

46^E CONGRÈS DE L'AAQ

Cartographier l'évolution du contenu de la revue *Archives* au moyen des techniques de fouille de textes et de bibliométrie

DOMINIC FOREST

Professeur agrégé, Université de Montréal

SABINE MAS

Professeure agrégée, Université de Montréal

VALÉRIE RIOUX

Assistante de recherche, Université de Montréal

VINCENT LARIVIÈRE

Professeur agrégé, Université de Montréal

BENOÎT MACALUSO

Professionnel de recherche, Observatoire des sciences et des technologies, Université du Québec à Montréal

INTRODUCTION

Publiée par l'Association des archivistes du Québec (AAQ) depuis 1969, la revue *Archives* contribue à faire rayonner l'expertise professionnelle et scientifique de ses membres au sein de la francophonie. À cet égard, elle rend disponible sur son site Web le contenu des articles publiés depuis 1994. Ainsi, à l'occasion du 50^e anniversaire de l'AAQ, nous avons réalisé une analyse du contenu de la revue *Archives* à partir d'une approche fondée sur des techniques de fouille de textes que nous avons complétée par une analyse bibliométrique. Ces approches permettent le traitement de grandes quantités de données afin d'en extraire automatiquement certaines caractéristiques qui renseignent tant sur l'évolution thématique de la revue que sur la place qu'occupe sa production dans l'espace archivistique scientifique et professionnel québécois et international.

Nous amorcerons cet article en présentant de manière synthétique les principaux travaux ayant un lien avec notre projet (section 1). Par la suite (section 2), nous présenterons les principes généraux de la démarche méthodologique employée, en insistant d'abord sur les processus de fouille de textes, puis sur la bibliométrie. Finalement, nous procéderons à une analyse détaillée des principaux résultats obtenus (section 3).

Notre contribution offre l'occasion de revenir sur les principales thématiques abordées dans la revue depuis les vingt dernières années en rendant compte de l'évolution des objets de recherche étudiés et, plus généralement, de la profession archivistique. Elle permet également de s'interroger sur les liens existant entre les auteurs et leurs institutions, en plus de faire le point sur l'académisation ou la professionnalisation de la revue. En se penchant sur le passé de la revue, il est possible de comprendre l'horizon vers lequel elle poursuit sa lancée.

1. CONTEXTE

Au cours des dernières décennies, Louise Gagnon-Arguin a publié trois analyses du contenu de la revue *Archives*, respectivement en 1988, 1994, et 2009-2010. L'article le plus récent, *La revue Archives à travers ses vingt dernières années d'existence, 1988-2008*, retrace les dix premières années du corpus que nous avons analysé dans le cadre de notre étude (voir section 2.1). Dans son article, Gagnon-Arguin offre

une analyse de l'évolution de l'archivistique québécoise en rappelant les faits marquants de la discipline. À travers une analyse approfondie des thématiques des articles, une recension des auteurs, des directeurs de la revue et des changements dans l'édition de la revue, elle parvient à dégager les grandes tendances de la discipline et de sa pratique et à rendre compte des changements au sein des institutions et des réseaux (Gagnon-Arguin, 2009-2010).

Au terme de son analyse, Gagnon-Arguin souligne l'importance que prend le milieu de la recherche au sein de l'ensemble des articles de la revue. De plus en plus d'actes de colloques et de congrès, dont la qualité a été évaluée par les pairs, y sont désormais publiés. Pour elle, il s'agit d'une valeur supplémentaire pour une revue qui se veut « à la fine pointe de l'actualité archivistique » (Gagnon-Arguin, 2009-2010).

De plus, Gagnon-Arguin souligne l'impact des changements dans le milieu éditorial archivistique. La revue *Archives* n'est désormais plus la seule à publier des articles faisant état de l'archivistique québécoise. Les archivistes peuvent publier et s'informer sur plusieurs plateformes informationnelles différentes. Cet aspect lui paraît être l'un des plus grands défis auxquels elle doit faire face (Gagnon-Arguin, 2009-2010).

Dans le cadre d'un mémoire de maîtrise déposé en 1989, Louis-René Dessureault s'est également intéressé à l'état du discours de la profession d'archiviste au Québec à travers une analyse de contenu de la revue *Archives*. Son analyse couvre la période des vingt premières années d'existence de la revue (1969-1988) et s'intéresse particulièrement aux citations afin de brosser le portrait scientifique de la revue. Bien que l'approche de Dessureault demeure manuelle et qu'elle ne s'attarde pas aux mêmes périodes que notre étude, celle-ci s'avère intéressante puisqu'elle nous permet de disposer de données bibliométriques comparatives. Dessureault remarque, entre autres, la faible importance des articles scientifiques au sein de la publication, l'importance grandissante de la collaboration ainsi que l'augmentation soutenue du nombre d'articles écrits par des femmes (Dessureault, 1989).

D'autres analyses se sont par ailleurs attardées à dresser le portrait des revues du milieu des sciences de l'information au Québec en comparant la revue *Archives* aux revues *Argus* et *Documentation et bibliothèques*. En 2008, Gagnon-Arguin soulève l'idée que la revue *Archives* traite de

la profession et de la discipline archivistique, alors que *Documentation et bibliothèques* et *Argus* s'intéressent aux enjeux touchant le milieu des bibliothèques, des techniciens en documentation et des bibliothécaires. Chacune des trois revues rend compte de l'évolution et des champs d'intérêts disciplinaires. Dans un autre ordre d'idées, Gagnon-Arguin fait remarquer que, toutes revues confondues, la majorité des auteurs des textes des revues provient soit de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal ou de la School of Information Studies de l'Université McGill (Gagnon-Arguin, 2008).

2. MÉTHODOLOGIE

La fouille de textes est un domaine de recherche interdisciplinaire dont l'objectif général est le développement de techniques, d'algorithmes, d'applications informatiques et de méthodologies applicatives afin d'extraire et de structurer automatiquement de nouvelles informations à partir de grands corpus de documents textuels non structurés (Forest et al., 2017). Une définition proposée par Hearst (2003) permet de cerner adéquatement les composantes caractéristiques de la fouille de textes :

La fouille de textes est la découverte [à l'aide d'outils informatiques] de nouvelles informations en extrayant différentes données provenant de plusieurs documents textuels. Un élément fondamental de ce processus réside dans les relations identifiées entre les informations extraites afin d'établir de nouveaux faits ou de nouvelles hypothèses à explorer. (Hearst, 2003, s.p., traduction libre)

La fouille de textes tire son origine du domaine de la fouille de données. Elle s'en distingue toutefois sur la base de la nature des données qu'elle traite. En effet, alors que les données traitées dans le domaine de la fouille de données sont normalement de nature numérique et structurée (bases de données, entrepôts de données, etc.), les processus de fouille de textes sont appliqués à des données textuelles (de grands corpus de documents) dont l'une des principales caractéristiques est d'être beaucoup moins structurées (Forest, Bastien, Legault-Venne, Brousseau et Lacombe, 2017).

Le traitement des documents textuels, en employant des algorithmes de fouille, est normalement réalisé en respectant une démarche méthodologique inspirée de celle que nous retrouvons au cœur

de nombreux projets de fouille de données. Cette démarche, de nature itérative, est composée de cinq principales étapes (Figure 1) : 1) la constitution du corpus, 2) le filtrage, 3) la transformation vectorielle, 4) l'application des algorithmes de fouille et 5) l'évaluation, l'interprétation et l'intégration des informations extraites.

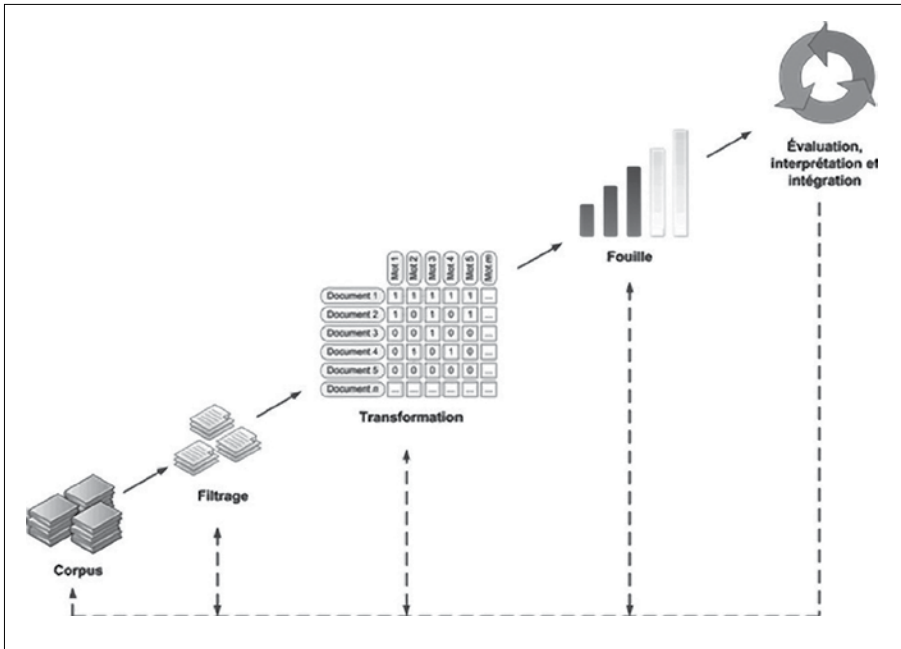


Figure 1 : La démarche générique de fouille de textes.

