

Note

« La création de la base de données du recensement canadien de 1852 »

Lisa Dillon

Cahiers québécois de démographie, vol. 34, n° 2, 2005, p. 297-313.

Pour citer cette note, utiliser l'adresse suivante :

<http://id.erudit.org/iderudit/014013ar>

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <http://www.erudit.org/apropos/utilisation.html>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : erudit@umontreal.ca

La création de la base de données du recensement canadien de 1852

Lisa DILLON *

Un échantillon du recensement effectué en 1852 dans le Canada-Est et le Canada-Ouest, sous forme de fichier lisible par machine, s'ajoutera bientôt à la séquence de bases de données en cours de construction à partir des recensements canadiens des 19^e et 20^e siècles. La base contiendra 259 000 personnes, soit 20 pour cent de la population recensée en 1852. Similaire aux recensements ultérieurs par la forme et le contenu, celui-ci présente cependant des difficultés particulières au chapitre de la représentation de toutes les parties du territoire et du repérage des chefs de ménage. Des étapes clés du déroulement du projet sont décrites ici : organisation de la saisie des données, résolution du problème de la perte de données sur les zones urbaines, repérage des chefs de ménage au moyen de l'information sur le type d'immeuble. Une première analyse des fréquences du recensement de 1852 et leur comparaison avec celles du recensement de 1871 laisse croire que l'échantillon construit est représentatif de la population rurale du Canada en 1852.

English abstract, p. 313.

Depuis plus de trente ans, des chercheurs travaillent à la création de bases de données qui permettront d'exploiter les recensements canadiens des 19^e et 20^e siècles. Il sera question ici de l'une des plus récentes réalisations inscrites dans ce grand projet, soit l'échantillon informatisé du recensement effectué en 1852 dans le « Canada-Est » et le « Canada-Ouest », équivalents du Québec et de l'Ontario d'aujourd'hui. Similaire aux recensements ultérieurs par la forme et le contenu, le recensement de 1852 soulève néanmoins des difficultés qui lui sont propres eu égard à la représentation de toutes les parties du territoire et au repérage des chefs de ménage. La création d'une base de données tirée de ce recensement suit son cours à l'Université de Montréal, dans le cadre des activités du Programme de recherche en démo-

* Département de démographie, Université de Montréal. Texte traduit de l'anglais par Gilbert Gendron.

graphie historique (PRDH) du Département de démographie. Le PRDH gère déjà le Registre de la population du Québec ancien (RPQA), base de données informatisée intégrant l'ensemble des faits d'état civil (baptêmes, mariages et sépultures) relatifs à la population française de la vallée du Saint-Laurent de 1621 à 1799; il s'emploie maintenant à enrichir ce fonds en y ajoutant les données censitaires du 19^e siècle. Notre travail sur le recensement de 1852 est financé par la Fondation canadienne pour l'innovation (volet « infrastructures »), dont la subvention quadriennale contribue à l'implantation des infrastructures matérielles et virtuelles et des réseaux nécessaires pour construire des bases de données sur les populations anciennes alimentées par divers canaux (saisie d'informations originales, acquisition de corpus existants, notamment de données généalogiques...), avec l'appui des organismes subventionnaires et de l'Université ¹.

LES DONNÉES DÉMOGRAPHIQUES DU 19^E SIÈCLE

La première étape de la création de la base de données du recensement de 1852 a consisté à préparer un échantillon de 20 pour cent de la population du Canada-Ouest et du Canada-Est touchée par le dénombrement. Nous aboutirons à un échantillon exploitable par machine qui se situera au début de la séquence de bases de microdonnées individuelles à grande diffusion tirées des recensements canadiens des 19^e et 20^e siècle, série quasi complète incluant les années de recensement 1871 à 1951 et 1971 à 2001 ². Comme les autres bases québécoises et canadiennes, celle-ci est conçue de manière à permettre les comparaisons internationales, en particulier avec les échantillons du recensement américain de 1850 et du recensement effectué en Angleterre et au Pays de Galles en 1851. L'intégration de ces trois recensements nationaux du milieu du 19^e siècle sera ensuite réalisée dans le cadre du North Atlantic Population Project (Roberts et al., 2003). En outre, nous espérons mettre en bonne voie les collaborations nécessaires pour jumeler les données du recensement de 1852 avec celles des registres paroissiaux québécois du début du 19^e siècle, à l'aide du fonds numérisé par l'Institut généalogique Drouin ³, que le PRDH a acquis afin de faciliter la réalisation de projets de ce genre. À l'échelle du Québec, la période 1800-1852 accuse pour l'instant un déficit de recherches en démographie historique, en grande partie à cause du manque de données couvrant le territoire entier : les chercheurs intéressés à la famille et aux structures sociales ont

pu étudier la fin du 19^e siècle et le début du 20^e siècle, périodes sur lesquelles on trouve des données détaillées, ou encore des régions — comme le Saguenay et Charlevoix — pour lesquelles l'information sur le début du 19^e siècle est disponible. Mais le manque de microdonnées systématisées sur les provinces et le Canada nous prive à la fois d'une vue d'ensemble et d'une bonne connaissance de la population et de la société canadiennes du début du 19^e siècle.

