

CHAPITRE B

LE NOUVEAU MODE DE RECENSEMENT - ÉLABORATION A TRAVERS LES TESTS

Le titre V (« Des opérations de recensement ») de la loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité confie à la commune la préparation et la réalisation des enquêtes de recensement, tandis que l'Institut national de la statistique et des études économiques (Insee) a pour mission d'organiser et de contrôler la collecte des informations.

Les enquêtes de recensement sont elles-mêmes réalisées selon des modalités différentes selon que les communes ont moins de 10 000 habitants ou, au contraire, atteignent ou dépassent cette taille³.

C'est pourquoi ce chapitre distingue les communes selon leur position, en termes de population, par rapport au seuil de 10 000 habitants voulu par le législateur.

B.1 La collecte de l'information dans les communes de moins de 10 000 habitants

Les communes de moins de 10 000 habitants sont désormais recensées une fois tous les cinq ans, par roulement. A cet effet, elles ont été réparties en cinq groupes constitués selon des règles précises, qui assurent le même effectif et la même structure démographiques à chaque groupe (voir [§ B.3.1](#)). Chaque année, la totalité de la population et des logements de toutes les communes d'un même groupe est recensée, si bien que, au bout de cinq ans, l'ensemble des communes de moins de 10 000 habitants du pays⁴ est recensé. La sixième année, le tour des communes recensées la première année revient, et ainsi de suite.

Jetons un coup d'œil à la carte qui suit. Elle représente la région de Lorraine. Chaque commune de moins de 10 000 habitants appartient à l'un des cinq groupes. Chaque groupe a été identifié par une couleur distincte. Les communes de 10 000 habitants ou plus ne sont pas classées dans ces groupes ; elles figurent en gris sur la carte.

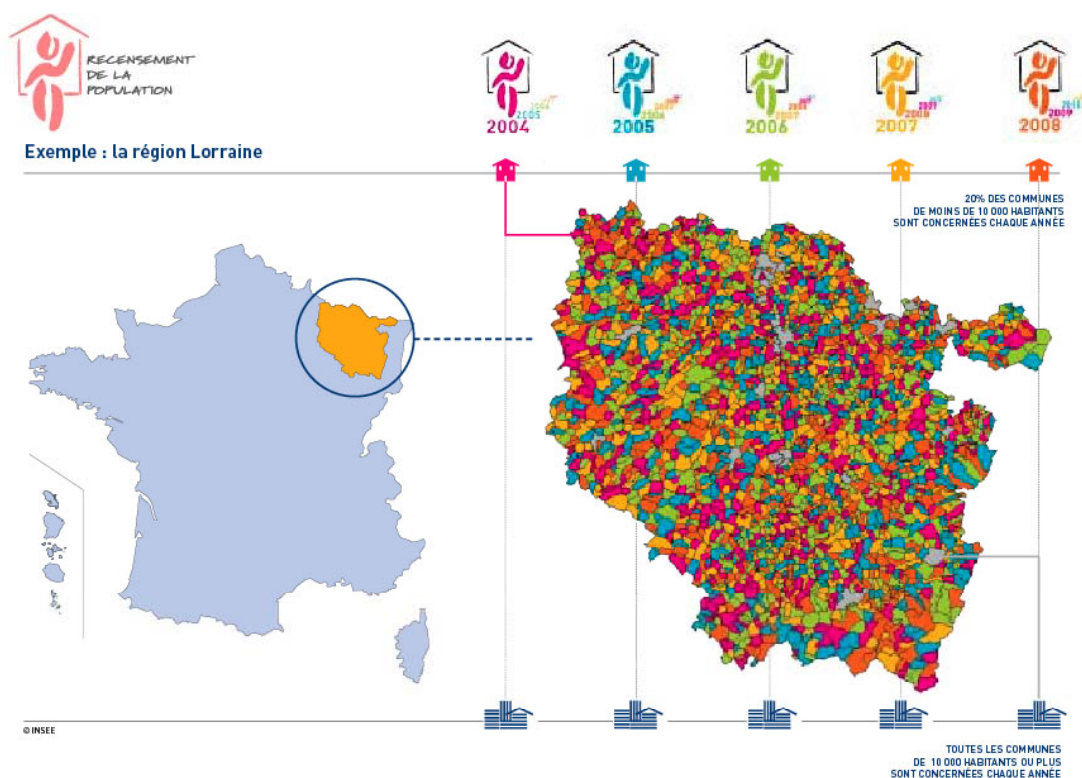
En 2004, tous les habitants et tous les logements des communes du groupe "rose" ont été exhaustivement recensés. Ceux des communes du groupe "bleu" l'ont été en 2005, etc. Les

³ Il s'agit de la population sans doubles comptes issue du recensement général de la population de 1999 puis, à partir du premier décret d'authentification qui suit, de la population municipale. Cf. [Article 20 du décret n° 2003-485 du 5 juin 2003](#).

⁴ Très précisément, il s'agit des communes de moins de 10 000 habitants de métropole, des départements d'outre-mer (Guadeloupe, Guyane, Martinique et La Réunion) et de Saint-Pierre-et-Miquelon.

communes recensées en 2004 le seront à nouveau en 2009, puis en 2014, en 2019, etc. Celles du groupe 2005 seront à nouveau recensées en 2010, 2015, 2020, etc. Et le mécanisme est le même dans chacune des régions. On notera, au passage, qu'une personne vivant dans l'une de ces communes sera recensée très régulièrement une fois tous les cinq ans aussi longtemps qu'elle ne déménagera pas.

On voit, par ailleurs, que chaque groupe de communes de moins de 10 000 habitants est dispersé sur l'ensemble du territoire de la région. Cette dispersion a une raison précise : elle permet d'obtenir que chacun des groupes constitue un échantillon représentatif de l'ensemble des communes de moins de 10 000 habitants de la région.⁵



Combien de personnes sont-elles recensées, de la sorte, chaque année ? Il se trouve que la moitié environ de la population, soit 30 millions de personnes, vivent dans les communes de moins de 10 000 habitants. Il en est donc recensé un cinquième chaque année, ce qui équivaut à 6 millions de personnes, ou encore 10% de la population vivant en France.

⁵ L'intérêt de cette propriété sera développé et défini plus loin dans ce chapitre (voir [§ B.3.1](#)) et au [chapitre D](#).

B.2 La collecte de l'information dans les communes de 10 000 habitants ou plus

Les communes de 10 000 habitants ou plus procèdent d'une tout autre manière puisqu'elles réalisent chaque année une enquête de recensement auprès d'une partie de leur population. La collecte annuelle porte sur un échantillon d'adresses⁶ tirées au hasard et représentant environ 8 % des logements de la commune, donc environ 8% de sa population. Toutes les personnes vivant dans les logements des adresses ainsi tirées au sort sont recensées.

Le territoire de chacune de ces communes doit être pris en compte en totalité au cours d'une période de cinq ans. Autrement dit, au cours de cinq enquêtes de recensement successives, chacune des adresses de la commune devra avoir été prise en considération dans la confection des échantillons⁷. Au bout de cinq ans, 40 % des logements de la commune auront donc été recensés.

Dans la pratique, comment s'y prend-on ?

La constitution des échantillons annuels d'adresses à enquêter repose, dans les communes de métropole⁸, sur le répertoire d'immeubles localisés (le RIL : [voir l'encadré](#)). Ce répertoire est la liste des édifices résidentiels, institutionnels ou commerciaux de la commune, localisés individuellement par leurs coordonnées géographiques de façon à pouvoir créer une cartographie numérisée, c'est-à-dire une cartographie sur support informatique.

A partir du RIL, on a réparti au hasard, dans chaque commune, les adresses d'habitation en cinq groupes comptant le même nombre total de logements et respectant certaines conditions de représentativité (voir plus bas, [§ B.3.2](#)). Chaque groupe rassemble donc 20% des logements de la commune. Chaque année, tour à tour, un échantillon d'adresses est tiré dans l'un des groupes. La taille de l'échantillon est de 40 % des logements de ce groupe. Comme le groupe représente un cinquième des logements de la commune, l'échantillon compte $40\%:5=8\%$ des logements de la commune. Au bout de cinq ans, on aura ainsi recensé $8\% \times 5 = 40\%$ des logements de la commune, donc 40% environ de sa population.

Le schéma présenté plus loin permet de mieux comprendre comment les choses se passent.

⁶ Il s'agit de l'adresse au sens ordinaire du terme : un édifice d'habitation repéré par un numéro (éventuellement assorti d'un indice de répétition) dans une voie portant un nom.

⁷ L'utilité de procéder ainsi sera développée plus loin dans ce chapitre (voir [§ B.3.2](#)) et au [chapitre D](#).

⁸ Le cas des communes de 10 000 habitants ou plus des départements d'outre-mer est traité à la fin du [§ B.3.2](#), question 9.

Encadré 1 : Le répertoire d'immeubles localisés (RIL)

Le répertoire d'immeubles localisés (RIL) (communes de 10 000 habitants ou plus de métropole)

*Le présent encadré donne une information sommaire sur le RIL.
Une information plus complète figure en [annexe B.3](#)*

Qu'est-ce que le RIL ?

Le RIL est un répertoire d'adresses localisées associé à une base communale de données géographiques.

Le répertoire contient et gère les informations suivantes :

- pour chaque adresse : les coordonnées géographiques, le type et le nom de la voie, le numéro dans la voie ;
- pour chaque ensemble immobilier implanté à une adresse : le type (immeuble d'habitation, d'activité, d'équipement urbain, ou mixte), la date de construction, la date d'entrée dans le répertoire ou la date de dernière modification ou de destruction, le nombre d'étages, le nombre de logements, le nombre d'établissements ou de communautés qui y sont installés et le type d'équipement urbain.

Le champ du RIL couvre toutes les communes de France métropolitaine ayant atteint 10 000 habitants au recensement de 1999. Cela représente environ 900 communes.

Statut juridique du RIL

La constitution et la mise à jour du RIL par l'Insee ont été autorisées par un arrêté du 19 juillet 2000 du ministre de l'économie, des finances et de l'industrie pris après avis de la Commission nationale de l'informatique et des libertés (CNIL).

Pour les besoins du recensement de la population et suivant les dispositions de la loi du 27 février 2002 relative à la démocratie de proximité, l'arrêté du 19 juillet 2000 a été modifié par l'arrêté du 9 octobre 2002 afin d'autoriser la communication aux communes ou aux établissements publics de coopération intercommunale des informations du RIL qui les concernent.

Constitution et mise à jour du RIL

Le RIL a été constitué par géocodage des adresses des immeubles d'habitation et des communautés recensées en 1999 ainsi que de celles des établissements immatriculés au répertoire Sirene (Système informatisé du répertoire des entreprises et des établissements).

Le répertoire est mis à jour en continu pour identifier et localiser les adresses nouvelles et les adresses détruites. Cette mise à jour se fait à partir :

- des informations communiquées par les communes concernant les modifications de la voirie et des adresses ;
- de fichiers administratifs : permis de construire, adresses de La Poste, taxe d'habitation ;
- des informations communiquées par les communes dans le cas de différences constatées entre les diverses sources d'actualisation ; les communes sont consultées sur la base de listes d'adresses à confirmer ;
- des constatations faites sur le terrain pendant la réalisation des enquêtes de recensement ;
- de vérifications sur le terrain, le cas échéant.

L'expertise du RIL par la commune

L'expertise du répertoire est une opération préliminaire aux enquêtes de recensement ; elle est destinée à donner à celles-ci le meilleur niveau de qualité possible.

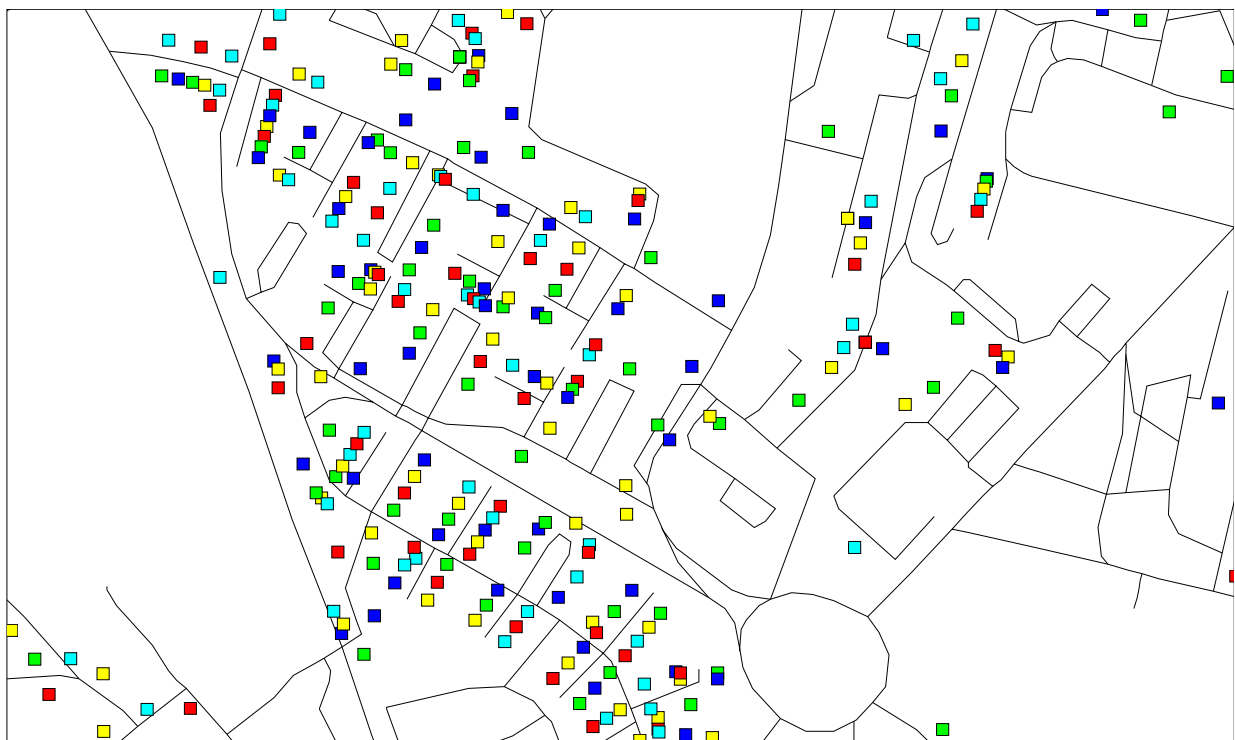
L'expertise intervient après intégration des informations fournies en continu à l'Insee par la commune (arrêtés municipaux, plans, extraits du système d'information géographique – SIG – communal), des informations venant de sources externes et des réponses de la commune aux listes d'adresses à confirmer établies par l'Insee suite à l'utilisation des diverses sources de mise à jour.

Elle permet à l'Insee d'utiliser une base de sondage aussi à jour que possible pour échantillonner les adresses qui vont être enquêtées dans la commune en janvier et février suivants.

Une date butoir (30 juin de l'année n), est fixée avant laquelle les remarques de la commune peuvent être traitées dans le cadre de l'édition "n" du RIL qui sera utilisé pour le tirage de l'échantillon de l'enquête de recensement de l'année n+1. Au-delà de cette date, les remarques seront traitées dans l'édition n+1 du RIL.

La mise à jour du RIL en continu et l'opération d'expertise offrent matière à concertation entre la commune et l'Insee sur la qualité du répertoire.

Les groupes d'adresses (communes de 10 000 habitants ou plus)



Ce schéma représente les voies d'un quartier quelconque d'une commune de grande taille. Les petits carrés représentent les adresses d'habitation le long des voies. Chaque adresse a été affectée à un groupe identifié par sa couleur. Il y a cinq groupes au total dans la commune et tous ont le même nombre de logements. Ces groupes sont stables dans le

temps et sont mis à jour uniquement en fonction des créations et des destructions d'adresses. Une adresse qui a été placée dans un groupe y demeure, du moins jusqu'à sa disparition.

On décide, par exemple, que l'échantillon des adresses recensées en 2005 sera tiré dans le groupe "rouge" et celui des adresses recensées en 2006 dans le groupe "jaune". Dans ces conditions, le groupe d'adresses "rouge" fera l'objet d'un nouveau tirage d'échantillon en 2010, puis 2015, puis 2020, etc., tandis que le groupe "jaune" sera concerné en 2011, puis 2016, puis 2021, etc.

Comme on le voit, une personne vivant dans une commune d'au moins 10 000 habitants et ne déménageant pas a une chance sur 2,5 (40 chances sur 100) d'être recensée au cours d'une période de cinq ans. Si elle est recensée en 2006, elle n'a aucune chance d'être recensée avant 2011 puisque son adresse ne sera pas remise dans l'urne du tirage au sort avant cette date. De cette manière, comme dans les communes de moins de 10 000 habitants, un même logement ne sera pas enquêté plus d'une fois dans un intervalle de cinq ans.

Puisque 30 millions de personnes vivent en France dans des communes ayant au moins 10 000 habitants, il en est donc recensé environ 2,5 millions chaque année dans ces communes.

Au total, toutes catégories de communes réunies, chaque vague annuelle d'enquêtes de recensement portera sur 8,5 millions de personnes (6 dans les communes de moins de 10 000 habitants et 2,5 dans les autres communes), ce qui représente environ 14% de la population française.

B.3 Le choix des personnes et des logements enquêtés lors d'une enquête de recensement

Comme on l'a vu, une enquête de recensement a lieu chaque année. Elle concerne toute la population vivant dans un cinquième des communes de moins de 10 000 habitants et environ 8% de la population de chacune des communes de 10 000 habitants ou plus.

Pourquoi ces choix ont-ils été faits et comment se traduisent-ils dans la pratique ?

La réponse à la première question résulte des considérations suivantes :

- d'une part, il s'agissait de produire une information plus régulière et en moyenne plus récente, sans pour autant accroître la dépense publique ; l'étalement dans le temps des collectes d'information et la technique des échantillonnages répondaient à cette préoccupation ; la proposition gouvernementale d'une durée de cinq ans pour prendre en compte la totalité du territoire est apparue raisonnable au législateur, compte tenu du rythme des changements des structures démographiques et sociales de la population ;
- d'autre part, il fallait tenir compte de l'extrême dispersion de la taille des communes⁹, qui interdisait de généraliser la technique des enquêtes de recensement sur échantillon. L'expérience montre qu'il faut atteindre une taille de 10 000 habitants pour passer avec une bonne sûreté de l'observation exhaustive à l'observation sur échantillon ; c'est le seuil que le législateur a retenu.

La manière dont ces choix se traduisent dans la pratique est donc très différente de part et d'autre du seuil de 10 000 habitants. Pourtant, une même logique est à l'œuvre dans les deux univers. En effet, comme on le verra au [chapitre D](#), les informations collectées lors de

⁹ Sur un total de près de 36 600 communes, on compte en France métropolitaine 28 000 communes de moins de 1 000 habitants et environ 6 800 communes de 1 000 à 4 999 habitants.

cinq enquêtes de recensement successives servent à produire la population de chaque circonscription administrative du pays et des statistiques détaillées à tous les niveaux d'organisation du territoire, y compris aux niveaux communal et infra-communal. Les résultats de l'enquête de recensement d'une seule année, portant donc sur un seul groupe (de communes ou d'adresses), servent quant à elles à produire des données statistiques nationales et régionales, valables pour cette même année.

Cette double exigence conduit à définir les critères de représentativité auxquels doit souscrire chacun des cinq groupes (de communes ou d'adresses).

En effet, les estimations nationales et régionales seront d'autant plus fiables que les groupes seront individuellement l'image exacte de l'ensemble dont ils sont extraits. L'argument vaut aussi pour les estimations détaillées qui prennent en compte une enquête dans chacun des cinq groupes : c'est un enseignement fort de la théorie statistique que les estimations construites par l'agrégation de plusieurs parties sont d'autant plus précises que ces parties sont homogènes.

Regardons de plus près ce qu'il en est.

B.3.1 Les groupes de rotation¹⁰ de communes de moins de 10 000 habitants

Dans toutes les régions, les cinq groupes de communes de moins de 10 000 habitants ont été constitués pour être, chacun, l'image aussi fidèle que possible de l'ensemble des communes de moins de 10 000 habitants de sa région.

La méthode statistique utilisée est celle des échantillons équilibrés¹¹. Généralisant la notion de stratification, elle consiste à choisir des structures de référence et à construire des échantillons reproduisant, le plus fidèlement possible, ces structures.

Dans le cas présent, les structures de référence étaient à choisir parmi les variables démographiques et les catégories de logements. Les valeurs cibles ont été établies à partir des résultats du recensement général de la population de mars 1999. En d'autres termes, on a fait l'hypothèse qu'un ensemble de communes dont la population avait par exemple, en 1999, une structure par âge identique à celle de l'ensemble des communes de moins de 10 000 habitants conserverait, au moins pendant un certain temps, une bonne qualité de représentativité sur ce critère.

Les critères retenus ont été les suivants :

- le nombre de logements ;
- le nombre de logements en immeuble collectif ;
- la population des personnes de moins de 20 ans ;
- la population des personnes de 20 à 39 ans ;
- la population des personnes de 40 à 59 ans ;
- la population des personnes de 60 à 74 ans ;
- la population des personnes de 75 ans ou plus ;
- la population des femmes ;
- la population des hommes ;

et, pour chacun des départements, la population totale.

A ce stade, quelques questions subsistent :

Question 1 : *Pour une représentation plus fine du territoire, ne pouvait-on réaliser des groupes équilibrés par sexe, âge, etc. au niveau des départements ?*

¹⁰ Sur la notion de groupe de rotation, voir l'[encadré](#).

¹¹ Voir l'[annexe B.1](#)

Réponse : Pour pouvoir le faire, compte tenu de la grande variété de taille des communes (de 1 à 9 999 habitants), il faudrait que tous les départements aient un nombre suffisant de communes à répartir. Pour faire image, c'est comme si les communes étaient des billes d'inégale grosseur à répartir entre cinq balances de manière que toutes les balances affichent le même poids. Si l'on veut un équilibrage précis, l'ensemble des billes doit être suffisamment fourni pour permettre des dosages fins. C'est loin d'être le cas pour nombre de départements, dans lesquels il y a relativement peu de communes se situant au-dessous du seuil de 10 000 habitants, d'où le choix de la région comme territoire d'équilibrage complet. On s'assure cependant que les départements sont représentés dans chaque groupe proportionnellement à leur nombre d'habitants.

Question 2 : *Comment ont été choisies les variables d'équilibrage ?*

Réponse : L'expérience des recensements précédents et celle de nombreuses enquêtes montrent que la plupart des renseignements individuels recueillis dans le recensement sont en bonne corrélation avec l'âge, le sexe et le type d'habitat. On est donc en droit d'espérer que toutes les variables seront bien représentées.

Question 3 : *Au bout de cinq ans, doit-on « rebattre les cartes », c'est-à-dire faire de nouveaux groupes ?*

Réponse : Non, les communes enquêtées en 2004 seront enquêtées à nouveau en 2009 et ainsi de suite. On a vérifié que les groupes représentatifs de leur région d'après les résultats du recensement de 1999 étaient également représentatifs d'après les résultats de celui de 1990. Les grandes structures n'évoluent que lentement. Cela dit, la qualité de cette partition en cinq groupes sera régulièrement appréciée et si, à long terme, la nécessité de rééquilibrer les groupes se fait sentir, on aura tous les éléments pour le faire puisque toutes ces communes auront été recensées moins de cinq ans plus tôt.

