

À la recherche de l'histoire dans les bibliothèques numériques Les leçons de *Notre mémoire en ligne*

Donald Fyson

Volume 59, numéro 1-2, été–automne 2005

URI : <https://id.erudit.org/iderudit/012721ar>

DOI : <https://doi.org/10.7202/012721ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Institut d'histoire de l'Amérique française

ISSN

0035-2357 (imprimé)

1492-1383 (numérique)

[Découvrir la revue](#)

Citer cette note

Fyson, D. (2005). À la recherche de l'histoire dans les bibliothèques numériques : les leçons de *Notre mémoire en ligne*. *Revue d'histoire de l'Amérique française*, 59(1-2), 95–113. <https://doi.org/10.7202/012721ar>

Note critique

À la recherche de l'histoire dans les bibliothèques numériques Les leçons de *Notre mémoire en ligne*¹

DONALD FYSON

Département d'histoire
Université Laval

TABLES.

Mesures des Liquides.

2 roquilles	font	1 septier ou demiard ;
2 demiards	-	1 chopine ;
2 chopines	-	1 pinte ;
2 pintes	-	1 pot ;
2 pots	-	1 gallon ;

Michel Bibaud, *L'arithmétique à l'usage des écoles élémentaires du Bas-Canada* (Montréal, 1832 ; ICMH 48734), 28.

Cet extrait apparemment banal d'un manuel scolaire du Bas-Canada accessible dans la bibliothèque numérique *Notre mémoire en ligne* (NML) me sert d'exemple dans le cours obligatoire de méthodologie de la recherche que je donne chaque automne, à l'Université Laval, aux étudiants de 3^e année en histoire. Avant de le présenter, je leur raconte une

1. Une première version anglaise de ce texte est parue dans le bulletin de Canadiana.org (autrefois l'Institut canadien de microreproductions historiques), *Facsimile*, 27 (2005) : 27-34, dont la traduction française est dans *Fac-similé*, 27 (2005) : 29-38. Mes remerciements à Canadiana.org pour la permission de reproduire l'article ici ; à Christine Veilleux pour la traduction de la version anglaise ; et à Jean-Philippe Garneau et au lecteur anonyme pour leurs commentaires.

histoire. Il y a quelques années, j'ai reçu un courriel d'un étudiant confronté au problème de l'identification de diverses mesures qu'il avait retrouvées dans des sources du début du XIX^e siècle concernant la consommation d'alcool. Il devait, en particulier, déterminer la capacité d'un gallon, d'une pinte, d'un demiard et d'un pot de rhum. Le mot « pot », surtout, soulevait un problème, car il n'arrivait pas à le retracer. N'ayant pas la réponse à cette question, j'aurais autrefois conseillé à l'étudiant de consulter un dictionnaire, d'autant plus que je répondais à son courriel depuis la maison, loin de la bibliothèque. Toutefois, puisque la requête m'a été adressée dans l'année qui a suivi le lancement de NML, j'ai simplement effectué une recherche rapide pour repérer les pages contenant les mots « pinte », « demiard » et « pot » (une recherche que je reprends en classe) et j'ai trouvé le tableau ci-dessus. Cette démarche m'a permis de lui répondre presque immédiatement, de chez moi, tout en augmentant à la fois mes propres connaissances métrologiques et mon étonnement face aux capacités des consommateurs d'alcool du début du XIX^e siècle.

Cet exemple sert non seulement à démontrer l'utilité fondamentale des bibliothèques numériques portant sur l'histoire sur le Web, ce qui est assez évident, mais également à illustrer concrètement comment les capacités de la recherche plein texte à travers des documents historiques peuvent modifier les méthodes de travail des historiens et même la nature des connaissances historiques. En plus de me permettre d'élucider la question du volume des pots de rhum depuis le confort de mon foyer, la recherche plein texte à travers des ressources telles que NML a radicalement transformé ma pratique historique.

Comment vérifier la véracité d'une date ou d'un événement obscur cité dans un article ou une thèse? Comment saisir le sens du mot « fermier » dans le Québec du XVIII^e siècle, ou les modifications dans l'usage du mot « habitant » au cours du XIX^e siècle? Des tâches qui, auparavant, auraient pris des jours, ou qui n'auraient pas été remplies, faute de temps, peuvent maintenant être effectuées en quelques minutes ou, tout au plus, quelques heures. Et, de façon plus générale, l'accès facile à une collection quasi complète des lois, journaux et appendices des législatures du Québec et du Bas-Canada, de l'Acte constitutionnel à la Confédération, permet non seulement d'éviter de nombreux séjours à la bibliothèque et d'innombrables heures à faire des photocopies, mais aide également à intégrer plus facilement ces publications gouvernementales (et, par conséquent, le discours officiel) dans mes travaux.

Au-delà de la simple appréciation des bibliothèques numériques qui offrent des moyens à portée de la main pour accéder aux sources, les historiens doivent maintenant mesurer les conséquences que peuvent avoir ces ressources sur la constitution des connaissances historiques et la nature de la recherche. Les historiens universitaires semblent avoir dépassé à la fois les premières réactions de résistance face aux nouvelles technologies et l'étape de l'émerveillement devant les possibilités offertes par de telles ressources, des attitudes qui semblaient encore assez répandues lors des sondages menés au début du projet de NML. Toutefois, les historiens universitaires ont très peu écrit sur ces questions. La transition du codex imprimé au livre électronique a, bien sûr, été largement explorée. Les spécialistes en sciences de l'information ont également commencé à analyser comment les outils numériques affectent la recherche en archives. À part les commentaires d'ordre général dans les comptes rendus concernant des ressources particulières, peu d'historiens universitaires se sont engagés dans une réflexion de fond sur l'impact des sources en ligne sur leur enseignement et leur recherche.

