

*Le jumelage des données nominatives dans les recensements : problèmes et méthodes**

par Raymond ROY**, Christian POUYEZ** et François MARTIN***

L'utilité des listes nominatives des recensements pour la connaissance des sociétés n'a plus à être démontrée: toutes les branches des sciences sociales sont susceptibles d'utiliser ces sources qui ont l'avantage, lorsqu'elles sont bien faites, de procurer aux chercheurs un « instantané » de la société, à un moment donné de son évolution. Le rapprochement de plusieurs de ces instantanés permet, sinon de reconstituer le film de cette évolution — ce qui n'est possible qu'avec un véritable registre de population — au moins d'en saisir le sens, d'en préciser les aspects les plus importants.

L'étude de la mobilité géographique est certainement un des domaines où les listes nominatives s'avèrent les plus utiles. Elles seules — et plus particulièrement celles des recensements — permettent d'analyser les migrations à un niveau de raffinement bien supérieur à ce que révèlent les analyses de soldes migratoires, à partir de données agrégées. C'est d'ailleurs la constatation qui s'impose d'emblée à tous ceux qui ont travaillé sur la question: sous l'apparente stabilité des chiffres de population et des soldes migratoires se cache souvent un extraordinaire mouvement des personnes, et cela aussi bien dans les communautés rurales que dans les villes. La comparaison systématique de plusieurs listes nominatives permet de saisir et mesurer ce mouvement, d'en évaluer les dimensions sociales et économiques et, d'analyser les causes de cette mobilité.

Une telle comparaison suppose qu'on effectue ce que, dans le jargon du métier, on appelle du jumelage ou encore du couplage de mentions nominatives. En effet, à la base de toutes les études de mobilité faites à partir des recensements, on

* Les travaux qui ont mené à la rédaction de cet article ont été subventionnés par le Ministère de l'Éducation du Québec (F.C.A.C.) et par le Conseil des recherches en sciences humaines du Canada. Nous les remercions, ainsi que les collègues qui ont bien voulu lire et critiquer la première version de ce texte.

Le présent texte a été rédigé dans le cadre du « Programme de recherches sur la société sague-nayenne ». Créé en 1972, à l'Université du Québec à Chicoutimi, ce programme réunit une équipe pluridisciplinaire formée d'informaticiens, de démographes, de médecins, de sociologues et d'historiens. Il a pour but de constituer, pour l'ensemble d'une société régionale et pour la période allant de 1840 à nos jours, un registre informatisé et universel fondé sur le jumelage automatique de données nominatives.

Le registre prend la forme d'un fichier-réseau, ou *data base*, il contient des données à caractère économique, social, culturel, démographique, génétique, médical, etc. et il donne lieu à des analyses et enquêtes dans chacune de ces directions de recherche.

** Département des sciences humaines, Université du Québec à Chicoutimi.

*** Centre d'informatique, Université du Québec à Chicoutimi.

retrouve cette hypothèse que toute personne présente sur une liste d'habitants de la localité « x » et encore présente sur une liste postérieure est une personne sédentaire. Les personnes qu'on ne retrouve pas, d'une liste à l'autre sont considérées comme émigrantes¹. Il suffit donc de comparer une à une les mentions nominatives de deux listes quelconques pour identifier les sédentaires et les migrants, et procéder ensuite, sur chacun de ces groupes, à toutes les analyses voulues. Théoriquement très simple, cet exercice de jumelage se heurte, en pratique, à trois ordres de difficultés²: la première, dont nous reparlerons plus longuement ci-dessous, tient à l'existence d'homonymes; une comparaison portant uniquement sur des individus isolés est donc soumise au risque de confondre des personnes différentes. La deuxième difficulté vient du fait que les mentions d'une même personne dans deux listes différentes varient souvent. Il ne s'agira parfois que de variations mineures, affectant par exemple une seule lettre de la mention, mais souvent, il peut s'agir de variations graves, véritables mutations de prénoms ou même de patronymes. Le problème ne vaudrait pas la peine qu'on s'y attarde longuement si la masse des données à traiter n'introduisait des difficultés d'un autre ordre. Dans une population de taille réduite, en effet, le chercheur apprend vite à connaître ses familles et, à l'intérieur des familles, pratiquement chaque individu. Dans ce cas — qui est celui de la plupart des monographies portant sur un petit village — les variations orthographiques et les mutations nominatives ne posent guère de problèmes. Cependant, dès que l'étude porte sur une population de quelque importance — une ville, un comté, une région — le jumelage doit être fait par ordinateur, à cause de la lenteur du travail manuel; or, chacun sait que l'ordinateur a besoin de règles très précises, lui indiquant la marche à suivre en toutes circonstances. En deux mots, le recours à l'ordinateur suppose la mise au point d'algorithmes de jumelage automatique des mentions nominatives, algorithmes qui soient assez souples pour rapprocher et jumeler des mentions comportant des différences mineures, et assez sévères pour rejeter tout faux jumelage.

Il n'existe pas de solution entièrement a priori à ce problème d'équilibre. Il est donc essentiel de procéder à des études exploratoires, de façon à reconnaître empiriquement la nature exacte des problèmes et les limites à l'intérieur desquelles se trouvent les solutions. C'est dans ce contexte qu'il faut situer le présent article : à partir des données nominatives des recensements fédéraux de 1852 et 1861 pour toute la région, nous avons effectué un jumelage semi-automatique portant sur l'ensemble des familles du Saguenay. Les résultats de cet exercice feront l'objet de trois articles: l'un portant essentiellement sur la construction d'un algorithme de jumelage automatique des données des recensements, un autre présentant les principaux résultats de l'analyse de la mobilité géographique³. Le présent article, enfin, a pour objectif de présenter les méthodes suivies lors de ce jumelage semi-automatique, de discuter certains problèmes inhérents à toute étude de mobilité, qu'elle

¹ Il est clair que la valeur de cette hypothèse est d'autant plus grande que les listes nominatives utilisées sont rapprochées dans le temps: le risque d'émigration temporaire est alors réduit en proportion.

² Nous ne signalons ici que les difficultés propres au jumelage proprement dit. D'autres problèmes, tel le repérage et la mesure de la mortalité, sont également très importants. Nous les abordons plus loin.

³ Ces deux articles sont en préparation (date prévue pour la publication: 1980). Les résultats ont entretemps fait l'objet d'une communication au colloque annuel de l'Institut d'histoire de l'Amérique française en octobre 1979.

soit ou non réalisée par ordinateur, et d'examiner dans une perspective comparative quelques-uns des résultats acquis sur l'émigration extra-régionale.

I. — LES MÉTHODES

La plupart des études de mobilité géographique, utilisant des listes nominatives (recensements, *city directories*), procèdent de la façon suivante: à partir d'une liste de recensement, dans laquelle on trouve généralement les noms, prénoms, sexe, âge, profession et résidence, et parfois aussi lieu de naissance et religion, on essaie de retrouver, dans les listes de recensement subséquentes les individus qui étaient déjà dans la première liste. On doit noter tout de suite que, le plus souvent, ce sont seulement les chefs de ménage qui sont ainsi repérés, et que les résultats obtenus sont généralement attribués à l'ensemble de la population, ce qui, on le constatera ci-dessous, ne va pas de soi. Quand tous les individus repérables ont été retrouvés, on tente parfois d'évaluer, rapidement, l'importance de la mortalité, et on conclut que tous les individus qui n'ont pas été retrouvés sur la deuxième liste et qui ne sont pas morts entre les deux dates de relevé, sont des émigrants.

