr>
<tr>
<td>8.000</td>
<td>122,09</td>
</tr>
<tr>
<td></td>
<td>(158,92)</td>
</tr>
<tr>
<td>6.000</td>
<td>121,06</td>
</tr>
<tr>
<td></td>
<td>(132,40)</td>
</tr>
<tr>
<td>2.000</td>
<td>119,07</td>
</tr>
<tr>
<td></td>
<td>(95,99)</td>
</tr>
<tr>
<td>1.000</td>
<td>117,45</td>
</tr>
<tr>
<td></td>
<td>(80,68)</td>
</tr>
</tbody>
</table>
Tableau 3 - Tests de la normalité des distributions des appréciations de sa situation pour divers indicateurs possibles

(Tous dossiers renvoyés)

<table>
<thead>
<tr>
<th>Indicateur</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>KSL test de normalité</th>
</tr>
</thead>
<tbody>
<tr>
<td>WI</td>
<td>25.5</td>
<td>836.4</td>
<td>0.365</td>
</tr>
<tr>
<td>WI - MEAN</td>
<td>23.1</td>
<td>833.5</td>
<td>0.307</td>
</tr>
<tr>
<td>WI/MEAN</td>
<td>34.8</td>
<td>1341.3</td>
<td>0.367</td>
</tr>
<tr>
<td>Log WI</td>
<td>0.1</td>
<td>9.0</td>
<td>0.180</td>
</tr>
<tr>
<td>Log (WI/MEAN)</td>
<td>- 0.1</td>
<td>9.1</td>
<td>0.064</td>
</tr>
<tr>
<td>MEAN</td>
<td>19.5</td>
<td>496.2</td>
<td>0.315</td>
</tr>
<tr>
<td>Log MEAN</td>
<td>0.3</td>
<td>31.0</td>
<td>0.116</td>
</tr>
</tbody>
</table>
Tableau 4 – Eléments sur le pouvoir explicatif d’un modèle de base, pour divers indicateurs séparément.

<table>
<thead>
<tr>
<th>Indicateur</th>
<th>$R^2$</th>
<th>Sexe</th>
<th>Age</th>
<th>Niveau d'étude</th>
<th>Revenu</th>
<th>Épargne</th>
<th>PCS</th>
<th>Type de ménage</th>
<th>Lieu de résidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_1$</td>
<td>0.03</td>
<td>0.32</td>
<td>0.83</td>
<td>0.42</td>
<td>0.01</td>
<td>0.07</td>
<td>0.95</td>
<td>0.91</td>
<td>0.63</td>
</tr>
<tr>
<td>$W_1 - MEAN$</td>
<td>0.04</td>
<td>0.52</td>
<td>0.91</td>
<td>0.71</td>
<td>$\varepsilon$</td>
<td>0.29</td>
<td>0.92</td>
<td>0.70</td>
<td>0.61</td>
</tr>
<tr>
<td>$W_1/MEAN$</td>
<td>0.03</td>
<td>0.34</td>
<td>0.81</td>
<td>0.87</td>
<td>0.04</td>
<td>0.85</td>
<td>0.07</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>$Log W_1$</td>
<td>0.29</td>
<td>0.56</td>
<td>0.79</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>0.02</td>
<td>0.03</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>$Log (W_1/MEAN)$</td>
<td>0.33</td>
<td>0.59</td>
<td>0.26</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>0.02</td>
<td>0.003</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.04</td>
<td>0.73</td>
<td>0.86</td>
<td>0.37</td>
<td>0.003</td>
<td>0.03</td>
<td>0.83</td>
<td>0.15</td>
<td>0.19</td>
</tr>
<tr>
<td>$Log (MEAN)$</td>
<td>0.05</td>
<td>0.27</td>
<td>0.48</td>
<td>0.11</td>
<td>0.005</td>
<td>0.005</td>
<td>0.86</td>
<td>0.03</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note – $W_1$ est l’appréciation que l’on fait de sa propre situation MEAN est la moyenne des appréciations données sur les autres situations, celles proposées à évaluation. Les situations proposées à évaluation constituent approximativement un échantillon représentatif des situations envisageables dans la société française.
Graphique 1. Taux de non-réponse suivant l’âge de la personne proposée à évaluation
(Tous dossiers renvoyés)
Graphique 2. Répartition des notes inférieures à 1000
Graphique 3. Note maximale en fonction de la note minimale (après standardisation)
Graphique 4. Note moyenne attribuée à une question en fonction de sa place dans le questionnaire
LE DEVELOPPEMENT DE L'ENQUETE SUR
LA SANTE DANS LES COLLECTIVITES
CANADIENNES

Y. BELAND et J. DUFOUR

Statistique Canada - Division des méthodes des enquêtes auprès des ménages

SOMMAIRE

Dans le but de remédier aux principales lacunes statistiques en ce qui a trait aux déterminants de la santé, à l'état de santé et à l'utilisation du système de santé de la population canadienne à l'échelle des régions sociosanitaires, une nouvelle enquête, l'Enquête sur la santé dans les collectivités canadiennes (ESCC), est présentement en cours d'élaboration à Statistique Canada. L'ESCC est une enquête composée de deux volets distincts : une enquête à l'échelle régionale la première année et une enquête à l'échelle provinciale la deuxième année. L'enquête régionale, dont la collecte a débuté en septembre 2000, a comme but premier de produire des estimations transversales pour 136 régions sociosanitaires au Canada à partir d'un échantillon de plus de 130 000 répondants. L'échantillon de ménages a été sélectionné à partir de deux bases de sondage : une base aréolaire et une base de composition aléatoire de numéros de téléphone. Une personne âgée de 12 ans ou plus est choisie aléatoirement dans chaque ménage sélectionné et, selon certains critères, une deuxième personne (groupe d'âge 12 à 19 ans) pourrait être choisie dans le but d'augmenter la représentativité de ce groupe dans l'échantillon.

L'enquête provinciale, dont la collecte des données débutera en janvier 2002, a comme objectif principal de produire, à l'échelle provinciale, des estimations transversales portant sur différents aspects liés à la santé mentale et au bien-être des Canadiens. Un échantillon de 30 000 répondants est visé pour cette enquête. Cet article expose principalement les développements méthodologiques reliés à l'enquête à l'échelle régionale. Les options présentement à l'étude concernant le plan d'échantillonnage de l'enquête à l'échelle provinciale seront décrites de façon sommaire.

INSEE Méthodes 285
1. Introduction

En 1998, le Conseil consultatif sur l'infrastructure de la santé du ministre de la Santé fédéral, l'Institut canadien d'information sur la santé (ICIS) et Statistique Canada ont réuni plus de 500 personnes incluant des administrateurs des services de santé, des chercheurs, des dispensateurs de soins, des fonctionnaires, des groupes de revendication en matière de santé et des consommateurs pour mettre en place une stratégie permettant d'obtenir une meilleure information sur le système de santé au Canada. Le système d'information canadien actuel en matière de santé fournit des données fragmentées et incomplètes, des données qui ne peuvent être partagées facilement et qui ne peuvent être analysées en profondeur et, finalement, des résultats qui n'atteignent pas les Canadiens de façon régulière (ICIS, 1999a). Les consultations auprès d'experts ont mené à la création de l'Initiative du Carnet de route de l'information sur la santé (ICIS, 1999b) qui présente une vision canadienne commune pour moderniser et normaliser le système d'information sur la santé dans les diverses régions du pays.

Les principaux plans d'action du Carnet de route de l'information sur la santé comprennent : i) la réalisation de consultations continues avec les intervenants clés au pays, ii) l'obtention de données harmonisées et de techniques standardisées afin d'assurer une cohérence et une comparabilité entre la collecte, l'échange et l'interprétation des données sur la santé, iii) l'identification des priorités en ce qui a trait aux lacunes de données pour les services de santé et les coûts connexes, l'état de santé et les déterminants non médicaux de la santé, iv) la création d'un nouvel Institut de recherche en santé et v) la réalisation d'études spéciales et de rapports réguliers, faciles à comprendre, sur la santé des Canadiens et le système de santé au Canada.

En réponse à cette initiative, le gouvernement canadien a investi 95 millions de dollars canadiens dans le Carnet de route pour une période de quatre années. Cet investissement permettra la réalisation de nombreux projets qui seront gérés par l'ICIS, Santé Canada, Statistique Canada et différents partenaires au Canada. Parmi ces projets figure l'obtention d'estimations transversales à jour des déterminants de la santé, de l'état de santé et de l'utilisation du système de santé à l'échelle infraprovinciale, c'est-à-dire à l'échelle des régions sanitaires.

Au printemps 1999, Statistique Canada a mené des consultations à travers le pays auprès de plus de 225 intervenants clés du domaine de la santé afin de mieux cerner les besoins en matière de données. Ces consultations ont également permis d'établir certaines exigences qui ont dicté le développement méthodologique de l'ESCC. Statistique Canada se doit de développer un instrument de sondage flexible à trois volets : un contenu commun qui s'adresse à tous dans toutes les régions sanitaires du pays; un contenu optionnel, qui est adapté aux besoins de ces mêmes régions; et un contenu thématique, comprenant des sections sur des sujets d'actualité tels que la nutrition, les mesures physiques, la santé mentale et le
bien-être de soi. (Les contenus thématiques ont comme objectif de combler des besoins provinciaux.) Par sa flexibilité, l’ESCC doit offrir la possibilité d’augmenter la taille d’échantillon lorsque désiré, d’élargir le contenu si nécessaire et de servir de véhicule national pour la réalisation d’enquêtes sur la santé tout en tenant compte de l’infrastructure de Statistique Canada, des politiques en place et de sa capacité à réaliser des enquêtes. Il est évident que la mise sur pied d’une telle enquête d’envergure entraîne des changements majeurs à l’orga- nisme statistique.

La première section de cet article présente les différentes options qui ont été envisagées à Statistique Canada pour réaliser l’ESCC étant donné plusieurs contraintes imposées par l’infrastructure de l’Agence. Suit ensuite une discussion en profondeur sur le développement méthodologique de l’enquête à l’échelle régionale. Un aperçu du développement de l’enquête à l’échelle provinciale ainsi qu’une description des développements méthodologiques futurs complètent le tout.

2. Stratégie d’enquête du cycle biénal de l’ESCC

Afin de répondre efficacement aux besoins des utilisateurs tels qu’exprimés lors des consultations et en tenant compte de l’infrastructure existante de l’Agence, un cycle biénal de collecte de données se veut un choix logique. En effet, la nécessité de produire i) des estimations pour un vaste éventail de sujets pour 136 régions sociosanitaires et ii) des estimations pour un sujet thématique à l’échelle provinciale exige un échantillon d’une taille telle que recueillir l’information sur une plus courte période est opérationnellement très difficile à gérer.

Parmi les différents scénarios envisagés, deux ont fait l’objet d’une évaluation sérieuse (Statistique Canada, 1999a). L’option A requiert le développement de deux enquêtes distinctes et séparées : une enquête à l’échelle de la région sociosanitaire la première année auprès d’un échantillon d’un peu plus de 130 000 répondants et une enquête provinciale la deuxième année portant sur un échantillon de 30 000 répondants. L’enquête régionale aurait un contenu très diversifié qui serait adapté aux besoins régionaux alors que l’enquête provinciale traiterait en profondeur d’une thématique particulière.

L’option B représente une version annualisée de l’ESCC où une enquête principale à l’échelle de la région sociosanitaire est menée sur une période de deux ans auprès d’un échantillon de 160 000 répondants. À cette enquête, un supplément provincial (avec thématique particulière) serait administré auprès d’un sous-échantillon de 30 000 répondants. Tout comme l’option A, l’enquête principale serait adaptée aux besoins régionaux.

Les lignes directrices mises en place à Statistique Canada concernant la qualité des données (Statistique Canada, 1998) ont dicté le choix de la stratégie d’enquête pour l’ESCC. Des critères de qualité tels la pertinence, l’exactitude, l’intelligibilité, la
cohérence mais surtout l’actualité des données ont été les facteurs déterminants dans le choix de l’option A. Le cycle biennal de l’ESCC est donc formé de deux enquêtes : une enquête régionale la première année et une provinciale la deuxième. Pour la première année de réalisation, le volet régional est noté *cycle 1.1* et celui à l’échelle provinciale *cycle 1.2*.

3. Plan de sondage de l’enquête du cycle 1.1

Comme il a été mentionné précédemment, l’objectif premier de l’ESCC vise d’abord et avant tout à fournir des estimations transversales fiables à l’échelle de la région sociosanitaire en ce qui a trait aux déterminants de la santé, à l’état de la santé et à l’utilisation du système de santé. En juin 1999, les représentants provinciaux et territoriaux de concert avec Statistique Canada ont dénombré 136 régions sociosanitaires (RSS) : 133 RSS réparties dans les 10 provinces auxquelles s’ajoutent une RSS pour chacun des trois territoires. Comme l’enquête dans les trois territoires utilise un plan de sondage différent, la présente section ne fait référence qu’au plan utilisé dans les 10 provinces dans le but d’alléger l’écriture.

3.1 Formation des régions sociosanitaires

Les RSS sont des régions. Il existe au total 114 RSS telles que définies par les provinces, et plusieurs d’entre elles possèdent des subdivisions administratives supplémentaires. Pour des raisons d’ordre pratique et politique, Statistique Canada, avec l’accord des provinces, a révisé quelque peu les limites de certaines RSS selon la géographie du Recensement de la population de 1996 pour ainsi permettre la production de projections démographiques pour différents groupes âge-sexe (Wannell, 2000). Pour des fins statistiques, 133 RSS (définies selon les RSS originales et les subdivisions administratives supplémentaires) réparties dans les 10 provinces sont considérées pour l’enquête du cycle 1.1. Pour cette enquête, une subdivision géographique basée sur les RSS se veut un critère logique pour stratifier la population étant donné les profils sociodémographiques relativement homogènes de ces RSS.

3.2 Population cible

Pour l’enquête du cycle 1.1, l’ESCC ne vise que les personnes vivant dans des logements privés qui sont âgées de 12 ans ou plus. Sont exclus du champ de l’enquête les personnes qui vivent sur les réserves indiennes et sur les terres de la Couronne, les résidents des établissements, les membres à temps plein des Forces armées canadiennes et les résidents de certaines régions éloignées. L’ESCC couvre environ 97% des 12 ans ou plus de la population canadienne. Il est bon de noter que l’Enquête nationale sur la santé de la population (ENSP) possède déjà un volet institutionnel qui traite des résidents à long terme de centres hospitaliers et d’établissements de soins pour bénéficiaires internes (Statistique Canada, 1996).
3.3 Taille et répartition de l'échantillon

Afin de fournir des estimations fiables aux 133 RSS et en tenant compte du budget alloué à l'enquête du cycle 1.1, un échantillon net de base de 130 750 répondants est visé. Bien que l'objectif de produire des estimations fiables à l'échelle de la RSS soit primordial, la qualité des estimations à l'échelle provinciale pour certaines caractéristiques clés demeure tout de même un but à atteindre. Divers scénarios ont été envisagés pour répartir cet échantillon entre les provinces et les RSS. Comme la taille et le nombre des RSS varient considérablement d'une province à l'autre, il est difficile d'établir un équilibre entre les besoins régionaux et provinciaux. Une approche axée principalement vers les RSS pénalise les provinces fortement peuplées ayant un nombre relativement petit de RSS tandis qu'une approche axée vers les provinces procure l'effet inverse. La démarche qui a été adoptée est une démarche en trois étapes qui permet d'accorder une importance relativement égale aux RSS et aux provinces. Les deux premières étapes visent à répartir l'échantillon total entre les provinces en fonction de leur population respective et du nombre de RSS qui s'y trouve; à la troisième étape, on s'attarde à répartir l'échantillon entre les RSS de chaque province.

À la première étape, on répartit entre les provinces à peu près la moitié de l'échantillon de 130 750 répondants en fonction du nombre de RSS qu'elles renferment. Puisque l'ESCC a principalement pour objet de fournir des estimations régionales (RSS) fiables, quelque 500 unités d'échantillonnage sont attribuées à chaque RSS. Le chiffre de 500 est considéré comme le nombre minimal de répondants à prévoir pour chaque RSS de façon à conférer un niveau minimal de fiabilité aux principales caractéristiques relatives à la santé. Une exception toutefois : les RSS où la fraction de sondage résultante est trop élevée; auquel cas le taux d'échantillonnage est majoré à 1 ménage sur 20. Un total de 65 830 unités d'échantillonnage ont ainsi été réparties suite à cette première étape. La deuxième étape consiste à répartir le reste de l'échantillon total (64 920 unités) entre les provinces proportionnellement à la taille de population. Une exception : un minimum de 1 000 unités est attribué à la province de l'Île-du-Prince-Édouard. La taille totale de l'échantillon dans une province donnée s'obtient en additionnant les tailles obtenues aux deux premières étapes. Le tableau 1 fournit une répartition détaillée par province de l'échantillon de l'enquête du cycle 1.1.
Tableau 1 - Taille des échantillons provinciaux pour l’enquête du cycle 1.1 de l’ESCC

<table>
<thead>
<tr>
<th>Province</th>
<th>Taille de la population</th>
<th>Nombre de RSS</th>
<th>1ère étape 500 / RSS</th>
<th>2e étape Prop. à X</th>
<th>Échantillon total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terre-Neuve</td>
<td>551 000</td>
<td>6</td>
<td>2 780</td>
<td>1 230</td>
<td>4 010</td>
</tr>
<tr>
<td>Île-du-Prince-Édouard</td>
<td>135 000</td>
<td>2</td>
<td>1 000</td>
<td>1 000</td>
<td>2 000</td>
</tr>
<tr>
<td>Nouvelle-Écosse</td>
<td>909 000</td>
<td>6</td>
<td>3 000</td>
<td>2 040</td>
<td>5 040</td>
</tr>
<tr>
<td>Nouveau-Brunswick</td>
<td>738 000</td>
<td>7</td>
<td>3 500</td>
<td>1 650</td>
<td>5 150</td>
</tr>
<tr>
<td>Québec</td>
<td>7 139 000</td>
<td>16</td>
<td>8 000</td>
<td>16 280</td>
<td>24 280</td>
</tr>
<tr>
<td>Ontario</td>
<td>10 714 000</td>
<td>37</td>
<td>18 500</td>
<td>23 760</td>
<td>42 260</td>
</tr>
<tr>
<td>Manitoba</td>
<td>1 114 000</td>
<td>11</td>
<td>5 500</td>
<td>2 500</td>
<td>8 000</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>990 000</td>
<td>11</td>
<td>5 400</td>
<td>2 320</td>
<td>7 720</td>
</tr>
<tr>
<td>Alberta</td>
<td>2 697 000</td>
<td>17</td>
<td>8 150</td>
<td>6 050</td>
<td>14 200</td>
</tr>
<tr>
<td>Colombie-Britannique</td>
<td>3 725 000</td>
<td>20</td>
<td>10 000</td>
<td>8 090</td>
<td>18 090</td>
</tr>
<tr>
<td>Canada</td>
<td>29 000 000</td>
<td>133</td>
<td>65 830</td>
<td>64 920</td>
<td>130 750</td>
</tr>
</tbody>
</table>

La troisième et dernière étape de la démarche consiste à répartir l'échantillon provincial entre les RSS de la province proportionnellement à la racine de la taille de la population des régions, tout en s'assurant d'attribuer à chaque RSS un échantillon minimal de 500 unités. Il est à noter que la règle de la fraction de sondage maximum de un ménage sur 20 dans une RSS s'applique toujours. Cette démarche en trois étapes permet d'avoir pour chaque RSS un échantillon suffisant tout en ne perturbant que très légèrement la répartition provinciale des tailles d'échantillon. Le tableau 2 donne un aperçu sommaire de la répartition des RSS selon leur population, ainsi que de la taille moyenne de l'échantillon prévue par catégorie de RSS.
Tableau 2 – Taille moyenne des échantillons par catégorie de RSS

<table>
<thead>
<tr>
<th>Taille de la population</th>
<th>Nombre de RSS</th>
<th>Taille moyenne des échantillons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petites RSS</td>
<td>Moins de 75 000</td>
<td>41</td>
</tr>
<tr>
<td>RSS moyennes</td>
<td>75 000 - 240 000</td>
<td>60</td>
</tr>
<tr>
<td>Grandes RSS</td>
<td>240 000 - 640 000</td>
<td>25</td>
</tr>
<tr>
<td>Très grandes RSS</td>
<td>640 000 +</td>
<td>7</td>
</tr>
</tbody>
</table>

La taille de l’échantillon finale visée pour l’enquête du cycle 1.1 de l’ESCC est de 130 750 répondants et ce, malgré l’opportunité qu’ont eue les RSS et les autorités provinciales d’augmenter les tailles d’échantillon en achetant des unités supplémentaires; les tailles allouées ont été jugées satisfaisantes. Ces tailles ont été gonflées avant d’effectuer la collecte pour tenir compte de la vacance et de la non-réponse.

3.4 Choix de la base de sondage

Il existe plusieurs bases de sondage à Statistique Canada qui sont utilisées par les différentes enquêtes auprès des ménages pouvant être souhaitables pour une enquête sur la santé. La base aréolaire mise en place pour l’Enquête sur la population active (EPA) à l’intérieur de laquelle il est possible de sélectionner soit un échantillon de logements nouveaux (non sélectionnés par l’EPA ou par d’autres enquêtes), soit un échantillon de logements échus de l’EPA est certes une base de choix pour les enquêtes auprès des ménages. Par contre, le registre des adresses (RA) de Statistique Canada, la base de composition aléatoire (CA) de numéros de téléphone et les fichiers administratifs provinciaux de la santé sont également des bases potentielles. (Il est bon de noter que pour l’EPA, les logements sélectionnés demeurent dans l’échantillon pour une durée de six mois consécutifs et sont par la suite remplacés; ces logements deviennent alors disponibles pour d’autres enquêtes.)