Cependant, un certain nombre de projets de recherche en démographie historique actuellement en cours vont contribuer à combler les lacunes. À l'Université du Québec à Chicoutimi, la base de données BALSAC, qui contient déjà les informations tirées des registres paroissiaux du 19^e siècle pour les régions du Saguenay et de Charlevoix, incorporera à terme les données des actes de l'état civil de l'ensemble de la population du Québec, du 17^e siècle à nos jours (Bouchard, 1992). De conception un peu similaire, le Registre de la population du Québec ancien intègre les faits d'état civil antérieurs à 1800 consignés dans les registres paroissiaux, et le démographe Bertrand Desjardins est en train d'y ajouter les actes de sépulture des catholiques décédés à un âge avancé entre 1800 et 1852. L'échantillon de 20 pour cent de la population recensée en 1852 dans le Canada-Est et le Canada-Ouest prolonge ces efforts de constitution d'un fonds de données historiques propre à éclairer la réalité démographique du Canada de la première moitié du 19^e siècle.

ÉCHANTILLONNAGE ET SAISIE DES DONNÉES

La première bande-échantillon tirée d'un recensement du Canada a été créée à York University, à la fin des années 1970, dans le cadre du Canadian Historical Social Mobility Project de l'Institute of Social Research. Il s'agissait d'un échantillon d'un pour cent du recensement de 1871 (Darroch et Ornstein, 1994). Cette expérience et d'autres similaires, tels l'Integrated Public-use Microdata Series de l'Université du Minnesota et le Projet de recherche sur les familles canadiennes, ont fait ressortir un consensus chez les créateurs de bases de données eu égard aux caractéristiques des échantillons : représentativité nationale, stratification géographique, informations sur les individus, sélection aléatoire de grappes de logements et de ménages (Ruggles, 1995; Ornstein, 2000 : 195). Les données originales du recensement de 1852 remplissent la première condition, dans la mesure où elles fournissent un décompte de la population

des deux colonies britanniques — le Canada-Est et le Canada-Ouest — réunies en 1840 par l'Acte d'Union (Conrad et al., 1993; Curtis, 2001 : 98-100). Des recensements ont été effectués au même moment dans deux autres colonies britanniques gouvernées séparément, les actuelles provinces du Nouveau-Brunswick et de la Nouvelle-Écosse (Curtis, 2001 : 256). Faute de temps et de moyens, aucune des deux n'a été incluse dans notre projet. Dans le premier cas, les pages de recensement contiennent un nombre très variable de personnes dénombrées, ce qui complique l'échantillonnage par pages. Dans le second, le dénombrement a porté uniquement sur les chefs des ménages et non sur tous leurs membres.

Afin de choisir une stratégie d'échantillonnage, nous avons examiné les méthodes utilisées pour les projets apparentés au nôtre. Ayant décidé de sous-traiter la saisie de données pour ses projets les plus récents, le Minnesota Population Center était passé de l'échantillonnage de logements à l'échantillonnage de pages de recensement, afin de faciliter le travail et d'en assurer la qualité. Les échantillons demeuraient représentatifs au plan national, aléatoires, stratifiés géographiquement et prélevés par grappes, mais la sélection de logements à intervalles de dix était remplacée par la sélection de fenêtres de vingt-cinq lignes séparées par un certain nombre de pages.

Nous avons retenu cette dernière stratégie, parce que c'était la mieux adaptée aux difficultés posées par les relevés (manuscrits) des énumérateurs de 1852. Tout d'abord, à la différence des recensements ultérieurs, celui-ci ne prévoyait pas, dans la présentation des documents, de colonnes destinées à la numérotation des logements et des ménages. Les répondants sont groupés de façon claire par logements et ménages; la présentation des informations sur le nom, le statut matrimonial, le sexe et l'âge permet de regrouper les familles. et la colonne prévue pour la description de l'immeuble contient l'information sur le nombre de logements. Mais les deux ensembles de renseignements ne sont pas reliés par une numérotation systématique. Impossible donc de prendre des logements comme points d'échantillonnage, comme l'ont fait les créateurs de l'échantillon du recensement de 1901. Chaque page du recensement de 1852 contient cinquante lignes (et un individu par ligne). Chaque ligne se subdivise en quarante et une colonnes, réparties sur quatre pages, lesquelles occupent trois images numérisées tirées du microfilm original. Les traits horizontaux séparant les informations sur les cinquante individus sont parfois à-demi effacés, et les lignes ne sont

numérotées que sur la première page de l'original, tandis que la seule variable permettant de distinguer entre les ménages, soit la colonne consacrée à l'information sur l'immeuble, apparaît à droite en deuxième page. Pour échantillonner les logements il aurait fallu, au moment de la saisie, beaucoup de va-et-vient et de vérification sur les images numérisées pour associer à chaque logement la séquence de noms appropriée. C'est pourquoi, plus simplement, nous avons opté pour l'échantillonnage par pages. Nous avons échantillonné (jusqu'ici) 240 836 individus de la population dénombrée au recensement de 1852. L'échantillon final totalisera quelque 259 000 personnes, y compris 20 pour cent de l'échantillon de 100 pour cent des citoyens de Québec en 1852 construit par Marc St-Hilaire et Richard Marcoux, du Centre interuniversitaire d'études québécoises (CIEQ) de l'Université Laval (Marcoux et al., 2003).