La principale critique qu'on peut faire de cette façon de procéder, c'est qu'elle n'accorde pas assez d'importance aux mutations nominatives⁴. Dans les meilleurs cas, c'est-à-dire lorsque le jumelage porte sur deux listes de recensement, les éléments sur lesquels peut porter le jumelage sont les noms et prénoms, l'âge, le sexe, la religion et le lieu de naissance. La résidence et la profession ne peuvent évidemment être utilisées, puisque la mobilité géographique (et souvent, sociale) fait l'objet même de l'enquête. Des autres éléments, le lieu de naissance, la religion et le sexe ne sont pas très discriminants; seuls demeurent les noms et prénoms, et l'âge. Dans les études qui sont faites à partir des *city directories*, c'est-à-dire dans presque toutes les études faites sur les Etats-Unis⁵, on ne dispose même pas de l'âge. C'est donc dire que les éléments essentiels au jumelage sont les noms et prénoms. Or, l'expérience démontre que ce sont là des réalités très fluides, et que par conséquent, un jumelage qui ne porterait que sur le nom et le prénom d'un individu présenterait des risques considérables, à cause des phénomènes d'homonymie, qui amènent à sous-évaluer l'émigration, et à cause des mutations nominatives qui, en sens inverse, amènent à la surévaluer. La plupart des auteurs qui ont travaillé sur le sujet reconnaissent l'existence du problème; mais il y en a très peu qui ont tenté d'en mesurer l'ampleur. Ainsi, Chudacoff⁶ qui affiche pourtant une très grande confiance dans les *city directories* admet que la source d'erreur la plus

⁴ Les travaux effectués sur la banque de données nominatives du PRSS ont permis à Gérard Bouchard d'établir une typologie des mutations nominatives. On distingue ainsi: a) les variations graphiques au sens strict; b) les variations phonétiques; c) les patronymes composés; d) les prénoms composés; e) les prénoms usuels multiples. G. BOUCHARD, P. BRARD et Y. LAVOIE, « Un code de transcription phonétique pour la reconstitution automatique des familles saguenayennes » soumis à *Population*.

⁵ Dans la plupart des études américaines, on part d'une bonne liste de recensement, et on utilise, pour les années subséquentes, les *city directories*. Voir, par exemple, H. P. CHUDACOFF, *Mobile Americans. Residential and Social Mobility in Omaha, 1880-1920*, New York, Oxford University Press, 1972. Le procédé a été repris par de nombreux autres chercheurs. S. THERNSTROM, *The Other Bostonians. Poverty and Progress in the American Metropolis*, Cambridge, Mass., Harvard University Press, 1973 utilise très largement les *city directories*.

⁶ CHUDACOFF, *Mobile Americans...*, p. 25.

fréquente est la variation dans l'orthographe des noms. Alwyn Barr, plus sévère, souligne quelques-unes des difficultés dues à l'usage des *directories*: erreurs d'omission, assez fréquentes; variations orthographiques des noms et des prénoms; valeur douteuse des jumelages portant sur des noms communs, à cause des cas d'homonymie⁷.

Ces difficultés ne sont pas propres aux seuls *directories*: dans une étude basée sur les recensements américains de 1850 à 1880 pour les deux villes de Binghamtown et Johnstown, Dennis Kelly a constaté que les noms de famille variaient dans une proportion de 18 à 36% des cas, selon les périodes⁸. Dans notre propre enquête sur le Saguenay au milieu du 19^e siècle, on a pu établir que, chez les hommes, les noms et prénoms n'étaient identiques, d'une liste de recensement à l'autre, que dans 78% des cas; chez les femmes, cette proportion tombe à 62% (Tableau I). C'est donc dire qu'une méthode de jumelage qui serait basée sur le nom et le(s) prénom(s) des individus et qui exigerait leur parfaite identité s'exposerait à une surévaluation qui, au Saguenay, atteindrait en moyenne 30% des individus « jumelables »: parmi les 1,230 individus considérés au Tableau I, il y en a 370 qui viendraient s'ajouter aux véritables émigrants.

Tableau I

IDENTITÉ DES ÉLÉMENTS NOMINATIFS, SELON LE SEXE,
DANS DEUX LISTES DE RECENSEMENT, 1852 ET 1861

Sexe	Éléments nominatifs identiques dans les deux listes									
	Nom et prénom		Nom		Prénom		Aucun		Total	
	N.A.	%	N.A.	%	N.A.	%	N.A.	%	N.A.	%
Hommes	478	78	82	13	33	5	22	4	615	100
Femmes	382	62	116	19	75	12	42	7	615	100
Ensemble	860	70	198	16	108	9	64	5	1230	100

N.A. = Nombre absolu

À l'inverse, les cas d'homonymie peuvent facilement amener à sous-évaluer l'émigration en créant de faux jumelages: le fait, par exemple, de trouver un Louis TREMBLAY en 1861 ne signifie pas nécessairement que c'est le même Louis TREMBLAY qui était présent en 1852, et cela même si les âges concordent de façon satisfaisante. Ce problème se pose avec tous les noms communs, qu'il s'agisse des SMITH ou des JONES, dans les pays anglophones, des DUPONT et des DURAND en France, ou des TREMBLAY, SIMARD et BOUCHARD au Saguenay. Là encore, il importe de le souligner, le problème s'aggrave à mesure que grossit la taille de la population considérée: dans le cadre d'un village, la probabilité de trouver deux Louis TREMBLAY du même âge est relativement faible. Cette même probabilité est évidemment beaucoup plus forte si on considère un ensemble régional.

⁷ A. BARR, «Occupational and Geographic Mobility in San Antonio, 1870-1900 », *Social Science Quarterly*, vol. 51 (septembre 1970), p. 398, n. 8.

⁸ D. KELLY, « Linking Nineteenth-Century Manuscript Census-Records: A Computer Strategy », *Historical Methods Newsletter*, vol. 7, n° 2 (March 1974), p. 74.

Ces difficultés, auxquelles se heurtent toutes les études basées sur des listes nominatives, se posent avec acuité lorsque le jumelage porte sur des individus. Si, au lieu de considérer des individus isolés, on tient compte des personnes avec qui vivent ces individus, les risques découlant des mutations nominatives et de l'homonymie disparaissent presque totalement. Un seul exemple suffira à illustrer notre propos: en 1852, au Saguenay, on comptait trois François BOUCHARD; en 1861, on en comptait cinq. Même en tenant compte des âges, il est très difficile de jumeler correctement les trois individus de 1852 à trois des cinq individus de 1861. Par contre, si on fait intervenir le nom du conjoint, la situation devient très simple; la liste suivante, qui énumère les huit mentions de couples impliquant un François BOUCHARD en 1852 et en 1861 montre bien que l'utilisation du nom du conjoint est très discriminante: elle permet de jumeler avec sûreté deux des trois mentions de 1852, et de refuser le jumelage de la troisième. Il en va de même pour les cas de mutations nominatives: le nom du conjoint et, dans les cas les plus difficiles, les noms des enfants permettent de régler tous les cas.

1852	François BOUCHARD	Marie DUCHÊNE
1861	François BOUCHARD	Marie DUCHÊNE
1852	François BOUCHARD	Olympe GOBEIL
1861	François BOUCHARD	Olympe GOBEIL
1852	François BOUCHARD	Edyle TREMBLAY
1861	François BOUCHARD	Justine ABOR
1861	François BOUCHARD	Victoire GAGNÉ
1861	François BOUCHARD	Asseline SAILLANT

On a donc décidé, dès le départ, de tenter de jumeler non pas des individus, mais des couples, ce qui permet de surmonter les difficultés évoquées ci-dessus. Avant d'indiquer plus précisément comment on a procédé, il faut dire tout de suite ce que nous fait perdre cette façon de faire: on ne saisit ainsi que les couples, et non pas l'ensemble de la population, ce qui signifie, concrètement, que les personnes ne faisant pas partie d'un couple tel que défini ci-dessous échappent à l'analyse. C'est une perte importante, puisqu'il s'agit le plus souvent de jeunes adultes célibataires dont la mobilité est plus importante que celle de leurs aînés; aussi tentera-t-on, après avoir évalué la mobilité des couples, de transposer ces résultats à l'ensemble de la population. On notera également que la très grande majorité des études de mobilité, et spécialement celles qui sont faites à partir des *directories*, ne concernent que les chefs de ménage. En ne considérant que les couples, nous en faisons donc plus que la majorité des auteurs.