2.1. Constitution du corpus

La première étape de ce projet a consisté à recueillir les documents que nous souhaitons soumettre au processus d'analyse. Le corpus que nous avons analysé dans le cadre de ce projet est composé de 21 volumes, soit les volumes 26 (1994-1995) à 46 (2016). À la différence de la démarche de Louise Gagnon-Arguin, nous avons uniquement retenu les textes scientifiques et professionnels (bilans d'expériences, mémoires de l'Association des archivistes du Québec présentés au gouvernement, etc.), laissant ainsi de côté les comptes rendus de lecture et les notes de la rédaction. Le corpus ainsi constitué est composé de 238 documents.

Afin de permettre l'analyse automatisée de nos fichiers, certaines opérations de conversion et de nettoyage ont dû être préalablement réalisées. Une fois la conversion des fichiers originaux (en format pdf) effectuée vers un format de traitement de texte (.doc), un nettoyage systématique des marques de mise en pages (en-têtes, pieds de page) a dû être effectué. Les traits d'union de fin de ligne et les espaces superflues ont également été supprimés. Les notes de bas de page, dont la mise en pages originale a rendu particulièrement difficile le travail de conversion, ont été exclues du corpus. Les fichiers nettoyés ont par la suite été enregistrés en format .txt pour être facilement parcourus par les algorithmes de fouille de textes. Les bibliographies ont été nettoyées et analysées séparément. Les informations bibliographiques des articles, c'est-à-dire les titres, résumés, auteurs, volumes, numéros et dates de publications ont été exportées des fichiers initiaux pour être intégrées dans le processus d'analyse. Ainsi, le titre, le résumé et le texte des articles, considérés comme des variables dépendantes, ont été croisés avec les années, par tranches de 5 ans, et avec les auteurs, considérés comme des variables indépendantes. Le Tableau 1 représente la division du corpus en quatre périodes. À noter que la dernière période comprend les numéros des volumes 45 (2013-2014) et 46 (2016). La parution irrégulière de la revue depuis 2015 nous a contraints à ajouter le volume 46 à la dernière période de notre analyse. Le nombre d'articles du volume 46 est de six.

Période	Nombre d'articles
1995-1999	61
2000-2004	55
2005-2009	57
2010-2016	65

Tableau 1 : Répartition des articles par période.

2.2. Filtrage du lexique

Une fois les fichiers préparés et les différentes variables identifiées, nous avons procédé à l'extraction du lexique nécessaire pour les opérations de fouille de textes. Nous avons supprimé les mots fonctionnels en utilisant

un antidictionnaire. Une attention particulière a par ailleurs été accordée aux mots contenant des traits d'union ou des apostrophes. Nous avons ensuite lemmatisé le lexique du corpus ; la lemmatisation est une opération consistant à ramener les formes fléchies (conjuguées, plurielles) à des formes standards (infinitif ou singulier).

Un traitement statistique automatisé a par la suite été appliqué. La fréquence des mots a été pondérée selon la mesure *tf-idf* (*term frequency x inversed document frequency*). Cette mesure nous a permis de mettre de côté les mots fréquents apparaissant dans plus de 75 % des cas, par exemple : archive, document, information, pour ne garder que les mots les plus typiques des périodes étudiées dans le cadre de notre projet. Cependant, cette méthode possède quelques limites. Par exemple, dans un numéro thématique traitant fortement de l'évaluation des archives, le terme « évaluation » a été écarté, car trop fréquent et donc peu discriminant, au profit de termes associés à cette thématique. Les mots supprimés en raison de leur fréquence pondérée trop élevée ont toutefois été récupérés lors de l'analyse des résultats. À la suite de l'extraction et du filtrage du lexique de notre corpus, une liste de 250 mots-clés, dont la fréquence pondérée était la plus élevée, a été conservée. Celle-ci représente environ 1 % des formes présentes dans l'ensemble du corpus.

À titre informatif, le corpus initial contenait 37 929 formes, soit 37 929 mots différents, et 1 791 514 occurrences, soit le nombre total de mots du corpus. Une fois normalisé et filtré selon les méthodes mentionnées précédemment, le corpus était composé de 20 676 formes et 11 785 420 occurrences. Le Tableau 2 présente la répartition des formes et des occurrences par période analysée.

Période	Nombre de formes (mots différents)	Nombre d'occurrences (total de mots)
1995-1999	8 761	432 196
2000-2004	9 052	408 828
2005-2009	11 504	533 593
2010-2016	10 181	410 803

Tableau 2 : Fréquences des formes et des occurrences normalisées et filtrées par période.

2.3. Transformation

Des matrices de données ont été produites en croisant les 250 mots-clés retenus pour décrire les 238 documents de notre corpus en fonction des années (périodes de cinq ans) et en fonction des auteurs. Cette opération est réalisée en structurant les documents du corpus en une matrice de vecteurs dans laquelle chaque document est représenté par l'absence ou la présence pondérée de chaque mot retenu à l'étape précédente. Les données de départ, ainsi transformées en matrices, ont été soumises aux processus de fouille.

2.4. Fouille thématique

Une principale technique non supervisée de fouille de textes a été privilégiée pour cette recherche : le regroupement automatique (*clustering*). Cette technique consiste à regrouper les documents d'un corpus dans un certain nombre de regroupements sur la base d'un ou de plusieurs critères de similarité. L'objectif de cette technique est de délimiter des regroupements de documents thématiquement comparables. Ce processus de fouille de textes réalisé à l'aide de l'application Wordstat a été appliqué sur le corpus afin d'identifier en diachronie les principaux thèmes abordés et les proximités thématiques entre les auteurs de la revue.

Afin de faciliter l'analyse et l'interprétation des résultats générés par les processus de fouille de textes, nous avons opté pour une représentation des résultats sous forme de cartographies thématiques. En procédant ainsi, il nous a été possible de rendre compte de la structure des relations entre les mots-clés retenus. Il s'agit d'une démarche classique pour ce type d'analyse (Spence et Press, 2000 ; Fayyad, Grinstein et Wierse, 2001 ; Mokrane et al., 2004). Dans le cadre de nos analyses, les matrices ont été importées dans le logiciel de visualisation Netdraw. Pour parvenir à faire ressortir les relations entre les différents nœuds formant le réseau, une optimisation du nombre de liens a été nécessaire. Le réseau a par ailleurs été structuré en fonction du degré de centralité qu'un nœud entretenait avec l'ensemble du réseau. Finalement, nous avons réussi à générer des cartographies thématiques pour l'ensemble des auteurs et des périodes couvertes par notre étude. Il est à noter que la longueur des liens entre les différents nœuds n'est pas déterminante dans les figures que nous avons produites, afin de garder un souci de lisibilité textuelle.

2.5. Bibliométrie

Ayant pour origine l'élaboration des collections de périodiques dans les universités américaines au début du XX^e siècle, alors que les bibliothécaires faisaient le décompte manuellement des références citées par les chercheurs afin de déterminer les revues les plus importantes, la bibliométrie s'est principalement développée dans les années 1960 avec l'avènement des larges bases de données bibliographiques et des indices de citations, tels que la *Science Citation Index*. D'une façon générale, la bibliométrie peut être définie comme étant la mesure des documents scientifiques et de leurs différentes caractéristiques, principalement fondée sur leurs métadonnées. Ses applications sont multiples, allant de la description de l'évolution d'une discipline, comme c'est le cas dans le cadre de notre projet, à la sociologie des sciences et l'évaluation de la recherche.

Dans le cadre de cet article, les méthodes bibliométriques utilisées ont été appliquées aux métadonnées des articles de la revue *Archives* extraites à partir des textes. Plus spécifiquement, nous avons :

1. isolé les prénoms des auteures et auteurs afin de leur attribuer un genre et un statut (récemment diplômés et étudiants, professionnels, professeurs) et de mesurer la participation de différents groupes sociodémographiques aux activités de la revue ;
2. déterminé les principales institutions d'attache des contributeurs à la revue ;
3. analysé les références citées par les auteurs afin de déterminer leur nombre moyen par article ainsi que leur âge.

3. RÉSULTATS ET DISCUSSION

Cette étude, rappelons-le, visait à documenter et surtout à cartographier de manière circonscrite l'état de la discipline archivistique à travers l'analyse quantitative de la revue *Archives*. La méthodologie employée est nettement différente des statistiques descriptives combinées à l'analyse qualitative réalisées par Louise Gagnon-Arguin lors de ses trois études rétrospectives, riches et documentées, portant sur le contenu de la revue *Archives*. Nous ne pouvons ainsi comparer ces deux types d'approches et leurs résultats qui sont, par ailleurs, largement complémentaires.