Question 4 : *Au cours du temps, il arrive que des communes fusionnent ou se séparent. D'autre part, certaines communes franchiront, à la hausse ou à la baisse, la barre des 10 000 habitants. Que fera-t-on alors ?*

Réponse : Annuellement, de tels événements ne sont pas nombreux. Ils seront traités au fil du temps en tenant compte des disparitions de communes et en affectant au mieux chaque nouvelle commune à un groupe, dans le respect des équilibres. De la même manière, les franchissements de seuil (qui sont régis en droit par l'article 29 du décret n° 2003-485 du 5 juin 2003) conduiront à revoir à la marge la composition des cinq groupes de rotation.

Question 5 : *La méthode est-elle la même pour les communes des départements d'outre-mer ?*

Réponse : C'est la même méthode. Toutefois, le faible nombre de communes de moins de 10 000 habitants dans ces départements ne permet pas de les répartir en cinq groupes aussi strictement équilibrés sur les mêmes variables qu'en métropole. La solution apportée est pragmatique et fondée sur un seul critère d'harmonisation des groupes : leur nombre d'habitants constaté en 1999.

Encadré 2 : La notion de groupe de rotation

La notion de groupe de rotation

Pourquoi l'expression « groupe de rotation » ?

La notion de groupes de rotation découle du rythme quinquennal indiqué dans la loi. Dans les communes de moins de 10 000 habitants « les enquêtes sont exhaustives et ont lieu chaque année par roulement au cours d'une période de cinq ans ». Dans les autres communes, « une enquête par sondage est effectuée chaque année ; la totalité du territoire de ces communes est prise en compte au terme de la même période de cinq ans ».

Même si cela n'est pas précisé par la loi, il est naturel de concevoir que, dans les communes de 10 000 habitants ou plus, les échantillons des enquêtes annuelles de recensement portent sur des unités différentes pendant la période de cinq ans dont il s'agit : les adresses d'une année ne peuvent être à nouveau échantillonnées que cinq ans plus tard.

Le territoire national est donc réparti en cinq ensembles, chaque ensemble faisant l'objet d'une enquête tous les cinq ans par roulement. Ou bien il s'agit de cinq ensembles de communes de moins de 10 000 habitants faisant l'objet, tour à tour, d'une enquête de recensement exhaustive ; ou bien il s'agit, dans chaque commune de 10 000 habitants ou plus, de cinq ensembles d'adresses dans lesquels sont tirées au sort, tour à tour, les adresses constituant l'échantillon de l'enquête de recensement de l'année.

Ces ensembles (de communes ou d'adresses) sont appelés **groupes de rotation**.

L'ensemble enquêté une année donnée est ainsi constitué, d'une part, d'un groupe de communes de moins de 10 000 habitants et, d'autre part, d'un groupe d'adresses tirées au sort dans chacune des autres communes.

Dans toutes les régions, les groupes de communes de moins de 10 000 habitants sont l'image aussi fidèle que possible de l'ensemble des communes de moins de 10 000 habitants de leur région en termes de population par âge, par sexe, par type de logement (individuel ou collectif).

Dans les communes de 10 000 habitants ou plus, les enquêtes par sondage portent sur des groupes d'adresses qui respectent des critères semblables au niveau de la commune pour la population et les logements.

B.3.2 Les groupes de rotation d'adresses dans les communes de 10 000 habitants ou plus

Dans les communes de 10 000 habitants ou plus, l'échantillon de chaque enquête annuelle de recensement est constitué d'adresses d'habitation, tous les logements d'une adresse échantillonnée étant recensés de façon exhaustive.

Pour éviter d'interroger souvent une même personne, on a réparti les adresses d'habitation de chaque commune en cinq groupes, qui sont échantillonnés à tour de rôle.

Le plan de répartition des adresses et de tirage des échantillons est conçu selon trois principes :

- tenir compte de la variabilité des adresses selon le nombre de logements ;
- s'adapter aux modifications annuelles, à savoir les suppressions et créations d'adresses ;
- garantir une bonne représentativité communale et infracommunale des groupes et des échantillons annuels d'adresses pour les logements et la population.

C'est pourquoi on classe les adresses en trois catégories (ou strates) : les adresses de grande taille, les adresses nouvelles, les autres adresses :

- les adresses de grande taille d'une commune sont celles qui, à la fois, sont les plus grandes en nombre de logements, ont au moins 60 logements chacune et qui, réunies, cumulent au maximum 10% des logements de la commune¹². Elles constituent la première catégorie. Elles sont réparties en cinq groupes pour être enquêtées exhaustivement en cinq ans ;
- les adresses nouvelles sont les adresses qui apparaissent année après année. Celles qui sont de grande taille rejoignent la catégorie précédente. Les autres constituent la deuxième catégorie ;
- les autres adresses d'habitation regroupent toutes les autres adresses où il existe au moins un logement. Elles constituent la troisième catégorie. Les adresses nouvelles qui forment la deuxième catégorie sont réparties entre les groupes de la troisième catégorie de manière à être toutes recensées en cinq ans.

Pour les mêmes raisons qui ont conduit à équilibrer les cinq groupes de communes de moins de 10 000 habitants¹³, mais cette fois en raisonnant au niveau de la commune et non plus au niveau de la région, chaque groupe d'adresses doit être aussi représentatif que possible de l'ensemble de sa catégorie dans la commune et chaque échantillon d'adresses doit être aussi représentatif que possible de l'ensemble des adresses de la commune. Les critères retenus sont les mêmes que pour l'équilibrage des groupes de communes de moins de 10 000 habitants : nombre de logements individuels¹⁴, nombre de logements collectifs, population par sexe et âge (la référence étant le recensement de 1999 à l'initialisation du répertoire).

Chaque année, l'observation du parc des logements est menée par l'Insee en concertation étroite avec la commune. Les immeubles détruits sont naturellement enlevés de la base d'adresses, les constructions nouvelles y sont introduites avec leur nombre de logements. Les groupes d'adresses ayant été mis à jour, on sélectionne un échantillon aléatoire dans le groupe de l'année, avec comme critères d'équilibrage le nombre de logements total, le nombre de logements collectifs, le poids actuel des IRIS-2000®¹⁵ en nombre de logements. Ce dernier critère est là pour s'assurer que l'échantillon est bien réparti sur tout le territoire de la commune (voir l'[encadré](#)).

Le taux de sondage d'une année est ajusté de manière que la proportion de logements enquêtés cette année-là dans la commune, toutes adresses confondues, soit aussi proche que possible de 8 %¹⁶.

¹² Cette définition a été retenue au terme de multiples simulations visant l'optimisation du plan de sondage.

¹³ Voir le [§ B.3.1](#)

¹⁴ Dans les communes de 10 000 habitants ou plus, le recensement de 1999 a montré qu'il y a en moyenne 30% de logements individuels.

¹⁵ Les IRIS-2000® (îlots regroupés pour l'information statistique) constituent une partition en quartiers d'environ 2 000 habitants du territoire des communes de 5 000 habitants ou plus.

¹⁶ La méthode de tirage des échantillons des enquêtes annuelles de recensement dans les communes de 10 000 habitants ou plus est exposée plus en détail en [annexe B.2](#).

Encadré 3 : Les IRIS-2000

Les IRIS-2000®

(Marque déposée)

Les **IRIS-2000®** (îlots regroupés pour l'information statistique) constituent une partition en quartiers d'environ 2 000 habitants du territoire des communes de 5 000 habitants ou plus. Ils ont été définis en 1998 et 1999, en concertation avec les communes intéressées, pour les besoins de la diffusion des résultats du recensement général de la population de 1999. Le découpage s'est appuyé sur les **îlots** au sens du recensement.

L'îlot est l'unité géographique de base :

- en zone bâtie dense, il représente le plus souvent un pâté de maison, éventuellement scindé en cas de limite communale ou cantonale traversant le pâté de maison (ou petit groupe de pâtés de maison) ;

- en zone " périphérique ", l'îlot est un ensemble limité par des voies (ou autres limites visibles) découpant cette zone en plusieurs morceaux (plus finement en 1999 qu'en 1990).

Les îlots peuvent être vides d'habitants (par exemple une gare). Les îlots ont été définis par l'Insee en concertation avec les communes. Il existait un découpage en îlots à chaque recensement. Pour le découpage 1999, il y a une table de passage 90-99.

On appelle îlots 99 les îlots définis par l'Insee pour le recensement de 1999.

Les IRIS sont les briques de base en matière de diffusion de données locales. Leur taille en nombre d'habitants garantit à la fois la robustesse des statistiques produites à ce niveau et la protection de l'anonymat des personnes. La France en compte environ 50 800 (50 100 en métropole et 700 dans les DOM). L'IRIS correspond à :

- l'IRIS-2000 pour toutes les communes urbaines d'au moins 10 000 habitants et la plupart des communes de 5 000 à 10 000 habitants (16 000 IRIS-2000 en France, dont 15 400 en métropole) ;

- la commune pour les communes non découpées (34 800 communes) ; il s'agit pour l'essentiel des communes de moins de 5 000 habitants.

Chaque IRIS-2000 forme un "petit quartier" qui se définit comme un ensemble d'îlots contigus.

Les IRIS-2000 se déclinent en trois types de zones :

- IRIS d'habitat : IRIS-2000 dont la population se situe entre 1 800 et 5 000 habitants ; ils sont homogènes quant au type d'habitat ;

- IRIS d'activité : IRIS-2000 qui regroupent plus de 1 000 salariés et comptent deux fois plus d'emplois salariés que de population résidente ;

- IRIS divers : IRIS-2000 de superficie importante à usage particulier (bois, parcs, zones portuaires...).

Le découpage de 1998-1999 en IRIS-2000 a été soumis à l'approbation de la Commission nationale de l'informatique et des libertés (CNIL), notamment en ce qui concerne les IRIS comptant moins de 2 000 habitants. La Cnil a accepté qu'un toilettage du découpage en IRIS intervienne en 2006 ou 2007 dans les communes où les informations issues des premières enquêtes de recensement montreront que le découpage de 1998-1999 n'est plus pertinent. Cela permettra de produire dans un découpage actualisé les résultats infracommunaux prévus à partir de fin 2008.

Quelques questions méritent qu'on s'y arrête.

Question 6 : *Pourquoi avoir retenu le même taux de sondage dans toutes les communes de 10 000 habitants ou plus ? Les très grandes communes paraissent avantagées.*

Réponse : Les résultats ne sont pas requis seulement au niveau de la commune. Il faut aussi produire de l'information au niveau infracommunal. La "brique de base" de diffusion est l'IRIS-2000. Comme la taille des IRIS est voisine de 2 000 habitants, un taux de sondage uniforme s'imposait pour égaliser la précision des résultats au niveau infracommunal.

Question 7 : *Quel est l'intérêt d'enquêter exhaustivement les adresses de grande taille ? On semble privilégier les quartiers à grands immeubles.*

Réponse : Le problème est celui qu'on appelle en statistique le phénomène des effets de grappe. Dans le ou les immeubles qu'on trouve à une adresse donnée, il y a une forte probabilité de ressemblance des logements qui les composent et de leurs habitants quant aux variables relatives au confort du logement, à la catégorie sociale des habitants, à leur nationalité, etc. Si on pratiquait un sondage, le fait que cette adresse soit tirée ou pas dans l'échantillon aurait un impact certain sur l'estimation de ces variables. L'impact serait d'autant plus fort que l'adresse correspondrait à de nombreux logements et que son poids dans le domaine géographique où sont produites les estimations serait grand : l'impact serait perceptible au niveau de la commune et encore plus à celui de l'IRIS ou de n'importe quelle zone incluant cette adresse. Le fait qu'en cinq ans toutes les adresses au-delà d'une certaine taille seront recensées élimine, par construction, cet inconvénient. L'effet de grappe subsiste pour les adresses de moindre taille mais, parce qu'elles sont plus homogènes, son impact est moins accusé.

En recensant tous les logements des adresses de grande taille, on évite par ailleurs d'avoir à tirer au sort un échantillon de leurs logements, avec tous les problèmes que cela poserait (voir plus loin le [§ B.4.2.2.2](#) relatif aux tests effectués sur les plans de sondage envisagés), y compris quant à la multiplicité des taux de sondage dans une même commune.

Des essais par simulation ont été faits en comparant des plans de sondage sans et avec stratification. La conclusion est que, dans les communes où il n'y a que peu d'adresses de grande taille, la précision des résultats est très semblable ; en revanche, dans les communes où les grandes adresses ont un poids élevé, la stratification est très nettement préférable.

Question 8 : *Comment a été choisi le critère pour définir la catégorie des adresses de grande taille ? Pourquoi ne pas avoir retenu un même nombre absolu pour toutes les communes, 100 logements par exemple ?*

Réponse : Là encore, la réponse tient à l'exigence de garantir des précisions aussi homogènes que possible pour les statistiques infracommunales. Dans certaines communes (en particulier en Île-de-France et dans certaines métropoles régionales), un seuil absolu (de 100 logements par exemple) absorberait l'essentiel des moyens dévolus à l'échantillon de la commune (40 % dans un cycle de cinq ans). On n'aurait donc plus grand-chose, voire rien, pour représenter les logements des autres adresses. Les IRIS ne contenant pas les adresses de grande taille seraient défavorisés pour leur description statistique.

Une solution de compromis raisonnable est de garder le seuil en valeur relative (10% des logements) avec un plancher de 60 logements. Ces seuils ont été choisis en réalisant un grand nombre de simulations concernant plusieurs communes de configurations variées. Autrement dit, on définit de manière pragmatique les adresses de grande taille comme les

adresses les plus grandes en nombre de logements, qui ont au moins 60 logements chacune et qui, réunies, cumulent au maximum 10% des logements de la commune

Question 9 : *La méthode est-elle la même pour les communes des départements d'outre-mer ?*

Réponse : Il n'y a pas encore dans les DOM de RIL permettant de constituer une base de tirage des échantillons d'adresses. La solution apportée est donc la suivante, qui utilise une répartition des îlots et un sondage en deux phases :

- on répartit les îlots¹⁷ d'une commune en cinq groupes équilibrés pour les mêmes variables que celles que l'on a utilisées en métropole (tranches d'âge, sexe, logements, etc.), avec les valeurs observées au recensement de 1999 ;
- chaque année, un groupe est sélectionné (par roulement) et fait l'objet d'une « enquête cartographique » visant, dans chacun des îlots de ce groupe, à localiser toutes les constructions et à dénombrer les logements de celles-ci (et par conséquent à fournir les éléments constitutifs d'un RIL¹⁸) ;
- dans le groupe ainsi mis à jour, on tire un échantillon d'adresses avec un taux de sondage de 40% (en logements).

B.3.3 Les personnes ne vivant pas dans des logements

Toutes les personnes doivent être recensées mais toutes ne vivent pas dans des logements. Il convient donc d'organiser la collecte de manière à pouvoir dénombrer :

- les personnes vivant dans des communautés ;
- les personnes résidant habituellement dans des habitations terrestres mobiles ;
- les personnes sans abri ;
- les marinières et les personnes vivant sur les bateaux de ces derniers.

B.3.3.1 Les personnes vivant dans des communautés¹⁹

La collecte a lieu la même année que les enquêtes de recensement dans les communes recensées exhaustivement et une fois tous les cinq ans dans les communes enquêtées par sondage. Ces dernières sont, dans chaque région, réparties en cinq groupes de manière que la charge de collecte dans les communautés de ces communes soit sensiblement la même chaque année. Pour certaines communes de 10 000 habitants ou plus, qui comptent un grand nombre de communautés ou des communautés hébergeant beaucoup de personnes, la collecte est cependant étalée sur deux années ou plus.

L'échantillon annuel est donc composé, d'une part, des communautés des communes de moins de 10 000 habitants du groupe de rotation concerné par les opérations de recensement de l'année et, d'autre part, des communautés d'une partie des communes de 10 000 habitants ou plus. Ainsi chaque commune ne sera sollicitée qu'une fois tous les cinq ans pour l'expertise de la liste des communautés, à l'exception de la vingtaine de communes qui, du fait de leur volume de communautés, sont réparties sur plusieurs années.

¹⁷ L'îlot est la maille de base d'une partition fine du territoire communal. En ville, c'est le plus souvent un pâté de maisons.

¹⁸ Sur les enquêtes cartographiques dans les DOM, voir l'[encadré](#) au chapitre C, § C.1.1.

¹⁹ Pour la notion de communauté, voir le chapitre C, [§ C.3](#).

Dans les DOM, l'enquête cartographique permet de mettre à jour la liste des communautés implantées dans les communes de moins de 10 000 habitants du groupe de rotation concerné, respectivement dans les îlots du groupe de rotation concerné pour les communes au-dessus de ce seuil. Dans ces dernières, le mode de collecte retenu consiste à interroger toutes les communautés sises dans les îlots du groupe de rotation de l'année.

B.3.3.2 Les personnes résidant habituellement dans des habitations mobiles terrestres

Ces personnes sont recensées sur le territoire de la commune et par ses soins les deux premiers jours de la collecte des enquêtes de recensement. La situation est appréciée à la date du premier jour de la collecte à zéro heure.

La collecte se fait en même temps que les enquêtes de recensement dans les communes de moins de 10 000 habitants et, pour les autres, tous les cinq ans à partir de 2006.

B.3.3.3 Les personnes sans abri

Les personnes sans abri sont recensées sur le territoire de la commune et par ses soins les deux premiers jours de la collecte des enquêtes de recensement. La situation est appréciée à la date du premier jour de la collecte à zéro heure.

La collecte se fait en même temps que les enquêtes quinquennales de recensement dans les communes de moins de 10 000 habitants et, pour les autres, tous les cinq ans à partir de 2006 sur l'ensemble de leur territoire.

B.3.3.4 Les mariniers et les personnes vivant sur les bateaux de ces derniers

Pour les mariniers et les personnes qui vivent sur les bateaux de ces derniers, la collecte est assurée par l'Insee avec le concours du Service des voies navigables. Elle a lieu une fois tous les cinq ans, à partir de 2006, sur l'ensemble du territoire français, selon des modalités analogues à celles du recensement de 1999 : recensement aux écluses et aux frontières.

.

*

* *

Le tableau qui suit présente les principales différences entre le recensement général de la population et le nouveau recensement.

Encadré 4 : Comparaison entre l'ancien et le nouveau recensement

	Le recensement de 1999	Le recensement à partir de 2004
La méthode d'enquête	<ul style="list-style-type: none"> Recensement exhaustif par comptage traditionnel 	<ul style="list-style-type: none"> Techniques d'enquêtes annuelles de recensement
La fréquence de la collecte	<ul style="list-style-type: none"> Tous les 8 ou 9 ans 	<ul style="list-style-type: none"> Chaque année dans les communes de 10 000 habitants ou plus Tous les 5 ans dans les communes de moins de 10 000 habitants
La durée de la collecte	<ul style="list-style-type: none"> 1 mois 	<ul style="list-style-type: none"> 4 ou 5 semaines selon la taille de la commune
Les distinctions entre communes	<ul style="list-style-type: none"> Aucune 	<ul style="list-style-type: none"> Pour les communes de moins de 10 000 habitants : recensement exhaustif quinquennal Pour les communes de 10 000 habitants ou plus : enquête annuelle auprès d'un échantillon de logements
Le nombre de communes concernées par chaque collecte	<ul style="list-style-type: none"> La totalité des 36 679 communes 	<ul style="list-style-type: none"> Environ 8 000 communes chaque année. Toutes les communes ne sont pas recensées la même année. Cela dépend de la taille de la commune
Le pourcentage de la population recensé à chaque collecte	<ul style="list-style-type: none"> 100 % 	<ul style="list-style-type: none"> Dans les communes de moins de 10 000 habitants : 100% des habitants dans une commune sur cinq chaque année Dans les communes de 10 000 habitants ou plus : 8% des habitants chaque année, 40% au bout de cinq ans Au total : 14% de la population par an
Les acteurs du recensement	<ul style="list-style-type: none"> L'Insee prépare et contrôle la collecte. Il traite les données et diffuse les résultats. La commune réalise la collecte 	<ul style="list-style-type: none"> Un partenariat renforcé entre les communes et l'Insee : L'Insee organise et contrôle la collecte des informations. Il traite les données et diffuse les résultats. La commune prépare et réalise les enquêtes de recensement
La fraîcheur de l'information	<ul style="list-style-type: none"> 6 ans en moyenne, compte tenu de l'espacement entre deux recensements 	<ul style="list-style-type: none"> 3 ans en moyenne
Le dispositif de communication	<ul style="list-style-type: none"> Campagne ponctuelle d'information nationale et locale 	<ul style="list-style-type: none"> Campagne nationale d'information annuelle Communication ciblée vers le public et les zones concernés chaque année par le recensement

B.4 Les tests menés de 2000 à 2003

La mise au point de l'ensemble du dispositif qui vient d'être décrit a nécessité de nombreux tests, dont la réalisation s'est échelonnée tout au long des années 2000 à 2003. L'[encadré](#) joint en donne une présentation synoptique et les développements qui suivent en décrivent, année par année et thème par thème, les modalités et les résultats.

B.4.1 Les tests de l'année 2000

B.4.1.1 Choix de l'unité primaire d'échantillonnage (printemps 2000)

Le tout premier plan de sondage prévoyait, dans les communes de 10 000 habitants ou plus, la collecte de renseignements auprès d'un échantillon de logements tirés dans la liste exhaustive des logements d'un échantillon « d'immeubles du recensement ». Cette liste découlait elle-même d'un « prérecensement » consistant à inventorier et localiser précisément tous les logements de chacun des immeubles sélectionnés.

La notion-clé, ici, était « immeuble du recensement », c'est-à-dire l'ensemble des logements desservis par la même cage d'escalier²⁰. A la limite, il pouvait s'agir d'une maison individuelle comme d'une tour de quarante étages.

Un test de repérage a eu lieu au printemps 2000 afin de voir dans quelle mesure il était possible de repérer le bon « immeuble ». Le test, organisé dans trois communes de grande taille en milieu densément construit, a montré la difficulté d'utiliser la notion même « d'immeuble » : près de 3% des immeubles n'ont pas pu être retrouvés avec certitude, même en recourant aux documents du recensement de 1999 (RP99).

Un tel taux de défaillance a fait retenir l'adresse comme unité primaire d'échantillonnage. L'adresse est une notion d'usage général ; dans la majorité des cas, elle correspond à un édifice d'habitation ou d'activité repéré par un numéro (éventuellement assorti d'un indice de répétition) dans une voie portant un nom. Voir à ce sujet l'[encadré](#) joint.

