



N° 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

**Symposium 2004 : Méthodes
innovatrices pour enquêter
auprès des populations
difficiles à joindre**

2004



LES DÉFIS MÉTHODOLOGIQUES D'UNE ENQUÊTE PORTANT SUR LA DIVERSITÉ ETHNIQUE ET CULTURELLE DE LA POPULATION CANADIENNE

Valérie Bizier, Jennifer Kaddatz et Danielle Laroche¹

RÉSUMÉ

Mener une enquête dans une population culturellement diversifiée représente de nombreux défis surtout lorsque la population en question est aussi multiculturelle que celle du Canada. Cette communication présente les problèmes méthodologiques auxquels ont dû faire face l'équipe de l'Enquête sur la diversité ethnique lors du choix du plan de sondage, de l'élaboration du questionnaire, de la collecte des données, de la pondération et de l'estimation de la variance ainsi que les différentes méthodes mises en place afin de relever les défis d'une telle enquête.

MOTS CLÉS : Conception de questionnaire, enquête postcensitaire, origines ethniques, populations rares.

1. GÉNÉRALITÉS

1.1 Pourquoi avoir mis sur pied l'Enquête sur la diversité ethnique ?

La composition ethnique et culturelle de la population du Canada a beaucoup évolué au cours du dernier siècle car l'immigration joue un rôle très important dans l'accroissement de la population canadienne. Jusqu'en 1971, ceux qui choisissaient le Canada comme terre d'accueil provenaient majoritairement des pays d'Europe. Mais, depuis, le nombre d'immigrants en provenance de l'Asie, de l'Afrique, des Caraïbes, de l'Amérique centrale et de l'Amérique du Sud a augmenté considérablement accélérant ainsi la diversité culturelle et raciale de la société canadienne.

Face à ce phénomène, les dirigeants du pays ont besoin d'une information objective sur l'appartenance ethnique afin d'appuyer des décisions stratégiques pour assurer la cohésion sociale et l'inclusion de tous les Canadiens dans l'ensemble de la société. Jusqu'à maintenant, la question portant sur l'origine ethnique ou culturelle des ancêtres d'une personne au recensement était la principale source de données concernant la diversité ethnique et culturelle au Canada. L'ennui avec cette question est qu'elle ne mesure qu'un seul aspect de l'appartenance ethnique. L'ethnicité d'une personne est influencée par plusieurs autres facteurs tels que l'identité ethnique, la race, la nationalité, le profil linguistique et la religion.

De plus, une variation dans la façon dont les personnes déclarent leurs origines a été remarquée au cours des dernières années. Le nombre de personnes qui déclarent avoir des ancêtres d'origine canadienne ne cesse d'augmenter de recensement en recensement ce qui laisse croire que la question est souvent mal comprise.

C'est donc dans le but de mesurer plus adéquatement l'appartenance ethnique d'une personne et combler certaines lacunes statistiques reliées à l'ethnicité que Statistique Canada et Patrimoine Canadien se sont réunis pour mettre sur pied l'Enquête sur la Diversité Ethnique (EDE).

¹Valérie Bizier, Statistique Canada, Immeuble R.H.Coats 15 Q, 120 av. Parkdale, Ottawa ON, Canada, K1A 0T6;
Jennifer Kaddatz, Statistique Canada, 600-300 W. Georgia Street, Vancouver BC, Canada, V6B 6C7;
Danielle Laroche, Statistique Canada, Immeuble R.H.Coats 15 Q, 120 av. Parkdale, Ottawa ON, Canada, K1A 0T6.

L'enquête sur la diversité ethnique poursuivait 4 objectifs:

- explorer diverses façons de mesurer l'ethnicité afin d'éclairer les futures collectes de données;
- mieux comprendre comment les Canadiens de différents antécédents ethniques interprètent et déclarent leur ethnicité;
- fournir de l'information sur la diversité ethnique au Canada;
- examiner comment les antécédents d'une personne influencent sa participation à la vie sociale, économique et culturelle du Canada.

1.2 Population cible

La population ciblée par l'EDE était constituée des personnes de 15 ans et plus vivant dans les logements privés des dix provinces canadiennes, ce qui incluait les citoyens canadiens, les immigrants reçus et les résidents non permanents.

Les territoires et les régions éloignés ainsi que les logements collectifs ont dû être exclus de la population cible pour des contraintes de collecte et de coûts. Les autochtones et les réserves indiennes ont également été exclus de la population cible de l'EDE car une enquête portant sur les peuples autochtones était effectuée presque au même moment et Statistique Canada voulait minimiser le fardeau de réponse pour les autochtones qui auraient pu se retrouver dans les deux enquêtes. En 2001, on estimait à 23 092 645 le nombre de personnes incluses dans la population cible de l'enquête.

