

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

AUDIT ALGORITHMIQUE DE LA PLATEFORME NUMÉRIQUE GOOGLE ACTUALITÉS : LA PART DES
MÉDIAS LOCAUX ET RÉGIONAUX QUÉBÉCOIS DANS LES RÉSULTATS DE RECHERCHE D'AOÛT À
DÉCEMBRE 2021

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN COMMUNICATION

PAR
NICOLAS ST-GERMAIN
JUIN 2023

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens à remercier toutes les personnes qui m'ont soutenu tout au long de ce projet qui s'est amorcé il y a maintenant plus de deux ans. Mes amis, ma famille et ma conjointe, Camille Beausoleil, merci de m'avoir poussé à continuer! Un grand merci également à Marie-Pier Frappier, ma gestionnaire chez Groupe Contex/Les Affaires, qui m'a poussé à prendre une demi-journée par semaine pendant plus d'un an pour continuer mon cursus et écrire mon mémoire. Cette aide m'a permis de compléter ce mémoire, et ce malgré mon horaire chargé de travailleur à temps plein. Merci également à Camille Payant qui m'a envoyé l'étude américaine sur laquelle se base ce mémoire. Et finalement, un merci très spécial pour mon directeur de recherche qui était plus que présent et qui n'a pas hésité à me donner un coup de main pour organiser ma pensée et concevoir les programmes informatiques.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
TABLE DES MATIÈRES	iii
LISTE DES FIGURES	v
LISTE DES TABLEAUX	vii
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES	viii
RÉSUMÉ.....	ix
ABSTRACT.....	x
INTRODUCTION	1
I. La bataille de l'attention et des revenus publicitaires	3
II. La bataille de la conversation publique et des données publicitaires	4
CHAPITRE 1 PROBLÉMATIQUE	1
1.1 État de la recherche sur la curation algorithmique de l'information.....	4
1.2 La réalité de la presse locale	8
1.3 Et si le lecteur était en cause?.....	11
1.4 Question générale de recherche et pertinence	12
CHAPITRE 2 CADRE D'ANALYSE	14
2.1 Cadre conceptuel	14
2.1.1 Algorithme	14
2.1.2 Audit	15
2.1.3 Référencement	16
2.2 Cadre théorique	17
2.2.1 Responsabilité/gouvernementalité algorithmique.....	19
2.2.2 Et Google Actualités dans tout ça?	21
2.3 Questions spécifiques et hypothèses de recherche.....	22
CHAPITRE 3 MÉTHODOLOGIE.....	24
3.1 Structure des requêtes automatisées	26
CHAPITRE 4 RÉSULTATS.....	29
4.1 Explications et détails de la base de données.....	32
4.1.1 L'entièreté de la base de données.....	33

4.1.2	Dans les cinq premiers résultats	38
4.1.3	Comparaison avec la base de données de Fischer <i>et al.</i>	40
4.2	Réponses aux hypothèses de recherche.....	40
4.2.1	Hypothèses de recherche primaire.....	41
4.2.2	Répartition des résultats pour le mot-clé « élection OR campagne »	44
4.3	Comparaison avec la recherche de Fischer <i>et al.</i> ainsi que les critiques de Magnusson	46
4.3.1	Reproduction de la recherche de Fischer <i>et al.</i>	47
4.3.1.1	Hypothèse 1.....	47
4.3.1.2	Reproduction de l'hypothèse 1 avec les critiques de Magnusson.....	56
4.3.1.3	Hypothèse 2.....	59
4.4	Le cas de Nécrologie Canada.....	65
4.5	La temporalité des contenus affichés	66
	CONCLUSION	75
	ANNEXE 1 Exemple de résultats de recherche à partir du mot-clé « accident » et de la localisation « près de Laval ».	77
	ANNEXE 2 Exemple de document CSV contenant les résultats de requêtes effectuées le 18 décembre 2021.	78
	ANNEXE 3 Les 26 mots-clés utilisés pour la présente recherche ainsi que les 32 utilisés par Fischer <i>et al.</i> , par type de mot-clé.....	79
	ANNEXE 4 Listes des 104 MRC ou territoires équivalents et leurs chefs-lieux ou villes les plus peuplées. 80	
	ANNEXE 5 Liste des 20 médias, filtrés par catégorie, les plus souvent affichés par Google Actualités dans la base de données complète.....	83
	ANNEXE 6 Liste des 20 médias les plus fréquemment affichés par Google Actualités dans la base de données complète, selon le type de mots-clés.	84
	ANNEXE 7 Liste des 62 villes où se trouvent les sièges sociaux des médias locaux et régionaux répertoriés par Google Actualités.	86
	ANNEXE 8 Liste des 122 médias locaux et régionaux les plus retournés par Google Actualités, avec leurs sièges sociaux.	88
	ANNEXE 9 Poids relatif des différentes catégories de médias et catégorie la plus souvent retournée par Google Actualités, par ville.....	91
	RÉFÉRENCES	94

LISTE DES FIGURES

Figure 1. Répartition de l'ensemble des résultats de recherche sur Google Actualités selon le rang de chaque catégorie de média dans ces résultats.....	29
Figure 2. Répartition, par type de mot-clé, des résultats de recherche sur Google Actualités selon le rang de chaque catégorie de média dans les résultats.....	30
Figure 3. Indice de Gini pour les types de mots-clés local, national et mixte selon la position dans les résultats de recherche.....	31
Figure 4. La figure 1 de la recherche Fischer et al.	48
Figure 5. Première reproduction de la figure 1 de Fischer et al. Coefficient de Gini pour les mots-clés orientés national (jaune) et mixte (bleu). Chaque mot-clé est étiqueté avec le média retourné le plus souvent et sa catégorie.	50
Figure 6. Première reproduction de la figure 1 de Fischer. Coefficient de Gini pour les mots-clés orientés localement. Chaque mot-clé est étiqueté avec le média retourné le plus souvent et sa catégorie.	51
Figure 7. Deuxième reproduction de la figure 1 de Fischer. Coefficient de Gini pour chaque mot-clé orienté localement. Moyenne pour les quatre catégories de média. L'étiquette indique la catégorie de média la plus fréquemment retournée.	53
Figure 8. Deuxième reproduction de la figure 1 de Fischer. Coefficient de Gini pour chaque mot-clé orienté mixte (bleu) et national (jaune). Moyenne pour les quatre catégories de média. L'étiquette indique la catégorie de média la plus fréquemment retournée.	54
Figure 9. Écart des coefficients de Gini calculés par Magnusson et par Fischer et al. pour les recherches avec les mots-clés locaux.....	57
Figure 10. Écart des coefficients de Gini calculés par Magnusson et par Fischer <u>et al.</u> pour les recherches avec les mots-clés nationaux.....	57
Figure 11. « Figure 2 : The relative frequency of local and national outlets for the top n results of each search query. The first sixteen purple-labeled terms are locally-oriented terms. The last sixteen orange-labeled terms are general terms. » (Fischer et al., 2020, p. 23).....	60
Figure 12. Fréquence relative des résultats pour les requêtes contenant les 13 mots-clés locaux.....	61
Figure 13. Fréquence des résultats pour les requêtes contenant les six mots-clés mixtes	63
Figure 14. Fréquence des résultats pour les requêtes contenant les sept mots-clés nationaux	64
Figure 15. La fraîcheur en jours des articles selon la catégorie pour les 50 premières villes étudiées.....	68
Figure 16. La fraîcheur en jours des articles selon la catégorie pour les 55 dernières villes étudiées.....	69

Figure 17. La fraîcheur en jours des articles, par catégorie, pour les mots-clés locaux.71

Figure 18. La fraîcheur en jours des articles, par catégorie, pour les mots-clés nationaux.72

Figure 19. La fraîcheur en jours des articles, par catégorie, pour les mots-clés mixtes.....73

LISTE DES TABLEAUX

Tableau 1. Parts du marché publicitaire au Québec (CEM, 2020).....	4
Tableau 2. Répartition de la base de données par catégorie de médias	33
Tableau 3. Répartition de la base de données par catégorie de média selon le type de mots-clés	35
Tableau 4. Répartition des résultats de recherche pour les 20 médias les plus souvent affichés par Google Actualités, selon la catégorie de média. Le poids dans les résultats reflète le poids de chaque catégorie de média dans l'ensemble de la base de données (488 009 résultats).	35
Tableau 5. Répartition des résultats de recherche pour les 20 médias les plus affichés par Google Actualités selon la catégorie de média et le type de mots-clés	36
Tableau 6. Répartition des cinq premiers résultats sur Google Actualités par catégorie de média.	39
Tableau 7. Comparaison du poids des médias les plus retournés entre la base de données complète et celle comprenant les cinq premiers résultats de recherche sur Google Actualités	40
Tableau 8. Comparaison entre la présente recherche et celle de Fischer et al.	40
Tableau 9. Nombre de villes où chaque catégorie de média est la plus fréquemment retournée.....	42
Tableau 10. Les villes où la catégorie de média la plus fréquente dans les résultats de recherche sur Google Actualités n'est pas « local/régional ».	43
Tableau 11. Poids de différents groupes des médias les plus fréquemment retournés par Google Actualités (20 premiers, 10 premiers, cinq premiers, trois premiers et le premier) dans les recherches ne comprenant que le mot-clé « élection OR campagne » par rapport à l'ensemble des recherches (base de données complète).	45
Tableau 12. Répartition des résultats pour le mot-clé « élection OR campagne » selon la catégorie de média	45
Tableau 13. Reproduction du calcul de l'indice de Gini selon le mot-clé en filtrant par villes.	58
Tableau 14. Fraîcheur en jours des articles retournés par Google Actualités selon le mot-clé pour les deux recherches effectuées à partir de Montréal, ainsi que pour l'ensemble de la base de données.	74

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

GAFAM: Google, Amazon, Facebook, Apple, Microsoft

RÉSUMÉ

Cette recherche analyse le fonctionnement de l'algorithme de Google Actualités afin de déterminer la place qu'il accorde aux médias locaux et régionaux dans ses résultats, le tout dans un contexte de crise médiatique. L'audit algorithmique, basé sur un corpus de plus de 450 000 résultats, observe les médias retournés par Google Actualités au Québec lors de requêtes comprenant un des 26 mot-clé catégorisé comme local, mixte ou national ainsi qu'une des 104 localisations. À la lumière des résultats, il semble clair que les médias locaux et régionaux ne sont pas désavantagés par la plateforme. Au contraire, ils sont souvent les plus retournés, quelle que soit la catégorie de mots-clés. En revanche, l'algorithme semble avoir un problème lorsqu'il s'agit d'afficher du contenu récent, car le jour moyen entre la date de publication et les dates d'entrée est très élevé. De plus, Google News affiche de nombreux résultats qui ne sont pas des médias ou qui sont situés en dehors de la province, dont beaucoup en France.

