

# EXTRAITS DU RAPPORT FINAL DES TRAVAUX EFFECTUÉS PAR CAROLINE SAURIOL DE JANVIER 2004 À DÉCEMBRE 2006

POUR LE TRIO DE RECHERCHE  
DANIELLE GAUVREAU (CONCORDIA),  
SHERRY OLSON (MCGILL)  
& PATRICIA THORNTON (CONCORDIA)

## Table des matières complète

FÉVRIER 2007 À MONTRÉAL

<b>1 ACCÈS AUX FICHIERS</b> .....	<b>3</b>
1.1 SOMMAIRE DES TRAVAUX SUR ET À PARTIR DES BD .....	4
1.2 GÉNÉRALITÉS.....	4
1.2.1 <i>Nomenclature</i> .....	4
1.2.2 <i>Regroupements de variables</i> .....	4
1.3 ÉCHANTILLON DE 1881.....	5
1.3.1 <i>Recensement 1881 (1881 combiné modifié)</i> .....	5
1.4 ÉCHANTILLONS DE 1901 .....	5
1.4.1 <i>Échantillon CFP 1901 (CFP modifié 1901)</i> .....	5
1.4.2 <i>Échantillon de Danielle (montreal, men &amp; mtl2)</i> .....	5
1.4.3 <i>Échantillon structures familiales 1901 (1901 family structure)</i> ....	5
1.5 ÉCHANTILLONS PRODUITS DU JUMELAGE.....	6
1.5.1 <i>Échantillon d'occupations 1901 (1901 ech occ)</i> .....	6
1.5.2 <i>Échantillon de fils 1901 pairés avec leur père en 1881 (1901-1881 link)</i> .....	6
1.5.3 <i>Registres de mariages 1881-1901 (MKEEP modifié)</i> .....	6
<b>2 RECENSEMENT DE 1881 – DÉTAILS DES TRAVAUX</b> .....	<b>7</b>
2.1 PROCESSUS DE CONSOLIDATION.....	7
2.1.1 <i>Consistency checks</i> .....	8
2.1.2 <i>Nettoyage et création de variables</i> .....	9
2.2 VARIABLES DU RECENSEMENT.....	10
2.2.1 <i>Cas invalides (INVALID)</i> .....	10
2.2.2 <i>Commentaires du manuscrit (COMMENT)</i> .....	10
2.3 VARIABLES GÉOGRAPHIQUES .....	11
2.3.1 <i>Wards: District, subdistrict and subdistrict name</i> .....	11
2.3.2 <i>Division, neighbourhood &amp; geocode</i> .....	11
2.3.3 <i>Cadastral lot &amp; localisation (TRES &amp; TPROPID)</i> .....	12
2.4 VARIABLES DU MÉNAGE .....	14
2.4.1 <i>Identification des ménages (SERIAL)</i> .....	14
2.4.2 <i>Variables quantitatives des ménages et des familles</i> .....	14
2.4.3 <i>Types de ménages non familiaux (HHTYPE)</i> .....	14
2.4.4 <i>Types de ménages familiaux (HHTYPE)</i> .....	17
2.5 VARIABLES INDIVIDUELLES .....	18
2.5.1 <i>Numéro UNIQUE d'identification individuel (ID)</i> .....	18
2.5.2 <i>Révision des prénoms et noms de familles</i> .....	18
2.5.3 <i>Identification des noms de jeune fille (NAMMAID &amp; HHMAIDEN)</i> .....	19

2.5.4 Genre, âge, et statut matrimonial (SEXE, AGE, MARST) .....	20
2.5.5 Infirmités (INFIRM) .....	21
2.5.6 Membres d'un ordre religieux (RLGORDR) .....	22
2.5.7 Employés résidents & Ménages avec employés (DOMRES & HHDOMRES) .....	22
2.5.8 Rôles individuels (INDSIT) .....	22
2.6 VARIABLES SUR LES COMMUNAUTÉS CULTURELLES .....	23
2.6.1 Lieu de naissance des individus (BPLC) .....	23
2.6.2 Origine (ORIGIN) .....	23
2.6.3 Religion (RLGN) .....	24
2.6.4 Communauté culturelle (ETH & ETHGRP) .....	24
2.6.5 Types de mixité ethnique (HHETH, HHETHY, HHETHDOM) .....	25
2.7 VARIABLES OCCUPATIONNELLES .....	28
2.7.1 Occupations textuelles (OCCSTRNG, OCCNOTE & OCC_CL) .....	28
2.7.2 Occupation HISCO .....	29
2.7.3 Occupation (3-lettres) (OCC) .....	31
2.7.4 Fréquentation scolaire et grandes classes d'occupation (SCHOOL.05 & SCHOOL.5 & HHSCHOOL) .....	32
<b>3 RECENSEMENTS DE 1901 &amp; JUMELAGE AVEC 1881 – DÉTAIL DES TRAVAUX .....</b>	<b>33</b>
3.1 ÉCHANTILLON DANIELLE .....	33
3.2 ÉCHANTILLON CFP .....	34
3.3 JUMELAGE POUR LA MOBILITÉ SOCIALE .....	34
<b>4 ANALYSES DIVERSES .....</b>	<b>36</b>
4.1 1881 POPULATION INSTITUTIONNELLE .....	36
4.2 1881 LES JEUNES AU 19ÈME SIÈCLE (KATHY PROVOST) .....	36
4.3 1901 EXEMPLES (BERLIN 2005) .....	36
4.4 1881-1901 ACFAS 2005 .....	36
4.5 1881-1901 MOBILITÉ SOCIALE (LYON 2006) .....	36
4.6 1881-1901 YOUTH (AMSTERDAM 2006) .....	36
4.7 FEMMES ET DÉMOGRAPHIE (AUTOMNE 2005) .....	37
4.8 ENTRETIENS JACQUES CARTIER (DÉCEMBRE 2005) .....	37
4.9 MISE À JOUR ET STANDARDISATION DES BD DANS REFERENCE MANAGER .....	37
<b>5 ANNEXES .....</b>	<b>38</b>
5.1 LISTE DES VARIABLES 1881 .....	38
5.2 LISTE DES VARIABLES DE 1901 (CFP SAMPLE) .....	41
5.3 LISTE DES MÉNAGES SÉPARÉS ET RENUMÉROTÉS .....	46
5.4 MÉNAGES POTENTIELLEMENT DÉDOUBLÉS (IDENTIFIÉS PAR SHERRY – OCTOBRE 2006) .....	47
5.5 INDIVIDUS (SERVANTES) POTENTIELLEMENT DÉDOUBLÉS (IDENTIFIÉS PAR SHERRY – OCTOBRE 2006) .....	48

[...]

## 1.1 SOMMAIRE DES TRAVAUX SUR ET À PARTIR DES BD

### 1.2 GÉNÉRALITÉS

#### 1.2.1 Nomenclature

La première opération effectuée sur chaque nouveau fichier a été d'indiquer des noms longs aux variables, et d'encoder numériquement les valeurs qui pouvaient l'être. Ceci allait faciliter les analyses subséquentes en standardisant les valeurs des données et en clarifiant leur contenu. Les titres de variables ne sont donc pas nécessairement originaires de l'auteur de chaque

variable.

Lorsque des variables de différentes sources étaient agrégées dans un même fichier, un suffixe a été ajouté aux noms des variables tel que suit :

xxx.0 Variables provenant du site internet de NAPP. Ces variables n'ont pas été modifiées.

xxx.05 Variables qui ont été relevées par cette équipe à partir des microfilms.

xxx.1 Variables transmises par le Canadian Families Project (CFP) à cette équipe

xxx.2 Variables créées ou modifiées par Pat Thornton, à partir de la version CFP 2001.

xxx.3 Variables créées ou modifiées par Sherry Olson, à partir de la version CFP 2001.

xxx.5 Variables créées par Caroline Sauriol, soit pour combiner les modifications faites par Pat & Sherry, soit pour effectuer des corrections, soit pour créer de nouvelles variables facilitant l'analyse.

### **1.2.2 Regroupements de variables**

Les variables de la plupart des fichiers de données ont été regroupés selon qu'elles décrivaient le ménage, la position géographique, ou encore l'individu (démographie, ethnicité, occupation ou relation avec les autres membres). Pour faciliter le repérage visuel de ces données, ces sous-groupes sont souvent précédés d'une variable vide dont le nom est en majuscules et toutes les valeurs sont absentes. Les sections du présent document sont organisées autour de ces thèmes.

## **1.3 ÉCHANTILLON DE 1881**

### **1.3.1 Recensement 1881 (1881 combiné modifié)<sup>2</sup>**

*2 Le fichier 1881 Census Common \_ 3 est obsolète, mais je ne suis pas arrivée à l'effacer. Cette version ne doit pas être utilisée, même si elle apparaît dans le répertoire.*

*3 J'ai malheureusement perdu la variable qui fait le lien entre les deux, je ne sais plus comment le serial a été créé, et ne peux donc pas réintégrer les données.*

À partir des données extraites du site de NAPP, les opérations suivantes ont été faites :

1. Agréger les fichiers de NAPP, de Sherry et de Pat pour inclure toutes les variables ayant été créées lors d'analyses antérieures (momloc, geocode, etc.)
2. Analyser les discordances entre les versions et déterminer les valeurs à retenir
3. Standardiser certaines variables d'analyse (ethnicité, occupation, etc.)
4. Valider les données étranges ou incomplètes, en retournant au microfilm (sexe, marst, age, etc.)
5. Ajouter certaines variables non transcrites par les Mormons (scolarité, infirmité, commentaires, etc.)
6. Faire les liens avec les registres de décès et de baptêmes pour en tirer les noms de jeune fille. (voir 2.5.3 Identification des noms de jeune fille (NAMMAID & HHMAIDEN))
7. Créer de nouvelles variables pour analyses futures (type de ménage, ethnicité mixte, etc.). Plus de détails sur ces travaux se trouvent dans les prochaines sections.