B.4.1.2 Test des principaux questionnaires (juillet et novembre 2000)

Les deux principaux questionnaires du recensement de la population (bulletin individuel et feuille de logement) ont fait l'objet d'un test qualitatif en juillet puis en novembre 2000. Ce test a intégré les attentes du groupe de travail du Conseil national de l'information statistique (CNIS) alors en activité (voir chapitre F, [§ F.6.3.2](#)), qui désirait expérimenter un certain nombre de questions. De ce fait, la taille des questionnaires n'avait pas été limitée et le test s'est déroulé avec des bulletins individuels comportant quatre pages.

B.4.1.2.1 Les objectifs du test

La méthode de collecte du recensement est dite par dépôt-retrait. Cela signifie que l'agent recenseur dépose les questionnaires dans les logements et que les personnes recensées les remplissent elles-mêmes.

²⁰ Cette notion est une extension de celle utilisée lors du recensement général de la population de 1999 où l'immeuble ne pouvait compter plus de 60 logements ; un « immeuble » était donc l'ensemble des logements desservis par une même cage d'escalier qu'on pouvait lister sur le « Dossier d'immeuble collectif », lequel comptait 60 lignes... Pour les besoins du nouveau recensement, les différents « immeubles » correspondant à la même cage d'escalier ont été regroupés en un seul immeuble.

Encadré 5 : Les tests (hors site expérimental)

Présentation d'ensemble des tests (hors site expérimental)					
Campa- gne de tests (année)	Domaine de test				
	Question- naires	Plan de sondage	Protocole de collecte	Validation par les communes	Applications Informati- ques et organisation s
2000					
Jun		Test du choix entre immeuble et adresse			
Juillet et novembre	Tests qualitatifs				
2001					
Février			Liste de référence		
Mai et juin	Groupes de discussion				
Mars à juin	Test quantitatif	GC : sondage liste ou grappe	GC : faisabilité technique du prérecense- ment		
Novembre et décembre	Test quantitatif	GC : sondage grappe	GC : faisabilité technique		
2002					
Mars			PC : faisabilité technique		
Mars			PC : faisabilité organisation- nelle	GC : faisabilité organisation- nelle	
Mars à juin	Question- naires DOM	GC : sondage DOM	GC : faisabilité technique DOM		
2003					
Février					Répétition
Jun	Question- naires des commu- nautés				Organisation pour les communau- tés

GC = commune de 10 000 habitants ou plus

PC = commune de moins de 10 000 habitants

Encadré 6 : La notion d'adresse dans le RIL

La notion d'adresse dans le RIL

Une adresse regroupe un ou plusieurs bâtiments contenant des logements d'habitation, des établissements d'activité économique, des communautés ou des équipements

Adresse normalisée

Une adresse normalisée est caractérisée, dans une commune, par un nom de voie et un numéro dans la voie, auquel s'ajoute parfois un indice de répétition (bis, ter...).

Par exemple : 3 rue Condorcet – 7bis rue St Antoine

Dans les communes de 10 000 habitants ou plus, 95% des adresses sont normalisées.

On recense l'ensemble des logements situés à une adresse normalisée, quels que soient le nombre de bâtiments et le type de construction :

- un bâtiment d'habitation d'un seul logement ;
- plusieurs bâtiments d'habitation d'un seul logement ;
- un bâtiment d'habitation de deux logements ou plus ;
- plusieurs bâtiments d'habitation de deux logements ou plus ;
- un bâtiment administratif avec un ou plusieurs logement d'habitation.

A une adresse normalisée correspond une seule adresse au sens du recensement. Tous les logements recensés à l'adresse portent le même rang d'adresse, quel que soit le nombre de bâtiments recensés à cette adresse.

Cas particulier au concept d'adresse normalisée : les adresses dégroupées

Cela concerne les adresses de grande taille (en nombre de logements) qu'il apparaît possible de scinder en deux ou plusieurs éléments, ainsi que les bâtiments ayant une adresse comportant une lettre ou un nom et clairement repérables sur le terrain. On crée ainsi deux ou plusieurs adresses là où l'adresse normalisée était unique.

Par exemple, le 3 rue Lamartine dessert une résidence formée de deux bâtiments A et B ; on pourra distinguer deux adresses 3A et 3B.

Pour alerter les acteurs de la collecte sur ce cas, dans la zone « indice de répétition » des listes d'adresse, un # suivi d'un ou deux caractères renvoie sur une zone "complément d'adresse" dans laquelle est précisée la partie de l'adresse normalisée à recenser.

Adresse non normalisée

Une adresse sans numéro dans la voie, voire sans nom de voie (lieu-dit, nom de résidence ou de lotissement) est une adresse non normalisée. On trouve ce type d'adresse principalement dans les communes rurales ou en périphérie des villes.

Par exemple : impasse du Parc, chemin de l'Observatoire, hameau Bourbon, lotissement du Moulin.

Lorsqu'une adresse n'est pas normalisée, chaque bâtiment de l'adresse correspond à une adresse au sens du nouveau recensement et porte donc un rang d'adresse distinct. Pour localiser le bâtiment, un complément d'adresse est introduit afin de préciser sa position. Dans une voie sans numéros comportant plusieurs bâtiments, il y a autant d'adresses au sens du recensement que de bâtiments.

Par exemple : Impasse du Parc, maison à droite ; lotissement du Moulin, bâtiment A.

Les questionnaires doivent donc être simples, tant dans leur formulation, qui doit être facilement comprise par tous, que dans le cheminement entre les questions. Les questionnaires doivent aussi être courts et ne pas comporter de questions jugées sensibles par les enquêtés, susceptibles d'entraîner un rejet de l'intégralité du questionnement.

Les objectifs du test qualitatif étaient :

1. de vérifier la bonne compréhension par les personnes enquêtées de questions nouvelles ou modifiées par rapport au recensement de 1999 ;
2. de vérifier la pertinence des formulations des questions et des notes explicatives ;
3. de valider la structure générale des questionnaires (place des questions, lisibilité de la maquette, cheminement entre les questions).

B.4.1.2.2 Les modalités du test

Le test de juillet a été organisé en Lorraine auprès de 200 logements ; celui de novembre en Lorraine et en Haute-Normandie²¹ auprès de 400 logements.

Pour ces tests, on ne recherchait pas la représentativité de l'échantillon au sens statistique du terme, mais simplement que toutes les grandes composantes de la population soient présentes (habitat urbain/rural, habitat HLM/non HLM, personnes âgées/jeunes, actifs/inactifs, familles/personnes seules, etc.).

Le test a été réalisé par entretiens d'environ une heure en face à face. Son protocole comprenait des phases distinctes et d'égale importance :

1. La personne enquêtée remplissait d'abord les questionnaires de façon auto-administrée, sans aucune intervention de la part de l'enquêteur. Pendant qu'elle remplissait les questionnaires, l'enquêteur notait les hésitations, le recours à des documents extérieurs, la manipulation des questionnaires (retours en arrière, etc.) ;
2. L'enquêteur l'interrogeait ensuite sur sa compréhension des questions, les difficultés rencontrées, les modalités de réponse inadaptées, etc., en l'incitant à expliciter le cheminement qui l'avait conduite à apporter une réponse plutôt qu'une autre, voire à ne pas répondre à la question. Pour ce faire, l'enquêteur disposait d'un guide d'entretien.

L'enquête était réalisée dans tous les logements de l'échantillon occupés lors de la période de collecte. Chaque fois, seule une personne adulte était interrogée. Dans le cas de ménages de plus de deux personnes, seuls deux bulletins individuels étaient remplis.

Les questionnaires testés ont été élaborés en repartant des questionnaires du recensement de 1999 et en tenant compte des nouveaux besoins d'information statistique locale exprimés par les utilisateurs et pouvant relever d'un recensement. Ces besoins se sont exprimés en particulier à l'occasion de la concertation conduite sous l'égide du CNIS et lors des concertations organisées par les directions régionales de l'Insee au niveau local (voir le [chapitre F](#)). Les questionnaires intégraient également les modifications induites par la nouvelle méthode de recensement envisagée.

Deux versions de la feuille de logement et du bulletin individuel ont été testées en juillet ; deux nouvelles versions de chacun de ces deux questionnaires ont été testées en novembre 2000. Les versions de novembre tenaient compte des enseignements du test de juillet 2000.

B.4.1.2.3

²¹ La direction régionale de Haute-Normandie gérant également le département du Val-d'Oise (95), le test a aussi été réalisé dans ce département.

Les enseignements du test

Le test de juillet 2000 a permis de dégager de nombreux axes d'amélioration des questionnaires.

Les réactions générales spontanées des personnes enquêtées ont porté sur la longueur de remplissage des questionnaires et sur le fait que des questions rétrospectives n'avaient pas leur place dans un questionnaire de recensement, qui devait à leurs yeux s'intéresser uniquement à la situation présente des habitants.

Concernant le contenu même des questionnements, les principaux enseignements ont été les suivants :

Dans la feuille de logement :

- les tableaux de composition des ménages demandaient à être mieux hiérarchisés afin de ne pas noyer sous les cas particuliers les personnes en situation simple ;
- la question sur les emplacements de stationnement était à revoir (forme et vocabulaire) ;
- les personnes enquêtées faisaient état d'une méconnaissance ou de fortes incertitudes sur le nombre de pièces du logement²², sa surface et son année d'achèvement.

Dans le bulletin individuel :

- la question sur l'existence d'un second logement était bien comprise, mais sa présentation demandait à être améliorée ;
- l'ajout d'une question sur la vie en couple était bien perçu par les personnes interrogées dans le cadre du test, qui estimaient satisfaisant le fait de n'être pas contraintes de décrire leur situation par le seul état matrimonial légal ;
- la question sur l'année de fin de formation leur demandait un long temps de réflexion, cette année n'étant pas connue dans la majorité des cas ;
- les questions sur l'emploi étaient perçues comme trop techniques (catégorie professionnelle notamment) ;
- la tentative de poser les mêmes questions sur l'emploi en cours (pour les personnes en emploi) et sur le dernier emploi occupé (pour les personnes ne travaillant pas ou plus) était un échec quelle que soit la version proposée. Le taux de non réponse était élevé. L'objectif d'améliorer la qualité de la catégorie socioprofessionnelle pour les personnes au chômage ou retraitées sans augmenter le nombre de questions ne pouvait donc être atteint dans un questionnaire de recensement. Il fallait revenir à des questions spécifiques pour les personnes en emploi et d'autres questions pour les personnes ayant travaillé.

Enfin, le test a également permis de proposer des améliorations de formulation de plusieurs questions.

Les questionnaires proposés au test de novembre ont été nettement plus faciles à remplir par les personnes enquêtées que ceux du test de juillet. Cela s'est traduit par un meilleur taux de réponse aux questions, moins de difficultés de compréhension, moins d'erreurs dans les réponses et le cheminement et par un remplissage plus rapide.

Les points les plus positifs des questionnaires du test de novembre ont été :

- l'introduction d'un tableau spécifique pour les étudiants dans les tableaux de composition des ménages ;
- la meilleure formulation de la question alors envisagée sur la double résidence ;

²² Dans un tiers des logements, le nombre de pièces déclaré au test est différent de celui déclaré au RP99.

- l'introduction d'une question sur la vie en couple ;
- la meilleure présentation de la question sur la situation principale (emploi/chômage, etc.) ;
- le meilleur cheminement dans la partie activité/emploi.

En revanche, des difficultés sensibles ont persisté sur les sujets suivants :

- les consignes sur les résidences secondaires (notamment celles occupées plusieurs mois dans l'année) dans le tableau de composition des ménages ;
- la question sur l'emplacement de stationnement ;
- la question sur le statut d'occupation du logement antérieur ;
- la question sur la catégorie d'emploi.

Le test a également apporté des indications précieuses sur la présentation des questionnaires.

Les résultats du test qualitatif de l'année 2000 demandaient à être confortés et complétés par des tests quantitatifs auprès d'un grand nombre de personnes afin de valider, notamment, la pertinence et la qualité de l'information collectée sur l'emploi. Cette préoccupation a été prise en compte dans les tests de l'année 2001, à la fin de laquelle le contenu et la forme des questionnaires ont été arrêtés, et dans les tests des années suivantes, qui ont permis d'affiner la présentation visuelle des questions.

B.4.2 Les tests de l'année 2001

Les tests prévus en 2001 étaient de deux types : des tests qualitatifs concernant les questionnaires et la collecte et des tests quantitatifs de collecte. Ils ont été réalisés par l'Insee sans participation directe des communes.

B.4.2.1 Tests qualitatifs (février, mai et juin 2001)

B.4.2.1.1 Test de comparaison des maquettes des questionnaires (mai et juin 2001)

Une réflexion sur les maquettes des principaux questionnaires²³ du recensement avait été engagée dès l'été 2000 avec l'appui de graphistes et de sociétés de lecture automatisée. Son objectif principal était de prendre en compte le plus en amont possible les contraintes de manipulation lors de la collecte et du traitement en lecture automatisée (dite lecture optique). Lors des tests de 2001, il s'agissait de comparer la perception, par les personnes enquêtées, des graphismes très différents de deux maquettes. De plus, différents pays utilisant pour leur recensement des questionnaires regroupés dans un cahier comprenant le questionnaire relatif au logement et plusieurs questionnaires individuels, ce test devait aussi permettre de mesurer l'acceptabilité par les enquêtés d'un tel questionnaire par rapport aux questionnaires traditionnels avec feuille de logement (FL) et bulletins individuels (BI) séparés.

²³ Feuille de logement et bulletin individuel.

Confié à un prestataire spécialisé dans la technique des groupes de discussion, le test a été réalisé en mai-juin 2001 et complété par une expertise des différents types de questionnaires pour la lecture optique.

Quatre types de questionnaires ont été retenus pour le test :

- deux structures de questionnaires :

Structure 1 : questionnaires séparés (reconduction du choix des recensements précédents en France). Sont remplis : une feuille de logement (4 pages A4) pour chaque logement et un bulletin individuel (2 pages A4) pour chacune des personnes devant être recensées dans le logement correspondant ;

Structure 2 : questionnaire-cahier. Cette option est retenue dans plusieurs pays. Chaque questionnaire-cahier avait 24 pages A4 et regroupait une feuille de logement et six bulletins individuels. Un questionnaire-cahier devait être renseigné pour chaque logement ;

- deux maquettes de questionnaires :

Un travail spécifique sur le graphisme des questionnaires a également été réalisé dans la perspective de faciliter la lecture optique des données. Il a conduit à deux maquettes différentes, à la fois par les couleurs mais aussi par le traitement des zones à renseigner (défoncé dans un cas, filets de tour de case dans l'autre cas) ou des consignes de remplissage.

Les résultats :

Les groupes de discussion ont permis de choisir une maquette parmi les deux qui étaient proposées.

Les tests ont conduit à retenir pour les bulletins individuels la forme traditionnelle de feuilles indépendantes de format A4 insérées dans le questionnaire-logement (feuille A3 formant chemise). En effet, la formule consistant à regrouper les questionnaires d'un même ménage dans un cahier a été moins appréciée, sauf par les ménages nombreux, car jugée trop lourde.

De plus, ces tests ont permis de recueillir de nombreuses suggestions sur le cheminement dans le questionnaire (place des exemples, formulation des filtres, etc.).

B.4.2.1.2 Expertise technique des différentes maquettes pour la lecture optique (juin 2001)

Dans le choix entre les deux maquettes de questionnaires, la facilité et le coût du traitement nécessaire pour l'acquisition des données en lecture optique sont des critères déterminants. Il s'agissait d'obtenir, à partir des normes admises dans la profession et en recueillant l'avis de plusieurs prestataires utilisant des logiciels différents de reconnaissance de caractères, des recommandations concernant la forme des questionnaires. Cette analyse, confiée à un cabinet spécialisé dans le domaine de la lecture optique et achevée fin juin 2001, a porté à la fois sur les modalités de mise en place d'une chaîne de traitement en fonction des différents questionnaires, sur les améliorations possibles et sur les coûts.

Les résultats :

La maquette choisie l'a été après qu'a été vérifiée sa compatibilité avec toutes les exigences techniques de la lecture automatisée des questionnaires.

B.4.2.1.3 Test de faisabilité d'une liste de référence en commune de moins de 10 000 habitants (février 2001)

Pour une commune de moins de 10 000 habitants, la rénovation ne changeait pas beaucoup la manière de réaliser le recensement. Mais des améliorations pouvaient être apportées sur certains points, notamment en ce qui concerne le contrôle d'exhaustivité de la collecte et la mise à disposition d'outils permettant son suivi et facilitant le travail de numérotation et de récapitulation des imprimés collectés.

C'est avec cet objectif qu'a été imaginé l'établissement d'une "liste de référence" d'adresses pour guider l'agent recenseur lors de la collecte. Cette liste serait constituée, juste avant la collecte, en rapprochant le fichier de la taxe d'habitation de la commune, trié par adresse, d'une liste des adresses établie par l'agent recenseur pour son secteur lors d'une tournée de reconnaissance précédant de peu la collecte. A l'issue de l'opération, on disposerait ainsi d'un fichier d'adresses représentant la cible à atteindre. La collecte continuerait d'être réalisée sur un mode aréolaire par district (c'est-à-dire que tous les logements de chaque adresse de l'aire constituée par le district seraient à recenser), mais pourrait être suivie et contrôlée en permanence grâce à cette liste. En effet, celle-ci fournirait les adresses à recenser, permettrait la pré-numérotation des questionnaires, la production d'indicateurs de suivi en cours de collecte et le contrôle d'exhaustivité de cette collecte.

Avant d'être mis en œuvre, un tel protocole, qui s'inspirait de celui développé au RP99 dans les communes de grande taille, nécessitait d'être expérimenté puis testé dans toutes les étapes de sa réalisation.

L'objectif du test de 2001 a été de vérifier la possibilité de créer une telle liste en partant d'un relevé d'adresses réalisé par l'agent recenseur lors de sa tournée de repérage et en le complétant par des informations issues du fichier de la taxe d'habitation.

Le programme de tests prévoyait :

- une première expérimentation dans trois communes en février 2001 en région Basse-Normandie ;
- un test de collecte de plus grande ampleur en octobre-novembre 2001 en région Rhône-Alpes ;
- un test quantitatif si nécessaire.

Les résultats :

Un premier travail en bureau a permis de vérifier la faisabilité et la qualité d'un appariement entre les données de la taxe d'habitation et d'une liste d'adresses établie à partir d'informations relevées sur le terrain.

Ce résultat encourageant a conduit à l'expérimentation de février 2001 qui s'appuyait cette fois directement sur un relevé d'adresses effectué sur le terrain. Il est alors apparu un taux d'appariement qui, tout en étant élevé, était insuffisant pour que la liste d'adresses en résultant puisse être utilisée en l'état par l'agent recenseur au cours de sa collecte, en particulier pour permettre la numérotation des questionnaires.

Avant de lancer un test sur un nombre important de communes, il fallait donc procéder à un approfondissement du mode opératoire à tester. Cette réflexion a débouché sur un test réalisé en 2002 (voir plus bas [§ B.4.3.1](#)).

B.4.2.2 Tests quantitatifs (mars à décembre 2001)

Les tests quantitatifs prévus au cours de l'année 2001 étaient au nombre de six. Ils devaient se dérouler en trois vagues selon le schéma suivant, indépendamment de la première expérimentation de la liste de référence en février 2001:

Première vague

Champ : communes de 10 000 habitants ou plus

Période du pré-recensement et de la collecte : mars à juin 2001

Ampleur de la collecte : deux échantillons de 2 000 logements chacun répartis sur huit communes

Objectifs :

- comparaison des plans de sondage ;
- faisabilité technique du sondage sur liste ;
- évaluation des procédures de contrôle ;
- évaluation des questionnements.

Deuxième vague

Champ : communes de 10 000 habitants ou plus ;

Période du pré-recensement et de la collecte : septembre à décembre 2001 ;

Ampleur de la collecte : 7 000 logements dans 11 communes différentes et trois régions (Aquitaine, Bretagne et Ile-de-France) ;

Objectifs :

- mise au point du protocole "adresses" (notamment la tournée de reconnaissance des adresses à recenser, la pré-numérotation des questionnaires et les mesures de charge) ;
- évaluation des procédures de contrôle ;
- évaluation des questionnements.

Au total, 4 635 logements ont été effectivement recensés dans 2 061 adresses. Les tests ont permis de conforter le mode opératoire prévu pour la tournée de reconnaissance des adresses en début de collecte et le processus de suivi de la collecte par la tenue d'indicateurs hebdomadaires.

Troisième vague

Champ : communes de moins de 10 000 habitants ;

Période de la collecte : fin 2001 ou début 2002 ;

Ampleur de la collecte : environ 18 000 logements dans 30 communes différentes ;

Objectifs :

- protocole de collecte ;
- mesures de charge ;
- évaluation des procédures de contrôle ;
- évaluation des questionnements.

Le détail des tests effectivement réalisés est exposé ci-dessous.

B.4.2.2.1 Faisabilité du prérecensement dans le cas de sondage sur liste en commune de 10 000 habitants ou plus (mars 2001)

Le changement d'unité primaire d'échantillonnage décidé après le test de l'an 2000 nécessitait la reprise des tests de prérecensement puisque se posait désormais la question : peut-on toujours retrouver avec sûreté une unité identifiée par une adresse (numéro, indice de répétition, nom de voie) ?

Par ailleurs, après le séminaire d'octobre 2000 organisé par la Société française de statistique (SFdS) et l'Insee²⁴, il fallait répondre à la question de l'efficacité du sondage sur liste (40% des logements d'un groupe de rotation d'adresses sont enquêtés) plutôt que par grappe (tous les logements d'une même adresse sont enquêtés) et par conséquent mesurer les taux de réponse selon les deux méthodes de sondage. On attendait du test une comparaison des avantages et inconvénients des deux méthodes et une évaluation du mode de sondage donnant le meilleur taux de réponse.

Dans le cas du sondage sur liste, on voulait en outre :

- d'une part mesurer la qualité du repérage des adresses et des logements par les agents précenseurs, la possibilité pour les agents recenseurs de trouver le logement à enquêter à partir des informations de localisation fournies par les agents précenseurs et l'intérêt de fournir ou non aux agents recenseurs des informations sur les logements voisins pour les aider à retrouver le bon logement ;
- d'autre part, tester la faisabilité d'une numérotation des questionnaires au moment du dépôt chez le ménage et le remplissage par l'agent recenseur des éléments d'adressage du logement et de sa catégorie.

Le test mis sur pied pour 2001 a été organisé en deux temps.

- d'abord, un test de précensement où l'on vérifierait que l'adresse était plus utile sur le terrain (en tout cas pas moins utile !) que l'immeuble au sens du recensement ;
- ensuite, un test de collecte à la fin du printemps 2001. Outre les objectifs mentionnés plus haut, ce test devait permettre de tester le contenu et l'ergonomie de différents carnets de tournée. Cette partie du test est reprise en détail au [§ B.4.2.2.2.](#)

La mise en œuvre d'un sondage sur liste nécessite un précensement pour dresser la liste de l'ensemble des logements des adresses du groupe de rotation de l'année et identifier leur emplacement dans l'immeuble. Le tirage de l'échantillon des 40% de logements à enquêter se fait dans cette liste. Pour des raisons tenant aux délais de traitement, le précensement doit avoir lieu assez en amont de la collecte. Il est en effet nécessaire d'assurer la saisie des informations de localisation des logements de l'échantillon qui doivent être imprimées par la suite sur le carnet de tournée de l'agent recenseur. Avec un carnet de tournée papier comme celui utilisé dans le test, un délai de deux mois était un minimum incompressible.