2. PLAN DE SONDAGE

2.1 Choix de la base de sondage

Afin d'atteindre les objectifs d'enquête, il était essentiel d'étudier à la fois les groupes ethniques majoritaires et les minorités ethniques présents au Canada. Les bases de sondage utilisées habituellement tel que les bases de sondage aréolaires ou de logements étaient donc inadéquates pour l'enquête parce que les minorités ethniques sont pour la plupart des populations rares et dispersées.

L'approche postcensitaire semblait l'option la plus adéquate pour créer la base de sondage de l'enquête. Cette approche consiste à utiliser la base de données du recensement pour repérer la population cible. Elle comporte plusieurs avantages. C'est un moyen économique de recueillir des données auprès de sous-groupes restreints et dispersés de la population dans le sens où le recensement permet de repérer plus facilement les répondants en vue de la collecte. Et puisqu'il est plus facile de repérer les petites sous populations, l'approche postcensitaire permet d'obtenir des estimations pour des domaines de petite taille. Enfin, le recensement comprend une foule d'informations connexes qui permet de mieux préparer l'échantillon pour la collecte et permet également d'améliorer les méthodes d'estimation.

Au Canada, le recensement de la population a lieu tous les 5 ans et le dernier a eu lieu le 15 mai 2001. Ce recensement est dit «de jure» modifié car, à part quelques exceptions, les personnes sont recensées dans le logement où elles habitent habituellement. En général, deux questionnaires sont distribués à la population. Un questionnaire court qui comporte 6 questions, principalement des questions démographiques, est distribué à 4 logements sur 5 et, un questionnaire long qui comporte une soixantaine de questions est distribué à tous les 5 logements. Les questions incluses dans ce questionnaire long portent, entre autre, sur la démographie, l'origine ethnique, l'utilisation des langues, la race, la religion, la scolarité, les activités, le revenu et le logement.

Puisque les variables d'intérêts pour l'identification de la population cible et la sélection de l'échantillon pour l'EDE sont recueillies par le questionnaire long, la base de données de ce questionnaire a donc été utilisée pour tirer l'échantillon de l'EDE.

2.2 Plan d'échantillonnage

Le plan d'échantillonnage utilisé pour l'EDE est un plan stratifié à deux phases. La phase I correspondait à la distribution du questionnaire long du recensement à 1 ménage sur 5 au Canada de façon systématique à l'intérieur de chaque secteur de dénombrement (SD). Et, pour la phase II, les répondants de la phase I étaient répartis selon des strates et un échantillon systématique de personnes était choisi dans chacune des strates.

Comme il a été mentionné plus tôt, pour atteindre les objectifs de l'enquête, il était fondamental de cibler les groupes ethniques majoritaires et minoritaires dans la population cible, c'est-à-dire de s'assurer que ceux-ci soient bien représentés dans l'échantillon. Une autre condition pour atteindre ces objectifs était d'obtenir une bonne représentativité des individus qui se déclarent d'origine canadienne ainsi que des minorités ethniques d'origine européenne et non européenne puisque ces minorités ont des caractéristiques différentes. Ainsi, à partir des réponses données au recensement à la question sur les origines ethniques et culturelles des ancêtres d'une personne, les 7 strates suivantes ont été créées. (Il faut noter ici que les réponses multiples contenant à la fois des origines européennes et non européennes ont été classées dans la strate des individus d'origine non européenne en raison de la dominance de la présence des minorités visibles)

Groupes ethniques majoritaires (3 strates)

1. Individus d'origine canadienne seulement
2. Individus d'origine canadienne et d'origine britannique et/ou française
3. Individus d'origine britannique et/ou française seulement

Minorités ethniques (4 strates)

4. Individus d'origine canadienne et d'origine européenne
5. Individus d'origine canadienne et d'origine non européenne
6. Individus d'origine européenne
7. Individus d'origine non européenne

Un autre aspect important à considérer lorsque l'on étudie l'appartenance ethnique est le moment où le répondant ou ses ancêtres sont arrivés au Canada pour la première fois. Ainsi, les réponses aux questions sur les lieux de naissance du répondant et de ses parents ont également servi lors de la stratification. En effet, les répondants ont été répartis selon qu'ils étaient des immigrants de première génération, des personnes de deuxième génération ou encore qu'ils étaient au Canada depuis 3 générations ou plus. Le terme «immigrants de première génération» désigne les répondants nés à l'extérieur du Canada, que le terme «personnes de deuxième génération» désigne les répondants nés au Canada ayant au moins un parent né à l'extérieur du Canada, et que le terme «Canadien depuis 3 générations ou plus» désigne les répondants nés au Canada de parents également nés au Canada.