Partant de tous les couples existant en 1852, on a tenté de retrouver ces couples sur la liste du recensement de 1861. Ont été considérés comme couples ou comme noyaux familiaux en 1852 *toutes les unités formées d'au moins deux personnes unies par un lien conjugal ou un lien de filiation*. Ainsi, un veuf vivant avec son fils a été considéré comme constituant un couple; par contre, un homme ou une femme seule, même s'ils sont dits "mariés", ne forment pas un couple. Par exemple, les hommes qui, au moment du recensement de 1852, vivaient dans des chantiers forestiers n'ont pas été considérés comme formant des couples, même si bon nombre d'entre eux étaient, de fait, des hommes mariés. Par contre, les conjointes de ces hommes, dans la mesure où elles avaient au moins un enfant, ont été considérées comme formant un noyau familial.

Pour être considérés comme présents en 1861, ces couples devaient répondre aux mêmes exigences qu'en 1852 (c'est-à-dire ≥ 2 personnes, lien conjugal, lien de filiation); en outre, tout couple où se produit le décès d'un conjoint, suivi du remariage du survivant, et tout remariage de veuf ou de veuve pendant la période 1852-1861 sont comptés à part. En effet, si un couple de 1852 perd un de ses conjoints, et que le survivant se remarie, ou si un veuf ou une veuve se remarie, tout en demeurant au même endroit, le jumelage ne pourra être fait, puisque le couple à l'arrivée sera différent du couple de départ, et le couple risque alors d'être considéré, à tort comme émigrant. D'où la nécessité d'identifier soigneusement et de classer à part tous ces cas. On notera au passage que cette identification n'est possible que grâce à un usage intensif des registres d'état civil, qui sont le complément essentiel des recensements, dans toute étude de mobilité.

A première vue, le fait de ne travailler que sur des couples limite considérablement la portée de l'enquête: en 1852 en effet, on ne dénombre que 916 couples (Tableau II), sur une population totale de 5 359 personnes. En tenant compte du nombre de conjoints, plutôt que du nombre de couples, c'est 32% de la population que l'on atteint. Si on tenait compte de tous les enfants appartenant à ces couples, c'est 84% de la population qu'on rejoindrait. On ne peut cependant pas présumer que tous les enfants restent dans leur famille pendant toute la période 1852-1861: à partir du moment où ils atteignent 15 ans, et parfois même avant, ils sont susceptibles de se déplacer sans leurs parents. Dès lors, la proportion de la population qu'on atteint, en travaillant seulement sur les couples tels que définis, se situe en-deçà de 84%, mais largement au-dessus de 32%. Nous reviendrons plus loin sur ces proportions.

Tableau II

REPRÉSENTATIVITÉ DE LA POPULATION DES COUPLES EN 1852

1		2	3	4	5	6	7
<i>Nombre de couples</i>		<i>Nombre total de conjoints</i>	<i>Nombre d'enfants de ces couples</i>	2 + 3	<i>Population totale 1852</i>	2/5	4/5
<i>Mariés</i>	<i>Veufs(ves)</i>						
812	104	1 728	2 788	4 516	5 359	32%	84%

Il n'est pas sans intérêt d'examiner l'influence des mutations nominatives sur la méthode de jumelage par couple. En effet, en travaillant sur deux personnes au lieu d'une, on rend encore plus rare la coïncidence parfaite des éléments nominatifs, puisque l'on exige non plus deux, mais quatre paires d'éléments identiques (nom et prénom, de l'homme et de la femme), ce qui en principe devrait rendre le jumelage plus difficile. Cependant, en considérant les deux membres du couple, et leurs enfants le cas échéant, on se dote d'une série de renseignements qui permettent de jumeler très sûrement, même lorsque les mentions nominatives diffèrent.

Tableau III

IDENTITÉ DES ÉLÉMENTS NOMINATIFS CHEZ LES COUPLES
DANS DEUX LISTES DE RECENSEMENT, 1852 ET 1861

FEMMES		HOMMES									
		<i>Éléments nominatifs identiques dans les deux listes</i>									
		<i>Nom et prénom</i>		<i>Nom seulement</i>		<i>Prénom seulement</i>		<i>Aucun</i>		<i>Ensemble</i>	
		<i>N.A.</i>	<i>%</i>	<i>N.A.</i>	<i>%</i>	<i>N.A.</i>	<i>%</i>	<i>N.A.</i>	<i>%</i>	<i>N.A.</i>	<i>%</i>
<i>Nom et prénom</i>		295	48,0	55	8,9	18	2,9	14	2,3	382	62,1
<i>Nom seulement</i>		96	15,6	10	1,6	6	1,0	4	0,6	116	18,9
<i>Prénom seulement</i>		56	9,1	13	2,1	3	0,5	3	0,5	75	12,2
<i>Aucun</i>		31	5,0	5	0,9	6	1,0	0	0,0	42	6,8
<i>Ensemble</i>		478	77,7	83	13,5	33	5,4	21	3,4	615	100,0

Le Tableau III illustre bien l'importance de ce qu'on pourrait appeler les éléments contextuels dans les décisions de jumelage des couples⁹: si on exigeait que les noms et prénoms de deux conjoints soient parfaitement identiques, on effectuerait seulement 48% des jumelages possibles. En prenant appui sur des variables contextuelles telles que les âges des conjoints, les noms des enfants et leurs âges, on parvient à doubler l'efficacité du jumelage et à rapprocher des mentions de couple dont les éléments nominatifs sont discordants. Parmi les variables contextuelles, la résidence et la profession ne sont jamais utilisées.

Concrètement, les opérations de jumelage ont été menées de la façon suivante: à partir des données enregistrées sur bandes magnétiques on a produit deux listes: l'une qui énumérait tous les couples de 1852 et de 1861, par ordre alphabétique du nom de l'époux¹⁰, et l'autre, qui donnait la même information, mais classée par ordre alphabétique du nom de l'épouse (voir l'extrait de liste en Annexe I). C'est la première liste qui a été utilisée comme liste de travail, la deuxième ne servant qu'à des fins de contrôle.

Pour chacun des 916 couples présents en 1852, on a vérifié s'il était encore présent en 1861, et, s'il ne l'était plus, quelle était la cause de l'absence. La démarche peut être résumée ainsi¹¹:

⁹ Les données du Tableau III proviennent du jumelage des couples de 1852 et 1861. Seuls les couples composés de deux conjoints ont été pris en considération. La comparaison des mentions nominatives une à une a été faite par l'ordinateur.

¹⁰ Le classement des mentions de couple a été fait dans l'ordre suivant: nom de l'époux, prénom, nom de l'épouse, prénom.

¹¹ A l'avenir, pour alléger le texte, on référera à chacune des étapes par son numéro d'ordre (en chiffres romains, à gauche).

- I Existe-t-il une mention de couple strictement identique? S'il y en a une, elle est nécessairement placée à côté de la mention de 1852, en raison de l'ordre alphabétique.

OUI Jumelage

NON

- II Y a-t-il une mention comportant une ou des variations sur le nom et/ou le prénom de la conjointe, tout en étant compatible avec la mention de 1852?

OUI Jumelage

NON

- III Y a-t-il une mention, compatible avec la mention de 1852, mais comportant une ou des variations sur le nom et/ou le prénom du conjoint?

OUI Jumelage

NON

- IV Les patronymes de la mention de 1852 sont-ils susceptibles d'être des parties de patronymes composés? Par exemple, un Gauthier-Larouche peut être appelé Gauthier en 1852 et Larouche en 1861. Vérification des deux parties du patronyme¹².

OUI Jumelage

NON

- V L'un des deux conjoints est-il devenu veuf, sans se remarier, entre 1852 et 1861? Si oui, la mention apparaîtra en début de la liste 1 (cas des veuves) ou de la liste 2 (cas des veufs).

OUI Jumelage

Si, au terme de ces cinq premières étapes, on n'avait toujours pas trouvé de mention jumelable, la mention de départ était alors comparée aux listes d'actes de l'état civil, afin de vérifier s'il y avait eu décès des deux conjoints, ou décès de l'un et remariage de l'autre.

- VI Les membres du couple sont-ils décédés pendant la période 1852-1861?

OUI Décès

NON

- VII Si un seul des membres du couple est décédé entre 1852 et 1861, le survivant s'est-il remarié?