Le test de précensement a été réalisé du 19 au 30 mars 2001. Il a porté sur 1 475 adresses. On en donne ici les principales conclusions :

Les résultats du test de précensement :

La première constatation est que le repérage à l'adresse ne lève pas toutes les difficultés : certaines adresses n'ont pas d'adresse, du moins pas d'adresse structurée selon les normes de La Poste ! En outre, il s'est révélé impossible de faire appliquer partout avec la même rigueur les consignes de précensement (et notamment le mode de tournée dans les étages).

Certaines adresses n'ont pu être précensées :

Le taux d'échec au précensement s'est établi à près de 8% si l'on s'en tient à la collecte brute par les agents précenseurs. L'intervention des superviseurs n'a guère permis de réduire ce taux qui est resté important (un peu moins de 6%). Ces échecs ont plusieurs origines :

- une faible part (1%) est due à la qualité insuffisante des adresses fournies aux agents précenseurs, qui ne permettait pas ou permettait difficilement l'identification du

²⁴ Voir au chapitre F [§ F.5.3](#)

bâtiment correspondant sur le terrain. Dans certains cas, les éléments de l'adresse étaient même inconnus dans la commune ;

- une part plus importante des échecs est due à l'impossibilité d'entrer dans les immeubles : 4,6 % après intervention des superviseurs²⁵.

Le fait que l'on ait utilisé pour ce test des adresses brutes issues de la saisie des documents remplis par les agents recenseurs de 1999 a sans doute contribué à accroître le nombre de ces cas. Le géocodage qui a accompagné ultérieurement la constitution du RIL a permis en effet de repérer des anomalies d'adresses qui ont été corrigées à ce stade. De même, la comparaison du RIL avec des fichiers administratifs a permis ultérieurement d'identifier des adresses suspectes qui ont pu être vérifiées. Par ailleurs, la disponibilité de plans positionnant les adresses a permis, lors des enquêtes de recensement en vraie grandeur, de diminuer l'incertitude de repérage des constructions sur le terrain.

Mais il n'est pas certain que toutes ces anomalies auraient pu être repérées et traitées en amont de la réalisation d'un précensement. Ainsi en aurait-il été par exemple d'adresses qui présentaient toutes les apparences de "bonnes adresses" et qui pourtant n'ont pu être retrouvées sur le terrain (une vingtaine dans le test).

Le précensement a présenté des erreurs :

Les erreurs de précensement ont été relativement moins nombreuses. Toutes catégories confondues, elles ne dépassaient pas 2%. Par ailleurs, les procédures de contrôle proposées permettaient assez facilement de les détecter et de les redresser.

Ces erreurs portaient principalement sur le repérage sur le terrain, dès lors que le tissu urbain était complexe ou l'adresse imprécise.

Les cas d'erreur sur le décompte du nombre de logements ont été rares²⁶. De fait, la correspondance globale sur l'échantillon entre le nombre de portes précensées et le nombre de logements recensés au RP99 a été bonne.

Une localisation des logements fréquemment basée sur le nom de l'occupant

C'est la facilité avec laquelle l'agent recenseur retrouvera le logement lors de la collecte qui permet de juger de la qualité des informations de localisation consignées par l'agent précenseur sur son carnet de tournée (voir ci-dessous [§ B.4.2.2.2](#)).

Pour plus de 60% des portes précensées, un nom, la plupart du temps celui inscrit sur la porte, a été relevé par l'agent précenseur et vient compléter les éléments de localisation. C'est une proportion importante et utile pour les contrôles mais loin des 100%. On peut signaler quelques cas où, du fait de l'impossibilité dans laquelle s'est trouvé l'agent précenseur de pénétrer dans l'immeuble, il a réparti les noms inscrits sur l'interphone entre les différents logements, en donnant des localisations au hasard et sans vérifier le nombre total de logements de l'adresse...

En conclusion, le précensement est pour plusieurs raisons une opération complexe et délicate à mettre en œuvre :

1) Il est difficile de pénétrer dans les immeubles au seul motif de dresser la liste des logements avec le nom des occupants. De nombreux immeubles munis d'un digicode ou d'un interphone nécessitent un contact avec au moins un occupant pour y pénétrer. Au total, 7 % des adresses n'ont pu être précensées avant intervention d'un agent de l'Insee ;

²⁵ Compte tenu de l'absence de campagne de communication et de soutien logistique de la part des communes, les conditions du test n'étaient pas aussi favorables qu'en grandeur réelle. Mais l'expérience montre que c'est la nécessité d'avoir à enquêter qui permet finalement de franchir les portes qui ne veulent pas s'ouvrir. L'annonce d'un précensement comportant une simple visite de repérage des logements n'a pas la même force de persuasion vis-à-vis des habitants.

²⁶ L'agent précenseur disposait de l'information sur le nombre de logements de l'adresse au RP99.

2) La localisation du logement par les soins de l'agent précenseur à l'aide d'informations descriptives (situation sur le palier, etc.) s'est révélée à l'étape suivante malcommode à utiliser par l'agent recenseur et les schémas n'aidaient pas davantage. La localisation des logements s'effectue principalement à l'aide du nom, ce qui n'est pas satisfaisant dès lors que le nom n'est présent que dans 60 ou 70 % des cas, qu'il peut changer entre le moment du précensement et la date de la collecte et qu'il peut y avoir des homonymies dans l'immeuble ;

3) La saisie des informations du précensement, le tirage et la restitution de la liste des logements à enquêter, posent de redoutables problèmes d'organisation. La saisie des informations du carnet-papier s'avère difficile en raison de l'écriture peu déchiffrable de l'agent précenseur, qui le plus souvent a dû remplir le carnet debout dans l'immeuble ou la rue.

B.4.2.2.2 Comparaison de plans de sondage en commune de 10 000 habitants ou plus (mai et juin 2001)

Le test a été réalisé auprès de 4 000 logements répartis dans 48 IRIS²⁷, huit communes et trois régions (Franche-Comté, Haute-Normandie et Lorraine). Pour des raisons de logistique, les logements ont été échantillonnés dans des communes proches des villes sièges des directions régionales de l'Insee car l'encadrement était confié à des agents permanents de ce dernier.

Les IRIS ont été préalablement classés par type de construction dominant afin de prendre en compte la variété de l'habitat lors de la collecte. Quatre types ont été définis : IRIS à dominante d'immeubles anciens, IRIS à dominante d'habitat individuel, IRIS à forte concentration de grands immeubles et IRIS hétérogènes. Chacun des quatre types d'IRIS était représenté approximativement à part égale dans l'échantillon final des 48 IRIS.

Deux modes de sondage ont été expérimentés : le mode « liste », qui consiste à effectuer un précensement dans le groupe de rotation d'adresses de l'année afin de dresser la liste complète des logements, à en tirer 40%, puis à confier à un agent recenseur la liste et la description des logements à recenser ; et le mode « adresses », qui conduit à enquêter tous les logements des adresses tirées au sort parmi le groupe de rotation d'adresses de l'année.

La moitié des IRIS (soit 24) a été collectée selon un mode de sondage « liste » : on a tiré au hasard 1/5 de leurs adresses, qu'on a précensées, puis on a sélectionné 40% des logements de ces adresses, que l'on a recensés. De plus, pour la moitié de ces IRIS (soit 12), les agents recenseurs disposaient de l'information de localisation concernant les logements voisins des logements à recenser, l'agent recenseur n'ayant d'information que sur « le » logement à recenser dans l'autre moitié des IRIS.

Dans les 24 autres IRIS, on a effectué un sondage en grappe de logements, l'îlot étant considéré comme une grappe de logements individuels et une adresse étant une grappe pour les adresses de 2 logements ou plus. On a choisi les adresses de manière à obtenir un taux de sondage de 8% environ dans chaque IRIS.

Les secteurs d'agent recenseur étaient composés en général de deux IRIS contigus et de même type : un IRIS était interrogé en mode « liste » et un IRIS interrogé en mode « grappe ». La taille de ces secteurs était à peu près de 2 000 logements correspondant à 160 logements à recenser.

²⁷ IRIS pour "îlots regroupés pour l'information statistique". Les communes de 5 000 habitants ou plus sont découpées en IRIS (ou quartiers) d'environ 2 000 habitants.

Compte tenu des contraintes de temps et de budget, ce plan ne permettait pas de tester tous les effets principaux et leurs interactions. En particulier, tous les effets « agent recenseur » n'étaient pas maîtrisés. En revanche, l'effet « type d'IRIS » était pris en compte du fait de la multiplicité des IRIS et parce que chaque mode de collecte était mis en œuvre sur chacun des quatre types d'IRIS.

Dans chaque direction régionale, deux agents prérecenseurs ont participé à l'opération. Dans chaque IRIS du mode "liste", un cinquième des adresses a été tiré au sort. L'échantillon total comprenait 1 477 adresses. Toutes les adresses étaient prérecensées lors de la première semaine. Puis 114 d'entre elles, sélectionnées en raison de leur complexité, étaient refaites lors de la deuxième semaine par l'autre agent prérecenseur de la DR.

Dans chacun des modes, 2 000 logements ont été recensés, chaque agent recenseur ayant à travailler pour moitié selon chacun des deux modes afin d'éliminer les effets liés à l'enquêteur. Le test avait été déclaré d'intérêt général mais il n'était pas obligatoire d'y répondre.

Pour la collecte, huit agents recenseurs différents de ceux ayant réalisé le prérecensement ont été recrutés. Ils n'avaient pas d'expérience particulière des enquêtes. Une formation d'une journée leur a été dispensée et un matériel de communication en direction de la population a été mis à leur disposition. Ce matériel comprenait des affichettes à apposer dans les immeubles, des lettre-avis et des avis de passage.

La collecte, quel que soit le type de sondage, était prévue pour quatre semaines, du 14 mai au 9 juin 2001. Elle a été prolongée d'une semaine compte tenu des difficultés rencontrées par les agents recenseurs pour contacter les ménages en cette période de "ponts".

Des contrôles de collecte ont été réalisés du 25 juin au 6 juillet avec l'aide d'enquêteurs pigistes de l'Insee.

Les résultats :

Les principales conclusions du test de prérecensement ont été exposées au [§ B.4.2.2.1](#)

Les principales conclusions du test de sondage sur liste

A l'issue du prérecensement, un fichier de 4 664 logements a été constitué. Un échantillon a été tiré avec un taux de sondage de 40%. C'est sur cet échantillon de 1 856 logements que la collecte s'est déroulée.

Certains agents recenseurs n'ayant pu terminer leur travail en temps voulu, quelques IRIS ont été exclus de l'analyse qui a porté en définitive sur 1 618 logements.

Les conclusions du test quant à la localisation des logements de l'échantillon sont, pour l'essentiel, les suivantes :

- 1) Les agents recenseurs ont rencontré des difficultés pour repérer les logements à enquêter sur la base des informations issues du prérecensement mais ils sont parvenus dans l'ensemble à les surmonter.
- 2) Lorsqu'il figurait comme information complémentaire aux éléments de localisation, le nom de l'occupant du logement a principalement été utilisé pour ce repérage. Cette manière de faire est justifiée puisque peu de temps sépare le prérecensement du recensement.
- 3) Lorsqu'à une même adresse nom et localisation ne concordaient pas, la consigne était que le nom devait l'emporter. Cette non-concordance était en effet interprétée comme la conséquence d'une erreur au prérecensement (erreur de positionnement sur le palier,

confusion entre niveau d'habitation et étage) ou d'un prérecensement incomplet²⁸. Elle pose problème pour deux raisons : que vaut la localisation des logements pour lesquels il n'y a pas de nom ? et n'y a-t-il pas de risque de contentieux avec les communes à ce sujet ?

4) Le nombre de logements non trouvés est très faible : quelques unités sur l'ensemble de l'échantillon-liste. Parmi ces logements non trouvés, deux ont pu être repérés par les enquêteurs lors des contrôles.

5) Les contrôles font apparaître un très faible nombre d'erreurs de repérage, qu'il s'agisse de résidences principales, de résidences secondaires ou de logements vacants.

6) La disponibilité d'informations sur les logements voisins n'est pas jugée d'une grande utilité par les agents recenseurs pour repérer le logement et l'intérêt des schémas, qu'ils soient de bâtiment ou de palier, est controversé ; leur manque de lisibilité (a fortiori leur inexactitude) peut prêter à confusion.

La collecte proprement dite en mode liste pose également des difficultés :

1) L'argumentaire vis-à-vis des personnes enquêtées est plus délicat : il faut pouvoir répondre à la question « pourquoi moi et pas mon voisin ? » ;

2) L'obtention de renseignements sur le caractère vacant ou non d'un logement à recenser est plus difficile lorsqu'elle se fait auprès de voisins non enquêtés. Les agents recenseurs ont largement préféré le mode « adresses » et l'ont trouvé plus facile à réaliser ;

3) Le taux de réponse est un peu plus faible en mode liste, de l'ordre de trois points.

4) Les logements recensés en mode liste nécessitaient évidemment plus de déplacements que les logements recensés en mode grappe.

Les principales conclusions du test de sondage en grappe

Sur le plan du repérage des adresses, les difficultés rencontrées avec le plan de sondage en grappe sont comparables à celles mises en évidence pour le prérecensement du sondage sur liste.

Sur l'ensemble des 488 adresses de l'échantillon, après intervention des superviseurs en cas de difficulté, seules 3 adresses n'ont pu être retrouvées sur le terrain. Ce pourcentage d'échec (0,6 %) est très proche de celui du prérecensement (1%).

Ce bon résultat final, lié au travail d'expertise fait en direction régionale pour corriger des adresses que les agents recenseurs n'arrivaient pas à trouver, rend néanmoins mal compte des problèmes rencontrés quant à la qualité des adresses. Hormis les adresses imprécises, fort peu nombreuses dans l'échantillon, ces difficultés peuvent se répartir en diverses catégories :

1) Des erreurs de saisie dans l'adresse issue des documents de collecte établis par les agents recenseurs du recensement de 1999. De telles erreurs ont pu, pour certaines d'entre elles, être détectées et corrigées lors du géocodage qui a accompagné la création du répertoire d'immeubles localisés (RIL) ;

2) Des oublis d'indice de répétition lors de l'établissement des documents de collecte du recensement de 1999. Ces erreurs sont difficilement détectables sans retour au terrain. L'utilisation de fichiers administratifs d'adresses pour analyser la qualité du RIL permettra sans doute de repérer l'oubli de l'indice de répétition ;

3) Des numéros ou lettres de bâtiment, à l'intérieur d'une adresse, transformés en indice de répétition dans les documents de collecte du recensement de 1999. Ces divergences peuvent provenir d'une différence entre l'adressage théorique et l'adressage usuel. Là encore, une comparaison avec un fichier d'adresses devrait permettre de repérer ces anomalies.

²⁸ C'est le cas par exemple lorsque, n'ayant pu pénétrer dans un immeuble, l'agent prérecenseur a dressé la liste des logements de l'immeuble uniquement à partir des noms figurant sur les sonnettes ou sur le digicode, sans vérification de leur localisation.

D'une manière générale, la qualité de l'adresse indiquée par les habitants sur leur feuille de logement est meilleure que celle notée par les agents recenseurs sur les documents de collecte lors du recensement de 1999.

En ce qui concerne la collecte proprement dite, les principales conclusions sont les suivantes :

1) Au prérecensement, les difficultés de repérage des adresses s'étaient doublées de difficultés pour pénétrer dans les immeubles afin d'établir la liste des logements. Pour l'échantillon en grappe, les agents recenseurs ont rencontré des difficultés analogues mais les ont mieux résolues : sept adresses seulement n'ont pu être enquêtées, au surplus concentrées sur un seul agent recenseur sans qu'on ait su faire la part entre l'impossibilité dans laquelle il aurait été d'entrer dans l'immeuble et son abandon du recensement faute de temps.

La raison de cette meilleure performance du sondage en grappe sur le prérecensement (4,6% des adresses n'avaient pu être prérecensées) est simple : il est plus aisé de se faire ouvrir une porte en annonçant dans l'interphone qu'on vient réaliser un recensement de tous les logements de l'immeuble qu'en expliquant que l'on vient juste repérer les logements ;

2) La communication en direction des habitants est plus aisée en mode "adresses" qu'en mode "liste" ;

3) Le taux de réponse est un peu plus faible en mode liste, de l'ordre de 3 points, ce qui est insuffisant pour justifier sur ce seul élément un rejet du mode « liste » mais constitue un élément négatif. La manière d'organiser la collecte, le suivi des agents recenseurs, l'expérience et la disponibilité de ces derniers, sont autant de facteurs qui peuvent jouer sur la qualité de la collecte et ce bien avant les difficultés dues au bâti. La manière de sonder est secondaire dans l'explication de la qualité de réponse des ménages ou de catégorisation des logements.

Conclusion générale

Du point de vue de leur mise en œuvre, la comparaison entre les deux modes de sondage a conduit à préférer le sondage par grappe. En voici les raisons.

Le passage de l'unité immeuble à l'unité adresse a permis de résorber une partie des difficultés liées au repérage des bâtiments, mais cela joue de manière identique sur les deux modes de sondage. De même, les difficultés rencontrées avec les adresses imprécises ou erronées ne sont pas spécifiques à un mode de sondage.

En revanche, le sondage sur liste génère un biais important lors de la phase préliminaire de prérecensement : celui-ci ne peut être réalisé dans un nombre significatif d'adresses du fait de l'impossibilité de pénétrer dans les immeubles. De ce fait, le prérecensement introduit un biais au niveau de la création de la base de sondage des logements.

Ces difficultés sont bien connues des personnels de collecte, mais elles sont mieux résolues quand l'objectif est d'enquêter un ou plusieurs logements que lorsqu'il s'agit uniquement de décrire l'emplacement de ceux-ci.

Certes, le test a montré qu'il est possible, avec des agents recenseurs, de procéder à une enquête dont l'échantillon est composé de logements préalablement identifiés. Mais ce résultat n'a pu être obtenu qu'au prix d'une intervention de l'encadrement de collecte dans quelques cas de prérecensement litigieux et en donnant la règle de privilégier l'information relative au nom relevé au prérecensement, même dans le cas où elle est incohérente avec la localisation. Dans le cadre d'un partenariat de collecte avec la commune, une telle situation aurait pu alimenter la contestation sur la méthode lors de la validation de la collecte et, plus tard, de la publication des chiffres de population légale. De ce point de vue, le mode de sondage par liste ne répondait pas au critère de robustesse qui s'imposait pour la mise en

œuvre du nouveau recensement. De tels problèmes ne se rencontrent pas lors de la collecte avec le sondage par grappe.

Par ailleurs, dans le sondage par grappe, les agents recenseurs obtiennent plus aisément auprès des voisins des informations sur les personnes difficiles à joindre.

En conclusion, le mode liste apporte une réduction de la variance mais engendre un biais non maîtrisable. La précision globale s'en trouve donc affectée, dans une mesure certes difficile à apprécier mais de nature à fragiliser le processus de collecte. La décision a donc été prise de privilégier le tirage d'adresses complètes afin de renforcer la robustesse du dispositif et de faciliter les opérations de collecte. Le plan de sondage a été optimisé dans cette direction.

B.4.2.2.3 Mesures de charge en commune de 10 000 habitants ou plus (novembre et décembre 2001)

Le test devait permettre de mesurer les charges de travail pour les agents recenseurs et les agents chargés de contrôler la qualité de la collecte. Il a été organisé en Aquitaine, Bretagne et Ile-de-France de mi-novembre à mi-décembre 2001 et a porté sur un échantillon relativement important de 7 000 logements dans 16 communes différentes, soit environ 23 000 questionnaires.

Les résultats :

Les fiches de mesure du temps de travail remplies chaque semaine par les agents recenseurs indiquaient pour chaque phase les heures de début et de fin ainsi que le volume de documents traité (par exemple le nombre de feuilles de logement déposées puis récupérées). Elles ont fourni de précieuses indications. Ainsi, la charge de la tournée de reconnaissance était en général de 0,5 jour à 1,5 jour suivant l'étendue du secteur et la nature de l'habitat. En commune de 10 000 habitants ou plus, elles ont permis d'évaluer la charge de travail de l'agent recenseur à environ $\frac{3}{4}$ de temps pour les trois premières semaines et un mi-temps les deux dernières semaines.

Ces mesures détaillées ont nourri les recommandations faites plus tard aux communes pour la préparation des enquêtes de recensement.

Par ailleurs, ce test a permis de conforter l'ensemble des procédures de collecte dans cette catégorie de communes.

B.4.2.2.4 Définition du protocole de collecte en commune de moins de 10 000 habitants (report à 2002)

On a vu (cf. [§ B.4.2.1.3](#)) que le test qualitatif de février 2001 conduisait à penser qu'une liste de référence pouvait être mise en place dans les communes de moins de 10 000 habitants mais que seuls des tests quantitatifs permettraient de confirmer cette option et de définir le protocole de collecte en commune de moins de 10 000 habitants.

Il s'agissait de vérifier la faisabilité, dans ce type de communes, d'une tournée préparatoire de l'agent recenseur qui permettrait de contrôler l'exhaustivité de la collecte le plus en amont

possible et de procéder à la numérotation des questionnaires au moment de leur dépôt auprès des ménages.

Pour des raisons de moyens, les tests sur les protocoles de collecte dans les communes de moins de 10 000 habitants, initialement prévus en 2001, ont été reportés au printemps 2002 (voir ci-dessous [§ B.4.3.1](#)).

B.4.2.2.5 Evaluation des procédures de contrôle de la qualité de la collecte (novembre et décembre 2001)

Le test prévu était destiné à évaluer l'efficacité des procédures de contrôle de la collecte envisagées en bureau et sur le terrain et la charge de travail qu'elles représenteraient en bureau.

Les résultats :

Repérage des logements dans le sondage sur liste (pour mémoire puisque ce mode de sondage a été écarté)

Dès lors qu'un nom d'occupant du logement avait été enregistré par l'agent précenseur, le test a permis de vérifier dans 80% des cas la possibilité de contrôler le bon repérage du logement en utilisant le fichier de la taxe d'habitation et l'annuaire téléphonique.

Sur l'échantillon des résidences principales proposées au contrôle, 31% ne présentaient aucun nom d'occupant au précensement. Dans ce cas, seule une contre-enquête sur le terrain permettait de vérifier le bon repérage.

Il faut toutefois signaler que les contrôles en bureau peuvent être mis en défaut si l'agent recenseur invente de toutes pièces les questionnaires. Dans le test, cette fraude a été détectée car aucun nom n'avait été relevé pour ces logements au moment du précensement et que des enquêtes ont été menées sur le terrain. Si tel n'avait pas été le cas et si les questionnaires inventés avaient porté le nom enregistré lors du précensement, la fraude n'aurait pu être décelée.