Cela étant dit, les résultats découlant de l'analyse du corpus réalisée au moyen des techniques de fouille de textes et de bibliométrie sont nombreux. Dans le cadre de cet article, nous présentons quelques-uns des résultats liés à l'étude des mots-clés (relation, centralité, évolution), des auteurs (genre, affiliation, statut) et des références citées (nombre et âge moyens, principales revues citées) selon trois périodes de cinq ans (1995 à 1999, 2000 à 2004, 2005 à 2009) et une période de six ans (2010 à 2016).

D'emblée, l'examen de la Figure 2 permet de relever les mots-clés communs à trois ou quatre périodes qui se retrouvent au centre et les mots communs à deux périodes qui se retrouvent en périphérie. Les mots-clés qui se retrouvent à l'extérieur sont associés à leur période d'attache.

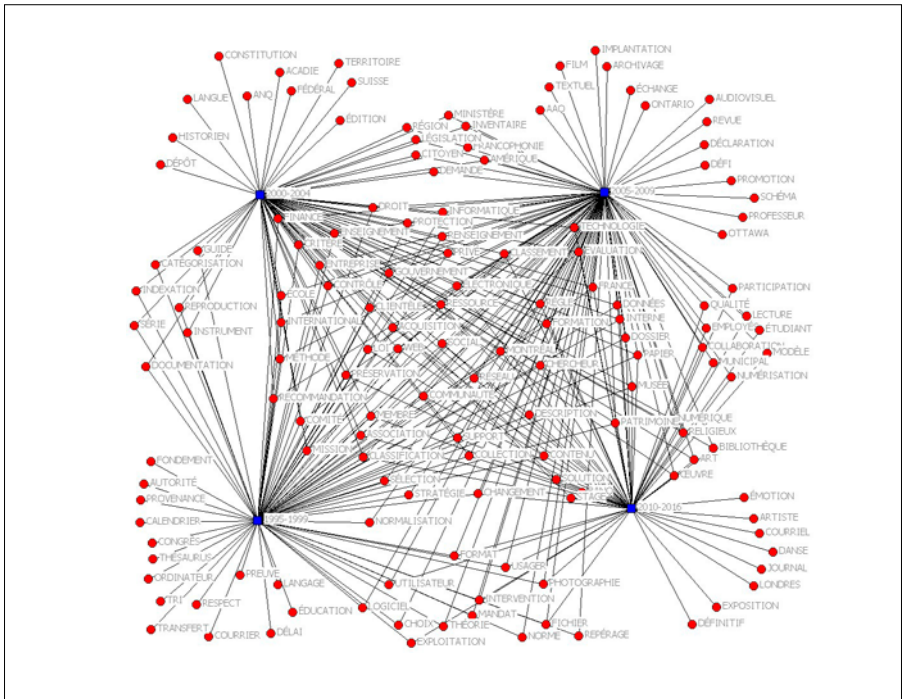


Figure 2 : Les mots-clés répartis selon toutes périodes (1995-2016).

Par exemple, les fonctions archivistiques, telles que l'acquisition, l'évaluation, la classification et la description, sont des thématiques régulièrement abordées par les auteurs des articles de la revue *Archives* au cours des vingt dernières années. Parallèlement, apparaissent des mots-clés plus marginaux, tels que « Acadie », « déclaration », « émotion », souvent à la suite de publications ponctuelles de numéros thématiques ou d'actes de congrès. Nous détaillerons dans les sections qui suivent la relation entre les mots-clés et entre les mots-clés et les auteurs pour chacune des quatre périodes étudiées.

3.1. Mots-clés et auteurs pour la période 1995-1999

La première période étudiée (1995-1999) est caractérisée par l'émergence du courrier électronique dans la sphère privée et surtout institutionnelle; il y est également question de gestion électronique des documents (GED) et de la gestion des documents électroniques (GDE).

Les nouvelles technologies de l'information et la GED commencent à représenter, au milieu des années 1990, un enjeu important pour les archivistes et font l'objet de deux numéros spéciaux dans la revue *Archives*: « Technologies de l'information » (*Archives*, vol. 26, nos 1-2, 1994-1995) et « Gestion électronique des documents » (*Archives*, vol. 29, no 1, 1997-1998). Dans un environnement de travail de plus en plus informatisé, on s'intéresse non seulement au traitement du contenant-médium (support, ordinateur, etc.), mais aussi au contenu-message (évaluation, etc.) ainsi qu'aux modalités de ses mises en forme et de ses représentations (logiciel, format). Le développement de l'informatique et des archives « électroniques » obligent les auteurs de la revue *Archives* à revoir leurs concepts et leurs méthodes de travail (classification, description, indexation, préservation, etc.).

ce qui apparaît comme discriminant. Pour la même raison, les termes « archives », « archivistique » et « gestion » sont également mis de côté.

Dans la Figure 4, l'application des techniques de fouille révèle le degré de centralité des auteurs et des mots-clés en fonction du nombre de relations qui les relie.

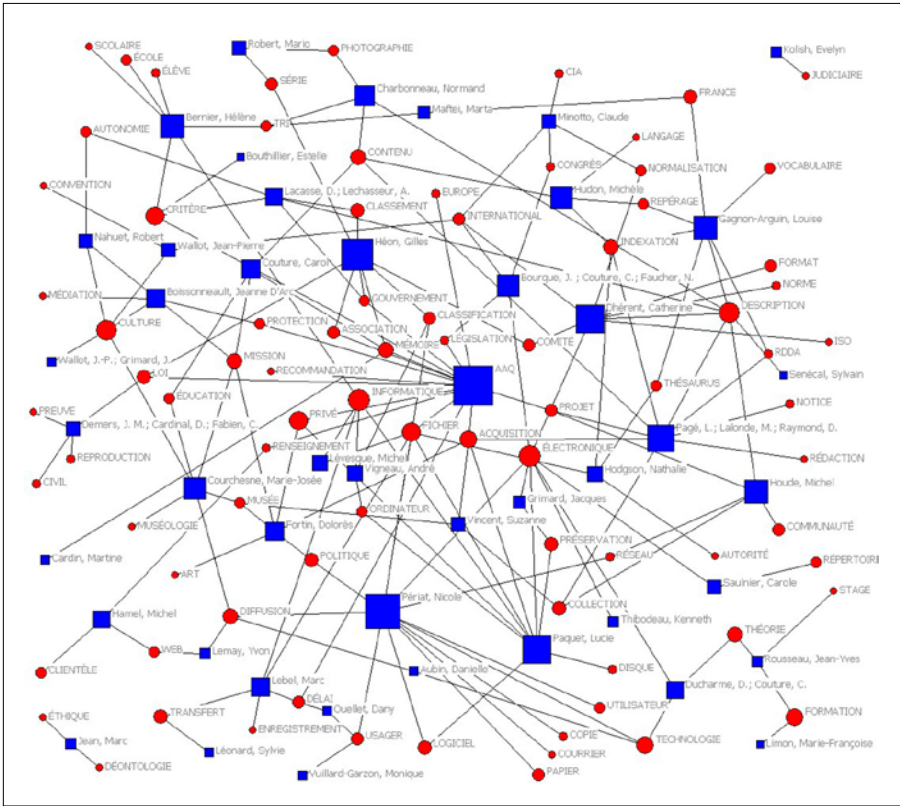


Figure 4: Les mots-clés en fonction des auteurs pour la période 1 (1995-1999).

La centralité est obtenue en fonction du nombre de liens qui relie chaque auteur au reste des mots-clés les plus représentatifs de la période ; elle n'est aucunement liée à la qualité ou au nombre d'articles publiés par chaque auteur. Certains auteurs se trouvent ainsi un peu plus isolés même s'ils ont publié plusieurs articles ; d'autres sont centraux tout en ayant publié qu'un seul article.

3.2. Mots-clés et auteurs pour la période 2000-2004

Cette période est fortement associée à l'archivistique pratiquée à l'extérieur du Québec notamment par la publication du premier numéro d'un volume thématique sur l'« Archivistique comparée » (volume 34, numéro 1, 2002-2003) et d'un autre portant sur les « Archives des francophonies nord-américaines » (volume 36, numéro 1, 2004-2005) ce qui explique sans doute la plus forte présence d'auteurs étrangers (Figure 6) et d'entités nommées représentant des nations (France, Suisse, etc.) ou des régions (Acadie, Nouvelle-France, Louisiane, etc.) (Figure 5). Le volume sur l'« Archivistique comparée » offre une série d'articles :

...sur les systèmes et pratiques archivistiques de cinq pays avec lesquels les archivistes nord-américains francophones du Québec et du Canada ont l'occasion d'entretenir des rapports professionnels ponctuels ou suivis. Ce sont : la Belgique, la France, l'Italie, la Suisse et la Tunisie. (Grimard, 2002-2003, p. 5)

Quant à lui, le volume thématique sur les « Archives des francophonies nord-américaines » regroupe des articles qui couvrent les archives de la Nouvelle-France, l'Ouest canadien, la Louisiane, la Nouvelle-Angleterre, les archives françaises des États-Unis, l'Ontario français (Gagnon-Arguin, 2009-2010). Les articles publiés dans ces quatre numéros de la revue *Archives* ont non seulement « été circonscrits à la fois chronologiquement et géographiquement » (Note de la rédaction, volume 36, numéro 2, p. 3), mais également thématiquement puisque les auteurs devaient couvrir un certain nombre d'aspects : la profession d'archiviste, la nature des archives (publiques, privées, domaines d'activité représentés, etc.), l'archivistique (formation, discipline) et quels étaient les particularismes sur le plan de l'acquisition, de la conservation, de la diffusion, etc. La France prend aussi une importance considérable dans les écrits associés aux années 2000 à 2004. Nous observons et comparons avec la France les concepts d'indexation, de patrimoine religieux, de formation. L'électronique prend moins d'espace ; la description et l'indexation aussi. La *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* (RLRQ, chapitre A-2.1) demeure un sujet d'actualité et est toujours bien représentée (Figure 6).