Mots-clés : Audit, algorithme, Google Actualités, référencement, médias locaux et régionaux, médias nationaux, curation algorithmique, gouvernamentalité algorithmique, responsabilité algorithmique, transformation numérique

ABSTRACT

This research analyzes the functioning of the Google News algorithm in order to determine the place it gives to local and regional media in its results, all this in a context of media crisis and increasing closure of local media. The algorithmic audit, based on a corpus of more than 450,000 results, observes the media returned by Google News in Quebec when queries including keywords categorized as local, mixed or national as well as a location are submitted. Considering the results, it seems clear that local and regional media are not disadvantaged by the platform. On the contrary, they are often the most returned regardless of the keyword category. On the other hand, the algorithm seems to have a problem when it comes to displaying recent content as the average day between the date of publication and the dates of entry is very high. Moreover, Google News displays many results that are not media or that are located outside the province, including many in France.

Keywords: Audit, algorithm, Google News, SEO, local media, regional media, national media, algorithmic accountability, data scrapping

INTRODUCTION

Le recours au web fait partie du quotidien des citoyens. Au Québec, en 2021, ce sont 93 % des foyers qui disposent d'un accès à Internet, une hausse de 17 points de pourcentage en 10 ans, selon l'étude NETendances (Académie de la transformation numérique, 2021, p. 11). Cette recherche nous apprend aussi que 90 % des adultes québécois utilisent Internet quotidiennement, dont 82 % le font « plusieurs fois par jour » (*Ibid.*, p. 6). Ces chiffres sont étrangement en baisse par rapport à 2020, où 97 % des foyers mentionnaient avoir un accès Internet et où 93 % utilisaient Internet quotidiennement. Aucune raison n'est évoquée par NETendances.

Par ailleurs, les Québécois ont admis que la pandémie de COVID-19 avait eu des répercussions sur le temps qu'ils passent devant un écran en 2020. En effet, 63 % jugent que celui-ci a « un peu augmenté » (32 %) ou « beaucoup augmenté » (31 %) (Académie de la transformation numérique, 2020, p. 16). Quant aux usages accomplis sur le web, il faut se tourner vers les chiffres produits en 2018 par l'enquête NETendances qui démontre que 83 % des adultes québécois utilisaient au moins un réseau socionumérique (CEFRIQ, 2018, p. 6). Du lot, « près de huit adultes québécois sur dix (79 %) ont suivi l'actualité ou les nouvelles sur ces plateformes » (*Ibid.*, p. 13). Les résultats de l'enquête démontrent également que les usages des réseaux socionumériques sont en hausse constante depuis 2015. Cela laisse donc croire qu'en 2021 — le moment où ce projet s'est amorcé —, les chiffres seraient encore plus élevés, surtout en raison de la pandémie.

Ce constat concernant l'usage des réseaux socionumériques chez les adultes québécois inclut également les journalistes. En effet, près de 93 % des journalistes québécois utiliseraient les réseaux socionumériques dans le cadre de leur travail — Facebook (96 %) et Twitter (85 %) en tête (Dubois, 2021, p. XV). Un constat logique lorsque l'on remarque la place du numérique dans les foyers au Québec. Les usages des réseaux socionumériques par les journalistes ne font que « refléter les usages de la population » (Roy, 2015, p. 6).

Cette prémisse permet d'établir une chose : l'arrivée du numérique a mené à un changement dans la manière de s'informer des citoyens. En effet, ceux-ci ont délaissé les médias traditionnels comme les journaux imprimés, la télévision et la radio au profit de plateformes numériques. Par plateformes, nous entendons à la fois des agrégateurs de nouvelles en ligne, comme Apple News ou Google Actualités, et des réseaux socionumériques tels que Facebook ou Twitter.

Cette transformation a été accompagnée de changements économiques. La publicité, principale source de revenus des entreprises de presse privées, a migré vers les plateformes numériques. Ces transformations ont causé, en partie, la crise médiatique vécue actuellement par les médias traditionnels. En effet, au cours des deux premières décennies du XXI^e siècle, ceux-ci ont vu leurs revenus fondre à vue d’œil au détriment des géants du web (Google, Apple, Facebook, Amazon, Microsoft, Netflix, Spotify) (CEM, 2020). De plus, la présence de ces nouveaux outils numériques a fait perdre aux médias autant nationaux que locaux et régionaux leur monopole sur l’information. Par conséquent, cela force ces organisations à revoir leur modèle d’affaires et à tenter de se faire une place dans ce nouvel univers où tous les contenus ont la même valeur lorsqu’ils sont diffusés sur les plateformes numériques (Nocetti, 2017).

Cette réalité concernant la presse dans sa globalité affecte davantage les petits joueurs, au niveau local, qui n’ont pratiquement pas de ressources pour se battre face aux géants du numérique qui accaparent une grande partie de l’attention des Québécois. La presse locale se voit donc touchée de plein fouet par des coupes alors que son rôle est primordial à une saine démocratie locale. La disparition de nombreux médias locaux crée un « appauvrissement des nouvelles locales » au point où certaines localités n’ont plus accès à une information de qualité (Ryerson University, 2021).

Si les changements engendrés par l’arrivée du numérique affectent à ce point le secteur des médias, c’est que sa relation avec l’innovation technique et technologique en est une d’influence. À ce propos, le chercheur Christian Delporte écrit :

L’information s’approprie l’innovation technique, la détourne de sa destination première, l’applique à ses besoins; car l’information, par ses caractères mêmes, doit impérativement se saisir des derniers progrès dans le domaine de la technologie de la communication. (2005, p. 204)

C’est ainsi que les médias ont développé une relation de dépendance envers les plateformes numériques. Effectivement, avec près de 77 % des Canadiens qui disaient s’informer en ligne en 2021 et 55 % par l’entremise des médias sociaux, la présence des médias traditionnels sur les plateformes numériques, qu’ils soient nationaux, régionaux ou locaux, semble nécessaire pour rejoindre leur public (Newman *et al.*, 2022, p. 119).

Certains médias tentent néanmoins de briser cette dépendance. Ils ne croient plus que les réseaux sociaux puissent leur permettre de faire croître leur audience et leurs revenus de manière substantielle (Rashidian *et al.*, 2020, p. 8). Cela ne veut pas dire qu'ils vont délaisser les plateformes, mais plutôt qu'ils vont réorienter leurs efforts vers leurs lecteurs loyaux (*Ibid.*, p. 12). Rashidian *et al.* constatent un changement dans la stratégie de certaines organisations médiatiques. Au lieu de pourchasser les clics sur les plateformes, elles vont plutôt mettre en valeur le contenu se trouvant dans leur site web au moyen d'infolettres et miser sur les abonnements ou sur des contenus partenaires ou de marque pour générer des revenus (*Ibid.*).

Il n'en demeure pas moins que les bouleversements causés par le déplacement de la consultation médiatique au Québec vers le numérique ont eu des répercussions sur les médias traditionnels qui peuvent s'expliquer par quatre batailles, toutes remportées par les plateformes numériques : la bataille de l'attention, la bataille des revenus publicitaires, la bataille de la conversation publique et la bataille des données personnelles (Proulx, 2020).

I. La bataille de l'attention et des revenus publicitaires

De 2016 à 2021, le pourcentage de Canadiens qui s'informent par le biais des médias imprimés a chuté de 36 % à seulement 16 % de la population (Newman *et al.*, 2021, p. 119). Pendant ce temps, l'utilisation des réseaux sociaux comme source d'information est passée de 48 % à 55 %. Facebook s'inscrit au sommet des plateformes utilisées avec 41 %. Le réseau attire 69 % des Canadiens pour tout usage (Newman *et al.*, 2021, p. 119). Au Québec, le Centre d'études sur les médias (CEM) précise, dans un rapport datant de 2022, que quelque 52 % des lecteurs réguliers du Québec s'informent exclusivement par le numérique. « La proportion des lecteurs réguliers qui n'utilisent maintenant que l'ordinateur, la tablette ou le téléphone cellulaire a gagné 19 points de pourcentage depuis 2017 » (Giroux, 2022, p. 35). Le pourcentage est en hausse dans pratiquement tous les groupes d'âge.

Ce déplacement de l'attention a pour effet de déplacer les investissements publicitaires dans cet univers numérique. Tout simplement, l'argent a changé de mains et se retrouve désormais dans les poches des GAFAM (Google, Amazon, Facebook, Apple, Microsoft). Les chiffres ne mentent pas. En 2003, c'était à peine 3 % du marché publicitaire canadien qui se trouvait en ligne, un pourcentage qui a grimpé à 53 % en 2019 selon le CEM (2020). Toujours en 2019, 78 % des investissements publicitaires numériques au Canada

ont été enregistrés chez Google et Facebook (Roy, 2020). De plus, en même temps que le numérique gagnait en importance, tous les autres secteurs perdaient des parts du marché.