OUI Remariage

C'est seulement au terme de ces sept étapes qu'un couple était considéré comme ayant probablement émigré. Il demeurerait cependant possible que certains de ces émigrants ne le soient pas véritablement. En effet, si on ne peut retrouver, en 1861, un couple de 1852, ce peut être simplement parce que le recensement de 1861 est incomplet. Par ailleurs, ce peut être aussi parce que le couple est décédé, mais que les décès n'ont pas été enregistrés. Il y a donc à la fois un risque de sous-dénombrement et un risque de sous-enregistrement des décès. Ce dernier risque, probablement assez faible, puisqu'il suppose le non enregistrement du décès des deux conjoints pendant la période intercensitaire, n'a pu être évalué. Quant au sous-dénombrement, on en a tenu compte de la façon suivante:

¹² Ces patronymes composés ont été relevés à partir du dépouillement des registres d'état civil pour la période 1842-1911. On en a constitué une liste comptant cent cinq cas.

- VIII La mention de couple non retrouvée en 1861 apparaît-elle au recensement de 1871? L'apparition d'un couple de 1852 en 1871 ne peut être automatiquement interprétée comme preuve de sous-dénombrement en 1861, mais, si le phénomène est assez fréquent, il constitue un bon indice. OUI Faux émigrants
- IX La mention de couple apparaît-elle dans l'état civil postérieurement à 1861? L'apparition du couple, après 1861, à l'occasion d'un baptême ou d'un décès, ne constitue pas une preuve absolue de sous-dénombrement en 1861, mais c'est un indice sérieux. OUI Faux émigrants

Ces deux dernières vérifications appellent un certain nombre de remarques: 1) le qualificatif de « faux émigrants » doit être interprété en tenant compte du contexte. On sait que parmi ces couples, il y a sans doute quelques émigrants temporaires, mais comme on n'a pas les moyens de les distinguer, on les considère tous comme de faux émigrants; 2) l'étape n° IX ne constitue pas un test exhaustif, en ce sens que les couples trop vieux pour procréer et les couples sans enfants n'apparaîtront pas dans l'état civil, sauf si au moins l'un des deux conjoints décède. Les autres couples, cependant, sont susceptibles d'apparaître fréquemment à l'état civil, à l'occasion des naissances et des décès d'enfants. Comme ces couples sont les plus nombreux, la vérification demeure très utile.

Telle est donc la méthode qui a été suivie lors de cette étude sur la mobilité au Saguenay. Elle aboutit à une répartition de la population de départ, en l'occurrence les 916 couples de 1852, en cinq sous-groupes: 1) les couples *jumelés*; 2) les couples *décédés*; 3) les couples où le décès d'un des conjoints a été suivi d'un *remariage* pendant la période intercensitaire; 4) les « *faux-émigrants* »; 5) le résidu constitue le groupe des *émigrants*, c'est-à-dire ceux qui ont quitté la région du Saguenay entre 1852 et 1861.

II. — MESURE DE L'ÉMIGRATION EXTRA-RÉGIONALE

Il est maintenant possible de mesurer l'ampleur de l'émigration extra-régionale, et, moyennant certains ajustements et certaines évaluations, de comparer les résultats obtenus à ceux d'autres études de même nature. Concrètement, on procédera en trois temps: on montrera d'abord comment, et avec quels résultats, les méthodes exposées ci-dessus ont été appliquées. Etant donné que ces résultats concerneront des *couples*, et non des individus, on tentera de les extrapoler à l'ensemble de la population, ce qui rendra plus facile, dans un troisième temps, la comparaison avec les résultats obtenus par d'autres chercheurs.

1. Les résultats

Des 916 couples existant en 1852, 648 ont été retrouvés sur la liste du recensement de 1861, soit une proportion de 71%. Il n'est pas sans intérêt d'examiner comment se répartissent ces 648 couples, par rapport aux cinq premières étapes de jumelage:

Nombre total de couples en 1862:	916	100%
Étape I:	295	32%
Étape II, III, IV:	320	35%
Étape V:	33	4%
Total, étapes I à V:	648	71%

Il restait donc, au terme des cinq premières étapes, 268 couples, soit 29% de l'ensemble. Ces couples ont été jumelés aux listes de décès et de mariage de la période 1852-1861, avec les résultats suivants :

Étape VI (décès):	19 couples	2%
Étape VII (remariage):	33 couples	4%
Total, étapes VI et VII:	52 couples	6%

Ce sont donc 52 couples qui sortent d'observation à la suite de décès ou de remariages, laissant un résidu de 216 couples dont on pouvait penser qu'ils avaient quitté la région entre 1852 et 1861.

Dans un contexte de parfait enregistrement des événements d'état civil et de dénombrement exhaustif de la population, on aurait pu arrêter là les vérifications et considérer que les 216 couples étaient effectivement des émigrants. Cependant, le recensement de 1861 ayant la réputation — bien méritée — d'être incomplet, on a procédé aux deux dernières vérifications, et on a ainsi pu retrouver 35 couples de « faux-émigrants »:

Étape VIII:	23 couples	3%
Étape IX:	19 couples	2%
Moins: 7 couples retrouvés à l'étape IX et qui l'avaient déjà été à l'étape VIII:	- 7 couples	- 1%
Total des étapes VIII et IX:	35 couples	4%

Le tableau IV résume l'essentiel des opérations de jumelage et de vérification des sorties d'observation. Il illustre bien l'importance des étapes VI à IX, au terme desquelles le nombre de couples « non retrouvés » a été réduit de 268 à 181, une diminution de 32%. En d'autres mots, faute d'effectuer la vérification des décès, des remariages et du sous-enregistrement, le nombre de couples d'émigrants aurait été de 268, soit 29% de la population de départ. Négliger ces vérifications, c'est s'exposer à une surestimation de l'émigration de l'ordre de 10%.

Tableau IV

RÉSULTATS DU JUMELAGE DES COUPLES DE 1852 ET 1861

<i>Destin des couples de 1852</i>	<i>Nombre de cas</i>	<i>%</i>
Population de départ	916	100,0
dont: Présents sur la liste du recensement 1861	648	70,7
Décédés (décès des deux conjoints)	19	2,1
Remariés suite au décès d'un conjoint	33	3,6
Retrouvés dans l'état civil et/ou en 1871	35	3,8
Emigrés	181	19,8

On doit noter cependant que malgré toutes les précautions qui ont été prises, cette mesure de l'émigration des couples est susceptible de pécher à la fois par défaut et par excès. Par excès, dans la mesure où on ne contrôle pas le sous-enregistrement des décès. S'il y avait réellement un tel sous-enregistrement, il viendrait gonfler le chiffre des Décédés (voir Tableau IV), et diminuerait d'autant le nombre

d'émigrants. Notre mesure pêche aussi par défaut, en ce sens que, parmi les 35 couples retrouvés dans l'état civil ou au recensement de 1871, il peut y avoir eu un certain nombre de cas de migrations temporaires.

Quoi qu'il en soit, c'est la mesure la plus précise à laquelle on puisse parvenir dans l'état actuel des sources. On peut donc dire qu'entre 1852 et 1861, un couple sur cinq quitte la région du Saguenay. S'agissant de couples, donc, en réalité, de familles formées, et non de célibataires, ce chiffre paraît assez élevé. Néanmoins, l'interprétation d'un tel chiffre ne va pas sans difficulté, car la plupart des études sur le sujet montrent que la mobilité des jeunes adultes célibataires est plus forte que celle des familles; on ne peut donc extrapoler les chiffres obtenus sur l'émigration des couples, même s'ils sont très sûrs, à l'ensemble de la population, et la comparaison de nos résultats avec ceux d'autres études qui portent sur l'ensemble de la population devient alors assez problématique; c'est pourquoi nous avons tenté, à partir des données disponibles, une évaluation des proportions minimum et maximum d'émigrants dans l'ensemble de la population.