En croisant les groupes formés par l'origine ethnique et les lieux de naissance, 21 strates ont été formées. Cependant, afin d'obtenir un nombre suffisamment élevé de personnes dans chaque strate, certaines strates formées par les générations ont été regroupées. Par exemple, les strates formées par les personnes de première et deuxième génération ayant déclaré être d'origine canadienne seulement ont été regroupées. Puis, pour des raisons de couverture, une strate de non classés a été ajoutée. Cette strate contenait les individus qu'il était impossible de répartir dans une autre strate parce qu'ils n'avaient pas répondu à l'une des questions nécessaires à la stratification. Finalement, ce sont 15 strates qui ont été formées pour la sélection de l'échantillon de deuxième phase. Ces strates sont décrites dans le tableau 1 de la section 2.3.

Une fois les strates établies, la base de sondage a été triée selon la province, la circonscription électorale, le secteur de dénombrement et le numéro du ménage. Ainsi en tirant l'échantillon de façon systématique dans chaque strate, on s'assurait d'avoir une assez bonne répartition géographique de l'échantillon tout en réduisant le risque d'avoir plus d'une personne sélectionnée à l'intérieur d'un même ménage.

2.3 Allocation de l'échantillon

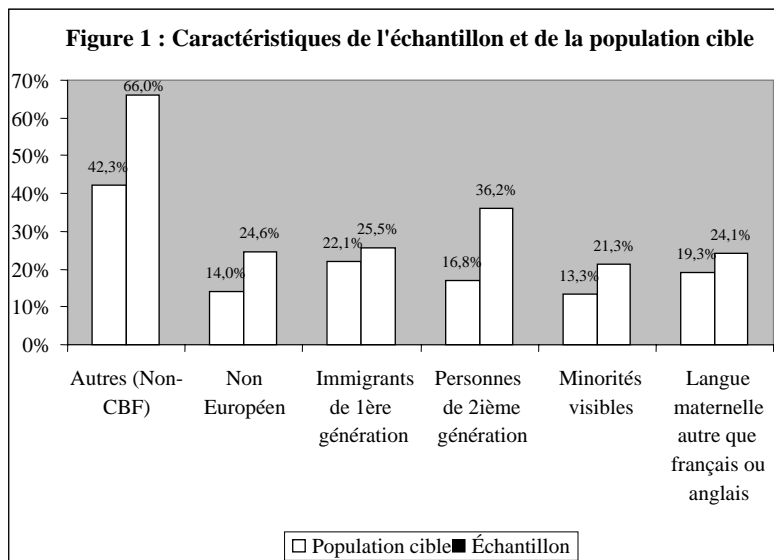
L'enquête visait l'obtention d'au moins 40 000 répondants. Puisqu'un taux de réponse de 70 % était prévu au départ, un échantillon d'environ 57 200 personnes était nécessaire pour atteindre cet objectif. Le coefficient de variation (c.v.) a été utilisé comme mesure de fiabilité. Dans chaque strate, la taille initiale de l'échantillon a été déterminée en fonction d'une proportion minimale à estimer de 4 %, d'un c.v. maximum de 12,5 %, d'un effet de plan de 1,2 et d'un taux de réponse estimé à partir des résultats du test pilote de septembre 2001.

Le tableau 1 montre le nombre de personnes tirées dans chaque strate lors de la deuxième phase ainsi que les fractions de sondage dans chaque strate.

Tableau 1 : Répartition de l'échantillon de deuxième phase par strate et fractions de sondage. Enquête sur la diversité ethnique (2002).