Médias	2003	2013	2019
Hebdomadaires	10 %	7 %	4 %
Quotidiens	31 %	20 %	8 %
Magazines	9 %	6 %	1 %
Radio	14 %	13 %	11 %
Télévision	36 %	31 %	23 %
Numérique	1 %	24 %	53 %

Tableau 1. Parts du marché publicitaire au Québec (CEM, 2020)

Ces chiffres (tableau 1) sont encore plus révélateurs lorsqu'on tient compte de la dépendance des médias traditionnels envers la publicité. En effet, toujours selon le CEM, « les stations de radio y puisent 97 % de leurs revenus, la télévision généraliste 85 %, les hebdomadaires régionaux 93 %, les quotidiens 70 % et les magazines 44 % ».

Même si les médias sont de plus en plus présents dans l'univers virtuel (sites Internet, applications pour mobiles) et que cela entraîne des dépenses supplémentaires pour eux, ils ne récoltent que 7 % des dollars consacrés aux annonces numériques (CEM, 2020).

Ce chiffre ne fait aussi que descendre depuis 2012, année au cours de laquelle les médias traditionnels récoltaient 16 % des investissements publicitaires sur Internet. Le déplacement des revenus publicitaires vers les GAFAM a eu pour effet de diminuer la capacité des médias à effectuer leur travail.

II. La bataille de la conversation publique et des données publicitaires

« Les médias traditionnels ont perdu à jamais le monopole de la conversation publique », souligne le directeur du *Devoir* Brian Myles (Proulx, 2020, p. 10). Il ajoute que :

Pendant une grande partie du XX^e siècle, les coûts d'acquisition d'une station de radio, d'une station de télé ou d'un journal étaient tellement élevés que cela concentrait la propriété des médias entre peu de mains. Cela rendait difficile l'émergence de voix alternatives ou discordantes (Ibid.).

L'arrivée d'Internet et, plus tard, des réseaux sociaux numériques a eu pour effet de démocratiser le « savoir et l'accès à la parole publique », poursuit Myles. Ce faisant, n'importe qui peut devenir un « communicateur public » (*Ibid.*). Cela inclut également les annonceurs qui peuvent utiliser ces plateformes pour rejoindre leur public directement sans passer par les médias traditionnels. Les sources d'information se sont donc multipliées venant diminuer la part des organisations médiatiques dans le débat public. Les réseaux sociaux numériques comme Facebook, Twitter et LinkedIn, ainsi que les algorithmes qui les régissent, ont fait apparaître ce que le chercheur Julien Nocetti nomme « l'horizontalité des sources ». Cela fait en sorte que le contenu journalistique a la même valeur que tout autre contenu publié sur ces réseaux, et ce, peu importe, qui le publie.

Devant désormais passer par un intermédiaire pour rejoindre leur public — qui utilise à hauteur de 55 % les réseaux sociaux numériques pour s'informer —, les médias traditionnels n'ont plus la mainmise sur la hiérarchisation des nouvelles. Cela engendre, dans cet univers en ligne, une nouvelle dynamique dans la circulation, la consultation et l'interprétation de l'information (Nechushtai et Lewis, 2019 ; Newman *et al.*, 2021). En effet, les contenus publiés sur les réseaux sociaux numériques sont mis de l'avant automatiquement par l'entremise d'un ou plusieurs algorithmes dont personne ne connaît la recette (Diakopoulos, 2014b). C'est ainsi que les médias ont perdu la bataille de la conversation publique (Proulx, 2020).

En parallèle, les plateformes numériques utilisent les données de leurs utilisateurs pour créer des « profils riches et précis, [ce qui est] exactement ce que recherchent les annonceurs désireux d'optimiser leurs investissements publicitaires » (Proulx, 2020, p. 7). Le problème est que même si les médias tentaient de reproduire ce modèle, sur leur site Internet, par exemple, la quantité de données dont ils disposent n'est pas la même. Ainsi, il est impossible pour les médias de rivaliser directement avec cette technique. De même, cet usage de masse des réseaux sociaux numériques par les citoyens crée un cercle vicieux où plus les utilisateurs fréquentent les plateformes, plus celles-ci récoltent des données sur eux, ce qui leur permet d'offrir du contenu de mieux en mieux ciblé qui incite les utilisateurs à y revenir.

Bref, ces quatre batailles ont des incidences sur le contenu proposé à l'utilisateur alors que le contenu journalistique doit rivaliser avec tous les autres contenus pour atteindre son public. Qui plus est, non seulement les médias ne savent pratiquement rien du fonctionnement des algorithmes de ces plateformes, elles sont à la merci de changements fréquents dans ceux-ci (Rashidian *et al.*, 2020, p. 12).

Les médias locaux et régionaux sont ceux qui ont le plus souffert des bouleversements décrits jusqu'ici. C'est ce qui explique pourquoi le mémoire cherchera à déterminer la place que l'algorithme de l'agrégateur de nouvelles Google Actualités leur réserve dans ses résultats de recherche.

Il sera question dans les prochains chapitres, de la problématique en lien avec les changements dans les habitudes d'information des citoyens, de la question générale de ce mémoire ainsi que ses pertinences scientifique et sociale. Le cadre théorique, incluant les objectifs de la recherche, ses questions spécifiques et ses hypothèses, sera détaillé. La méthodologie utilisée pour mettre en place la recherche et les techniques de collecte et d'analyse seront aussi abordés, de même que les résultats de l'audit algorithmique mené sur Google Actualités durant la deuxième moitié de l'année 2021.

CHAPITRE 1

PROBLÉMATIQUE

Plusieurs études en communication ont été consacrées au rôle de « *gatekeeping* » historiquement joué par les journalistes, mais désormais assumé de plus en plus par la machine par l'entremise des algorithmes des plateformes numériques. Par « *gatekeeping* », il est question ici du processus par lequel quelqu'un ou quelque chose filtre une information avant qu'elle se rende au public (Bro, 2019). Ce type de filtrage, subjectif, est fréquent dans les salles de rédaction, souvent en lien avec la ligne éditoriale de la publication. Si une contrainte d'espace limite les contenus affichés dans les médias imprimés, l'apparition des sites de nouvelles sur Internet permet une diffusion en continu « sans restrictions sur la longueur et la quantité d'articles » (Bro, 2019, p. 3). Néanmoins, ce ne sont pas tous les contenus qui sont découverts étant donné la quantité d'informations disponibles. Des agrégateurs de contenu algorithmiques ou humains ont donc vu le jour pour aider à filtrer le contenu diffusé en ligne, que ce soit sur les plateformes numériques ou directement sur les sites de nouvelles, comme l'outil Sophi du *Globe and Mail* (The Globe and Mail Inc, 2020). Le rôle de « *gatekeeping* » a donc évolué. Il n'est plus seulement mis à contribution avant la production des contenus, mais bien une fois que ceux-ci sont produits.

Un agrégateur de contenu agit comme intermédiaire entre le public et le média. Cette relation qui date du début des années 1990 s'inscrit dans l'infomédiation, un concept pensé « dès les années 1970 comme un recours à la médiation des machines pour gérer une surcharge d'information grandissante avec les réseaux informatisés, dépassant les capacités de l'esprit humain » (Rebillard et Smyrnaio, 2019, p. 257). Celle-ci « comporte des composantes à la fois sociale (guidage des internautes), économique (commercialisation, notamment publicitaire), et éditoriale (“packaging” ou mise en forme des contenus) » (*Ibid.* p.257). Elle est vite devenue « l'activité centrale dans les filières se développant sur Internet » (Rebillard et Smyrnaio, 2019, p. 258). Les plateformes de Facebook, Google Actualités, Apple News et autres se « sont progressivement imposées comme des points de passage incontournables, situés entre les producteurs et les éditeurs de contenu d'un côté et les publics de l'autre » (Smyrnaio et Rebillard, 2011, p. 7).

Les algorithmes qui constituent ces infomédiaires, que ce soit le fil de nouvelles de Facebook, de Google Actualités ou d'Apple News, décident du contenu proposé à l'utilisateur (Trielli et Diakopoulos, 2019). Ils font une curation de masse des contenus sur leurs plateformes et possèdent un pouvoir éditorial au même

titre qu'un média traditionnel lorsqu'il effectue des choix éditoriaux dans le type de contenu qu'il présente (Mosco, 2017). Si les médias traditionnels font des choix en fonction de leur ligne éditoriale, le processus décisionnel des algorithmes des plateformes numériques, lui, est pratiquement inconnu. Ce manque de transparence fait que ces algorithmes sont considérés comme des boîtes noires (Diakopoulos, 2014b).

D'ailleurs, les médias traditionnels interagissent sur les plateformes numériques sans trop connaître la portée de leurs contenus après avoir appuyé sur « publier » (Mosco, 2017 ; Rashidian *et al.*, 2019). L'une des seules façons de comprendre le processus algorithmique des plateformes, outre le fait d'avoir des lois favorisant la transparence, est de se rabattre sur la recherche empirique. C'est le principal objectif de cette recherche. La reddition de compte algorithmique¹ analyse les résultats de la curation algorithmique afin de mieux comprendre la façon dont les infomédiaires décident du contenu affiché et d'en faire ressortir d'éventuels biais (Diakopoulos, 2014b).

Ce type de recherche, publié surtout aux États-Unis (Diakopoulos, 2013 ; Soltani, 2012), est encore rare au Québec. Il est important de préciser qu'il ne permet pas d'établir avec certitude le fonctionnement des algorithmes. Il sert plutôt à dégager les grandes lignes de ce fonctionnement. La reddition de compte algorithmique consiste à mieux comprendre le processus décisionnel des algorithmes et de mettre en lumière d'éventuelles iniquités ou erreurs de la machine.