La saisie des données du recensement de 1852 a été facilitée par les images numérisées des documents censitaires produites à notre intention par la Division des services en ligne de Bibliothèques et archives Canada, à partir des microfilms tirés des originaux. La Division a ainsi réalisé 86 706 reproductions (numérotées par bobine de microfilm)⁴. Il nous a fallu beaucoup d'heures, au début du projet, pour classer ces images par province, district et subdivision de district afin d'assurer la stratification géographique de l'échantillon (la sélection des pages de recensement devant se faire par subdivision). Une fois les images classées, nous avons construit une base de données dans laquelle chacune d'elles était identifiée, et notre programmeur a mis au point un système de sélection des pages permettant de tirer un échantillon d'images présélectionnées. Les préposés à la saisie des données pouvaient alors télécharger les images, par séries de cinq : un bloc de trois images contenant les quatre côtés de pages de recensement à inclure dans l'échantillon et les deux images subséquentes, relatives au groupe suivant de cinquante individus. Si la fin du bloc échantillonné ne coïncidait pas avec la fin de l'information relative à un logement, il fallait compléter cette dernière tranche d'information à l'aide des images subséquentes. Parallèlement, dans l'échantillon final, nous éliminerons les fins d'information apparaissant au début des blocs échantillonnés (cas où l'information relative à un logement commence sur une image non échantillonnée et se poursuit sur une image échantillonnée). Pour certaines subdivisions où moins de 250 personnes ont été dénombrées, le taux d'échantillonnage a dépassé 20 pour cent.

Pour afficher les images numérisées à l'écran, nous avons utilisé le logiciel Adobe Acrobat 6.0, qui permet d'agrandir les informations à déchiffrer, tels les noms et les prénoms. Comme beaucoup de renseignements se répétaient, par exemple sur le lieu de naissance et la religion, et qu'il n'y avait souvent rien d'inscrit dans certaines cases (où auraient pu figurer des noms de magasins, d'édifices publics, de lieux de culte), les préposés ont gagné du temps en saisissant l'information par colonne plutôt que par ligne. Vu l'absence de numérotation des lignes sauf à la première page de chaque bloc, notre programmeur a fait en sorte que les autres pages s'affichent sur l'écran à droite de colonnes statiques, à fond gris, reprenant l'information sur le nom, le statut matrimonial, le sexe et l'âge (information apparaissant automatiquement dans les colonnes grises dès sa saisie). On pouvait ainsi à tout moment, tandis que les pages défilaient sur l'écran à mesure que de nouvelles données s'y inscrivaient, voir à quelle personne celles-ci se rapportaient. De plus, des codes couleurs (deux tons de bleu pour les veufs et les veuves, de rose pour les célibataires des deux sexes, d'orange pour les hommes et les femmes mariés, du lie-de-vin pour les enfants) permettaient de garder plus aisément le fil des informations saisies. Le classement des images par dossiers (districts et subdivisions de districts) a facilité la supervision des travaux et le contrôle de leur qualité, en simplifiant le repérage des pages dont nous avons besoin en cas de problème. Au stade du contrôle de la cohérence et du nettoyage des données, il nous a permis d'effectuer rapidement des vérifications à l'aide des relevés du recensement.

DONNÉES MANQUANTES

À la différence des relevés du recensement de 1871 (et des recensements subséquents sur lesquels s'effectue le travail de création de bases de données), les documents manuscrits du recensement de 1852 ont été en partie détruits avant que les Archives nationales du Canada entreprennent de les mettre sur microfilm, dans les années 1950. Selon l'inventaire dressé conjointement par le personnel des Archives et notre équipe, il manque 348 des 1274 subdivisions et divisions que comptaient le Canada-Ouest et le Canada-Est en 1852, c'est-à-dire, d'après les données agrégées publiées par le Bureau du recensement, 498 844 personnes, sur une population totale de 1 828 482 personnes. Ainsi, 27 pour cent de la population recensée en 1852

serait disparue avec ces documents originaux, et notre opération de transcription des données aurait porté sur seulement 73 pour cent de la population recensée, soit 1 295 505 personnes.

Notre échantillon de 20 pour cent représente par conséquent 14 pour cent environ de la population totale du Canada-Ouest et du Canada-Est en 1852 (et moins encore après élimination des fins d'information présentes au début des blocs de pages échantillonnés). Les subdivisions et divisions sur lesquelles subsiste la documentation manuscrite n'étant pas représentatives de l'ensemble, l'échantillon est faussé. Par exemple, les renseignements sur la plupart des grandes villes canadiennes, notamment la quasi-totalité de Montréal et de Toronto, outre quelques villes de taille moindre comme Kingston, London et St. Catharines, ont été perdus. Nous ne pouvons pallier l'absence des habitants des grandes villes dans notre échantillon, si ce n'est en ajoutant à notre base de données l'échantillon (pondéré adéquatement) de 100 pour cent de la population de la ville de Québec créé par le CIEQ, et peut-être en suréchantillonnant le quartier électoral montréalais de Saint-Louis, toute la ville de Hamilton et les quartiers électoraux d'Ottawa-est et d'Ottawa-ouest (si nous en avons le temps et les moyens).