Strates	Taille de l'échantillon (n _i)	Taille de la population (N _i)	Fraction de sondage (n _i /N _i)
Non classés (non-réponse au recensement)	2 630	1 835 218	1/697
Origines canadiennes seulement – Générations 1 et 2	3 196	268 199	1/83
Origines canadiennes seulement – Générations 3 et +	3 218	4 406 439	1/1369
Origines canadiennes et britanniques et/ou françaises – Générations 1 et 2	2 970	311 703	1/104
Origines canadiennes et britanniques et/ou françaises – Générations 3 et +	2 986	1 727 043	1/578
Origines britanniques et/ou françaises – Générations 1 et 2	2 986	1 587 629	1/531
Origines britanniques et/ou françaises – Générations 3 et +	2 988	3 713 414	1/1242
Origines canadiennes et européennes – Générations 1 et 2	4 365	255 984	1/58
Origines canadiennes et européennes – Générations 3 et +	4 248	583 874	1/137
Origines canadiennes et non européennes – Toutes les générations	4 314	103 539	1/24
Origines européennes – Génération 1	5 032	1 750 284	1/347
Origines européennes – Génération 2	4 258	1 636 840	1/384
Origines européennes – Générations 3 et +	4 258	1 971 425	1/462
Origines non européennes – Génération 1	5 324	2 493 333	1/468
Origines non européennes – Générations 2 et +	4 379	447 721	1/102
Total	57 152	23 092 645	-

Étant donné que les fractions de sondage varient considérablement entre les différentes strates, l'échantillon de l'EDE présente plusieurs caractéristiques spécifiques. Afin d'illustrer ce que la stratification et l'allocation de l'échantillon a eu comme impact sur la composition de l'échantillon, voici un graphique (figure 1) qui montre les proportions de caractéristiques présentes dans l'échantillon par rapport à celles de la population cible.



On constate, par ce graphique, que la proportion de personnes avec au moins une origine autre que canadienne, britannique ou français est plus élevée dans l'échantillon de l'EDE que dans la population cible. (66,0% dans l'échantillon contre 42,3% dans la population).

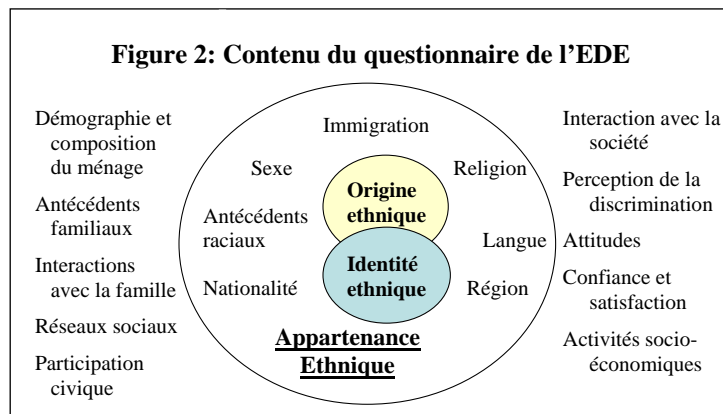
On remarque également que l'échantillon contient, en proportion, plus d'immigrants de première génération, plus d'individus de deuxième génération, plus de personnes faisant partie d'une minorité visible et plus de personnes dont la langue

maternelle n'était ni l'anglais ni le français.

3. CONCEPTION DU QUESTIONNAIRE

Lorsque l'on conçoit un questionnaire, on doit tenir compte de plusieurs aspects importants liés à l'enquête. Il faut tenir compte des caractéristiques particulières de la population cible afin de s'assurer que les questions soient pertinentes, claires et faciles à répondre pour tous les individus de cette population. Comme on a pu le constater dans la section précédente, la population visée par l'EDE était très diversifiée. Notamment, il fallait tenir compte des différences ethniques et culturelles des individus, de la présence de nombreux immigrants parmi les répondants et des concepts liés à l'ethnicité, tel que l'ascendance et l'identité ethnique, qui peuvent être difficiles à comprendre pour certaines personnes. Il fallait également considérer le fait que certains répondants, parmi ceux au Canada depuis plusieurs générations, se sentaient très peu ou pas du tout concernés par les questions liées à l'ethnicité.

Un autre aspect important dont on doit tenir compte lors de la conception d'un questionnaire est le caractère délicat de l'information que l'on cherche à recueillir. Dans le cas de l'EDE, certaines questions importantes pour l'enquête, telles que la religion, la perception de la discrimination et les crimes haineux, pouvaient être sensibles pour certaines personnes. Il fallait donc s'assurer que la formulation de ces questions ne constituait pas un trop grand fardeau de réponse pour les individus. Il fallait également établir un climat de confiance avec les répondants pour pouvoir espérer obtenir une réponse adéquate de la part de ceux-ci.



Afin d'élaborer le contenu du questionnaire, des comités d'experts ont été formés et consultés à plusieurs reprises. Ceux-ci ont aidé, entre autre, à établir le cadre conceptuel de l'appartenance ethnique, à déterminer les autres sujets susceptibles d'influencer ou d'être influencés par l'appartenance ethnique et à formuler les questions pour qu'elles soient claires et pertinentes pour les répondants. Ils ont également recommandé que chaque question sensible de l'enquête soit précédée par une introduction justifiant pourquoi une telle question était posée.