C'est cette réflexion que ce texte souhaite encourager, à partir d'une analyse des sources en histoire du Canada offertes par NML². Le texte se penche notamment sur deux questions fondamentales. La première est celle de la sélectivité des bibliothèques numériques. NML, ainsi que cette ressource se présente aujourd'hui, tout comme plusieurs autres grandes bibliothèques numériques, offre une collection très accessible et étendue, mais en fin de compte très sélective, qui fournit un aperçu tronqué de l'histoire canadienne. La deuxième question concerne l'impact de la recherche plein texte sur la recherche historique, une capacité qui est de plus en plus reconnue comme étant en mesure de révolutionner les manières de faire³. Explorer ces deux questions nous permettra d'aborder à la fois la promesse et les périls des collections de sources en ligne pour les historiens universitaires.

2. Je laisse de côté ici les ressources éducatives du site NML qui s'adressent davantage à l'enseignement préuniversitaire.

3. Voir, par exemple, les comptes rendus sur l'une des ressources américaines équivalentes à NML, *Early American Imprints, 1639-1800* (désignée aussi sous le nom de *Evans Digital Edition*), par Katherine Stebbins McCaffrey, Cathy N. Davidson et Jay Fliegelman dans *Common-Place*, 3,3 (2003) : <<http://www.common-place.org/vol-03/no-03/>>.

SÉLECTIVITÉ ET REPRÉSENTATION DE L'HISTOIRE DU CANADA

Il y a quelques années, au cours d'une table ronde à la Société historique du Canada, j'ai suggéré que «l'imprimé est loin de disparaître, surtout en ce qui concerne les historiens. Trouvez-moi le premier Code criminel sur Internet, ou les débats du Parlement des années 1880. Ce sont là des ressources non commercialisées et il y a peu d'espoir qu'elles soient diffusées sur Internet sans une subvention gouvernementale.» Je dois maintenant ravalier mes paroles, car tant le premier Code criminel que de nombreux débats parlementaires des années 1880 sont maintenant disponibles dans NML.

En effet, on peut dire que NML, dans sa version complète (celle offerte aux abonnés et aux écoles), représente la plus importante collection en ligne de sources imprimées portant sur l'histoire canadienne. Évidemment, d'autres collections importantes d'imprimés numérisés existent. Je pense particulièrement aux journaux et autres matériaux disponibles sur *Our Future, Our Past: The Alberta Heritage Digitization Project* ou encore aux livres de la collection numérique de la Bibliothèque nationale du Québec. Il existe également des dépôts d'archives virtuels de grande taille qui diffusent des documents d'archives, comme les manuscrits de Bibliothèque et Archives Canada et ceux des Archives nationales du Québec qui sont disponibles par le biais du site *Nouvelle-France: nouveaux horizons*, ou encore les collections de sources non textuelles, telles que l'art documentaire, les cartes ou les photographies que des institutions comme Bibliothèque et Archives Canada, la Bibliothèque nationale du Québec ou les Archives de la Colombie-Britannique proposent. NML appartient à une catégorie qui lui est propre, tout comme *Gallica* en France ou *Early American Imprints, 1639-1800* et *Making of America* aux États-Unis. Un tel accès aux nombreux *Canadiana* procure aux historiens privilégiant les documents officiels un «nirvana de l'érudition», pour reprendre la formule d'une chercheuse commentant une collection semblable en ligne⁴.

En effet, bien qu'à l'origine la numérisation des microfiches de l'Institut canadien de microreproductions historiques (ICMH) ait été un projet pilote, NML va bien au-delà. Comme l'ICMH d'autrefois, NML est l'une des principales pistes empruntées par les étudiants et autres pour accéder aux documents originaux sur l'histoire canadienne. En fait, sa disponibilité sur le Web permet de rejoindre un nombre beaucoup plus grand d'utili-

4. C. N. Davidson, «From Movable Type to Searchable Text», *Common-Place*, 3,3 (2003): <<http://www.common-place.org/vol-03/no-03/>>.

sateurs que les microfiches de l'ICMH qui se retrouvent principalement dans les grandes bibliothèques universitaires. Et pourtant, contrairement à l'ICMH, mais comme la plupart des bibliothèques numériques qui comprennent plus qu'un seul fonds ou collection, NML n'est pas une ressource exhaustive, et il n'y a aucun projet dans l'immédiat pour qu'elle le devienne. Ses 11 000 titres ne représentent qu'une fraction des anciens *Canadiana*, éclipsés comme ils le sont par les quelque 90 000 titres qui constituent la collection de l'ICMH, elle-même limitée à seulement certaines catégories de publications. L'accès à l'histoire du Canada que procure NML est donc inévitablement lié aux choix qui sous-tendent la création de la ressource. La compréhension des critères de sélection qui façonnent une collection est une démarche importante de la critique de sources et constitue, par conséquent, un exercice essentiel que doit entreprendre tout étudiant ou chercheur désireux d'aller au-delà de la simple consultation des documents individuels et d'utiliser plutôt NML comme une source de base pour une recherche substantielle. C'est également un élément clé pour comprendre la représentation de l'histoire canadienne diffusée dans le monde entier par l'entremise d'une collection comme NML. Et du point de vue de la critique des sources et du corpus (tel que je l'enseigne à mes étudiants), cette représentation des sources de l'histoire du Canada est assez tronquée. Tronquée non pas au sens négatif d'une défiguration délibérée de l'histoire du Canada - je crois que la plupart des choix qui ont été faits sont compréhensibles et justifiés. Toutefois, il faut bien reconnaître que la collection privilégie certains aspects plutôt que d'autres et est donc loin de présenter un accès neutre aux sources de l'histoire du Canada⁵.

Le site Web de NML énumère neuf collections différentes qui composent la ressource. Du point de vue de l'historien cherchant à comprendre la nature de ce corpus, il est cependant plus utile de considérer qu'il renferme trois grandes collections de documents. La première collection est composée d'un échantillon très sélectif de documents tirés de la collection originale de l'ICMH, couvrant des aspects choisis de l'histoire canadienne du XVI^e au XIX^e siècle (environ 3500 titres). La deuxième rassemble des publications gouvernementales de toutes sortes produites en Amérique du Nord britannique ou portant sur l'histoire de cet espace entre 1760 et

5. L'analyse dans cette section est basée en partie sur tous les titres contenus dans la base de données NML et dérivés des résultats obtenus lors d'une recherche sur le site de NML en janvier 2005, ainsi que ceux de la base de données de la collection des microfiches de l'ICMH provenant du catalogue de la bibliothèque de l'Université Laval.