Les algorithmes sont des boîtes noires, dont ni le fonctionnement ni les données qui les nourrissent, ne sont généralement rendus publics. Ce faisant, il est fort probable que l'algorithme perpétue des stéréotypes dans sa curation des résultats (Castets-Renard, 2018). L'algorithme apprend à classer et à répertorier les éléments qui lui sont fournis à partir des définitions et des critères que les personnes qui le programment lui soumettent. De ce fait, il y a un risque d'introduire des biais dans le processus (Diakopoulos, 2014b). Ainsi, les algorithmes ne sont pas neutres, malgré un certain fétichisme selon lequel la technologie est plus objective que l'humain (Broussard, 2019 ; Castets-Renard, 2018). Au contraire, les algorithmes peuvent reproduire des biais qui se trouvent dans les bases de données fournies pour l'alimenter (Castets-Renard, 2019).

1 Traduction de l'expression anglaise *algorithmic accountability*.

Une enquête du média en ligne *ProPublica* a montré que des systèmes utilisés aux États-Unis pour évaluer le risque de récidive des accusés étaient discriminatoires envers les personnes noires (Angwin *et al.*, 2016). Chaque décision de ces outils est accompagnée d'une cote entre un et dix. Plus le chiffre est grand, plus le risque de récidive est censé être élevé. Les journalistes ont analysé les cotes de 7 000 personnes arrêtées en 2013 et en 2014 dans le comté de Broward, en Floride. Ils ont remarqué que les personnes noires se voyaient généralement attribuer des cotes plus élevées que les personnes blanches pour des crimes similaires. Les journalistes ont poussé l'analyse plus loin en regardant si les accusés avaient commis de nouveaux crimes dans les deux ans suivant leur arrestation — le barème utilisé par les créateurs de l'algorithme. Ils ont remarqué que les prédictions de l'outil étaient exactes 61 % du temps.

But blacks (44.9%) are almost twice as likely as whites (23.5%) to be labeled a higher risk but not actually re-offend. [The assessment tool] makes the opposite mistake among whites (47.7%): they are much more likely than blacks (28.0%) to be labeled lower risk but go on to commit other crimes (*Ibid.*).

Autrement dit, parmi les personnes noires ayant reçu une cote élevée, 44,9 % n'ont pas récidivé contre 23,5 % pour les personnes blanches. À l'opposé, 47,7 % des personnes blanches ayant reçu une cote faible ont récidivé contre 28,0 % pour les personnes noires. Les journalistes concluent également que les crimes passés ou le type de crime pour lequel les accusés ont été arrêtés n'ont pas d'influence dans la disparité des résultats. La firme qui a mis au point le système a contesté les résultats de *ProPublica* lorsque les journalistes l'ont contactée. La société n'a pas accepté de rendre public le fonctionnement exact de son algorithme nommé COMPAS prétextant le secret industriel. Elle a néanmoins précisé que le programme utilisait divers critères comme le niveau d'éducation et le fait que l'accusé avait un travail ou non. La cote est construite en partie sur un questionnaire de 137 questions soumises à l'accusé ou à partir de son dossier criminel. La race ne fait pas partie des questions.

Si des logiciels utilisés par la justice comportent des biais, il y a lieu de se demander si des biais n'influencent pas également la curation et la diffusion algorithmique de l'information. En d'autres mots : les algorithmes qui distribuent les contenus journalistiques en ligne sont-ils subjectifs? La question mérite d'être posée, puisque la subjectivité s'oppose à l'une des valeurs cardinales du journalisme : l'objectivité.

Bien sûr, la quête de l'objectivité longtemps mise de l'avant par le journalisme occidental est un idéal inatteignable. Comme le précise Phillipa Chong :

Even if fact-based reporting remains a popular benchmark of good journalism, media scholars have pointed out that news-making is intrinsically selective, partial, and thereby subjective in its practice (2019).

Ces biais en soi ne sont pas catastrophiques lorsque l'article partage fidèlement des faits « et des idées [...] sans contraintes ni entraves » (Fédération professionnelle des journalistes du Québec, 1996). Dans le préambule de son guide déontologique, la FPJQ précise que :

Le rôle essentiel des journalistes est de rapporter fidèlement, d'analyser et de commenter le cas échéant les faits qui permettent à leurs concitoyens de mieux connaître et de mieux comprendre le monde dans lequel ils vivent [...]. Les journalistes servent l'intérêt public et non des intérêts personnels ou particuliers (Ibid.).

Il y a donc une notion de transparence dans cette citation. Elle sous-entend que les journalistes doivent se placer au-dessus de leurs intérêts personnels et éviter le plus possible de mettre de l'avant leurs biais, même si les journalistes font des choix éditoriaux selon la ligne de leur média dans le cadre de leur rôle de « *gatekeeping* ».

Les algorithmes, de leur côté, tendent à favoriser les contenus jugés « populaires », les proposant en masse aux utilisateurs des plateformes. De ce fait, comme le contenu est filtré par un algorithme, comportant possiblement des biais, mettant de l'avant le « populaire » et où toutes les informations se valent, il peut devenir difficile pour le citoyen de distinguer le fait de l'opinion. Cela a pour effet que « le mensonge, la rumeur et les ragots se répandent quasi instantanément » sur ces plateformes, ce qui peut avoir des répercussions sur les démocraties en plus de « décrédibiliser le concept d'information » (Nocetti, 2017, p. 139). Étant désormais dans un univers qui tend à se numériser, le nombre d'algorithmes ne fait qu'augmenter influençant de plus en plus les utilisateurs dans leurs décisions (Diakopoulos, 2014a ; Newman *et al.*, 2020). Il convient donc de les analyser pour mieux les comprendre et proposer différentes modifications.

1.1 État de la recherche sur la curation algorithmique de l'information

La recherche empirique sur la relation entre les médias traditionnels et les plateformes numériques montre que les algorithmes ont tendance à promouvoir les articles jugés « populaires ». En effet, plusieurs études effectuées sur l'information présentée par Apple News, par le moteur de recherche de Google et

par Google Actualités, en sont venues à cette conclusion en ajoutant même qu'il y a un manque flagrant de diversité dans les sources présentées (Bandy et Diakopoulos, 2020 ; Fischer *et al.*, 2020 ; Nechushtai et Lewis, 2019 ; Trielli et Diakopoulos, 2019).

Dans la recherche effectuée sur Apple News (Bandy et Diakopoulos, 2020), les chercheurs ont analysé et comparé les articles présents dans les sections *Trending Stories* (« populaires »), sélectionnées par la machine, et *Top Stories* (« à la une ») où la curation est faite par l'humain. Ils ont observé que la localisation et la personnalisation ont très peu d'effet dans le type de contenu proposé à l'utilisateur par l'algorithme. Selon les chercheurs, ce type de décision de la part de l'algorithme d'Apple permet de diminuer les risques de bulles de filtre et de chambres d'écho (Haim *et al.*, 2018 ; Pariser, 2012). Toutefois, les sources présentées sont souvent les mêmes, ce qui semble contredire l'effet espéré. Par ailleurs, la section produite par l'algorithme a proposé davantage de *soft news*² où le contenu est axé sur les célébrités et le divertissement, alors que la partie sélectionnée par des éditeurs humains présentait plutôt du *hard news*³ avec des articles sur l'actualité internationale et politique. De plus, les trois sources les plus souvent présentées dans la section « populaires » représentaient 45 % des occurrences totales, contre 23,7 % pour « à la une ». Les humains proposent donc une plus grande diversité de sources que l'algorithme.

La recherche s'attardant au cas de la section *Top Stories* (« actualités ») du moteur de recherche Google a elle aussi conclu que les mêmes sources sont le plus souvent mises de l'avant (Trielli et Diakopoulos, 2019). L'étude montre que les 20 sources les plus souvent répertoriées dans les résultats comptent pour 52,1 % des impressions totales. Les trois premières (CNN, *The New York Times* et *The Washington Post*) concentrent même 23 % des nouvelles présentées aux utilisateurs. Pour les chercheurs, ces résultats concordent avec d'autres recherches qui ont montré la domination de CNN et du *New York Times* lors de différents audits algorithmiques aux États-Unis. Ils mentionnent par ailleurs que les sources présentées sont majoritairement des médias nationaux, les médias locaux ou régionaux se retrouvant en minorité. Malgré ces résultats, les auteurs précisent qu'il est difficile de déterminer si les résultats obtenus sont la

2 Ce terme fait référence à une nouvelle qui n'est pas politique, qui est épisodique et qui se concentre sur les conséquences individuelles tout en étant émotionnelle et personnelle. (Reinemann *et al.*, 2012)

3 Ce terme fait référence à une nouvelle politique, thématique et qui se concentre sur les conséquences sociales d'un événement tout en étant impersonnelle et sans émotion. (Reinemann *et al.*, 2012)

preuve que l'algorithme a un biais dans sa curation ou si les sources présentées sont plus habiles pour mettre leur contenu de l'avant grâce à un référencement⁴ efficace.

Deux recherches effectuées spécifiquement sur Google Actualités en viennent sensiblement au même constat (Fischer *et al.*, 2020 ; Nechushtai et Lewis, 2019). Nechushtai et Lewis (2019) ont étudié les résultats présentés par Google Actualités à 168 participants qui recherchaient des nouvelles sur « Hillary Clinton » et « Donald Trump » durant la campagne présidentielle américaine de 2016. Ils ont conclu que les allégeances politiques et la localisation ne semblaient pas avoir d'influence sur les résultats présentés par l'algorithme. Les participants étaient pratiquement tous exposés aux mêmes sources. En effet, les cinq médias les plus souvent proposés (*The New York Times*, CNN, Politico, *The Washington Post*, HuffPost) représentaient 49 % des recommandations totales. Les auteurs concluent que dans certains cas l'algorithme présente le même contenu à tous les utilisateurs. Selon les experts, le référencement pourrait avoir un effet dans les résultats de recherche présentés.