La figure 2 montre les différents thèmes abordés par le questionnaire de l'EDE. A l'intérieur du cercle, on retrouve les thèmes faisant partie du cadre conceptuel de l'appartenance ethnique et à l'extérieur du cercle, on retrouve les autres thèmes abordés par l'enquête susceptibles d'influencer ou d'être influencés par l'appartenance ethnique. Afin de mettre le questionnaire à l'essai, des études qualitatives sous forme de groupes de discussion et d'entrevues face à face avec rétroaction ont été effectuées. Ces études ont eu lieu à 5 reprises en français et en anglais à travers tout le Canada. Chaque série d'étude permettait de mieux comprendre la façon dont les gens répondaient aux questions, d'identifier certains problèmes et d'améliorer le questionnaire. Au total, une vingtaine de groupes de discussion et environ 150 entrevues personnelles auprès de personnes avec différents antécédents ethniques ont été effectuées.

Un test pilote à l'échelle nationale a également été effectué en septembre 2001. Lors de ce test, près de 1500 entrevues en anglais et en français ont été réalisées. Ce test pilote a permis d'identifier des problèmes qui n'avaient pas été remarqués jusque là, de peaufiner davantage le questionnaire et de mettre à l'essai les procédures de collecte. Suite à ce test, il a été décidé que pour l'enquête principale, une lettre d'introduction serait envoyée aux personnes sélectionnées avant la collecte afin qu'elles puissent mieux comprendre le but de l'enquête et se sentir plus concernées par celle-ci.

Comme le lecteur a pu le constater à la figure 2 ci-dessus, certains des thèmes abordés par le questionnaire étaient très spécifiques, par conséquent les questions résultantes devaient être posées directement aux répondants. Ceci représentait un défi de plus puisque plusieurs personnes sélectionnées pour l'enquête ne parlaient ni l'anglais ni le français. De plus, lors du test pilote, il a été remarqué que le nombre de non-réponse en raison d'une barrière

linguistique était plus élevé que ce que le recensement prédisait. Il a donc été décidé que le questionnaire de l'EDE serait traduit en 7 langues non officielles soit le mandarin, cantonais, italien, panjabi, portugais, vietnamien et espagnol. Le choix de ces langues s'est fait à partir des réponses données, lors du recensement, aux questions sur la connaissance des langues officielles et sur les langues parlées à la maison. Les 7 langues choisies correspondaient aux langues les plus souvent parlées à la maison par les personnes échantillonnées qui déclaraient ne connaître ni l'anglais, ni le français.

Traduire un questionnaire dans des langues non officielles comporte plusieurs défis. Ce ne sont pas toutes les langues qui ont des mots pour décrire les mêmes concepts. Le fait qu'aucun membre de l'équipe de l'EDE ne parlait les langues non officielles rendait la mise à l'essai de ces questionnaires difficiles. Heureusement, d'autres employés de Statistique Canada qui parlent ces langues ont accepté de participer à la mise à l'essai des questionnaires. De plus, certaines langues choisies sont extrêmement difficiles à lire même pour des individus qui les parlent parfaitement. L'embauche d'interviewers a été un peu difficile car plusieurs personnes qui parlaient très bien une ou plusieurs langues non officielles mais n'étaient pas capables de les lire. Finalement, les effectifs nécessaires ont été recrutés pour la collecte des données.

4. COLLECTE DES DONNÉES

La collecte des données s'est faite sous forme d'interviews téléphoniques assistées par ordinateur (ITAO) à partir des bureaux régionaux de Statistique Canada à Halifax, Sherbrooke, Sturgeon Falls, Toronto, Winnipeg et Vancouver. Elle a eu lieu entre les mois d'avril et d'août 2002, soit presque un an après le recensement, puisqu'il fallait attendre que la variable du recensement sur l'origine ethnique des ancêtres soit codée. Certaines composantes de l'ITAO qui avaient été peu ou pas utilisées auparavant dans des enquêtes de Statistique Canada ont été mises en place pour l'EDE. D'abord, l'application ITAO de l'EDE incluait un planificateur d'appels qui permettait de mieux gérer les rendez-vous et les cas de langues officielles afin d'assurer que chaque individu ait le maximum de chance d'être interviewé. Puis, étant donné la période écoulée entre le recensement et la collecte des données de l'EDE, une composante pour le dépistage a été incorporée à l'application ce qui facilitait le travail des personnes qui tentaient de retracer les répondants déménagés depuis le recensement.