Nous devrions cependant pouvoir compenser les pertes qui concernent les régions rurales en construisant un échantillon de microdonnées pour étudier celles-ci. Nous avons d'abord envisagé de suréchantillonner les subdivisions géographiquement et sociodémographiquement proches des subdivisions « perdues », en utilisant pour certaines variables des techniques d'imputation de données manquantes inspirées des programmes de *hot-decking* utilisés notamment pour le projet IPUMS (sur ce sujet, voir Dillon, 2000). Après discussion avec des collègues de l'Institut national de la recherche scientifique (INRS-Urbanisation, Culture et Société), de l'Université York et de l'Université de Toronto, nous avons écarté cette possibilité. Premièrement, il serait difficile de dresser une liste de critères pour choisir un ensemble de subdivisions à utiliser comme source de données puisque chaque utilisateur de la base l'interrogera en fonction de ses propres préoccupations. Consacrer du temps et de l'argent à la mise en banque de données sur des unités géographiques choisies en fonction de certains critères ne résoudrait le problème du déséquilibre de l'échantillon que pour les chercheurs dont les questions de recherche correspondent à ces critères. Deuxièmement, les données agrégées du recensement de 1852 — à partir desquelles seraient choisies les subdivisions — posent

problème. En les utilisant pour créer une série de feuilles de calcul Excel, nous sommes tombés sur des cas où, pour certaines variables, les sous-totaux n'étaient pas égaux à la population totale de la subdivision, ou encore les totaux se rapportaient à plusieurs subdivisions. Comme la précision des données agrégées n'est pas assurée pour l'ensemble des subdivisions, nous avons renoncé au suréchantillonnage de subdivisions voisines des subdivisions perdues.

Pour résoudre le problème, nous avons plutôt décidé de faire la liste des types de données manquantes, puis de faire une comparaison des fréquences entre nos microdonnées et les données agrégées, pour en déduire des pondérations à appliquer aux premières afin de les calibrer en fonction des secondes. Cela ne corrigera pas le déséquilibre causé par l'absence des grandes villes, mais pourrait pallier le manque de données relatives aux régions rurales. L'avantage des pondérations est leur flexibilité : il est possible de les modifier en fonction des questions posées par les chercheurs. Néanmoins, nous demeurerions obligés de nous limiter aux variables pour lesquelles les données agrégées paraissent sûres.

TRAVAIL À VENIR

La prochaine étape de notre travail sera le contrôle de la cohérence des données, leur nettoyage et leur codification. L'une des grandes difficultés qui se posent à nous est celle de diviser les 50 personnes de chaque page de recensement en groupes reliés chacun à un logement, à l'aide de l'information sur l'immeuble notée par les agents recenseurs. Voici des exemples de ce qu'on trouve à la 31^e colonne des relevés : maison en rondins, maison à charpente de bois, cabane, cabane en rondins, maison en pièces, maison en bois, maison en pierre. Lors de la préparation de notre procédure d'échantillonnage, il nous a semblé que ces descriptions avaient probablement été fournies par les personnes qui étaient les chefs de ménage, c'est-à-dire, le plus souvent, des hommes mariés d'âge adulte.

Au cours de la saisie des données, il a parfois fallu choisir nous-mêmes la ligne à laquelle associer l'information sur le type d'immeuble, étant donné la quasi-invisibilité des séparations entre les lignes et l'absence de numérotation de ces dernières. Ayant le choix entre quatre lignes (plus ou moins), les préposés à la saisie ont tendu à choisir celle où figurait une personne qui, selon toute apparence, était le chef de ménage.

À l'heure actuelle, nous avons accumulé les cas transcrits en nombre suffisant pour faire le portrait des personnes qui ont fourni l'information sur leur immeuble lors du recensement et sont inscrites à la même ligne que cette information. Nous avons comparé les caractéristiques de ces « chefs de ménage présumés » à celles de l'échantillon de population de 1852 dans son ensemble, ainsi qu'aux chefs de ménage « avérés » de l'échantillon du recensement de 1871, plus faciles à reconnaître du fait que cette année-là, sur les relevés, une colonne était prévue pour la numérotation des logements. Puisque, dans les circonstances, l'échantillon de la population énumérée en 1852 dans le Canada-Ouest et le Canada-Est est en grande partie rural, la comparaison présentée au tableau 1 porte sur les caractéristiques démographiques des chefs de ménage « présumés » qui vivaient dans les régions rurales de ces deux colonies en 1852 et des chefs de ménage qui vivaient dans les régions rurales de l'Ontario et du Québec en 1871⁵. Il en ressort que la plupart des chefs de ménage présumés de 1852 étaient bel et bien chefs de ménage. Par exemple, dans l'ensemble de l'échantillon de 1852, on compte 52 pour cent d'hommes et 48 pour cent de femmes, mais le taux de masculinité des chefs de ménage présumés (94 pour cent) est voisin de celui des chefs de ménage avérés de 1871 (93 pour cent). On pouvait s'attendre à ce que la plupart des chefs de ménage soient des hommes, exception faite de quelques veuves. Les deux groupes de chefs se ressemblent également par l'âge (figure 1) : la plupart se situent entre 20 ans et 59 ans (85 pour cent en 1852, 82 pour cent en 1871). De même, la plupart sont mariés : 87 pour cent (comparativement à 31 pour cent pour l'ensemble de la population rurale) en 1852 et 85 pour cent en 1871.