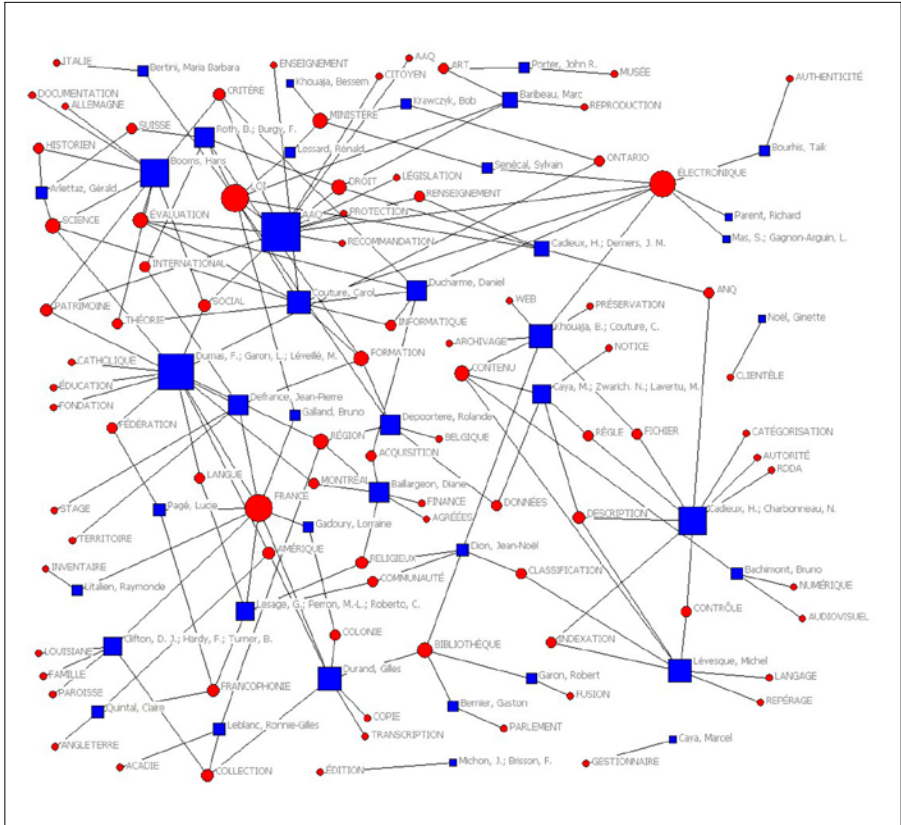


Figure 6 : Les mots-clés en fonction des auteurs pour la période 2 (2000-2004).

3.3. Mots-clés et auteurs pour la période 2005-2009

Deux principaux événements touchant l’archivistique québécoise caractérisent les années 2005 à 2009 : tout d’abord, la *Déclaration québécoise sur les archives* est lancée officiellement en avril 2006, puis l’Association des archivistes du Québec célèbre son 40^e anniversaire en 2007. Un espace important est laissé dans la revue *Archives* pour souligner ces deux événements. Les activités ou discours se rapportant à la *Déclaration* font l’objet d’un dossier spécial dans la revue *Archives*. De même, la commémoration du 40^e anniversaire de l’Association des archivistes du Québec fait l’objet d’un autre numéro spécial (volume 39, numéro 1, 2007-2008) qui comporte des articles sur la vie associative à partir du témoignage d’archivistes provenant du Québec et d’ailleurs.

Par ailleurs, le deuxième volet du projet « Archives des francophonies nord-américaines » (volume 36, numéro 2, 2004-2005) donne lieu une nouvelle fois à la publication d'une série d'articles provenant d'une diversité de communautés et de milieux d'archives. Les auteurs sont invités à couvrir plusieurs aspects tels que la présentation de leurs fonds et collections d'archives, leurs pratiques en matière d'acquisition, de traitement et de diffusion de leurs archives, la formation (Figure 7).

La revue *Archives* diffuse régulièrement les mémoires que l'AAQ élabore dans le cadre de consultations et d'audiences publiques ; la période 2005 à 2009 en est particulièrement prolifique (Figure 8). Un premier mémoire, *Accès aux documents et protection des renseignements personnels. Agir pour plus de transparence*, présente les commentaires de l'AAQ concernant le projet de loi modifiant la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels et d'autres dispositions législatives*. Un second mémoire porte sur *l'Avenir du patrimoine archivistique religieux au Québec* et a été présenté lors des auditions publiques de la Commission de la culture. Un troisième mémoire rédigé dans le cadre de la consultation sur la révision de la *Loi sur les biens culturels* propose des recommandations dans le but de faire reconnaître les archives comme biens culturels à part entière (Notes de la rédaction, volume 37, numéro 2, p.3 et volume 39, numéro 2).

Cette période voit la publication d'un numéro spécial sur les archives universitaires dans une perspective québécoise (volume 38, numéro 2, 2006-2007) en raison du « dynamisme de ce milieu archivistique depuis de nombreuses années et l'importante contribution de ses membres au développement de la théorie et de la pratique archivistique » (Note de la rédaction, p. 2). Les thèmes abordés par les auteurs travaillant dans les services d'archives universitaires portent sur l'acquisition de fonds et de collections d'archives privées, la gestion des documents et des dossiers numériques, la numérisation, les différents types de créateurs, d'utilisateurs ou de bénéficiaires (étudiants, employés, professeurs). Les mots-clés « fonds » et « collection » sont centraux au sein de la Figure 7.

Enfin, la revue *Archives* ne s'intéresse pas seulement aux dimensions théoriques et pratiques de l'archivistique, mais cherche également à s'ouvrir à d'autres besoins que ceux provenant de la gestion et de la

recherche avec un article qui présente trois projets de collaboration entre archivistes et « artistes » et un autre article s’intéressant aux « usagers indirects » des archives.

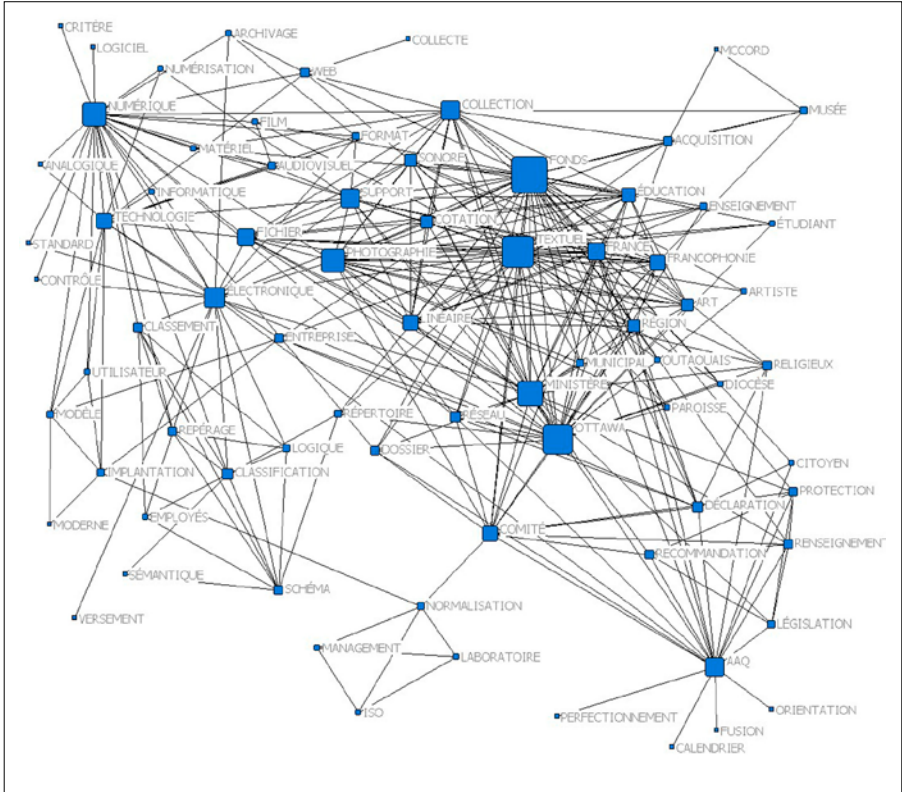


Figure 7 : Les liens entre les mots-clés pour la période 3 (2005-2009).