[...]

## **2 RECENSEMENT DE 1881 – DÉTAILS DES TRAVAUX**

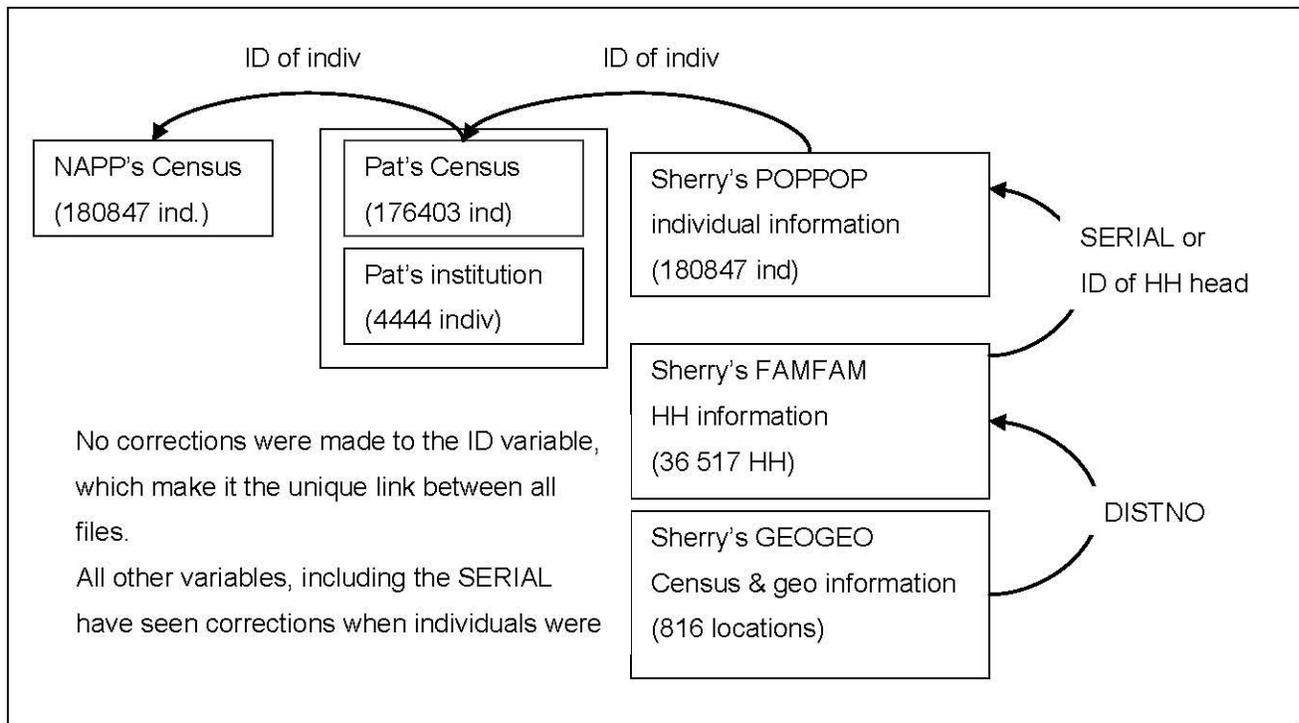
This section relates the work done on the 1881 census data. This first sub section describes the consolidation process of the various versions, while the other sub sections relate to the detailed work on existing and/or newly created variables. Because there has been, in the course of the two years, much detailed work performed on almost all variables, the text is organised around groups of variables and reviews each and every one to present their meaning, corrections, and creation

process when they were not part of initial DB.

## 2.1 PROCESSUS DE CONSOLIDATION<sup>4</sup>

When this project started, there were already a few census files being used within the team. A first version was that of Pat Thornton, whom had started with a file from the CFP team and had enhanced and made corrections to it. It must be noted that the CFP file includes information not available from NAPP<sup>5</sup> such as actual string values of variables that were later encoded (such as origin, religion and the like)

This augmented file was then transferred to Sherry, who made further enhancements and corrections to it. Using different BD software, Sherry split the file into a three: one for individual information, one for HH information and one for geographical information. Pat, for her part, removed HHs that appeared to be institutions (hh size > 29 individuals), which meant 4 444 individuals were set aside in a different file.



The first step of the consolidation process was thus to recombine Pat's two files and Sherry's three files, and have two files that could be compared one to the other, and then to compared to the NAPP file. Because NAPP team had done further work on the census file since the work originally started in 2001, a new version of the NAPP file<sup>5</sup> was downloaded from their website and was used as the baseline. It must be noted that NAPP team had since then recalculated the SERIAL numbers in order to standardise it throughout countries, which caused minor problems right from the start, but was overcome.

<sup>5</sup> The NAPP version and various methodological and contextual information can be found on their web site at : <http://www.nappdata.org>

All variables were combined into one file, using the suffix as a marker of its originator. Versions of similar variables were compared when identical, the NAPP version remained and others were removed from the file.

When variables were similar but with minor differences, each correction was analysed to assess

which version of data should be kept. Microfilm was often used to validate corrections. When one version of the variable seemed to encompass all other corrections, this version was kept and the others removed. When all variables had been corrected and none could be thought generally better than the other ones, a new variable was created (with a .5 suffix) that encompassed all retained corrections.

When variables reflected a similar phenomenon, but its structure was different (such as geographical location), Pat & Sherry were consulted to choose which version should remain for future analysis.

The next step was to standardise the format of the variables that remained in the file. For example, some string information could be encoded as ordinal variables (such as subdistrict name), value labels could be entered for ordinal variables, some numeric variables could be cleaned (for example ages with a decimal), and missing values could be formally declared instead of remaining empty. This process of standardisation of variables would facilitate further analysis.

The resulting file encompassed all variables that NAPP, Pat & Sherry had developed, as well as all corrections that had been further made by each of these researchers. The last step of the consolidation would be to validate if these variables were coherent with one another, as described in the next section.

### **2.1.1 Consistency checks**

The variables from the census have been created based on the input provided by enumerators and the best read and understanding that data entry volunteers had of the manuscripts. However, it happens that some variables for a given household or individual are not coherent with other variables regarding these same household or individuals. This section lists the consistency checks that have been done between variables in order to have a coherent view of households and individuals. Such checks were performed between original variables, in order to provide a solid base for variables construction and analysis. Modifications have been made in recoded variables (.2, .3 or .5), while all original NAPP variables (.0) have been left intact.

Variables constructed from the original variables have also been modified, but these checks are considered part of the creation process and are not part of these initial consistency checks. The validation checks are detailed in the section describing the creation of these variables.

#### *2.1.1.1 Census variables: No checks*

#### *2.1.1.2 Geography variables*

- a) Subdistrict name and letters

#### *2.1.1.3 Household variables*

- a) Serial number and household number breaks
- b) Household size and index number within household
- c) Number of families must be equal or less than number of individuals in household (not done)
- d) Number and size of families and household size (not done)
- e) Family size and index number within family (not done)

#### *2.1.1.4 Demographic variables*

- a) Sex with first names (identified, but not validated. 700 cases to check)
- b) Age of „under 1 year’ with month of birth (agemonth and/or dbirthmonth and/or birthmonth)

c) Marital status with age

#### 2.1.1.5 Cultural variables

- a) Dual origins
- b) Dual religions

#### 2.1.1.6 Occupational variables

- a) Age and occupation

### 2.1.2 Nettoyage et création de variables

Comme les sections suivantes le décriront, plusieurs variables ont fait l'objet de minutieuses vérifications directement sur les microfilms<sup>6</sup>. Les variables « .5 » contiennent ces modifications de même que les modifications apportées par les chercheuses lors des différents processus de jumelage ou d'analyses. Ces variables sont donc nettoyées et fiables, à moins d'avis contraire dans les lignes qui suivent.

*6 La plupart des documents utilisés pour capter les informations lors de visites à la bibliothèque ont été conservés. Ils pourront servir de référence futures quant aux modifications effectuées aux variables. La plupart sont regroupés par thèmes, selon les variables qui faisaient l'objet de vérification.*

## 2.2 VARIABLES DU RECENSEMENT

La plupart de ces variables (version, country, reel, volume, page number, line number) servent de référence aux manuscrit et microfilm et n'ont été ni corrigées, ni modifiées.

### 2.2.1 Cas invalides (INVALID)

Some of the cases included in the database should be considered invalid because individuals registered should not have been registered at the time of the census, because they had moved out of the city, were dead or not born on the date of April 1st. In contextual analysis, these „marginal” individuals can remain because their situation is relevant to analysis of the general environment. It also happened that birth, death and immigration information were registered in various fields. There are also cases when individuals or Households were enumerated twice. These families may have moved at the time of census or two enumerators may have confused the boundaries of their section to register. When two households were thought to be identical, one of the family was identified as invalid, based on which family's information seemed more accurate. It also happened that individuals were registered but were then after crossed out. There is sometimes an indication of the reason of the crossing, which can help us determine if the case should be kept or not. When there was no mention of a cause, we assumed that it was a mistake and have removed the individual from the census. When it was mentioned that these individuals were dead or had born after the date, they were kept because the purpose of our project is about population characteristics and not absolute count.