2. *Extrapolation à l'ensemble de la population*

Les calculs qui vont suivre portent sur trois populations, dont il importe de préciser les effectifs:

- A. *Les couples.* Des 916 couples qui formaient la population de départ, on a soustrait les 35 couples qualifiés de « faux-émigrants » (Tableau IV): en effet, ces 35 couples, normalement, appartiennent à la population des couples jumelés, en ce sens qu'ils n'ont pas quitté la région pendant la période considérée; on ne peut cependant les intégrer sans difficultés aux 648 couples jumelés, car on n'a aucune information sur la structure du couple ou de la famille en 1861: on sait qu'elle existe, sans plus. Désormais, donc, ces 35 couples et leurs enfants seront éliminés des calculs concernant les couples et leurs enfants. On les conservera toutefois dans la population totale de 1852. La population des couples est ainsi composée, au total, de 881 couples, dont 181 couples d'émigrants. Ces couples comptent 1 662 individus au total, dont 310 composent les 181 couples d'émigrants.
- B. *Les enfants.* Les 881 couples de 1852 ont au total 2 698 enfants. Quelle hypothèse retenir à propos de la mobilité de ces enfants? Si l'on peut sans risque associer le destin des plus jeunes à celui des parents, il en va autrement pour les autres, dont le degré d'autonomie peut devenir, au cours de la période inter-censitaire, s'il ne l'est pas déjà en 1852, suffisant pour justifier un traitement particulier. Il faut donc tracer une ligne de démarcation quelque part, ce qui a été fait de la façon suivante: ont été « laissés » avec les parents tous les enfants qui sont demeurés au dessous de quinze ans pendant la majeure partie de la période séparant les deux recensements, ce qui revient à fixer la limite à onze ans exactement et à considérer que 2 097 des 2 698 enfants ont suivi leurs parents. Nous sous-estimerons ainsi la mobilité des plus âgés de ce groupe, mais en revanche, celle des onze ans et plus se trouvera surestimée dans une proportion probablement analogue. Par ailleurs, étant donné que nous ne disposons pas de chiffres sur le nombre d'enfants de dix ans et moins selon le destin de la famille, nous supposerons que ce nombre est sensiblement le même dans toutes les familles, soit $2\ 097 \div 881 = 2,38$ enfants.

C. *Les individus isolés.* Quiconque ne fait pas partie, en 1852, d'un couple tel que défini ci-dessus, et n'est pas uni par un lien de filiation à ces couples, est ici considéré comme individu isolé. Il s'agit donc d'un groupe résiduel, formé pour l'essentiel de ménages de célibataires, mais comprenant aussi toutes les autres personnes qui ne peuvent être classées en A ou en B. Au total, ce sont 999 personnes qu'on retrouve dans cette catégorie.

L'ensemble de ces chiffres est résumé au Tableau V. A partir de ces données, on a tenté une évaluation de l'émigration pour l'ensemble de la population. Il s'agit en fait de calculer trois hypothèses à l'intérieur desquelles on fait varier l'émigration: une hypothèse forte et une hypothèse faible, qui constituent les maximum et minimum théoriquement possibles, et une hypothèse moyenne qui se situe à l'intérieur de la fourchette ainsi définie.

Tableau V

COMPOSITION DE LA POPULATION DE 1852

Couples, après élimination des faux-émigrants	881
Nombre de personnes composant ces couples	1 662
Couples d'émigrants	181
Nombre de personnes composant ces couples	310
Nombre total d'enfants des couples	2 698
Nombre d'enfants de dix ans et moins	2 097
Nombre d'enfants de plus de dix ans	601
Nombre d'individus isolés	999
Population totale, 1852	5 359

Hypothèse 1: on suppose que tous les enfants de plus de dix ans et tous les individus isolés quittent la région entre 1852 et 1861, et s'ajoutent donc aux familles d'émigrants déjà identifiées. On ne tient pas compte, dans ce cas, des sorties d'observation par décès ou remariage. Il s'agit donc d'une hypothèse forte, bien au-dessus de la réalité. Le calcul de cette hypothèse donne les résultats suivants:

a) Nombre d'individus dans les couples d'émigrants:	310
b) Nombre d'enfants de dix ans et moins appartenant à ces couples: 181 couples, à 2,38 enfants par couple, en moyenne:	431
c) Nombre d'enfants de plus de dix ans dans l'ensemble de la population:	601
d) Nombre d'individus isolés dans l'ensemble de la population:	999
Total	<u>2 341</u>
e) Proportion d'émigrants ($2\ 341 \div 5\ 359$):	44%

Il s'agit là, on l'a dit, d'une hypothèse forte: si on tenait compte des sorties d'observation chez les dix ans et plus, en supposant le même taux de sortie que pour la population des couples, soit 10%, la proportion d'émigrants serait réduite à 41%¹³.

¹³ Les chiffres seraient alors: a) 310; b) 431; c) 601 moins 60 = 541; d) 999 moins 100 = 899. Au total, 2 181 personnes, soit 40,7% de l'ensemble de la population.

Hypothèse 2: on suppose alors que les enfants de plus de dix ans et les individus isolés émigrent dans la même proportion que les couples et leurs enfants de dix ans et moins. Sachant que les 310 émigrants représentent 18,7% des 1 662 personnes composant l'ensemble des couples, le calcul de cette hypothèse est le suivant:

a) et b) sont identiques, soit 310 + 431:	741
c) 18,7% des 601 enfants de plus de dix ans:	112
d) 18,7% des 999 individus isolés:	187
	<hr/>
Total:	1 040
e) Proportion d'émigrants (1 040 ÷ 5359):	19%

Il s'agit là, évidemment d'une hypothèse conservatrice, qui suppose que la mobilité est la même pour toutes les catégories d'individus.

Hypothèse 3: une troisième hypothèse consiste à estimer la mobilité des enfants de plus de dix ans et des individus isolés au double de la mobilité des couples. Les résultats sont alors:

a) et b) identiques, soit:	741
c) 37,4% des 601 enfants de plus de dix ans:	225
d) 37,4% des 999 individus isolés:	374
	<hr/>
Total:	1 340
e) Proportion d'émigrants (1 040 ÷ 5 359):	25%

Ces estimations, il va sans dire, laissent subsister une part d'incertitude; elles ont cependant le grand avantage de fixer des limites, à l'intérieur desquelles se trouve la véritable mesure. Ainsi, on peut affirmer sans crainte de se tromper que c'est au maximum 41% de la population qui pourrait avoir émigré entre 1852 et 1861, et plus vraisemblablement environ 25%. Un tel chiffre, on le verra plus loin, paraît sans doute faible, comparativement aux chiffres obtenus dans les études canadiennes et américaines: il signifie quand même, et il importe de le souligner, qu'une personne sur quatre environ quittera le Saguenay entre 1852 et 1861, et cela, quelques années à peine après que cette région eût été ouverte à la colonisation. Le régime démographique excédentaire de la population saguenayenne ne saurait à lui seul expliquer cette émigration: il faudrait, pour cela, supposer une quelconque saturation des terroirs régionaux, ou un problème d'épuisement des ressources. Sans doute est-ce du côté des difficultés d'adaptation à un milieu qui, en 1850-1860, n'avait rien d'accueillant, qu'il faut chercher l'explication de ce mouvement d'émigration¹⁴.

¹⁴ Sous le terme plutôt vague de « difficulté d'adaptation », nous regroupons aussi bien des facteurs d'ordre psychologique et familial (par exemple, l'éloignement du réseau de parenté) que des facteurs plus concrets, tels que la plus ou moins grande accessibilité aux terres, l'absence de voies de communication, l'éloignement des marchés, etc.

3. *Evaluation et comparaison*

Les résultats obtenus sur la mobilité au Saguenay ne prennent tout leur relief qu'en comparaison avec la mobilité observée dans d'autres régions. Si on ne dispose pas, pour le Québec, d'études comparables, l'Ontario par contre fournit deux points de repère intéressants: le comté de Peel, un milieu rural de peuplement encore récent en 1850-1860, et la ville de Hamilton. Aux Etats-Unis, de très nombreuses études ont été réalisées, mais elles portent le plus souvent sur des villes et utilisent des sources différentes. On en dira cependant un mot, bien que l'essentiel de la comparaison porte sur les études de David Gagan sur Peel¹⁵ et de Michael Katz sur Hamilton¹⁶.