1867. Toutefois, elle se limite surtout aux documents produits par les organismes administratifs coloniaux et supérieurs et exclut en grande partie des catégories importantes, telles que les gazettes officielles, les rapports judiciaires et les publications éphémères comme les affiches (environ 6000 titres à l'heure actuelle). Enfin, la troisième collection regroupe des publications du gouvernement fédéral de 1867 à 1900 (environ 2000 titres, mais ce chiffre est en nette progression). À la différence de la première collection, la plupart des titres des deux dernières sont des numérisations nouvelles à partir des documents originaux, bien que certains se trouvaient déjà dans la collection originale de l'ICMH. Enfin, en marge de ces trois grandes collections, il faut ajouter quelques ensembles de titres, notamment certains documents gouvernementaux concernant le Canada publiés avant 1760 et un petit nombre de publications datant d'après 1900. Dans l'ensemble, ces trois grandes collections (l'échantillon de l'ICMH, les publications gouvernementales de l'Amérique du Nord britannique et les publications du gouvernement fédéral) constituent la plus grande part du contenu de NML.

Ainsi, NML est une bibliothèque de sources plutôt hétérogènes en histoire du Canada, dont chacune des grandes collections introduit une série de critères de sélection affectant diversement la représentativité de l'ensemble. Prenons la première collection de documents, l'échantillon de l'ICMH. À la base, il est composé des quatre collections sélectionnées dans le cadre d'un projet pilote pour tester la faisabilité de la numérisation et qui correspondent à quatre aspects précis de l'histoire du Canada : Histoire des femmes canadiennes, Littérature canadienne-anglaise, Histoire du Canada français et Études autochtones. Évidemment, au départ, il n'était pas question de créer un échantillonnage représentatif des documents de l'ICMH, et encore moins des sources pour l'histoire canadienne de façon plus générale. Toutefois, ce qui, à l'origine, devait être un projet pilote a maintenant acquis un statut plus permanent et on a même rajouté trois autres petites collections, soit les *Relations des jésuites* (traduites par Thwaites), la Compagnie de la Baie d'Hudson et les anciens gouverneurs généraux du Canada.

Les titres de cet échantillon sont, en général, les seuls de NML qui sont versés directement dans les catalogues des bibliothèques universitaires et qui, par conséquent, apparaissent à l'écran lors des recherches bibliographiques effectuées le plus souvent par les étudiants. De plus, l'accès aux documents par un simple clic de la souris, souvent offert par les catalogues, signifie que ce sont là les sources d'histoire canadienne qui sont les

plus visibles et facilement accessibles. Plus important encore, cette collection constitue la plus grande partie de ce qui est offert gratuitement sur le Web (selon mes calculs, près de 80 % des pages dans la partie gratuite de NML). Par conséquent, les choix qui ont été faits pour le projet pilote affectent inévitablement la représentation de l'histoire du Canada, tant chez les étudiants que sur le Web en général. Et cette représentation de l'histoire canadienne contient des distorsions importantes.

Par exemple, l'échantillon de l'ICMH sur NML penche lourdement en faveur des documents publiés après la Confédération. Malgré des documents qui couvrent le *xvi*^e siècle jusqu'aux années 1910, près des deux tiers des textes datent du dernier tiers du *xix*^e siècle (1867-1900), alors que la période s'étalant de 1840 à 1866, qui est tout aussi importante dans l'histoire canadienne, n'en comprend que 15 %. En fait, cela représente assez bien la répartition du contenu de la collection initiale de l'ICMH, et donc vraisemblablement de la production livresque du Canada en général, mais cela signifie aussi que ceux qui recherchent des textes autres que les documents gouvernementaux sont inévitablement dirigés vers la période d'après la Confédération, perpétuant ainsi une tendance lourde dans l'étude de l'histoire du Canada depuis plusieurs années⁶. Le fort penchant pour les matériaux d'après la Confédération offerts dans la partie gratuite de NML y est tempéré par l'inclusion des journaux législatifs de la période d'avant la Confédération ; pourtant même en les rajoutant, la répartition de l'ensemble des titres disponibles gratuitement donne un peu plus de 5 % pour la période d'avant 1799, environ 35 % pour la période 1800-1866 et 60 % pour la période d'après la Confédération (dont presque tous les titres appartiennent aux dernières trente années du *XIX*^e siècle).

D'autres critères de sélection introduisent des distorsions semblables dans l'échantillonnage de l'ICMH offert dans NML. Ainsi, la décision d'inclure uniquement des textes français dans la collection « Histoire du Canada français » est bien compréhensible, étant donné l'objectif d'atteindre un équilibre linguistique, comme la présentation de la collection le mentionne. En effet, alors que seulement un cinquième des publications dans l'ensemble de la collection de l'ICMH est en français⁷, la proportion des textes français offerts dans la partie gratuite de NML atteint plutôt le tiers. En tant qu'enseignant dans une université francophone, j'apprécie

6. Allan Greer, « Canadian History: Ancient and Modern », *Canadian Historical Review*, 77,4 (1996) : 575-590.

7. Basé sur la classification selon la langue utilisée dans le catalogue de la bibliothèque de l'Université Laval.

cet état de fait. Mais cela signifie également que les textes en anglais concernant l'histoire du Québec sont grandement sous-représentés (sauf dans la mesure où ils apparaissent dans les autres collections), ce qui donne une vision nettement réduite de l'histoire du Québec. Qui plus est, les textes sur le Canada français publiés à l'extérieur du Québec semblent sous-représentés, étant donné la décision apparente de n'inclure dans l'« Histoire du Canada français » que les « publications relatives au Québec⁸ ». Certes, NML est loin d'être la seule bibliothèque numérique où la langue motive de tels choix et conduit à des distorsions importantes au niveau de la collection. Ainsi, la collection de livres et de partitions musicales dans la collection numérique de la Bibliothèque nationale du Québec, comprenant quelque 1500 publications québécoises des XIX^e et XX^e siècles, est aussi composée uniquement de titres français. Cela veut dire que la collection exclut, sur la base de leur langue, des ouvrages importants publiés au Québec, tels que l'*History of Canada* de William Smith et l'*History of the Late Province of Quebec* de Robert Christie, deux titres qui sont également absents de NML.