We wanted to retain as much information as possible on each household and thus we have not removed these cases from the database. For precise analysis of population, they should be removed because they should not have been enumerated, „as of the rule”. The variable „invalid cases” and/or the comment field are indication of such situations, and it is left to the user to determine whether they should be kept or removed.

### 2.2.2 Commentaires du manuscrit (COMMENT)

Les commentaires inscrits dans la colonne „Commentaires” du manuscrit n'avaient pas été

relevés par les personnes ayant fait l'entrée de données. Cette équipe a donc consulté l'ensemble des microfilms, de manière à relever cette information additionnelle sur les individus et les ménages, colligée dans la variable COMMENT.5

□ Certains commentaires généraux qui avaient été entrés dans d'autres variables, telles l'occupation et le nom, ont été transférés vers cette variable-ci afin de regrouper tous les commentaires en une seule variable. Dans ces cas, la provenance a été inscrite entre parenthèses à la suite du commentaire (ex. « blind (occupation) »).

□ Le „ x' précédant un commentaire indique qu'il avait été rayé sur le manuscrit. Parce que le rayage ne semble pas être dû à des erreurs, mais bien à un jugement de leur pertinence postérieur à

l'inscription, ces commentaires ont tout de même été enregistrés dans la variable. Par exemple, pour certaines sections, tous les commentaires ont été rayés sans discrimination.

□ À la suite de vérifications ou d'analyses diverses, certains ménages ont été segmentés en deux groupes, ou certains enfants mal placés ont été assignés à des parents d'un autre ménage. Ces ménages ont été renumérotés en séquence, à partir de 99000. Pour ces nouveaux ménage, un commentaire dans la variable COMMENT indique, entre parenthèses, le numéro de série antérieur des ces ménages. Par ailleurs, les ménages qui ont „perdu' des membres ont aussi reçu un commentaire indiquant le nouveau numéro de série des membres séparés. Ces indications réciproques permettront à l'analyste de reconstituer les ménages et ainsi faire le lien avec d'autre BD. Dans tous les cas, un tri séquentiel par ID permettra de replacer les ménages dans leur séquence originale et officielle.

## 2.3 VARIABLES GÉOGRAPHIQUES

### 2.3.1 Wards: District, subdistrict and subdistrict name

The combination of the District (90 or 91) and Subdistrict (from a to s) correspond to the 29 different wards of Montreal at the time of census. The following table provides the list of ward numbers and names.

90 a Ste-Anne (Quartier/Ward)	91 a St-Jean-Baptiste (Village)	91 j Cote-St-Paul (Village)
b West Ward/Quartier Ouest	b Mile End (Village) & St-Louis	k NDG (Ouest/West)
c Centre (Quartier/Ward)	c Coteau-St-Louis (Village)	l Cote-St-Antoine (Village)
d East Ward/Quartier Est	d Cote-de-la-Visitation (Village)	m Cote-des-Neiges (Village)
e St-Louis (Quartier/Ward)	e Hochelaga (Village)	n Cote-St-Paul
f St-James (Quartier/Ward)	f St-Henri	o Longue-Pointe
g St-Mary's	g Ste-Cunegonde	p Pointe-aux-Trembles

	(Quartier/Ward)	(Village)	
h	St-Antoine (Quartier/Ward)	h St-Gabriel (Village)	q Riviere-des-Prairies
i	St-Lawrence (Quartier/Ward)	i Verdun (Village)	r Sault-aux-Recollets
			s Outremont (Village)

**À noter** The original NAPP file contained two ‘St-Mary’s (Quartier/ward)’, with a slight spelling difference. To facilitate any further utilisation of this variable, it has been recoded into the same variable (SDSTNAMCA.0).

### **2.3.2 Division, neighbourhood & geocode**

Wards can be segmented into different neighbourhoods, which can be done by using the division number. The variable GEOCODE was constructed as concatenation of three geographical variables (DISTRCT, SDSTRCT and DIVNOCA), and results into 99 neighbourhoods.

The GEOCODE is very similar to the variable NEIGH, which goes a step further and refers to the names of the different neighbourhoods. Pat and Sherry having given different names to these neighbourhoods, Sherry’s coding convention<sup>7</sup> was kept and value labels were defined to keep track of Pat’s naming system. A three-

*7 It was coded as follows: NEIGH = SDISTLET + DIVNO (adding ‘0’ to ensure that length is same for all values).*

capital-letters name has been assigned when neighbourhoods (mostly in district 91) were considered as suburb and not in the city of Montreal. Neighbourhoods can be located on the map of the next page.

There are thus three naming schemes (GEOCODE, NEIGH value and NEIGH label) for the identification of specific geographical areas.

### **2.3.3 Cadastral lot & localisation (TRES & TPROPID)**

These two variables originate from Sherry’s location in tax roll and provide the detailed location of household on cadastral map.

TPROPID is the property identification and thus identifies the lot on map (n=10.000).

TRES is the key to household id point on map.

## **2.4 VARIABLES DU MÉNAGE**

### **2.4.1 Identification des ménages (SERIAL)**

Serial were originally sequentially attributed by the NAPP team. They were assigned based on the “family” column (#6) of the census track and reflect a common HH, which may actually comprise more than one family.

□ Analysis showed that some numbers had been wrongly entered in the database, or some HH steps did not match the family numbers from the census. It also happened that the family column had not been entered at all, and that a single serial comprised more than one HH

□ In cases where whole HH were affected, a new serial was created (ranging above 90000) and a note was registered in the COMMENT to keep track of the prior SERIAL assignment of this HH. When single or a few individuals were affected, their individual serial was corrected. These HH are listed in appendix.

### **2.4.2 Variables quantitatives des ménages et des familles**

A household is defined as all those living in a same house or apartment. They are identified through the family steps of the census manuscript.

A family is defined as those living together and bearing the same family name. Additional information, such as relational information or maiden name, may indicate kinship between individuals with different last names. In such cases, these related members were included in the family. (not systematic)

These variables reflect the quantitative composition of the HH and the family(ies) within.

HHSIZE gives the number of individuals in the HH

NFAM gives the number families in the household

FPERNUM provides the rank of each family within a HH

FAMID is a unique identifier of families, created by concatenating SERIAL and FPERNUM.

FAMSIZE gives the number of individuals in each the family

### **2.4.3 Types de ménages non familiaux (HHTYPE)**

This variable was elaborated and created by this team to qualify the nature of relations in each HH.

#### **2.4.3.1 Typology**

10 Family pension (1-2 non-rel)

15 Boarding house (3 + non rel)

20 Hotel

30 Hospital

40 Indust. house, etc.

60 Orphanage

70 School

75 Seminary

80 Monastery / Nunery

85 Presbytere

90 Prison

99 Undetermined instit

The variable first separates between institutions (<100) and familial HH.

□ Institutions were classified into different types: Hospital, Industry house, Asylum, Hospice, Refuge, Seminary, School, Orphanage, Monastery / Nunery, Presbytery, Prison, and when such classification was not possible, were put under: Undetermined institution.

□ When enough information indicated a Hospitality type of HH, these were also segregated into Family pension (1-2 non-related individuals), Boarding house (3 + non related individuals) and Hotel. Attention must be put when analysing this subgroup because their identification relied on occupation of tenant, and may not include all such hospitality HH. A great number of familial HH may also correspond to this profile, and should be considered if and when Hospitality HH are analysed.

#### **2.4.3.2 Pensions de famille et maisons de pension (à partir des occupations)**

Les ménages qui sont des maisons de pension ou des pensions de famille ont été identifiés comme suit :

1. Les personnes dont l' occupation est de tenir un maison de pension (OCC.5=BHK) ont été identifiés, et leur ménage a été extrait pour analyse.

2. Les ménages qui étaient constitués par une seule famille (un seul nom de famille) ont

été retirés de la liste des maisons des pensions, de même que les ménages constitués d' une seule famille avec un ou des serviteurs. Ces ménages ont été classés comme ménage simple, en attendant une structuration plus étoffée des ménages non institutionnels.

3. Les ménages constitués d' une famille principale (dont un tenancier) et de un ou deux autres individus apparemment non liés et non domestique ont été classés comme des pensions de famille.

4. Les ménages constitués d' une famille principale (dont un tenancier) et de trois ou plus individus apparemment non liés et non domestiques, ont été classés comme maison de pension.

**Note :** *On doit faire attention lorsque ces groupes sont utilisés pour analyse puisqu'ils pourraient se recouper avec certains ménages à familles multiples dont aucun membre n'aurait une occupation de Boarding House keeper ou Innkeeper. On doit donc porter une attention particulière si on souhaite analyser ces ménages.*

#### 2.4.3.3 Hôtels

Les ménages qui sont des hôtels ont été identifiés comme suit :

5. Les personnes dont l' occupation est de tenir un hôtel ou une auberge (OCC.5=HOT ou INN) ont été identifiés, et leur ménage a été extrait pour analyse.