Le recensement de 1852 contient aussi des informations socioéconomiques qui laissent supposer que les personnes qui ont communiqué les renseignements sur les immeubles étaient vraiment les chefs de ménage. On s'attend à ce que les chefs de ménage aient un emploi. Or, en 1852, alors que les deux tiers des ruraux n'ont déclaré aucun métier, seulement 9 pour cent des chefs de ménage présumés sont dans le même cas (8 pour cent en 1871 : voir le tableau 1). L'information de 1852 sur les métiers n'ayant pas encore été codifiée, l'analyse économique ne peut être approfondie à ce stade-ci. Néanmoins, sur les 25 premières déclarations de la liste que l'on obtient pour l'ensemble de la population, six ne se rapportent pas au monde du travail (ex. « aucun », « épouse », « femme célibataire », « fille »). Par con-

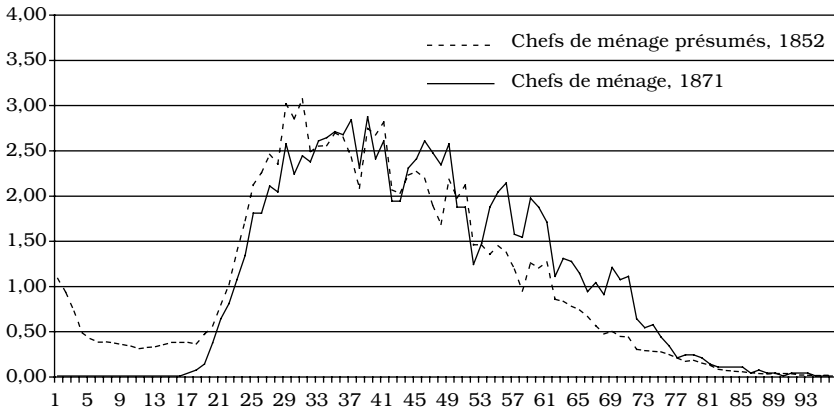
TABLEAU 1. *Caractéristiques des chefs de ménage présumés (1852), des chefs de ménage (1871) et de l'ensemble de la population (1852 et 1871), Canada, pour la population rurale*^a

| | Recensement de 1852 | | Recensement de 1871 | |
|----------------------------|---------------------|--------------------------|---------------------|-----------------|
| | Tous | Chefs de ménage présumés | Tous | Chefs de ménage |
| % hommes:% femmes | 52:48 | 94:6 | 51:49 | 93:7 |
| % mariés | 31 | 87 | 32 | 85 |
| % célibataires | 67 | 7 | 66 | 6 |
| % âgés de 1 à 19 ans | 55 | 2 | 54 | 0 |
| % âgés de 20 à 59 ans | 41 | 85 | 41 | 82 |
| % sans profession déclarée | 66 | 9 | 74 | 8 |
| % nés hors du Canada | 26 | 45 | 17 | 40 |
| N | 236 753 | 36 380 | 16 872 | 2 836 |

a. Comparaison entre la population rurale recensée dans le Canada-Ouest et le Canada-Est en 1852, et la population rurale recensée en Ontario et au Québec en 1871.

tre, les 25 premières déclarations faites par les chefs présumés désignent toutes des métiers (agriculteur, ouvrier, menuisier, cordonnier, forgeron, marchand, tailleur, tonnelier, tisserand...). D'autre part, on pouvait s'attendre à ce que la proportion de personnes nées à l'étranger soit plus élevée chez les chefs de ménage que chez les « non-chefs », parmi lesquels figurent en principe les enfants nés au Canada. Or près de la moitié des chefs de ménage présumés de 1852 ne sont nés ni dans le Canada-Ouest, ni dans le Canada-Est, ni dans les colonies voisines du Nouveau-Brunswick et de la Nouvelle-Écosse (comparativement à 26 pour cent de l'ensemble des Canadiens : voir le tableau 1).

Étant donné que les préposés à la saisie des données ont dû, à l'occasion, attribuer un immeuble à un individu, il nous faudra voir s'il y a des différences entre leurs suppositions respectives. La comparaison des caractéristiques des personnes désignées comme chefs de ménage par chacun d'eux peut nous renseigner sur les hypothèses qu'ils ont faites pour choisir la ligne où placer l'information sur l'immeuble. On sait qu'un logement pouvait abriter plusieurs ménages ou familles. Lorsque nous aurons fini de relier logements et groupes d'habitants et de repérer les chefs de ménage, nous concevrons un programme pour tâcher de repérer les seconds chefs de ménage (et autres), par une série de critères.



Sources : Lisa Dillon et Alexandre Bujold, *1852 Canadian Public-Use Microdata Sample*, Historical Demography Research Infrastructure, 2005; Gordon Darroch et Michael Ornstein, *1871 Canadian Census Data* (Canadian Historical Mobility Project), Institute of Social Research, York University, 1979.

Note. Ajustement des distributions par âge : moyennes mobiles de trois ans.