Dans le cadre du « Peel County History Project », Gagan a étudié la mobilité géographique et sociale à partir du jumelage des trois listes nominatives des recensements de 1852, 1861 et 1871. Afin de déterminer jusqu'à quel point les résultats saguenayens sont comparables à ceux du comté de Peel, on doit d'abord exposer et commenter les caractéristiques principales de la méthode utilisée par Gagan.

1. Le jumelage ne porte que sur les chefs de ménage (*householders*), et laisse de côté les célibataires âgés de 18 à 25 ans.
2. On suppose que le destin de la famille est le même que celui du chef du ménage; par conséquent, les résultats sont présentés comme s'appliquant à toute la population, en soulignant toutefois que l'absence des 18-25 ans amène une sous-estimation de la mobilité.
3. Cette sous-estimation serait compensée par le fait qu'on ne tient pas compte de la mortalité¹⁷.
4. Le jumelage des données nominatives a été exécuté automatiquement, à partir des programmes mis au point par Ian Winchester pour le projet de Katz¹⁸. Dans un court appendice, Gagan résume l'essentiel des conditions exigées pour qu'un jumelage soit effectué¹⁹:
 - a) traitement des noms au code SOUNDEX; pour que le jumelage de deux mentions soit accepté, le code SOUNDEX du nom de famille doit être identique;
 - b) les prénoms, qui ne sont pas traités au code SOUNDEX, doivent être identiques;

¹⁵ D. GAGAN, «Geographical and Social Mobility in Nineteenth-Century Ontario: A Microstudy », *Canadian Review of Sociology and Anthropology*, vol. XIII, n° 2 (May 1976), pp. 152-64.

¹⁶ M. KATZ, *The People of Hamilton, Canada West. Family and Class in a Mid-Nineteenth-Century City*, Cambridge, Mass., Harvard University Press, 1975.

¹⁷ GAGAN, «Geographic and Social Mobility... », p. 156.

¹⁸ I. WINCHESTER, « The Linkage of Historical Records by Man and Computer; Techniques and Problems », *Journal of Interdisciplinary History*, vol. I, n° 1 (1970-1971), pp. 107-24. Une analyse aussi précise que les textes nous la permettaient, nous incite à croire que le programme de jumelage utilisé par Gagan diffère quelque peu de celui utilisé par Katz: le traitement des âges, par exemple, est assez différent chez les deux auteurs.

¹⁹ GAGAN, « Geographical and Social Mobility... », pp. 162-63.

- c) les autres variables considérées sont le lieu de naissance, la religion, l'âge du chef de ménage et l'âge de son conjoint. Ces variables doivent correspondre, dans les deux mentions à jumeler, dans une certaine mesure: cette mesure prend la forme d'une pondération accordée à chacune des variables; l'addition des divers poids doit donner un score en deça duquel les jumelages sont refusés;
- d) en ce qui concerne les âges, les écarts tolérés sont assez restreints. En effet, d'un recensement décennal à un autre, l'âge peut normalement varier de + 9 ou + 10, selon la date à laquelle le deuxième recensement est effectué. Or, dans les programmes de Gagan, on permet à l'âge de varier de + 7 à + 12 seulement.

La mobilité géographique observée par Gagan dans le comté de Peel est extraordinairement élevée: en l'espace de vingt ans à peine, soit de 1852 à 1871, 94,9% de la population de Peel émigre. Pendant les neuf ans correspondant à notre période (1852-1861), 62,5% des habitants quittent le comté, et la situation empire entre 1861 et 1871, où 69,8% de la population fait de même²⁰.

Il ne saurait être question de nous livrer, dans le cadre limité de cet article, à une critique minutieuse de ces chiffres. Il nous semble néanmoins important de formuler quelques remarques d'ordre méthodologique sur les techniques de jumelage utilisées, ces techniques ayant un impact direct et majeur sur les résultats. D'une manière générale, on peut dire que les méthodes de jumelage utilisées par Gagan paraissent trop rigides, et ne tiennent pas suffisamment compte de l'état très imparfait des sources du 19^e siècle²¹.

Par exemple, le fait d'exiger que les prénoms des chefs de ménage soient identiques d'une mention à l'autre pour que celles-ci puissent être jumelées constitue une règle beaucoup trop restrictive. Au Saguenay, on a pu constater que chez les hommes, 17% des jumelages effectués comptent une variation sur les prénoms: chez les femmes cette proportion passe à 26%. Cette grande volatilité des prénoms n'est d'ailleurs pas particulière au Québec: elle a été constatée aussi bien en Ontario qu'au Texas²². Etant donné que Gagan ne traite pas les prénoms au code SOUNDEX, mais exige leur identité, lettre par lettre, il est clair que cette règle lui fait manquer des jumelages. Appliquée au Saguenay, une règle aussi restrictive aurait eu pour résultat de faire rejeter près de 20% de jumelages sûrs.

Par ailleurs, sur la question des âges, un double problème se pose. Problème d'interprétation tout d'abord: lorsque Gagan écrit: « In census linkage, for example, the individual must 'age' no more than twelve and not less than seven years during the census interval »²³, est-ce qu'on doit comprendre que tout écart situé hors de ces limites amène un rejet du jumelage, ou doit-on plutôt comprendre qu'en

²⁰ *Ibid.*, p. 156.

²¹ On trouvera dans KELLY, «Linking Nineteenth-Century Manuscript... », pp. 73-74, une brève mais excellente discussion du problème de la qualité des sources du 19^e siècle, en particulier les recensements fédéraux des Etats-Unis.

²² WINCHESTER, «The Linkage of Historical Records... », pp. 113-14. BARR, «Occupational and geographic mobility in San Antonio... », p. 398, n. 8.

²³ GAGAN, « Geographical and Social Mobility... », p. 162.

dehors de ces limites, un poids négatif est attribué à la variable « âge »? Le texte laisse à penser que le jumelage est alors rejeté. Quelle que soit la réponse donnée à ce problème d'interprétation, le problème de fond demeure le même: s'agissant d'âges déclarés, au milieu du 19^e siècle, une fourchette qui autorise une variation de deux ans seulement de part et d'autre de l'âge normal est beaucoup trop étroite. En examinant ce qui se produisait, au Saguenay entre 1852 et 1861, on a pu constater que les écarts d'âge couvraient un registre beaucoup plus étendu²⁴. Le Tableau VI, qui résume l'essentiel de notre enquête sur ce point, montre qu'en imposant la même règle que Gagan, on aurait refusé 23% de jumelages sûrs.

Tableau VI

ÉCARTS, EN ANNÉES, ENTRE LES ÂGES DÉCLARÉS
EN 1852 ET EN 1861, ENSEMBLE DES ÂGES

Écart d'âge (en années)	Nombre de personnes	
	N.A.	%
3 et moins	74	6
4	22	2
5	18	1
6	35	3
7	83	7
8	192	15
9	293	23
10	224	17
11	127	10
12	62	5
13	37	3
14	25	2
15 et plus	71	6
Ensemble	1 263	100

Là encore, la situation n'est pas particulière au Québec: Katz et Tiller ont montré, dans leur jumelage du recensement de 1861 au rôle d'évaluation de la même année, que les âges ne pouvaient être considérés comme exacts (± 1 an) que dans 51,7% des cas. En appliquant la même règle que Gagan, ils auraient effectué 69,1% de leurs jumelages²⁵. A Binghamtown et Johnstown, Kelly a constaté que les âges étaient exacts (± 1 an) dans 56 à 63% des cas seulement, et que les variations d'âges entre deux recensements, dans les jumelages sûrs, allaient jusqu'à ± 8 ans²⁶.

Un troisième aspect des méthodes utilisées par Gagan doit être souligné: l'impact de la mortalité sur la mesure de l'émigration n'a pas été évalué, l'auteur affirmant que la surestimation qui en résultait était compensée par l'absence, dans son analyse, des célibataires de 18-25 ans. Est-il besoin de rappeler qu'il s'agit là

²⁴ Ces écarts d'âge sont attribuables non seulement aux erreurs de déclaration, mais aussi aux erreurs possibles du recenseur, du releveur et du codeur. Ce type d'erreur ne peut apparaître qu'au moment du jumelage.