6. Les ménages qui étaient constitués par une seule famille (un seul nom de famille) ont été retirés de la liste des hôtels, de même que les ménages constitués d' une seule famille avec un ou des serviteurs. (à l' exception du ménage 350034, qui avait 4 serviteurs et qui a été classé comme hôtel) Ces ménages ont été classés comme ménage simple, en attendant une structuration plus étoffée des ménages non institutionnels.

7. Les ménages constitués d' une famille principale (dont un hôtelier) et de un ou deux autres individus apparemment non liés et non domestique ont été classés comme des pensions de famille.

8. Les ménages constitués de deux familles dont l' une contient un hôtelier et l' autre a plus de deux individus apparentés ont été classés comme maison de pension.

9. Les ménages constitués d' une famille principale (dont un hôtelier) et de trois ou plus individus apparemment non liés et non domestiques, ont été classés comme hôtel.

#### 2.4.3.4 Institutions

La classification des individus en institutions s' est faite à partir d' identification antérieure faite par l' un ou l' autre des chercheuses. Des analyses faites auparavant avait amené une certaine mise à l' écart des ménages qui semblaient être des institutions, sans que cette procédure ne soit très méthodique ou raffinée. C' est toutefois à partir de ces identifications préalables que la classification a été réalisée.

Un ensemble de près de 8000 individus avait été identifié comme vivant potentiellement dans une forme ou une autre d' institution. Les commentaires provenant du recensement, le statut religieux d' un individu, ou encore l' identification du lot à partir de la liste des taxes foncières ont permis d' identifier quelques institutions additionnelles. À partir de cet ensemble, voici comment ont été répartis les ménages selon les types:

10 Family pension (1-2 non-rel)

15 Boarding house (3 + non rel)

20 Hotel

- 30 Hospital
- 40 Indust. house, etc.
- 60 Orphanage
- 70 School
- 75 Seminary
- 80 Monastery / Nunery
- 85 Presbytere
- 90 Prison
- 99 Undetermined instit

Analyse du # de religieux ou de personnes de même sexe ou d' un même groupe d' âge. En regardant la liste des numéros de série des ménages, on peut voir que certains ménages semblent former un ensemble institutionnel, comprenant plusieurs unités. Certaines unités peuvent séparer hommes et femmes, religieux et employés, enfants et adultes, etc., mais pour réaliser la classification, nous les avons d' abord regroupés séquentiellement. Certains ménages intercalés n' avaient pas été identifiés comme institutionnel (parfois en raison de leur petite taille, mais ont été intégrés lorsque leur vocation semblait liée à celle de l' institution l' entourant.

Toutes les analyses ont été faites en faisant abstraction des domestiques résidents<sup>8</sup>. En les retirant de la liste, cela permettait de mieux évaluer le profil démographique des institutions et de ne pas altérer des proportions entre laïques et religieux, entre sexes ou entre les groupes d' âges. Aussi, l' identification et l' analyse de la population religieuse<sup>9</sup>, en comparaison de la population laïque d'un établissement a facilité l'identification de la vocation du ménage

<sup>8</sup> Pour plus de détail sur cette identification, voir la section sur DOMRES.5.

<sup>9</sup> Pour plus de détail sur cette identification, voir la section sur CLERGY.5..

<sup>10</sup> Plus de détail peut être trouvé dans le fichier décrivant spécifiquement les travaux faits relativement aux structures familiales.

#### 2.4.3.5 Identification des institutions (INST\_NAME)

Lorsqu'il a été possible de les identifier, les noms d'institutions ont été enregistrés dans cette variable. Ces identifications peuvent provenir d'un commentaire du manuscrit (ex. asyle) ou encore d'un titre individuel (ex. sœur de Notre-Dame) ou encore d'une identification ultérieure basée sur la composition du ménage et leur localisation géographique. Ce travail d'identification géographique a été entamé par l'équipe de Sherry, mais certains résultats sont contradictoires et ils n'ont pas tous pu être inclus. La variable est actuellement incomplète et il reste du travail à faire mais elle est incluse pour ne pas perdre l'information amassée.

#### **2.4.4 Types de ménages familiaux (HHTYPE)<sup>10</sup>**

This typology centers on the simple family, but also recognizes that this family might have shared the residence with other members, related or not, as well as it included unmarried people living alone, or with other people, related or not related. At this stage, the residence is the unit of analysis, including all those members that might have been sharing it. The relationship between these families and these additional members can often be inferred based on last names, ages or occupations, but maternal kinship or distant relatives may not necessarily have shared the same last name. Keeping in mind that „unrelated' implies „apparently unrelated', the typology segregates families based on the demographic profiles of the HH and of all its members (number of people, gender, marital status, age and for some categories, the occupation was also used), reflecting the strategy enacted by the families within, at a particular moment in their life cycle. The typology can be summarised as follows:

- 110 Solit widow married

- 210 Co-res siblings
- 220 Co-res relatives
- 230 Co-res non relatives
- 240 Co-res rel & non-rel
- 310 Married couple
- 320 Married couple w children
- 321 Married couple w children, maternal
- 330 Rem parent w related children
- 331 Rem parent w rel children, maternal
- 400 Base fam ext laterally
- 401 Base fam ext laterally, maternal
- 500 Base fam ext vertically (& lateral)
- 501 Base fam ext vert (& lat) maternal
- 610 Base fam w unrel children
- 620 Base fam w 1-2 non rel to CU
- 621 Base fam w 1-2 non rel, ext. maternal
- 630 Base fam w 3+ non rel to CU
- 631 Base fam w 3+ non rel, ext. maternal
- 650 Mult base fam non-rel
- 651 Mult base fam non-rel ext. maternal
- 999 Undetermined HH

La présence ou non de domestiques dans une ménage n'ayant que peu d'effet sur la configuration même du ménage, cette information n'est pas requise dans la classification de base des ménages. Elle peut toutefois être utile pour identifier des personnes non apparentées dans les ménages, ou encore fournir de l'information sur le statut et le fonctionnement du ménage. Pour mieux comprendre ce processus, voir la section 2.5.7 Employés résidents& Ménages avec employés

## **2.5 VARIABLES INDIVIDUELLES**

### **2.5.1 Numéro *UNIQUE* d'identification individuel (ID)**

Le numéro est la seule variable qui n'a subi aucune correction et est resté identique de NAPP à aujourd'hui. C'est le lien entre tous les fichiers, toutes les versions, tous les extraits. C'est un numéro séquentiel qui identifie uniquement chaque individu, et qui n'a pas été modifié, même quand un individu était réassigné à un autre ménage.

### **2.5.2 Révision des prénoms et noms de familles**

Les individus inscrits au manuscrit ont donné leur nom de différentes façons, qui ont par la suite été interprétés et retranscrits de façons diverses aussi. La révision des noms a servi à valider les noms qui étaient illisibles ou incompréhensibles, de même qu'à standardiser l'organisation des composantes des différents types des noms.

Les prénoms et noms de famille ont été revus tel que suit :

1. Éliminer les titres divers compris dans les prénoms et noms de famille
2. Remplacer les inversions de prénoms et noms, et/ ou extraire du prénom des 2èmes noms de famille.
3. Corriger les accents et autres caractères étranges dans les prénoms et noms de famille
4. Interpréter les noms et prénoms dont une ou deux lettres seulement déparaient un nom familier (faute d'orthographe ou typographique)
5. Revoir tous les noms et prénoms qui n'apparaissaient seulement qu'une fois dans le recensement et valider s'il étaient censés ou malheureusement inscrit.

## 6. Retourner au manuscrit pour quelque 700 noms illisibles ou incohérents.

La variable NAMTTL.5 a été créée afin de contenir des informations liées aux noms, mais qui ne font pas partie du prénom ou du nom de famille. Entre autres, les prénoms de femme qui étaient en fait celui de leur mari précédé de madame ou mrs ont été déplacés dans cette variable, laissant leur prénom en blanc.

La variable NAMLST2.5 a été créée afin de contenir les deuxièmes noms de famille de certains individus et les alias propres à certains noms de famille. Le contenu de cette variable pourra provenir de plusieurs sources (autre que le recensement) mais sert à situer les individus dans son contexte familial plus large. Le nom qui est conservé dans la variable NAMLST.5 est celui de la famille immédiate, du ménage avec qui l'individu réside, alors que le deuxième nom de famille le met en relation avec d'autres individus.

Des caractères spéciaux ont été utilisés afin d'indiquer certaines situations particulières :

- o Tous les doubles noms pouvant être associés à un alias portent la mention „ (dit) „
- o Les noms qui diffèrent entre les corrections de Pat & Sherry ont un «\* » à la fin
- o Les noms dont l'orthographe semble étrange ont un «? » à la fin.

Les religieux ayant parfois des prénoms et noms composés ou non conformes à une structure de nom habituelle, avaient été encodés de diverses façons. Leur noms ont donc été standardisés comme suit :

- o Les mentions frère, sœur, révérend ont été transférées dans la variable NAMTTL.5.
- o Les prénoms et les noms de familles usuels sont encodés dans les variables de prénom et de nom de famille. Dans le cas où seulement l'une des deux composantes était inscrite, elle a été encodée dans la colonne pertinente, laissant l'autre en blanc.
  - ex. frere marceau est devenu NAMTTL.5= frere et NAMLST.5=marceau.
  - Dans les cas où le nom est un référent catholique composé (aimée du tres st sacrement), il a été préservé en un tout dans la variable du prénom NAMFRST.5.