La deuxième recherche publiée par Fischer *et al.* (2020) a analysé les résultats affichés par l'agrégateur à la suite d'une requête soumise dans sa barre de recherche. Les requêtes avaient la forme suivante « <mot-clé> near <comté> » et étaient formulées automatiquement par un programme informatique. Les mots-clés provenaient d'une liste de 32 termes identifiés par les auteurs comme étant de nature locale (*governor, police, death, weather*) ou générale (*politics, syria, shutdown, president*). Les comtés étaient chacun des quelque 3 143 comtés des États-Unis. Lors de chacune des saisies, le programme informatique récoltait les 100 premiers résultats. Un total de 12,29 millions de résultats a été recueilli.

Les chercheurs ont utilisé le coefficient de Gini⁵ pour déterminer s'il y avait des inégalités dans la présentation des résultats. Cela leur a permis de répondre à leur question principale de recherche à savoir

4 Traduction de l'expression anglaise *search engine optimization* (SEO).

5 « L'indice (ou coefficient) de Gini est un indicateur synthétique permettant de rendre compte du niveau d'inégalité pour une variable et sur une population donnée » (INSEE, 2020). Le résultat est un chiffre compris entre 0 à 1, où 0 signifie une égalité parfaite et 1 une inégalité parfaite. Généralement, la mesure est utilisée pour déterminer les écarts de richesse au sein d'un pays. Les auteurs de l'étude s'en sont servis pour mesurer le degré de diversité des sources d'information présentées dans les résultats de recherche de Google Actualités. « *The Gini Index tells us whether a few outlets dominate search results or whether outlets appear at similar levels* » (Fischer et

si Google Actualités pouvait contribuer au déclin des médias locaux aux États-Unis. Ils ont constaté que les médias nationaux sont davantage mis de l'avant par l'algorithme de la plateforme au détriment des médias locaux. Le constat vient du fait que les requêtes plus générales n'affichent pratiquement que des reportages de médias nationaux. Pour que du contenu local soit affiché, les requêtes doivent être liées à la gouvernance locale avec des mots-clés comme « maire » ou « police ».

Ils précisent que l'algorithme est une boîte noire, qu'il est impossible de savoir si Google Actualités tient compte du référencement de chaque média dans la présentation de ses résultats. Le fait, pour Google Actualités, de favoriser les médias nationaux au détriment des médias locaux et régionaux semble donc être un biais inhérent à l'algorithme. Cette étude est celle dont le présent mémoire s'inspire.

Un **éc**Erreur! Signet non défini. onomiste de chez Google, Evan Magnusson, considère cependant que la recherche de Fischer *et al.* comporte plusieurs lacunes dans son interprétation des résultats et dans son usage du coefficient de Gini pour y calculer les inégalités (2020). Premièrement, il souligne le fait que Fischer *et al.* n'aient pas produit le même nombre de résultats pour chacun des termes calculant donc les inégalités à partir de bases différentes. En effet, 10 des 16 mots-clés locaux n'ont été enregistrés que lors d'une ou deux saisies sur un total de trois, alors que 15 des 16 mots-clés nationaux ont été inclus dans les trois saisies qu'ils ont effectuées. Dans leur rapport, Fischer *et al.* précisent que les résultats manquants étaient tout simplement corrompus.

Deuxièmement, Magnusson mentionne que le calcul relatif au coefficient de Gini aurait dû être fait par comté, soit de la même façon que les résultats ont été récoltés, et non pas à la grandeur des États-Unis. Cette méthode aurait permis de tenir compte des spécificités propres à chaque région et de mieux ancrer dans les faits les résultats obtenus. Ce faisant, l'employé de Google obtient des résultats où les médias locaux sont moins marginalisés que dans l'étude de Fischer *et al.* (Magnusson, 2021).

al., 2020, p. 3). Un coefficient qui se rapproche de 1 signifie que Google Actualités privilégie un petit nombre de médias (ou de catégories de médias : local/régional par rapport à national) dans ses résultats.

Cette réponse de l'économiste de Google semble s'inscrire dans les efforts publics de l'entreprise pour soutenir les médias locaux. En août 2021, elle a organisé un « Community News Summit » afin d'aider des médias locaux à accroître et monétiser leur audience. Olivia Ma, directrice de la Google News Initiative, y a dit que « *community journalism is key to [help organize the world's information and make it accessible to everyone]* » (Google News Initiative, 2021).

1.2 La réalité de la presse locale

Le changement dans les façons de s'informer des populations et le déplacement des investissements publicitaires ont eu des effets néfastes sur les médias locaux. Plusieurs localités dans le monde ont vu leurs médias locaux disparaître ou réduire leurs effectifs. Par ailleurs, leur importance a décliné dans les conversations publiques. Pourtant, les médias locaux ont une importance dans leurs communautés. Ils augmentent le taux de participation aux élections, enquêtent sur de possibles corruptions et surtout tiennent les gouvernements responsables de leurs actions, au même titre que les médias nationaux. « Sans ces petits journaux locaux, ces stations de radio ou de télévision communautaires, les citoyens vivant en région ne sauraient pas ce qui se passe chez eux » (Martel, 2018, p. 18), sans compter qu'ils jouent également un « rôle de gardien de la mémoire collective au sein de la société » (*Ibid.*, p. 36).

Aux États-Unis, une étude montre que les effectifs des salles de nouvelles des médias locaux ont été réduits de moitié entre 2000 et 2013 passant de 50 journalistes en moyenne à 25 (Jennings et Rubado, 2019, p. 3). La partie quantitative de la recherche permet de constater qu'à tous les endroits où les salles de rédaction ont diminué, le taux de participation et la compétition lors des élections ont suivi la même courbe. La partie qualitative de l'étude, basée sur des entrevues avec des journalistes œuvrant dans le secteur local, fait remarquer qu'il est difficile de dissocier la réduction des salles de nouvelles et les changements dans l'industrie causés par la technologie et la hausse de la demande du numérique (*Ibid.*, p. 6). En 2018, une étude du *Reuters Institute for the Study of Journalism* sur les médias locaux en Europe mentionne que les petits médias ne peuvent pas rivaliser avec les géants du web qui ont accaparé les revenus publicitaires en étant moins coûteux et plus efficaces pour cibler les utilisateurs (Jenkins et Nielsen, 2018, p. 7). La journaliste Margaret Sullivan mentionne que plusieurs études constatent la baisse dans l'implication politique locale lorsque les médias locaux disparaissent, ainsi qu'une augmentation des emprunts des municipalités et une diminution dans l'efficacité des gouvernements de proximité (Sullivan, 2020, p. 22) :

Many of the same discouraging results follow when dependable news sources cut their staffs or disappear altogether: less civic engagement, more political polarization, more potential for government corruption. These are global trends, and global problems (*Ibid.*, p. 79).

Le Canada et le Québec n’y échappent pas. Au Québec, entre 1998 et 2018, le nombre d’hebdomadaires a chuté de plus de la moitié passant de 200 à 88, selon Influence communication (Infopresse, 2018). L’Association canadienne des radiodiffuseurs mentionnait pour sa part que plus de 200 stations de radio et de télévision pourraient disparaître au pays au cours des trois prochaines années (Agence QMI, 2020).

Le Local News Research Project de l’Université Ryerson⁶ nous apprend pour sa part que 450 organisations médiatiques locales ont fermé leurs portes entre 2008 et le 1^{er} décembre 2021 dans 324 communautés au Canada (Lindgren et Corbett, 2021, p. 3). Du lot, 74 se trouvent au Québec, à égalité au deuxième rang avec la Colombie-Britannique, 60 derrière l’Ontario. Pendant cette période, 177 nouveaux médias ont été lancés dans 125 communautés au pays. Le Québec se situe au 4^e rang avec 14 nouveaux médias. Depuis le 11 mars 2020, soit le jour où l’Organisation mondiale de la santé a déclaré la pandémie de la COVID-19, 64 organisations ont toutefois mis la clé sous la porte de façon permanente partout au Canada (*Ibid.*).

Selon April Lindgren, qui pilote le Local News Research Project, « *the danger [...] is that the trend of net loss of small-market news outlets will lead to pockets of “local news poverty” across the country, threatening the wellbeing of communities and the functioning of local democracies* » (Ryerson University, 2021). Cette tendance entraîne ce qu’elle appelle un appauvrissement des nouvelles locales⁷ : « *Situations where communities don’t have access to the timely, relevant, independently produced, verified news they need to navigate daily life* » (*Ibid.*).

La chercheuse préfère utiliser le terme « appauvrissement » plutôt que l’expression « désert d’information » généralement utilisée pour illustrer le problème. Selon elle, cette dernière décrit mal la situation, car le journalisme local fait face à un problème de qualité de l’information, non pas seulement

6 Aujourd’hui connue sous le nom d’Université métropolitaine de Toronto.

7 Traduction de l’expression *local news poverty*.

de quantité d'information. Cela reflète mieux les nombreuses compressions dans les salles de rédaction (Hoffman, 2020).

Entre 2005 et 2015, selon Statistique Canada, le nombre de journalistes aurait diminué de 10 % au Québec. Mais la diminution n'est pas la même partout. Alors qu'elle a été de 2,2 % dans la grande région de Montréal, elle a atteint 16 % dans les autres régions (Martel, 2018, p. 201). Au-delà des chiffres, la réalité de la presse écrite locale au Québec s'explique aussi par une guerre entre deux géants des médias qui a culminé dans un éclatement complet en 2017. En effet, « entre 2010 et 2014, Sun Media, une division de Québecor, et TC Media (Transcontinental), se sont disputé le marché des hebdomadaires, marché lucratif d'un point de vue publicitaire » (Martel, 2018, p. 39). Cette concurrence a mené à la création de plusieurs médias, faisant soudainement du Québec un endroit « excessivement bien desservi par de nombreux titres [...] trop bien desservi, même, pour la capacité économique du marché » (*Ibid.*, p. 40).