D'autres distorsions sont évidentes dans les deux autres grandes collections (portant sur les publications gouvernementales d'avant et après la Confédération) et cela malgré un désir évident de couvrir l'ensemble de la période de manière aussi complète que possible. Exemple le plus frappant de distorsions dans NML, les documents gouvernementaux d'après la Confédération sont essentiellement ceux produits par le gouvernement fédéral. Bien que la décision d'exclure les documents provinciaux soit parfaitement compréhensible compte tenu des contraintes de temps (et sans doute budgétaires)⁹, elle implique qu'en pratique, les sources concernant l'histoire canadienne d'après la Confédération reflètent largement les activités, les compétences et les intérêts d'un seul niveau de gouvernement. En effet, sans le vouloir, NML présente une vision de l'histoire canadienne qui passe de la période des colonies disjointes à celle de la nation unifiée. Il reste à voir comment cela affectera la recherche et l'enseignement de l'histoire basés sur les sources de NML. Ce choix signifie également que la collection des documents gouvernementaux d'après la Confédération, du moins telle qu'elle se présente maintenant, exclut largement Terre-Neuve, puisque les documents du gouvernement fédéral du Canada ont évidemment peu de choses à dire au sujet de Terre-

8. Ce sont les mots utilisés sur la page qui décrit la collection, <http://www.canadiana.org/francais/collection_hist_fr.html>.

9. <http://www.canadiana.org/francais/collection_govdocs.html>.

Neuve et que NML écarte les documents de la colonie de Terre-Neuve après 1867; les rares documents sont donc avant tout ceux qui ont été produits à Londres.

De plus, même les publications gouvernementales d'avant la Confédération excluent la plupart des productions des premières municipalités canadiennes. Pourtant, les administrations municipales constituaient souvent le niveau de gouvernement le plus intimement lié à la vie quotidienne, se chargeant d'un vaste ensemble de responsabilités depuis la police jusqu'aux œuvres de bienfaisance. Encore une fois, le raisonnement voulant que de telles publications soient extrêmement difficiles à identifier de façon systématique est fort compréhensible¹⁰, mais conduit néanmoins à l'élaboration d'une collection partielle.

Enfin, tout comme dans l'ensemble de NML, on ne trouvera que des imprimés dans les publications gouvernementales. Ici encore, il ne faut guère s'en étonner puisque NML, en tant que source complémentaire aux microfiches de l'ICMH, ne touche que les *Canadiana* imprimés. De ce choix découle tout de même une sous-représentation importante des documents gouvernementaux plus anciens, car avant la fin du XVIII^e siècle, la plupart n'étaient jamais imprimés. Prenons comme exemple les documents gouvernementaux du Régime français. Les textes de l'époque figurant dans NML sont essentiellement des documents publiés en France, qui sont beaucoup moins importants pour l'histoire de la colonie canadienne que les ordonnances, délibérations et autres documents du Conseil supérieur et de l'Intendant de la Nouvelle-France: jamais imprimés à l'époque, ces documents sont donc ignorés par NML. Quelques-uns furent publiés plus tard et certaines de ces publications se retrouvent dans NML, mais la couverture reste très inégale. Ainsi, les jugements et délibérations du Conseil supérieur sont absents, sauf ceux pour la période d'avant 1716 qui connurent une publication à la fin du XIX^e siècle. Même situation pour les ordonnances du Régime français, présentes uniquement dans des réimpressions partielles d'après la Conquête. Cela vaut aussi pour le début du Régime britannique au Québec, notamment pour les journaux des Conseils législatif et exécutif d'avant 1791 qui n'existent qu'en forme manuscrite et qui sont donc entièrement absents de NML (bien qu'on y retrouve les ordonnances, par le biais d'une publication du

10. « Publications gouvernementales fédérales, provinciales et coloniales d'avant 1920 », rapport du Comité consultatif sur les publications gouvernementales de l'ICMH (1999), <<http://web.archive.org/web/20030707232013/www.nlc-bnc.ca/cihm/icmhgov.htm>> (ce rapport ne semble plus disponible directement sur le site de Canadiana.org).

début du xx^e siècle)¹¹. C'est en effet la grande majorité des documents gouvernementaux de cette partie du xviii^e siècle québécois qui manque dans NML, ce qui fait contraste avec la collection presque complète des lois et des journaux législatifs de l'Amérique du Nord britannique du xix^e siècle.

Rien de cela ne diminue son intérêt ou son utilité, ni ne m'incite à remettre en cause le raisonnement qui sous-tend les politiques de sélection, limitées comme elles le sont par des questions de budget ou d'échéancier et par le désir évident d'assurer une juste représentation en termes de genres, races, ethnies, et ainsi de suite. L'impact de ces choix sur le produit final, considéré dans son ensemble comme une bibliothèque numérique portant sur l'histoire du Canada, est néanmoins indéniable ; et, avec sa facilité d'accès et sa portée mondiale, le contenu de NML ne peut qu'avoir un impact sur la représentation de l'histoire du Canada. Comme se demande Roy Rosenzweig par rapport à *Making of America*, une ressource américaine équivalente à NML,

La numérisation créera-t-elle un nouveau critère de recherche historique dans lequel les historiens auront recours beaucoup plus régulièrement à des œuvres qui se retrouvent en ligne et qui permettent aisément la recherche plein texte plutôt que celles qui se retrouvent dans des dépôts plus à distance ? [...] En arriverons-nous à penser un jour que, s'il ne se trouve pas sur le Web, peut-être que cela n'est jamais arrivé¹² ?