Voici quelques exemples illustrant le résultat :

senior	George	legault	deslauriers (dit)
veuve jacques		trepanier	
soeur	marie de st nazaire		

### **2.5.3 Identification des noms de jeune fille (NAMMAID & HHMAIDEN)**

Les travaux de Pat sur la fertilité, la natalité et le mortalité infantile a permis d'élaborer une BD comprenant tous les registres de baptêmes de tous les enfants de moins de 2 ans, avec le nom de leur père et mère, de même qu'une BD sur les registres de décès de la période entourant l'année 1881. En plus de l'information sur la personne décédée, on y trouve aussi des informations sur l'époux et les parents. Ces deux BD ont été mises à profit afin de retracer les noms de jeune fille de certaines femmes mariées et ou d'enfants issus d'un premier lit, de même qu'il a permis de mettre à jour certaines relations familiales entre individus qui semblaient non apparentés.

#### **2.5.3.1 Jumelage**

Les travaux de Pat avaient déjà établi des liens entre les données des registres et du recensement. Ce lien est donné par les variables bapid et did des fichiers baby79 et 81deaths, et qui se retrouvent aussi dans les fichiers originaux de Pat.

Par ailleurs, tous les ménages qui avaient été identifiés « à famille multiple » lors de la classification par type de ménage, ont été spécifiquement recherchés dans les BD de baptêmes et de décès. Ces recherches ont été faites à partir du nom prénom et nom de famille de la mère, qui, combinés, devenait un lien entre les deux fichiers. L'appariement entre le recensement et les registres a pu se faire tant au niveau des individus visés par le registres, que par le biais des parents qui y été identifiés.

Les informations nominales qui ont pu être colligées ont été indiquées au niveau individuel dans la variable NAMMAID, et les ménages pour lesquels une telle information est disponible ont une valeur de 1 à la variable de HHMAIDEN. Les types de ménages ont de plus été modifiés conséquemment cette nouvelle information et ces ménages reclassés ont été mis dans une sous catégorie de la catégorie principale (ex. 401 au lieu de 400 pour extension latérale), avec la mention « maternal » dans la description. Ceci permet de différencier les types déterminés avec seulement les infos du recensement et ceux issu d'une information additionnelle.

#### **2.5.3.2 Résultats**

En tout, 6113 noms de jeunes filles provenant de 6076 ménages ont pu être identifiés par l'exercice de jumelage. Ces ménages comprennent 35829 individus ce qui augmente la portée de l'exercice en donnant plus d'information contextuelle pour un grand nombre d'individus. Notamment, ces informations additionnelles ont permis de mieux classier 457 ménages, qui avaient originellement été catégorisés non-familiaux (7,5% des ménages pour lesquels une info additionnelle a été trouvée). Ceci représente 6,5% des ménages jugés non familiaux au départ.

#### **2.5.4 Genre, âge, et statut matrimonial (SEXE, AGE, MARST)**

Parce que les vérifications de cohérences (voir section 2.1.1 Consistency checks) ont révélés plusieurs incohérences, et parce qu'il y avait de nombreuses valeurs „0' ou „unknown', les genre, âge et statut matrimonial ont été minutieusement vérifiés en retournant au manuscrit. Près de 400 sexes ont été modifiés, 114 âges ont été modifiés (dont 46 étaient originellement 0), et près de 700 statuts matrimoniaux ont été modifiés (dont 456 étaient indéterminés). Dans ce dernier cas, les erreurs semblaient plus fréquentes dans certains secteurs de la ville, liées soit à l'énumérateur ou la personne ayant fait l'entrée de données. Pour référence future, la liste de ces secteurs problématiques est donnée ici :

Secteurs de la ville avec données étranges (plus de 2% des individus) :

90 a 6  
90 g 4  
90 i 3  
91 e 2  
90 e 3  
90 e 8  
91 a 3.

Avec bcp de problèmes d'entrés de données :

91 e 2  
91 f 1  
90 g 6  
90 g 7

Aussi, mais moins pire :

90 g 5  
90 g 8  
90 h 2  
90 h 5  
90 h 15  
91 p.

#### **2.5.4.1 Age des bébés (AGEMONTH & BMONTH)**

The values for this variable were written in the age column of the manuscript, as a fraction of a year: 7/12 for seven months old. The month of birth was also inscribed for those who were not yet aged one year.

Some values were recoded further after a cross check with the age and validation on microfilm.

#### **2.5.4.2 Groupes d'âge (RECAGE & AGEGROCC)**

Afin de faciliter les analyses différenciées selon les âges, deux variables permettent les regroupements par groupes d'âges :

- RECAGE : Cette variable regroupe les individus par tranches de 5 ans, de 0 à 79, avec une catégorie de 80 et plus.
- AGEGROCC : Cette variable sert à regrouper les individus par phases de vie occupationnelle.

Ainsi, les regroupements vont de 15 à 59 ans, avec une catégorie pour les plus de 60 ans, qui doit être utilisée avec réserve. En effet, il a été noté que nombre d'individus de plus de 60 ans avaient déclaré une occupation et avaient par ailleurs déclaré être retraité, ce qui brouille les analyses qui pourraient être faites sur ce groupe d'âge.

#### **2.5.5 Infirmités (INFIRM)**

Le manuscrit contient trois colonnes où les recenseurs devaient indiquer l'infirmité des individus, le cas échéant. Ces colonnes ont été remplies par les énumérateurs, mais n'ont pas été enregistrées lors de l'entrée de données.

La variable INFIRM.05 donne relevé des données de recensement, sans ajustement aucun selon d'autres sources. La variable INFIRM.5 reprend l'information de INFIRM.05, mais a été ajustée en fonction d'informations contenues dans d'autres variables.

Les codes utilisés pour représenter les infirmités dans les deux variables sont associés au numéro de la colonne de feuilles de manuscrit, avec :

17 : sourd muet

18 : aveugle

19 : pour aliénation mentale

Les double infirmité ont été indiquées par les deux chiffres à la fois (ex. : 1719). Les codes dans la série 20 ont été ajoutés pour représenter d'autres types d'infirmités qui avaient été nommées dans les commentaires du recensement.

#### **2.5.6 Membres d'un ordre religieux (RLGORDR)**

Afin d'identifier et classier toutes les personnes appartenant à un ordre religieux, la variable RLGORDER.5 a été créée à partir de la variable OCC.0, et amendée selon des champs textes, comme suit :

1) Création initiale à partir d'un recodage des personnes dont l'occupation HISCO était dans la catégorie religieuse : 14100 à 14999.

2) Validation des occupation telles que transcrites afin d'assurer qu'aucune n'ait été oubliée, élimination des enfants de moins 5 ans.

3) Reclassification d'individus en fonction des champs de commentaires SPECIAL.3 & COMMENT.2, de même que certains noms et prénoms. Ceci requiert l'analyse de ménages entiers, puisque plusieurs religieux sont „cachés' dans des institutions.

En tout, 2078 individus ont été identifiés comme appartenant à un ordre religieux. Les seules occupations n'en identifiaient au départ que 1898, un amélioration de 9.5%.

### **2.5.7 Employés résidents & Ménages avec employés (DOMRES & HHDOMRES)11**

11 Plus de détail peut être trouvé dans le fichier décrivant spécifiquement les travaux faits relativement aux structures familiales.

12 *idem*

Cette variable permet d'identifier tous les individus ayant un emploi qui pourrait potentiellement les amener à résider chez leur employeur. Les ménages des individus ainsi identifiés ont ensuite été analysés pour séparer les employés résidant dans leur propre famille et ceux résidant chez leur employeur.

Tous les ménages ayant au moins un employé résidant dans le ménage ont été catégorisés comme ménages avec domestiques (HHDOM.5).

### **2.5.8 Rôles individuels (INDSIT)12**

This typology is based on „parenting' as the epiphenomenon around which HHs get organised. Hence, instead of situating individuals in relation to the traditional notion of „Head', the typology emphasizes the relation of individuals to all other members of the HH, with a specific focus on parenting or potential parenting.

10 Parent gen

30 Offspring

50 G-Parent gen

60 Relative to parent

65 Relative to g-parent

68 Related Independent

70 Independent

90 Res employee

91 Relative of res empl

99 Undeterminate

## **2.6 VARIABLES SUR LES COMMUNAUTÉS CULTURELLES**

Le fichier de NAPP contient des valeurs encodées pour les variables de lieu de naissance, d'origine et de religion. Le fichier CFP, qui provient d'une version antérieure de l'équipe NAPP, contient, en plus des valeurs encodées, les valeurs textuelles de ces variables, et parfois une valeur encodée seconde (ex. : religion & religion2). Ces secondes valeurs servent à capturer les informations textuelles doubles (ex : English Irish, encodée origin1= english & origin 2 = Irish). Les travaux qui ont été faits sur ces variables culturelles ont d'abord servi à valider ces dédoublements et encoder plus simplement les variables.

### **2.6.1 Lieu de naissance des individus (BPLC)**

Raison de l'encodage :

□ Pour identifier le lieu de naissance, deux variables principales existent : une qui donne les lieux de naissance (BPLCA.0), et ayant un grand nombre de catégories et une qui donne le pays de naissance (BPLCOUNTRY.0), avec un nombre moindre de catégories, mais qui ne discerne pas entre le Qc et le Canada.