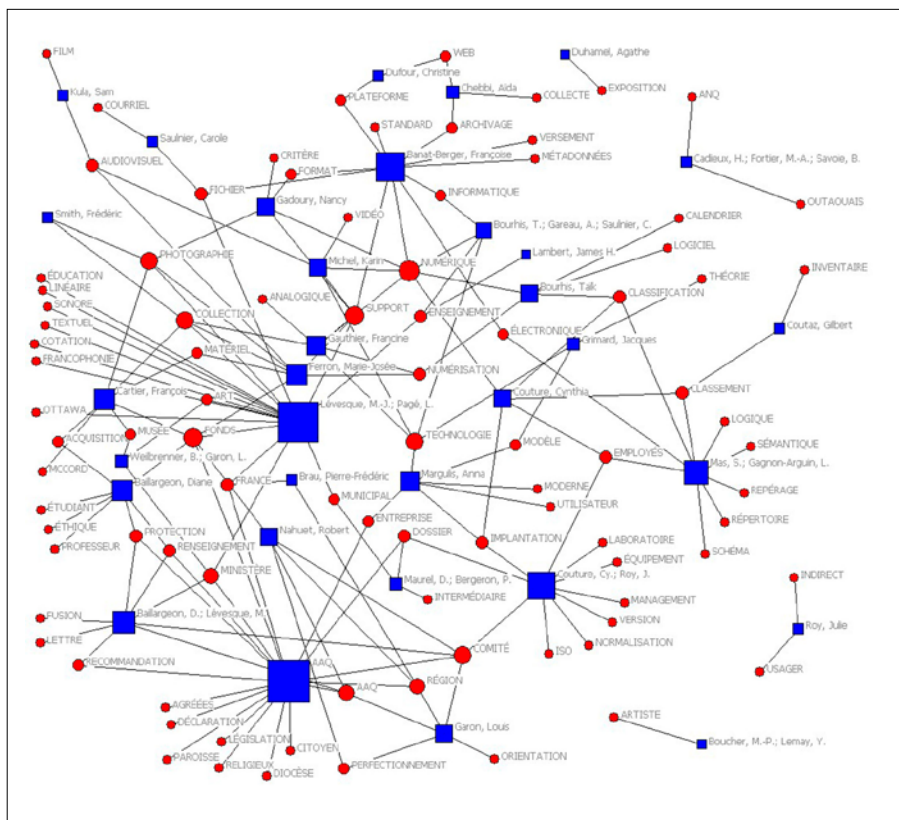


Figure 8 : Les mots-clés en fonction des auteurs pour la période 3 (2005-2009).

3.4. Mots-clés et auteurs pour la période 2010-2016

Afin d'assurer l'actualité des sujets offerts aux lecteurs de la revue *Archives*, cette dernière publie régulièrement les textes des conférences du congrès annuel de l'Association des archivistes du Québec; ainsi en est-il des textes du 41^e (volume 43, numéro 2, 2011-2012) et du 42^e congrès (volume 44, numéro 2, 2012-2013). Durant cette période, la revue *Archives* diffuse, parallèlement et en partenariat avec les organisateurs, les textes de colloques comme celui tenu lors du 81^e Congrès de l'Association francophone pour le savoir à l'Université Laval les 9 et 10 mai 2013 (volume 45, numéro 1, 2013-2014) sur la question de la mise en valeur et de l'exploitation des archives à l'ère numérique par des universitaires et des archivistes professionnels œuvrant dans différents milieux.

L'analyse des mots-clés associés à la quatrième et dernière période analysée est plus difficile que pour les précédentes périodes, car les textes abordent une multitude de thématiques, explicitement mentionnée par la note de la rédaction :

La revue *Archives* s'intéresse à tous les aspects de la gestion de l'information organique et consignée. Qu'elles soient définitives, intermédiaires ou courantes, qu'elles soient privées ou publiques, sur support traditionnel ou sous forme numérique, toutes les archives, dont elle emprunte le nom, sont pour elle objet d'études. (*Archives*, 43(2), p. 3)

Les articles publiés durant cette période illustrent bien cette diversité dans les champs d'intérêt des auteurs de la revue *Archives*. Les thèmes « numérique » et « courriel » semblent toutefois revenir en force.

Cette période débute avec un numéro spécial sur les archives municipales (volume 42, numéro 1, 2010-2011), milieu fort représenté parmi la communauté archivistique. Toujours soucieuse de comparer les pratiques, la revue *Archives* invite neuf auteurs québécois et européens à témoigner de la gestion des documents et des archives dans plusieurs villes au Québec et à l'étranger (Figure 10). Dans un contexte de fusion et défusions municipales, les thèmes traités par les auteurs sont variés : politique de gestion intégrée des documents, création de calendriers de conservation, numérisation, formation du personnel impliqué dans la gestion documentaire, défis liés à la gestion des fichiers numériques, développement de logiciels et systèmes informatiques en soutien à la gestion, la consultation et la valorisation des documents et des archives, parmi d'autres sujets.

Il est révélateur de porter l'analyse non seulement sur les mots-clés centraux, plus récurrents dans les textes des auteurs de la revue *Archives*, mais également sur les mots-clés périphériques, plus marginaux. Le mot-clé « artiste », révélé dans la période de 2004 à 2009, demeure bien ancré entre 2010 et 2016 et le mot-clé « émotion » y apparaît pour la première fois. La raison est très certainement liée à la publication des actes du sixième symposium du Groupe interdisciplinaire de recherche en archivistique intitulé *Les archives, de l'information à l'émotion* dans un numéro de la revue (volume 42, numéro 2, 2010-2011), mais aussi aux travaux d'étudiants universitaires de premier, deuxième ou troisième

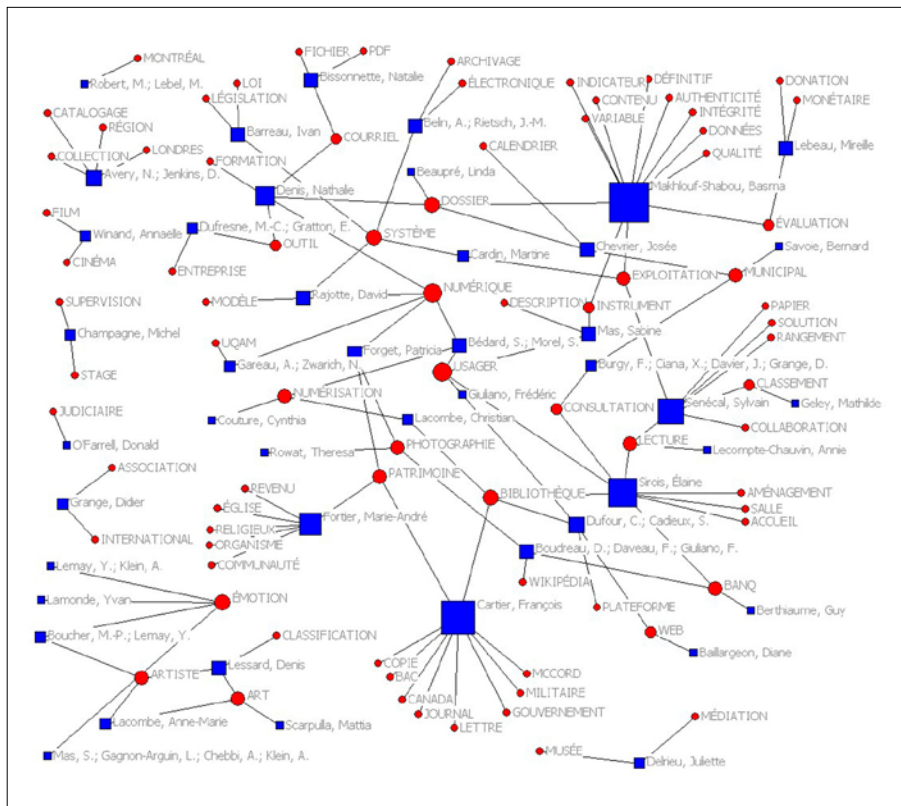


Figure 10: Les mots-clés en fonction des auteurs pour la période 4 (2010-2016).

3.5. Proportion d'auteurs et d'auteurs, 1995-2016

L'analyse bibliométrique des auteurs et auteurs de la revue révèle que la proportion d'articles écrits par des hommes est à la baisse depuis 1995 (Figure 11). Louis-René Dessureault observait déjà à la fin des années 1980 une augmentation du nombre d'articles écrits par des femmes et, plus précisément, que 37,5 % des articles sont écrits par des femmes pour la période 1985 à 1988. Cette proportion se maintient pour la période 1988-2008 (Gagnon-Arguin, 2009-2010, p. 36). Depuis 2005, les femmes sont plus nombreuses que les hommes à écrire des articles. Ce résultat pourrait venir appuyer le constat fait en sciences de l'information où l'on observe une féminisation des professions reliées aux milieux documentaires (Leroux et al., 2009).