Le défi pour les étudiants et les chercheurs qui utilisent de telles collections pour tout autre chose que l'accès facile à des documents est de comprendre comment la collection a été façonnée par ces divers critères de sélection, et d'en tenir compte dans leurs analyses. Ainsi, toute analyse du vocabulaire ou de l'usage des mots par le biais d'une recherche plein texte dans NML doit tenir compte non seulement de la nature (très éli-

11. Jugements et délibérations du Conseil Souverain/Supérieur de la Nouvelle-France (Québec, 1885-1891 ; ICMH 94437-94442) ; François-Joseph Cugnet, *Extraits des édits, déclarations, ordonnances et reglemens de Sa Majesté très chrétienne : des reglemens et jugemens des gouverneurs généraux et intendans concernans la justice, et des reglemens et ordonnances de police rendues par les intendans...* (Québec, 1775 ; ICMH 34905) ; *Édits, ordonnances royaux, déclarations et arrêts du Conseil d'État du roi, concernant le Canada* (Québec, 1803 ; ICMH 40523) ; *Ordonnances des intendans et arrêts portant réglemens du Conseil supérieur de Québec...* (Québec, 1806 ; ICMH 40524) ; *Ordonnances rendues pour la province de Québec depuis l'établissement du gouvernement civil jusqu'à 1767 et Ordonnances pour la province de Québec, rendues par le gouverneur et le conseil de ladite province, de 1768 à 1791*, dans *Rapport sur les travaux relatifs aux Archives publiques*, 1913 et 1914-1915 (Ottawa, 1915 et 1917 ; ICMH 9_04403 et 9_04404).

12. Roy Rosenzweig, « The Road to Xanadu: Public and Private Pathways on the History Web », *The Journal of American History*, 88,2 (2001) : 548-579 (traduction libre).

tiste) des publications, mais aussi du fait qu'environ les deux tiers de la collection sont composés de documents gouvernementaux, faisant en sorte que les mots employés, et le discours en général, sont surtout ceux des législateurs du XIX^e siècle, dans des publications à faible diffusion. Cette situation diffère beaucoup d'une recherche dans les journaux numérisés du XIX^e siècle, avec leur diffusion plus large et leurs auteurs et lectorat plus diversifiés. Cela m'amène à aborder mon deuxième point : l'impact de la recherche plein texte sur la recherche historique.

LA RECHERCHE PLEIN TEXTE ET LA RECHERCHE HISTORIQUE

Tel que le suggère l'exemple donné au début de cet article, une des particularités les plus frappantes et utiles des bibliothèques numériques comme NML est la possibilité d'effectuer des recherches plein texte à travers les centaines de milliers de pages qu'elles contiennent, grâce aux techniques de reconnaissance optique de caractères qui convertissent des images de lettres et de mots en un texte lisible par ordinateur. Il y a, en effet, une différence fondamentale entre les bibliothèques numériques de grande taille n'offrant pas les possibilités de la recherche plein texte – comme la majeure partie des documents de *Gallica* ou les collections numériques de la Bibliothèque nationale du Québec – et celles qui offrent ce service – comme NML, *Making of America* et *Early American Imprints, 1639-1800*. Par exemple, dans ma pratique personnelle, les collections en ligne de *Gallica* et de la Bibliothèque nationale du Québec fournissent avant tout un moyen rapide d'accéder au fac-similé électronique d'un document que j'ai déjà identifié : c'est commode, sans plus. Pour une recherche plus approfondie, je me tourne vers NML ou d'autres bibliothèques numériques portant sur l'histoire canadienne, telles que *Our Roots/Nos racines*, qui offrent la recherche plein texte. Et cette prédilection n'est nullement exceptionnelle : dès les débuts, les sondages auprès des utilisateurs de NML ont démontré qu'environ 75 % d'entre eux obtenaient des résultats en utilisant les fonctions de recherche, et qu'environ 80 % de ces recherches étaient plein texte¹³. Ma manière de mener une telle recherche présente toutefois des différences tangibles avec mes démarches dans les bibliothèques ou les centres d'archives : dans mes questions, dans ma méthodologie, dans le degré de certitude découlant des résultats et

13. Wendy M. Duff et Joan M. Cherry, « Notre mémoire en ligne : résumé des conclusions de l'étude sur l'utilisation », *Fac-similé*, 22 (1999) : 19 ; « Use of Historical Documents in a Digital World » ; « Studying Digital Library Users Over Time ».

même dans ma relation avec les sources. En fait, je crois que l'intégration de la recherche plein texte à travers les documents historiques dans les pratiques de la recherche universitaire nécessite, sinon l'émergence de nouveaux paradigmes, du moins une nouvelle réflexion sur les structures de la recherche historique et de la validation scientifique.

Prenons les questions que les bibliothèques numériques avec recherche plein texte nous incitent à poser. Se fondant sur l'utilisation de mots clés spécifiques, elles s'adaptent beaucoup mieux à des sujets précis qu'à des enquêtes plus vastes. Par exemple, une recherche utilisant les mots « woman » ou « femme » dans NML relève 48 286 occurrences (« pages trouvées ») dans 3992 volumes différents, alors que « parliament » ou « parlement » relève 57 937 occurrences dans 9420 volumes¹⁴. Il est donc évident que cette façon de faire présente peu d'utilité pour obtenir une vue d'ensemble du fonctionnement des rapports de genre ou de la politique au Canada durant le XIX^e siècle. Pour cela, il ne reste que l'approche classique qui consiste à constituer un corpus, puis à le lire. Toutefois, une recherche dans NML avec « prostitute » ou « prostituée » produit des résultats beaucoup plus gérables avec 239 occurrences dans 150 titres, et même un concept d'ordre général tel que « royal prerogative » [prérogative royale] recueille 616 occurrences dans 321 documents, présentant une faisabilité certaine. Les étudiants peuvent donc aborder des questions portant tant sur les genres que sur le gouvernement en employant ces catégories plus limitées. La recherche plein texte est particulièrement utile pour la recherche biographique ou prosopographique. Ainsi, en développant une banque de données sur les magistrats du Bas-Canada au début du XIX^e siècle, j'ai énormément fait appel à NML (ainsi qu'à d'autres sources électroniques semblables), afin de rassembler des miettes et des fragments d'informations biographiques sur ces officiers relativement peu connus. Et, bien sûr, c'est un moyen idéal pour tester l'utilisation des mots (1373 occurrences pour « Briton » [Britannique], 625 pour « Celt » [Celte]) ou la première référence canadienne à un concept ou à une idéologie (selon mes recherches dans NML, en 1851 pour communisme), bien que ce dernier type de recherche soit plus difficile en raison de l'interface qui ne permet pas d'indiquer des limites temporelles ou de trier les résultats par date (essayez d'identifier le premier texte à contenir le mot Confédération!). Des participants à H-Canada, le principal forum de discussion en

14. Recherche faite en juin 2005.

ligne destiné à l'histoire du Canada¹⁵, ont d'ailleurs eu recours à des analyses lexicographiques à partir de NML pour jeter un nouvel éclairage sur des problèmes méthodologiques.