La variable du lieu de naissance agrégé (BPLC.5) a été créée comme suit :

1. Création à partir de regroupements de la variable BPLCNTRY.0
2. Répartition des lieux de naissance Québec et Ontario dans des catégories plus spécifiques que „Canada’ (à partir de la variable BPLCA.0)
3. Ajustement en fonction des lieux de naissance inscrits dans DBRTHPL.2 (string)

### **2.6.2 Origine (ORIGIN)**

This variable is defined as the country of birth of the first male ancestor that immigrated to Canada. This variable thus describes the cultural community the individuals belong to.

The variable DORIGIN.2 provides the actual census inscription of the origin of each individuals. The NAPP variable ORIGIN.0 encoded these origins into categories that could be compiled. Where there had been two origins stated, the variable ORIGIN2CA.0 was used to reflect the dual origin.

In some cases, the origin of the individuals seemed problematic when compared to other household members or to the language of the name they bore. When the manuscript was checked, it was noticed that inverted commas had been used to indicate the origin of groups of people. Some of these inverted commas may have been put hastily and seemed to introduce mistakes in the origin of some families. Some other peculiar cases were validated against the manuscript and were amended. The recoded variable ORIGIN1.5 includes these corrections to the original NAPP variable.

The classification of origin was reviewed and modified due to problematic coding of many dual origins. The original variables mistakenly codified some single origins in the second origin, which made this variable unuseful. The second origin was completely recoded based on the origin string, and, when appropriate, the first origin was recoded accordingly. In total, 400 individuals have been recoded.

### **2.6.3 Religion (RLGN)**

The variable DRELIGN.2 provides the census statement of the individuals’ religion. The variable RELIGION.0 encodes these religions into categories that can be compiled and the variable RELIGION2.0 allows for dual religion by encoding a second religion stated on the manuscript.

Similarly to the situation described for the origin statement, some cases also seemed oddly coded and were validated and corrected into the recoded religion 1 & 2 variables. Some religion statement seemed dubious and had been stated using inverted commas and some cases showed a second religion when the string only showed one religion and they were recoded to blank. The recoded variable depicts the few individuals that stated a dual religion.

#### **2.6.3.1 Religion group (RLGN1GRP)**

This variable groups the different protestant religions into a smaller number of categories. The Montreal context is such that religion is an important factor, but in its opposition of Catholics and Protestants. Inner differences within Protestantism are not a key element in analysing this context and sub groups have thus been regrouped to simplify analysis on religion.

### **2.6.4 Communauté culturelle (ETH & ETHGRP)**

Le contexte du projet de recherche souhaite que les individus soient regroupés en trois grands

groupes culturels : Canadiens français, Irlandais catholiques et Anglo-protestants, mais aucune variable ne classifie les individus de cette façon.

La variable ETH.5 a été créée en combinant les variables d'origine et de religion, qui identifie 11 groupes différents (plus un groupe d'indéterminés).

La variable ETHGRP sert à regrouper les groupes en trois grandes familles suffisamment nombreuses et pertinentes aux analyses de cette équipe de recherche. Ces regroupements de ETH en ETHGRP ont été faits comme suit :

10	FC	10	French catholic
20	IC	20	Irish catholic
		30	English catholic
		40	Scottish catholic
		50	Other catholic
30	PR	15	French protestant
		25	Irish protestant
		35	English protestant
		45	Scottish protestant
		55	Other protestant
50	OT	59	Other
		99	To be reclassified

Par la suite, les cas de divergence entre les dossiers de Pat et de Sherry (43 exceptions) ont été analysés un à un pour assurer une cohésion.

### **2.6.5 Types de mixité ethnique (HHETH, HHETHYP, HHETHDOM)**

Élaboration d'une typologie à deux niveaux pour définir les types de mixité ethnique des ménages. Typologie appliquée au recensement de 1881, de même qu'à l'échantillon CFP.

#### **2.6.5.1 Processus de création de HHETH**

Afin d'identifier et quantifier les types de mixité ethnique des ménages, une variable détaillée a été créée comme suit :

1. Identifier les ménages dans lesquels il n'y a *qu'une seule ethnicité* et leur donner une valeur d'ethnicité unique selon leur groupe d'appartenance.

2. Des *ménages avec domestiques*, isoler les domestiques et valider si le ménage de base (sans les domestiques) a plus d'une ethnicité ou non. Si le ménage n'a qu'une ethnicité, assigner une valeur qui indique que la mixité provient des domestiques, selon le groupe d'appartenance du ménage principal et des domestiques qui s'y trouvent. Si le ménage a plus d'une ethnicité, assigner un code décrivant la mixité du ménage, incluant celle des domestiques, tel que décrit à la prochaine sous-section.

3. Aux *ménages mixtes sans lien matrimoniaux* ou parentaux, assigner une valeur de mixité complexe aux. (Puisque l'intérêt de cette variable est surtout de comprendre la mixité matrimoniale, la variable ne décrira pas en détail tous les cas de mixité qui n'engagent pas un couple, mais les placera plutôt dans une catégorie à part, pour des analyses futures le cas échéant.)

4. Des ménages comprenant *une famille* (par le nom de famille) *et des personnes indépendantes* (identifiables en utilisant la variable des types de ménages), valider si la famille prise isolément est mixte, ou si la mixité du ménage provient des personnes isolées. Dans ce

dernier cas, assigner une valeur qui indique que la mixité provient des personnes indépendantes, selon le groupe d'appartenance du ménage principal et des personnes

additionnelles qui s'y trouvent. Si la mixité provient de la famille, assigner une valeur de mixité complexe, sans tenter de décrire plus avant la forme de mixité.

5. Pour les ménages dont la mixité provient du lien matrimonial ou des enfants (ménages restants), assigner un code décrivant la mixité du ménage

#### 2.6.5.2 Codes

o Les codes de valeurs ont été élaborés selon la logique suivante :

1= Canadien-français, 2= Irlandais, 3= Anglo-protestant, 4 mixte, 5= autre.

o Codes à 2 et 3 chiffres : ménages sans domestique, et dont l'éventuelle mixité provient d'un couple ou d'une famille du ménage, et non pas seulement des personnes additionnelles à une famille de base.

o Codes à 4 chiffres : ménages avec domestiques et/ou dont la mixité provient de l'ajout d'une ou quelques personnes indépendantes d'une famille du ménage.

o 1er chiffre du code indique le groupe ethnique du mari/père. S'il est absent, le code est à deux chiffres.

o 2ème chiffre du code indique l'ethnicité de la femme/mère. Si elle est absente, le deuxième chiffre est un 0.

o 3ème chiffre du code indique l'ethnicité des enfants. S'il n'y a pas d'enfants, la valeur est 0. Si les enfants sont eux-mêmes mixtes, la valeur est 4.

o 4ème chiffre du code indique la mixité des domestiques. S'ils sont mixtes, la valeur est 4.

Cas particuliers: Si un ménage comporte une famille non mixte et des personnes indépendantes d'un ou de plusieurs groupes ethniques différents, le premier chiffre indiquera le groupe de la famille, le deuxième chiffre sera un 7, pour indiquer l'ajout de personnes additionnelles, le 3ème chiffre indiquera l'ethnicité de ces personnes additionnelles (un 4 si elles sont mixtes) et le 4ème chiffre indiquera l'ethnicité d'éventuels domestiques.

La résultante est une typologie à 190 codes différents, ce qui est complexe à gérer, mais permet de multiples regroupements selon les aspects de la mixité qui est étudiée. Les prochaines sections présentent deux variables qui permettent chacune un regroupement analytique.

#### 2.6.5.3 Type de mixité des ménages (HHETHYTP)

La variable HHETHYTP permet de regrouper les ménages selon le type de mixité de l'unité matrimoniale ou parentale :

- 1 Single ethnicity
- 2 Missing parent, mixed
- 3 Mix couple, no child
- 4 Mix couple, ch as father
- 5 Mix couple, ch as mother
- 6 Couple & child mix
- 7 Single eth + mix add.
- 9 Unclassifiable mix

#### 2.6.5.4 Mixité des domestiques et des ménages (HHETHDOM)

La variable HHETHDOM permet de regrouper classer les ménages avec domestiques :

- 10 FC fam, no dom
- 11 FC fam, FC dom
- 12 FC fam, IC dom
- 13 FC fam, AP dom
- 14 FC fam, mix dom
- 20 IC fam, no dom
- 21 IC fam, FC dom
- 22 IC fam, IC dom
- 23 IC fam, mix dom
- 25 IC fam, OT dom
- 30 AP fam, no dom
- 31 AP fam, FC dom
- 32 AP fam, IC dom
- 33 AP fam, AP dom
- 34 AP fam, mix dom
- 40 Mix fam, no dom
- 41 Mix fam, FC dom
- 42 Mix fam, IC dom
- 43 Mix fam, AP dom
- 44 Mix fam, mix dom
- 50 OT fam, no dom
- 51 OT fam, FC dom
- 52 OT fam, IC dom
- 53 OT fam, AP dom
- 54 OT fam, mix dom
- 55 OT fam, OT dom
- 70 Mix fam & add, no dom
- 71 Mix fam & add, FC dom
- 72 Mix fam & add, IC dom
- 73 Mix fam & add, AP dom
- 74 Mix fam & add, mix dom

## 2.7 VARIABLES OCCUPATIONNELLES

Raisons de l'encodage :

- o La variable occupationnelle est un point d'ancrage de plusieurs analyses subséquentes et il importe qu'elle soit juste et traitée de façon cohérente lorsqu'elle est utilisée pour des variables spécifiques.