²⁵ M. KATZ and J. TILLER, « Record Linkage for Everyman; A Semi-Automated Process », *Historical Methods Newsletter*, vol. 5, n° 4 (1972), p. 148.

²⁶ KELLY, « Linking Nineteenth-Century Manuscript... », pp. 74-76.

de deux phénomènes qui ne sont pas comparables: la mortalité des chefs de ménage n'a rien à voir avec le risque d'émigration des célibataires de 18-25 ans, et rien n'autorise à considérer que ces deux phénomènes affectent des populations de même importance. Enfin, il aurait évidemment fallu tenir compte du sous-dénombrement au recensement de 1861.

Dès lors, comment comparer les résultats obtenus au Saguenay à ceux que Gagan obtient pour le comté de Peel? Y a-t-il eu 62,5% d'émigrants entre 1852 et 1861, ou s'agit-il d'une mixture de jumelages manqués, de personnes décédées et de véritables émigrants, dans des proportions qui restent à définir? Cette situation est d'autant plus regrettable qu'il s'agit là du seul exemple d'étude comparable portant sur une région rurale au Canada.

A défaut d'un point de comparaison rural canadien, l'étude de Katz sur la ville de Hamilton vaut d'être examinée de près, ne serait-ce que du point de vue méthodologique. La population de Hamilton, comme celle du comté de Peel, semble extrêmement mobile: le jumelage des recensements de 1852 et 1861, pour l'ensemble de la population, indiquerait que 68,7% des hommes et 66,7% des femmes présents en 1852 ont quitté la ville dix ans plus tard. Les chefs de ménage, s'ils sont moins mobiles, disparaissent quand même très vite: il en reste 40% en 1861²⁷.

Katz est l'un de ceux qui ont le mieux souligné la difficulté de comparer les résultats d'études basées sur le jumelage de mentions nominatives, du fait de la négligence des auteurs à exposer leurs méthodes. Or, Katz l'a très justement souligné, « the method of record linkage which is used directly affects the rates of persistence which are found²⁸ ».

Dans un appendice, que d'aucuns jugeront trop bref, Katz expose l'essentiel des méthodes de jumelage automatique qu'il a suivies²⁹. Il serait trop long, et sans doute peu à propos de détailler ici toutes les étapes du jumelage; on n'en retiendra que l'essentiel:

1. Tous les noms de famille sont traités au code SOUNDEX, tel que modifié par Winchester.
2. Pour être considérées comme candidates au jumelage, deux mentions doivent avoir un code SOUNDEX identique, le même sexe, un écart d'âge qui ne soit pas plus important que l'intervalle intercensitaire ± 12 ans, et la même religion, si la religion déclarée en 1852 était catholique.
3. Les mentions répondant aux conditions exposées en 2 sont alors comparées; la comparaison porte sur huit variables: lieu de naissance, religion, âge, origine ethnique, initiale du prénom, initiale du conjoint (ou des parents) et état matrimonial des femmes. Trois de ces huit variables doivent être identiques, et deux de ces trois doivent nécessairement être le lieu de naissance et/ou le nom et/ou l'âge, avec une variation égale à l'intervalle intercensitaire ± 3 ans.

²⁷ KATZ, *The People of Hamilton...*, pp. 122-23.

²⁸ *Ibid.*, p. 120.

²⁹ De regrettables erreurs d'impression, qui peuvent prêter à confusion, se sont glissées dans cet appendice. Ainsi, en page 352, on doit lire, parmi la liste des variables énumérées, RACE et non pas NAME. De même, toujours en page 352, sous la variable AGE, la ligne a) doit se lire: « If 1851 + 9 = 1861 ± 3 », et non « If 1851 + 9 = 1851 ± 3 ».

4. Les mentions candidates qui ont réussi à passer ces tests préliminaires sont ensuite comparées, variable par variable, et un score est attribué à la comparaison, selon un système de poids assez complexe: poids positif si les variables sont identiques, négatif si elles diffèrent, nul si la variable est absente. Tout score en deçà de + 4 amène un rejet du jumelage.
5. Lorsque tous les jumelages possibles ont été effectués, Katz évalue la mortalité des chefs de ménage en calculant le nombre de femmes mariées en 1852, et qui sont dites veuves en 1861.

Telles sont, pour l'essentiel, les règles suivies par Katz pour le jumelage des données nominatives à Hamilton. Ces règles, si elles sont relativement faciles à appliquer, sont difficiles à évaluer de façon précise, car elles laissent place à plusieurs alternatives; elles appellent cependant un certain nombre de remarques qui, sans mettre en cause l'essentiel des résultats obtenus Katz, invitent néanmoins à s'interroger sur l'ampleur réelle de la mobilité observée.

Tout d'abord, examinons de près la règle d'admissibilité au jumelage énoncée en 3, ci-dessus. Cette règle exige que deux mentions, pour avoir le droit d'être comparées en vue du jumelage, présentent au moins trois variables identiques (sur un total de huit), et deux de ces trois variables doivent être le lieu de naissance, le nom (orthographe originale) et l'âge additionné de l'intervalle intercensitaire ± 3 ans. Or il a été constaté, que pour une même personne, le lieu de naissance peut varier, d'un recensement à l'autre, dans 12 à 20% des cas³⁰. Katz a pallié à ce problème, dans une certaine mesure, en établissant des équivalences entre un certain nombre de lieux de naissance susceptibles d'être déclarés indifféremment (par exemple: « England » et « Wales »), mais il ne fait aucun doute qu'un certain nombre de mentions ont pu échapper à ce filet d'équivalences dont les mailles sont assez lâches.

La deuxième exigence est beaucoup plus restrictive: on demande en effet que le nom de famille, dans son orthographe originale, soit identique: or chacun sait à quel point sont variables les graphies d'un même nom de famille: c'est d'ailleurs pour cela que le code SOUNDEX a été mis au point. M. Katz lui-même a constaté, dans son jumelage semi-automatique du recensement de 1852 au rôle d'évaluation de 1851, que les noms n'étaient identiques que dans 68,1% des cas³¹; à Binghamtown et Johnstown, Kelly a observé des variations sur les noms dans 18 à 36% des cas³². Au Saguenay, on a pu constater que l'application d'une telle règle aurait amené à rejeter, au départ, 9% des hommes et 19% des femmes³³.

On objectera peut-être que, dans les cas où le nom diverge, il suffit que l'âge satisfasse aux conditions stipulées (en prenant pour acquis que le lieu de naissance

³⁰ KELLY, « Linking Nineteenth-Century Manuscript... », p. 75.

³¹ KATZ and TILLER, « Record Linkage for Everyman... », p. 148.

³² KELLY, « Linking Nineteenth-Century Manuscripts... », pp. 73-74.

³³ La friabilité des mentions nominatives n'est d'ailleurs pas propre aux recensements. Même dans les registres d'état civil, les noms et prénoms varient très fréquemment: dans le cadre du PRSS, Bouchard a mis au point un code phonétique qui réduit à une forme unique les variations orthographiques d'un même nom ou prénom. L'application de ce code à la banque de données du PRSS a permis de réduire de 18% le nombre de mentions nominatives. BOUCHARD *et al.*, « Un code de transcription phonétique... ».

coïncide) pour que les mentions soient jugées candidates au jumelage. Il faut cependant rappeler ici que le fait de ne permettre qu'une variation de ± 3 sur l'âge constitue une règle trop sévère: Katz et Tiller ont montré que 30% des mentions ne répondaient pas à ce critère³⁴, et nous avons montré, plus haut, que la variation est effectivement beaucoup plus large que ± 3 ans.