Cette guerre a fini par provoquer une chute du prix des publicités. À Laval, lorsque Québecor a lancé *L'Écho de Laval*, une page de publicité s'y vendait 500 \$ dans un marché où le *Courrier Laval* la vendait 3 000 \$. Si le *Courrier Laval* a « résisté », des dires de l'éditeur de l'époque, Claude Labelle, certains plus petits journaux n'y sont pas parvenus. Cet affrontement a fait diminuer le nombre de journaux indépendants (Collard, 2012).

Québecor a finalement capitulé en 2014 : 74 titres ont été rachetés par TC Media. **TErreur! Signet non défini.**rois ans plus tard, c'était au tour de TC Media de se retirer du marché de la presse écrite, mettant en vente 93 titres qui trouveront preneur chez différents joueurs créant de nouveaux groupes de presse plus petits (Transcontinental, 2017). « Ces multiples ventes et acquisitions ont ramené l'industrie dans une situation similaire à ce qui prévalait il y a une dizaine d'années, c'est-à-dire que les médias appartiennent à plusieurs intérêts plutôt qu'à deux gros joueurs » (Martel, 2018, p. 47).

Ces différents problèmes économiques de la presse locale, causés en partie par la baisse des prix de la publicité lors de la guerre entre Québecor et Transcontinental, affectent les hebdomadaires pour qui la publicité représente 93 % des revenus contre 70 % pour les quotidiens. Par conséquent, la chute de 10 % (2003) à 4 % (2019) dans leur part du marché publicitaire au Québec est substantielle (CEM, 2020). Tout cela, alors que « 81 % des Canadiens affirment que les nouvelles locales sont importantes pour eux » (CRTC, 2016) et que l'actualité locale et régionale est la rubrique d'information qui suscite le plus d'intérêt

(Charlton *et al.*, 2013, p. 40). Malgré cela, l'information locale continue de décliner, et ce, même si l'intérêt du public pour celle-ci semble toujours présent.

Les médias régionaux ont vécu une crise à la fois médiatique, « induite par les transformations technologiques de grande ampleur, [et] économique mondiale qui est venue fragiliser davantage des médias en pleine mutation à l'ère du numérique » (Bizimana et Kane, 2019, p. R146).

À toute cette question de la presse locale, il faut aussi prendre en compte l'enjeu de la « montréalisation » de l'information. En effet, les médias favoriseraient l'information en provenance de Montréal ou concernant celle-ci au détriment des enjeux régionaux et locaux (Carignan, 2011). Cela a comme effet de diminuer l'importance des régions dans la couverture médiatique.

1.3 Et si le lecteur était en cause?

Depuis le début, ce mémoire fait le lien entre les transformations numériques et leurs effets sur le journalisme en général, plus précisément sur la presse locale. Mais serait-il possible que ce soit aussi parce que les lecteurs quittent le navire? C'est du moins l'opinion du journaliste senior chez *Politico* Jack Shafer (2021). Au cœur de son argument : un rapport du Pew Research Center publié en 2009 mentionnant que 42 % des Américains ne seraient pas affectés (« *not at all* ») ou très peu (« *not much* ») par la disparition de leur média local, et cela bien que 74 % croient que la disparition de la presse locale affecterait la vie civique.

Selon Shafer, le problème en serait donc un de demande, à savoir que les gens ne veulent pas payer ou même consulter les médias locaux bien qu'ils les jugent importants : « *It's not that nobody wants to read local news; it's just that not enough people do to make it a viable business* », écrit-il après avoir mentionné que seulement 14 % des répondants à une étude datant de 2018 du Pew Research Center avaient donné de l'argent à un média local.

En 2022, ce sont 19 % des Américains qui ont dit avoir payé pour des nouvelles en ligne (Newman *et al.*, 2022, p. 113). Au Canada, ce chiffre glisse à 15 %, dont 16 % pour du contenu en français (*Ibid.*, p. 118), mais ces habitudes sont peut-être en train de changer. Les Coops de l'information, qui produisent six journaux régionaux, ont instauré un mur payant en 2020. Après un an, 25 000 personnes ont choisi de s'abonner. Il reste à voir si le groupe sera en mesure de faire croître ce chiffre (Roy, 2021).

Shafer croit également que le déclin des médias locaux est lié à la baisse d'audience expliquée en partie par la multiplication des supports technologiques, au fil des ans, permettant la diffusion de l'information. Le problème ne serait pas seulement l'Internet, mais aussi la radio, la télévision et toutes ces autres plateformes non informationnelles qui accaparent l'attention des citoyens. L'attention devient ainsi plus difficile à capter. Le média local ne se bat plus seulement avec son concurrent direct, mais aussi avec des plateformes qui n'ont aucun lien avec son objectif. Tout cela parce que l'attention est limitée.

1.4 Question générale de recherche et pertinence

Ce mémoire a l'intention de répondre plus spécifiquement à la question suivante :

- Quelle est la part des médias locaux et régionaux québécois dans les requêtes formulées sur la plateforme Google Actualités?

Autrement dit, en prenant compte de tous les enjeux énumérés dans la problématique, il est question d'analyser le comportement de l'algorithme de Google Actualités dans le but de mieux comprendre les résultats qu'il présente à l'utilisateur québécois.

Ce projet s'inspire de l'étude s'étant attardée à cette question aux États-Unis (Fischer *et al.*, 2020). L'objectif ici est de reproduire cette dernière, tout en tenant compte des particularités propres au Québec, notamment sa langue, son territoire et ses enjeux comme la « montréalisation » de la nouvelle. Bien que tous les résultats sont analysés, une attention plus particulière est portée sur les médias locaux et régionaux.**Erreur! Signet non défini.**

La pertinence scientifique de ce mémoire est de voir s'il y a des différences dans la curation algorithmique en comparant les résultats obtenus avec ceux des États-Unis. De même, comme la reddition de compte algorithmique est pratiquement inexistante au Québec, autant dans la littérature que dans le travail journalistique, ce mémoire permet d'ouvrir la voie à de futures recherches qui voudront s'attarder à ce type de pratique.

La pertinence sociale de ce mémoire repose sur son choix d'analyser plus particulièrement les médias locaux et régionaux. En effet, ceux-ci ne visent pas nécessairement la production de contenu « populaire », ne ciblant qu'une partie précise de la population. Ce faisant, tout porte à croire que leur contenu n'est pas mis de l'avant adéquatement par l'algorithme de Google Actualités, même si les responsables de la Google News Initiative prétendent leur accorder une grande importance. De même, ces médias sont davantage touchés par la transition vers le numérique et la perte de revenus causés par le départ des publicitaires vers le web (Bizimana et Kane, 2019 ; Fischer *et al.*, 2020 ; Proulx, 2020). Toutefois, il est probable, bien

que ce ne soit pas l'hypothèse principale de ce projet, que le problème des médias locaux ne réside pas dans leur mise en valeur inadéquate par Google Actualités, mais plutôt dans le fait que les lecteurs ne soient pas au rendez-vous.

Enfin, la pertinence communicationnelle de ce mémoire repose sur son examen de la curation algorithmique de l'information locale. En effet, les médias régionaux sont essentiels à la survie des démocraties locales. De ce fait, si leur contenu est marginalisé sur les plateformes numériques, il y a un risque que ceux-ci disparaissent venant ainsi affecter les communautés déjà mal desservies par les médias nationaux et que cela témoigne, donc, d'un appauvrissement de l'information locale (Hoffman, 2020 ; Lindgren et Corbett, 2021). S'il s'avère finalement que Google Actualités met de l'avant adéquatement le contenu, ce projet demeure pertinent puisqu'il mènera à un approfondissement des connaissances et à la naissance possible d'une autre étude sur les raisons justifiant le manque d'intérêt des lecteurs envers les contenus locaux.

CHAPITRE 2

CADRE D'ANALYSE

Ce chapitre compte deux sections. Les recherches dont ce mémoire s'inspire ne présentent pas de cadre théorique à proprement parler, se basant plutôt sur une série de concepts qui seront décrits dans la première. La seconde analyse les implications normatives du concept de « *gatekeeping* » et s'attarde aux enjeux de transparence algorithmique en puisant dans le concept de « gouvernamentalité algorithmique ».

2.1 Cadre conceptuel

2.1.1 Algorithme

Le chercheur Nicholas Diakopoulos définit un algorithme de la façon suivante : « *A series of steps undertaken in order to solve a particular problem or accomplish a defined outcome* » (Diakopoulos, 2014b, p. 3). Un algorithme peut autant être produit par un humain qu'une machine. La recette d'un gâteau est considérée comme un algorithme, nous dit Diakopoulos. Déterminer l'ordre du contenu sur la page d'une plateforme numérique l'est aussi.

Les algorithmes gérés par des machines qu'on retrouve dans l'univers numérique sont des boîtes noires. L'analogie du gâteau peut être reprise ici. Prenons par exemple quelqu'un qui mange un gâteau fait par un ami. Ce dernier peut déduire qu'il contient certains ingrédients, mais il ne peut pas déterminer la recette exacte ayant permis sa réalisation à moins d'en obtenir une copie. Toutefois, son ami pourrait refuser de partager son secret, surtout s'il produit son gâteau dans sa pâtisserie. Il en va de même pour les algorithmes numériques. Les plateformes numériques sont des entreprises privées qui refusent généralement de divulguer leurs recettes.

Cependant, un gâteau n'a pas les mêmes conséquences sociales que les plateformes numériques, surtout que la fréquentation de ces dernières ne va pas en diminuant. Il est important de noter, comme le précise Nicholas Diakopoulos, que les algorithmes ne sont pas seulement négatifs. Ils peuvent procurer plusieurs bienfaits à la société. Comme nous l'avons évoqué dans la problématique, il faut toutefois rappeler que ceux-ci peuvent comporter des biais et faire des erreurs (Diakopoulos, 2014b). Le chercheur ajoute également que ce qui demeure inconnu des algorithmes, c'est la façon dont ils exercent leur pouvoir sur les utilisateurs au quotidien.