Une autre des particularités de l'application de l'EDE concerne les questions où un grand nombre de choix de réponse était possible telles que les questions référant à un groupe ethnique ou culturel, à une langue ou à un pays. Puisque le codage des réponses écrites est une étape qui demande énormément de temps et d'énergie, l'EDE voulait réduire au minimum le nombre de questions ne comportant pas de choix de réponses pré-établies. Pour les questions ne comportant que peu de choix possibles, les choix de réponse étaient incorporés sur le même écran que la question et la réponse appropriée pouvait facilement être repérée. Cependant, pour les questions comportant plus d'une trentaine de choix de réponse, cette approche était impossible. Pour ces questions, les programmeurs de l'application de l'EDE ont dû emmagasiner les choix de réponse séparément de l'écran contenant la question et créer un outil de recherche permettant de trouver la réponse appropriée dans la longue liste de réponses possibles. Cet outil de recherche a été conçu de telle façon qu'il permettait de repérer la réponse appropriée en tapant les premières lettres de la réponse (recherche en mode «alphabétique») ou en tapant un groupe d'au moins trois lettres contenues dans la réponse appropriée (recherche en mode «triagramme»).

De plus, l'application de l'EDE a été conçue de sorte que certaines réponses données étaient insérées dans des questions subséquentes. Ainsi, l'application facilitait l'approfondissement ou le suivi des réponses et limitait certaines questions, par exemple celles sur les langues, à quelques choix de réponse. Par exemple, pour mieux comprendre comment l'appartenance ethnique d'une personne influence sa vie sociale, les origines ethniques les plus importantes du répondant selon celui-ci étaient insérées dans des questions de suivi tel que «A votre connaissance, combien de vos amis ont une ascendance».

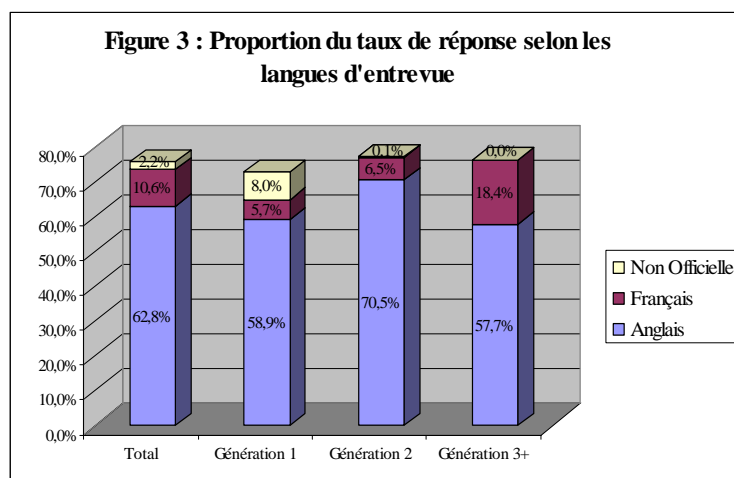
Cependant, étant donné les ressources que cela aurait exigées, l'application n'a pas été développée dans les langues non officielles. Les interviewers affectés aux entrevues dans ces langues possédaient une copie papier du questionnaire traduit et devaient effectuer l'entrevue à l'aide de cette copie tout en saisissant simultanément en anglais ou en français les renseignements recueillis dans l'application ITAO.

Une partie des cas nécessitant une entrevue dans une langue non officielle a pu être identifiée avant la collecte grâce aux renseignements fournis lors du recensement. Ces cas ont donc été acheminés directement aux bureaux régionaux de Toronto ou de Vancouver d'où avaient lieu la majeure partie des entrevues dans les langues non officielles. Cependant, comme il avait été constaté lors du test pilote, une partie des cas nécessitant une entrevue dans une langue non officielle ne pouvait être identifiée à l'aide des renseignements fournis lors du recensement. Il a donc été nécessaire de développer une stratégie différente pour traiter ces cas puisqu'ils n'étaient pas envoyés dans les bureaux régionaux où les effectifs linguistiques étaient suffisants et que la technologie de Blaise ne permet pas l'envoi de cas d'un bureau régional à un autre. Pour ces cas, les entrevues ont été effectuées par l'entremise d'une séance de «Net Meeting». Ce système demandait plus de coordination entre les différents bureaux régionaux puisqu'il consistait à établir une connexion, par le biais du réseau interne de Statistique Canada, entre un ordinateur duquel était effectué l'entrevue avec l'individu et un ordinateur situé dans un autre bureau régional où le cas de l'individu en question pouvait être accédé.