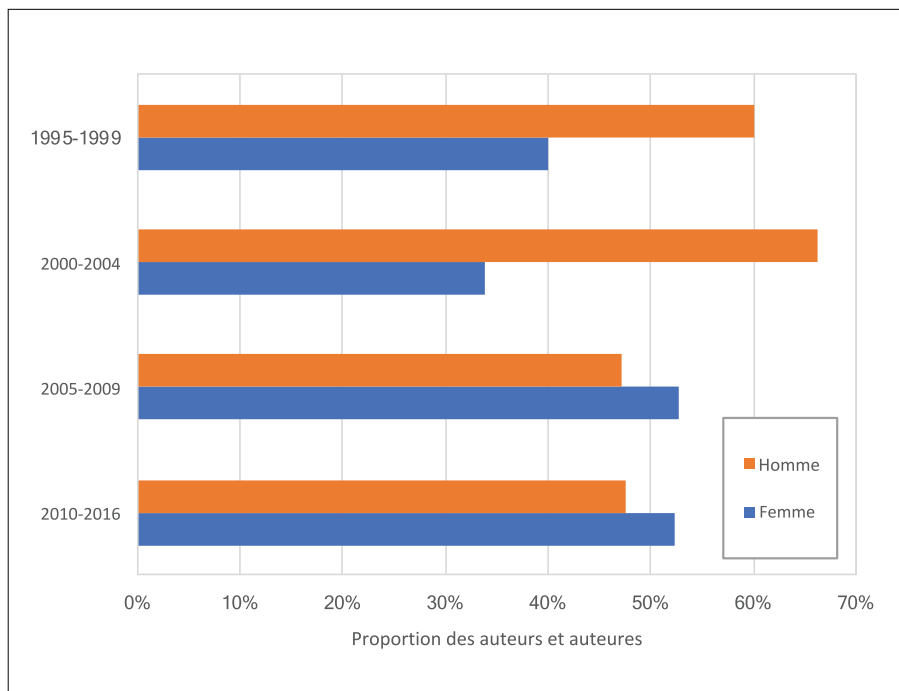


Figure 11: Proportion des auteur(e)s en fonction de leur genre.

3.6. Proportion d’auteurs selon le statut, 1995-2016

Comme l’a souligné Louise Gagnon-Arguin (2009-2010), il est difficile d’opérer une classification des postes occupés par les auteurs des articles puisque les fonctions mentionnées à la fin des articles sont reliées au titre donné par l’institution d’attache et diffèrent d’une institution à l’autre. Toutefois, il a été possible de faire deux principales distinctions parmi les auteurs à partir des titres mentionnés à la fin des articles : ceux travaillant dans le milieu des archives nationales et des services d’archives (les archivistes professionnels) et ceux en provenance du milieu de l’enseignement et de la recherche universitaires (les étudiants et les professeurs en archivistique).

La pénurie d’articles observée ces dernières années serait attribuable, selon les notes de rédaction de la revue *Archives*, au fait que les professeurs publient leurs textes dans d’autres revues. Cela est vrai pour la période 1995 à 2009 où nous observons une proportion régulièrement à la

baisse de textes écrits par des étudiants ou des universitaires. Or, comme l'indique la Figure 12, les résultats révèlent, pour la dernière période analysée (2010-2016), une proportion plus faible de professionnels et, en contrepartie, une proportion plus élevée d'universitaires, de professeurs et d'étudiants de certificat, de maîtrise et de doctorat, ce qui renforce le caractère scientifique de la revue *Archives*. Nous pouvons cependant nous questionner sur la durée de ce phénomène à plus long terme, car, toutes disciplines confondues, la tendance est au retrait des universitaires des revues professionnelles et à la diversification des lieux de publication (Larivière, 2015) explicitement mentionnée par le mot du directeur de la revue *Archives* (volume 43, numéro 2, p.2). Parallèlement à la publication dans des revues savantes ou professionnelles, les étudiants et professeurs universitaires utilisent d'autres plateformes de diffusion que les revues traditionnelles: les dépôts institutionnels, les revues en accès libre, les blogues, etc. À titre d'exemple, les étudiants de l'École de bibliothéconomie et des sciences de l'information (EBSI) publient régulièrement le résultat de leurs travaux sur le dépôt institutionnel *Papyrus* (<https://papyrus.bib.umontreal.ca/xmlui/handle/1866/559>). De leur côté, les enseignants en archivistique de l'Université de Montréal publient sur le blogue *Archives 21* (<http://archives21.ebsi.umontreal.ca/>).

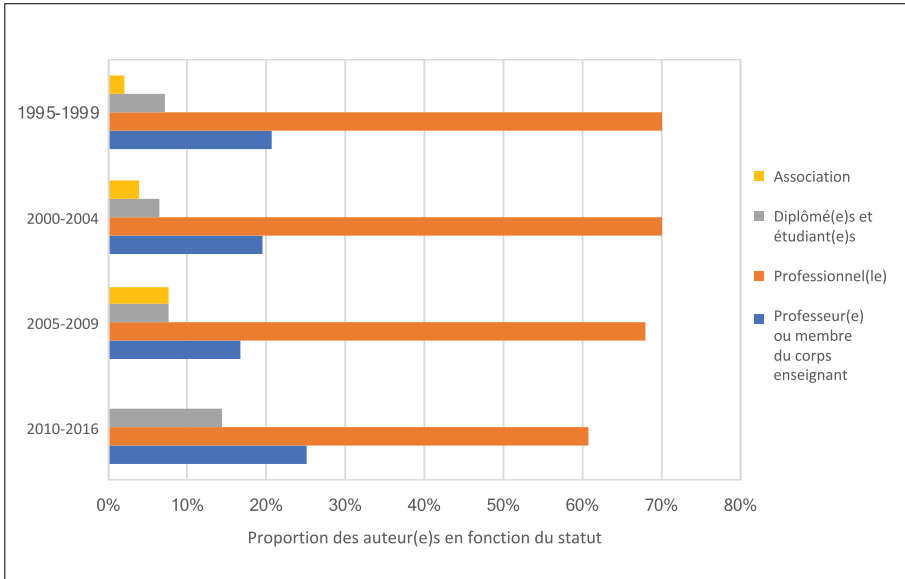


Figure 12: Proportion des auteur(e)s en fonction du statut, 1995-2016.

Il a été possible d'identifier et de comptabiliser les institutions d'affiliation des auteurs (Figure 13). Lors de l'analyse des institutions ayant contribué à cinq articles et plus nous constatons que, par exemple, l'Université de Montréal a contribué à la publication de 70 articles, les Archives nationales du Québec/BANQ à 42 articles et les auteurs sans affiliation, à 19 articles.

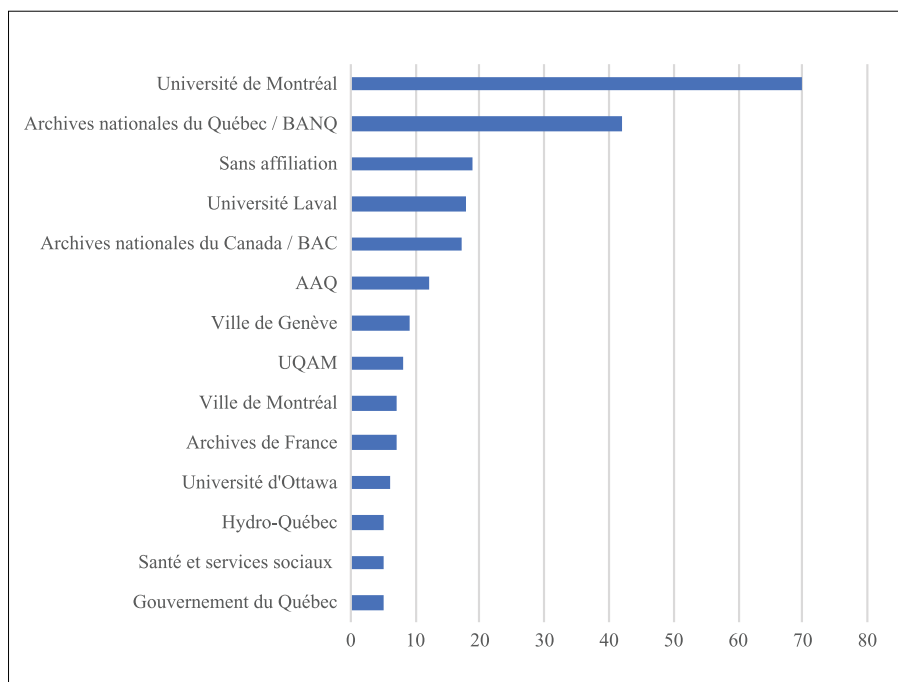


Figure 13: Institutions d'affiliation des auteurs ayant contribué à cinq articles et plus, 1995-2016.