2.1.2 Audit

C'est donc pour combler ce manque de connaissance concernant les algorithmes que les audits algorithmiques existent. L'objectif de ce type d'audit est de déterminer si l'algorithme comporte certains biais ou présente des discriminations dans ses résultats (Bandy et Diakopoulos, 2020). Plus précisément, « *audit studies are field experiments meaning a research design involving the random assignment to groups in a controlled setting in order to isolate causation* » (Sandvig *et al.*, 2014, p. 6). Par « expérience terrain⁸ », Sandvig et ses collègues précisent que l'objectif est d'avoir des critères et des paramètres d'analyse qui permettent de généraliser la situation à une expérience réelle d'utilisation, sur les plateformes numériques dans le cas qui nous intéresse ici.

Concrètement, les exemples donnés dans la section sur « la mise en valeur du contenu populaire » dans la problématique résument bien ce qui est entendu par audit. Le but est de reproduire une situation réelle et d'observer les résultats obtenus. Il y a donc nécessairement une collecte de données pouvant être réalisée de plusieurs façons. Dans le cas de ce mémoire, il s'agit de la méthode du *scraping audit* où un programme informatique est conçu pour récolter les données (Sandvig *et al.*, 2014, p. 12). L'audit algorithmique postule que tout algorithme mérite d'être analysé. Cette démarche s'inscrit dans ce que Nicholas Diakopoulos nomme la reddition de compte algorithmique (2014b). L'objectif est donc d'analyser une plateforme et de rendre responsable cette dernière si des problèmes sont observés.

Un exemple récent est celui du *Wall Street Journal* avec sa série d'articles nommée « The Facebook Files » (Hagey et Horwitz, 2021). Dans ce dossier, les journalistes ont eu accès à de nombreux documents internes montrant comment un récent changement dans l'algorithme de Facebook ayant pour but d'augmenter les « interactions sociales significatives »⁹ a plutôt eu l'effet inverse en augmentant les interactions négatives. Le plus surprenant dans ce dossier, c'est qu'il a permis de montrer que les dirigeants de l'entreprise étaient au courant. En effet, les chercheurs au sein de Facebook ont envoyé des mémos. « *“Misinformation, toxicity, and violent content are inordinately prevalent among reshares”, researchers noted in internal memos* » (*Ibid.*). Des solutions ont été proposées par les experts de l'entreprise pour pallier ce problème, mais certaines auraient été rejetées par Mark Zuckerberg, le PDG de l'entreprise Meta, parce qu'il

8 Traduction libre de l'expression : *field experiment*.

9 Traduction libre de l'expression *meaningful social interactions*.

s'inquiétait de voir baisser l'engagement des utilisateurs sur sa plateforme. Une des raisons pour changer l'algorithme résidait justement dans le but d'inverser la tendance selon laquelle les utilisateurs interagissaient moins avec le réseau socionumérique. Bref, ce qu'illustre cet exemple, c'est qu'un examen approfondi d'un algorithme consiste à rendre ses concepteurs responsables des conséquences de leur outil.

2.1.3 Référencement

Le référencement est central dans le monde numérique. Il s'agit d'un ensemble de pratiques mises en place par les sites Internet pour les aider à avoir un meilleur positionnement dans les résultats de recherche de Google ou de tout autre moteur de recherche (Carroll, 2018). Elles comprennent l'insertion de métadonnées dans le code HTML d'une page web (mot-clé, description, etc.), l'insertion de liens vers des pages bien référencées, la simplification de la structure du code HTML, etc. Plus de 200 facteurs détermineraient l'ordre du contenu dans les résultats « organiques » (non publicitaires) présentés par un moteur de recherche quand un utilisateur y fait une requête (Dean, 2021). Il y aurait également plus de 500 changements permanents aux critères de référencement par année, selon le site Moz qui les répertorie depuis 2000 (Moz, s. d.). Bien que certains critères du référencement soient connus, l'importance individuelle de chacun de ces critères dans le classement des pages ne l'est pas. Il est donc difficile d'établir avec certitude l'effet de ce référencement dans les résultats lors d'un audit.

Ce concept du référencement trouve une partie de ses racines dans l'initiative de Sergey Brin et Larry Page, les fondateurs de Google, de réformer la façon dont les contenus sont filtrés en proposant de mesurer « la force sociale de la page [analysée] dans la structure du web » (Cardon, 2015, p. 25). Ces derniers vont donc établir un système de méritocratie où « les sites les mieux classés sont ceux qui ont reçu le plus de liens hypertextes venant de sites qui ont, eux-mêmes, reçu le plus de liens hypertextes des autres » (*Ibid.*). Le système PageRank considère « que les liens hypertextes enferment la reconnaissance d'une autorité » qu'elle soit négative ou positive (*Ibid.*). Un marché va naître où des entreprises vont proposer d'améliorer le design et la structure des sites pour que les robots des moteurs de recherche les comprennent mieux et donc les référencent mieux, menant parfois à une autorité artificielle (*Ibid.*, p. 27). Cardon (2015) pose deux critiques au processus PageRank de Google. La première fait part d'effets d'exclusion et de centralisation de l'autorité causée par l'agrégation du jugement des pairs. « À force d'être cités par tous, les plus reconnus deviennent aussi les plus populaires et reçoivent en conséquence le plus de clics », un cercle vicieux dont la mécanique ressemble à celle de la promotion des contenus populaires, décrite dans

la problématique. La deuxième critique aborde l'effet censitaire des mesures d'autorité : « Ne participent au classement de l'information que ceux qui publient des documents comportant des liens hypertextes, comme les détenteurs de sites ou les blogueurs; les autres sont ignorés » (*Ibid.*, p.28). Par « les autres », il est question ici des interactions sur les réseaux sociaux numériques qui ne sont pas promues dans les résultats de recherche sur Google.

En plus du classement par liens hypertextes, Google a ajouté une fonction qui « enregistre le lien sur lequel a cliqué l'internaute parmi la liste de réponses proposées à un mot-clé » (*Ibid.*, p.63). Ainsi, « si l'utilisateur ne revient pas sur la page pour cliquer sur un autre lien, l'algorithme va conclure que l'internaute est satisfait par la réponse apportée » (*Ibid.*).

2.2 Cadre théorique

Abordé dans la problématique, le concept du « *gatekeeping* » est largement utilisé dans la sélection des nouvelles par des journalistes ou, comme dans le cas présent, par des algorithmes (Shoemaker et Vos, 2009, p. 9). Pour rappel, le concept se définit de la façon suivante : le processus par lequel quelqu'un ou quelque chose filtre une information avant qu'elle se rende au public (Bro, 2019).

Le travail associé à ce concept est encore présent dans le quotidien des journalistes, qui doivent filtrer quotidiennement l'actualité afin de déterminer ce qui est digne de mention. Dans la notion du « *gatekeeping* », on peut également inclure la curation faite par les algorithmes qui filtrent les informations sur les médias sociaux numériques comme Facebook, Twitter ou LinkedIn, ou bien sur les plateformes de nouvelles comme Google Actualités, Apple News et autres agrégateurs de contenu. Ceux-ci déterminent selon leurs critères ce qui mérite d'être montré aux utilisateurs et dans quel ordre.

Alors que les journalistes filtrent l'information sur la base d'un jugement humain avant de la publier, les moteurs de recherche (ainsi que Google News) filtrent a posteriori une information déjà publiée sur la base des jugements humains émis par l'ensemble des internautes qui publient sur le web. (Cardon, 2015, p. 26)

La une d'un média papier relève également du « *gatekeeping* » où une ou plusieurs personnes ont filtré le contenu de la journée et ont déterminé ce qu'ils jugeaient le plus pertinent à montrer en premier au public. C'est également le même principe pour les bulletins de radio et de télévision où l'information est ordonnée et rapportée selon l'importance accordée par les journalistes. C'est aussi ce qui s'applique pour la une d'un média sur Internet : une ou des personnes mettent à jour le contenu qui est affiché sur la page d'accueil

faisant ainsi passer les nouvelles de la journée par un premier filtre selon l'importance accordée par la rédaction, encore une fois selon des critères qui sont propres à chaque média. Sur Google Actualités, le principe demeure le même, mais il est fait par une machine qui détermine quels contenus journalistiques sont affichés.

Les critères qui déterminent ce qui franchit les portes du « *gatekeeper* » ne sont ni fixes ni arrêtés dans la littérature. D'un côté, David Manning White, l'un des premiers, avec Kurt Lewin, à théoriser le « *gatekeeping* » en 1950, croit que les valeurs de la personne qui fait la curation ont un rôle à jouer dans le choix d'écartier ou non une nouvelle (Shoemaker et Vos, 2009, p. 6). Dans une analyse de 16 journaux dont la curation est faite par des éditeurs par télégraphe, Walter Gieber statue de son côté, en 1956, que la subjectivité de la personne n'a pas d'influence. « *Personal subjectivity was less important in gatekeeping than structural considerations, including "the number of news items available, their size and the pressures of time and mechanical production" » (Ibid.).*

Bref, ce travail journalistique découle de ce qu'on nomme en anglais le « *newsworthiness* » — ou le contenu « digne d'intérêt », en français. « *Newsworthiness is a cognitive construct that only partially predicts which events make it into the news media and how those events are covered » (Ibid., p. 14).* Bien que les auteurs ne s'entendent pas sur des critères fixes pour reconnaître une nouvelle digne d'intérêt, certains ressortent tout de même du lot :

There are many lists of news attributes, but they generally include some or all of the following: timeliness; proximity; importance, impact, or consequence; interest; conflict or controversy; sensationalism; prominence; and novelty, oddity, or the unusual (*Ibid.*).

De plus, selon Nisbett et Ross,

humans are more likely to store and remember vivid rather than pallid information, implying that vivid events and issues are more likely to enter a channel. Vivid information grabs hold of the imagination and is exciting, whereas pallid information is dry and unappealing. (*Ibid.*, p.15)

Une information est donc plus susceptible d'être transmise au public si elle est « *(a) emotionally interesting, (b) concrete and imagery-provoking, and (c) proximate in a sensory, temporal, or spatial way » (Nisbett et Ross, 1983, p. 45).* Shoemaker et Vos (2009) précisent que les informations soumises par les « *gatekeepers* » ne sont pas dignes d'intérêt en soi, mais le deviennent parce que ces derniers l'ont décidé.