Le choix d’une base de sondage pour tirer un échantillon dépend de plusieurs facteurs mais la base doit d’abord et avant tout correspondre le plus possible à la population cible de l’enquête. De plus, la création, l’utilisation, la mise à jour et la vérification de la base de sondage doivent respecter les contraintes opérationnelles et budgétaires de l’enquête. Pour l’enquête du cycle 1.1, il a été décidé d’utiliser deux bases de sondage chevauchantes : la base aréolaire mise en place pour l’EPA et la base CA de numéros de téléphone. Dans le cadre de l’ESCC, la base aréolaire est utilisée comme base principale alors que la base CA agit comme complément à la base principale dans certaines RSS.

Outre le fait que la population cible de l’ESCC soit la même que celle de l’EPA, les avantages liés au choix de la base aréolaire mise en place pour l’EPA pour tirer l’échantillon de l’ESCC sont indéniables. L’infrastructure déjà en place pour
effectuer les mises à jour en fonction des nouveaux logements, des logements démolis et des unités hors du champ d’observation de même que tout le processus d’évaluation de la couverture de la base sont des atouts certains. De plus, étant donné que plusieurs autres enquêtes auprès des ménages de Statistique Canada utilisent également cette base aréolaire, le contrôle des chevauchements d’échantillon entre les enquêtes est facilité. Pour plusieurs raisons mais principalement pour une question de fardeau de réponse, il a été décidé de sélectionner un échantillon de logements nouveaux et non un échantillon de logements échus de l’EPA.

Bien que les limites reliées à l’utilisation de la base CA soient évidentes (sous-couverture des ménages sans téléphone (~2%) ou avec téléphone cellulaire seulement (estimé entre 1% et 2% selon les secteurs), taux de réponse généralement plus faibles et la nécessité d’effectuer plusieurs appels avant de contacter un ménage valide), différentes raisons justifient ce choix. En fait, il y a trois facteurs principaux : i) minimiser les coûts de collecte des données dans les régions où la disponibilité des intervieweurs est restreinte, ii) pallier l’incapacité de la base aréolaire de fournir l’échantillon nécessaire pour certaines RSS et iii) disposer en permanence d’une structure flexible de collecte de données par téléphone. Comme la base aréolaire de l’EPA a été conçue en fonction de l’économie du marché du travail et de l’économie canadienne en général, sa principale stratification géographique est basée sur les régions économiques d'assurance-emploi telles que définies par les provinces. Ces limites géographiques sont souvent différentes de celles délimitant les RSS. Cela a pour effet de limiter le rendement possible de la base aréolaire dans certaines RSS forçant par le fait même le recours à la base CA pour sélectionner les ménages pour ainsi obtenir les tailles d’échantillon désirées. De plus, disposer en permanence d’une structure de collecte de données par téléphone permet de répondre rapidement aux requêtes régionales et/ou provinciales quant à l’achat d’unités d’échantillonnage et/ou de contenu spécifique qui peuvent survenir à tout moment durant la période de collecte. Le grand défi que pose le recours à la base CA aux fins de l’ESCC réside dans l’association des RSS aux numéros de téléphone à l’aide de divers fichiers administratifs; cet aspect sera abordé dans une section ultérieure. La stratification de la base CA en RSS, jumelée aux avantages de la base aréolaire de l’EPA, font en sorte que les deux bases de sondage répondent très bien aux besoins de l’enquête du cycle 1.1.

3.5 Stratégie d’échantillonnage

L’échantillon total visé par l’ESCC est de 130 750 répondants; la majorité (115 000 répondants) provient de la base aréolaire où des interviews personnelles sont menées alors que le reste provient de la base CA où, bien sûr, des interviews téléphoniques sont réalisées. Les répondants de la base aréolaire proviennent d’un échantillon de 97 000 ménages où une personne par ménage est sélectionnée dans 79 000 d’entre eux alors que deux personnes sont sélectionnées à l’intérieur des 18 000 autres ménages afin d’augmenter la représentativité du groupe d’âge 12 à 19 ans. On complète l’échantillon en sélectionnant 15 750 ménages à partir de la base CA dans
lesquels seulement une personne est sélectionnée. La présente section décrit en détails les mécanismes de sélection des ménages mis en place pour chacune des bases de sondage. La stratégie pour sélectionner les personnes à l’intérieur des ménages est également décrite.

3.5.1 Plan d’échantillonnage des ménages de l’EPA

Le plan de base de l’EPA est un échantillon stratifié à plusieurs degrés dans lequel le logement est l’unité d’échantillonnage finale (Gambino, Singh, Dufour, Kennedy et Lindeyere, 1998). Pour les fins du plan, chaque province est répartie en trois types de régions : grands centres urbains, villes urbaines, et régions rurales. Des strates géographiques ou socioéconomiques sont formées au sein de chaque grand centre urbain. À l’intérieur des strates, de 150 à 250 logements sont regroupés pour former des grappes. Certains centres urbains ont des strates distinctes pour les appartements et pour les secteurs de dénombrement (SD) du recensement où le revenu moyen par ménage est élevé. On choisit ensuite, dans chaque strate, six grappes ou immeubles résidentiels (parfois 12 ou 18 appartements) par une méthode d’échantillonnage aléatoire avec probabilité proportionnelle à la taille (PPT), la taille correspondant au nombre de ménages. Le plan de l’EPA prévoit un renouvellement du sixième de l’échantillon à tous les mois.

Les autres villes et régions rurales de chaque province sont stratifiées sur une base géographique d’abord, puis selon les caractéristiques socioéconomiques. Dans la plupart des strates, six grappes (habituellement des SD du recensement) sont sélectionnées au moyen de la méthode PPT. Dans l’éventualité d’une densité faible de la population, on a recours à un plan à trois degrés en sélectionnant d’abord deux ou trois unités primaires d’échantillonnage (UPÉ), qui correspondent habituellement à des groupes de SD, et en divisant chaque UPÉ en grappes, dont six sont échantillonnées. La sélection se fait à chaque degré au moyen de la méthode PPT.

Une fois les opérations de listage d’une nouvelle grappe terminées, on obtient l’échantillon grâce à un échantillonnage systématique de logements. Le tableau 3 donne un aperçu des types d’UPÉ utilisés pour l’ensemble de l’échantillon de l’EPA. Le rendement est le nombre de ménages sélectionnés dans le cadre de l’EPA pour un mois donné. Les taux d’échantillonnage étant déterminés d’avance, il existe souvent un écart entre la taille prévue de l’échantillon et les chiffres effectivement obtenus. Le rendement de l’échantillon, par exemple, est parfois excessif. Cela se produit surtout dans les secteurs où un accroissement du nombre de logements occasionné par les nouvelles constructions par exemple est observé. Pour réduire le coût de la collecte, on corrige un rendement excessif en éliminant systématiquement au hasard une partie des unités sélectionnées à l’origine, et en modifiant les poids du plan d’échantillonnage. Cette opération, habituellement réalisée à un niveau agrégé, est appelée stabilisation de l’échantillon. De plus, on augmente la taille requise des échantillons, en ménages, pour tenir compte des logements, l’expérience ayant prouvé que 12% de tous les logements ne sont pas occupés par des ménages faisant
partie du champ d’observation (certains logements étant vacants ou occupés sur une base saisonnière, d’autres étant occupés par des ménages non visés par l’enquête). Le plan d’échantillonnage donne, pour l’EPA, environ 68 000 logements sur une base mensuelle parmi lesquels approximativement 60 000 ménages constituent l’échantillon.

### Tableau 3 – Principales unités primaires de l’EPA, tailles et rendements

<table>
<thead>
<tr>
<th>Secteurs</th>
<th>Unités primaires d’échantillonnage (UPÉ)</th>
<th>Taille (nombre de ménages par UPÉ)</th>
<th>Rendement (nombre de ménages dans l’échantillon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toronto, Montréal, Vancouver</td>
<td>Grappe</td>
<td>200-250</td>
<td>6</td>
</tr>
<tr>
<td>Autres grands centres urbains</td>
<td>Grappe</td>
<td>150-200</td>
<td>8</td>
</tr>
<tr>
<td>Base d’appartements</td>
<td>Immeubles d’appartements</td>
<td>variable</td>
<td>5</td>
</tr>
<tr>
<td>Secteurs ruraux, petits centres urbains</td>
<td>Secteurs de dénombrement</td>
<td>300</td>
<td>10</td>
</tr>
</tbody>
</table>

---

### 3.5.2 Adaptation de la stratégie de l’EPA pour l’ESCC

Non seulement l’échantillon de l’ESCC est choisi à partir de la base aréolaire mise en place pour l’EPA tel que mentionné précédemment, mais le mécanisme de sélection des ménages de l’EPA est également utilisé. Afin de rencontrer les exigences de l’ESCC en termes de tailles d’échantillon à l’échelle de la RSS, des modifications ont toutefois été apportées à ce mécanisme de sélection.

Pour obtenir un échantillon visé de 97 000 ménages, 123 000 logements doivent être sélectionnés à partir de la base aréolaire (en tenant compte des logements vacants et des ménages non-répondants). Le plan de l’EPA donne environ 68 000 logements répartis dans les régions économiques alors que l’ESCC requiert un échantillon total de 123 000 logements répartis dans les RSS qui possèdent des limites géographiques différentes de celles des régions économiques de l’EPA. À l’échelle globale, l’ESCC nécessite près du double de logements que ce que génère le mécanisme de sélection du plan de l’EPA, soit un facteur de gonflement de 1,8 (123/68). À l’échelle des RSS cependant, les facteurs de gonflement varient de 0,6 à 6,0, nécessitant donc certains ajustements.

Les modifications apportées au mécanisme de sélection dans une RSS varient selon l’importance des facteurs de gonflement. Pour les RSS qui ont un facteur inférieur ou égal à 1, une simple opération de stabilisation telle que décrite plus haut est
appliquée à l’échantillon de logements. Pour celles ayant un facteur supérieur à 1 et inférieur ou égal à 2, le processus d’échantillonnage des logements à l’intérieur d’une UPÉ est répété une seconde fois et ce, pour toutes les UPÉ sélectionnées faisant partie d’une même RSS. Pour les RSS ayant un facteur supérieur à 2 et inférieur ou égal à 4, le processus d’échantillonnage des UPÉ ainsi que celui des logements dans une UPÉ sont répétés une seconde fois. Pour les RSS ayant un facteur entre 4 et 6, le processus d’échantillonnage des UPÉ est répété non pas une mais deux fois tandis que celui des logements n’est répété qu’une seule fois. Dans tous les cas où l’approche choisie occasionne un surplus inutile de logements, une stabilisation est effectuée.

À noter que les modifications apportées au mécanisme de l’EPA se résument à, au plus, tripler le nombre d’UPÉ sélectionnées et, au plus, doubler le nombre de logements sélectionnés dans les UPÉ; d’où le facteur de gonflement maximum de 6,0. À l’échelle des RSS, les facteurs de gonflement ont été volontairement majorés à 6,0 pour deux raisons : limiter les opérations de listage de grappes (chaque nouvelle UPÉ sélectionnée nécessite un listage) et éviter les effets de grappes possibles occasionnés par un trop grand nombre de logements sélectionnés dans une même UPÉ. Cette limite imposée au facteur de gonflement de certaines RSS a, par ricochet, dicté le nombre nécessaire de ménages provenant de la base CA. Le lecteur trouvera des détails supplémentaires dans Morano, Lessard et Bélard (2000).

3.5.3 Sélection des ménages de la base CA

La stratégie d’échantillonnage des ménages par composition aléatoire (CA) de numéros de téléphone utilise la méthode d’élimination des banques inactives (MÉBI). Une banque (indicatif régional à trois chiffres plus les cinq premiers chiffres d’un numéro de téléphone qui en compte sept) est considérée comme admissible aux fins de l’échantillonnage si elle comprend au moins un numéro de téléphone. L’information est obtenue à partir des fichiers de facturation fournis par les compagnies de téléphone ainsi que d’autres fichiers administratifs publiques. Les banques admissibles sont regroupées pour former des strates CA. À l’intérieur d’une strate CA, une banque est choisie aléatoirement et un numéro entre 00 et 99 est généré de façon aléatoire pour former un numéro de téléphone complet à 10 chiffres (3+7). Ce processus est répété jusqu’à ce que le nombre de numéros de téléphone souhaités à l’intérieur de la strate CA soit atteint. Souvent, le numéro généré correspond à un numéro de téléphone qui est hors du champ de l’enquête ou à un numéro qui n’est pas en service. Afin d’atteindre les objectifs visés quant à la taille d’échantillon nécessaire, de nombreux numéros de téléphone additionnels doivent être générés. Dans le cadre de l’ESCC, les taux de succès varient de 15% à 61% parmi les RSS où un échantillon de ménages de la base CA est nécessaire. À l’échelle globale, un échantillon d’un peu plus de 51 000 numéros de téléphone est donc nécessaire (en tenant également compte des ménages non-répondants) pour les besoins de l’ESCC afin d’obtenir les 15 750 ménages répondants requis.
Le grand défi que pose le recours à la base CA aux fins de l'ESCC réside dans le regroupement des banques admissibles pour former des strates CA qui correspondent le plus près possible aux limites géographiques des RSS pour ainsi pouvoir contrôler efficacement la sélection de l'échantillon. Cet exercice nécessite l'utilisation de fichiers administratifs tels que les fichiers de facturation des compagnies de téléphone ainsi que les bottins téléphoniques électroniques. Les codes postaux présents sur ces fichiers servent de clés pour apparier ceux-ci avec les différents fichiers de la Division de la géographie de Statistique Canada dans le but d'assigner une région sociosanitaire aux numéros de téléphone. Le nouveau lien numéro de téléphone / RSS ainsi obtenu est utilisé pour regrouper les banques en strates CA.

Pour contrôler parfaitement la sélection de l'échantillon de ménages à partir de la base CA en utilisant la MÉBI, les secteurs géographiques couverts par les strates CA doivent coïncider exactement avec les RSS. Parmi les quelques 67 RSS ayant recours à la base CA, cela se produit pour deux RSS seulement. Pour les autres RSS, le taux de couverture des strates CA en termes de numéros de téléphone dans une RSS donnée (selon les fichiers administratifs) varient de 92% à 99%. Dans le cadre de l'ESCC, ces taux de couverture ne soulèvent aucune préoccupation car la base CA ne sert qu'à compléter la base aréolaire qui elle, par définition, couvre la presque totalité des secteurs visés. (À noter que la faiblesse de la base CA de certaines RSS sera prise en compte lors de la pondération des unités dans l’étape du traitement des données.) Les détails complets reliés à la stratification de la base CA en RSS sont fournis dans Morano et coll. (2000).

3.5.4 Sélection des personnes interviewées

L'ESCC cible les personnes vivant dans des logements privés occupés qui sont âgées de 12 ans ou plus. Différentes options pour définir la règle de sélection des personnes à l’intérieur des ménages échantillonnés ont été étudiées. Interviewer plus d’une personne dans un même ménage permet une économie des coûts de collecte; une grande partie de ces coûts étant en effet attribuable aux démarches pour rejoindre le ménage, interviewer plus d’une personne ne coûte guère plus. Par contre, les fortes similitudes observées chez les membres d’un même ménage peuvent occasionner un certain effet de grappe non souhaitable pour quelques caractéristiques importantes de l’enquête (une des composantes clés de l’ESCC est l’utilisation du système de santé). De plus, le fardeau de réponse du ménage se trouve accru.

D’un autre côté, sélectionner une seule personne par ménage représente un accroissement significatif des coûts de collecte car un plus grand nombre de ménages doit être échantillonné. Et encore, comme les chances de faire partie de l’échantillon sont inversement proportionnelles au nombre de personnes à l’intérieur du ménage, certaines catégories de personnes (selon l’âge) sont soit sous, soit surreprésentées. Le tableau 4 présente les distributions par groupe d’âge observées du Recensement de la population de 1996 (colonne 2), d’un échantillon de l’EPA où toutes les

296 INSEE Méthodes
personnes d’un même ménage sont sélectionnées (colonne 3) ainsi que la distribution moyenne de 100 échantillons simulés de l’ESCC où une personne par ménage est sélectionnée (colonne 4). Les résultats démontrent clairement qu’en sélectionnant une seule personne par ménage, le groupe des 12 à 19 ans est grandement sous-représenté alors que les personnes âgées sont surreprésentées.

**Tableau 4 – Distribution de l'échantillon par groupe d'âge**

<table>
<thead>
<tr>
<th>Groupe d'âge</th>
<th>Recensement 1996</th>
<th>Échantillon EPA</th>
<th>*Échantillon ESCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col.1</td>
<td>Col. 2</td>
<td>Col. 3</td>
<td>Col. 4</td>
</tr>
<tr>
<td>12-19</td>
<td>13,2</td>
<td>13,7</td>
<td>8,5</td>
</tr>
<tr>
<td>20-29</td>
<td>16,4</td>
<td>14,4</td>
<td>14,3</td>
</tr>
<tr>
<td>30-44</td>
<td>30,8</td>
<td>28,7</td>
<td>29,1</td>
</tr>
<tr>
<td>45-64</td>
<td>25,8</td>
<td>28,0</td>
<td>27,9</td>
</tr>
<tr>
<td>65+</td>
<td>13,8</td>
<td>15,2</td>
<td>20,2</td>
</tr>
</tbody>
</table>

* Moyenne des distributions sur 100 simulations

Durant le processus de consultation pancanadienne, les représentants régionaux et provinciaux ont exprimé leurs préoccupations quant à la nécessité d’augmenter la représentativité des jeunes et des personnes âgées dans l’échantillon de l’ESCC comparativement à la distribution de la population canadienne observée lors du recensement de 1996. Le but étant de permettre des analyses plus approfondies de ces deux importants sous-groupes de la population. Donc implanter une règle de sélection où une personne par ménage serait aléatoirement choisie ne répondrait qu’à une partie de la requête.

Plusieurs options sont possibles pour augmenter la représentativité des 12 à 19 ans dans l’échantillon telles que changer les probabilités de sélection des personnes dans un ménage, augmenter la taille d’échantillon des ménages pour ensuite mettre en place une méthode de rejet (Tambay et Mohl, 1995), ou bien implanter un plan à deux phases. Pour des raisons de coût, d’efficacité du plan, du fardeau de réponse ainsi que des contraintes opérationnelles, la règle de sélection des personnes mise en place pour l’enquête du cycle 1.1 se veut un compromis.

Pour tous les ménages provenant de la base CA, une seule personne âgée de 12 ans ou plus est choisie parmi tous les membres du ménage. Pour les ménages provenant de la base aréolaire, une seule personne (12 ans ou plus) est sélectionnée dans 79 000 ménages alors que deux personnes (12 ans ou plus) sont choisies dans 18 000 ménages. La règle de sélection des personnes pour les ménages de la base aréolaire est définie en fonction de la composition du ménage et elle a pour but d’augmenter la représentativité des 12 à 19 ans dans l’échantillon (Thivierge, 1999).
Le tableau 5 décrit la règle de sélection. À noter que cette règle ne s’applique pas aux ménages provenant de la base CA car des interviews téléphoniques y sont menées. (Étant donné le contenu, il a été jugé que deux interviews téléphoniques occasionneraient des problèmes opérationnels non négligeables.)

Tableau 5 – Règle de sélection des personnes selon la composition des ménages – base aréolaire seulement

<table>
<thead>
<tr>
<th>Nombre de 12-19 ans</th>
<th>Nombre de personnes âgées de 20 ans ou plus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>3+</td>
<td>A</td>
</tr>
</tbody>
</table>

A : sélection aléatoire d’une personne âgée de 12 ans ou plus
B : sélection aléatoire de deux personnes âgées de 12 ans ou plus
C : sélection aléatoire d’une personne dans le groupe 12-19 ET sélection aléatoire d’une personne de 20 ans ou plus

Le tableau 6 présente la distribution anticipée de l’échantillon de l’ESCC en sélectionnant deux personnes (selon la règle décrite au tableau 5) dans 18 000 ménages et une personne de façon aléatoire dans tous les autres ménages sélectionnés. Les résultats montrent que les deux sous-groupes d’intérêt (12 à 19, et 65 et plus) sont maintenant surreprésentés dans l’échantillon sans toutefois pénaliser outre mesure les autres groupes d’âge.