Bien sûr, tout cela aura potentiellement pour effet d'inciter les étudiants, en particulier, à choisir des sujets qui présentent une plus grande facilité au niveau de la recherche par mots clés, et d'éviter des sujets plus larges pour lesquels, par conséquent, cette technique serait plus difficile. Et cela affecte même mes propres activités de chercheur. Ainsi, dans ma recherche prosopographique informatisée, je rassemble inévitablement moins d'informations biographiques sur James Taylor ou William Baker que sur Loop Odell ou Joseph-Octave Bastien; non pas parce qu'il y a moins d'information disponible, mais simplement parce que je n'ai pas le temps de passer à travers les 643 occurrences trouvées dans NML pour Taylor ou même les 207 pour Baker, alors que les 53 réponses liées à Odell ou, mieux encore, les 11 occurrences pour Bastien sont plus faciles à gérer. Même si une interface de recherche plus raffinée permettait de réduire encore davantage le flot des résultats, en fixant des limites au niveau des dates ou des lieux, il n'en reste pas moins que la recherche historique est en train de passer d'une culture de rareté à celle de l'abondance, comme le souligne Roy Rosenzweig au sujet des sources numériques produites par notre société contemporaine¹⁶. Pas tant par rapport aux sources mises à notre disposition (il y a toujours eu des masses de sources potentielles), mais en raison plutôt de leur plus grande accessibilité. Puisque la recherche plein texte constitue l'un des principaux moyens de gérer cette abondance d'information, les connaissances historiques elles-mêmes en sont façonnées par ses capacités et ses limites.

En effet, cette technique affecte directement les méthodologies de la recherche. D'abord, comme dans le cas des études prosopographiques, cela change le rythme en privilégiant la rapidité plutôt que l'examen méthodique et laborieux d'un petit corpus fixe. Et comme l'ont souligné plusieurs comptes rendus de bibliothèques numériques permettant la recherche plein texte, cela favorise aussi la consultation plus holistique d'une variété de documents qui défie les méthodes préétablies de catégo-

15. Voir, par exemple, les messages affichés par José Igartua, 16/9/1999 et 13/5/2002, ou ma propre intervention dans un débat au sujet de l'utilisation du mot « habitant », 21/10/2002, le tout étant disponible par le biais du journal de bord des discussions sur le site H-Canada à <<http://www.h-net.org/~canada/>>.

16. « Scarcity or Abundance? Preserving the Past in a Digital Era ».

risation et d'examen des sources. Ainsi que l'a fait remarquer Jay Fliegelman à propos de *Early American Imprints, 1639-1800*, de telles recherches sont, de par leur nature, interdisciplinaires. Les recherches par mots clés, phrases ou termes rapprochés permettent de se rendre compte des contextes innombrables dans lesquels apparaît un terme clé, sans porter préjudice à ce terme, comme étant fondamentalement juridique, social, théologique, économique ou autrement spécifique à une discipline. Ainsi, une recherche offre la possibilité de constater les liens archéologiques, les discours plus étendus ou les métaphores qui sous-tendent des textes en apparence disparates¹⁷.

De fait, mon propre usage des bibliothèques numériques me porte à consulter davantage de littérature, d'historiographie du XIX^e siècle, de documents publiés à l'extérieur du Québec, et ainsi de suite, que je ne l'aurais fait si je m'étais limité à mes territoires habituels d'historiographie courante et de fonds d'archives québécois d'avant la Confédération. Par ailleurs, la recherche plein texte cible des sources présentant un intérêt jusqu'alors insoupçonné. Par exemple, la référence intrigante au communisme en 1851, émanant de la plume d'un marchand de bois francophone de La Malbaie et de celle d'un membre anglophone de la Chambre de commerce de la ville de Québec, se retrouve dans une correspondance concernant l'incorporation des pilotes du fleuve Saint-Laurent au Québec (à laquelle les deux hommes s'opposaient, argumentant que cela décourageait l'initiative individuelle), correspondance elle-même enfouie dans l'appendice NN du dixième volume des *Journaux de l'Assemblée législative de la province du Canada* ; en somme, pas le premier endroit où l'on songerait à regarder...

À un autre niveau, la recherche plein texte à travers les documents historiques exige aussi que les étudiants et les chercheurs acquièrent de nouvelles compétences méthodologiques, tout comme l'avaient demandé auparavant les premiers catalogues de bibliothèques informatisés et les banques de données bibliographiques. Au-delà des principes de base de la logique booléenne, que la plupart des étudiants ont déjà abordés en entreprenant une recherche bibliographique, et la maîtrise de l'interface de recherche qui, dans le cas de NML est bien expliquée dans la page d'aide, les périls et les pièges doivent être clairement exposés. Ainsi, les utilisateurs doivent comprendre les particularités des logiciels de recon-

17. « An MRI of Early America », *Common-Place*, 3,3 (2003) : <http://www.common-place.org/vol-03/no-03/tales/tales-1.shtml>.