Contrôle du remplissage des listes A, B et C de la feuille de logement

Un simple contrôle "sur pièces" des questionnaires permet dans près de 80 % des cas de conclure à une bonne répartition des personnes du ménage entre les différentes listes de la feuille de logement²⁹. Les erreurs sur les listes A, B et C sont exceptionnelles.

Les cas litigieux nécessitent une enquête terrain pour être résolus.

Contrôles en bureau

Les contrôles en bureau ont été réalisés en comparant les feuilles de logement ou les carnets de tournée aux fichiers de la taxe d'habitation (TH). Ces contrôles fonctionnent bien : pour 91,5 % des adresses, on constate un écart sur le nombre de logements nul ou compris dans les limites admises (elles-mêmes variables selon le nombre de logements de l'adresse).

Contrôles sur le terrain

Les contrôles sur le terrain ont porté sur les adresses non validées en bureau (8,5% des adresses)³⁰. L'entretien, quand il a eu lieu, a duré moins de dix minutes. Certains enquêteurs ont pu le réaliser par téléphone.

²⁹ Il aurait été utile de vérifier par enquêteur la validité de ce contrôle sur pièces, mais les moyens disponibles ne l'ont pas permis.

Les enquêteurs ont constaté quelques problèmes de localisation de logements et des erreurs dans la catégorie de logement : une part non négligeable des logements qualifiés de vacants sont en fait des résidences principales. A l'inverse, mais en proportion moins importante, certains logements dont les occupants supposés ont été classés en "impossible à joindre" sont en réalité des logements vacants.

Au total, les tests ont permis de valider les procédures de contrôle envisagées, d'en améliorer le détail et d'en optimiser le volume.

B.4.2.2.6 Evaluation des questionnements des deux principaux questionnaires (mai et juin 2001)

La compréhension du contenu des questionnaires par la population avait été testée en 2000. Il s'agissait en 2001 de valider ou d'amender plusieurs des questionnements envisagés dans la feuille de logement et le bulletin individuel à partir des taux de réponse et de la cohérence des réponses apportées par un large échantillon de personnes.

Les tests devaient notamment porter :

- sur les variables nécessaires aux traitements lourds que sont les codages de l'activité économique de l'établissement employeur en nomenclature NAF et de la catégorie socioprofessionnelle et de la profession en nomenclature PCS ;
- sur la qualité du remplissage des tableaux de composition des ménages à l'intérieur de la feuille de logement et de la réponse à la question sur la double résidence.

Les tests de 2001 ont été organisés avec des bulletins individuels de 2 pages, puisque le comité directeur de la rénovation du recensement avait décidé que telle serait leur taille maximale. Cela entraînait de fortes contraintes sur le nombre de questions et des choix ont dû être opérés pour les questionnaires proposés au test du printemps 2001 :

- suppression des questions :
 - o ayant un taux de non-réponse élevé au RP99 ;
 - o dont la compréhension n'était pas bonne au test de juillet ou de novembre 2000 et qui étaient difficilement améliorables (mode d'occupation du logement précédent) ou pouvaient l'être mais au prix d'un questionnement plus long (informations sur l'emploi antérieur pour les chômeurs) ;
 - o concernant un trop petit nombre de personnes (certaines pouvant faire l'objet d'enquêtes complémentaires réalisées auprès de la catégorie de population concernée) ;
- maintien des questions :
 - o structurelles des recensements précédents ;
 - o nouvelles ou modifiées, ne posant pas de problème de compréhension mais demandant à être confortées lors des tests quantitatifs (c'était le cas en particulier pour la question très délicate sur la double résidence) ;
 - o nécessaires pour coder dans les deux grandes nomenclatures (NAF et PCS), mais en version allégée. La pertinence du choix de ces questions, de leur positionnement et de leur formulation devait être testée ;

³⁰ Il est intéressant de noter que ces contrôles sont globalement bien perçus par la population, qui y voit la marque du « sérieux » de l'Insee.

- réintroduction d'une question sur le combustible de chauffage (celle-ci ayant été demandée parmi les premières priorités lors de la dernière réunion du groupe de travail du CNIS en octobre 2000, elle n'avait pas été intégrée lors des tests précédents).

Il était prévu que d'autres choix pourraient être proposés lors des tests de fin 2001.

Ces tests constituaient un élément essentiel pour les décisions de la maîtrise d'ouvrage du programme de rénovation du recensement de la population, dont il était prévu qu'elles seraient prises fin 2001 sur la base des avis et recommandations exprimés lors des consultations de 1999 et de 2000, des résultats des tests et des dernières consultations de l'année 2001.

Les résultats en termes de qualité de l'information collectée :

La catégorie de logement

La responsabilité de l'agent recenseur est directement engagée dans la détermination de la catégorie de logement. Un contrôle a donc été opéré pour vérifier la qualité de cette variable. Aucune erreur de catégorie n'a été repérée lors de ces contrôles pour les logements classés par l'agent recenseur comme résidence principale et pour lesquels il a pu obtenir une réponse des habitants.

Les erreurs de catégorisation des logements ne concernaient que les logements classés vacants ou résidences secondaires par les agents recenseurs. Sur les 78 logements de ce type dans l'échantillon sur liste, 12 ont été reclassés en résidence principale lors du contrôle. Rapporté au nombre total de résidences principales, le taux d'erreur était ainsi de 1,2 %. Ce taux est proche de celui observé à l'enquête postcensitaire du recensement de 1990 : 0,9 % pour l'ensemble des logements mais 1,3% pour les logements des communes appartenant à une unité urbaine de plus de 200 000 habitants (hors région parisienne).

Pour le sondage par grappe, le taux de logements classés à tort en résidence secondaire ou en logement vacant était du même ordre de grandeur que pour le sondage sur liste.

Le recensement des personnes

On s'intéressait ici au bon décompte des personnes "vivant habituellement dans le logement" et plus particulièrement à leur bonne répartition entre les listes A, B et C de la feuille de logement (voir [annexe F.6](#)).

Les agents recenseurs avaient consigné de vérifier qu'un bulletin individuel avait été rempli pour chaque personne inscrite en liste A et seulement pour ces personnes. Ils n'avaient pas à vérifier que le questionnaire était correctement rempli mais devaient s'assurer qu'il n'était pas complètement vierge. Aucun contrôle des informations concernant les personnes inscrites en liste B ou C ne leur était demandé.

Le contrôle de la qualité du classement entre les listes A, B et C des personnes vivant habituellement dans le logement a été mené sur un échantillon au 1/3 des résidences principales du seul échantillon-liste.

Sur les 335 résidences principales ainsi contrôlées, 13 (4%) présentaient une erreur sur au moins une des trois listes :

- 9 présentaient une erreur sur le nombre de personnes inscrites en liste A (habitants permanents du logement). Dans 6 cas , le nombre de bulletins individuels remplis ne correspondait pas à celui de la liste³¹ ;
- 3 sont en erreur sur le nombre de personnes en liste B³² ;

³¹ Cette non-concordance entre le nombre de personnes inscrites en liste A et le nombre de BI renseignés aurait dû être détectée par l'agent recenseur.

- 1 est en erreur sur le nombre de personnes en liste C³³.

Ce faible nombre d'erreurs était rassurant quant à la compréhension par la population mais ne permettait pas des analyses très fouillées. Quelques indications peuvent néanmoins être données :

Analyse des erreurs en liste A

Les erreurs repérées en liste A sont à peu près autant des oublis que des personnes inscrites à tort.

La place dans le ménage de la personne oubliée est variable : mari, grand-mère, enfant, personne elle-même. Ces personnes oubliées n'ont pas d'autre résidence que celle enquêtée, même lorsque qu'aucun bulletin n'a été rempli pour elles.

Les personnes inscrites à tort en liste A sont exclusivement des enfants ayant quitté la maison. Dans deux cas, un bulletin individuel a été rempli pour ces jeunes. La question sur la double résidence (incomplètement renseignée) ne permet pas de repérer l'erreur.

Analyse des erreurs en liste B

Les erreurs repérées en liste B sont :

- 1 cas d'erreur de liste : enfant inscrit à tort en liste A ;
- 3 cas d'inscription erronée en liste B : enfants ne faisant plus d'études, pouvant parfois relever de la liste C.

Analyse des erreurs en liste C

Les erreurs repérées en liste C sont des cas d'oubli (enfant inscrit en liste B, mari en maison de retraite). D'autres situations, complexes, ont été interprétées par l'enquêteur comme devant conduire à une inscription en liste C mais l'arbitrage, douteux, aurait mérité d'être explicité par l'agent recenseur.

B.4.3 Les tests de l'année 2002

Les tests opérationnels de l'année 2002 ont été consacrés à vérifier la capacité des différents acteurs à mettre en œuvre sur le terrain les propositions techniques élaborées par les concepteurs de la rénovation du recensement. Les actions correspondantes se sont déployées dans trois domaines :

- les tests initialement prévus en 2001 dans les communes de moins de 10 000 habitants ont eu lieu en mars ;
- un site expérimental constitué auprès de la direction régionale de l'Insee de Rhône-Alpes a permis d'approfondir avec quelques communes volontaires (Anse, Lyon 5e arrondissement et Villefranche-sur-Saône) la répartition des rôles entre les communes et l'Insee et d'affiner, grâce à des tests de collecte et d'organisation, les protocoles à mettre en place ;
- enfin, une série de tests ont eu lieu dans les DOM où les spécificités du plan de sondage appelaient diverses validations de procédures.

B.4.3.1 Définition du protocole de collecte en commune de moins de 10 000 habitants (mars 2002)

³² Une feuille de logement en erreur sur la liste A et sur la liste B est ici dénombrée dans les erreurs de la liste A.

³³ Ne sont dénombrées ici que les feuilles de logement en erreur uniquement sur la liste C.

Le protocole de collecte des communes de moins de 10 000 habitants a été testé entre le 25 février et le 20 mars 2002 auprès de 12 000 ménages enquêtés dans douze communes réparties dans six régions (Bourgogne, Corse, Franche-Comté, Languedoc-Roussillon, Lorraine et Basse-Normandie).

L'objectif du test était de vérifier la faisabilité d'un protocole de collecte qui permettrait de contrôler l'exhaustivité de la collecte le plus en amont possible, en totalisant les logements attendus (dénombrés au cours d'une tournée de reconnaissance faite par l'agent recenseur avant la collecte, sur le modèle de celle réalisée en commune d'au moins 10 000 habitants) et en comparant ce nombre aux résultats des recensements précédents et au nombre total de logements dans les rôles de la taxe d'habitation, de manière à disposer d'une référence pour le contrôle de l'avancement de la collecte. Etaient en outre testés l'établissement d'indicateurs hebdomadaires de suivi et la mise en place d'un contrôle post-collecte.

Les résultats :

Intérêt de la tournée de reconnaissance

L'utilité d'une tournée de reconnaissance avec relevé des adresses d'habitation et décompte du nombre de logements à l'adresse est apparue indiscutable. Cette étape préliminaire est essentielle pour la réussite de la collecte. Elle permet notamment à l'agent recenseur de bien repérer son secteur, d'organiser son travail avec un itinéraire logique, d'avoir une idée de sa charge de travail et de se faire connaître par les habitants et les commerçants du quartier. L'établissement d'un relevé d'adresses est au surplus nécessaire pour pouvoir numéroter les questionnaires au moment de leur dépôt.

Difficulté à apprécier le nombre de logements

Le décompte du nombre de logements à partir des boîtes aux lettres ou des sonnettes n'est pas toujours significatif.

Pour les résidences sécurisées, le nombre de logements est estimatif, il provient des informations des voisins : "dans l'enceinte, il y a à peu près tant de logements". L'estimation du nombre de logements est médiocre dans les centres-villes anciens où il y a des boîtes aux lettres non utilisées et des bâtiments commerciaux considérés comme abritant des logements.

Efficacité de la vérification sur le nombre de logements

La tournée de reconnaissance débouche sur un relevé des adresses et l'estimation du nombre de logements à chaque adresse. Les contrôles effectués sur le nombre de logements par secteur d'agent recenseur à partir des données du RP99 ont permis de déceler là où il y avait des manques ou des chevauchements de secteurs.

Dans le cas de rénovation du tissu urbain (centre-ville ancien réhabilité) ou de constructions nouvelles (ville pavillonnaire ou dortoir), la comparaison avec 1999 au niveau global ne suffisait pas : il fallait affiner la recherche au niveau de la voie pour contrôler le nombre de logements. Mais ce qui est possible au cours d'un test ne l'est pas au cours d'une enquête de recensement. En pareil cas, la connaissance du terrain est très importante et les communes ont cette connaissance.

La conclusion pratique du test a été de recommander aux communes de faire exécuter la tournée de reconnaissance entre les deux sessions de formation des agents recenseurs, sur une durée de trois à quatre jours au maximum, avec un premier rendez-vous pendant la tournée pour vérifier la bonne compréhension des consignes, puis une vérification

d'exhaustivité au retour des relevés avant de donner le feu vert pour la collecte et le remplissage du carnet de tournée.

Le test a permis par ailleurs de parfaire l'ergonomie du carnet de tournée et son adaptation aux communes de moins de 10 000 habitants et de poursuivre l'expérimentation d'un suivi de collecte par l'enregistrement des résultats de la collecte et l'établissement d'indicateurs d'avancement.

B.4.3.2 Site expérimental de la direction régionale de Rhône-Alpes

Le site expérimental mis en place de fin 2001 à l'été 2003 à la direction régionale de Rhône-Alpes de l'Insee avait pour raison d'être de servir de laboratoire aux communes et à l'Insee pour mettre au point les procédures et les organisations les plus appropriées à la réalisation des enquêtes de recensement par les soins des communes. Cette expérience originale a été menée avec le concours des communes de Lyon (5e arrondissement), Villefrance-sur-Saône et Anse.

Le site a donc été chargé de réaliser, en collaboration avec les trois communes volontaires, l'enchaînement des opérations de collecte de manière à contrôler la pertinence et l'efficacité de l'organisation prévue. Il ne s'agissait pas seulement, comme lors des tests, de lever des options techniques précises, mais aussi et surtout de s'assurer que la collaboration entre les communes et l'Insee pouvait se dérouler conformément au schéma convenu, avec un bon assemblage des applications informatiques.

Les enseignements tirés des travaux du site en 2001, qui avaient surtout porté sur la mise à jour du répertoire d'immeubles localisés (RIL), ont dicté les opérations à lui confier en 2002. On s'est efforcé d'y tester précisément l'organisation et les outils de suivi de la collecte, d'évaluer les charges incombant aux différents partenaires, de définir précisément les plans de formation des acteurs, et aussi d'effectuer une première validation de l'intégration des composants informatiques développés par l'Insee.

Le site expérimental a participé à de nombreux travaux. Il a notamment mis en place, début 2002, un test portant sur les trois communes du site, réalisé par les communes elles-mêmes dans les conditions de la loi du 27 février 2002.

Les objectifs de ce test étaient :

- de valider les hypothèses de travail (méthodologiques, statistiques, informatiques et organisationnelles) élaborées par l'Insee concernant la préparation et la réalisation des enquêtes de recensement en vérifiant leur acceptabilité par les communes et leur cohérence d'ensemble ;
- de valider les hypothèses de travail concernant le suivi de la collecte et le contrôle des enquêtes de recensement par l'Insee ;
- de mettre au point le référentiel d'organisation de la collecte qui régirait le partenariat entre les communes et l'Insee ;
- de mieux évaluer les charges des communes.

Le test devait permettre aux communes d'apprécier l'adaptation du **protocole cible**, d'en rechercher les modalités les plus efficaces et d'expérimenter des organisations.

En commune de 10 000 habitants ou plus, ce protocole était celui mis au point à la suite des tests réalisés par l'Insee en 2001. Il se caractérisait par :

- un sondage d'adresses entières, échantillonnées par l'Insee au sein d'un des cinq groupes de rotation d'adresses constitués dans la commune, de telle sorte que le nombre de logements recensés soit proche de 40% du nombre total de logements de ce groupe ;

- une tournée de reconnaissance préalable à la collecte, permettant à l'agent recenseur de vérifier l'existence sur le terrain des adresses à recenser et d'en estimer le nombre de logements ;
- le recensement de tous les logements de ces adresses.

En commune de moins de 10 000 habitants, le protocole mis en œuvre était très proche de celui du RP99. La modification portait sur la réalisation par l'agent recenseur, peu avant la collecte, d'une tournée de reconnaissance de chaque district de son secteur d'intervention.

Au cours de cette première phase, l'agent recenseur établit la liste des adresses de chaque district en mentionnant pour chacune le nombre indicatif de logements attendus.

La récapitulation au niveau district, éventuellement IRIS et commune, de ce nombre des logements attendus permet un premier contrôle d'exhaustivité par comparaison avec les résultats obtenus au précédent recensement³⁴ ou, pour ce qui est du seul niveau communal, avec les données de la taxe d'habitation. Outre cette indication sur l'exhaustivité de la collecte disponible le plus en amont possible, cette procédure permet de déterminer une "cible" à la collecte de chaque agent recenseur, par rapport à laquelle il serait possible de calculer des indicateurs d'avancement³⁵.

Le test devait permettre aussi d'expérimenter l'utilisation par les communes d'un **outil de suivi de la collecte**. Ce suivi consistait :

en commune de 10 000 habitants ou plus :

- à l'issue de la tournée de repérage, à pointer dans la liste des adresses fournies par l'Insee celles ne devant pas être recensées (démolitions, changement d'affectation, etc.) ;
- au cours de la collecte, à saisir, au fur et à mesure des retours de questionnaires par les agents recenseurs, le résultat de la collecte : catégorie de logement et, en cas de résidence principale, nombre de bulletins individuels (BI) remplis, non-réponse éventuelle analysée par nature ;
- selon un rythme hebdomadaire, à produire un état d'avancement par IRIS de la collecte réalisée par les agents recenseurs et à le transmettre, ainsi qu'un récapitulatif communal, à l'Insee ;
- en fin de collecte, à produire et transmettre à l'Insee, avec l'ensemble des questionnaires collectés, des états récapitulatifs par IRIS et au niveau de la commune : pour chaque adresse de la liste initiale, nombre de logements recensés par catégorie et nombre de questionnaires individuels collectés, nombre de fiches d'enquête non aboutie remplies ;

en commune de moins de 10 000 habitants :

- au cours de la collecte, à saisir, au fur et mesure des retours de questionnaires par les agents recenseurs, le résultat de la collecte : catégorie de logement et, en cas de résidence principale, nombre de BI remplis, non-réponse éventuelle analysée par nature ;
- selon un rythme hebdomadaire, à produire par district un état d'avancement de la collecte réalisée par les agents recenseurs et à le transmettre, ainsi que le récapitulatif communal, à l'Insee ;
- en fin de collecte, à produire et transmettre à l'Insee, avec l'ensemble des questionnaires collectés, des états récapitulatifs par district et au niveau de la commune : nombre de logements recensés par catégorie et nombre de questionnaires individuels collectés, nombre de fiches d'enquête non aboutie remplies.

³⁴ La comparaison au niveau district n'est possible que si la délimitation de celui-ci n'a pas changé dans l'intervalle.

³⁵ Ce protocole était envisagé pour les plus grandes de ces communes, entre 5 000 et 10 000 habitants.

Les questionnaires utilisés pour le test ont été ceux utilisés dans le cadre des tests de novembre 2001, modifiés en fonction des décisions prises par le comité directeur du programme de rénovation du recensement en conclusion de la concertation menée dans le cadre du groupe de travail du CNIS (voir chapitre F, [§ F.6.3.2](#)).

La mise en œuvre des procédures de collecte

Les deux communes de plus de 10 000 habitants (Lyon 5e et Villefranche-sur-Saône), ainsi que la commune de moins de 10 000 habitants (Anse) ont participé au test. La collecte d'informations auprès des logements³⁶ a été confiée à des agents recenseurs recrutés par chaque commune et dont l'encadrement a été assuré par les responsables communaux désignés par les maires. Le rôle des superviseurs (appui technique à la commune et contrôle de la collecte) a été tenu par les responsables du site au sein de la direction régionale de l'Insee. Les contrôles ont été centrés sur les défauts de couverture et sur la non-réponse.

Dans les deux communes les plus grandes, le test a porté sur près de 3 200 logements au total (dont 2 000 à Lyon 5e) et occupé 14 agents recenseurs durant 5 à 6 semaines ; dans la commune d'Anse, la collecte s'est déroulée sur une portion seulement de son territoire. Le test y a porté sur 750 logements et a occupé 3 agents recenseurs pendant 4 à 5 semaines.

La collecte a été réalisée entre fin février et fin mars 2002 et les contrôles ont eu lieu en avril. Ces contrôles combinaient l'utilisation du fichier de la taxe d'habitation et le recours à des enquêtes sur le terrain menées par des enquêteurs pigistes du réseau de l'Insee.

Ces contrôles ont porté en premier lieu sur les défauts de couverture :

- *en commune de 10 000 habitants ou plus* : vérification que toutes les adresses de la liste ont été effectivement enquêtées, qu'il n'y a pas eu d'erreur de repérage et qu'aucun logement n'a été oublié ;
- *en commune de moins de 10 000 habitants* : vérification que toutes les adresses du district ont été effectivement enquêtées et qu'aucun logement n'a été oublié.

En second lieu, et pour toutes les communes, des contrôles ont été opérés :

- sur la non-réponse (vérification de la qualité de résidence principale des logements correspondants et de la liste des personnes indiquées) ;
- sur la catégorie du logement ;
- et, pour les résidences principales, sur la répartition des habitants entre les listes A, B et C de la feuille de logement.

Parce qu'elles ont fait l'objet d'un dialogue continu avec les personnels que les communes avaient désignés pour travailler avec les agents de l'Insee, les actions du site expérimental ont grandement contribué à la mise au point de protocoles de collecte susceptibles d'être mis en œuvre par les communes sans leur poser de problèmes majeurs sur les plans administratif et organisationnel.

B.4.3.3 Tests dans les départements d'outre-mer (DOM) (janvier à septembre 2002)

Dans les départements d'outre-mer, une campagne de tests a eu lieu en 2002 en Guyane et à La Réunion. En effet, dans les communes de 10 000 habitants ou plus, les méthodes envisagées dans les DOM différaient trop sensiblement de celles prévues en métropole pour

³⁶ Il n'était pas prévu de collecte dans les communautés, ni auprès des personnes vivant dans une habitation mobile ou des personnes sans abri.

penser que les tests menés en métropole et les conclusions qui en seraient tirées seraient directement transposables. Les tests réalisés dans les DOM en 2002 n'ont donc concerné que les communes situées au-dessus du seuil. Ils visaient quatre objectifs principaux :

- s'assurer que le plan de sondage envisagé était opérationnel ;
- valider l'enchaînement des procédures allant des phases préparatoires (enquête cartographique de localisation des constructions dans les îlots du test, tirage de l'échantillon d'adresses) à la collecte elle-même (y compris son suivi et son contrôle) ;
- étudier la possibilité de réaliser la collecte par dépôt-retrait des questionnaires en s'assurant de la bonne compréhension des questionnaires par les personnes ;
- valider la possibilité de mettre à jour, voire de créer un répertoire d'adresses proche de celui de la métropole.