Tableau 6 – Distribution anticipée de l’échantillon de l’ESCC par groupe d’âge

<table>
<thead>
<tr>
<th>Groupe d’âge</th>
<th>Recensement 1996</th>
<th>*Échantillon ESCC (1 personne seulement)</th>
<th>*Échantillon ESCC (1 et 2 personnes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-19</td>
<td>13,2</td>
<td>8,5</td>
<td>14,9</td>
</tr>
<tr>
<td>20-29</td>
<td>16,4</td>
<td>14,3</td>
<td>13,1</td>
</tr>
<tr>
<td>30-44</td>
<td>30,8</td>
<td>29,1</td>
<td>28,1</td>
</tr>
<tr>
<td>45-64</td>
<td>25,8</td>
<td>27,9</td>
<td>26,3</td>
</tr>
<tr>
<td>65+</td>
<td>13,8</td>
<td>20,2</td>
<td>17,6</td>
</tr>
</tbody>
</table>

* Moyenne des distributions sur 100 simulations
3.6 Conception du questionnaire

Un des objectifs du processus de consultation tenu au printemps 1999 était de créer un consensus pour l’établissement des priorités en ce qui a trait aux exigences des utilisateurs de l’information sur la santé afin de concevoir le questionnaire de l’enquête du cycle 1.1. Ce processus de consultation comprenait une série d’ateliers d’une journée, de discussions et l’utilisation d’une feuille de travail pour ordonner les sujets pouvant figurer dans le questionnaire par ordre de priorité (Statistique Canada, 1999b). Les consultations ont révélé une grande variabilité dans les priorités et les besoins de données pour les différentes régions et provinces, mais l’information recueillie a toutefois permis d’établir, pour chaque RSS, une liste de sujets par priorité (Statistique Canada, 1999c). Avec cette information en main, il s’est avéré évident qu’il était impossible de satisfaire pleinement les besoins en données des 136 RSS en adoptant une seule version d’un questionnaire. (À noter que la durée maximale d’une interview a été fixée à 45 minutes pour des raisons de fardeau de réponse.)

Il est important de mentionner qu’étant donné le temps de développement alloué à cette enquête, il existait déjà un ensemble de questions pour tous les sujets proposés. Ces ensembles de questions, qui sont appelés modules, ont tous été utilisés dans d’autres enquêtes, procurant ainsi un temps d’interview individuel associé à chaque module. (Ces modules proviennent de l’ENSP de Statistique Canada, du Behavioral Risk Factor Surveillance System des États-Unis, de l’Enquête sociale et de santé du Québec, de l’Enquête sur la promotion de la santé et de l’Enquête sur la santé en Ontario.)

Afin d’adapter le plus possible le questionnaire de l’enquête du cycle 1.1 aux besoins régionaux, une stratégie innovatrice et unique à l’ESCC a été mise en place. Le développement du contenu final du questionnaire est divisé en deux parties : un contenu commun d’une durée de 35 minutes et un contenu optionnel d’une durée de 10 minutes. En combinant toutes les priorités des RSS pour ne créer qu’une seule liste de priorités, le contenu commun a été identifié comme étant ni plus ni moins que les modules jugés prioritaires par la majorité des RSS et ce, jusqu’à concurrence d’un temps total de 30 minutes. Le contenu commun est complété par un module renfermant les questions sociodémographiques essentielles de base d’une durée de cinq minutes. Les questions relatives au contenu commun sont posées à toutes les unités échantillonnées afin de recueillir sur une base continue les données élémentaires de base sur la santé.

Pour ce qui est du développement du contenu optionnel, chaque RSS a eu la possibilité de personnaliser leur questionnaire en fonction de leurs besoins régionaux (Statistique Canada, 1999d). Chaque RSS a donc choisi une combinaison de modules d’au plus 10 minutes parmi 28 modules de durée variable (les modules optionnels offerts aux RSS étaient tous les modules ne faisant pas partie du contenu.
commun). La durée des modules optionnels varie de 20 secondes à six minutes et tous les modules sont indivisibles. (Il est à noter que l'utilisation du mode d'interview assistée par ordinateur (IAO) combinée à l'aspect modulaire du questionnaire offre la possibilité de créer facilement plusieurs versions différentes du questionnaire.) L'annexe A fournit la liste de tous les modules communs ainsi que les 28 modules optionnels offerts aux RSS.

Il est intéressant de mentionner que la majorité des provinces ont décidé de gérer cette sélection afin d'uniformiser le plus possible le contenu retenu dans leur province. À la fin de ce processus, 25 versions différentes du questionnaire ont été identifiées pour l'enquête du cycle 1.1. Le lecteur est référé au site internet de l’ESCC pour obtenir la liste des modules optionnels tels que sélectionnés par chaque RSS ainsi que la version complète du questionnaire de l’enquête du cycle 1.1.

Mis à part les nombreuses mises à l’essai opérationnelles, plusieurs essais qualitatifs (groupes de discussion) ont eu lieu sur le terrain afin de valider le questionnaire, la durée de l’interview et de recueillir les commentaires de certains répondants et intervieweurs (Lévesque, 2000). Ces essais ont également permis de mettre à l’épreuve l’ordonnancement des modules et d’évaluer la compréhension et la formulation des questions. Suite aux résultats de cette démarche, plusieurs modifications ont été apportées au questionnaire.

3.7 Collecte des données

La collecte des données, qui a débuté en septembre 2000, s'échelonne sur une période de 12 mois afin d'étendre dans le temps la charge de travail des intervieweurs sur le terrain et d'éliminer les effets saisonniers possibles sur certaines caractéristiques relatives à la santé (les activités physiques par exemple). Toutes les RSS sont visitées à chaque mois de collecte ce qui permettra de diffuser des données préliminaires sur une base trimestrielle lorsque les bases de données accumulées dans le temps seront jugées suffisantes. Toutes les interviews, qu’elles soient personnelles ou téléphoniques, sont assistées par ordinateur. Tous les intervieweurs de Statistique Canada sont formés expressément pour mener des interviews selon la méthode IAO. Durant les mois de collecte, une structure de surveillance et de contrôle est également en application.

Quelques semaines avant de contacter pour la première fois un ménage provenant de la base aréolaire, une lettre d’introduction est envoyée pour préparer la venue d’un intervieweur et souligner l’importance de l’enquête ainsi que pour solliciter la coopération du ménage (aucune lettre n’est envoyée aux ménages provenant de la

1 www.statcan.ca/enquetes_sante
base CA). Tous les cas de non-réponse sont traités sur le terrain à la faveur d'un réexamen le mois suivant. Les cas de non-réponse après trois mois consécutifs sont soumis aux intervieweurs principaux et un 13ᵉ mois de collecte est prévu à la fin pour effectuer un suivi supplémentaire auprès de ce groupe.

4. Aperçu de l’enquête du cycle 1.2

L’enquête du cycle 1.2 a comme objectif principal de produire des estimations transversales portant sur différents aspects reliés à la santé mentale et au bien-être des Canadiens à l’échelle provinciale. De plus, l’enquête recueillera des données sur les déterminants de la santé mentale tant positive que négative, sur l’utilisation des services de soins de santé mentale ainsi que des données sur les impacts sociaux et les coûts associés à la santé mentale. Le tout sera complété par la collecte d’une série de caractéristiques sociales et démographiques.

L’échantillon visé pour cette enquête est de 30 000 répondants répartis dans les dix provinces¹. La population cible comprend les personnes vivant dans les ménages privés qui sont âgées de 15 ans ou plus (les exclusions sont les mêmes que celles de l’enquête du cycle 1.1). La collecte des données débutera en janvier 2002 et s’échelonnera sur une période de 12 mois. Le développement du questionnaire est présentement en cours et seulement des interviews personnelles (approximativement d’une heure) seront menées.

Différents plans d’échantillonnage sont présentement à l’étude pour le cycle 1.2. Bien entendu, sélectionner les personnes faisant partie de l’échantillon du cycle 1.1 selon un plan à deux phases est une option certes très attrayante mais qui comporte deux inconvenients de taille : le fardeau de réponse et le dépistage des personnes sélectionnées. D’une part, le fardeau supplémentaire (on vise quand même une interview d’une durée moyenne d’une heure sur la santé mentale) imposé à un certain nombre de répondants de l’enquête du cycle 1.1 s’en trouverait démesuré. D’autre part, mettre en place des procédures de dépistage dans le but de retracer les personnes sélectionnées est un exercice fastidieux et coûteux. Qui plus est, avec le nombre grandissant de préoccupations soulevées par la population quant au respect de la vie privée de la part des organismes gouvernementaux, la possibilité de contacter à nouveau les mêmes répondants n’a pas été retenu.

Deux options sont toujours possibles. La première est d’utiliser la base aréolaire mise en place pour l’EPA et d’y sélectionner un échantillon de logements nouveaux à l’intérieur desquels une personne âgée de 15 ans ou plus serait choisie

¹ Les trois territoires canadiens ne font pas partie de la population visée par cette enquête par souci du fardeau de réponse.
aléatoirement (approche très semblable au plan de l’enquête du cycle 1.1). C’est une option simple, sans surprise, qui procurerait un tout nouvel échantillon; une particularité quand même souhaitable dans une enquête lorsque l’on désire obtenir des mesures de prévalence sur un vaste répertoire de caractéristiques.

La deuxième option est d’utiliser les logements (et non les personnes) sélectionnés dans l’enquête du cycle 1.1 comme base de sondage à partir de laquelle un sous-échantillon serait choisi selon un plan à deux phases, et où une personne âgée de 15 ans ou plus serait sélectionnée aléatoirement à l’intérieur des ménages vivant dans ces logements. (À noter que la possibilité de sélectionner le même répondant serait toujours possible sous cette option mais l’impact global sur le fardeau de réponse serait tout de même moindre.) Le défi que représente la deuxième option consiste à identifier des caractéristiques à l’échelle des ménages et/ou logements telles que recueillies dans l’enquête du cycle 1.1 qui permettraient d’accroître l’efficacité de l’échantillon de l’enquête du cycle 1.2; une étude sur ce point est présentement en cours.

En fonction des objectifs de l’enquête, plusieurs facteurs sont considérés pour analyser ces deux options. L’efficacité du plan (en termes de précision et biais), le fardeau des répondants de même que les contraintes opérationnelles (incluant les coûts) sont tous des facteurs pour lesquels une évaluation soignée est nécessaire afin d’arrêter un choix éclairé.

5. Développements méthodologiques futurs

Les activités de collecte de la première composante (cycle 1.1) de l’ESCC ont débuté en septembre 2000 et s’étendront sur une période de 12 mois. Durant l’année 2000-2001, les méthodologistes se concentreront sur les nombreux défis qu’entraîneront la gestion mensuelle d’un tel échantillon, l’estimation, la variance, l’analyse de la qualité et la diffusion des données de cette première enquête. (Comme il a été mentionné précédemment, l’ESCC diffusera des données sur une base trimestrielle, semestrielle et annuelle.)

La deuxième composante de l’ESCC, soit le cycle 1.2 de l’enquête sera mise en œuvre dès janvier 2002. La méthodologie s’affairera dans les mois à venir à finaliser le plan d’échantillonnage et à participer activement au processus d’élaboration du questionnaire. Cette fois-ci, ce dernier sera commun à toutes les régions du pays.

Pour la première partie du cycle 2, soit en 2003, il s’agira de répéter l’exercice du cycle 1.1 en réalisant de nouveau une enquête régionale. Pour l’année 2004, la thématique particulière retenue est celle de la nutrition. Le but de cette enquête sera de fournir de l’information précise et actuelle sur la consommation d’aliments et la nutrition afin d’informer et guider les activités des programmes fédéral et provinciaux. Le choix du meilleur véhicule de collecte n’est pas une tâche facile et
encore moins la détermination du contenu du questionnaire. Des consultations sont présentesment en cours avec des experts dans le domaine pour trouver le meilleur moyen de recueillir l’information et pour établir un consensus en ce qui a trait au contenu. Un autre volet important de cette réalisation est le jumelage de la composante nutrition (soit un sous-échantillon de celle-ci) avec une composante sur les mesures physiques. Il s’agit en fait de recueillir quelques mesures physiques appropriées pour évaluer plus en profondeur l’état nutritionnel. Il peut s’agir par exemple de mesures anthropométriques (taille, poids, tour de taille), de mesures biologiques (échantillon de sang, cheveu, etc.) ou encore de tests physiques (vision, audition, etc.).

Toutefois, avant de pouvoir réaliser une telle enquête nutrition-mesures physiques, il n’en demeure pas moins que plusieurs enjeux en matière d’éthique doivent être résolus : i) quels mécanismes doivent être utilisés pour obtenir le consentement du répondant, ii) doit-on et dans l’affirmative comment faire part des découvertes aux répondants et iii) est-il possible de remiser des échantillons biologiques pour analyses futures. Des consultations sont en cours avec Santé Canada, des agences concernées, des groupes d’experts et le comité consultatif en santé de Statistique Canada. Les enjeux en discussion présentement comprennent l’élaboration du contenu de l’enquête, le mode de collecte et le rationnel s’y rattachant, les protocoles de l’enquête, l’éthique et finalement les coûts associés à cette nouvelle façon de recueillir de l’information.

REMERCIEMENTS

Les auteurs tiennent à souligner l’excellent travail des méthodologistes suivants durant le processus de développement de cette enquête : Suzanne Lessard, Marianna Morano, Orane Saint-Denis et Sylvain Thivierge. Ils tiennent également à remercier France Biloq et Jean-Louis Tambah pour leurs précieux commentaires lors de la rédaction de cet article.
BIBLIOGRAPHIE


Annexe A – Liste des modules de l’enquête du cycle 1.1

<table>
<thead>
<tr>
<th>Modules:</th>
<th>Commun</th>
<th>Optionnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables du dossier du ménage</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Etat de santé général</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Changements pour améliorer la santé</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Taille et poids</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Incapacité au cours des deux dernières semaines</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Utilisation des soins de santé</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Services à domicile</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Limitation des activités</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Problèmes de santé chroniques</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Consommation de médicaments</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Tension artérielle</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Test PAP</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Mammographie</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Examen des seins</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Auto-examen des seins</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Visite chez le dentiste</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Examen de la vue</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Examen général</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Vaccins contre la grippe</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Test de l’antigène spécifique prostatique</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consommation de fruits et légumes</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Activités physiques</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Activités sédentaires</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Utilisation des précautions</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Blessures</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Indice de l’état de santé (HUI)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Stress au travail</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Estime de soi</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Contrôle</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Usage du tabac</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Outils pour cesser de fumer</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Exposition à la fumée des autres</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Variantes du tabagisme</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Consommation d’alcool</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>La conduite automobile et la consommation d’alcool</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Dépendance à l’égard de l’alcool</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Allaitement</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Comportement sexuel</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Soutien social (questions relatives à l’étude sur les préoccupations médicales)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Spiritualité</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Modules</td>
<td>Commun</td>
<td>Optionnel</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>Utilisation des services-santé mentale</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Humeur (échelle d’équilibre affectif de Bradburn)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Détresse</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Dépression</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Suicide</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Renseignements Sociodémographiques</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Population active</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Revenu</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Précarité alimentaire</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Administration</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
 UTILISATION DES DONNÉES DE LA TAXE SUR LES PRODUITS ET SERVICES DANS LE REMANIEMENT DE L’ENQUÊTE MENSUELLE DU COMMERCE DE GROS ET DE DETAIL DE STATISTIQUE CANADA

M. BRODEUR et H. BÉRARD
Statistique Canada - Division des méthodes d’enquêtes auprès des entreprises

RÉSUMÉ

L’Enquête mensuelle du commerce de gros et de détail (EMCGD) est une enquête très importante menée par Statistique Canada. Elle vise à mesurer les ventes mensuelles pour divers domaines d’intérêt tels que les provinces et les groupes de commerce (regroupements de codes de classification type des industries). Les estimations mensuelles sont utilisées dans le calcul du Produit intérieur brut (PIB). L’estimation de tendance représente un indicateur économique majeur et la somme des estimations mensuelles est utilisée pour le calcul des paiements de péréquation (des transferts de fonds du gouvernement fédéral vers les provinces les plus pauvres).

L’EMCGD recourt à un plan de sondage stratifié aléatoire simple sans remise. Le dernier remaniement complet de l’enquête remonte à 1988. Plusieurs améliorations ont été apportées à l’enquête au fil des ans. Toutefois, la venue d’une nouvelle classification industrielle, le besoin de réduire les coûts d’enquêtes et l’utilisation de systèmes informatiques vétustes justifient le besoin de faire un remaniement. Au même moment, Statistique Canada a initié plusieurs projets pour utiliser davantage les données administratives provenant de la collecte de la Taxe sur les produits et services (TPS). Deux projets préparatoires au remaniement ont été amorcés. Le premier projet vise à étudier la possibilité d’utiliser les données des fichiers de TPS comme variable de stratification. Le deuxième propose de développer un estimateur par calage en employant les données de TPS comme variable auxiliaire. Un tel estimateur pourrait éventuellement remplacer l’actuel estimateur par dilatation. Cette communication présente les principales étapes du remaniement et traite des résultats préliminaires des deux projets d’utilisation de la TPS.
1. Introduction

L'Enquête mensuelle du commerce de gros et de détail (EMCGD) est une enquête très importante de Statistique Canada (SC) et son plan d’échantillonnage, conçu en 1988, a servi de modèle à plusieurs enquêtes. L’enquête fait présentement l’objet d’un remaniement. Le plan d’échantillonnage demeurera sensiblement le même. Cependant, les unités d’échantillonnage seront modifiées et le processus de renouvellement simplifié. Toutefois, un des problèmes majeurs de l’enquête est relié aux problèmes de classification des unités et le fait qu’au fil du temps, les unités se retrouvent dans de mauvaises strates. Les problèmes proviennent surtout de la variable utilisée lors de la stratification. Cet article traitera plus en détails de deux projets préparatoires au remaniement. Le premier projet vise à étudier la possibilité d’utiliser les données des fichiers de la Taxe sur les produits et services (TPS) comme variable de stratification. Il s’agit d’une taxe sur la valeur ajoutée. Le deuxième projet propose de développer un estimateur par calage en employant les données de TPS comme variable auxiliaire. Un tel estimateur pourrait éventuellement remplacer l’actuel estimateur par dilatation.

Pour bien comprendre l’impact de ces deux projets, il est important de faire une mise en contexte de l’EMCGD. La section 2 fournira une description du plan d’échantillonnage et de l’estimateur. La section 3 dressera un bref historique de l’enquête depuis le dernier remaniement de 1988, tandis que la section 4, présentera brièvement les objectifs du remaniement. La section 5 traitera du projet d’utilisation de la TPS dont, la sous-section 5.3 qui parlera de l’utilisation de la TPS comme variable de stratification. Les sous-sections suivantes parleront d’estimation par calage et de résultats.
2. Description du plan d’échantillonnage

Le dernier remaniement majeur de l’enquête remonte à 1988. L’EMCGD est une enquête mensuelle conçue pour mesurer principalement les ventes par groupe de commerce (regroupement de codes de classification type industrielle (CTI 1980)) à trois ou quatre chiffres par province et pour certaines régions métropolitaines de recensement (RMR). L’échantillon est sélectionné à partir du Registre des entreprises (RE) de Statistique Canada. Le RE contient toutes les entreprises connues opérant au Canada. Sur le RE, la structure de chaque entreprise est hiérarchisée en fonction des besoins statistiques et comporte quatre niveaux qui sont dans l’ordre l’entreprise, la compagnie, l’établissement et l’emplacement. Une entreprise se compose d’une ou plusieurs compagnies. Une compagnie se compose d’un ou plusieurs établissements et ainsi de suite. Pour plusieurs entreprises simples toutefois, l’entreprise, la compagnie, l’établissement et l’emplacement coïncident. Les entreprises dites complexes peuvent œuvrer dans plusieurs provinces et dans différents secteurs industriels et ont très souvent un revenu très élevé. La population cible de l’enquête varie pour le commerce de gros et de détail. Pour le commerce de détail, elle est définie comme étant toute compagnie comportant au moins un emplacement œuvrant dans le commerce de détail alors que pour le commerce de gros, elle est définie comme étant toute compagnie comportant au moins un établissement œuvrant dans le commerce de gros. L’unité d’échantillonnage est la compagnie et seulement les compagnies ayant des employés font partie de l’enquête.

La taille d’échantillon pour le commerce de détail est d’environ 16 000 compagnies statistiques vivantes provenant d’une population de 137 000. La taille d’échantillon pour le commerce de gros est de 7 000 compagnies statistiques vivantes provenant d’une population de 58 000. La population est stratifiée par province, territoire, certaines RMR et par groupe de commerce. Chaque combinaison de groupe de commerce et de région géographique forme une strate. Chaque strate est divisée en trois sous-strates selon la taille : une à tirage complet et deux à tirage partiel contenant respectivement les moyennes et les petites compagnies. Les strates à tirage complet englobent toutes les compagnies qui ont une structure complexe, i.e., opérant dans plus d’un groupe de commerce ou région géographique, ou qui ont un revenu brut d’entreprise (RBE) supérieur à un seuil donné. Les autres compagnies sont classées dans les strates à tirage partiel selon leur RBE et les seuils des strates. En 1988, les seuils entre la strate à tirage complet et les strates à tirage partiel ont été calculés à l’aide de la méthode d’Hidiroglou (1986). Les seuils entre les deux strates à tirage partiel correspondent au niveau du RBE délimitant les unités simples et complexes sur le RE. Le coefficient de variation des ventes visé se situe à 1,5 % au niveau canadien alors qu’il est de 2,5 % au niveau provincial et de 3,5 % au niveau des groupes de commerce. La méthode de répartition de l’échantillon utilisée est celle de la racine carrée du RBE.