naissance optique des caractères lorsqu'ils s'appliquent aux anciens imprimés : le logiciel de NML interprète la forme allongée « f » de la lettre s, souvent rencontrée jusqu'au début du XIX^e siècle, comme un f, ce qui signifie que toute recherche portant sur la police au XVIII^e siècle doit nécessairement inclure le terme « conftable » (99 occurrences). De manière plus générale, le succès d'une stratégie de recherche dépend avant tout de la formulation, laquelle est beaucoup plus complexe dans le cas des documents historiques que dans celui des banques de données bibliographiques modernes qui comportent un langage relativement plus simple. En effet, elle réclame des connaissances historiques assez considérables qui ne peuvent être acquises qu'avec la pratique. Cette stratégie peut exiger le terme exact (avant 1840, « Canadien » plutôt que « Canadien français »), mais doit aussi tenir compte des caprices de l'orthographe – pour prison et geôlier, les termes anglais *gaol* et *gaoler* plutôt que *jail* et *jailer* ou, pour utiliser un exemple tiré de mes propres recherches, « M' » ainsi que « Mc » lorsque j'effectue une recherche sur les Écossais (une omission qui, à un moment donné, m'a amené à sous-estimer la présence parlementaire d'un personnage important pour mon travail, Thomas McCord (43 occurrences), alias « Thomas M Cord » (97 occurrences) selon le moteur de recherche de NML).

Encore plus fondamentalement, la recherche plein texte dans les bibliothèques numériques, dont les moteurs de recherche sont basés sur des textes générés par la reconnaissance optique des caractères sans vérification, oblige bien des historiens à reconsidérer la primauté qu'ils accordent à la certitude empirique. Cette forme de reconnaissance optique des caractères, qui est indispensable à la majorité des grands projets de numérisation de textes anciens avec recherche plein texte financés par le secteur public (étant donné l'importance des ressources nécessaires pour la vérification manuelle des pages), exige que l'historien abandonne l'idéal des recherches précises et systématiques et qu'il accepte des résultats moins sûrs et apparemment plus sujets à l'erreur. La recherche historique empirique valorise habituellement la précision et l'exactitude et les historiens empiriques se targuent de leurs stratégies qui reposent sur une démarche méthodique. Par exemple, dans un groupe de recherche dont je fais partie, des chercheurs consultent méthodiquement les journaux sélectionnés pour y trouver des articles renfermant un certain nombre de mots clés ou de concepts déterminés par les membres du groupe. Les résultats sont très précis, étant donné la compétence des chercheurs, mais demandent aussi beaucoup de temps ; au point d'avoir recours à des échantillonnages.

Avec *Paper of Record* ou d'autres projets de numérisation de journaux, ces stratégies soulèvent de plus en plus de débats ; tout comme il devient de moins en moins opportun de procéder à ce que nous appelions auparavant un « balayage de la législation » (opération qui consiste à parcourir systématiquement les lois d'une période quelconque dans le but d'identifier les références à des termes ou des concepts spécifiques), puisque la recherche plein texte est maintenant possible à travers une grande partie de la législation du XIX^e siècle par le biais de NML. En contrepartie, toutefois, les recherches plein texte basées sur des textes générés par la reconnaissance optique des caractères sans vérification, surtout celles portant sur des documents historiques, ne sont pas particulièrement précises, une critique qui revient maintes fois dans les comptes rendus de ces ressources¹⁸.

Cette situation, combinée aux caprices de la formulation des stratégies de recherche, fait en sorte que l'historien devra simplement accepter que certaines informations pertinentes ne seront pas repérées. Et, comme corollaire, la preuve par la négative devient, en fait, une tâche épineuse – ainsi, l'absence d'une référence au communisme ou aux communistes avant 1851 dans NML n'indique en rien qu'une telle référence n'existe pas (et, en effet, une recherche menée dans *Paper of Record* relève des références dans le *Journal de Québec* à compter de 1845, dans des articles publiés à l'origine dans les journaux français ; mais également, elle retrace d'autres références encore plus anciennes, datant des années 1830, qui ne sont que des interprétations erronées par le logiciel de reconnaissance optique des caractères du terme anglais « cannot » ou du mot « commercial »!).

Donc, paradoxalement, alors que les bibliothèques numériques de documents historiques permettant la recherche plein texte encouragent la recherche empirique en raison de leur facilité d'utilisation, elles assouplissent du même coup les normes en les rendant plus « floues ». On pourrait bien sûr s'interroger sur le niveau d'exactitude de la recherche empirique classique. Mais c'est un tout autre débat...

Finalement, comme plusieurs l'ont remarqué, le recours à la recherche plein texte affecte fondamentalement le rapport entre chercheur et source en éliminant le contact avec le document et son contexte¹⁹. Puisque le

18. Par exemple, C. N. Davidson, « From Movable Type to Searchable Text », ou Tobias Higbie, « Web Site Review: Making of America <<http://moa.umdl.umich.edu/>> », *The Journal of American History*, 89,2 (2002): <http://www.historycooperative.org/journals/jah/89.2/wr_2.html>.

19. Du point de vue de l'historien, par rapport aux sources historiques, cette question est explorée à fond dans Bonnie Mak, *(Re)Defining the Page for a Digital World*, thèse de doctorat, University of Notre Dame, 2004.

résultat habituel est une page plutôt que le document dans son ensemble, cela réduit inévitablement le champ de vision des chercheurs. C'est un peu comme si la pratique pédagogique de distribuer des photocopies de pages ou de courts extraits de documents était reproduite à grande échelle et constituait le moyen privilégié d'accéder aux sources du passé. Malgré tous les efforts des développeurs de sites (entre autres par la présentation des résultats de recherche regroupés par volume), le lien à la source originale et à son contexte bibliographique ou archivistique est inévitablement affaibli. En fait, dans certains cas, le champ de vision des chercheurs peut être réduit encore davantage, au contexte immédiat du mot recherché, avec des systèmes qui mettent les résultats en surbrillance (*Paper of Record* par exemple); la tentation est alors très forte de ne lire que les phrases entourant ces mots.