FIGURE 1. Distribution selon l'âge des chefs de ménage présumés de 1852 (Canada-Ouest et Canada-Est) et des chefs de ménage de 1871 (Ontario et Québec), en milieu rural (échantillons de données censitaires, %)

Une question du recensement de 1852 ayant trait à l'appartenance au ménage pourrait nous aider à discerner les familles et les ménages à l'intérieur des logements. D'après la feuille de route des agents recenseurs, les personnes n'appartenant pas à une famille étaient « celles qui se sont arrêtées dans la maison cette nuit-là, mais qui ne sont pas membres de la famille, entre autres les voyageurs, pensionnaires, ecclésiastiques et serviteurs ». Malheureusement, on ne précise pas la manière de classer les familles et ménages occupant à plusieurs un même logement. Une première analyse des déclarations pour la variable « membre de la famille, ou non » révèle que certains recenseurs classaient certaines personnes comme non-membres, tandis que d'autres inscrivaient toutes les personnes présentes comme membres. En ce qui a trait aux seconds chefs de ménage, nos déductions reposeront probablement sur des variables simples comme le nom, l'âge, le statut matrimonial, le métier et la position occupée dans la liste des occupants du logement (sur le relevé). Ce travail sera concomitant à la recherche du « lien avec le chef de ménage » et des autres variables représentant les relations familiales.

TABLEAU 2. Nombre de réponses différentes fournies pour certaines variables au recensement de 1852

| | |
|-----------------------------------|------|
| Profession | 4106 |
| Lieu de naissance | 4956 |
| Religion | 2010 |
| Lieu de résidence si hors limites | 1545 |
| Type de maison | 1352 |

Une analyse préliminaire des fréquences de diverses variables du recensement de 1852 a donné les résultats attendus, mais aussi causé quelques surprises. La codification de certaines variables prendra du temps. À titre d'exemple, la question sur le métier a suscité 4106 déclarations différentes, la question sur le lieu de naissance 4956, et la question sur la religion 2010 (tableau 2). Comme prévu, la question sur la « couleur » n'a rien donné, car 1042 personnes seulement ont déclaré être noires, et 1399 personnes se sont dites aborigènes. Le nombre de réponses a été faible aussi pour les questions sur la surdité (180) et la cécité (124); un peu plus de gens (349) ont été déclarés déments, idiots ou aliénés. Moins de gens allaient à l'école en 1852 (12 pour cent) qu'en 1871 (19 pour cent). La fécondité ayant diminué légèrement entre les deux dates, on pouvait s'attendre à l'évolution inverse, ou à une stagnation, ce qui laisse supposer que cette information n'a pas été recueillie convenablement en 1852.

Il nous faudra analyser attentivement la signification de quatre variables, en relation les unes avec les autres : Résidence si elle est hors des limites, Membres de la famille, Non-membres de la famille, Membres de la famille absents. Bruce Curtis a attiré l'attention sur le manque de cohérence des critères servant à juger de l'appartenance à un ménage dans le recensement de 1852 (Curtis, 2001 : 115). Les recensements canadiens de la fin du 19^e siècle sont dits *de jure*, c'est-à-dire que les personnes devaient être inscrites à leur lieu de résidence habituel, qu'elles s'y trouvent ou non lors du passage de l'énumérateur. Les fonctionnaires responsables des recensements estimaient que dans un vaste pays où une partie de la population allait travailler en forêt ou en mer pendant une partie de l'année, on avait ainsi plus de chances d'enregistrer tout le monde. Cependant, en 1852, l'approche *de facto* est additionnée à l'approche *de jure* : « Vous écrirez dans la première colonne le nom de chaque personne qui a séjourné dans la maison pendant la nuit du dimanche 11 janvier, tant les étrangers que les membres de la famille,

ainsi que les membres de la famille qui sont absents temporairement mais qui résident d'habitude en ce lieu », lit-on dans les instructions (Gagan, 1974). On voulait en somme que les gens soient recensés à leur domicile, mais aussi que les étrangers soient mentionnés. Ces derniers devaient également donner leur lieu de résidence habituel : « 5e col. — Quand il se trouve qu'une personne a reçu le gîte dans une maison pendant la nuit du 11 janvier, vous aurez à noter si possible le lieu où elle réside d'habitude : dans beaucoup de cas, cela sera impossible et il vous faudra écrire le mot "inconnu" dans la colonne ». Curtis est d'avis que ces directives entraînaient le risque de recenser deux fois certaines personnes : les personnes en visite chez des amis pendant la journée et la nuit du recensement peuvent avoir été déclarées aussi bien chez elles que chez leurs hôtes (Curtis, 2001 : 116, 126-127). Pour résoudre définitivement ce problème, il faudrait pouvoir analyser une base de données contenant tout le recensement de 1852. L'un de nos objectifs serait d'ailleurs de porter de 20 à 100 pour cent notre couverture des données de ce recensement, en bénéficiant de collaborations bénévoles comme celles qui ont permis la transcription du recensement de 1901⁶. Moins de trois ans après que les Archives nationales du Canada eurent affiché sur Internet les reproductions des documents de ce recensement, des bénévoles travaillant en ligne avec l'Automated Genealogy.com 1901 Canadian Census Indexing Project avaient transcrit 99,99 pour cent du contenu du recensement, soit 5 642 088 lignes réparties sur 112 797 pages. La publication en ligne des images numérisées des documents du recensement de 1852 par les Archives nationales pourrait susciter des initiatives similaires. Entre-temps, nous nous intéresserons aux caractéristiques des individus inscrits comme étrangers et à leurs lieux de résidence habituels; des études de cas pourraient nous aider à mesurer l'étendue réelle du double dénombrement.