INSEE Méthodes 309
L'enquête emploie un plan de sondage stratifié aléatoire simple sans remise. Les compagnies d'une même strate $h$ à tirage partiel sont réparties aléatoirement en un certain nombre $P_h$ de grappes, appelées panels, de taille égale à une unité près. Chaque mois, toutes les compagnies d'une sélection contiguë de $p_h$ panels forment l'échantillon. Les valeurs $P_h$ et $p_h$ sont des fonctions de la fraction de sondage désirée, mais aussi du nombre de mois maximal qu'une compagnie peut demeurer dans l'échantillon et du nombre minimal qu'elle doit demeurer exclue de l'échantillon une fois sortie. En effet, une rotation est appliquée chaque mois à l'échantillon en substituant un panel de l'échantillon par un nouveau panel qui était hors échantillon. Plus de détails sont disponibles dans Hidiroglou et coll. (1991). De plus, des compagnies nouvelles au commerce de détail (naissances) sont identifiées chaque mois par le RE et ajoutées systématiquement aux panels des strates auxquelles elles appartiennent. Les naissances tombant dans les panels échantillonnés font par conséquent partie de l'échantillon. Il est à noter que chaque mois les naissances sont ajoutées systématiquement aux panels de la population de sorte que l'on obtienne, dans les faits, le nombre espéré de naissances dans l'échantillon.

Les compagnies ayant cessé leurs opérations dans le commerce de détail ou de gros (cessations) apparaissent mensuellement elles aussi. Des cessations peuvent se rapporter à une compagnie statistique dont les activités ont cessé, donc retirée des affaires, ou dont les principales activités ne sont plus le commerce de détail ou de gros, donc hors du champ d'observation ou inactives. Elles sont identifiées à l'aide de sources administratives permettant la mise à jour du RE et à l'aide de mises à jour relatives aux résultats des opérations de collecte des enquêtes, y compris l'EMCGD. Les cessations sont d'abord codées comme telles dans le RE, puis dans la base de sondage de l'EMCGD. Les résultats des enquêtes sont un moyen beaucoup plus rapide d'identification des cessations puisque les sources administratives peuvent comporter un délai d'un an. Ainsi, pour avril 1999, la proportion des cessations identifiées dans la population du commerce de détail est de 10,3 % et la proportion de celles dans l'échantillon est de 18,7 %. Pour le commerce de gros, nous avons 26,8 % de cessations identifiées dans la population et 34,0 % dans l'échantillon. Ces faits ne sont pas des évidences que notre échantillon est biaisé pour autant. Nous croyons qu'il existe des unités ayant cessé leurs activités et non identifiées dans le volet non échantillonné qui sont jugées actives dans le RE. Ces unités ont une mesure de taille positive. Puisque l'EMCGD met à jour indirectement sa propre base de sondage (source qui est donc non indépendante) et puisque nous n'avons aucune façon de distinguer facilement dans le RE les cessations identifiées à l'aide de sources administratives ou de sources d'enquête, les cessations ne sont pas automatiquement éliminées de l'échantillon et de la base de sondage de l'EMCGD. Lavallée (1996) a montré l'ampleur du biais que susciterait l'utilisation, pour la mise à jour de la base de sondage, d'une source d'information non indépendante du plan de sondage de l'enquête. C'est pourquoi, une ou deux fois par année, des cessations sont éliminées de la base de sondage et de l'échantillon d'une façon non biaisée (Trépanier et coll. 1998).
Malgré l’utilisation des panels, on peut considérer à toutes fins pratiques le plan de sondage comme un plan de sondage stratifié aléatoire simple d’éléments sans remise, dont la fraction de sondage est approximativement \( n_h / N_h \), où \( N_h \) est le nombre de compagnies dans la strate au niveau de toute la population et \( n_h \) est le nombre de compagnies dans la strate au niveau de l’échantillon. En effet, l’estimateur utilisé est un estimateur post-stratifié classique où les poids de sondage sont \( P_h / p_h \); les post-strates sont lestrates telles que définies en 1988 et les totaux de contrôle sont les tailles de population \( N_h \) des strates \( h \). L’estimateur des ventes, \( \hat{Y} \), pour un domaine \( d \) peut donc s’écritre :

\[
\hat{Y}(d) = \sum_{i=1}^{n_h} \left( \frac{N_h}{p_h} \frac{P_h}{p_h} \right) y_{ih}(d) = \sum_{h} \left( \frac{N_h}{p_h} \frac{P_h}{p_h} \right) y_{h}(d) = \sum_{h} \frac{N_h}{p_h} \sum_{i=1}^{n_h} y_{ih}(d)
\]

où \( y_{ih}(d) = y_{ih} \), c.-à-d., les ventes de la compagnie \( i \), si celle-ci est dans le domaine \( d \), et \( y_{ih}(d) = 0 \) sinon. On retrouve donc la forme de l’estimateur par dilatation où le poids de sondage est \( N_h / n_h \). C’est d’ailleurs à cet estimateur que nous ferons référence dans la suite de cet article. De façon plus générale, cet estimateur peut s’écritre :

\[
\hat{Y}(d) = \sum_{i \in S} a_i y_i(d)
\]

où \( a_i \) représente le poids de sondage de \( i \). Lorsqu’une compagnie fait partie de l’échantillon, un questionnaire est généré pour l’ensemble des composantes de la compagnie qui œuvre dans le même groupe de commerce. Sur ce questionnaire, on recueille les ventes mensuelles au détail effectuées dans chaque région géographique d’intérêt et dans certains cas les inventaires. Une compagnie peut recevoir plus d’un questionnaire si, en autres, elle œuvre dans plusieurs groupes de commerce. Cette procédure nous permet de produire des estimations fiables par groupes de commerce et régions géographiques. On peut donc considérer à toutes fins pratiques que le plan de sondage est un plan stratifié aléatoire simple sans remise par grappes. La grappe est cette fois-ci représentée par la compagnie statistique et les éléments de la grappe sont les différents questionnaires générés pour cette compagnie. Cette analogie sera utile plus tard lorsqu’on parlera d’estimation par l’entremise du Système généralisé d’estimation de Statistique Canada.
3. Développement de l'EMCGD depuis 1968

Après le remaniement de 1988, un certain nombre de projets furent mis en place pour augmenter la taille de l'échantillon afin d'améliorer les estimations de certains groupes de commerce problématiques. Ces projets visaient à maintenir un niveau de qualité acceptable dans l'EMCGD. Toutefois, il demeure que sans mise à jour complète de la stratification pendant toutes ces années et sans méthode d'estimation utilisant de l'information auxiliaire pour compenser, la qualité de certaines estimations de l'enquête s'est graduellement détériorée. En effet, la valeur de la strate et du domaine différant de plus en plus (p. ex.: le groupe de commerce de stratification et de domaine) et des poids de sondage élevés étant associés à certaines compagnies ayant pris de l'importance qui avaient été classées dans les strates à tirage partiel à l'époque rendaient les estimations instables dans le temps et faisaient grimper certains coefficients de variation.


La restructuration a été très bénéfique et a permis de corriger de nombreux problèmes reliés à une mauvaise classification. Cependant, l'estimateur par calage n'a pas donné les résultats escomptés surtout à cause de la faible corrélation de la variable auxiliaire avec les ventes. Les résultats étaient intéressants au niveau agrégé mais, étaient quelquefois aberrants au niveau des groupes de commerce. Nous avons donc conservé l'estimateur par dilatation pour toutes les strates.
4. Objectifs du remaniement

Tel que mentionné auparavant, l’EMCGD entre dans une période de remaniement. La présente méthodologie de l’enquête est basée en outre sur la classification type des industries de 1980 et il est impératif que l’EMCGD soit remaniée en vertu du nouveau Système de classification des industries de l’Amérique du Nord (SCIAN). L’environnement informatique de l’EMCGD repose principalement sur l’ordinateur principal et une migration vers un environnement de micro-informatique semble inévitable pour réduire les coûts d’exploitation, favoriser l’intégration des activités d’enquêtes et accroître la flexibilité. Le remaniement ouvre la porte vers l’utilisation des produits généralisés de SC.

L’EMCGD ne couvre présentement que les compagnies ayant des employés. Un ajustement au niveau de l’estimation est fait pour contrer ce problème mais, il a tendance à faire fluctuer les estimations. Le RE possède maintenant une liste exhaustive des entreprises au Canada. Ainsi, l’EMCGD doit modifier sa couverture pour ajouter les entreprises sans employés et les détaillants hors magasins qui sont présentement sondés séparément. D’autres raisons méthodologiques justifient le remaniement de l’EMCGD. L’avènement du Projet d’amélioration des statistiques économiques provinciales (PASEP) et, en conséquence, l’intégration des enquêtes annuelles sur le commerce de gros et détail à l’Enquête unifiée des entreprises (EUE) forcent l’EMCGD à harmoniser certaines de ses caractéristiques (p. ex. unité visée, stratification) avec celles de l’EUE pour assurer la cohérence entre les deux sources de données statistiques. Dans la même veine, le remaniement de l’EMCGD permet à celle-ci de mettre en place les seuls d’exclusion Royce-Maranda permettant de réduire le fardeau de réponse des petits établissements en les exemptant de compléter un questionnaire.

Plusieurs années de renouvellement mensuel du l’échantillon du commerce de détail ont démontré que la méthodologie actuelle de renouvellement dans l’EMCGD rend difficile la gestion des unités incorrectement stratifiées selon la taille. De nouvelles approches de renouvellement de l’échantillon doivent être examinées. La méthodologie de l’EMCGD remaniée doit aussi porter une attention particulière aux conséquences de certaines méthodes sur l’ampleur des révisions aux estimations.

La disponibilité des données de la Taxe sur les produits et services (TPS) offre une occasion unique d’utiliser une source de données administratives dans le remaniement de l’EMCGD pour accroître la qualité des estimations, réduire les coûts ou diminuer le fardeau de réponse. Les données administratives peuvent être utilisées pour des totalisations directes; des estimations indirectes; comme variable de stratification; pour valider et corriger les réponses lors de la vérification et imputation et pour le remplacement de données d’enquête.
Deux projets préparatoires au remaniement ont été amorcés en 1999. Le premier projet vise à étudier la possibilité d'utiliser les données des fichiers de TPS comme variable de stratification. Le deuxième propose de développer un estimateur par calage en employant les données de TPS comme variable auxiliaire. Un tel estimateur pourrait éventuellement remplacer l'actuel estimateur par dilatation. Ces deux projets visaient essentiellement à se familiariser avec les concepts de la TPS. Il a été jugé trop prématuré de considérer la TPS comme remplacement des données d'enquête. Les sections suivantes traitent de l'utilisation des données de la TPS dans le cadre du remaniement de l'EMCGD.

5. Projet d'utilisation de la TPS

5.1 Description de la TPS

En 1991, le Gouvernement fédéral a décidé d'introduire une nouvelle taxe sur la valeur ajoutée que l'on nomme Taxes sur les produits et services (TPS). Chaque entreprise doit remettre la valeur de la TPS à l'Agence de douanes et du revenu du Canada (ADRC). Un fichier administratif contenant un ensemble de variables reliés aux ventes de TPS est maintenant accessible à Statistique Canada. Pour notre étude, nous nous intéressons à la variable Fourniture taxable. Pour cette étude, le montant de Fourniture taxable sera considéré comme la variable de ventes de TPS.

Chaque entreprise ayant 30 000$ ou plus en ventes annuelles doit s'enregistrer auprès de l'ADRC. La fréquence de remise dépend de la taille de l'entreprise. Les entreprises ayant plus de 6 millions$ en ventes annuelles doivent remettre mensuellement. Celles ayant entre 500 000$ et 6 millions$ doivent remettre à chaque trimestre alors que les autres, celles ayant moins de 500 000$ remettent annuellement. Environ 78% des détaillants et des grossistes remettent sur une base trimestrielle.

Présentement, Statistique Canada reçoit un fichier de remise de l'ADRC. Ce fichier regroupe les codes d'activité, les ventes totales de TPS et la valeur de la TPS perçue. Le fichier est reçu sur une base irrégulière. Il s'écoule 90 jours du moment où une entreprise remet sa taxe au moment où les données sont acheminées à Statistique Canada qui fait alors une vérification et imputation sommaire des données manquantes et en erreur.

La première étape de ce projet a été de créer un fichier de ventes TPS pour chaque compagnie statistique de notre population. Cette opération s'est avérée très complexe et a demandé beaucoup de travail. Au cours de cette étape, qui fut nommée « réconciliation avec le registre des entreprises », nous avons fait plusieurs découvertes importantes quant à la qualité et au contenu du fichier de données de
ventes de TPS. Citons ici quelques exemples : le fichier initial de données de ventes de TPS ne contient pas de date de début de période; les périodes de référence varient d'une compagnie à l'autre; la fréquence de remise peut varier d'une compagnie à l'autre et même d'une fois à l'autre pour une même compagnie; enfin le fichier de ventes de TPS ne tient pas compte des transactions intra-entreprises. De plus, les délais d'acquisition du fichier sont longs et le traitement est limité. Les unités de l'univers du registre des entreprises ne trouvent pas toutes un enregistrement correspondant sur le fichier de ventes de TPS et finalement, il y a des différences entre les concepts de ventes de TPS et les ventes mensuelles de l'EMCGD.

Pour cette étude, nous avons fait la somme de 12 mois de données (ventes annalisées) de avril 1998 à mars 1999. Dans environ 8% des cas, la valeur annualisée des ventes de TPS peut en réalité représenter une période de moins de 12 mois ou de plus de 12 mois. En effet, un répondant peut subitement remettre de façon annuelle alors qu'il remettait de façon trimestrielle. Comme la période visée par la remise n'est pas indiquée sur le fichier, la somme directe des 12 dernières remises peut mener à des sous-estimations ou des surestimations des ventes annuelles de TPS.

Nous avions la somme des ventes de TPS au niveau du sommet de la structure opérationnelle, donc de l'entreprise statistique qu'il fallait répartir aux niveaux statistiques inférieurs. Cette étape a demandé beaucoup de temps et d'efforts pour l'établissement des liens entre le numéro de compte de TPS et le numéro d'entreprise (NE : indicateur unique du RE). Cette tâche n'était pas triviale dans le cas d'entreprises complexes. Cette étape était nécessaire pour faire la répartition des ventes de TPS au niveau de la structure statistique. Le RBE a été utilisé pour faire la répartition des ventes de TPS au niveau de l'établissement statistique pour ensuite se rapporter au niveau de la compagnie statistique. Le diagramme suivant illustre bien le processus de répartition pour une entreprise complexe.
5.2 Variable de stratification

Dans l'EMCGD actuelle, les compagnies sont stratifiées selon le RBE. Celui-ci est estimé à tous les mois à partir de différentes sources de données fiscales. Par exemple, pour les employeurs, le RBE est estimé à partir d'une valeur annualisée du total des remises de paye. Étant donné les divergences entre le RBE et les ventes réelles, environ 17% des détaillants et 23% des grossistes sont classés dans la mauvaise strate de taille. Dans le cas des naissances, l'écart est beaucoup plus grand, seulement 36% des nouveaux détaillants et 30% des nouveaux grossistes se retrouvent dans la strate appropriée. Les changements de taille sont surtout des changements entre les deux strates à tirage partiel et très peu d'unités passent de strate à tirage partiel à strate à tirage complet. Cependant, quelques unités mal classées peuvent avoir un effet dévastateur sur les estimations lorsque leurs ventes sont multipliées par un poids de sondage ne reflétant pas leur taille.

La différence entre le RBE et les ventes réelles s'explique par une différence de concept entre la mesure de taille utilisée, le revenu brut et la valeur collectée, les ventes. Le RBE a tendance à sous-estimer le niveau des ventes. Dans le cas des naissances, le RBE varie beaucoup les premiers mois et ce changement est
généralement à la hausse. Après quelques mois, le RBE se stabilise. Cette fluctuation est reliée en partie à l'utilisation dans le modèle d'une valeur annualisée basée parfois sur très peu de mois de données. De plus, dans bien des cas, la première remise de paye servant à calculer le premier RBE ne reflète pas la taille réelle de l'entreprise.

5.2.1 Étude de la variable de stratification

Cette étude vise à évaluer les différentes sources d'informations auxiliaires afin de mieux stratifier les unités. Les paragraphes qui suivent comparent l'efficacité lors de la stratification de l'utilisation des ventes TPS par rapport à l'utilisation du RBE.

Dans un premier temps, nous avons analysé la relation entre le niveau des ventes réelles (somme de 12 mois de avril 1998 à mars 1999) et les deux variables à l'étude. Le tableau 1 présente ces résultats. Le niveau de la TPS est très proche du niveau des ventes alors que le RBE sous-estime le niveau des ventes. Les corrélations observées sont plus élevées pour le commerce de détail que pour le commerce de gros. En considérant toutes les compagnies, les corrélations sont autour de 97% pour les deux variables dans le cas du commerce de détail. Pour le commerce de gros, les corrélations sont de 93% avec la TPS et de 91% avec le RBE.

<p>| Tableau 1 |
| Corrélations* entre les ventes de TPS et les ventes annuelles reportées de l'EMCGD |
| Commerce de détail | Commerce de gros |</p>
<table>
<thead>
<tr>
<th>TPS</th>
<th>RBE</th>
<th>TPS</th>
<th>RBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toutes les compagnies**</td>
<td>0.97</td>
<td>0.97</td>
<td>0.93</td>
</tr>
<tr>
<td>&gt;10M$**</td>
<td>0.97</td>
<td>0.97</td>
<td>0.92</td>
</tr>
<tr>
<td>1-10M$**</td>
<td>0.89</td>
<td>0.73</td>
<td>0.64</td>
</tr>
<tr>
<td>&lt;1M$ **</td>
<td>0.70</td>
<td>0.58</td>
<td>0.21</td>
</tr>
</tbody>
</table>

* corrélations avec données pondérées
** exclus les valeurs aberrantes (<1% des détaillants; 1% des grossistes)

Le niveau de corrélation observé varie en fonction des groupes de commerce considérés. À quelques exceptions près, les corrélations entre les ventes des détaillants et les ventes de TPS pour chaque groupe de commerce demeurent autour
de 90% pour les compagnies avec des revenus annuels supérieurs à 1 million. Toutefois, pour les compagnies dont le revenu annuel est inférieur à 1 million les corrélations observées sont en général plus faibles et varient entre 27% (distributeurs d'alcool, de vin et de bière) et 91% (distributeurs de vêtements).

Chez les grossistes, la corrélation entre les ventes et la TPS varie entre 84% et 97% pour les compagnies aux revenus annuels supérieur à 10 millions. Les corrélations observées pour les compagnies dont le revenu annuel est entre 1 et 10 millions sont plus faibles et varient entre 48% et 82%. Dans le cas des compagnies dont le revenu annuel est inférieur à 1 million, les corrélations sont inférieures à 50% pour la très grande majorité des groupes de commerce.

Dans un deuxième temps, nous avons comparé le taux de mauvaise classification qui pourrait survenir en utilisant comme variable de stratification soit le RBE ou la TPS. Les résultats ont été validés en utilisant les ventes obtenues des répondants de l'EMCGD. En terme de nombre d'unités bien classées, l'utilisation de la TPS mène à un gain, par rapport au niveau actuel basé sur le RBE, de 6% pour le commerce de détail et de 9% pour le commerce de gros (Tableau 2).

**Tableau 2**

Pourcentage d'unités bien classées

<table>
<thead>
<tr>
<th>Commerce de Détail</th>
<th>RBE</th>
<th>TPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commerce de Gros</td>
<td>77%</td>
<td>86%</td>
</tr>
</tbody>
</table>

1 % d'unité assignées à la bonne strate selon les ventes réelles.

Le tableau 3 présente les résultats pour les unités les plus importantes de l'enquête. Pour le commerce de détail, le nombre d'unités mal classées est de 5.3% avec le RBE et baisse à 2.6% avec les ventes de TPS. On observe la même tendance chez les grossistes.
Tableau 3

Taux de mauvaise classification des unités importantes

<table>
<thead>
<tr>
<th></th>
<th>RBE</th>
<th>TPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commerce de Détail</td>
<td>5.3%</td>
<td>2.6%</td>
</tr>
<tr>
<td>Commerce de Gros</td>
<td>7.8%</td>
<td>4.4%</td>
</tr>
</tbody>
</table>

1 % unités assignées à une strate à tirage partiel alors quelles auraient dû être assignées à une strate à tirage complet selon leurs ventes réelles.

Toutefois, le RBE demeure plus efficace que les ventes de TPS pour classifier les nouveaux détaillants et les nouveaux grossistes (Tableau 4). La moins bonne performance de la TPS pour identifier le niveau des ventes des naissances est probablement en partie due à la méthode actuelle pour calculer les ventes annualisées (somme des remises disponibles pour les 12 derniers mois). Depuis août 2000, la période couverte pour les remises sera disponible sur le fichier TPS. Nous serons donc en mesure de produire une valeur annualisée qui sera sans doute supérieure.