Mis à part les auteurs affiliés aux institutions nationales d'archives, nous observons une importante contribution du milieu universitaire (archivistes, étudiants, professeurs). La raison est liée à la publication d'un numéro spécial portant sur les archives universitaires durant une des périodes analysées (volume 38, numéro 2, 2006-2007). De plus, la reconnaissance institutionnelle encourage le partage de l'expertise de ses membres par le biais de rédaction et de publication d'articles. Les auteurs en provenance du milieu universitaire sont régulièrement encouragés, voire incités à publier par leur institution dont la mission est

d'encourager le développement et la diffusion des connaissances. Par ailleurs, les premiers finissants du programme de doctorat en sciences de l'information, créé en 1997 à l'Université de Montréal, et dont leur thèse porte sur un sujet dans le domaine archivistique, marquent leur milieu en privilégiant très souvent la revue *Archives* comme premier lieu de publication des résultats de leur recherche doctorale.

3.7. Nombre moyen de références par article et par période

Le nombre moyen de références par article et par période (Figure 14) est au-dessous de la moyenne (environ 50 références par article) lorsque nous comparons l'archivistique avec toutes les autres disciplines confondues (Larivière, 2015). Ces résultats montrent même une certaine baisse au cours de la dernière décennie, ce qui est contraire à la tendance observée dans l'ensemble des disciplines. Chose certaine, la tendance du nombre moyen de références par articles n'est manifestement pas à la hausse, ce qui semblerait *a priori* être révélateur d'une sous-académisation de la revue *Archives* ou, à tout le moins, d'une plus faible intégration dans la littérature existante.

Plusieurs explications sont possibles, liées soit aux choix méthodologiques réalisés dans le cadre de cette étude, soit au caractère hétérogène du corpus, qui tendent à diminuer le nombre moyen de références par article durant les périodes étudiées. Tout d'abord, le corpus comprend plusieurs mémoires d'une quarantaine de pages déposés en commission parlementaire par l'AAQ et dont les références bibliographiques, présentes et parfois nombreuses, sont « cachées » dans les notes de bas de pages exclues de l'analyse bibliométrique. Ensuite, les textes des communications ou des conférences d'ouverture et de clôture de congrès et de colloques publiés dans la revue *Archives* ne comportent généralement aucune référence bibliographique.

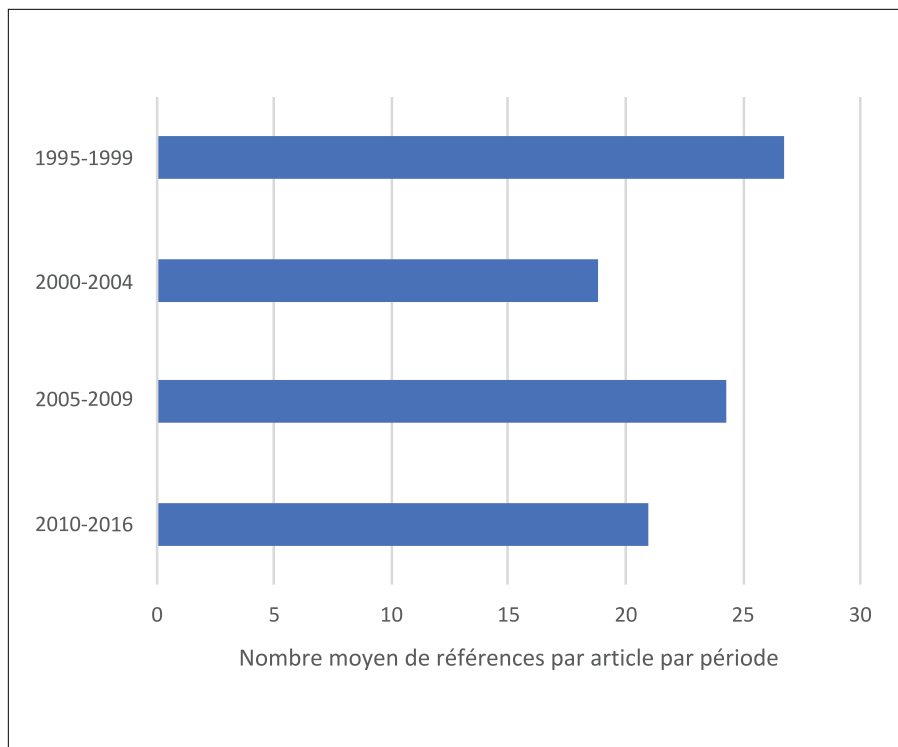


Figure 14: Nombre moyen de références par article, 1995-2016.

3.8. Âge moyen des documents cités

Dans l'ensemble, les travaux savants citent en moyenne des documents de plus en plus âgés (autour de 13-14 ans pour les années récentes). Cette tendance est due à deux facteurs: le passage d'une croissance exponentielle de la science à une croissance plus linéaire, ainsi que l'avènement de l'ère numérique et de l'indexation rétrospective d'articles de périodiques qui ont ainsi eu une seconde vie (Larivière, Gingras et Archambault, 2008). Nous pourrions nous attendre à un âge moyen plus élevé pour les documents cités par les auteurs de la revue *Archives* grâce, d'une part, à son objet d'étude associé au passé, mais également, à la numérisation et à l'accès en ligne aux anciens numéros de la revue. Or ce n'est pas le cas: l'âge moyen des documents cités par la revue *Archives* est inférieur à 10 ans, ce qui est au-dessous de la moyenne de l'ensemble des disciplines (Figure 15).

Une seule exception est observée pour la période 2000-2004, qui correspond à un numéro spécial sur l'archivistique comparée (archives en Italie, Suisse, Belgique, etc.), et où nous observons un plus grand nombre d'articles qui développent un aspect plus historique des archives et des institutions d'archives. Une explication possible serait que la connaissance contenue dans les articles en archivistique n'est pas éternelle. Autrement dit, la validité des connaissances en archivistique est en fonction d'une époque, d'un contexte.

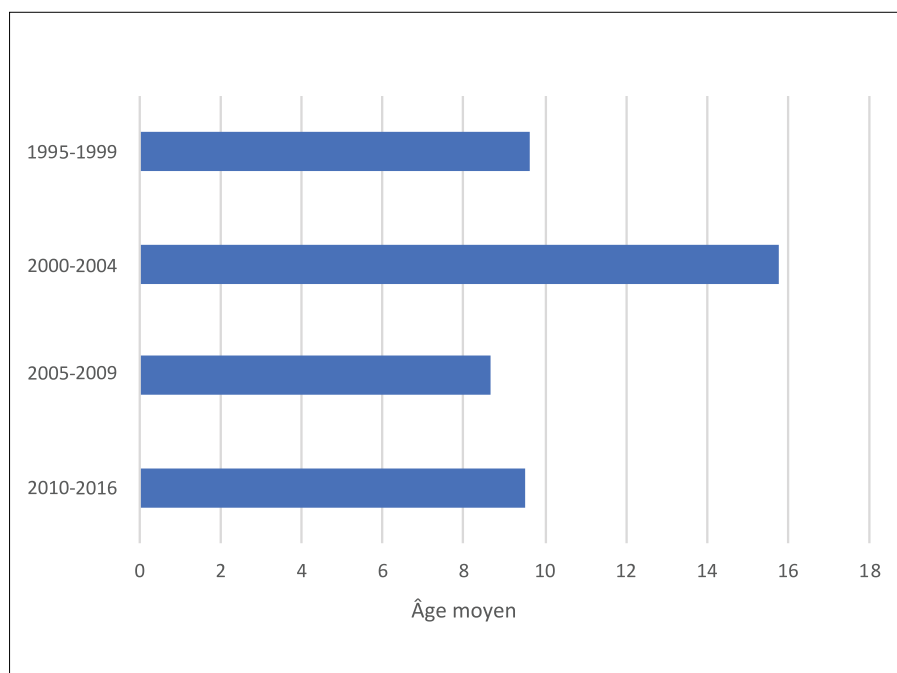


Figure 15: Âge moyen des documents cités par la revue *Archives*, 1995-2016.

3.9. Principales revues citées

Les résultats obtenus pour la recherche montrent que la grande majorité des références faites par la revue *Archives*, soit 60 %, ne vont pas à des articles de revues, mais plutôt à des monographies et à de la littérature grise (rapports institutionnels, par exemple). Cette tendance est similaire à celle observée pour d'autres domaines relevant

davantage des humanités, telle que l'histoire, où plus de 60 % des références sont faites à des documents autres que des articles de revues (Larivière, Archambault, Gingras et Vignola-Gagné, 2006). Pour les revues citées, nous observons, au fil des périodes, un nombre de plus en plus important de citations de revues nord-américaines (*Archives*, *American Archivist*, *Archivaria*) (Figure 16). La revue *Archives* est toujours bien représentée comme source d'information pour les auteurs de celle-ci. Cependant, la *Gazette des archives*, produite par l'Association des archivistes français, est relativement peu citée. Il serait intéressant de vérifier cela dans les prochaines années en raison de l'accès récent sur *Percée* à l'intégralité des textes publiés dans cette revue depuis les années 1930.