De plus, aucune composante du planificateur des appels permettant de rediriger automatiquement les cas de langue non officielle aux intervieweurs compétents n'a pu être développée pour l'enquête. Les intervieweurs affectés aux entrevues dans les langues non officielles ont donc dû accéder à leurs cas de façon manuelle et gérer eux-mêmes leurs rendez-vous. Somme toute, bien que les entrevues dans les langues non officielles prenaient plus d'organisation et de temps pour les administrer (50 minutes contre 35 à 45 minutes en français ou en anglais), il a été utile de traduire le questionnaire en plusieurs langues comme en fait foi les taux de réponses.

Au total, un taux de réponse de 75,6% a été obtenu, ce qui est excellent si on considère que l'enquête avait lieu un an après le recensement et que l'échantillon était composé de gens plutôt mobiles. Ce taux de réponse a été atteint grâce au dépistage effectué avant le début de la collecte, aux lettres d'introduction envoyées aux individus sélectionnés, à la formation qui a été donnée aux interviewers et à la traduction des questionnaires dans plusieurs langues.

La figure 3 donne une idée de l'impact qu'a eu la traduction des questionnaires sur les taux de réponse. Il montre la proportion du taux de réponse qui est dû aux entrevues en français, en anglais et dans une langue non officielle au niveau global et par génération.



On peut voir que les entrevues dans les langues non officielles ont permis d'augmenter le taux de réponse global de 2.2 %, ce qui est tout de même non négligeable. Pour les immigrants de première génération, on peut voir que la traduction des questionnaires dans les langues non officielles a eu beaucoup plus d'impact. Dans cette population, elle a permis d'augmenter le taux de réponse de 8%.

Notons que des quelques 13 000 personnes qui n'ont pas répondu à l'EDE, un peu plus de 900, pour la plupart des immigrants de première génération, ne l'ont pas fait en raison d'une barrière de langue (environ 7 %

de non-réponse). Ce nombre de non-répondants aurait pu être diminué de 40% si la période de collecte avait été prolongée car, à la fin de la collecte, 365 cas étaient en liste pour une entrevue dans une des langues traduites.

5. PONDÉRATION

Le poids final attribué à chaque répondant de l'enquête a été calculé à partir du poids du plan de sondage auquel un facteur d'ajustement pour la non-réponse et un facteur d'ajustement pour la post-stratification ont été appliqués. Le poids du plan de sondage, appelé également le poids initial, correspond à l'inverse de la probabilité d'être choisi pour l'enquête. Pour l'EDE, cette probabilité correspond au produit de la probabilité d'avoir reçu un questionnaire

long au dernier recensement, c'est-à-dire environ une chance sur 5, et de la probabilité de sélection à la deuxième phase qui correspond au nombre de personnes choisies dans la strate par rapport au nombre de total de personnes dans cette strate.

Puisque beaucoup d'information sur les non-répondants était disponible grâce au recensement, la méthode de propension à répondre a été utilisée afin d'ajuster les poids pour la non-réponse. Cette méthode consiste à prédire à l'aide d'un modèle de régression logistique la probabilité qu'a chaque personne de répondre à l'enquête. Ensuite, les individus qui ont des probabilités de réponse semblables sont regroupés en classes et les poids des individus sont ajustés par l'inverse du taux de réponse obtenu dans leur classe.

Tel que mentionné plus tôt, 13 000 personnes n'ont pas répondu à l'EDE. Ces non-répondants avaient différentes raisons de ne pas répondre à l'enquête. Il a été remarqué que ceux-ci pouvaient être classifiés selon deux groupes : les non-contactés et les autres non-répondants. Étant donné que ces deux groupes avaient des caractéristiques distinctes, il a été décidé que deux ajustements pour la non-réponse, un pour chaque groupe, seraient effectués.

Au niveau de la post-stratification, les poids d'enquête ont été ajustés aux totaux du recensement pour la géographie, certains groupes d'âge, le sexe et les strates de sélection. Le facteur d'ajustement pour la post-stratification correspond au rapport entre le total connu de la population et le total des poids après ajustement de la non-réponse dans chaque post-strate formée par le croisement de la géographie, des groupes d'âge, du sexe et des strates de sélection.