Cette diminution du champ de vision contextuel, utile à certains égards, suscite néanmoins plusieurs effets potentiellement négatifs. D'abord, la consultation fragmentaire d'une source à partir des résultats d'une recherche plein texte réduit à la fois l'acquisition d'une vision plus large de l'histoire et les trouvailles inattendues qui font le bonheur des historiens, deux bienfaits qui découlent de la lecture méthodique d'une source. La recherche traditionnelle dans les journaux, qui implique la lecture rapide de centaines ou de milliers de pages, est une des meilleures façons d'initier les étudiants non seulement à la nature des journaux historiques, mais aussi à la culture historique de la période étudiée. Et combien de projets de recherche ont commencé par une découverte au hasard d'une source, tandis que l'on cherchait patiemment autre chose? Ces deux sous-produits de la recherche méthodique sont perdus, ou du moins limités, quand tout ce qui est examiné est la page ou, encore plus, le paragraphe ou la phrase dans lequel se trouve le mot clé recherché. Ainsi, les bienfaits de la recherche à travers des sources plus diversifiées sont peut-être contrebalancés par la perte de ce regard historique plus large qui dépasse la question initiale.

La consultation d'une source à partir des résultats d'une recherche plein texte nécessite aussi un effort particulier, afin de saisir toute l'importance du document consulté. Car, au lieu du mouvement naturel consistant à passer de la bibliothèque ou du dépôt d'archives au document, puis à la page, le mouvement se fait de la page au document, et de là au contexte bibliographique ou archivistique. Pour les étudiants en particulier, commencer au niveau de la page peut entraîner beaucoup de confusion. Prenons comme exemple l'une des sources les plus considérables disponibles

dans NML : les appendices des *Journaux de l'Assemblée législative du Canada-Uni* et leur suite fédérale, les *Documents de la session*. Vu leur étendue et la diversité des sujets qui y sont traités, ils sont vraiment remarquables. Par contre, ce sont aussi des documents très complexes. Destinés surtout aux législateurs et à leur personnel, ils sont composés de toute une série de rapports, de sous-rapports, de pièces jointes, et ainsi de suite, présentés par divers ministères ou autres organismes administratifs, et regroupés dans des volumes reliés (jusqu'à une douzaine et plus par année à la fin du XIX^e siècle). Comprendre le raisonnement administratif derrière l'organisation de ces volumes est une tâche ardue. Entre autres, dans les documents de la session du gouvernement fédéral à compter des années 1880, on retrouve une fascinante série de statistiques sur la criminalité au Canada, laquelle est présentée comme un sous-rapport inclus dans le rapport du ministre de l'Agriculture (puisque la cueillette des statistiques relevait alors de la responsabilité de ce ministère). Toutefois, l'accès dans NML, et c'est compréhensible, est organisé par volume plutôt que par rapport. Ainsi, quelqu'un qui trouverait par hasard cette série de statistiques sur la criminalité, peut-être par le biais d'une recherche avec les mots « drunk and disorderly » ou « ivrognerie et conduite désordonnée », ne pourrait se baser que sur le titre énigmatique « Documents de la session [vol. 13, no 7 (1880)] » pour établir le contexte du document. Après avoir d'abord lancé directement des étudiants dans une recherche à travers les documents de la session par le biais de NML, j'ai compris qu'ils ne s'y retrouvaient pas, en dépit de mes explications préalables sur la série, car ils ne pouvaient visualiser le document au complet et ne comprenaient pas comment les pages trouvées s'inséraient dans un système documentaire et administratif plus large. J'ai donc plutôt commencé par les amener à la bibliothèque afin de leur montrer, rayon après rayon, les documents physiques, des volumes de couleur jaunâtre. Cet exercice leur a permis de concevoir mentalement les textes trouvés. En d'autres termes, un détour vers les sources originelles s'est avéré nécessaire pour comprendre les sources provenant de NML.

En somme, la recherche plein texte affecte fondamentalement le processus de la recherche et la création de la connaissance historique. Certains de ces effets présentent des dangers, comme la tentation d'adopter des méthodes de recherche rapides ou de définir les sujets de recherche en fonction des capacités des moteurs de recherche ; d'autres offrent peut-être plus de liberté, comme le fait d'accepter que la recherche historique n'est pas, et n'est pas obligée d'être, une science exacte. Mais, sans l'ombre d'un

doute, ma pratique de l'histoire aujourd'hui n'est plus la même, pas plus que les connaissances historiques que je tire de NML.

CONCLUSION

Les bibliothèques numériques de sources historiques comme NML sont souvent louées avant tout pour l'accès qu'elles procurent aux documents rares qui évoquent le passé. Pour les écoles, les institutions de petite taille et les utilisateurs à l'extérieur du Canada, NML a évidemment révolutionné l'accès à ces sources. Mais les chercheurs et les étudiants dans la plupart des grandes universités canadiennes avaient déjà plusieurs de ces documents à portée de main, soit par l'intermédiaire de l'ICMH, soit dans la section des publications gouvernementales des bibliothèques universitaires. NML a certainement facilité leur approche ; sûrement aussi qu'il a ajouté de l'eau au moulin ; mais ce n'est pas là que réside la vraie révolution mise en œuvre par les bibliothèques numériques. Celle-ci tient plutôt dans la recherche plein texte à travers ces masses de sources de l'histoire du Canada ; une abondance qui implique des changements fondamentaux dans les méthodes de travail des historiens et la formation offerte aux étudiants. Toutefois, cette abondance n'est pas sans dangers. D'abord, il y a la question de la représentation du passé à travers des bibliothèques numériques obligatoirement sélectives. Au-delà des distorsions inévitables, on peut se demander si la place centrale des imprimés dans de telles ressources ne nous ramène pas aux sources traditionnelles de l'histoire canadienne, notamment pour la période d'avant le xx^e siècle. Après tout, en se concentrant sur les documents imprimés des xviii^e et xix^e siècles, NML privilégie nécessairement la voix du pouvoir, de ceux ayant accès à l'éducation et à la culture de l'imprimé. Des voix discordantes se font entendre dans les imprimés de l'époque, même dans les documents gouvernementaux, mais ces voix sont assourdies dans ces sources. Ces considérations devraient nous amener à bien réfléchir lorsque nous nous lançons dans l'exploitation de ces outils merveilleux que nul ne peut désormais se permettre d'ignorer.