- o Il existe plusieurs variables ayant trait à l'occupation, mais elles ne sont pas toutes cohérentes. Il arrive que certaines occupations inscrites n'ont pas été encodées dans la variable principale d'encodage (HISCO),. Ces individus sont de plus considérés comme sans occupation dans les autres variables.

- o Les apprentis, les religieux et les étudiants sont parfois encodés comme tels, et parfois comme „sans occupation”, et ce, non pas de façon cohérente, mais plutôt aléatoire ou selon la qualité de l'information encodée ou la localisation de l'information.

- o Certaines inscriptions „non-occupationnelles (ex. aveugle, écossais) ont été encodées de diverses façons dans diverses variables.

### **2.7.1 Occupations textuelles (OCCSTRNG, OCCNOTE & OCC\_CL)**

Les occupations données dans OCCSTRNG n'étant pas toujours compréhensibles et certaines occupations ayant été encodées comme „missing’ il est apparu que la variable avait besoin de nettoyage. Des vérifications ont donc été faites sur le manuscrit, et les modifications ont été enregistrées dans la variable OCCNOTE.5, créée pour contenir les corrections faites à la description des occupations, et l'information relevée lors des vérifications des occupations inscrites sur le manuscrit.

- o La majeure partie des inscriptions a été validée à partir du manuscrit.
- o Les valeurs „ \* „ indiquent que le manuscrit a été vérifié et que l'inscription est telle que lue originellement ou illisible.
- o Les valeurs entre parenthèses indiquent des commentaires, et on a une inscription réelle sur le manuscrit. Les cas „ (marital status) „ indiquent que la valeur indiquée dans la base de données reflète en fait la colonne adjacente, qui contient le état matrimonial des individus.

Plus valide grâce à ce travail, la variable OCCSTRNG reste toutefois difficile à utiliser parce que les occupations y apparaissent telles que déclarées lors du recensement, ce qui signifie par exemple que les comptables, accountants, accountant, accounting forment tous des titres occupationnels différents. Il y avait donc un besoin pour une variable qui :

- o standardiserait les titres occupationnels dans une seule langue (l'anglais)
- o serait nettoyée des commentaires superflus
- o permettrait des regroupements logiques (tous les comptables ensemble, et les commis, quel que soit leurs titres).

La variable OCC\_CL sert donc à standardiser la variable OCCSTRNG (incluant les notes prises dans OCCNOTE). Pour ce faire, certains titres sont parfois inversés (ex. broker insurance, vs insurance broker), sans qu'il n'y ait perte d'information. Cette variable a permis de passer de 3676 titres occupationnels à 1828 titres (50% réduction).

### **2.7.2 Occupation HISCO13**

*13 For more information on this coding scheme, consult their web site at: <http://hisco.antenna.nl>.* Cette variable a été créée à partir de la variable de NAPP OCC.0, et a été modifiée pour redresser des incohérences (ex. commis & clerk dans catégories différentes), préciser certaines catégories floues, refléter les vérifications faites sur le manuscrit et prendre en considération les commentaires faits par les différentes chercheuses ayant travaillé ces variables (selon commentaires inscrits dans OCCNOTE.5, COMMENT.2, SPECIAL.3, OCCNOTE).

Le tableau présente les opérations/remarques/conclusions faites sur des codes spécifiques HISCO, notamment tous les codes 0 et plus élevé que 99120 donnés par NAPP.

0 No occupation (coded as missing)	This code has been used for all individuals for which there was no mention of occupation, as well as for any registered comment into the occupation variable that did not relate to an occupation. Some examples of such cases were : marital status, „bourgeois’, „not given’, etc.
---------------------------------------	--

14000 religious order members	Many changes were done to this category since many sisters and brothers had not indicated their status as an occupation, but registered it as a comment or though their names. They were recoded into this category.
39000 Clerks, nfs	Special attention must be paid to analysis on these categories. The CLERK occupation was rarely stated what type of work the clerks were performing and one can be certain that office clerks, store clerks and other types of clerk are appropriately encoded. Furthermore, the English „clerk’ had been coded as 39000, while the French „commis’ had been encoded as 45130, which introduces a linguistic bias in regards to clerical vs. retail types of occupation. The French „commis’ were recoded as 39000 to reduce bias
45130 Clerks in shops and stores (1575 changes)	
99120 Labourers nfs	These three LABOURER codes represent a similar phenomena as for the clerks: even though a distinction has been made by original encoders, they should be used as a single category for analysis. The English expressions may be distinct, but there is not justification to categorize the French „journalier’ more in one than the other.
99130 Common or general labourers	
99140 Day labourers (e.g journalier)	
99150 Worker nfs	This code has been used when there was an indication of employment, but the type of work did not refer or could not be associated with factory work.
99200 Factory labourers (unspecified)	This code has been used when there was an indication of employment that could be associated with factory work, either through a mention of factory or a relevant type of work.
99300 Occupational title	This code has been used when the occupation was readable

unclassifiable	and understandable, but could not be classified into one established category. Apprentices have been classified into this category.
99420 Assistants nfs	XXX
99430 Helpers nfs	
99500 Ambiguous responses	This code has been used when an occupation was stated but it was either not readable or incomprehensible.

De plus, dans un effort de classification des statuts propres à chaque occupation et de standardisation avec des données de 1901, certains groupes occupationnels ont été regroupés ou modifiés. Les résultats est une variable HISCO non directement transférable avec les données de 1901, mais dont les regroupements peuvent s'apparenter et être comparés.

**Note :** *Les groupes occupationnels des commis/clerks, des journaliers/labourer/day labourer/char woman/etc et celui des travailleurs des ventes (marchants,/merchants/sales person, etc.) doivent être revus pour être catégorisés plus spécifiquement. Les dénominations de 1881 étant moins détaillées que celles de 1901, il a fallu les regrouper pour être en mesure de comparer. Mais un travail de re-classification des individus de ces groupes devrait être fait pour mieux représenter leur situation et divergences.*

Plus de 10 000 occupations ont été modifiées, dont 1571 clerks, 1132 joiners, 1989 drivers, 474 labourers.

#### 2.7.2.1 Statuts et secteurs occupationnels (OCCSTATUS & OCCSECTOR)

Les codes Hisco ont été regroupés selon deux axes :

- o En fonction du statut social de l'occupation :
  - A. Grande bourgeoisie
  - B. Petite bourgeoisie
  - C. Commis
  - D. Artisans
  - E. Ouvriers semi-qualifiés
  - F. Journaliers
- o Et en fonction du secteur industriel de cette occupation

- 1 Construction
- 2 Railroad
- 3 Transport other
- 4 Service pers.
- 5 Service food & Hotel
- 6 Service prot.
- 7 Finance
- 8 Merchants

- 9 Other trade
- 10 Professional
- 11 Whitecollar oth

- 12 Printing
- 13 Mfg metalwork
- 14 Mfg wood & paper
- 15 Mfg food & tob
- 16 Mfg shoe & leat
- 17 Mfg textile & clothing
- 18 Mfg other
- 19 Labourers
- 20 Farm, forest, fishing

Ces variables doivent donc être modifiées si et quand la codification HISCO est modifiée.

#### *2.7.2.2 Variables excluded from NAPP download*

The following variables have been removed from the file because the information they relate is inconsistent and should be used only as complementary information.

- o The OCSTATUS.0 variable highlights status information that had been registered in the description of the occupation (eg.: apprentice butcher). It is a complement of information to the HISCO codes, but is not a classification of the occupations codes in regards to their social status.

- o Same for OCRELATE, which reflects relational information that might have been registered in the occupation description. The frequency distribution shows that this information was not frequent and this available must be used with care since such information may not have been systematically collected.

These two variables contain information that could not be captured in the HISCO code, and represent additional information. They do not consistently capture specific situations, and have not been validated. They should be used as a complement of HISCO, but in their current status, never on their own.

- o Language occupation stated in

This variable provides the language in which the occupation of each working individual was registered on the census sheet. It was not a census variable, and does not represent the language in which people were working. It was collected by NAPP to complement occupational interpretation.

#### **2.7.3 Occupation (3-lettres) (OCC)**

Certains membres de l'équipe de recherche travaillent avec un autre système de classification que les codes HISCO. Cette méthode de classification est basée sur les loyers médians payés par les individus dans chaque titre occupationnel et permet donc une évaluation des statuts sociaux de chaque occupation. Cette typologie à 3 lettres contient environ 250 codes, qui peuvent être regroupés en 20 secteurs industriels (tel qu'appliqué à la variable HISCO) et en 6 statuts sociaux (A à F). Ces statuts n'ayant pas été déterminés de la même façon que ceux de HISCO,

ils ne sont pas équivalents, donc pas directement comparables.

La codification à trois lettres a été élaborée à partir des titres occupationnels et non en relation avec la codification HISCO. Les travaux faits sur la qualité de la variable HSCO ont donc dû être faits à nouveau, pour assurer que les clarifications et corrections des titres occupationnels étaient aussi reflétées dans cette variable.