Ces tests ont été réalisés par l'Insee sans solliciter la participation des communes. La concertation sur le nouveau recensement se déroulait par ailleurs dans le cadre général de la concertation et faisait l'objet de consultations locales.

B.4.3.3.1 Les tests se sont déroulés en deux phases :

La première concernait le prérecensement dans deux communes et a porté sur 100 îlots et environ 3 000 logements dans chacune d'elles ; elle permettait de dresser la liste des constructions et d'alimenter le répertoire d'adresses.

La seconde phase était consacrée au recensement des occupants d'un échantillon de 40% de logements appartenant à 100 îlots parmi les 200 qui avaient été prérecensés. Cet échantillon représentait environ 1 200 constructions et 2 400 logements. Durant quatre semaines, cette phase de collecte devait permettre de vérifier la faisabilité technique du protocole de collecte proposé et en particulier de s'assurer du bon repérage des constructions-échantillons et de l'exhaustivité du recensement des logements dans chacune d'elles. Cette phase devait aussi permettre de juger de la bonne compréhension des questionnaires et de la pertinence du mode de collecte en dépôt-retrait à la place de l'interview (qui a été pratiquée jusqu'au RP99).

Hormis la feuille de logement dont les pages 1 et 4 diffèrent de celles de la version métropolitaine et le carnet de prérecensement qui est spécifique aux départements d'outre-mer, tous les autres documents utilisés pendant les tests étaient identiques à ceux utilisés en métropole. Le carnet de prérecensement permet de répertorier par district les constructions susceptibles d'abriter des locaux à usage d'habitation, avec pour chacune le nombre de logements attendus.

B.4.3.3.2 Le contexte dans les DOM

La taille des communes

Chaque DOM est composé d'un petit nombre de communes. De plus, et contrairement à la situation connue en métropole, leur taille est relativement importante.

La cartographie

Les outils cartographiques disponibles n'offrent pas encore les mêmes possibilités qu'en métropole. Ils diffèrent au surplus d'un département à l'autre ; en la matière, La Réunion est en avance sur ses homologues antillo-guyanais.

L'adressage

Une autre particularité réside dans la qualité de l'adressage. Une vaste opération de dénomination des rues et de numérotation des immeubles dans chaque voie est en cours. Outre le fait que cette opération est très longue, les habitudes locales en matière d'adressage privilégient encore beaucoup les lieux-dits, les mornes ou autres écarts. La méthode de sondage retenue en métropole ne pouvait se transposer à l'identique aux DOM.

Le suivi de la construction neuve

Le développement rapide de la construction dans les DOM en rend très difficiles la maîtrise et le suivi par les services qui en ont la charge. Même si des progrès sont enregistrés et des espoirs permis, les opérations de collecte des recensements doivent nécessairement y être précédées par un prérecensement³⁷ qui permet de repérer l'ensemble des constructions. Ce prérecensement guide ensuite la collecte et en permet entre autres le contrôle d'exhaustivité.

Le mode de collecte

Alors qu'en métropole l'agent recenseur procède par dépôt-retrait, l'agent recenseur des DOM pratiquait, jusqu'au recensement de 1999 inclus, par interview, comme pour une enquête-ménage classique. Cette façon de faire s'expliquait par le nombre jugé élevé de personnes qui éprouveraient des difficultés à comprendre voire lire les questionnaires.

Le questionnement

Les questionnaires eux-mêmes diffèrent de ceux utilisés en métropole. En effet, les DOM ne réalisent pas la totalité des enquêtes auprès des ménages. Il s'ensuit un déficit d'informations sur le confort des logements et l'équipement des ménages. Traditionnellement, le recensement comblait pour partie ces manques en adaptant le questionnement au contexte des DOM.

Le bulletin individuel (BI) ne différait lors du recensement de 1999 que sur une question. Il s'agissait de la date d'arrivée dans le département (respectivement en métropole) pour les personnes nées hors du département (respectivement à l'étranger).

La feuille de logement (FL) intégrait quant à elle des questions sur le confort du logement et l'équipement du ménage. Les premières se justifiaient par des taux d'équipement n'ayant pas encore atteint ceux connus en métropole et nécessitant donc encore un suivi par le biais du recensement, les dernières par l'absence d'enquêtes auprès des ménages.

B.4.3.3.3 Les adaptations méthodologiques

Les adaptations envisagées pour le nouveau recensement dans les DOM concernaient donc le questionnement, le plan de sondage et la collecte.

Le questionnement

Deux solutions alternatives étaient envisagées pour la feuille de logement.

La première visait à retenir le même dessin de questionnaire qu'en métropole mais en transformant les questions peu adaptées (par exemple : mode de chauffage, principal combustible utilisé).

³⁷ Le terme de prérecensement utilisé ici correspond à la notion suivante. Le prérecensement vise à repérer sur le terrain l'ensemble des constructions pouvant abriter des logements. L'agent prérecenseur dispose des plans du secteur concerné et d'un carnet de tournée. A l'issue de ce travail, il aura complété les plans en y reportant les constructions neuves, en supprimant les constructions détruites, en plaçant les nouvelles voies ou places. Le carnet de tournée est lui aussi complété par diverses informations : le numéro d'ordre de la construction, sa localisation précise, le nombre de logements attendu, le type de construction. Entre deux constructions successives, de numéro d'ordre N et N+1, l'agent prérecenseur laisse libre autant de lignes du carnet qu'il y a de logements rattachés à la première.

La seconde consistait à concevoir une maquette propre aux DOM, dans l'esprit de ce qui avait effectué lors des précédents recensements. La maquette comporterait un tronçon commun et une partie spécifique. Cette dernière intégrerait des questions permettant d'obtenir des informations sur l'équipement du logement et du ménage, dont le suivi est précieux mais qui sont extérieures au champ habituel d'un recensement.

Le bulletin individuel pouvait, pour sa part, être identique à celui de métropole sous réserve de modifier la seule question qui différait en 1999. C'est le choix qui a été fait en retenant pour les DOM la version adoptée pour la métropole (« Si vous êtes né(e) à l'étranger, en quelle année êtes-vous arrivé(e) en France ? »).

Le plan de sondage

Des éléments signalés dans le paragraphe précédent, il ressortait la nécessité d'adapter le plan de sondage et la collecte au contexte des DOM.

En particulier, le plan de sondage devait connaître plusieurs adaptations³⁸.

Les communes situées au-dessous du seuil de 10 000 habitants étant peu nombreuses, elles sont réparties en cinq groupes de rotation de telle sorte que, chaque année, la population à recenser dans cette strate soit de l'ordre du cinquième de la population totale de la strate dans le département considéré.

Pour les communes situées au-dessus du seuil, compte tenu des problèmes d'adressage exposés précédemment, l'hypothèse majeure à tester était que les groupes de rotation ne seraient plus constitués d'adresses mais d'îlots du recensement de 1999. Chaque année, un groupe de rotation d'îlots serait prérecensé, c'est-à-dire qu'on relèverait sur un plan la localisation des constructions. Parmi les constructions ainsi localisées et contenant des locaux à usage d'habitation, il serait tiré un échantillon d'adresses représentant 40% des logements. Une enquête de recensement serait ensuite effectuée dans tous les logements de chacune des constructions échantillonnées.

La collecte

Les sources administratives et les évolutions sur le terrain en matière d'adressage amenaient à maintenir dans toutes les communes, quelle que soit leur taille, un prérecensement préalable aux opérations de collecte.

Enfin, l'évolution de la société observée dans les DOM permettait d'envisager le dépôt-retrait comme méthode principale de collecte, ce qui n'empêcherait pas l'agent recenseur de remplir lui-même les questionnaires en cas de nécessité.

B.4.3.3.4 Le déroulement du programme de tests

Les tests ont été menés en Guyane et à La Réunion. Dans chacun de ces départements, une commune a été retenue (Cayenne en Guyane et Saint-Paul à La Réunion). Le programme s'est déroulé en deux phases, la première portant sur les opérations préalables à la collecte, la seconde sur la collecte.

Les opérations préalables, échelonnées de janvier à avril 2002 en Guyane et de février à mai à La Réunion, se décomposaient ainsi :

³⁸ Pour mémoire, en métropole, les communes de moins de 10 000 habitants sont réparties en cinq groupes de rotation équilibrés sur le sexe et l'âge au niveau de la région. Chaque année un groupe de rotation est recensé exhaustivement. Dans les communes de 10 000 habitants ou plus, les adresses comprenant des locaux à usage d'habitation sont réparties entre cinq groupes de rotation équilibrés sur le sexe et l'âge. Chaque année on interroge un échantillon des adresses d'un groupe de rotation de telle sorte que la charge de collecte soit de 40% des logements du groupe. La constitution de ces groupes de rotation s'appuie sur la disponibilité d'un répertoire des immeubles localisés (RIL). La mise en œuvre de ce plan de sondage nécessite de pouvoir retrouver sans ambiguïté sur le terrain une adresse tirée dans l'échantillon.

- répartition des îlots en cinq groupes de rotation ;
- tirage d'un des groupes d'îlots ;
- préparation des documents nécessaires au prérecensement (plans, carnet de prérecensement) ;
- recrutement et formation des agents prérecenseurs ;
- prérecensement (annotation des plans, remplissage du carnet de prérecensement) ;
- tirage des adresses à recenser ;
- report de cette information sur les documents mis à la disposition de l'agent recenseur (plans, carnet de tournée, etc.).

Durant cette phase, des mesures de charge ont été effectuées ainsi qu'une estimation des coûts, en particulier pour l'impression des plans.

A la lumière de l'expérience du recensement de 1999, un protocole de suivi et de contrôle du travail de l'agent prérecenseur avait été préalablement élaboré. L'agent prérecenseur était accompagné sur le terrain par un agent de l'Insee afin de repérer les limites du secteur qui lui était confié et celles de chaque îlot le composant. Le contrôle du travail consistait à valider exhaustivement le prérecensement du premier îlot puis par sondage celui des autres îlots. Dans ceux-ci, les constructions de plus de dix logements étaient vérifiées systématiquement, ainsi qu'une sur cinq pour les constructions comportant moins de dix logements. Ces contrôles s'effectuaient par un retour sur le terrain et pouvaient amener l'agent prérecenseur à reprendre tout ou partie de son relevé en cas d'erreurs importantes.

A cette occasion, les informations nécessaires à la constitution et à la mise à jour d'un répertoire d'adresses étaient collectées (l'adresse, le nombre de logements, l'aspect du bâtiment, le type de construction, etc.). Ces informations permettaient d'alimenter le répertoire d'adresses et de préparer les étapes suivantes.

Ces opérations préalables ont porté sur 200 îlots regroupant 6 000 constructions et ont occupé 20 agents prérecenseurs pendant 1 semaine.

La deuxième phase a porté, en mai et juin 2002 en Guyane et en août et septembre 2002 à La Réunion, sur la réalisation des enquêtes de recensement au sein des immeubles retenus dans un sous-échantillon de districts.

Cette phase a permis de valider la procédure de travail de l'agent recenseur pour le repérage des immeubles échantillonnés et l'exhaustivité du recensement de leurs logements. Le contrôle du travail de l'agent recenseur a été effectué par du personnel de l'Insee grâce à des retours sur le terrain.

Par ailleurs, le bon remplissage des questions sur l'aspect des constructions, la composition du ménage et le type de logement a été vérifié sur un échantillon de ménages.

Cette phase a permis aussi de juger de la bonne compréhension des questionnaires par les personnes recensées et de la pertinence du passage de l'interview au dépôt-retrait des questionnaires comme mode principal de collecte.

Ce test de collecte a concerné 100 îlots, 1 200 constructions abritant 2 400 logements et il a occupé 20 agents recenseurs pendant 4 semaines.

Au total, les tests effectués ont permis de valider les procédures de sondage et de collecte dans les DOM et ont conforté les choix opérés quant au contenu des questionnaires et à leur mode d'administration (dépôt-retrait avec aide de l'agent recenseur en cas de besoin).

B.4.4 Les tests de l'année 2003

L'année 2003 a été marquée par une répétition effectuée avec plus de 90 communes volontaires dans les conditions des futures enquêtes de recensement (qui devaient commencer en 2004) et par un test d'enquête dans les communautés.

B.4.4.1 La répétition (janvier et février 2003)

Les tests de 2000, 2001 et 2002, réalisés sans implication des communes (à l'exception en 2002 de celles du site expérimental), avaient pour objectif de préciser les modalités du plan de sondage dans les communes de 10 000 habitants ou plus et les modes opératoires de collecte, quelle que soit la taille de la commune.

Une nouvelle série de tests ayant valeur de répétition a été organisée en janvier et février 2003, cette fois dans les conditions d'une véritable enquête de recensement, c'est-à-dire avec la pleine implication des communes volontaires pour réaliser ces tests.

Les objectifs visés et les modalités de mise en œuvre des tests de 2003 sont présentés ci-après.

B.4.4.1.1 Les objectifs de la répétition

Les tests des années précédentes avaient permis de valider les procédures de collecte. Ceux de 2003 devaient avant tout tester une dernière fois les protocoles de collecte et les imprimés correspondants, pour les communes de moins de 10 000 habitants d'une part et pour les communes de 10 000 habitants ou plus d'autre part, et vérifier la conformité aux spécifications des premiers livrables informatiques des projets du domaine "Collecte", à savoir :

- pour le projet "Pilotage" : la programmation, le suivi des opérations et la remontée d'indicateurs à différents niveaux d'agrégation (commune, zone de supervision, direction régionale de l'Insee et niveau national) ;
- pour le projet "Collecte-flux de questionnaires" : le poste de travail du superviseur sur micro-ordinateur portable, le poste de suivi de la collecte en commune et le poste de réception des questionnaires à l'Insee en fin de collecte.

Ces tests portaient aussi sur les formations des différents acteurs de la collecte (agents recenseurs, coordonnateurs communaux et superviseurs de l'Insee) et sur le rodage des organisations en direction régionale et dans les communes. Ils permettaient enfin de tester la dernière version des questionnaires, en particulier le bulletin individuel qui avait fait l'objet d'ultimes retouches quant à l'ordre de certaines questions pour diminuer la non-réponse observée lors des tests précédents.

Avec ces objectifs, les tests de 2003 ne nécessitaient pas une répétition générale impliquant toutes les directions régionales de l'Insee et de très nombreuses communes, ce qui de toute façon aurait excédé les moyens disponibles. Il fallait toutefois impliquer un nombre suffisant de communes et de directions régionales de l'Insee pour valider les outils mais ce nombre devait être limité afin que les questions techniques de déploiement des applications informatiques sur plusieurs sites ne viennent pas perturber l'observation des validations fonctionnelles et organisationnelles recherchées.

Il fallait également un nombre significatif d'acteurs, agents recenseurs et superviseurs, pour simuler correctement les consolidations dans l'application "Pilotage" ainsi que des questionnaires en nombre suffisant pour pouvoir tester l'application de contrôle des retours à l'Insee.

B.4.4.1.2 Les modalités de la répétition

Les tests ont été réalisés dans 92 communes (49 de 10 000 habitants ou plus, 43 de moins de 10 000 habitants) réparties dans 11 régions de métropole et à la Guadeloupe. Les régions de métropole concernées furent les suivantes : Alsace, Auvergne, Champagne-Ardenne, Limousin, Lorraine, Midi-Pyrénées, Nord - Pas-de-Calais, Pays-de-la-Loire, Picardie, Poitou-Charentes, Provence-Alpes-Côte d'Azur. Parmi elles, avaient été retenues celles qui n'avaient encore jamais participé à un test. Il convenait, par ailleurs, d'avoir un site dans les départements d'outre-mer pour tester les conditions spécifiques de mise en œuvre dans cet environnement. Au total, toutes les directions régionales auront ainsi participé à un test de recensement rénové avant la première enquête de recensement en 2004.

Dans chaque région, huit communes environ ont été sollicitées pour réaliser dans les conditions prévues par la loi les enquêtes-tests de recensement. Pour s'assurer du bon fonctionnement des applications testées dans différents environnements, le choix des communes respectait la répartition indicative ci-dessous en fonction de la taille, sans pour autant viser un échantillon représentatif.

Catégorie de communes	Nombre par région
10 000 habitants ou plus	4
de 5 000 à 9 999 habitants	2
de 1 000 à 4 999 habitants	1
de moins de 1 000 habitants	1

Dans chacune des communes, une partie du territoire seulement a été recensée, représentant en moyenne 400 logements, soit une charge de travail nécessitant le recrutement de deux agents recenseurs. Ce faible nombre d'agents recenseurs par commune ne posait pas de problème puisqu'aucune application informatique dont on prévoyait le test n'était dépendante de ce nombre. Pour les communes, la logistique à mettre en œuvre était légère mais, du même coup, les tests ne permettaient pas d'expérimenter, au delà de ce qui avait été fait fin 2002 dans le cadre du site expérimental de la région Rhône-Alpes, les organisations plus lourdes que les communes auraient à mettre en place pour la première enquête de recensement début 2004.

Les questionnaires utilisés ont été ceux arrêtés par l'Insee à la suite de la concertation menée au sein du CNIS et après validation des évolutions jugées souhaitables après les tests réalisés les années précédentes.

La collecte a commencé le 13 janvier 2003. Elle avait été précédée d'une tournée de reconnaissance au cours de la semaine du 6 au 11 janvier. Sa durée a été de quatre semaines pour les communes de moins de 10 000 habitants et de cinq semaines dans les communes au dessus du seuil de 10 000 habitants.

A la Guadeloupe, l'enquête cartographique précédant la collecte et nécessaire à la constitution de l'échantillon d'adresses à recenser s'était déroulée sur deux semaines à partir du 1^{er} octobre 2002.

A la suite de la collecte, des contrôles ont été opérés par l'Insee. Utilisant des fichiers administratifs ou reposant sur des enquêteurs de l'Institut, ces contrôles ne visaient pas le remplissage des questionnaires mais seulement l'exhaustivité de la collecte pour les communes de moins de 10 000 habitants et le bon repérage de l'adresse dans les communes de 10 000 habitants ou plus.

Au total 35 549 logements ont été recensés pour une collecte de 70 411 bulletins individuels.

Pour deux communes, la collecte a été prise en charge par un établissement public de coopération intercommunale (EPCI). Bien que tous les textes réglementaires régissant le nouveau recensement n'aient pas été encore publiés à la date du test, tous les acteurs en ont préfiguré l'application.

B.4.4.1.3 Les enseignements de la répétition

A l'issue de ces tests, un bon nombre de recommandations ont pu être formulées pour la mise en œuvre de la première vague d'enquêtes de recensement en janvier et février 2004. Les principales sont retracées ci-après :

- axer la communication en direction des élus communaux sur le nouveau rôle attribué aux communes pour la collecte des enquêtes de recensement et les obligations qui en résultent ;
- renforcer la formation des responsables communaux sur le soin à apporter au recrutement des agents recenseurs et sur l'importance de l'examen du carnet de tournée pour le suivi des agents recenseurs pendant la collecte ;
- donner une indication de la charge de travail par agent recenseur ;
- s'appuyer sur les associations de communes pour faire circuler entre les communes des exemples de pratiques de rémunération ;
- décrire finement le travail demandé à un agent recenseur ;
- insister, dans la formation des superviseurs, sur la spécificité de leur rôle qui diffère de celui des délégués du recensement de 1999 (en particulier, ils n'ont pas à donner d'instructions aux personnels des communes ni à encadrer les agents recenseurs) ; insister, notamment, sur la nature et la pratique de leur mission de contrôle en cours de collecte ;
- veiller à une bonne articulation entre campagne de communication nationale et communication locale ;
- mettre en place des dispositifs de soutien de l'Insee au superviseur en cas de mauvaise prise en charge de la collecte par une commune ;
- revoir la présentation de la lettre d'annonce du recensement à la population : il est souhaitable qu'elle soit accompagnée d'une lettre personnalisée du maire de la commune ou du président de l'EPCI ;
- attirer l'attention du maire sur les conditions d'utilisation de fichiers nominatifs dans le cadre de cette communication aux habitants ;
- expliciter le mode de calcul de la dotation forfaitaire de recensement et son mode d'actualisation par rapport au recensement de 1999 ;
- prévoir un temps suffisant pour la tournée de reconnaissance, dont l'utilité est apparue amplement confirmée par ces tests ;
- dans les communes de moins de 10 000 habitants, insister, lors de la formation des superviseurs et des agents recenseurs, sur l'utilité du relevé d'adresses ;
- dans toutes les communes, insister sur l'importance des indicateurs d'avancement de la collecte et de leur transmission hebdomadaire à l'Insee ;
- dans les communes de 10 000 habitants ou plus, recenser tous les bâtiments existant dans une voie d'adresses non normalisées si cette voie fait partie de l'échantillon ; fournir à toutes ces communes en 2004 des plans par IRIS ;
- instaurer une fiche-navette pour les échanges d'informations entre la commune et l'Insee, notamment pour consigner les décisions retenues en ce qui concerne le repérage des adresses ;
- adapter le calendrier en faisant démarrer la collecte le 3e jeudi de janvier et en la faisant se terminer le cinquième ou le sixième samedi suivant la date de début selon qu'il

s'agit d'une commune de moins de 10 000 habitants ou d'une commune de 10 000 habitants ou plus, ce qui se traduit par une durée effective de quatre semaines plus trois jours ou de cinq semaines plus trois jours selon la catégorie d'appartenance de la commune ;

- insister lors de la formation des agents recenseurs sur les raisons de la numérotation des questionnaires lors de leur dépôt ;
- insister sur le fait que la fiche de logement non enquêté est à établir uniquement pour une résidence principale.

Par ailleurs, les tests de 2003 ont permis d'apporter de très nombreuses améliorations à la présentation et au contenu des imprimés de gestion de la collecte.

Ils ont conduit, dans le domaine des applications informatiques, à différer à 2005 la mise en œuvre de certaines des fonctionnalités de l'application "Pilotage" (voir l'[encadré](#) correspondant, au chapitre C, § C.1.1), telles que la communication entre les communes et l'Insee via l'internet, la synchronisation automatisée de l'ensemble des postes de travail des acteurs de la collecte ou la messagerie au bénéfice des superviseurs. Mais ils ont confirmé l'utilité de l'outil de suivi de l'avancement de la collecte mis à la disposition des communes et éclairé les améliorations à apporter en termes de fonctionnalités ou d'ergonomie à plusieurs applications.

B.4.4.2 Le test auprès des communautés (juin 2003)

B.4.4.2.1 Objectifs du test

Un test auprès des communautés a été organisé en 2003, qui visait deux objectifs majeurs :

- la mise au point de la grille de rémunération qui serait proposée aux enquêteurs recrutés par l'Insee pour les opérations de collecte. Il s'agissait donc de mesurer, à travers différents cas, les temps d'interviews et les difficultés rencontrées, ainsi que la charge pour l'enquêteur et pour la communauté ;
- la mise au point du protocole de collecte et du livret d'instructions pour les enquêteurs.

Les informations collectées devaient aussi servir à mettre au point les procédures de traitement des données en aval de la collecte (codification de certaines variables, définition du poste de gestion en direction régionale, test de la prestation de lecture automatisée des données).