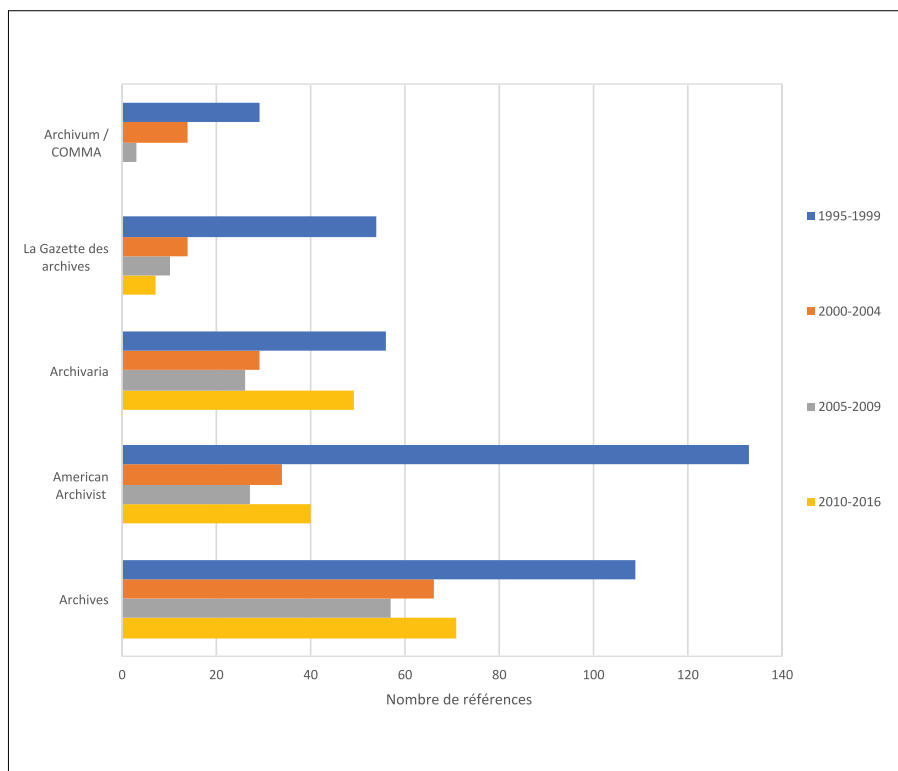


Figure 16 : Nombre de références faites par revue et par période, pour les 5 revues les plus citées, 1995-2016.

CONCLUSION

L'analyse du contenu de la revue *Archives* à l'aide de techniques de fouille et de bibliométrie a été possible uniquement par la mise en ligne et l'accès à la version numérique des textes de la revue depuis 1994 sur le site Web de l'AAQ. Il s'agit d'un premier essai, car sur le plan méthodologique les techniques de fouille et de bibliométrie portent habituellement sur des corpus beaucoup plus volumineux. Les résultats obtenus sont intéressants, malgré la taille réduite du corpus disponible. Bien entendu, les résultats de l'analyse automatisée ne peuvent être compris et interprétés sans une connaissance du contexte archivistique québécois, de la consultation des articles de la revue *Archives*, de la connaissance des choix éditoriaux par la consultation des notes de rédaction et de la connaissance de l'évolution des communications savantes.

Les résultats sont le reflet de quelques-unes des préoccupations et des défis des archivistes travaillant essentiellement au Québec. Dans un premier temps, l'étude des mots-clés a révélé non seulement la diversité des intérêts des auteurs de la revue, mais aussi l'évolution de l'archivistique québécoise elle-même à travers l'étude de sa terminologie : certains termes propres à une époque disparaissent, alors que d'autres termes demeurent ou deviennent plus spécifiques, ce qui est un trait caractéristique de l'évolution des disciplines dans la sociologie des sciences. Dans un deuxième temps, les résultats nous révèlent plusieurs facettes des auteurs : augmentation de la proportion d'articles écrits par des femmes, affiliation fortement associée au secteur public ou parapublic, intérêts professionnels ou de recherches variés, mais fortement ancrés par des problématiques concernant le numérique. Pour terminer, l'analyse des références citées permet de positionner l'archivistique québécoise au sein de l'ensemble des disciplines.

La diffusion récente d'un premier numéro de la revue *Archives* sur la plateforme *Érudit* (volume 46, numéro 1, 2016) encourage le rayonnement non seulement national, mais aussi international de la seule revue exclusivement dédiée à l'archivistique et publiée au Québec depuis bientôt cinquante ans. Très certainement, ce nouveau mode de diffusion des articles de la revue *Archives* contribuera non seulement à offrir une plus grande visibilité de l'expertise des archivistes québécois sur le plan international, mais pourrait aussi contribuer à favoriser les

échanges en attirant de nouveaux auteurs provenant de l'étranger dans un souci d'ouverture à d'autres pratiques et à d'autres milieux professionnels et scientifiques.

DOMINIC FOREST
SABINE MAS
VALÉRIE RIOUX
VINCENT LARIVIÈRE
BENOÎT MACALUSO

REMERCIEMENTS

Nous tenons à remercier Mme Louise Gagnon-Arguin pour ses remarques pertinentes sur l'ensemble de notre projet.

BIBLIOGRAPHIE

- BOUGOUIN, A. (2015). *Indexation automatique par termes-clés en domaines de spécialité*. (Thèse de doctorat), Université de Nantes Angers le Mans, Nantes. Repéré à <http://adrien-bougouin.github.io/publications/2015/thesis.pdf>
- DESSUREAULT, L.-R. (1989). *L'état du discours de la profession d'archiviste au Québec: une analyse de contenu de la revue Archives*. (Mémoire de maîtrise). Université de Montréal, Montréal.
- FAYYAD, U., GRINSTEIN, G.G. et WIERSE, A. (dir.). (2001). *Information visualization in data mining and knowledge discovery*. Burlington, Massachusetts: Morgan Kaufmann Publishers.
- FOREST, D., BASTIEN, F., LEGAULT-VENNE, A., BROUSSEAU, H. et LACOMBE, O. (2017). Les mots de la campagne: la fouille de textes appliquée à l'étude de la communication électorale. Dans P.-M. Daigneault et F. Pétry, (dir.). *L'analyse textuelle des idées, du discours et des pratiques politiques*. (97-120) Québec, Québec: Presses de l'Université Laval.

GAGNON-ARGUIN, L. (1988). Les 20 ans de la revue *Archives*: analyse des articles et des auteurs de 1969 à 1988. *Archives*, 20(1), 3-29.

GAGNON-ARGUIN, L. (1994). La revue *Archives* depuis 1988: étude de son évolution. *Archives*, 25(3), 3-22.

GAGNON-ARGUIN, L. (2008). Les revues professionnelles en sciences de l'information au Québec. *Documentation et bibliothèques*, 54(2), 189-192.

GAGNON-ARGUIN, L. (2009-2010). La revue *Archives* à travers ses vingt dernières années d'existence, 1988-2008. *Archives*, 41(2), 22-44.

GRAHAM, S., MILLIGAN, I. et WEINGART, S. (2015). *Exploring Big Historical Data: The Historian's Macroscope*. Londres, Royaume-Uni: Imperial College Press.

GRIMARD, J. (2002-2003). Un regard sur le monde archivistique: particularismes et similitudes. *Archives*, 34(1-2), 5-8.

HEARST, M. (2003). What is text mining? Document non-publié repéré à www.ischool.berkeley.edu/hearst/text-mining.html.

LARIVIÈRE, V., ARCHAMBAULT, É., GINGRAS, Y. et VIGNOLA GAGNÉ, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, 57(8), 997-1004.

LARIVIÈRE, V., ARCHAMBAULT, É. et GINGRAS, Y. (2008). Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900-2004). *Journal of the American Society for Information Science and Technology*, 59(2), 288-296.

LARIVIÈRE, V. (2015). *Inflation doctrinale? Numérique et modes de diffusion des connaissances*. Centre de recherche en droit public, Université de Montréal, [Vidéo en ligne]. Repéré à <https://www.youtube.com/watch?v=ZJV2vGcmNow>.

LEMAY, Y., et Boucher, M.-P. (2010-2011). L'émotion ou la face cachée de l'archive. *Archives*, 42(2), 39-52.

- LEROUX, E., LEMAY, Y., BERGERON, P., DUFOUR, C., GAZO, D., MARCOUX, Y., ... SALAÜN, J.-M. (2009). Les professions et les institutions. Dans J.-M. Salaün et C. Arsenault (dir.), *Introduction aux sciences de l'information* (p. 15-52). Montréal, Québec: Les Presses de l'Université de Montréal.
- MOKRANE, A., AREZKI, R., DRAY, G. et PONCELET, P. (2004). Cartographie automatique du contenu d'un corpus de documents textuels. Dans *Actes des JADT'04, Le poids des mots* (2, p. 816-823). Louvain-la-Neuve, Belgique: Presses Universitaires de Louvain.
- SPENCE, R. et PRESS, A. (2000). *Information visualization*. Boston, Massachusetts: Addison-Wesley.