Tableau 4

Taux de naissances bien classifiées selon la taille

<table>
<thead>
<tr>
<th></th>
<th>RBE</th>
<th>TPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commerce de Détail</td>
<td>36.6%</td>
<td>30.1%</td>
</tr>
<tr>
<td>Commerce de Gros</td>
<td>30.1%</td>
<td>29.2%</td>
</tr>
</tbody>
</table>

Nos analyses préliminaires nous permettent de conclure que l'utilisation des ventes TPS comme variable de stratification est très prometteuse. Le RBE a tendance à sous-estimer les ventes totales alors que le niveau des ventes TPS est en fait très proche des ventes réelles pour les compagnies avec des revenus supérieurs à 1 million. Ces compagnies représentent plus de 87% des ventes totales des détaillants et plus de 96% des ventes totales des grossistes. Dans le cas des plus petites compagnies et des naissances nous devons poursuivre nos recherches car la qualité de la relation reste faible pour les deux mesures. Nous explorons différents modèles de régression basés sur l'utilisation du RBE, des ventes TPS ou une combinaison des deux. Un des modèles des plus prometteurs est tout simplement l'utilisation d'une nouvelle mesure de taille définie comme étant le maximum de soit le RBE ou la TPS.
5.3 Estimation par calage


L’estimateur GREG suppose un modèle linéaire entre la variable d’intérêt $y$ et la variable auxiliaire $x$ :

$$y_k = \beta_0 + \beta_1 x_k + \varepsilon_k = \mathbf{x}'_k \mathbf{\beta} + \varepsilon_k$$

où $E(\varepsilon_k) = 0$, $\text{Var}(\varepsilon_k) = \sigma^2_k$, $\text{Cov}(\varepsilon_k, \varepsilon') = 0 \quad \forall k \neq \ell$. (5.1)

Dans notre étude, $y_k$ représente les ventes mensuelles de la compagnie $k$ dans l’EMCGD. La variable auxiliaire $x_k$ représente les ventes annuelles de biens assujettis à la TPS de la compagnie. Nous supposons que chaque résidu $\varepsilon_k$ comporte la même structure de variance ($\sigma^2_k = 1$ pour chaque unité $k$). L’estimateur par le quotient, c’est-à-dire sans ordonnée à l’origine ($\beta_0$) et avec $c_k = x_k$, a été examiné, mais il a été exclu du reste de l’étude ultérieurement puisque $\beta_0$ différerait appréciablement de 0 la plupart du temps.

5.3.1 Groupes modèles

Les groupes modèles forment une partition disjointe et exhaustive de la population à l’intérieur de laquelle il existe une bonne relation entre la variable auxiliaire et la variable d’intérêt. Les paramètres du modèle linéaire qui sous-tend le GREG sont estimés dans chaque groupe modèle. C’est pourquoi il est important qu’il y ait un nombre minimal d’unités échantillonnées dans chaque groupe modèle si l’on veut que l’estimation des paramètres de modèle soit suffisamment précise.

5.3.2 Estimateur GREG

L’estimateur GREG sert à fournir des poids « nouveaux » ($w_k$) les plus proches possible des poids du plan original ($a_k$) en fonction d’une mesure quelconque de la distance. Les $w_k$ se laissent exprimer sous la forme $w_k = a_k g_k$. Les $g_k$ sont calculés de façon que

$$\hat{X}_p = \sum_{k \in \mathcal{S}_p} a_k g_k x_k = \sum_{k \in \mathcal{S}_p} x_k = \hat{X}_p$$

(5.2)

320

INSEE Méthodes
où \( a_k \) est le poids du plan pour l'unité \( k \), \( g_k \) est le facteur \( g \) pour l'unité \( k \), \( x_k \) est le vecteur variable auxiliaire pour l'unité \( k \), \( \hat{X}_p \) est le total estimatif du vecteur variable auxiliaire dans le groupe modèle \( p \) et \( X_p \) est le total de contrôle pour le groupe modèle \( p \).

Pour calculer le facteur \( g \), il faut connaître les valeurs de la variable auxiliaire pour les unités de l'échantillon et le total de la population de contrôle pour chacun des groupes modèles. Dans notre étude, les totaux de contrôle pour les ventes annuelles de biens assujettis à la TPS sont obtenus en prenant la somme, dans chaque groupe modèle, des valeurs \( x_k \) figurant dans les fichiers administratifs de la TPS provenant de l'ADRC. Plus précisément, le facteur \( g \) (à supposer que \( c_\bar{F}=1 \)) se laisse exprimer sous la forme :

\[
g_k = l + \left( \sum_{k \in \cup \bar{P}} x_k' - \sum_{k \in \cup \bar{P}} a_k x_k' \right) \left( \sum_{k \in \cup \bar{P}} a_k x_k x_k' \right)^{-1} (x_k)
\]

(5.3)


### 5.4 Défis posés par la stratégie d’estimation de l’EMCGD

La présente section décrit l'application de l'estimateur GREG à l'EMCGD, de même que les diverses options envisagées. Il importe de noter que les unités de l'échantillon à tirage complet ont été exclues du calage. En ayant recours au calage dans l'EMCGD, notre but principal est de réduire l'influence des unités classées de façon erronée (relativement à leur taille) sur les estimations; or les unités de l'échantillon à tirage complet ne causent pas ce genre de problème puisqu'elles sont auto-représentatives.

#### 5.4.1 Création de groupe de modèle

En plus de répondre aux critères énoncés à la section 5.3.1, les groupes modèles devraient idéalement se rapprocher le plus possible des domaines pour lesquels on
produit des estimations afin que l'on puisse éviter des estimations élevées de la variance. Par conséquent, le groupe modèle d'une unité donnée se prête à des variations d'un mois à l'autre puisque les unités peuvent changer de domaine au fil du temps. Par contre, dans le contexte d'une enquête mensuelle comportant un changement important d'un mois à l'autre, les groupes modèles devraient être aussi stables que possible. On garantira ainsi que le changement des estimations d'un mois à l'autre ne sera pas attribuable au fait que certaines unités ne relèvent pas du même groupe modèle que pour le mois précédent. Il faut tenir compte de ces deux préoccupations opposées lors de la création des groupes modèles.

Dans notre étude, les groupes modèles sont des groupements d'unités relevant d'un certain groupe de commerce et d'un certain code géographique. Ils contiennent au moins 25 unités échantillonnées actives. Chaque unité de la population doit être attribuée à un groupe modèle, et nous devons établir le total de la population de contrôle pour chaque groupe modèle. On peut obtenir des renseignements sur le groupe de commerce et le code géographique de trois sources.

**Source 1 - Stratification :** Chaque unité d'échantillonnage de l'EMCGD est stratifiée selon le groupe de commerce, la région géographique et la taille. Les renseignements sur cette stratification établie en décembre 1997 peuvent servir à créer nos groupes modèles. Cette source a l'avantage d'être stable au fil du temps; la stratification n'est pas mise à jour fréquemment dans l'EMCGD. L'inconvénient de cette méthode est que plus l'on s'éloigne de la dernière stratification, plus la stratification est périmée et différente des renseignements sur les plus récents domaines, ce qui risque d'entraîner une variance plus élevée dans les estimations.

**Source 2 - Utilisation du plus récent univers du RE :** Chaque mois, le RE crée un fichier de l'univers des populations du commerce de détail et de gros, reflétant les mises à jour provenant de sources administratives et des enquêtes. Ce fichier permet à l'EMCGD d'identifier les naissances et les cessations et de mettre à jour sa base de sondage et son échantillon. Ce sont là les plus récents renseignements sur les groupes de commerce et les régions géographiques pour ce qui est du RE. Cette information offre l'avantage de se rapprocher de l'information sur les domaines.

**Source 3 - Utilisation des renseignements sur les domaines :** Pour chaque compagnie statistique d'un échantillon, on obtient une combinaison de groupe de commerce et de code géographique à partir de l'information recueillie. Celle-ci est légèrement plus à jour que le plus récent univers du RE puisqu'elle intègre la collecte de données d'enquête plus récentes. Même si l'information n'est disponible que pour des unités échantillonnées, il serait possible d'utiliser les totaux de contrôle fondés sur la source 2, à supposer que celle-ci ne soit pas trop différente de la source 3.
Après avoir examiné ces trois options, nous avons décidé d'avoir recours à l'information de type stratification (source 1) pour établir les groupes modèles. Tout d'abord, la stabilité au fil du temps est une préoccupation majeure, à cause de l'ampleur du changement d'un mois à l'autre. Deuxièmement, notre but principal, pour ce qui est du calage dans l'EMCGD, est de réduire l'influence des unités classées de façon erronée relativement à leur taille. On peut y arriver même avec des groupes modèles fondés sur le groupe de commerce et le code géographique selon la stratification puisque la variable auxiliaire (total des ventes annuelles de biens assujettis à la TPS) est une mesure de la taille. Troisièmement, nous n'avons pas observé une bien grande différence dans les estimations de la variance en utilisant l'une ou l'autre des trois sources, bien qu'il soit possible que l'on observe une différence à l'avenir si la stratification n'est pas mise à jour périodiquement.

5.4.2 Traitement des unités d'échantillonnage inactives (cessations dans la base de sondage)

Pour chaque cessation que l'on identifie dans l'échantillon, la valeur de la variable d'intérêt \( y_k \) (ventes mensuelles) est de 0. Pour les \( x_k \) toutes les unités jugées actives dans le RE (y compris les unités inactives et non identifiées comme telles) comportent habituellement une valeur \( x_k \) positive. La question est de savoir comment traiter les cessations puisque la base de sondage de l'EMCGD (voir la section 2) contient une plus forte proportion d'unités de cessations identifiées dans l'échantillon que hors échantillon. Intuitivement, on trouve raisonnable d'attribuer une valeur de 0 à la variable auxiliaire \( x_k \) pour les cessations (unités dans l'échantillon aussi bien que hors échantillon). Il en résulte cependant un problème au moment de calculer les totaux de contrôle de la variable auxiliaire.

Nous avons vu à la section 3 que, pour calculer les facteurs \( g_k \), nous devons connaître le total de la population de contrôle pour la variable auxiliaire de chaque groupe modèle. Nous les calculons en prenant la somme de la valeur de la variable auxiliaire \( x_k \) dans chaque groupe modèle, y compris les cessations identifiées et non identifiées dans le RE. Si nous attribuons \( x_k = 0 \) aux cessations identifiées (unités tant de l'échantillon que hors échantillon), ce total de contrôle est alors lui-même touché par le fait qu'il existerait proportionnellement plus de \( x_k = 0 \) dans l'échantillon qu'à l'extérieur de l'échantillon. L'application du calage à ce type de population sans porter davantage attention aux cessations identifiées entraînerait une surestimation systématique des ventes mensuelles. Comme nous pouvons le constater pour le facteur \( g \) de l'estimateur par le quotient (qui n'est pas l'estimateur que nous utilisons, mais qui est plus facile à observer), nous avons pour le groupe modèle \( p \),

\[
g_{kp} = \frac{X_p}{\sum_{k \in p} a_k x_k} = \frac{X_p}{\hat{X}_p}, \quad (5.4)
\]

INSEE Méthodes 323
où $a_k$ est le poids du plan pour l’unité $k$ et $x_k$ est la variable auxiliaire pour l’unité $k$. Ce facteur $g$ sera toujours supérieur à 1 car il y aura plus de $x_k=0$ dans l’échantillon qu’à l’extérieur de l’échantillon dans chaque groupe modèle $p$. L’objectif est d’avoir des facteurs $g$ qui sont proches de 1 en moyenne. Deux options sont offertes pour le traitement des cessations identifiées.

**Option 1** Exclusion des cessations identifiées du calage. On impose simplement 1 à leur $g_k$. On applique ensuite le calage aux unités « jugées actives » seulement. Toutefois, si nous excluons les cessations identifiées du calage, nous devons modifier les totaux de contrôle connus puisqu’ils englobent la contribution des unités inactives et non identifiées qui sont largement présentes dans le volet hors échantillon de la base de sondage, mais non dans le volet échantillonné. Autrement, nous allons surestimer les ventes mensuelles parce que le total de contrôle de la variable auxiliaire sera trop grand comparativement au total estimé.

Afin de surmonter cette difficulté, nous corrigeons (déflation) le vecteur total de contrôle de la variable auxiliaire de chaque groupe modèle à l’aide d’un facteur qui représente le rapport entre le nombre estimatif d’unités actives dans chaque groupe modèle et le nombre d’unités actives dans la base de sondage (y compris les cessations non identifiées), de façon à obtenir le total de contrôle corrigé pour chaque groupe modèle $p$ :

$$X^*_p = X_p \times \frac{\hat{N}_{p(actives)}}{N_{p(actives)}} \tag{5.5}$$

où $\hat{N}_{p(actives)} = \sum_{k \in S_p} a_k I_k$ ($I_k=1$ si $k$ est considéré comme une unité active; $I_k=0$ autrement) est le nombre estimatif d’unités actives dans le groupe modèle $p$, et $N_{p(actives)} = \sum_{U_p} I_k$ ($I_k=1$ si $k$ est une unité active; $I_k=0$ autrement) est le nombre d’unités actives dans la base de sondage pour le groupe modèle $p$. Le recours à ce facteur de déflation est un prolongement du processus proposé par Hidiroglou et coll. (1995). Le total de contrôle corrigé $X^*_p$ servira alors de total de contrôle dans l’estimateur de régression.

Le facteur proposé fait diminuer le total de contrôle de 20 % environ dans chaque groupe modèle du secteur du commerce de détail, et de 30 % environ dans chaque groupe modèle du secteur du commerce de gros. Cela nous permet d’avoir de meilleurs totaux de contrôle. Un inconvénient de cette façon de procéder est que les nouveaux totaux de contrôle sont maintenant des valeurs estimatives comportant une variabilité inconnue, et que nous ne tenons pas compte de cette variabilité dans...
l'estimation de la variance. En réalité, l'hypothèse qui sous-tend notre estimateur de
la variance est que les totaux de contrôle sont constants.

**Option 2 Inclusion des cessations identifiées dans le calage mais moyennant un
traitement spécial.** La variable auxiliaire (total des ventes annuelles de biens
assujettis à la TPS) est imputée à l'aide d'une valeur positive pour toutes les
cessations identifiées à moins qu'une valeur soit accessible (comme c'est possible
pour les unités hors du champ d'observation). Le processus d'imputation se déroule à
l'aide d'un modèle de régression fondé sur une valeur historique du revenu figurant
dans des versions antérieures du RE. Tous les totaux de contrôle sont alors
recalculés à l'aide de ces valeurs imputées pour des cessations identifiées et à l'aide
des valeurs x originales des autres unités.

Un avantage de cette façon de procéder est qu'il n'est pas nécessaire de rajuster les
totaux de contrôle et que le facteur g a de nouveau une valeur moyenne de 1. Un
inconvénient est que le réglage du modèle qui sous-tend l'estimateur de régression
est moins efficace puisque nous avons des unités inactives (y_k=0 et x_k>0) dans le
modèle. Il devrait en résulter des estimations plus élevées de la variance. De plus, il
est à noter que nous ne tenons pas compte de la variabilité des x_k imputés dans
l'estimation de la variance. Pour le moment, les deux options restent possibles pour
notre étude.

### 5.5 Résultats

Nous avons retenu deux stratégies pour l'estimation. Un estimateur de régression
pour lequel les groupes modèles se fondent sur la stratification initiale est utilisé
pour les deux stratégies, la différence entre les deux relevant du traitement des
cessations, suivant l'option 1 et l'option 2 décrites à la section 5.4.2. Nous effectuons
toujours une estimation par domaine. Pour les deux stratégies, nous avons calculé
des estimations et des estimations de la variance. Le tableau 5 indique les intervalles
de confiance pour un niveau de confiance de 95 % et les coefficients de variation
pour certains domaines du secteur du commerce de détail. Le tableau 6 présente les
résultats pour certains domaines du secteur du commerce de gros. Ces intervalles de
confiance s'appliquent à des estimations mensuelles des ventes pour le mois de
référence avril 1999. Toutes les valeurs sont exprimées en milliards. Les intervalles
de confiance obtenus de l'estimateur par dilatation simple sont inclus à titre de
comparaison. Comme il a été expliqué à la section 2, l'estimateur par dilatation
simple est non biaisé, tandis que l'estimateur de régression est approximativement
non biaisé. On peut constater que les intervalles de confiance des deux estimateurs

*INSEE Méthodes* 325
de régression donnent lieu à des estimations qui ne sont pas appreciablement différentes de celles de l'estimateur par dilatation simple.

Comme on pouvait s'y attendre, l'estimateur de régression, pour les deux options de traitement des cessations, produit dans presque chaque cas des estimations comportant des coefficients de variation inférieurs à ceux de l'estimateur par dilatation simple. Il en va de même pour presque chaque groupe de commerce et chaque région géographique des secteurs du commerce et de détail. De plus, lorsque l'on exclut les cessations du calage (option 1), le modèle de régression qui sous-tend l'estimateur de régression est meilleur, et les estimations correspondantes de la variance comportent donc une variance moins élevée que celle de l'estimateur de régression lorsque les cessations sont incluses dans le calage (option 2).

Tableau 5
Intervalles de confiance avec α=0,05 (IC) et coefficients de variation (CV) des estimations des ventes mensuelles de détail d'avril 1999

<table>
<thead>
<tr>
<th>Commerce de détail, Secteurs</th>
<th>Estimateur de régression Option 1</th>
<th>Estimateur de régression Option 2</th>
<th>Estimateur par dilatation simple</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IC (en milliards)</td>
<td>CV (%)</td>
<td>IC (en milliards)</td>
</tr>
<tr>
<td>Supermarché et épicerie</td>
<td>[4,26, 4,39]</td>
<td>0,74</td>
<td>[4,30, 4,45]</td>
</tr>
<tr>
<td>Pharmacies</td>
<td>[1,04, 1,11]</td>
<td>1,67</td>
<td>[1,08, 1,16]</td>
</tr>
<tr>
<td>Véhicules récréatifs et automobile</td>
<td>[5,91, 6,47]</td>
<td>2,34</td>
<td>[6,22, 6,94]</td>
</tr>
</tbody>
</table>

INSEE Méthodes
Tableau 6
Intervalles de confiance avec α=0.05 (IC) et coefficients de variation (CV) des estimations des ventes mensuelles de gros d'avril 1999

<table>
<thead>
<tr>
<th>Commerce de gros, Secteurs</th>
<th>Estimateur de régression Option 1</th>
<th>Estimateur de régression Option 2</th>
<th>Estimateur par dilatation simple</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IC (en milliards)</td>
<td>CV (%)</td>
<td>IC (en milliards)</td>
</tr>
<tr>
<td>Pièces et accessoires de véhicules automobiles</td>
<td>[6,15 , 6,48 ]</td>
<td>1,34</td>
<td>[6,12 , 6,57]</td>
</tr>
<tr>
<td>Alimentation</td>
<td>[4,36 , 4,50]</td>
<td>0,80</td>
<td>[4,43 , 4,63]</td>
</tr>
<tr>
<td>Electronique, ordinateurs et machines connexes</td>
<td>[2,78 , 2,96]</td>
<td>1,58</td>
<td>[2,69 , 2,96]</td>
</tr>
</tbody>
</table>

5.6 Estimation par calage sur plusieurs mois

L'étude sur un mois de données donnaient des résultats très intéressants. Nous avons donc décidé de poursuivre l'étude en utilisant plusieurs mois de données. La période de janvier 1999 à juillet 1999 a été considérée.

5.6.1 Méthodes

Pour poursuivre nos tests et donc soumettre les deux options retenues sur plusieurs mois de données, deux objectifs, en quelque sorte opposés, sont visés :

1) Stabilité dans le temps, c'est-à-dire éviter que le facteur de calage varie grandement d'un mois à l'autre. Pour ce faire, il faut que l'estimateur retenu et ses options soient le moins affectés par le passage d'un mois à l'autre. Par exemple, si on change le fichier de ventes de TPS d'un mois à l'autre, il est évident que le facteur de calage changera conséquemment d'un mois à l'autre. Nous ne voulons pas introduire des sauts dans la série qui seraient simplement
dus à des changements au facteur de calage. En résumé, cet objectif se préoccupe d'abord et avant tout de l'aspect longitudinal.

2) **Meilleur calage possible**, c'est-à-dire utiliser de façon optimale l'information auxiliaire disponible pour effectuer la meilleure estimation par calage. Cet objectif se préoccupe d'abord et avant tout de l'aspect transversal.

Pour limiter notre charge de travail, nous avons décidé, dans un premier temps, de viser le meilleur calage possible. Ainsi, pour chaque mois d'estimation, la source la plus à jour disponible est utilisée. Ceci signifie que la valeur de la variable auxiliaire associée à chaque unité échantillonnée peut varier d'un mois à l'autre, de même que les totaux de contrôle au niveau des groupes modèles. Pour les estimations des mois de janvier 1999 à juillet 1999, nous avons utilisé la TPS du mois précédent ou sinon celle la plus récente. Les estimations ne couvrent uniquement que la partie commerce de détail.

Pour le traitement des cessations, nous avons retenu les deux options des tests sur le mois d'avril c'est-à-dire, les inclure ou les exclure. Le traitement des naissances vient ajouter une nouvelle dimension à l'estimation par calage lorsque répétée sur plusieurs mois. Nous n'avions pas eu à tenir compte de cet aspect dans les tests sur le mois d'avril 1999. En ayant choisi de réaliser le meilleur calage possible, nous rencontrons immédiatement notre objectif d'inclure les naissances puisque toutes les étapes sont refaites pour chaque mois indépendamment des autres mois.

**5.6.2 Analyse des résultats**

Les résultats sont présentés dans les graphiques 1 et 2. Les variables REG1 et REG2 correspondent respectivement à l’estimateur par régression des options 1 et 2 pour le traitement des cessations. Rappelons que l’option 1 consiste à exclure les cessations du calage alors que ces dernières sont incluses dans l’option 2. Les deux estimations sont comparées à l’estimation publiée et l’on retrouve les limites supérieures et inférieures de l’intervalle de confiance (95%) de cette dernière.