Bref, les règles appliquées par Katz ont pour effet de sélectionner, d'une façon qui paraît trop sévère, les mentions candidates au jumelage: en particulier, le fait d'exiger l'identité de deux variables sur trois, tout en sachant que ces trois variables sont sujettes à de nombreuses variations, constitue une erreur méthodologique. Il faut souligner par ailleurs que les mentions ayant passé tous les tests d'admissibilité ne sont pas jumelées pour autant: elles doivent obtenir un score de ± 4 lors de la comparaison des huit variables retenues. Il se fait donc, du début à la fin du processus, une sélection qui nous paraît trop sévère. Les mentions éliminées lors de ces sélections sont toutes considérées comme des mentions d'individus ayant émigré.

Katz effectue une certaine correction pour tenir compte de la mortalité. Le procédé qu'il utilise cependant ne rend compte que d'une partie de la mortalité d'une partie de la population, à savoir les chefs de ménage. Les données concernant l'ensemble de la population ne sont pas corrigées pour tenir compte de la mortalité, non plus que des remariages ou encore du sous-dénombrement.

Enfin, on ne doit pas oublier ce que Winchester lui-même signalait dès 1971, à savoir que le procédé de jumelage automatique mis au point pour le projet de Katz donne un rendement de 95% par rapport au jumelage manuel³⁵. C'est donc dire que l'évaluation de la mobilité risque, au départ, d'être surestimée de 5%.

Sans doute Katz est-il conscient de ces problèmes, puisqu'il prend soin de souligner que la mesure de la mobilité qu'il obtient doit être considérée comme une estimation qui surévalue certainement la mobilité. Il doute cependant que cette surévaluation soit de plus de 10%, et croit qu'elle y est nettement inférieure³⁶. Considérant les arguments qui ont été avancés ci-dessus, il nous semble au contraire que la surestimation doit certainement être bien supérieure à 10%.

En dehors du Canada, de très nombreuses études de mobilité ont été faites, surtout aux Etats-Unis, mais aussi en Europe. Aux Etats-Unis, la majorité des études publiées portent sur des milieux urbains (Boston, Omaha, San Antonio, etc.). Il existe cependant un certain nombre de travaux sur des régions rurales, à commencer par l'étude pionnière de James C. Malin sur le Kansas³⁷. Les ouvrages classiques de M. Curti sur Trempealeau³⁸ et de J. T. Lemon sur Chester et Lancaster³⁹ fournissent également des données intéressantes. Il ne saurait être question, dans le

³⁴ KATZ and TILLER, « Record Linkage for Everyman... », p. 148.

³⁵ WINCHESTER, « The Linkage of Historical Records... », p. 123.

³⁶ KATZ, *The People of Hamilton...*, p. 122.

³⁷ J. C. MALIN, «The Turnover of Farm Population in Kansas », *Kansas Historical Quarterly*, vol. 4 (November 1935), pp. 339-72.

³⁸ M. CURTI, *The Making of an American Community*, Stanford, Stanford University Press, 1969.

³⁹ J. T. LEMON, *The Best Poor Man's Country. A Geographical Study of Early Southeastern Pennsylvania*, Baltimore, Johns Hopkins University Press, 1972.

cadre d'un article, d'analyser leurs méthodes et leurs résultats en détail. Il suffira de noter que la plupart de ces auteurs font du jumelage individuel (et non du jumelage de couples) entre deux listes de recensement ou plus: ils n'effectuent à peu près aucune correction pour tenir compte des sorties d'observation, et, par ailleurs, les variables-clés dans les décisions de jumelage sont toujours les noms et les âges, mais on ne sait pas comment les problèmes de mutations nominatives ou d'écarts d'âges ont été réglés. Il en va de même pour Peter Laslett, dans son étude comparée de Clayworth, Copenhoe, Longuenesse et Hallines⁴⁰.

Dès lors, comment peut-on espérer comparer nos résultats avec ceux de ces auteurs et comment pourrait-on même comparer ces auteurs entre eux? On l'a dit, les méthodes utilisées influencent directement les résultats: ainsi, la décision de jumeler par individus au lieu de jumeler par couples nous semble avoir des incidences très lourdes sur les résultats. En jumelant par individus, on s'expose en effet au risque de faire de faux jumelages (à cause des noms très communs); pour éviter ce risque, on édicte des règles très rigides (cf. Gagan, Katz) qui, à l'inverse, forcent à éliminer des jumelages sûrs. La mobilité qu'on observe alors est évidemment très forte, mais à quoi correspondent de tels résultats? Il nous semble évident qu'en faisant porter le jumelage non plus sur une personne, mais sur un couple, voire sur toute une famille, on se dote d'une batterie d'éléments « contextuels » qui permettent de réaliser des jumelages très sûrs, même en présence de mutations nominatives nombreuses et d'écarts d'âges importants. Quoi qu'il advienne de cette suggestion, une conclusion s'impose au terme de cet exercice de comparaison: les historiens doivent apprendre à faire preuve de plus de rigueur, et ils doivent s'efforcer de suivre des méthodes standardisées. On ne peut que regretter l'absence, pour les études de mobilité, d'un manuel qui serait aussi suivi que peut l'être celui de Fleury et Henry pour l'exploitation de l'état civil ancien.

⁴⁰ P. LASLETT, «Le brassage de la population en France au XVII^e et au XVIII^e siècle », *Annales de démographie historique*, 1968, pp. 99-109.

ANNEXE I

15708778		HISTOIRE SOCIALE DE LA POPULATION DU SAGUENAY 1842-1911 ***** LISTES SEQUENTIELLES DE FAMILLES *****			PAGE 1 125
	MERE	FOSTER	CHRISTINE	405202259992020807112	
	ENFA	DESHIENS	AURELIE	406200110256205025112	
	ENFA	DESHIENS	PHILOMENE	407200110256205025112	
DESHIENS	JOSEPH	FOSTER	CHRISTINE	2288 1	4861025152002011231
	PERE	DESHIENS	JOSEPH	10110405013220160111121	
	MERE	FOSTER	CHRISTINE	102203159992026011121	
	ENFA	DESHIENS	AURELIE	103201219992036011112	
	ENFA	DESHIENS	PHILOMENE	104201119992036011112	
	ENFA	DESHIENS	LOUIS	105100919992036011112	
	ENFA	DESHIENS	AURORE	106200719992036011112	
	ENFA	DESHIENS	ADELE	107200419992036011112	
	ENFA	DESHIENS	EMMA	108200219992036011112	
DESHIENS	JOSEPH	GIRARD	ELEONORE	697 1	4851041152022 0921
	PERE	DESHIENS	JOSEPH	1011025501322010801112	
	MERE	GIRARD	LEONORE	102202559992020141112	
DESHIENS	LOUIS	BOULLIANNE	JOSEPHINE	2695 2	4861 482001010082
	PERE	DESHIENS	LOUIS	20910285013220360111121	
	MERE	BOULLIANNE	JOSEPHINE	2102023599920260111122	
	ENFA	DESHIENS	NAPOLEON	211100319992036011112	
	ENFA	DESHIENS	JOSEPH	212100119992036011112	
DESHIENS	LOUIS	LAVOIE	MARIE	258 1	4851 052021 0701
	PERE	DESHIENS	LOUIS	1011047502562010541112	
	MERE	LAVOIE	MARIE	102203859992020751112	
	ENFA	DESHIENS	ELEONORE	103200819992030751112	
	ENFA	DESHIENS	JEAN BAPTISTE	1041002102562036041112	
DESHIENS	LOUIS	LAVOIE	MARIE	1456 1	4861 052005070511
	PERE	DESHIENS	LOUIS	1011058599920160111121	
	MERE	LAVOIE	MARIE	1022045599920260111121	
	ENFA	DESHIENS	LEONORD	103201719992036011112	01
	ENFA	DESHIENS	JEAN	104101019992036011111	
	ENFA	DESHIENS	ELISABETH	105200719992036011112	
	ENFA	DESHIENS	LOUIS	106100519992036011112	
	ENFA	DESHIENS	GUILLAUME	107100319992036011112	
DESHIENS	OCTAVE	GAUDREAU	AGATHE	1078 1	4851007482023 0371
	PERE	DESHIENS	OCTAVE	1011025501322010801112	
	MERE	GODREAU	AGATHE	102202159992020801112	
	ENFA	DESHIENS	MUTIE	1031003102562036041112	01