Parmi les surprises auxquelles nous avons fait allusion, on peut signaler la présence, dans la colonne des relevés où sont censés figurer les métiers, de nombreuses déclarations sans rapport avec cette rubrique. Certaines concernent les liens familiaux et leur analyse pourrait se révéler utile. Fait plus intéressant encore, environ un tiers des Canadiens échantillonnés ont répondu à la question sur le lieu de naissance en précisant non seulement leur province natale, mais encore leur ville, village ou localité d'origine. Si ces personnes ressemblent à d'autres égards aux personnes qui se sont bornées à déclarer leur province natale, il pourrait valoir la peine d'entreprendre une ana-

lyse plus poussée de leurs parcours migratoires. À un autre point de vue, lors du recensement de 1852, on a demandé aux Canadiens de donner le nom de leurs proches morts au cours de l'année écoulée. Environ 1 pour cent des Canadiens échantillonnés ont été déclarés décédés durant l'année du recensement. Il ne fait pas de doute qu'il y a eu sous-déclaration des décès en 1852. McInnis évalue à 20 pour mille le taux de mortalité au Canada pendant la première moitié du 19^e siècle. « Les recensements canadiens avaient recueilli des données sur la mortalité à partir de 1851, écrit-il, mais elles étaient reconnues comme incomplètes et insuffisantes »; plus loin il ajoute que « l'un des problèmes les plus sérieux résidait dans la sous-déclaration des décès de gens âgés n'ayant pas laissé de proche susceptible de rapporter leur disparition. La mortalité infantile était enregistrée de façon plus complète et plus précise » (McInnis, 2000 : 379, 404). On pourrait vérifier les conclusions de McInnis en analysant les caractéristiques des 2270 Canadiens dont l'âge et la cause du décès nous sont connus par l'échantillon du recensement de 1852.

Celui-ci ouvre des possibilités tout aussi riches de recherche sur la vie économique et le développement de la petite industrie dans les campagnes de l'époque. Quelles étaient les caractéristiques des Canadiens selon les maisons où ils vivaient (cabanes, maisons en rondins, à charpente de bois, en pierre, à étages) ? D'autre part, 1755 personnes ont déclaré avoir une place d'affaires attenante à leur maison, le plus souvent taverne, magasin, boutique de forge, commerce de semences, boutique de menuiserie, cordonnerie, boutique de tailleur, tannerie, auberge, sellerie, atelier pour chariots et atelier pour voitures. Un certain nombre (793) avaient une fabrique attenante, et 454 ont déclaré un nombre de salariés.

Une comparaison des caractéristiques de la population recensée dans le Canada-Ouest et le Canada-Est en 1852 et des populations rurales recensées au Québec et en Ontario en 1871 tend à montrer que dans l'ensemble l'échantillon de 1852 est représentatif de la population rurale du Canada. Le rapport hommes-femmes est légèrement plus élevé en 1852 (tableau 1), mais cela peut être dû à la vague d'immigration qui a précédé le recensement de cette année-là. Selon McInnis, l'immigration a été forte durant les années 1815-1861, alimentée par des citoyens britanniques qui fuyaient le chômage et la famine dont souffrait leur mère patrie (McInnis, 2000 : 378-384). Un afflux d'immigrés est susceptible de faire grimper le taux de mascu-

linité. Mais l'immigration a nettement baissé de 1852 à 1871, et on pouvait s'attendre à ce que le rapport de masculinité se soit rapproché de l'équilibre. Les deux années censitaires se ressemblent davantage pour les proportions de célibataires et de personnes mariées : près d'un tiers des Canadiens étaient mariés en 1852 et en 1871, les deux tiers étaient célibataires. La population était un peu plus jeune en 1852 (55 pour cent des gens n'avaient pas plus de 19 ans) qu'en 1871 (54 pour cent). La jeunesse de la population de 1852 pourrait s'expliquer par la proportion plus élevée de jeunes immigrés en son sein et par une fécondité supérieure. Il faudra procéder à de nouvelles vérifications pour savoir jusqu'à quel point l'échantillon du recensement de 1852 est faussé par l'absence de certaines zones urbaines. Les résultats déjà obtenus laissent toutefois supposer qu'il représente convenablement la population rurale du Canada-Est et du Canada-Ouest en 1852.

CONCLUSION

Cette première analyse des déclarations recueillies dans le Canada-Est et le Canada-Ouest à l'occasion du recensement de 1852 montre que l'échantillon de ce recensement est une précieuse source d'informations sur la vie familiale et économique des populations rurales du Canada au milieu du 19^e siècle. L'échantillon n'est pas complet; en particulier, une partie des populations urbaines n'y est pas représentée, car les relevés originaux la concernant ont été perdus, ce qui entraîne un déséquilibre en faveur de la population des campagnes. Le suréchantillonnage de zones urbaines sur lesquelles les données subsistent pourrait compenser en partie cette lacune. Il faudra aussi vérifier s'il y a bien eu double dénombrement en 1852 à cause d'une directive donnée aux énumérateurs où se mélangaient l'approche *de jure* et l'approche *de facto*. Quoiqu'il en soit, ainsi que l'a montré le Projet de recherche sur les familles canadiennes, pour travailler sur une base de données anciennes il faut saisir avec clarté les sources auxquelles elle remonte. « Il n'est désormais plus possible [...] de faire comme si les sources anciennes et l'information qu'elles livrent étaient une fenêtre transparente sur la réalité sociale du passé. Le recensement lui-même est objet de problématique, et il faut en exposer les tenants et aboutissants » (Sager, 2000 : 180). Cet échantillon de microdonnées du recensement canadien de 1852, ainsi que les bases de données censitaires sur la population du Canada qui

lui font suite, sont une mine d'information pour les chercheurs qui s'intéressent à l'évolution sociale, économique et démographique du Canada, du milieu du 19^e siècle à nos jours.