Le graphique 1 présente les résultats pour tous les groupes de commerce à l’échelle nationale. Nous constatons que les deux estimateurs se retrouvent dans les limites de l’intervalle de confiance de l’estimation publiée. Cependant, il est important d’étudier les résultats au niveau plus détaillé. Le graphique 2 présente les résultats pour les détaillants du groupe « autres magasins de vêtement ». La corrélation entre la variable TPS et les ventes de l’EMCGD est de 91%. Il s’agit du groupe de commerce qui possède la plus forte corrélation au niveau des petites compagnies à tirage partiel. Nous constatons que les résultats provenant des deux estimateurs par régression sont en dehors de l’intervalle de confiance et que le niveau des ventes estimées est supérieur à la valeur publiée. Cependant, pour plusieurs autres groupes
de commerce, les estimations par calage sont toutes à l’intérieur de l’intervalle de confiance.

L’EMCGD publie toujours au niveau des groupes de commerce. Il est donc important d’ avoir un estimateur très stable à ce niveau. Ainsi, il n’est pas clair si le problème est causé par l’estimation par calage ou par l’estimateur par dilatation. Nos analyses nous permettent de croire que le modèle de calage est adéquat mais que la variable auxiliaire n’est pas très stable et varie d’un groupe de commerce à un autre. Il est donc important de poursuivre les analyses pour mieux comprendre les enjeux que représentent l’utilisation de la variable TPS. L’une des avenues que nous explorons est de faire une expérience contrôlée et de fixer des paramètres. Il faudrait faire une simulation et fixer les ventes et calculer l’estimateur par calage dans le temps. Ceci permettrait de mesurer la robustesse de notre modèle.
Graphique 1
Comparaison des estimations calibrées avec les estimations publiées pour le commerce détail
Tous les groupes de commerce

Graphique 2
Comparaison des estimations calibrées avec les estimations publiées pour le commerce détail
Groupe 070 – Autres magasins de vêtements
6. Conclusion

L'utilisation des données de la TPS dans le cadre du remaniement de l'EMCGD semble être très prometteuse. Au niveau de la stratification, les gains sont importants puisque les problèmes d'unités mal classées représentent souvent une lourde charge de travail d'un mois à l'autre. Cependant, l'analyse des estimations provenant de plusieurs mois de données ne nous permet pas de comprendre les mécanismes qui régissent les deux types d'estimateurs, par dilatation et par calage. Il est nécessaire de poursuivre les analyses afin de mieux comprendre les faiblesses des données de la TPS. L'estimateur par calage réduit le coefficient de variation mais, la qualité varie beaucoup d'un groupe de commerce à l'autre à cause de différentes périodes de remise, du type de compagnies et, pour les compagnies complexes, de la réconciliation avec le RE. Il faudra éventuellement choisir l'une des deux options pour les cessations.

Le projet d'expérience contrôlée permettra de mieux évaluer les estimateurs. Parallèlement à ce projet, il est primordial d'améliorer la qualité des données de la TPS avant de se commettre à utiliser un estimateur par calage. D'ailleurs, Statistique Canada a entrepris un vaste programme pour améliorer la qualité. Ce nouveau programme vise à développer une base de données longitudinales, un système sophistiqué de vérification et d'imputation des données aberrantes et un processus pour simplifier la réconciliation avec le RE. Puisque le remaniement va être complété en 2003, il est fort possible que l'estimateur par calage soit reconsidéré lorsque tous les aspects de la TPS seront mieux maîtrisés.

7. Remerciements

Les auteurs tiennent à remercier tous ceux qui ont contribué à la réalisation de la partie méthodologique de cette enquête dont, Julie Girard, Martin St-Pierre, Michel Ferland et Nalma Gouzi. Les auteurs tiennent également à remercier Michel Hidiroglou et Jocelyn Tourigny pour leurs précieux commentaires.
8. Bibliographie


La qualité

Présidée par M. GLAUDE

INSEE, Direction Générale
Directeur des Statistiques Démographiques et Sociales
Présentation

(Michel GLAUDE)

« La qualité est un thème fortement émergent au cours des dernières années, au niveau national (plutôt dans le secteur privé, me semble-t-il, que dans le secteur public qui, comme toujours, suit avec un peu de retard les modes ou tendances...), mais surtout au niveau international. Bien évidemment, nous aurons l’occasion d’y revenir. Il semblerait que l’Insee et le système statistique public, sans y être vraiment réticents, aient des difficultés à intégrer cette approche de la qualité dans toutes ses dimensions, que nous allons ici détailler.

Alors, à moins de faire, comme M. Jourdain, de la qualité sans le savoir ni le dire, certains pensent que c’est une mise en scène dont on peut utilement se passer, ou que ce ne sont que des techniques managériales, dans lesquelles on dit : « c’est l’utilisateur qui compte, donc c’est de la qualité pour lui... ». 

Pour discuter de toutes ces questions, nous avons constitué un "panel de qualité", bien évidemment, réunissant des intervenants du secteur privé comme public, nationaux et internationaux. Interviendront donc Gilbert SAPORTA, professeur au CNAM, qui est titulaire de la chaire de statistique appliquée, Roberto BARCELLAN, responsable des comptes trimestriels à Eurostat, Yannick CARRIOU, Directeur Scientifique à la Sofres, Alain DESROSIERES, chef de la Division Méthodes Comparées, Insee, qui a beaucoup travaillé sur l’histoire de la statistique, Raoul DEPOUTOT, chef de la Division Statistiques Structurelles d’Entreprise, qui a passé un long moment à Eurostat et s’est occupé de qualité, et Michel BLANC, qui suit en particulier le "Leg qualité" au Département des Normes Statistiques et des Méthodes Comparées de l’Insee.

Je vous propose d’organiser notre discussion autour de trois noyaux importants et intéressants, qui sont :

- L’émergence des critères de qualité d’un produit statistique, c’est-à-dire la pertinence, la précision, l’actualité, la ponctualité, l’accessibilité, la clarté, la comparabilité, la cohérence, la complétude, et la question du coût ou du rapport coût-qualité. Ceux qui se sont intéressés à cette notion se rendent compte qu’il y a une sorte de consensus autour de ces critères de qualité d’un produit statistique qui ont peu à peu émergé. Ma première question sera : pourquoi ces critères-là ? Qu’est-ce qui a présidé à ces choix ? Quelle est l’histoire de l’émergence de ces critères ? Et comment faut-il les comprendre ?
- La qualité comme vecteur de management : dans les processus de management, la qualité apparaît comme un concept intéressant pour orienter les processus de production. G. SAPORTA interviendra sur les aspects plus industriels, et, du côté des équipes et du management d'équipe ou du management d'entreprise, nous aurons des interventions de Y. CARRIOU de la Sofres et de R. BARCELLAN d'Eurostat, qui nous présenteront l'approche qualité comme vecteur de management.

- Un ensemble de questions amenant une discussion plus ouverte, concernant la qualité dans le cadre du service public de la statistique : que veut dire la qualité dans un service public de statistique et non pas simplement dans un processus industriel, en particulier au niveau national et européen ? Il nous faudra aborder la question de la demande. Quelle demande y a-t-il ? Quelle appréciation de qualité y a-t-il de la part de nos utilisateurs ? Et quelle sanction prendre, en particulier, s'il n'y a pas de qualité ?

Quels sont les critères de qualité d'un produit statistique ? Pourquoi ces critères sont-ils apparus et comment sont-ils apparus ?

(Raoul DEPOUTOT)

« Le point de départ se situe en 1995. Eurostat s’est rendu compte, en faisant le bilan d’une certaine expérience dans l’harmonisation européenne, qu’il fallait arriver à concilier une normalisation des travaux tout en laissant place à une certaine flexibilité, une certaine subsidiarité. Il fallait donc arriver à gérer cette capacité, cette liberté laissée aux États membres, tout en s’assurant des résultats de l’opération. La façon la plus normale de le faire était de laisser à chaque État membre la liberté de s’organiser, dans chacun des règlements, au-delà d’un certain socle global défini et encadré. En échange de cette liberté et de cette spécificité nationales, l’État membre doit produire un rapport sur la qualité de ses statistiques. Pour dépasser ce stade des grands principes, il fallait être capable de préciser quelle est cette qualité, comment on peut la quantifier et éventuellement, par la suite, aboutir à définir des critères de qualité minimale que devraient respecter les statistiques nationales.

Pour faire cela, nous nous sommes reportés à la littérature. Par chance, en 1995, un colloque international avait été organisé sur ce thème à Bristol, au cours duquel avaient été invités les grands spécialistes de l’époque du monde développé, dont Cathy DIPPO, du Bureau of Labor Statistics (USA) et Lars LYBERG, de Statistics Sweden. Cette conférence a permis de prendre connaissance à la fois de travaux généraux s’appliquant à toutes les statistiques et de travaux plus particuliers dans certains domaines. Nous avons également fait l’historique de ce qui avait été publié.
dans le passé. Il semblait que le point d’ancrage de tous ces travaux, avec déjà beaucoup de précision dans l’approche, se situait dans les années 1975-1977, aux États-Unis, avec des travaux sur la qualité de l’enquête “Current population survey”. Les États-Unis avaient alors reçu une demande qui émanait de l’autorité politique, afin de mesurer, dans cette enquête, la précision exacte du niveau d’emploi estimé. Un Comité inter-agences statistiques avait défini les bons critères et avait publié un premier document qui faisait une synthèse de tout ce que l’on pouvait dire sur les spécificités de cette enquête, sa précision etc... Ce fut le premier rapport qualité publié.

La deuxième approche a été de rendre visite à nos collègues de Statistics Canada, institut qui s’était signalé dans le domaine par sa charte "qualité". Celle-ci s’avère en fait être essentiellement une référence culturelle pour les agents, un engagement moral de donner de l’information aux utilisateurs, beaucoup plus qu’un outil de management.

La troisième référence dans le monde était Statistics Sweden, beaucoup plus proche d’Eurostat, qui de plus était un nouveau membre de l’Union Européenne et qui avait intégré la qualité dans sa gestion.

Nous avons également bénéficié des travaux qui sont incontournables du point de vue universitaire, ceux de Bob GROVES qui a publié notamment un ouvrage intitulé « Survey cost and survey errors », où il montre dans une approche très pédagogique comment on peut appréhender la qualité. Il y avait une cohérence entre ces travaux et ce qui avait été publié officiellement par Statistics Sweden, qui était : on peut décrire la qualité d’une enquête - car à l’époque la formulation était en termes d’enquête - par quatre grandes composantes qui sont : le contenu de cette enquête, la précision des statistiques qui en sont dérivées, la dimension de rapidité et la possibilité de communiquer avec l’usager (que nous avons plus tard appelée clarté et accessibilité).

En partant de là, nous avons entamé à Eurostat une discussion, interne tout d’abord, aidés de « consultants » qui, en l’occurrence, étaient des fonctionnaires de Statistics Sweden. Nous avons peu à peu incorporé à la conception en vigueur à Statistics Sweden une dimension spécifique européenne.

Celle-ci se caractérise par deux choses :

- la distance par rapport à l’enquête, puisque, dans beaucoup de domaines, tous les pays n’utilisent pas forcément une enquête, certains ont des sources administratives, sont organisés autrement ; on cherchait donc à attacher cette fois-ci la qualité à une statistique, c’est-à-dire à un chiffre agrégé et produit.

- une dimension nouvelle qui est une obsession, lorsque l’on travaille à Eurostat, c’est-à-dire la cohérence du message entre statistiques différentes, la comparabilité des statistiques provenant d’États membres différents ou de services différents : par exemple, on peut parler de qualité quand des
statistiques mensuelles, trimestrielles et annuelles donnent, au bout du compte, des messages compatibles.

Nous avons beaucoup travaillé autour de cela et avons pris en compte un dernier élément qui est très inscrit dans l'histoire du système statistique européen autour d'Eurostat, qui est l'extension de champ de la statistique pour la plupart des Etats membres au moment de l'intégration de la dimension européenne, la nécessité de construire des statistiques dont ils n'avaient pas besoin auparavant pour leurs propres besoins nationaux. C'est en quelque sorte la complétude du système statistique.

Voilà comment s'est développée petit à petit la définition de la notion de qualité en sept composantes. »

Lecture historique

(Alain DESROSIERES)

« Je vais essayer de faire une lecture historique du thème de la qualité en statistique, qui s'articule avec celle de Raoul DEPOUTOT. Le terme « qualité » n'a surgi vraiment avec son sens actuel que depuis 6 ou 7 ans. Mon hypothèse est que c'est une façon de combiner trois choses assez différentes, en les articulant de façon spécifique. On peut comprendre cette articulation en en retraçant l'histoire.

Le premier point est le contrôle de fabrication, le contrôle de qualité au sens où G. SAPORTA va en parler tout à l'heure.

Le deuxième est le management participatif.

Le troisième point, c'est la montée de la contractualisation.

On pourrait raconter l'histoire comme celle d'un thème qui a fait le tour du monde, tout en se transformant. Pourquoi le tour du monde ? Le point de départ serait le travail, aux Etats-Unis dans les années 1930, d'un statisticien américain, DEMING, un des pionniers des enquêtes par sondage au Bureau of Census américain : il a fait, avec 4 ou 5 autres statisticiens, les premières grandes enquêtes sur l'emploi, c'est-à-dire l'application des méthodes de NEYMAN pour les sondages. Puis, dans les années 1940, il s'est intéressé à l'utilisation de ces méthodes d'échantillonnage probabiliste (qu'il avait d'abord utilisées pour faire des sondages dans les enquêtes socio-économiques) pour les transporter ensuite dans les entreprises pour faire du contrôle de fabrication, c'est-à-dire de l'échantillonnage de pièces, pour vérifier si ces pièces étaient de bonne qualité. Vous avez le premier sens de la qualité, qui est le contrôle de fabrication par des techniques statistiques.
Il est frappant que, ensuite, le thème ait fait le tour du monde mais que, à un moment donné, le fait qu’il était issu initialement de techniques statistiques a été oublié. DEMING a essayé d’implanter ses nouvelles techniques probabilistes dans les entreprises américaines, mais il n’a pas eu beaucoup de succès dans un premier temps. Cela n’a pas bien marché dans les années 50. Il est donc allé au Japon, qui l’a accueilli à bras ouverts. Il a pu développer dans les entreprises japonaises ses techniques, mais, petit à petit, il s’est rendu compte que, pour traduire concrètement le résultat de ses travaux, il fallait que le personnel, et notamment les ouvriers, soient impliqués. Il a donc transformé le thème « qualité » dans ce sens, avec une forme de mobilisation du personnel typiquement japonaise, qui s’est développée en gros dans les années 1960-70. Ceci a impliqué en particulier la mise en place des fameux "cercles de qualité", qu’on a vu arriver ensuite en Europe dans les années 1980. C’est pourquoi on peut dire que cela a fait le tour du monde.

Mais à ce moment-là, c’était devenu un discours de management et l’idée du contrôle de fabrication originel de DEMING était un peu passée dans l’ombre, à telle enseigne que, quand, dans les années 1980 à l’INSEE, on a vu ces techniques apparaître, il y a eu des gens qui ont été les pionniers de ces méthodes, mais le fait qu’à l’origine ces méthodes aient été pratiquées par des statisticiens avec des techniques statistiques, avait été complètement oublié. A ce moment-là, les cercles de qualité, c’était un moyen pour mobiliser les gens, les impliquer, les concermer, une réadaptation de l’idée un peu soixante-huitard de l’auto-gestion, qui était récupérée par le management des directions des entreprises - ça c’est notre deuxième thème -.

Ce thème, comme vous le savez, est un peu passé de mode maintenant, car il a été inclus dans quelque chose de plus vaste, qui est le fait que, de plus en plus, les rapports entre les entreprises et leurs clients, d’une part, et, d’autre part, entre les administrations et leurs utilisateurs - qui sont devenus des clients -, se sont développés sur des bases contractuelles, à partir de cahiers des charges. Il y a des appels d’offres, système dans lequel les Français ne sont pas encore très à l’aise mais qui est très important au niveau européen. Il y a aussi des sous-traitances, et toutes ces formes de relations de division du travail, division à la fois technique et sociale du travail, qui se font sur des bases de cahier des charges, donc de spécifications contractuelles. Ce qui veut dire qu’il est important de définir la qualité : c’est ce que le produit doit être contractuellement. Il me semble que c’est un point important. Si l’on n’a pas cela en tête, on ne comprend pas bien pourquoi le thème "qualité" a monté. En effet, le fait de dire : "oui, on a envie de présenter des produits de bonne qualité" semble équivalent de l’annonce évidente du poissonnier : "ici on vend du poisson frais". Cela n’a pas l’air d’être spécifique..., tandis que, si l’on voit que c’est lié à ce développement des formes de relations contractuelles, à des relations de confiance à instaurer et entretenir entre fournisseurs et clients, sur des produits dont les spécifications ne sont pas évidentes, on comprend un peu mieux.

Pour finir, on peut dire un mot sur les six critères que R. DEPOUTOT vient d’enumérer. Je trouve que dans les deux premiers critères, la façon dont ils sont formulés dans les documents d’Eurostat semble illustrer ce que j’ai dit sur l’idée de la
contractualisation des relations. Le premier critère, la "pertinence" ("relevance" en anglais), est traitée en cinq lignes, dans lesquelles il est dit : on fera des enquêtes auprès des utilisateurs, pour savoir s'ils sont satisfaits. Par contre, après, il y a "précision", "accuracy", là on retrouve les techniques de l'ingénieur, et là il y en a cinq pages, dans l'esprit des premiers travaux de DEMING.

Pour un épistémologue, un spécialiste de science sociale (un économiste, un sociologue ou un philosophe), le fait que l'idée de "pertinence" soit réduit à cela est surprenant, puisque cette idée de pertinence est cruciale quand on veut faire des mesures dans des sciences sociales, ou pour une action... Autrement dit, il semble que la seule façon de comprendre cela est à chercher du côté de la division du travail entre, d'un côté, les "subject matter spécialisés" qui sont en rapport avec la demande, et, d'un autre côté, les méthodologues. Cette division du travail est peut-être moins poussée en France que dans d'autres pays, ce qui est plutôt heureux. Le fait que la pertinence soit définie de cette façon minimale dans les documents Eurostat, en faisant abstraction tant des nombreuses réflexions des sciences sociales sur cette notion que d'une approche de consultation collective, façon CNIS, semble digne d'être examiné.

Je suggérerai pour conclure qu'il faudrait, d'une certaine façon, rétablir le continuum entre les questions de pertinence et les questions de précision. Il ne faudrait pas les séparer totalement : il suffit de prendre pour exemple la mesure du chômage, les problèmes du codage des trois critères du BIT. Ces problèmes peuvent être analysés des deux points de vue, selon que l'on se pose des questions de signification ou de mesure. On est là en plein, à la fois, dans des critères de pertinence et de précision, qu'on ne peut absolument pas scinder. »

(Michel GLAUDE)

« Il y a deux questions que je voulais poser :

Premièrement, à partir de ce qu'a dit A. DESROSIERES, comment va-t-on mesurer, ou mettre en œuvre, ces fameux critères ?

Et deuxièmement, je demanderai bien à nos amis de la Sofres ce qu'ils pensent de ces critères ? Est-ce qu'ils les connaissent ? Les appliquent-ils ? Ont-ils les leurs... ?

Mais, auparavant, R. BARCELLAN sur la compréhension de ces critères. »
La compréhension de ces critères

(Roberto BARCELLAN)

« En ce qui concerne la pertinence, il y a une évolution de ce concept. C'est vrai qu'il s'agit principalement des enquêtes auprès des utilisateurs, mais nous devons tenir compte, à Eurostat mais aussi dans les Offices Statistiques, de la nécessité de mesurer un phénomène, qu'il soit économique ou social...

On doit trouver des mesures qui puissent donner une image de ces phénomènes, et c'est exactement sur cette base que l'on définira les critères de qualité : cela veut dire tenir compte de ce que les utilisateurs demandent, car, pour mesurer un phénomène, il faut aussi une demande de la part des utilisateurs, dans l'optique de fournir un service, un service qui doit être de qualité. Alors, en suivant et appliquant les critères de qualité, on doit pouvoir cibler sur ce que les utilisateurs demandent.

Mais il ne faut pas oublier qu'il ne faut pas seulement atteindre les utilisateurs : il faut aussi prévoir les besoins qu'ils pourraient avoir dans quelques mois ou années. C'est sur cette base que la « relevance », c'est-à-dire la pertinence du concept statistique, est mesurée et évaluée à l'intérieur d'Eurostat, et, de plus en plus, à l'intérieur des Instituts Statistiques. »

(Yannick CARRIOU)

« Du côté de la Sofres, il est vrai que ces critères, on les reconnaît parfois sous des vocables un petit peu décalés. Ce qu'il est important de dire, c'est que, comme nous sommes dans une relation avec des clients, les clients ne reconnaissent pas forcément tous ces critères. Il est clair que la complétude, les délais et autres, sont des choses auxquelles ils sont tout à fait sensibles, la pertinence en principe aussi, mais, sur d'autres critères, c'est parfois plus discutable.