B.4.4.2.2 Modalités du test

Le test a eu lieu du 2 au 18 juin 2003 auprès d'une soixantaine de communautés dans quatre régions (Lorraine, Franche-Comté, Haute-Normandie et Provence-Alpes-Côte d'Azur) et a été précédé de quelques enquêtes réalisées mi-mai 2003 par ses responsables en vue de préparer les formations des enquêteurs. Chaque direction régionale encadrait deux enquêteurs qui avaient chacun huit communautés à traiter. Les communes concernées étaient Besançon, Metz, Marseille, Aix-en-Provence et Rouen.

La taille de l'échantillon avait été calculée de manière à répondre à une double contrainte : enveloppe budgétaire définie et obtention de renseignements suffisamment représentatifs pour caractériser les conditions de collecte auprès des différentes catégories de communautés.

La communauté, telle qu'elle est définie par l'article 1, § V, du décret n° 2003-485 du 5 juin 2003 (voir [annexe E.5](#)) est une entité très hétérogène : en 1999, 20% des communautés comprenaient moins de 5 personnes, 12% plus de 100 personnes (avec un effectif maximal de 3 000 personnes dans un établissement pénitentiaire).

Les communautés enquêtées lors du test ont été choisies de préférence parmi les communautés des catégories 1 (services de moyen ou de long séjour des établissements publics ou privés de santé, établissements sociaux de moyen et long séjour, maisons de retraite, foyers et résidences sociales ou assimilés), 4 (établissements hébergeant des élèves ou des étudiants, y compris les établissements militaires d'enseignement) et 6 (établissements sociaux de court séjour) qui sont non seulement les plus importantes en nombre mais aussi, pour certaines d'entre elles, plus difficiles à enquêter. Il convenait en effet de tester, selon les catégories de communautés, les organisations propres à faciliter le travail des enquêteurs et à réaliser les opérations de collecte avec la meilleure qualité possible.

Les personnes enquêtées étaient les membres des communautés et les personnes vivant dans les logements de fonction situés dans l'enceinte de celles-ci.

Le mode de collecte était, comme pour les enquêtes auprès des ménages, le dépôt-retrait des questionnaires. Toutefois, l'expérience montre que cette méthode n'est pas adaptée à tous les cas ; il faut parfois procéder à une enquête directe auprès des personnes.

B.4.4.2.3 Questionnaires utilisés

Les questionnaires individuels destinés aux membres des communautés sont de deux types :

- le bulletin individuel utilisé dans les établissements pénitentiaires (communautés de catégorie 5). Ce bulletin est un simple recto, ne comportant qu'un questionnement très allégé sur l'activité professionnelle antérieure. Il se distingue aussi du bulletin individuel des ménages par la suppression de la question sur la vie en couple et l'absence de question sur la résidence personnelle. Il comporte un identifiant (celui de la communauté), qui est renseigné par l'enquêteur ;
- le bulletin individuel utilisé dans toutes les autres catégories de communautés. Ce bulletin est identique au bulletin individuel des ménages (recto-verso), à deux différences près :
 - l'ajout d'une question concernant l'éventuelle résidence personnelle de la personne interrogée (nom de la commune et du département) ;
 - un identifiant de la communauté en tête du questionnaire, permettant d'attribuer chaque personne à la bonne communauté.

B.4.4.2.4 Résultats du test

Le test a répondu aux attentes et a permis de mettre au point l'ensemble des procédures de collecte qui ont été appliquées pour la première fois en mars 2004 (janvier à La Réunion).

ANNEXE B 1

Les échantillons équilibrés

Il existe divers moyens d'améliorer l'efficacité d'un sondage, c'est-à-dire de réduire la variance des estimateurs. Les plus connus sont la stratification et le tirage à probabilités inégales. D'une façon générale, on cherche à constituer un échantillon qui reflète du mieux possible l'hétérogénéité de la population sondée, en mobilisant sur cette même population des observations auxiliaires, que l'on suppose liées aux variables explorées dans le sondage.

Soit Y une variable d'intérêt du sondage et $T(Y) = \sum_{i \in U} Y_i$ son total pour la population de référence U . Un plan de sondage conduit à un estimateur $\hat{T}(Y)$ pour $T(Y)$ et à l'intervalle de confiance de cet estimateur, intervalle déduit de sa variance $V(\hat{T}(Y))$. L'expression de l'estimateur est fournie par la formule de Horvitz-Thompson :

$$\hat{T}(Y) = \sum_{i \in s} \frac{Y_i}{\pi_i}$$
, où s désigne l'ensemble des unités de l'échantillon, Y_i la valeur observée pour l'unité i et π_i sa probabilité d'inclusion accordée par le plan de sondage.

L'idée des plans équilibrés repose sur le fait qu'il existe des variables auxiliaires Z_k , dont on connaît exactement les totaux $T(Z_k)$ pour la population. On observe donc aussi ces variables dans le sondage et on regarde si l'échantillon reconstruit bien ces données connues. En d'autres termes, l'application de la formule de Horvitz-Thompson aux valeurs observées Z_{ki} devrait permettre de retrouver les totaux $T(Z_k)$:

$$\hat{T}(Z_k) = \sum_{i \in s} \frac{Z_{ki}}{\pi_i} = \sum_{i \in U} Z_{ki} .$$

On dit alors que l'échantillon est équilibré pour les variables Z_k . L'hypothèse naturelle est que ce qui est bon pour les Z_k doit l'être aussi pour Y compte tenu des corrélations qu'il y a entre elles.

Les plans de sondage « traditionnels » ne permettent pas d'atteindre l'adéquation voulue et on a donc recherché des méthodes pour y parvenir. Des solutions approximatives ont été avancées par Yates (1949), Thionet (1953), Deville, Grosbras et Roth (1988). Plus récemment, Jean-Claude Deville et Yves Tillé (2000) ont développé un algorithme, dit *méthode du cube*, qui présente de grands avantages :

- la méthode trouve la solution exacte au problème quand elle existe ;
- elle donne une approximation optimale quand une solution exacte est mathématiquement impossible ;
- elle fournit, dans la plupart des cas, un procédé simple de calcul pour les estimateurs et leur variance.

Références :

DEVILLE J.-C., GROSBRAS J.-M., ROTH N., 1988, « Efficient sampling algorithms and balanced sample », *COMPSTAT, Proceedings in computational statistics*, Physica-Verlag, P. 255-266.

DEVILLE J.C., TILLÉ Y., 2000, « Échantillonnage équilibré par la méthode du cube, variance et estimation de variance », *Enquêtes, modèles et applications*, sous la direction de Jean-Jacques DROESBEKE et Ludovic LEBART, DUNOD, p. 344-362.

ANNEXE B 2

Le plan de sondage dans les communes de 10 000 habitants ou plus

Unités échantillonnées, base et taux de sondage.

L'unité d'échantillonnage est l'adresse d'habitation, tous les logements d'une adresse appartenant à l'échantillon devant être recensés. Ce choix a été dicté par la nécessité de désigner sans ambiguïté à l'agent recenseur les logements qu'il doit recenser. Or, les tests effectués ont montré que toute autre notion, comme celle d'immeuble ou même de logement à l'intérieur d'un immeuble, introduisait des risques non négligeables de non repérage ou de substitution sur le terrain par l'agent recenseur. Cette contrainte est forte puisqu'on doit prendre en compte le niveau infracommunal, niveau auquel il faut produire des estimations d'une bonne précision.

Le taux de sondage retenu est tel qu'au terme d'une période de cinq ans 40% des logements de chaque commune ont fait l'objet d'une enquête de recensement. Les données recueillies au cours de ces cinq années sont combinées pour élaborer des résultats valides pour l'année médiane de cette période, extrapolés à l'intégralité des logements existant dans la commune cette même année (connus grâce au répertoire d'immeubles localisés - RIL). Le processus est glissant et se décale d'un an tous les ans.

La base de tirage des échantillons d'adresses (le RIL) pose deux problèmes statistiques pour la production des résultats : les effets de grappe inhérents aux adresses et la qualité des informations annuelles actualisant la base de tirage. Cela conduit à constituer trois strates parmi les adresses d'habitation : les adresses dites « de grande taille », les adresses dites « nouvelles » et les autres adresses.

Effets de grappe : le traitement des adresses de grande taille

Le problème majeur à résoudre est la variance de la taille des unités à échantillonner. En effet, la présence d'une adresse contenant parfois jusqu'à plusieurs dizaines de logements entraîne des effets de grappe : les estimations communales et infra-communales pour certaines variables peuvent être très sensibles à la présence ou non de ces adresses dans l'échantillon. Par exemple, dans le cas d'un échantillon aléatoire simple de grappes, on sait que la variance de l'estimateur du total d'une variable Y s'écrit :

$$V(\hat{T}(Y)) = M^2(1-t) \frac{S_g^2}{m}$$

M est le nombre total de grappes, m le nombre de grappes de l'échantillon, t le taux de sondage et S_g^2 la variance inter-grappes, c'est-à-dire :

$$S_g^2 = \frac{1}{M-1} \sum_i (Y_i - \bar{Y})^2 \text{ où } Y_i \text{ est le total de la variable dans la grappe } i.$$

Cette formule montre que la variance peut être élevée si la taille des grappes est très disparate.

C'est pourquoi il a été décidé de créer une strate particulière constituée des adresses comptant un nombre élevé de logements. L'enquête de recensement sera exhaustive dans cette strate au cours d'une période de cinq ans. Il n'y aura donc pas de composante due à l'échantillonnage dans le calcul de la variance au sein de cette strate pour les estimations détaillées. Ce choix a pour avantage déterminant d'améliorer la précision dans les IRIS-2000 contenant des adresses de grande taille. A budget global constant, la contrepartie est, évidemment, un taux d'échantillonnage un peu moindre dans le reste de la base.

Mais qu'appelle-t-on adresse de grande taille ?

Des travaux de simulation ont été effectués dans le but de déterminer le meilleur compromis pour fixer le seuil de définition des adresses de grande taille. Les simulations effectuées ont conduit à qualifier d'adresses de grande taille dans une commune donnée celles qui sont les plus grandes, qui regroupent au moins 60 logements chacune et qui, ensemble, représentent 10% au maximum du nombre de logements de la commune.

Pratiquement, l'opération consiste à classer les adresses de la commune par taille décroissante (en nombre de logements) et à retenir - pour être recensées exhaustivement - toutes celles dont le nombre cumulé de logements (en partant de la plus grande) atteint 10% au plus du nombre total de logements de la commune (nombre connu grâce au RIL), la dernière retenue devant compter au moins 60 logements.

Il s'agit donc d'un seuil relatif, propre à chaque commune. Si l'on avait retenu un seuil de taille absolu en nombre de logements, valable pour toutes les communes, ce seuil aurait été trop élevé dans certaines (la strate des adresses de grande taille aurait alors été vide pour ces communes) ou trop bas dans d'autres (la strate des adresses de grande taille aurait alors occupé tout l'échantillon annuel pour ces communes).

Prise en compte de l'apparition d'adresses nouvelles¹¹⁹

Les opérations de mise à jour du RIL, confortées par l'expertise des communes, prennent en compte les modifications du parc des logements. Les immeubles détruits sont enlevés de la base d'adresses (au vu des permis de démolir et des résultats des enquêtes sur le terrain), tandis que les constructions sont introduites avec leur nombre de logements supposé (figurant dans les permis de construire ou fourni par la commune dans le cas d'adresses issues des fichiers de La Poste). Or, les méthodes d'estimation utilisent le critère « nombre de logements » comme variable principale d'extrapolation. Il importe donc que l'exactitude de ce critère soit avérée, notamment par les vérifications opérées sur le terrain.

C'est pourquoi tous les logements des adresses nouvelles doivent être recensés¹²⁰. En flux annuel, les adresses nouvelles représentent en moyenne 1% des logements des communes. Les adresses nouvelles dont il est question ici comprennent également des adresses qui existaient déjà mais qui ont subi une modification importante de leur nombre de logements. Ces dernières adresses conservent leur groupe de rotation initial mais, comme les autres adresses nouvelles, elles sont vues exhaustivement au moment du passage de leur groupe.

Au démarrage du système, et notamment pendant les années 2004 à 2008, se pose le problème particulier du stock des adresses nouvelles apparues entre 1999 (date du

¹¹⁹ On préfère le terme d'adresses « nouvelles » à celui d'adresses « neuves » : cette dernière catégorie d'adresses ne reflète en effet pas exactement la construction neuve, dans la mesure où, chaque année, on y classe les adresses qui n'existaient pas dans le RIL précédent : on peut donc avoir une année donnée des adresses « nouvelles » au sens où elles étaient inconnues du RIL, mais qui existent sur le terrain depuis plusieurs années. Cela est surtout vrai les premières années où les progrès de gestion du RIL sont plus importants (participation active des communes, résorption du stock des adresses à vérifier, etc.).

¹²⁰ Certaines de ces adresses nouvelles seront peut-être des adresses de grande taille et seront, alors, traitées comme telles.

recensement de la population qui a permis l'initialisation du RIL) et 2003. Recenser de manière exhaustive dès 2004 la totalité de ce stock d'adresses nouvelles aurait, dans certaines communes, détérioré fortement le taux de sondage applicable aux autres adresses, et par conséquent la représentativité infracommunale. C'est pourquoi il a été décidé de répartir ce stock entre les cinq années 2004 à 2008 en visant, là aussi, à égaliser les groupes en nombre de logements et à optimiser la répartition spatiale. Ainsi toutes les adresses neuves construites entre 1999 et 2003 auront été visitées en 2008, année du calcul des premiers résultats détaillés du recensement.

A terme, les adresses nouvelles seront visitées exhaustivement dans les trois ans au maximum qui suivent l'entrée des logements neufs dans le RIL. L'objectif sera ainsi que, lors de l'élaboration des résultats détaillés du recensement, toutes les adresses nouvelles introduites dans le RIL de l'année médiane aient été visitées. *Exemple* : adresses nouvelles incorporées en 2009 dans le RIL, réparties sur les enquêtes de recensement de 2010, 2011 et 2012. Lors des résultats détaillés publiés fin 2012 à valeur 2010, toutes les adresses nouvelles présentes dans le RIL de référence auront été recensées.

Taitement de la strate des « autres » adresses

Les adresses de la strate "autres adresses" sont au départ réparties en cinq groupes de rotation équilibrés. Les critères d'équilibrage sont ceux qui ont guidé la constitution des groupes de rotation des communes de moins de 10 000 habitants, à savoir :

- le nombre de logements ;
- le nombre de logements en immeuble collectif ;
- la population des personnes de moins de 20 ans ;
- la population des personnes de 20 à 39 ans ;
- la population des personnes de 40 à 59 ans ;
- la population des personnes de 60 à 74 ans ;
- la population des personnes de 75 ans ou plus ;
- la population des femmes ;
- la population des hommes.

Les valeurs d'initialisation sont celles observées lors du recensement de 1999.

Chaque année, un échantillon aléatoire d'adresses est tiré dans le groupe de l'année. Pour ce tirage de deuxième phase, on introduit comme critères d'équilibrage le nombre de logements de la strate, le nombre de logements collectifs et le poids des IRIS en nombre de logements.

La construction pratique des groupes de rotation d'adresses

Au démarrage du système, les opérations sont les suivantes pour une commune :

A - On dresse la liste des adresses de grande taille (selon la définition donnée plus haut) ; elles sont, soit déjà connues (parce que recensées en 1999), soit nouvelles (et on connaît alors le nombre de logements déclaré dans le permis de construire ou fourni par la commune) ; on répartit ces adresses de grande taille en cinq groupes équilibrés (en nombre de logements) en utilisant la méthode du cube si elles sont nombreuses et en cherchant à optimiser la répartition spatiale dans la commune ; chacun de ces groupes sera traité à part et recensé exhaustivement, son année venue ;

B - Les adresses nouvelles qui ne sont pas de grande taille sont réparties de la même manière en cinq (en trois dans un stade plus avancé) groupes équilibrés (en nombre de logements d'après les permis de construire et les renseignements fournis par la commune) ;
C - On répartit toutes les autres adresses (toutes connues, par définition) en cinq groupes équilibrés par la méthode du cube sur l'ensemble des critères énumérés plus haut (voir Traitement de la strate des "autres" adresses).

L'échantillon d'adresses d'une année comportera ainsi :

- toutes les adresses de grande taille du groupe-année correspondant (cf. A) ;
- toutes les adresses nouvelles du groupe-année correspondant (qui exclut par définition les adresses de grande taille : cf. B) ;
- les adresses tirées au sort parmi les "autres" adresses du groupe-année correspondant (cf. C).

Le taux de sondage d'une année est ajusté de manière que la proportion de logements enquêtés dans la commune, toutes catégories d'adresses confondues, soit égale à 8%. On introduit toutefois une « clause de sauvegarde » pour la strate des autres adresses. En effet, dans quelques communes, le poids des adresses nouvelles alourdi par le stock 1999-2003 peut conduire à un taux de sondage trop faible dans la strate des autres adresses et, donc, détériorer la représentativité globale de certains IRIS. L'analyse menée à partir des tendances observées entre 1990 et 1999 a conduit à décider que le taux de sondage pratiqué dans le groupe des autres adresses d'une année donnée ne serait pas inférieur à 25%. Si cela devait être le cas, on ajusterait le taux en renonçant à l'exhaustivité des adresses nouvelles devant être enquêtées cette année-là. En 2004, une dizaine de communes ont été concernées. Il s'agit de communes dont la population est comprise entre 10 000 et 30 000 habitants et qui se situent dans la couronne de grandes agglomérations.

En régime courant, la disparition d'adresses tend à altérer l'équilibre des groupes. Celui-ci est rétabli par la ventilation des adresses nouvelles autres que celles de grande taille entre les cinq groupes des "autres" adresses. Cette ventilation a lieu à leur apparition, selon le nombre de logements renseigné, en rééquilibrant les groupes au mieux.

ANNEXE B 3

Le répertoire d'immeubles localisés

(communes de 10 000 habitants ou plus de métropole)

Le répertoire d'immeubles localisés (RIL) est l'un des éléments du système d'information géographique de l'Insee, qui compte au nombre des producteurs publics d'informations géographiques.

En effet, l'information géographique fait partie du système statistique. La localisation géographique est au cœur de ses travaux, que ce soit en amont dans les phases de collecte d'information, pendant les phases d'analyse intégrant la dimension spatiale de l'information ou dans la phase aval de diffusion. Elle ouvre de nouvelles possibilités de croisement d'informations et d'analyse et va donc prendre une importance croissante compte tenu de la souplesse d'utilisation et de la puissance des moyens informatiques de cartographie. C'est ainsi que les établissements économiques, depuis longtemps repérés par leurs coordonnées géographiques, peuvent être aisément localisés sur des cartes à grande échelle.

Le système d'information géographique de l'Insee va des pays aux adresses et gère de nombreux zonages de référence. L'Institut est responsable de l'attribution d'un numéro de code aux communes et aux différents échelons administratifs et gère à ce titre le « Code officiel géographique ». Il a mis en place, par ailleurs, une base de données numérisées appelée CICN2 (base de cartographie infracommunale numérisée, version 2, couvrant les communes de 10 000 habitants ou plus et répertoriant les voies, les noms des voies, les adresses et les zonages de référence pour la production et la diffusion de l'information statistique) et un système d'information géographique relatif aux noms de voies pour les besoins de Sirene (Système informatisé du répertoire des entreprises et des établissements).

Le nouveau recensement de la population s'appuie, dans les communes de 10 000 habitants ou plus de métropole, sur un répertoire d'immeubles localisés dans une base de données géographiques. Dans les départements d'outre-mer, où ce répertoire n'est pas encore en place, une [enquête cartographique](#) permet de constituer la base de sondage d'adresses nécessaire.

1 - Qu'est-ce que le RIL ?

Le RIL est un répertoire d'adresses localisées (en coordonnées géographiques) associé à une base communale de données géographiques.

Le champ du RIL couvre toutes les communes de France métropolitaine ayant atteint 10 000 habitants au recensement de 1999. Cela représente 892 communes (soit 2,5% du nombre total de communes) couvrant :

- 4,5% du territoire français ;
- 30 millions d'habitants (environ la moitié de la population) ;
- 5 millions d'adresses d'habitation correspondant à 14 millions de logements ;
- 2 millions d'adresses d'entreprise dont un bon nombre sont aussi des adresses d'habitation.

Le taux d'évolution 1999-2003 a été de + 4,5% d'adresses et + 4% de logements.

Le RIL a été constitué et il est mis à jour par les directions régionales de l'Insee. Chacune le gère sur son territoire. La répartition de la charge est très inégale puisqu'elles ont à gérer un nombre de communes qui va de 3 à 166 et que cinq d'entre elles gèrent la moitié des adresses du répertoire.

2 - Le contenu du RIL

2-1 Les **adresses** contenues dans le RIL sont celles des immeubles d'habitation, des communautés et des établissements du répertoire Sirene.

Par immeuble d'habitation, il faut entendre immeuble au sens du recensement, c'est-à-dire une construction qui comprend au moins un logement, quelle que soit la nature de cette construction.

Ultérieurement, le RIL intégrera les adresses des équipements au service du public .

Le répertoire gère les informations suivantes :

- pour chaque adresse : les coordonnées géographiques, le type et le nom de la voie, le numéro dans la voie, éventuellement un complément d'adresse ;
- pour chaque ensemble immobilier implanté à une adresse :
 - le type : immeuble d'habitation (en habitat individuel ou collectif ou mixte), communauté, immeuble d'activité, immeuble d'équipement urbain, ou mixte ;
 - la raison sociale de la communauté, de l'établissement ou de l'équipement public ;
 - l'année de construction de l'immeuble le plus récent de l'adresse ;
 - la date d'entrée dans le répertoire ou la date de dernière modification ou de destruction ;
 - le nombre de niveaux de l'immeuble le plus haut de l'adresse ;
 - le nombre de logements, le nombre d'établissements ou de communautés qui y sont installés et le type d'équipement urbain.

2-2 Le RIL est associé, sur chacune des communes de son champ, à une **base de données géographiques**, Base-Ilots, qui contient les voies, les dénominations de voies, les adresses aux carrefours, les limites d'îlots et d'IRIS-2000¹²¹, ainsi que des éléments de repérage (jardin, stade, bâtiment public remarquable, etc.).

Chaque adresse (habitation, communauté, établissement Sirene) est géocodée sur ce référentiel cartographique de la manière suivante : son positionnement le long du tronçon de voie est obtenu par interpolation linéaire et automatisée entre les numéros aux extrémités du tronçon qui, eux, sont associés au tronçon (chaque tronçon porte comme attribut, quatre adresses : début et fin de tronçon, à droite et à gauche). Un positionnement manuel peut également être opéré. Pour les adresses non normalisées¹²², à défaut d'informations plus précises, le positionnement se fait en cercle autour de la position du lieu-dit ou encore, par défaut, au centre de gravité de l'îlot auquel appartient l'adresse.

¹²¹ Pour les notions d'îlot et d'IRIS-2000 (îlots regroupés pour l'information statistique), voir l'[encadré](#) du chapitre B, § B.3.2

¹²² Pour la notion d'adresse, normalisée ou non normalisée, voir l'[encadré](#) du chapitre B, § B.4.1.1

3 - A quoi le RIL sert-il ?