#### **2.7.4 Fréquentation scolaire et grandes classes d'occupation (SCHOOL.05 & SCHOOL.5 & HHSCHOOL)**

##### *2.7.4.1 Fréquentation scolaire*

Lors du recensement, une question était posée à savoir si les jeunes d'un ménage allaient à l'école ou non. Cette information a été enregistrée dans la colonne 16 du manuscrit, mais n'a pas été retranscrite lors de l'entrée dans la base de données. Puisque cette information était requise pour els recherches concernant les jeunes de 15 à 29 ans, cette équipe de recherche a colligé cette information à partir des microfilms et a créé la variable SCHOOL.05 pour ce faire.

De 6 377 jeunes qui allaient à l'école (selon NAPP), dans les secteurs suivants :

90 e 3

90 e 4

90 e 6

90 e 7

90 e 8

90 f 1

90 f 2

90 f 3

90 f 4

90 f 5

90 f 6

90 f 7

90 f 8

90 f 9

90 h 14

90 h 15

nous sommes passés à 27 279 dans tous les secteurs, et en combinant cette information avec celle inscrite dans le champs d'occupation, à 27 824 étudiants.

La variable SCHOOL.5 reprend cette information colligée, et y ajoute celle des titres occupationnels. Et comme l'intérêt de cette variable était d'identifier l'occupation des jeunes, d'autres catégories se sont ajoutées, telles que apprenti, labourer, assistant, etc. Cette variable est devenue un peu fourre-tout et nécessitera quelques travaux pour préciser son utilité et clarifier les catégories. Mais c'est à ce jour la variable la plus fiable pour ce qui est des occupations des jeunes, dont la fréquentation scolaire et les apprentis.

##### *2.7.4.2 Taux de fréquentation scolaire du ménage*

Dans le cadre des études de mobilité sociale, il était intéressant de catégoriser les ménages selon le taux de fréquentation scolaire des enfants qui le composaient. Cela allait permettre de savoir si un jeune avait des chances d'aller à l'école, même s'il était lui-même trop jeune ou trop vieux pour déclarer y être. Le taux de fréquentation scolaire du ménage a été déterminé comme suit:

1- Afin de préciser les données sur lesquels calculer les rations, sélectionner seulement les ménages à famille simple et avec des enfants entre 6 et 13 ans (ceci concerne environ le tiers

de la BD : 65 301 individus).

2- Pour chaque ménage, compter le nombre d'enfants entre 6 et 13 qui sont à l'école, et diviser par le nombre d'enfants de 6 à 13 ans dans le ménage.

3- Regrouper les taux de fréquentation en 3 niveaux de fréquentation : no child in school, 10-50% in school, 60-100 % in school. La valeur a été appliquée à tous les individus du ménage

[...]

## 5 ANNEXES

### 5.1 LISTE DES VARIABLES 1881

Variable	Label	N	Minimum	Maximum	Mean	Std. Deviation	Measm'nt Level	Value labels	Missing Values	Comment
Census variables		0					Scale			
VERSION.2	Release date						Nominal			
COUNTRY.0	Country of residence	180847	124	124	124	0	Scale	yes		
NATYEAR.2	Nation and Year	180847	1881	1881	1881	0	Ordinal	yes		
VOLUMECA.0	Volume, Canada	180847	1	23	9,29	5,407	Scale			
REELNACA.0	Reel number, National Archives	180847	13217	13222	13219,17	1,49	Scale			
PGENUMCA.0	Page number, Canada	180847	1	156	46,56	31,659	Scale			
LNENUMCA.0	Line number, Canada	180847	1	26	12,97	7,207	Scale			
INVALID.5	Removed cases (1)	180847	0	1	0	0,013	Scale			
COMMENT.05	Comment from manuscript						Nominal			Collected by us
geographical		0					Scale			
PROVINCA.0	Province, Canada	180847	170	170	170	0	Scale	yes		
REGIONNA.0	Census region, North America	180847	15	15	15	0	Scale	yes		
DSTRCTCA.0	District, Canada	180847	90	91	90,22	0,415	Nominal			
SDSTCA.0	Subdistrict, Canada						Nominal			
SDSTNMCA.0	Subdistrict name, Canada						Nominal			
DIVNOCA.0	Division number, Canada	180847	0	15	4,73	3,728	Nominal			
GEOCODE.2	Geocode						Nominal			
NEIGHB.3	Neighborhood						Nominal	yes		
TPROPID.3	Cadastral lot	180847	0	92228	34987,66	34689,87	Scale			
TRES.3	Localisation	45320	81302	904884	624857,4	187996,6	Scale			
hhvariables		0					Scale			
SERIAL.5	Household serial number	180847	344997	900159	365488,5	34918,44	Scale			
HHSIZE.5	Household size	180847	1	577	10,67	39,323	Scale			
FPERNUM.5	Family index within HH	176405	1	26	1,54	1,843	Scale			
FAMID.5	Family identification number	176405	9000001	90015901	36533759	3421114	Scale			
NFAM.5	Number of families in HH	176405	1	23	1,69	1,543	Scale			
FAMSIZE.5	Family size	176405	1	25	5,08	2,71	Scale			
HHTYPE.5	HH Type	180847	10	999	386,67	161,072	Scale	yes		Created by CS

Variable	Label	N	Minimum	Maximum	Mean	Std. Deviation	Measm'nt Level	Value labels	Missing Values	Comment
INST_NAME.5	Institution name						Nominal			Incomplete information Created by CS
individualvariables	INDIVIDUAL	0					Scale			
PERNUM.5	Person number within HH	180847	1	577	5,83	22,909	Scale			
ID	Individual identification number	180847	5,61E+08	6,11E+08	5,83E+08	14767868	Scale			Unique, unchanged individual identifier
NAMTTL.5	Title in name						Nominal			Complement to NAMLAST.5 Created by CS
NAMFRST.5	First name, recoded						Nominal			
NAMLAST.5	Last name, recoded						Nominal			
NAMLAST2.5	Second-altern last name						Nominal			Complement to NAMLAST.5 Created by CS From Baby & Death files
NAMMAID.5	Maiden name						Nominal			Created by CS
HHMAIDEN.5	HH w maiden name	180847	1	9	7,42	3,188	Scale			Created by CS
SEX.5	Sex	180845	1	2	1,53067	0,49906	Nominal	yes	9	
MARST.5	Marital status	180841	1	6	4,275662	2,334442	Nominal	yes	9	
AGE.5	Age	180613	0	104	24,7	18,229	Scale		999	
RECAGE.5	Five-year age groups	180619	1	17	5,57	3,657	Scale		99	
AGEGRPOCC.5	Age grp for occ	54079	2	5	3,18	0,799	Scale	yes		Defined for occupational analysis Created by CS
INFIRMITY.5	Infirmary	729	9	1819	41,5	200,364	Scale	yes	9999	Collected by us
INDSIT.5	Individual situation in HH	180847	10	99	30,24	21,469	Scale	yes		Related to HHTYPE.5 Created by CS
CLERGY.5	Member of religious order	180847	1	9	8,91	0,854	Scale	yes		Created by CS
DOMRES.5	Residing domestic	180847	1	99	95,11	18,738	Scale	yes		Created by CS
HHDOMRES.5	HH with residing domestic	180847	1	99	81,41	36,843	Scale	yes		Created by CS
occupation	OCCUPATION	0					Scale			
OCCSTRNG.0	Occupation as transcribed						Nominal			
OCCNOTE.5	Occupation note						Nominal			Complements OCCSTRNG.5

Variable	Label	N	Minimum	Maximum	Mean	Std. Deviation	Measm'nt Level	Value labels	Missing Values	Comment
OCC_CL	Cleaned occupation titles									
OCCHISCO.5	Occ NAPP-modified HISCO	65272	1100	99500	67375,9	27290,28	Scale	yes	0	
OCCSTATUS.5										
OCC SECTOR.5										
OCC.5	Occupation code (sherry)						Nominal	yes		
SCHOOL.5	Going to school?	92821	10	90	37,03	18,092	Nominal	yes	99	Collected by us
ethnicity	ETHNICITY	0					Scale			
BPLC.5	Birth place	180488	10	70	13,5257	8,582027	Scale	yes	99	
ORIGIN1.5	Ethnic origin	180192	1100	8410	1634,335	824,4651	Scale	yes	9970, 9980, 9990	
ORIGIN2.5	Ethnic origin, second stated	274	1100	7301	1972,996	1079,294	Scale	yes	9970, 9980, 9990	
RELIGION1.5	Religion	180601	1100	9970	1798,451	1591,816	Scale	yes	9950, 9980, 9990	
RLGN1GRP.5	Religion groups						Nominal	yes		
RELIGION2.5	Religion, second stated	42	1200	8895	3699,762	1607,27	Scale	yes	9970, 9980, 9990	
ETH.5	Ethnicity (lg +rel)	180847	10	99	18,51	12,594	Scale	yes		3 main groups; basis for HHETH.5
ETHGRP.5	Ethnic group (3)	180847	10	50	16,42	8,621	Scale	yes		Created by CS
HHETH.5	HH ethnicity - detailed	180847	12	5734	605,34	1083,968	Scale	yes		Groups based on HHETH.5 Created by CS
HHETH TYP.5	HH ethnicity types	180847	1	9	1,67	2,025	Scale	yes		Groups based on HHETH.5 Created by CS
HHETHDOM.5	Ethnicity of HH res.empl.	180847	10	94	22,56	20,357	Scale	yes		Groups based on HHETH.5 Created by CS
Archivevariables	ARCHIVE	0					Scale			