Donc, la difficulté à gérer lorsqu'on défend une éthique professionnelle, c'est de dire parfois à certains clients satisfaits que l'étude n'est pas tout à fait satisfaisante ; c'est quelque chose qui se fait, qui émane d'une certaine éthique et d'un certain sérieux, mais il y a vraiment 2 niveaux : une qualité perçue et une qualité réelle. On peut surfer de temps en temps sur la vague de la qualité perçue mais la vision qu'on en a est beaucoup plus globale... »

(Michel GLAUDE)

« Monsieur SAPORTA, avez-vous une réaction sur ces critères ? »
(Gilbert SAPORTA)

« Je n'ai pas vraiment de réaction là-dessus, car, en fait, mon expérience est beaucoup plus en relation avec le domaine industriel, où là, a priori - mais ce n'est pas encore tout à fait exact - , on a des indices, des mesures de qualité qui sont relativement objectives. Quand on se rapproche effectivement de la qualité de service, le grand problème est la définition des mesures de qualité comme, par exemple, la qualité des réponses à un centre d'appel téléphonique ; là, il y a toute une réflexion sur les bonnes variables à mesurer ; une fois que l'on a toutes ces variables, ce n'est plus que de la technologie. »

(Michel GLAUDE)

« Avez-vous réfléchi sur les façons de mesurer ou de donner des indicateurs pour ces différents critères ? »

(Roberto BARCELLAN)

« Oui, Eurostat a réfléchi à cette dimension, c'est-à-dire comment mesurer ces indicateurs. L'idée est que l'on essaie d'avoir principalement des statistiques sur la façon dont ces critères sont perçus auprès des utilisateurs, mais aussi à l'intérieur de l'entreprise. Cela parce que l'idée de qualité sous-entend cette double dimension : qualité envers les utilisateurs mais aussi qualité perçue - c'était le parcours historique - à l'intérieur de l'entreprise. Nous avons créé par exemple un ensemble d'indicateurs qui permettent d'évaluer l'évolution que suit la qualité à l'intérieur d'Eurostat, mais aussi à l'intérieur des Instituts Statistiques nationaux, et cela en tenant compte de la réaction du grand public : les utilisateurs génériques, les journaux, les utilisateurs économiques ou sociaux. C'est sur cette base que nous sommes en train de travailler. »

(Michel GLAUDE)

« Nous l'aborderons donc dans la seconde partie autour de la qualité comme vecteur de management.

Peut-être quelques réactions de la salle sur ce premier ensemble relatif aux indicateurs de qualité "des produits statistiques", leur émergence, pourquoi retenir ceux-ci ? Leur mesure ?... »
(Daniel VERGER)

« Est-ce que ces critères sont parfois contradictoires ou non ? Y a-t-il une certaine hiérarchie en cas de contradiction ? »

(Raoul DEPOUTOT)

« Un certain nombre de composantes, que l'on a plus ou moins recensées, possèdent la particularité de dégrader l'une des autres composantes quand on cherche à les améliorer à coût constant. Dans les couples ennemis classiques, il y a : la rapidité et la précision, la pertinence et la précision (il est plus souvent facile de mesurer avec une grande précision des concepts moins pertinents, on a souvent du mal à saisir avec précision les concepts les plus pertinents). On sait aussi que, dans un certain nombre de cas, on a envie de définir des concepts de façon adaptée à un secteur, pays ou domaine particulier, et que c'est tout à fait contraire à la démarche de comparabilité. De la même façon, un autre exemple classique : c'est la comparabilité dans le temps. On a envie d'adapter nos critères de mesure à l'époque (pour améliorer leur pertinence) et l'on veut conserver la continuité des séries historiques (encore appelée comparabilité temporelle). Cela fait partie des contradictions permanentes avec lesquelles on doit jouer dans notre métier.

Sur la mesure de la qualité, je voulais dire qu'il y a à la fois une documentation plus détaillée, qui est disponible sur les sites d'Eurolstat ; il y a un "cours TES" qui est fait régulièrement sur les mesures de la qualité en statistique ; et il y a des règlements qui commencent à appliquer ce concept et à demander aux Etats membres de fournir un certain nombre d'informations sur la qualité. Donc, la mesure de la qualité est devenue obligatoire, au moins pour une partie de ces critères, dans le cadre du règlement sur les statistiques structurelles d'entreprise et également pour la Labour Cost Survey et pour la Labour Force Survey, je crois.

C'est une mise en œuvre effective qui s'appuie sur un mode opérationnel : il s'agit d'une très bonne étude de référence faite avec un financement Eurostat, par un groupe de travail commun entre l'Office Statistique Britannique, l'Office Statistique Suédois et l'Université de Southampton. Il a produit un gros pavé très bien documenté sur, à la fois, la théorie de la mesure de la qualité et quatre cas pratiques de mesure de qualité. »

(Michel GLAUGE)

« Est-ce que le coût pour un certain produit, c'est-à-dire le rapport qualité-coût, est un critère qui est pris en compte ou bien cela doit-il arriver plus tard dans l'analyse ? »
(Roberto BARCELLAN)

« Avant de vous parler du coût, très rapidement, quel est le critère le plus important ?

Naturellement, tous sont importants, mais cela dépend de la priorité qu'à un moment donné on attribue à chacun, et cela dépend beaucoup, dans la démarche qualité, des enquêtes que l'on fait auprès des utilisateurs. Exemple : quand les données sur le PIB européen sont-elles disponibles ? Doit-on attendre tous les pays pour avoir plus de précision ? Ou doit-on anticiper l'estimation pour avoir la disponibilité rapide de ces données ? Là, il faut satisfaire une demande, et, dans le cas spécifique, la priorité, en ce moment, est dictée par la demande du marché : actuellement, cela correspond à avoir plutôt de la rapidité et un peu moins de précision, avec naturellement des révisions successives.

Dans tout cela, l'élément "coût" joue un rôle très important. Est-ce un critère ou non ? C'est difficile à juger, mais on ne peut pas faire des statistiques de qualité, dans le sens de précision ou qui respectent tous ces critères, sans tenir compte de la dimension coût. Il faut toujours évaluer les coûts et les bénéfices qui ont trait à cette approche, et, ensuite, évaluer quelle est la solution la meilleure qui produise les résultats les plus intéressants sur la base de ces deux dimensions. Pour les Offices Statistiques publics, il y a une autre dimension : nous devons fournir un service public, un service d'information statistique. Il y a certaines statistiques que nous devons produire, avec de la qualité, même si le coût pourrait suggérer, dans un domaine privé, que ce n'est pas vraiment rentable. »

(Intervention du public)

« Il y a un mot que l'on utilise, c'est coût-efficacité ou coût-avantage. »

(Michel GLAUDE)

« Vous avez raison, un des éléments de la qualité, c'est l'efficience. »

(Intervention d'un représentant de la Division Commerce)

« Le coût, je le vois plutôt comme une contrainte, puisque ce que l'on cherche à faire, c'est améliorer la qualité à coût constant, si on prend le coût comme la charge de l'Office Statistique ; par contre, si on prend le coût comme le temps que le répondant a mis pour répondre au questionnaire, là cela peut être un critère de qualité de l'enquête statistique. Je voudrais savoir ce que pensent les orateurs à ce propos. »
(Raoul DEPOUTOT)

« Pour définir en commun la qualité des statistiques, il y avait un groupe de travail, dans l'appellation Eurostat, qui a examiné ces propositions afin de rechercher un accord des États membres. La question du coût a été posée. Dans la plupart des pays qui ont fait des travaux sur la qualité, ou chez les universitaires qui travaillent dans ce domaine, en général, on refuse d'intégrer le coût à la définition de la qualité : "c'est une étape ultérieure dans la relation avec l'utilisateur, c'est le rapport qualité-coût". Une analyse plus fine du délégué Insee qui représentait la France, Pascal RIVIERE, a effectivement distingué deux coûts en disant : "il y a le coût pour l'administration et le coût pour les enquêtés".

D'une certaine façon, une bonne enquête, c'est celle qui donnerait à la fois une bonne information sans charger les répondants. Cela devenait un peu plus subtil de rejeter cette partie-là du coût. Mais comme ce sont les approches simples et pragmatiques qui conduisent au consensus, cette distinction n'a pas été incorporée. Mais il est vrai que si l'on voulait poursuivre les travaux sur le plan théorique, il faudrait sans doute inclure la partie "coût sur les enquêtés" comme une dimension de la qualité. C'est bien ce qui justifie que l'utilisation des données administratives soit si encouragée. »

(Michel GLAUDE)

« Je vous propose de passer à la seconde partie qui concerne la qualité comme vecteur de management, comme l'a déjà introduit Alain DESROSIERES dans sa lecture historique.

On va d'abord parler de qualité des processus, des procès de fabrication (G. SAPORTA), puis on abordera les modèles de qualité totale, qui seront introduits par M. BLANC et commentés par deux expériences : à la Sofrè et à Eurostat. »

Les méthodes qui permettent de suivre la qualité dans les processus

(Gilbert SAPORTA)

« Je parlerai plutôt de ce que je connais, la qualité dans les produits manufacturés et cela sans revenir sur ce qu'a présenté A. DESROSIERES, qui était tout à fait pertinent sur le plan historique. La qualité, ça se mesure, c'est le premier message qu'il y a effectivement dans l'industrie, et l'on peut utiliser des outils extrêmement simples. »
Voilà deux outils qui sont utilisés dans les cercles de qualité, pour mesurer la non-qualité. Pour agir, il faut savoir identifier les défauts les plus fréquents et les plus importants ; pour cela, on utilise un diagramme, appelé le diagramme de Pareto, par allusion à la fameuse loi de Pareto dite des 80-20 (80% des problèmes viennent de 20% des causes), et si l'on réussit à les identifier proprement, on sait par où il faut commencer.

Un autre exemple, le diagramme en "arête de poisson" ou diagramme d'Ishikawa, très prisé dans les cercles de qualité japonais, qui est une façon de représenter en quelque sorte le cheminement des diverses causes de non-qualité ; c'est un outil assez simple d'analyse et de discussion.

Si l'on revient dans les problématiques, il est intéressant de savoir, comme l'a dit A. DESROSIERES, que les méthodes de statistique de la qualité ont fait le tour du monde ; elles sont revenues ensuite aux Etats-Unis mais il y a eu dans leur diffusion des choses assez surprenantes. On a commencé par faire, si l'on peut dire, le travail un peu à l'envers. Les premières méthodes qui sont les plus diffusées, étaient les
métodes de type plan de sondage ou d'échantillonnage ; c'était pendant la seconde guerre mondiale où il s'agissait de contrôler des lots de produits finis pour savoir si on les acceptait ou si on les refusait ; on a donc des techniques de plan d'échantillonnage plus ou moins compliquées, et qui sont conçues pour analyser des lots qui se présentent régulièrement ; on voit, par exemple, qu'avec tel type de plan, en moyenne, on est sûr d'avoir une qualité qui, en moyenne, je le souligne, ne sera pas inférieure au sommet de la courbe de qualité moyenne dite "transmise".

On a donc commencé par la fin car ça n'améliore pas la qualité puisque c'est finalement du jugement a posteriori.

Pour pouvoir faire de la qualité, bien souvent les gens se sont dit : "il faut savoir si on est capable de la faire ". On a alors les études de "capabilité", qui se représente avec un petit schéma où l'on a une distribution sur une caractéristique numérique, et l'on essaie de regarder si on arrive à coincer cette distribution dans un intervalle à plus ou moins 3 écarts-types, d'où le fameux slogan "les méthodes 6 sigma" ; donc, si on arrive à la coincer dedans, on est sûr que l'on aura environ 3 défauts sur 1 million de pièces produites. Ceci sous-entend que, pour que ça marche, il ne faut pas que la distribution bouge, donc il faut contrôler la moyenne ; il ne faut pas non plus que l'écart-type augmente, donc il faut contrôler l'écart-type et l'on débouche sur tout ce qui est carte de contrôle.
Il y a de nombreux débats quand on parle de spécification ou de tolérance, car cela veut dire que l'on accepte à l'avance qu'il y ait un certain taux de pièces non conformes et ceci n'est pas conforme à la philosophie de Taguchi, où l'on ne veut pas entendre parler de ça : ce que l'on veut, c'est tendre vers la qualité la plus grande, le fameux 0 défaux, qui n'est d'ailleurs qu'un slogan, ou en tout cas une asymptote ; mais admettre que l'on a des tolérances, c'est admettre que l'on puisse en sortir.

Les cartes de contrôle. Ce sont des outils qui datent des années 30, mis au point par SHEWHART aux laboratoires Bell. Il y a encore des entreprises qui ont du mal à les utiliser, sans parler de cartes un petit peu plus élaborées comme celles où l'on va utiliser l'information passée sous forme de "lissage" exponentiel pour détecter plus vite des dérives. La carte EWMA détecte ici un dérèglement d'un procédé d'emballage de paquets de thé au prélèvement n°14 alors que la carte de SHEWHART classique ne montre aucun point hors contrôle.

Dans le domaine industriel, après avoir fait le contrôle de la qualité une fois le processus achevé, et maîtrisé les variations en cours de production (c'est l'objectif des cartes de contrôle), on se dit : "si l'on avait des procédés, des méthodes de fabrication qui étaient déjà bonnes à l'avance, qui empêcheraient les procédés de se dérégler, on aurait économisé encore plus" : c'est l'objectif des plans d'expérience, des méthodes "Taguchi", qui, en faisant un certain nombre d'essais bien choisis, vont
permettre d'essayer de modéliser, de façon empirique, une réponse en fonction d'un certain nombre de caractéristiques et surtout d'obtenir un optimum qui soit stable. C'est là où l'on rentre dans des développement récents, comme les approches autour de la conception robuste qui consiste à trouver des points de fonctionnement qui soient relativement insensibles à des agressions, des bruits extérieurs.

Les recherches s'orientent aussi vers les méthodes multivariées : quand on a plusieurs caractéristiques de qualité simultanées et corrélées, comment peut-on prendre des décisions ? Quand arrêter ou régler le processus et régler quoi ? Et, de plus en plus, on a des travaux qui sont un peu en amont et en aval, pour relier qualité et satisfaction finale du consommateur ; exemple : avec des méthodes multivariées, on va regarder les corrélations entre les variables du procédé de fabrication et les mesures de qualité finale. Ceci se relie aussi, dans certaines approches, avec des mesures de satisfaction des consommateurs, mais cela a déjà été évoqué.

Qualité des données et MSP : "Maîtrise Statistique des Procédés" (Statistical process control). Ce sont des outils qui font partie de tout un ensemble : management de la qualité totale, qui a été évoqué et qui va sans doute l'être encore ; une question que je me pose est de savoir si l'on peut effectivement transposer ce qui est fait dans l'industrie (procédé de fabrication, par exemple, des pièces individualisées ou des process continus quand on fabrique des fluides, des gaz, etc.) ? Peut-on transposer ceci à cet autre processus qu'est le processus de production de l'information statistique ?

(Michel GLAUDE)

« Merci, je peux dire en tout cas qu'il y a eu du contrôle de qualité sur un certain nombre de processus du Recensement de Population : qualité de la saisie, qualité de la codification… qui repose sur ces méthodologies.

Nous allons poursuivre avec les aspects "qualité totale" et les discours et les présentations autour de ce thème, puis avec deux exemples. »

L’approche de "total quality management"

(Michel BLANCO)

« "Total quality management" : "Gestion de la qualité totale", en français. C'est un premier point, parce que l'on parle de qualité totale, on parle de gestion de cette qualité, mais "total quality management", ça se prête à d'autres traductions et l'on voit bien, dans ce que l'on a dit tout à l'heure, quel est le mot qui devient important : ce n'est ni "quality" ni "total" mais c'est "management".

348 INSEE Méthodes
On le voit, par exemple, dans des modèles venus de la sphère privée, des entreprises, qui s'étendent de plus en plus à l'ensemble du monde académique puis des administrations publiques, y compris françaises, et en particulier au sein du Ministère où nous nous trouvons. Il y a d'ailleurs une tradition officielle employée dans le cadre de la réforme de l'État qui est : « management par la qualité totale » ; on voit donc là une dérive sémantique qui est intéressante.

Un des modèles assez répandu de "total quality management" est le modèle dit « Modèle d'excellence EFQM » (European Fondation for Quality Management). Cette fondation est une association européenne créée en 1988 par un groupe de grandes entreprises européennes avec des appuis financiers de la Commission, et dont le but était de promouvoir la qualité dans les entreprises ; depuis lors, ses adhérents, ou ses membres, se sont étendus, comme je l'ai dit, au monde académique, à l'administration, y compris aux Instituts de Statistique.

La représentation la plus générale de ce modèle, qui est censé s'appliquer à toute organisation, toute entreprise, tout organisme, se présente sous forme de 9 critères.

**MODELE D'EXCELLENCE EFQM**

Dans les facteurs, le premier (le leadership) montre la manière dont l'organisme exerce le leadership, ce sans quoi il n'y a aucune démarche de qualité possible dans l'organisme. Ensuite il y a politique et stratégie : la manière dont l'organisme met en œuvre sa mission au travers de plans, d'objectifs, de processus... ; la manière dont l'organisme gère et développe son potentiel humain, la manière dont l'organisme définit, traite avec ses partenaires externes et aussi gère ses ressources internes (ici on trouve les ressources financières, notamment l'équipement, etc.) ; et la manière dont l'organisme définit, gère, améliore ses processus.

Les facteurs de résultats donnent lieu à des indicateurs généralement quantitatifs, les indicateurs relatifs aux cinq premiers facteurs étant plutôt qualitatifs. Dans les résultats, on peut mesurer ce que fait l'organisme vis-à-vis des clients ou des citoyens, dans le cas général de l'administration (pour nous, on dira plutôt des utilisateurs) ; il s'agit à la fois de la perception qu'ont les clients ou utilisateurs de la qualité ou des prestations de l'organisme, mais également d'indicateurs de performance. On retrouve le critère "personnel" que l'on avait en ressource (quelquefois on traduit "satisfaction du personnel" mais ce n'est pas uniquement cela) ; on trouve des résultats sous forme d'impact sur la société, au sens de l'environnement en général et non pas de l'entreprise : par exemple, si c'est une entreprise industrielle, on pourrait dire que ce sont notamment les efforts qu'elle fait pour lutter contre la pollution ; pour la statistique, ce sont ceux fait par l'Institut statistique pour préserver la confidentialité des données... Enfin les résultats, ce sont des indicateurs directement liés aux types de prestations visés par l'entreprise.

Ce modèle est supposé être adaptable, c'est un modèle d'auto-évaluation, on analyse, on peut mettre des scores, en pondérant chacun des critères ou des sous-critères. Tout cela est fait pour entreprendre des actions. Après les mesures d'indicateurs, on est censé améliorer ces facteurs. »

(Michel GLAUBE)

« Deux exemples, tout d'abord la mise en œuvre de ce type de modèle à Eurostat. »

La mise en œuvre de ce type de modèle à Eurostat

(Roberto BARCELLAN)

« L'idée de qualité, c'est quelque chose de présent au sein d'Eurostat, car la mission d'Eurostat est de fournir à l'Union européenne un service d'information statistique de qualité. Pour satisfaire cette mission, il faut que la qualité soit aussi à l'intérieur de l'organisation d'Eurostat et c'est de cette façon que le problème de la qualité a été approché au cours des 5 dernières années.
Il y a eu un processus qui a analysé la situation, qui a mis en lumière les points faibles et les points forts, qui a proposé des objectifs et qui a donné des instruments pour atteindre ces objectifs et des instruments pour contrôler l'état de la qualité à l'intérieur d'Eurostat.

L'idée de base était la création de ce que l'on appelle "corporate planning", ce qui veut dire : définir le plan d'entreprise pour l'Office statistique de la communauté européenne, Eurostat.

Pour ce faire, il y a eu un contact direct avec les gens qui travaillent à Eurostat. Au départ, c'était plutôt un travail à partir d'idées qui venaient de la littérature, de l'approche qualité totale, mais aussi des idées émanant du personnel qui travaillait à l'intérieur d'Eurostat. Avec la double dimension : prendre des suggestions émanant des personnes, mais aussi en faire vis-à-vis d'elles et exercer une cession du pouvoir aux personnes (l'idée de base, c'est ce que l'on appelle en anglais "empowerment", ce qui veut dire donner de la responsabilité aux personnes dans leur domaine, en reconnaissant ce qu'elles font et, naturellement, en contrôlant ce qu'elles font). C'est de ce point de vue-là que la qualité est ciblée particulièrement sur les unités de base d'Eurostat.

Nous avons commencé en créant le plan de développement d'unité (l'unité, c'est l'unité venant à la base d'Eurostat, qui s'occupe d'un certain nombre de domaines statistiques cohérents) ; l'idée, c'est : on analyse du point de vue de la qualité les processus que chaque unité est en train de développer. Cela implique une dimension qui ne se limite pas à la qualité du produit mais aux qualifications du service et du processus global, en partant des États membres. Cela veut dire faire retomber les effets de la qualité sur les Instituts statistiques pour la collecte des données (Eurostat rassemble les données des États membres). Après, il faut procéder au traitement en vue de l'harmonisation et de la satisfaction de tous les critères de qualité que l'on a cités, mais aussi, une fois que l'on a des bonnes données statistiques, les proposer au public. Et les proposer d'une façon qualitative, ce qui veut dire : transmettre un message au public, un message efficace qui peut répondre aux exigences des utilisateurs, mais pas seulement les exigences explicitées, même les exigences implicites. Une fois que toute cette phase est terminée, il faut considérer les effets : faire des enquêtes auprès des utilisateurs pour savoir s'ils sont satisfaits, et pas seulement les utilisateurs externes mais également les utilisateurs de l'Office et/ou le système statistique européen. C'est pour cette raison qu'une fois toute cette dimension analysée, Eurostat a réalisé le "corporate plan", le plan d'entreprise, en fixant des objectifs.