Le RIL est utilisé par l'Insee comme base de sondage du recensement de la population pour tirer les échantillons d'adresses d'habitation enquêtées chaque année dans les communes de 10 000 habitants ou plus de métropole. Il a aussi d'autres utilisations, notamment :

- géocoder tout fichier disposant d'une variable de type adresse, c'est-à-dire donner aux adresses des coordonnées géographiques ou associer à ces adresses les codes de leurs zonages infracommunaux d'appartenance ;
- fournir des données finement localisées à des fins d'études, sous réserve des conditions d'utilisation définies après avis de la CNIL ;
- constituer un cadre cohérent pour la diffusion de résultats statistiques.

4 - Statut juridique du RIL

La constitution et la mise à jour du RIL par l'Insee ont été autorisées par un arrêté du 19 juillet 2000 du ministre de l'économie, des finances et de l'industrie pris après avis de la Commission nationale de l'informatique et des libertés (CNIL).

Pour les besoins du recensement de la population et suivant les dispositions de l'article 156, § IX de la loi du 27 février 2002 relative à la démocratie de proximité, l'arrêté du 19 juillet 2000 a été modifié par l'arrêté du 9 octobre 2002 afin d'autoriser la communication aux communes ou aux établissements publics de coopération intercommunale des informations du RIL qui les concernent.

Ces textes figurent à l'[annexe E 22](#).

5 - Constitution et mise à jour du RIL

5-1 Initialisation

L'Insee a constitué le RIL de 2000 à 2002 par géocodage des adresses des immeubles d'habitation et des communautés recensées en 1999 et des adresses des établissements immatriculés au répertoire Sirene. Puis l'Insee l'a mis à jour en 2002 et 2003 pour constituer un RIL millésimé juin 2003, dans lequel ont été tirés les échantillons d'adresses de la première enquête de recensement, réalisée en janvier et février 2004. Cette mise à jour s'est faite à partir des sources suivantes qui ont contribué à l'apport de nouvelles adresses :

- fichiers de permis de construire : 1,2 % d'apports ;
- fichiers de La Poste : 3,3 %.

Elle a aussi bénéficié de l'indispensable contribution des communes, sous la forme de confirmation de l'existence d'adresses et de travaux d'expertise. Plus précisément, voici les travaux qu'elles ont réalisés :

- envoi en continu à l'Insee des copies des arrêtés municipaux qui intéressent la mise à jour du RIL ;
- réponse aux sollicitations ponctuelles de l'Insee sur l'avancement des permis de construire ;
- examen des listes d'adresses à confirmer soumises par l'Insee lorsque ce dernier ne pouvait pas conclure avec ses sources : adresses susceptibles d'être nouvelles ou détruites ou modifiées mais sans que cela soit confirmé par plusieurs sources. Lors des travaux d'initialisation, ces listes ont porté sur 20% des adresses, ce qui est considérable. Les trois quarts des communes ont répondu (avec un investissement inégal) et leur travail a porté, de mars à mai 2003, sur 68% du total d'adresses à confirmer ;
- enfin, l'Insee a sollicité les communes sur l'expertise du RIL complet, opération encadrée par arrêté (cf. ci-dessous). 47 % des communes ont répondu à la demande d'expertise. Ce chiffre est à interpréter comme une fourchette basse car des confusions

ont pu avoir lieu entre l'examen des listes d'adresses et l'expertise du RIL. En effet, les calendriers étaient très proches, voire se recouvraient.

5-2 Mise à jour

Le répertoire est, depuis lors, mis à jour en continu pour identifier et localiser les adresses nouvelles, les adresses détruites et les adresses modifiées. Ce régime permanent s'organise depuis 2003 en un cycle annuel allant de juillet à juin de l'année suivante, avec édition en juillet d'un état du RIL. La mise à jour se fait à partir :

- des informations communiquées par les communes concernant les modifications de la voirie et des adresses ;
- de fichiers administratifs : permis de construire, adresses de La Poste, taxe d'habitation ;
- des informations communiquées par les communes dans le cas de différences constatées entre les diverses sources d'actualisation ; les communes sont consultées de manière échelonnée de novembre-décembre jusqu'à avril sur des listes d'adresses à confirmer ;
- des retours d'expertise par les communes ;
- des constatations faites sur le terrain pendant la réalisation des enquêtes de recensement ;
- de vérifications sur le terrain, le cas échéant.

5-3 Expertise du RIL par la commune

L'expertise du répertoire est une opération préliminaire aux enquêtes de recensement ; elle est destinée à donner à celles-ci le meilleur niveau de qualité possible.

Le décret d'application de la loi du 27 février 2002 (décret n° 2003-485 du 5 juin 2003 relatif au recensement de la population) et l'arrêté interministériel du 5 août 2003 fixent les conditions dans lesquelles l'expertise se déroule :

- envoi par l'Insee à la commune avant fin mai de la liste de toutes les adresses du RIL ;
- transmission par la commune de ses remarques dans le mois qui suit la réception.

L'expertise intervient après intégration des informations fournies en continu à l'Insee par la commune (arrêtés municipaux, plans, extraits du système d'information géographique communal), des informations venant de sources externes et des réponses de la commune aux listes d'adresses à confirmer établies par l'Insee après utilisation des diverses sources de mise à jour.

L'expertise commence par l'envoi à la commune par l'Insee, avant le 25 mai, du résultat de ces travaux (fichier extrait du RIL accompagné de plans-images). La commune dispose d'un mois pour faire connaître ses remarques, qui seront alors prises en compte dans l'édition du RIL destinée à la préparation de l'enquête de recensement suivante.

Une date butoir (30 juin de l'année n), est fixée avant laquelle les remarques de la commune peuvent être traitées dans le cadre de l'édition "n" du RIL qui sera utilisée pour le tirage de l'échantillon de l'enquête de recensement de l'année n+1. Au-delà de cette date, les remarques seront traitées dans l'édition n+1 du RIL. En régime de croisière, les vérifications de la commune peuvent être centrées sur les mouvements de création, de disparition et de modification d'adresses, qui ne représentent qu'un faible pourcentage du "stock" d'adresses (environ 3 % au total des adresses du RIL).

Grâce à l'expertise de la commune, l'Insee peut utiliser une base de sondage aussi à jour que possible pour tirer l'échantillon des adresses qui vont être enquêtées dans la commune en janvier et février suivants.

La mise à jour du RIL en continu et l'opération d'expertise créent ainsi un cadre de concertation entre la commune et l'Insee sur la qualité du répertoire.

6 - La qualité du RIL

6-1 Nécessité d'une mesure de la qualité

La qualité du RIL est, évidemment, un facteur déterminant de la qualité du recensement puisque le répertoire est la base de sondage dans laquelle sont tirés chaque année les échantillons d'adresses à recenser dans les communes de 10 000 habitants ou plus. La qualité du RIL peut être déclarée parfaite si le répertoire est l'image rigoureusement fidèle du terrain, c'est-à-dire si toute adresse existant sur le terrain se trouve dans le RIL et si toute adresse présente dans le RIL se retrouve sur le terrain. C'est un idéal auquel il faut prétendre pour des raisons juridiques et techniques. Des raisons juridiques puisque la loi exige que la totalité du territoire des communes de 10 000 habitants ou plus soit prise en compte en cinq ans lors des enquêtes de recensement¹²³, ce qui ne serait pas le cas si des adresses existant sur le terrain étaient écartées du tirage au sort. Des raisons techniques puisque l'absence d'adresses dans le RIL entraînerait une sous-estimation de la population de la commune et la présence dans le RIL d'adresses inexistantes sur le terrain une surestimation de sa population. En outre, un excédent d'adresses pose des problèmes d'organisation de la collecte puisqu'il augmente le risque d'envoyer un agent recenseur à une adresse inexistante.

Bien sûr, dans la réalité, cet idéal de perfection ne peut être atteint pour de multiples raisons bien connues de tous les gestionnaires de grands répertoires, la plus importante étant que les informations de mise à jour ne sont jamais ni instantanées ni elles-mêmes parfaites. Il convient donc de mesurer de façon aussi précise que possible les écarts du RIL par rapport au terrain.

C'est pourquoi un dispositif de **mesure de la qualité** du RIL est mis en œuvre dans les directions régionales de l'Insee. L'objectif est de mesurer :

- le déficit d'adresses, c'est-à-dire la proportion des adresses présentes sur le terrain mais absentes du RIL ;
- l'excédent d'adresses, c'est-à-dire la proportion des adresses présentes dans le RIL mais n'existant pas sur le terrain ;
- la qualité de la variable nombre de logements dans le RIL.

Un dispositif d'évaluation par **comptage** est également mis en œuvre, qui consiste à rapprocher l'évolution du nombre d'adresses dans le RIL et dans les fichiers de la taxe d'habitation et à rechercher les causes des écarts.

L'objectif de démarrage est d'avoir un déficit maximum par commune de 2,5% d'adresses. Cet objectif est ambitieux puisqu'il consiste à ne pas accepter un risque de sous-estimation de la population supérieur aux sous-dénombrements observés en milieu urbain concentré lors des recensements généraux de la population. Globalement, il était atteint dès le premier RIL mais il faut l'atteindre pour chaque commune. L'idée est de l'atteindre pour le RIL de janvier 2006 sur lequel seront calés les premiers calculs de population légale dans les communes de 10 000 habitants ou plus (voir chapitre D, [§ D.3.1](#)), puis d'élever progressivement le niveau d'exigence. De cette manière, les communes dont le RIL qui a

¹²³ Article 156, § VI, 2e alinéa, de la loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité?

servi à tirer les échantillons des enquêtes de recensement de 2004 et 2005 n'était pas encore au niveau de qualité souhaité ne seront pas pénalisées en termes de population légale.

6-2 L'enquête de mesure de la qualité

La première enquête de mesure de la qualité a eu lieu début 2004. Elle concernait un échantillon de 225 communes sur les 892 du champ des communes de 10 000 habitants ou plus. Les 34 communes de plus de 100 000 habitants et les 45 arrondissements de Paris, Lyon et Marseille faisaient partie de l'échantillon.

Au sein de chaque commune sélectionnée, 8% des îlots ont été tirés de façon équilibrée sur les adresses d'habitation et les logements. L'équilibrage s'est fait également sur les adresses sans numéro et les adresses nouvelles, mais avec une priorité moindre¹²⁴.

L'échantillon tiré sur les 225 communes totalisait 138 000 adresses, soit environ 2,8% des adresses du RIL.

Tableau 1	Communes	Ilots	Adresses	Logements
Echantillon total de l'enquête	225	5 875	138 434	539 465
RIL total	892	193 887	5 069 744	14 636 925

Lors de l'enquête, les enquêteurs disposaient de la liste de l'ensemble des adresses de l'îlot figurant dans le RIL arrêté au 30 juin 2003 : adresses d'habitation et autres catégories d'adresses (établissements, équipements, autres). Ils vérifiaient l'existence de toutes ces adresses avec des consignes de priorité sur les habitations. Ils contrôlaient également l'adressage, le nombre de logements, les adresses avec plusieurs accès et relevaient les éventuelles confusions entre catégories d'adresses.

L'inévitable décalage entre la date de la collecte de l'enquête-qualité et celle de l'édition du RIL a quelque peu détérioré les chiffres d'exhaustivité, mais le phénomène devrait être limité : il portait sur les adresses en cours de construction avant l'édition du RIL (non prises en compte, de ce fait, dans l'enquête-qualité) et qui se sont achevées avant la collecte (et étaient donc comptées comme absentes du RIL), ainsi que des adresses détruites entre l'édition du RIL et la collecte (et donc comptées comme abusivement présentes dans le RIL).

Le traitement des cas ambigus rencontrés sur le terrain était centralisé afin d'en assurer l'homogénéité.

6-3 Les résultats de l'enquête sur le RIL 2003

Rappelons que l'objectif principal de l'enquête était le **contrôle de l'exhaustivité du RIL-habitation** en termes d'adresses et de logements et en particulier du **déficit** d'adresses (puisque'un éventuel déficit peut déboucher sur une sous-estimation de la population).

6-3-1 Le déficit du RIL 2003

¹²⁴ Sur la notion d'échantillon équilibré, voir l'[annexe B.1](#)

Le déficit d'adresses s'élevait mi-2003 à 2,3% au total, mais ce nombre recouvre de fortes disparités entre communes. Au total cela représente près de 120 000 adresses comprenant 325 000 logements

Le traitement par les communes des listes d'adresses à confirmer¹²⁵ est un facteur décisif pour la qualité du RIL. Les communes où ces listes ont pu être traitées ont un déficit près de trois fois inférieur (1,6% contre 4,7%). Pour les communes n'ayant pu traiter les listes, les adresses manquantes sont souvent dans la catégorie « autres » : cela veut dire qu'elles figurent dans le RIL mais dans une autre couche que la couche des adresses d'habitation. Une remise à niveau de ces adresses abaisserait le déficit d'environ 2,8 points dans les communes concernées et de près de 0,8 point sur l'ensemble du RIL.

Tableau 2	Listes d'adresses à confirmer non traitées		Listes d'adresses à confirmer traitées		Ensemble	
Nombre de communes	75		150		225	
Adresses totales	1 218 133	100%	3 851 611	100%	5 069 744	100%
Déficit	57 483	4,7%	60 248	1,6%	117 731	2,3%
<i>dont adresses connues du RIL mais non qualifiées d'habitation</i>	33 640	2,8%	7 951	0,2%	41 591	0,8%
Logements totaux	4 695 657	100%	9 705 400	100%	14 401 057	100%
Déficit	164 348	3,5%	160 175	1,7%	324 524	2,3%
<i>Dont à des adresses connues du RIL mais non qualifiées d'habitation</i>	86 885	1,9%	17 841	0,2%	104 726	0,7%

Pour les logements, les chiffres concernent uniquement les logements des adresses en déficit ou en excédent. Les erreurs du RIL sur le nombre de logements des adresses correctes étaient comptées séparément.

Plus du quart des communes dépassaient mi-2003 le plafond de 2,5% de déficit, mais cette proportion atteignait 50% pour les communes n'ayant pu traiter les listes d'adresses à confirmer.

Tableau 3	Listes d'adresses à confirmer non traitées	Listes d'adresses à confirmer traitées	Ensemble
Communes en dessous de 2,5% de déficit	113	529	641
Communes dépassant 2,5% de déficit	112	138	251
Total	225	667	892
% de communes dépassant 2,5% de déficit	50%	21%	28%

6-3-2 L'excédent du RIL 2003

Les excédents étaient un peu supérieurs dans les communes n'ayant pu traiter les listes d'adresses à confirmer mais avec une moindre ampleur que pour les déficits. Il semble que

¹²⁵ Listes d'adresses issues du processus de mise à jour pour lesquelles une confirmation et des informations des communes étaient sollicitées. Ces listes concernaient en particulier les adresses nouvelles apportées par le fichier de La Poste. En l'absence de réponse de la commune, ces adresses ont été intégrées au RIL mais, par mesure de prudence, dans une catégorie « autre » afin de ne pas ajouter à tort d'adresses d'habitation.

plusieurs des communes de ce type aient cumulé les handicaps : terrain difficile (beaucoup d'adresses sans numéro ou d'adresses nouvelles) et faible participation de la commune.

Tableau 4	Listes d'adresses à confirmer non traitées	Listes d'adresses à confirmer traitées	Ensemble
Nombre de communes	75	150	225
Adresses totales	1 218 133 100%	3 851 611 100%	5 069 744 100%
Excédent	30 627 2,5%	72 900 1,9%	103 527 2,0%
Logements totaux	4 695 657 100%	9 705 400 100%	14 401 057 100%
Excédent	109 840 2,3%	195 819 2,0%	305 658 2,1%

La décomposition de l'excédent montre que les doublons, les défauts d'actualité et les confusions de catégorie (par exemple : adresse d'habitation dans le RIL mais établissement sur le terrain) sont assez rares. Ces trois éléments n'expliquent qu'un quart de l'excédent, soit 0,5% des 2,0% d'excédent total.

L'analyse des excédents selon la source des adresses dans le RIL est retracée par le tableau 5.

Tableau 5	Listes d'adresses à confirmer non traitées	Listes d'adresses à confirmer traitées	Ensemble
Nombre de communes	75	150	225
Adresses totales	1 218 133 100%	3 851 611 100%	5 069 744 100%
Excédent	30 627 2,5%	72 900 1,9%	103 527 2,0%
<i>dont adresses du RP 1999</i>	23 680 1,9%	47 868 1,2%	71 548 1,4%
<i>dont adresses nouvelles des permis de construire (PC)</i>	2 910 0,2%	1 816 0,0%	4 725 0,1%
<i>dont adresses nouvelles issues des fichiers de La Poste</i>	4 037 0,3%	23 216 0,6%	27 253 0,5%

L'excédent sur les adresses issues du RP1999 est nettement plus élevé dans les communes où les listes d'adresses à confirmer n'ont pu être traitées. Il est probable que des destructions d'adresses n'ont pas pu être prises en compte mais un taux de 1,9% d'adresses disparues de 1999 à 2003 dans ces communes semble trop important ; il est vraisemblable que le RP1999 lui-même n'y a pas été de très bonne qualité.

En revanche, dans les communes où les listes d'adresses à confirmer ont pu être traitées, un certain nombre d'adresses issues des fichiers de La Poste ont été ajoutées à tort. Cela représente près d'un tiers de l'excédent de ces communes soit 0,6% de leur RIL.

Ceci permet de faire un bilan de la mise à jour pour les communes ayant pu prendre en compte les listes d'adresses à confirmer :

Tableau 6	Listes d'adresses à confirmer traitées	
Nombre de communes	150	
Adresses totales	3 851 611	100,0%
Adresses nouvelles PC	33 602	0,9%
<i>dont adresses PC en excédent</i>	1 385	0,0%
Adresses nouvelles La Poste	150 682	3,9%

dont adresses La Poste en excédent	21 270	0,6%
------------------------------------	--------	------

Il apparaît que l'apport du fichier de La Poste, qui a nécessité un gros travail de la part des communes, a permis de réduire sensiblement le déficit sans trop détériorer la qualité du RIL par l'ajout induit d'adresses. Le gain est d'environ 3% : 3,9% d'adresses ajoutées dont 0,6% à tort.

Le nombre d'adresses PC ajoutées à tort est quant à lui négligeable.

6-3-3 Les adresses « autre » (catégorie indéterminée)

Les éléments qui précèdent montrent l'intérêt des adresses apportées par les fichiers de La Poste. Même si ce n'était pas son objectif prioritaire, l'enquête menée sur le terrain a permis de tirer certains enseignements de ces fichiers. On se limite ici aux communes où les listes d'adresses à confirmer n'ont pu être traitées car ce sont elles qui concentrent la plupart des adresses « autre ».

Tableau 7	Listes d'adresses à confirmer non traitées	
	Nombre de communes	75
Adresses "autres" (adresses connues du RIL mais non qualifiées d'habitation)	156 996	100%
dont absentes sur le terrain	62 408	40%
dont habitation sur le terrain	33 640	21%
dont autre accès d'une habitation du RIL	5 926	4%
indéterminées	55 022	35%

Près d'un quart des adresses « autre » se sont révélées être des habitations sur le terrain (elles apparaissaient donc en déficit du RIL). Les accès secondaires (qui font courir le risque de créer des doublons si on les bascule en adresses d'habitation) étaient en nombre limité (4% des adresses « autre »). Les adresses fictives étaient en revanche nombreuses (40% environ), ce qui démontre l'importance du travail à réaliser pour les distinguer des « bonnes » adresses à ajouter au RIL. Enfin, pour 35% des adresses « autre », la collecte n'a pas apporté de réponse directe mais les commentaires des enquêteurs permettaient de penser qu'il s'agissait souvent d'adresses de garages, d'entreprises ou d'établissements scolaires ayant plusieurs adresses ou d'adresses d'équipements (piscine, stade, etc.). Un certain nombre de ces adresses indéterminées est aussi constitué d'adresses inexistantes sur le terrain (mais l'enquêteur ne l'a pas précisé) qu'il faudrait ajouter aux 40% d'adresses fictives.

Ces résultats montrent que le RIL des communes où les listes d'adresses à confirmer n'ont pu être traitées pourrait être significativement amélioré si celles des adresses de la couche « autre » qui sont des habitations étaient identifiées.

6-3-4 Les erreurs d'adressage

L'enquête-qualité avait pour objectif premier de mesurer l'exhaustivité du RIL-habitation mais les consignes données aux enquêteurs ont permis de repérer les erreurs d'adressage. Les comptages ont été effectués toutes communes confondues, le fait d'avoir traité ou non les listes d'adresses à confirmer n'ayant pas eu d'impact significatif sur la qualité de l'adressage. Le tableau suivant éclaire la situation :

Tableau 8

Nombre de communes	225	
Adresses totales	5 069 744	100%
Total des erreurs d'adressage	59 437	1,2%
Erreurs sur le libellé de voie	23 544	0,5%
Erreurs sur le type de voie	11 755	0,2%
Erreurs sur le numéro (adr. normalisées)	16 587	0,3%
Erreurs sur le numéro (adresses nonnormalisées)	2 459	0,0%
Erreurs sur le suffixe	5 092	0,1%

Les erreurs d'adressage restent relativement limitées : 1,2% des adresses du RIL au 30 juin 2003 étaient concernées. De plus, de nombreuses erreurs sont mineures. Un type de voie erroné (avenue à la place de boulevard, par exemple) ne pose, en général, pas de problème de repérage. Les changements de dénomination de voie sont souvent connus, ou bien les plans de collecte permettent de situer la voie sans ambiguïté et amènent l'enquêteur à faire une recherche complémentaire pour trouver le nouveau nom. Les erreurs de suffixe sont souvent des B transformés en Bis ou l'inverse et n'ont pas d'incidence majeure. Les erreurs qui empêchent véritablement le repérage de l'adresse ou entraînent sa confusion avec une autre adresse sont surtout les erreurs de numéro (0,3% des cas).

6-4 Conclusions

Dès mi-2003, la qualité du RIL apparaissait globalement satisfaisante mais recouvrait des disparités entre communes. Le travail effectué par celles-ci sur les listes d'adresses à confirmer apparaît comme un élément déterminant de la bonne qualité finale du RIL.

L'apport des deux principaux fichiers de mise à jour, permis de construire et fichier de La Poste, est très positif quand il se greffe sur le travail de vérification des communes. Les fichiers de la taxe d'habitation (TH) peuvent être utilisés comme source de mise à niveau mais leurs apports doivent être confortés par d'autres éléments pour éviter que les différences d'adressage entre TH et RIL n'aboutissent à la création de doublons. Dans ce domaine aussi, le partenariat entre les communes et l'Insee permet de réaliser un travail de qualité.

Des investigations sur la qualité du RIL doivent être menées régulièrement. L'enquête sur la qualité du RIL au 30 juin 2004 a été conduite d'août à octobre 2004 et ses résultats permettront de suivre les progrès découlant du travail accompli en partenariat entre les communes concernées et l'Insee.

Au total, l'objectif de qualité retenu pour le RIL de janvier 2006 apparaît réaliste.