NOTES

- ¹ Fondation canadienne pour l'innovation, « Historical Demography Research Infrastructure/Infrastructure de recherches en démographie historique (HDRI/IRDH) » (Projet no 7549).
- ² Les responsables de ces travaux, et le titre des divers volets, sont : Gordon Darroch et Michael Ornstein, *1871 Canadian Census Data* (Canadian Historical Mobility Project), York University, Institute of Social Research, 1979; Lisa Dillon, *The 1881 Canadian Public-Use Microdata Sample* (1881 Canadian Census Project et North Atlantic Population Project), Université de Montréal, 2004; Kris Inwood et al., *The 1891 Canadian Census*, Université de Guelph; Eric Sager et Peter Baskerville, *The 1901 Canadian Census Public File* (The Canadian Families Project), Université de Victoria, 2001; Chad Gaffield et al., *The Canadian Century Research Infrastructure*; Statistique Canada, *Initiative de démocratisation des données*.
- ³ Voir <http://institutdrouin.com>.
- ⁴ Cet ensemble comprend les recensements réalisés en 1852 en Nouvelle-Écosse et au Nouveau-Brunswick, outre les recensements du Canada-Ouest et du Canada-Est.
- ⁵ Au tableau 1, la définition de l'adjectif rural n'est pas tout à fait la même pour les données de 1852 et pour celles de 1871. Les données de 1852 comprennent toutes les personnes qui ont fait une déclaration apparaissant dans un formulaire de recensement rural. Les données de 1871 sur les campagnes excluent les personnes qui vivaient dans des agglomérations urbaines de 3000 habitants ou plus.
- ⁶ The 1901 Canadian Census Indexing Project (Automated Genealogy) : <http://automatedgenealogy.com/census/cache/NationalSummary.jsp> (contenu mis en ligne en mai 2002).

RÉFÉRENCES BIBLIOGRAPHIQUES

- BOUCHARD, G. 1992. « Current issues and new prospects for computerized record linkage in the Province of Québec », *Historical Methods*, 25, 2: 67-73.
- CONRAD, M., A. FINKEL et C. JAENEN, 1993. *History of the Canadian Peoples: Beginnings to 1867*. Toronto, Copp Clark Pitman, 425 p.
- CURTIS, B. 2001. *The Politics of Population: State Formation, Statistics and the Census of Canada, 1840-1875*. Toronto, University of Toronto Press, 385 p.
- DARROCH, G., et M. D. ORNSTEIN. 1994. *Coding and Data Processing for the Feasibility Study: Canadian Historical Mobility Project*. Texte écrit pour l'Institute for Behavioural Research, York University, Canada (version révisée).

- DILLON, Lisa Y. 2000. « Integrating Canadian and U.S. historical census microdata: Canada (1871 and 1901) and the United States (1870 and 1900) », *Historical Methods*, 33, 4 (automne) : 189-192.
- GAGAN, D. P. 1974. « Enumerator's instructions for the census of Canada, 1852 and 1861 », *Histoire sociale/Social History*, 7, 14 : 355-365.
- MARCOUX, R., M. ST-HILAIRE et C. FLEURY. 2003. « Ville et population en changement : transformations urbaines et ajustements familiaux à Québec au XIX^e siècle et au début du XX^e », *L'Ancêtre*, 29 : 227-230.
- MCINNIS, R. M. 2000. « The population of Canada in the nineteenth century », dans M. HAINES et R. STECKEL, éd. *Population History of North America*. Cambridge, Cambridge University Press : 371-432.
- ORNSTEIN, M. 2000. « Analysis of household samples: The 1901 census of Canada », *Historical Methods*, 33, 4 : 195-198.
- ROBERTS, E., S. RUGGLES, L. DILLON, O. GARDARSDOTTIR, J. OLDERVOLL, G. THORVALDSEN et M. WOOLLARD. 2003. « The North Atlantic Population Project: An overview », *Historical Methods*, 36, 2 : 80-88.
- RUGGLES, S. 1995. « Sample designs and sampling errors », *Historical Methods*, 28, 1 : 40-46.
- SAGER, Eric. 2000. « The Canadian Families Project and the 1901 Census », *Historical Methods*, 33, 4 (automne) : 179-184.

ABSTRACT

Lisa DILLON

THE HISTORICAL DEMOGRAPHY RESEARCH INFRASTRUCTURE: CREATING THE 1852 CANADIAN CENSUS DATABASE

This research note discusses one of the latest additions to Canada's series of historical census microdata, a machine-readable sample of the 1852 census of Canada East and Canada West. While similar to subsequent censuses in form and content, the 1852 Canadian census poses particular challenges in terms of national representativity and the identification of heads of households. The 20% sample of the 1852 census of Canada East and Canada West will feature a total of 259,000 persons. Data entry procedures, the problem of missing data for urban areas and the identification of household heads using building-type information are discussed. A preliminary analysis of the 1852 census frequencies and comparison of these frequencies with those derived from the 1871 Canadian census of rural dwellers suggests that this machine-readable sample is representative of the rural Canadian population in 1852.