Le premier objectif du plan d'entreprise est la satisfaction des utilisateurs tant internes (Institutions, Eurostat) qu'externes. Puis il faut développer le système statistique européen, car, comme on l'a vu, le processus est global, et les Instituts statistiques en font aussi partie. Mais le personnel qui travaille doit également être motivé, alors il faut agir sur le personnel d'Eurostat et sur la productivité des unités d'Eurostat. Pour cela, on a plusieurs instruments : principalement, des enquêtes.
ciblées sur des utilisateurs précis (exemple : les Instituts statistiques nationaux) ou des utilisateurs généraux. Nous avons des rapports de qualité statistique que chaque responsable de domaine/processus doit développer selon les critères cités ci-dessus. Nous avons également des rapports de qualité, mais pas sur les aspects statistiques ; exemple : diffusion, dissémination de l'information, disponibilité des données... Nous avons fait une analyse coût-bénéfice en ligne avec ce qui a été dit auparavant et l'on a développé un ensemble d'indicateurs qui nous permettent de voir si les produits que l'on fait répondent aux attentes et sont de bonne qualité. Chaque mois, par exemple, dans notre site Web, on examine combien de fois les pages Web relatives à un certain argument ont été visitées pour avoir un aperçu des mouvements du marché, et on en fait le rapport aux responsables des domaines pour qu'ils en tiennent compte dans leur développement.

La base de tout ça, c'est le personnel qui a désormais assimilé l'idée de la qualité : à Eurostat, les cours de qualité sont basiques et obligatoires, ainsi que la communication à l'intérieur de l'entreprise. Pour vous donner une idée, s'il y a cinq ans, "Qualistat" était quelque chose de plutôt sombre, maintenant tout personnel d'Eurostat l'a bien présente dans sa tête et sait exactement ce que ça veut dire. »

Approche globale à l'intérieur de la Sofres

(Yannick CARRIOU)

« Je suis obligé de resituer deux ou trois éléments sur la Sofres, pour bien vous faire comprendre le pourquoi de tout cela.

On ne devrait plus parler de Sofres, la Sofres a fusionné à la fin des années 90 avec un groupe anglais, c'est donc un groupe international franco-anglais qui est aujourd'hui quatrième groupe mondial pour les études de marché. N'empêche que, en France, il y a une espèce d'aurore, de légitimité autour du mot Sofres, à la fois du côté des entreprises mais aussi du côté des particuliers, de l'opinion publique.

Ce trésor de guerre, tout le monde est conscient à la Sofres qu'il faut le préserver et c'est un peu la motivation pour mettre en place une démarche qualité.

Quelques données de flux : la Sofres réalise par an à peu près 350.000 interviews en face à face, à peu près 800.000 par téléphone, et aujourd'hui à peu près 100.000 par Internet sur un marché qui est en train d'exploser. Avec ce volume de contacts et de travail, il est indispensable de se dire que l'on va faire de la qualité comme ça de manière "éthérée". Il a donc fallu mettre en place un grillage très fin, une trame très précise, pour que chacun se sente et soit obligé d'adhérer à cette démarche.
Il y a différents niveaux de discours autour de la qualité. Il y a des discours plus commerciaux, c'est la qualité affichée à l'extérieur. Ces discours-là se détectent par une teneur en superlatifs qui est un peu supérieure à la moyenne autorisée. Et des discours fortement internalisés, que je dirai un peu plus pragmatiques.

Exemple d'un discours plus communicant : quelque chose qui est gravé en lettre d'or au-dessus de notre bureau sous le nom de la "règle 4 E" (Expertise, Engagement, Ecoute, Ethique professionnelle). L'expertise renvoie à la spécialisation des Départements : à la Sofres, on essaie de donner à chaque acteur économique, en particulier, un interlocuteur qui connaît bien son marché. L'engagement signifie que l'on définit ensemble les processus, au même titre que l'écoute : il y a une référence à l'international car, de plus en plus (et c'est le moteur essentiel de l'internationalisation de la Sofres), il n'est pas d'étude qui ne soit pas internationale, très rare, en fait, d'avoir une étude auprès d'une grande entreprise qui ne soit que française, ça fait partie des engagements qualité. Ethique professionnelle, car il peut nous arriver de dire : "on a fait quelque chose qui ne répond pas à toutes les attentes, on a peut-être insuffisamment réfléchi sur le problème au départ....", il y a une véritable honnêteté de ce marché, en tout cas des grands Instituts d'étude de marché, vis-à-vis de leurs principaux clients.

Tout cela fait un peu la transition et aboutit entre l'interne et l'externe, à un processus de certification. La SOFRES fait l'objet de visites régulières d'organismes certificateurs (AFAQ) qui contrôlent la pertinence des objectifs qualité et le strict respect des procédures mises en place pour les atteindre. Cette discipline permet à la SOFRES d'être certifiée selon la norme ISO 9001. C'est quelque chose qui s'affiche mais qui se vit au quotidien avec de fortes contraintes, sur lesquelles je reviendrai. Le plus dur, ce n'est pas d'avoir la certification mais c'est de la maintenir avec des contrôles extrêmement rigoureux mais qui sont là pour justifier que l'on a vraiment mis en place toutes les actions pour atteindre le niveau de qualité exigé.

La gestion de la qualité au quotidien passe par la mise en œuvre de procédures multiples.

Structure de l'information :
Le manuel qualité constitue le discours communicant.

Et un ensemble de procédures (22 à la Sofres) déterminent, chacune, la conduite à tenir sur tel ou tel point de l'activité, incluant des annexes qui précisent les choses etc.

Voici neuf procédures qui ont trait principalement à l'élaboration d'une étude. Il y a tout ce qui est relatif aux études quantitatives, mais aussi qualitatives, au recueil sur Internet, etc. Chaque personne qui réalise une étude sur ces chapitres est censée se conformer à un petit cahier qui fait une quinzaine de pages.
Comment cela se passe-t-il concrètement ? A chaque phase d'une étude, on dispose en général de deux documents qui sont : le document qualité, c'est-à-dire la procédure qu'il faut suivre, et le document d'enregistrement de qualité, c'est-à-dire un document où quelqu'un s'engage sur le bon suivi des préconisations de cette démarche. Donc, on a un processus qui est très segmenté et où, à chaque fois, quelqu'un signe et valide la procédure.

L'objectif est, en fin d'étude, de pouvoir assurer une certaine traçabilité de l'étude et, en cas de problème, qu'on puisse revenir à la source et voir quel est le maillon qui a flanché. La sanction de tout cela, c'est un des problèmes je crois, est une sanction donnée par le client : aujourd'hui, on envoie systématiquement une fiche de satisfaction à tous les clients, sur laquelle il est demandé si la prestation rendue est conforme ou non à ce qu'il attendait. Ensuite, on réalise une analyse transversale, on vérifie où sont les non-conformités, on remonte l'algorithme, on vérifie quels étaient les maillons qui manquaient. Donc, on utilise tout cet algorithme, tout ce système un peu lourd, pour assurer cette traçabilité et identifier les lieux et les responsables.

Le petit souci par rapport à cela, c'est de ne pas se contenter d'une sanction qui se définit par rapport au client. C'est le client qui donne les principaux signaux de l'alerte. C'est quelque chose de très fort dans la maison, les 500 personnes qui travaillent à la Sofres à Montrouge y sont sensibilisées. Il reste à une équipe plus restreinte (le directeur de la qualité surtout, le directeur scientifique aussi) de veiller au maintien d'efforts pour ce qui n'est pas directement visible par le client. »

(Michel GLAUDE)

« Cela introduit bien la dernière partie : qu'est-ce que c'est que la qualité et quelle est son approche générale dans le système statistique public ?

Content d'apprendre qu'il n'y a pas que la satisfaction du client qui compte dans l'approche qualité, je me pose la question : compte tenu du monopole de production statistique qui existe au niveau national, que serait une sanction de non-qualité pour la production statistique ?

Se pose-t-on cette question ? Quel en est l'intérêt ? On voit bien, dans une dimension internationale, émerger ce que l'on appelle le "benchmarking", c'est-à-dire, grossièrement, la comparaison des différentes méthodes utilisées dans les différents pays, les meilleures pratiques... ; on dispose donc d'un éclairage des différentes façons de faire grâce au fait qu'il y a différents pays qui sont confrontés aux mêmes problèmes. Mais je souhaiterais que l'on réfléchisse à la sanction de la non-qualité au niveau de la production statistique nationale, par exemple. »

354 INSEE Méthodes
(Alain DESROSIERES)

« Un point qui a été soulevé au cours des exposés est la distinction entre la qualité du produit - on dit aussi qualité des données - et qualité du processus. M. SAPORTA nous a bien montré, à travers ces techniques, comment la qualité du processus est une condition pour avoir une certaine stabilité des produits. Au fond, l'idée est qu'il n'y ait pas de variation non maîtrisée dans ces processus, que les roulements à bille ou les enquêtes soient faits avec une qualité constante. Ce qui est important, c'est la régularité, et l'on y passe par le biais du verrouillage du processus de production. Mais, d'un certain point de vue, l'utilisateur s'intéresse à la qualité des données. On a donc une tension entre qualité des données et qualité des processus.

J'attire l'attention sur le fait que l'on a dit que tous ces problèmes étaient liés à l'harmonisation, que, dans le cadre d'Eurostat, on s'est dit qu'il fallait harmoniser les systèmes statistiques des différents pays européens. Je vous rappelle que, dans les discussions sur l'harmonisation, on a été amené à faire une distinction à peu près analogue entre harmonisation des produits et harmonisation des méthodes.

Harmonisation des produits : cela consiste à définir les variables, ou « concepts ». Les spécialistes des domaines définissent les variables et chaque pays se débrouille pour les mesurer comme il veut, ce que font les Comptables Nationaux depuis 50 ans.

Une procédure d'harmonisation plus exigeante est l'harmonisation des méthodes, ou des processus, où l'on essaie de faire en sorte que les questionnaires soient les mêmes (les gens qui font le panel européen connaissent cela : essayer de faire passer les mêmes questions dans tous les pays européens, avec toutes les difficultés que cela pose). Il est intéressant de réfléchir au fait que l'on trouve la même dualité sur le problème de la qualité et sur le problème de l'harmonisation.

Pour revenir à la question de Michel GLAUDE, on voit bien comment on a été amené à parler de clients ; dans le cas d'une entreprise comme la Sofres, c'est tout à fait légitime. Avec les Anglais, on parle de "customer oriented" mais nous, en tant que service public, on avait l'habitude de parler d'utilisateurs, d'usagers ; dans le cadre du groupe Leg qualité, on a des discussions sur le fait que l'on insiste pour dire "user" et, eux, ils veulent mettre "customer". Alors on voit bien comment les Offices statistiques s'orientent vers une certaine dualité de leur production. D'une part, il y a certaines productions qui sont directement orientées vers des clients spécifiés et qui sont vendues ; il y a un encouragement dans certains pays (Suède, Angleterre...) pour que les produits soient commercialisés, et, dans ce cas-là, pour que cela fasse des recettes financières : on est alors exactement dans le cas de figure que nous a décrit la Sofres.

Mais il y a quand même, dans le cas du service public, d'autres canaux par lesquels les besoins des utilisateurs sont connus, je pense à des canaux du genre du CNIS, des CRIES ; on a d'autres circuits, en tout cas en France, qu'on a développés, par le biais
desquels s'expriment les besoins de la société, représentée par ses organisations, ses associations, les syndicats, etc... C'est l'originalité de notre système français de mode de représentation, un peu issue des systèmes de planification, que ce suivi ne soit pas perdu et constitue un autre canal. Il est vrai que le langage de la qualité semble aller a priori dans une autre direction, mais il faut que l'on réfléchisse à cette synthèse entre ces deux types de besoin : les besoins de service public qui ne sont pas, a priori, solvables, et les besoins de type commerciaux.

Les aspects internationaux

*Michel BLANC*

« Je vais repartir de la question du coût de la non-qualité. La perception des utilisateurs ne se traduit pas forcément à partir des critères précis qu'on a évoqués, mais c'est une perception qui peut être aussi globale, et, notamment, l'image d'un Institut statistique, son label, c'est très important ; il faut bien se rendre compte qu'autant il est difficile de le gagner, autant il est facile de le perdre et, ensuite, très difficile de remonter la pente. Je pense que c'est arrivé à certains de nos collègues dans d'autres pays.

Quand adopte-t-on ce type de démarche de management de la qualité totale ? On constate que, dans certains pays européens, cela s'est fait à l'occasion de chocs externes. Dans certains cas, il s'agit de faire face à une perte de crédibilité et, d'autres fois, c'est simplement (si l'on peut dire) parce qu'il y a des coupes budgétaires, parce qu'on demande des réductions des effectifs, et on constate qu'une première réaction de l'Institut statistique dans ces cas-là est de développer la méthodologie d'une part, la qualité d'autre part. Ce n'est pas toujours vérifié mais on a pu le remarquer plusieurs fois.

Je parlerai rapidement de deux réflexions qui ont lieu au niveau européen. L'une a été largement mentionnée ; elle se traduit par un groupe de travail qui comporte tous les pays de l'Union, et même d'autres, et qui est managé, piloté par Eurostat ; c'est ce groupe qui a défini ces critères de qualité, et qui a pour but de définir les méthodologies à appliquer pour les rapports qualité qui doivent être élaborés conformément aux règlements dont on a parlé.

Ensuite, il y a une réflexion qui passe par le canal du "LEG". Un LEG, c'est un "leadership group", c'est un groupe d'un nombre restreint de pays et animé, non pas par Eurostat, mais par un pays membre, en l'occurrence la Suède et en l'occurrence quelqu'un qu'on a cité, Lars LYBERG, qui est un spécialiste de qualité. Il anime donc un groupe de quelques pays dont la France fait partie, et auquel j'appartiens, et qui a pour but notamment de donner des recommandations sur l'application de ces modèles à la statistique.
Simplement, une incursion dans les méthodes pour dire que le mandat initial de ce groupe comportait l'objectif suivant : il faut développer, au niveau européen, des "current best methods", mot magique dans les TQM (total quality management) statistiques, les « meilleures méthodes actuelles » qu'il faut imposer, en quelque sorte, à ses collègues, car bien sûr on peut toujours faire mieux et on a des méthodes à leur proposer. La discussion au sein du LEG a montré que les différences culturelles ou institutionnelles et les pratiques des pays sont telles que l'on a renoncé, et heureusement, à imposer ce genre de chose, mais, par exemple, on va dire : chaque pays peut les développer pour lui-même, donc il faut montrer comment on développe de telles bonnes pratiques, etc. »

(Michel GLAUDE)

« Juste deux mots sur cette non-qualité, pour prendre un exemple, en espérant ne pas dévoiler de secret. Au Comité de Direction de l’Insee ce matin, une des questions du Directeur général était précisément : il semble qu’actuellement, dans la presse, il y ait une contestation de la qualité des chiffres de l’Insee concernant la mesure de la croissance et de la productivité. Qu’allons-nous mettre en place pour essayer de répondre à cette question, c'est-à-dire existe-t-il des travaux ? Sinon, comment faut-il les conduire ? Comment faut-il les communiquer ? Par quelle méthodologie ?...

La question de la contestation est très importante car elle peut entraîner une perte de crédibilité. On a tous l’exemple anglais présent à l’esprit sur les questions de salaires (il y a un petit article dans le "Courrier des statistiques" très intéressant à ce sujet) ; sur des problèmes de crédibilité, il est fondamental de se demander : "Qu’est-ce qu’on met en œuvre pour démontrer la qualité de nos produits. »

Intervention de l'assistance

(Michel DUÉE - DR d'Alsace)

« Vous avez posé la question : que pourrait être une démarche qualité à l'Insee dans un service public ?

En région, on est très souvent confronté aux clients, on a donc une assez bonne connaissance de ce qu'ils demandent.

Un axe de qualité, d'amélioration, que je pourrai vous proposer pour l'Insee, serait le respect des délais, car très souvent la Direction générale nous annonce des dates que nous retransmettons à nos clients et quand ça dérape de six mois ou un an, c'est ensuite à nous de gérer la crise. Donc je pense que, de ce côté-là, il y aurait une amélioration à apporter.
D'autre part, vous avez parlé des canaux pour l'aspect service public ; il y a aussi les Directions régionales qui se plaignent régulièrement à propos des utilisations des enquêtes qui sont réalisées par l'Insee, qui généralement ne sont pas utilisables au niveau régional, et l'on a exactement le même problème pour un certain nombre de fichiers, je pense notamment aux fichiers entreprises : un certain nombre de redressements sont faits pour les grosses entreprises pour qu'au niveau national les fichiers soient propres, mais quand les régionaux essaient d'utiliser ces mêmes fichiers, les redressements qui sont faits ne sont pas suffisants et, au niveau local, on est toujours très embêté. Donc, je pense que ce serait là un deuxième axe de progrès qui serait possible. »

(Michel GLAUCDE)

« Je partage entièrement votre premier point de vue : on peut dire que l'Insee gère ses processus par les délais, ce qui est une mauvaise gestion, mais c'est un constat que l'on a fait de manière très générale. En effet, si l'on regarde comment l'institut fait face à ses difficultés, c'est en faisant déraper les délais, ce qui est regrettable. En revanche, sur le deuxième point, je pense que la période récente montre que l'on a fait (et cela se poursuivra) des efforts très importants pour donner des produits utilisables au niveau local et régional… »

(Jean-Claude DEVILLE)

« On a parlé de définition de la qualité, moi je veux bien ; on a parlé un tout petit peu, mais pas assez, de façons de l'améliorer et c'est là où je trouve que ça commence à coincer un petit peu. Par exemple, diffuser les "best practices", je veux bien, mais :

1. il faut les connaître ;
2. il faut trouver des gens capables de les mettre en œuvre, or ce n'est pas toujours très facile et l'on peut se poser des questions.

J'ai fait un pointage rapide sur l'ensemble des présentations de ces Journées de Méthodologie, je suis persuadé qu'une bonne moitié d'entre elles ont été faites par des personnes qui parlaient au titre de travaux qu'ils avaient faits dans leur ancien poste, et l'on peut se demander si ce changement de qualification, de qualité, est une politique optimale pour obtenir des travaux ayant une certaine qualité. »
Conclusion

(Michel GLAUDE)

« Que peut-on tirer de ces Journées de Méthodologie Statistique ? Ce qui frappe tout d’abord, c’est le nombre, la qualité des travaux méthodologiques présentés.

Cela témoigne d’abord d’une grande vitalité de la réflexion méthodologique, et je tiens à le saluer ; et aussi d’une grande diversité de ces réflexions méthodologiques couvrant un vaste ensemble de domaines, d’approches etc…

Personnellement, j’ai trouvé qu’il y avait un élément qui ressortait fortement et qui me semble aussi intéressant pour aller dans le sens de la qualité, c’est le fait de travailler en équipe, et je trouve cela tout à fait important, c’est-à-dire qu’on a l’impression que la méthodologie ne se fait pas toute seule (les statistiques non plus en général), et les équipes, les réseaux de relation me semblent forts et efficaces.

On peut aussi percevoir qu’il y a une demande de formation permanente très forte. Chacun est confronté à une diversité d’expériences dans d’autres domaines, on connaît mal ces autres domaines, c’est une ouverture d’esprit très grande, qui appelle la formation.

Je signale à ce propos que les Journées de Méthodologie continuent sous forme de séminaire de méthodologie régionale demain. Il y aura donc aussi partage d’expérience au niveau des méthodologies régionales.

J’ai également beaucoup apprécié l’ouverture internationale dans le sens des "best practices" dont on a parlé.

Je remercie bien sûr tous nos orateurs et collaborateurs étrangers qui ont participé à ces journées et j’ai l’impression que nous allons bénéficier de ces apports. Je souhaiterais que nos méthodologies français sachent aussi exporter leurs avancées dans d’autres pays (je trouve dommage qu’ils ne le fassent pas assez), dans le cadre de la coopération technique certainement, mais aussi dans le cadre des échanges internationaux.

Ma dernière remarque concerne la modestie qui doit exister par rapport à ces travaux. Modestie, car on ne fait pas de la méthodologie pour la méthodologie, on fait de la méthodologie pour un meilleur service auprès des utilisateurs, et il est important de garder cela en mémoire.

Je conclurai en remerciant tout le monde et d’abord les organisateurs : l’UMS, le secrétariat de l’UMS, la Communication Externe de l’Insee qui a aussi beaucoup contribué, l’atelier d’impression sans lequel vous n’auriez pas tous ces beaux documents, en temps et heure voulus, et le Ministère qui nous a accueillis.

INSEE Méthodes
Je rappelle que nous produirons les actes du colloque à partir des versions définitives des papiers, que les contributeurs sont donc invités à donner prochainement.

Enfin je finirai sur deux points :

1. Le rendez-vous suivant sera vraisemblablement en 2002, et je suis content que les Journées de Méthodologie Statistique continuent.

2. Il y avait, comme vous l'avez peut-être perçu, passage de témoin, transmission de "bébé" entre une équipe qui avait organisé les Journées de Méthodologie Statistique précédemment, et qui avait consacré beaucoup d'énergie, d'affection à cette organisation, et une nouvelle équipe qui les a prises en charge. Tout ça, c'est l'UMS, mais l'UMS bouge et change aussi ; eh bien, il semble que cette transmission de témoin s'est bien passée et que le "bébé" grandit dans des mains différentes. En tout cas, la paternité précédente n'a pas à rougir du projet tel qu'il a existé au cours de ces Journées. Je crois que c'est une belle leçon sur la façon de transmettre et de poursuivre un objectif.

Je vous remercie tous, j'ai été très content de l'organisation de ces Journées et encore une fois bravo. »