6. ESTIMATION DE LA VARIANCE

La méthode d'auto-amorçage, mieux connue sous le nom de méthode du «bootstrap», a été utilisée pour estimer la variance. La méthode d'auto-amorçage est une méthode de rééchantillonnage et consiste à :

- 1) sélectionner un nombre suffisant M de sous-échantillons aléatoires simples avec remise de taille $n_h - 1$ à partir de l'échantillon principal et ce, indépendamment pour chaque strate h ;
- 2) attribuer un poids aux individus sélectionnés dans les sous-échantillons en tenant compte du poids initial, du sous-échantillonnage avec remise et des différents ajustements apportés aux poids de l'enquête principale;
- 3) estimer, dans chaque sous-échantillon, la caractéristique pour laquelle une approximation de la variance est désirée;
- 4) calculer la variance empirique des M estimations ainsi calculées.

L'estimation de la variance par auto-amorçage pour la caractéristique étudiée correspond à la variance empirique de ses M estimations. On a donc :

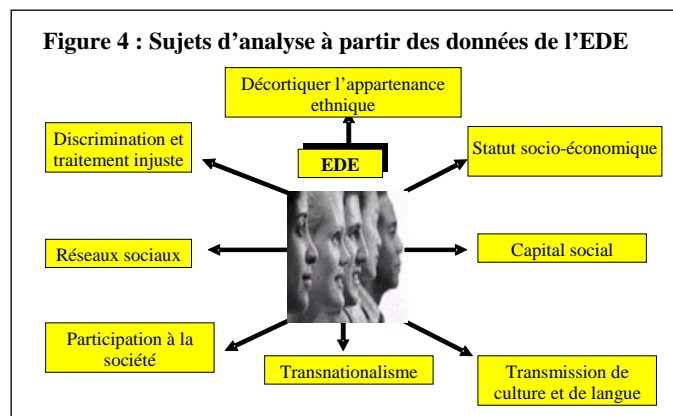
$$\hat{V}_B(\hat{\theta}) = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}_{Bi} - \hat{\theta})^2 \quad \text{où } \hat{\theta} \text{ et } \hat{\theta}_{Bi} \text{ sont respectivement l'estimation de la caractéristique obtenue en utilisant les poids d'enquête et celle obtenue à partir des poids d'auto-amorçage du sous-échantillon } i.$$

Le choix particulier de la taille, $n_h - 1$, dans chaque strate simplifie la formule des poids bootstrap initiaux (Rao, Wu et Yue (1992) et Rao et Wu (1988)). Puis, pour ce qui est de l'ajustement pour la non-réponse, les résultats de Langlet, Faucher et Lesage (2003) ont été considérés et seul un ajustement au niveau des classes de non-réponse obtenues lors de la pondération a été fait.

7. PROCHAIN DÉFI

Le prochain défi pour l'enquête sur la diversité ethnique est l'analyse des données. La figure 4 donne un aperçu de certains sujets d'analyse potentiels.

A cette fin, un fichier analytique a été mis à la disposition des chercheurs dans les centres de données de recherche de Statistique Canada. La diffusion d'un fichier de microdonnées à grande diffusion est également prévue pour mai 2005.



8. CONCLUSION

Mener une enquête dans une population diversifiée comportait plusieurs défis méthodologiques. Il fallait choisir un plan de sondage qui permettait de rejoindre certaines minorités de la population, tenir compte des différences culturelles, ethniques et linguistiques des répondants lors de l'élaboration du contenu et se donner les moyens de recueillir l'information voulue. C'est, entre autre, en choisissant de prendre le recensement comme base de sondage, en traduisant les questionnaires dans des langues non-officielles et en utilisant une application ITAO mieux adaptée pour recueillir des données complexes que l'enquête sur la diversité ethnique a pu être menée à bien. Devant la croissance imminente de la diversité ethnique, culturelle et linguistique au Canada, ces pratiques pourraient servir d'inspiration lors de l'élaboration d'enquêtes futures.

RÉFÉRENCES

- Langlet, E.R., Faucher, D. et Lesage, E. (2003), "An Application of the Bootstrap Variance Estimation Method to the Canadian Participation and Activity Limitation Survey", *2003 Proceedings of the American Statistical Association*, Section on Survey Research Methods, Alexandria, VA: American Statistical Association, pp. 2299-2306.
- Rao, J.N.K., Wu, C.F.J. et Yue, K. (1992). "Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes", *Techniques d'enquête*, 18, n° 2, pp. 225-234.
- Rao, J.N.K., et Wu, C.F.J. (1988), "Resampling inferences with complex survey data", *Journal of the American Statistical Association*, 83, pp. 231-241.