Comme « *gatekeepers* », les journalistes ont donc une influence sur la circulation de l'information alors qu'ils décident ce qui devrait ou non être couvert ou mis de l'avant par les médias. Les agrégateurs de contenu et les médias socionumériques interviennent a posteriori en décidant de manière automatique ce qui devrait ou non se rendre à leurs abonnés. Ces décisions automatisées pourraient ultimement avoir une influence démocratique dans les cas où certaines informations cruciales pourraient être « invisibilisées » par des algorithmes, bien qu'elles fassent l'objet d'une couverture par un ou des médias. C'est ce que ce mémoire tente d'observer en quantifiant les catégories de médias (local et régional, national, hors du Québec et les non-médias) affichés par l'algorithme de Google Actualités lors de recherches. Ceci est aussi en lien direct avec le concept de responsabilité et de gouvernementalité algorithmique.

2.2.1 Responsabilité/gouvernementalité algorithmique

La problématique a bien détaillé les effets de la transformation numérique sur l'économie, dont celle de l'attention qui s'est vue accentuée. Tout cela a fait de la collecte de données au détriment des citoyens une normalité. Présentée comme fragmentée, décontextualisée, désindexée et anonymisée, la donnée générée dans le monde numérique n'est désormais qu'un pur signal, qui lui confère une objectivité inégalée et une « croyance en son potentiel émancipateur » (Ouellet *et al.*, 2015, p. 597). Cette idolâtrie de la donnée transforme notre rapport à celle-ci lui permettant de prendre une place de plus en plus grandissante dans les décisions du quotidien.

La gouvernementalité algorithmique est définie par Antoinette Rouvroy comme « un mode de gouvernement qui se nourrit essentiellement de données numériques » (Gouritin, 2018) et donc qui modélise le monde social à partir de corrélations entre des bases de données anonymes pour créer des profils et établir des politiques de gestion de risque. Ce concept va plus loin qu'une simple application par les gouvernements au niveau politique, elle concerne pratiquement toutes les sphères de la société.

Les gouvernements récoltent [des données] à des fins de sécurité, de contrôle, de gestion des ressources, d'optimisation des dépenses...; les entreprises privées recueillent quantité de données à des fins de marketing et de publicité, d'individualisation des offres, d'amélioration de leur gestion des stocks ou de leurs offres de service, bref, en vue d'accroître leur efficacité commerciale et donc leurs profits...; les scientifiques collectionnent les données à des fins d'acquisition et d'amélioration de connaissances... ; les individus eux-mêmes partagent bénévolement « leurs » données sur les réseaux sociaux, les blogues, les « mailing lists ». (Rouvroy et Berns, 2013, p. 168)

La gouvernementalité algorithmique est donc intimement liée au concept du Big data qui rend possible « les capacités de contrôle et de surveillance » (Ouellet *et al.*, 2015, p. 598). Cette idéologie, et ce rapport à la donnée prôné par le Big data, est vue par ses artisans comme si « c'était le monde qui parlait spontanément à travers les données » (Gouritin, 2018). Autrement dit, c'est comme si la donnée était le reflet du monde, « comme si c'était le réel qui se gouvernait lui-même à travers les données » (*Ibid.*). Qui plus est, « l'application du Big data à l'ensemble des secteurs de l'activité sociale (santé, sécurité publique, finance, marketing, militaire, administration publique, etc.) permettrait de mettre en place des mécanismes de régulation sociale non coercitifs et efficaces grâce à des algorithmes autoapprenants qui seraient en mesure de prendre des décisions de manière automatisée et hautement rationalisée » (Ouellet *et al.*, 2015, p. 604).

Cette prise de décision des algorithmes est souvent prédictive, dans le but de concevoir des profils personnalisés « de consommateurs, de fraudeurs potentiels, de terroristes potentiels » à partir des données récoltées (Gouritin, 2018). Ces algorithmes tentent d'établir des corrélations dans des jeux de données massifs « plutôt que sur des catégories qui correspondent à des groupements humains socialement éprouvés » (*Ibid.*). La tendance est donc inversée : au lieu de chercher à confirmer ou infirmer des hypothèses, l'objectif est désormais de faire parler les bases de données pour établir des corrélations.

Bien que les données soient jugées objectives, car elles sont anonymisées et détachées de leurs origines, elles ne sont pas neutres pour autant. « [Les données] sont toujours des effets, des rapports de force, de domination [et] de situation qu'on n'a pas changée alors qu'on aurait pu les changer. » (*Ibid.*). Ce faisant, « la gouvernementalité algorithmique induit une forme de domination abstraite et dépersonnalisée puisque les sujets n'ont aucune autre liberté que de se soumettre et de s'adapter à cette dynamique objective » (Ouellet *et al.*, 2015, p. 607).

De même, l'algorithme n'est pas non plus objectif, du fait de sa programmation qui relève d'un humain qui y intègre du même coup ses biais (Broussard, 2019 ; Castets-Renard, 2018 ; Diakopoulos, 2014b).

Dès lors que les machines deviennent de plus en plus « autonomes » et « intelligentes », elles restent bien sûr dépendantes de leur design initial, des intentions, scripts ou scénarios en fonction desquels elles ont été imaginées. Elles sont, dès leur conception (et quelles que soient les formes qu'elles prennent ensuite), porteuses des visions du monde, attentes et projections conscientes ou inconscientes de leurs concepteurs (Rouvroy et Berns, 2013, p. 166).

Qui plus est, nombreux sont les exemples où la machine a accentué des stéréotypes étant donné la nature des données qui renferment d'une manière ou d'une autre certains rapports de force. Le cas en 2013 de Latanya Sweeney, une chercheuse en informatique afro-américaine, est très révélateur. Elle a constaté qu'en tapant son nom « dans le moteur de recherche de Google, elle voyait apparaître la publicité "Latanya Sweeney, arrested?" ». Cette publicité réfère à un service de consultation en ligne, instantcheckmate.com, qui permet, entre autres choses, de savoir si les personnes ont un casier judiciaire » (Cardon, 2015, p. 84). Elle remarque également que la publicité n'apparaissait pas lorsqu'elle écrivait le nom de ses collègues blancs. Elle se pose donc la question à savoir si le système publicitaire de Google était discriminatoire. « Après une enquête de rétro-ingénierie systématique sur un très grand nombre de requêtes, Latanya Sweeney a montré que l'algorithme n'a pas besoin d'avoir une intention classificatoire pour produire ce genre d'effets discriminatoires » (*Ibid.*, p.85). L'algorithme se base sur « les régularités statistiques » qui font que les noms des personnes noires sont plus souvent liés à des clics vers des recherches de casier judiciaire. Cardon (2015, p. 85) explique que « le calculateur s'appuie sur les comportements des autres internautes et contribue, "innocemment" si l'on ose dire, à la reproduction de la structure sociale, des inégalités et des discriminations ».

Cet exemple remet de l'avant encore le caractère opaque des algorithmes. Ce processus « reste pour nous mystérieux » alors qu'il « oriente des décisions, appaie des processus automatiques et justifie des choix politiques » (*Ibid.*, p.12). Il convient donc de se poser les questions suivantes :

Quelles sont les données qui servent au calcul? Comment l'information a-t-elle été quantifiée? Quels sont les principes de représentation qui animent le modèle statistique mis en œuvre pour classer tel objet plutôt que tel autre? Qui pilote le codage des calculs et quels sont ses objectifs? (*Ibid.*)

Une façon de les poser est de procéder à un audit des algorithmes afin de mieux les comprendre.

Les algorithmes hiérarchisent les informations, et c'est pour cela qu'ils sont utiles et même essentiels. Mais il est indispensable que les services puissent expliquer à l'utilisateur les priorités qui président aux décisions de leurs calculateurs; et qu'on puisse vérifier, en toute indépendance, que des intérêts cachés, des déformations clandestines ou des favoritismes n'altèrent pas le service rendu. (Cardon, 2015, p. 82)

2.2.2 Et Google Actualités dans tout ça?

C'est sur ce dernier point, celui de l'audit pour mieux comprendre les algorithmes que ce mémoire se concentre. Le précédent est révélateur de la force des biais qui ne sont pas directement inscrits dans la

programmation, mais plutôt entretenus par le comportement des utilisateurs qui eux en ont. Comme la machine se base sur les comportements d'un utilisateur, les risques augmentent. Google Actualités pourrait ne pas y échapper. Il convient donc de poser les questions que pose Cardon ci-dessus et de voir si l'on peut faire confiance à Google Actualités dans sa curation.

Les potentiels biais et les rapports de force sont au cœur de cette recherche où l'analyse du comportement de l'algorithme de Google Actualités permettra d'évaluer le « *gatekeeping* » de la plateforme et d'observer ou non des biais dans son fonctionnement. Si Google Actualités prend la décision de ne pas afficher adéquatement les médias locaux et régionaux comme le postule l'hypothèse générale de ce mémoire, les conséquences d'une telle décision peuvent être catastrophiques pour les démocraties locales. Ce que cette recherche cherche donc à observer, c'est comment l'algorithme de Google Actualités gouverne la diffusion de l'information. Est-ce que le travail seul de la machine suffit à offrir une vision sans biais? Est-ce que le référencement, et donc les décisions des humains derrière la mise en ligne des contenus journalistiques, a une influence dans l'affichage des résultats? Tout cela est lié à l'importance que Google accorde ou non à la donnée générée par l'utilisateur dans sa curation.

2.3 Questions spécifiques et hypothèses de recherche

Pour réaliser un audit algorithmique, concrètement, cette recherche est basée sur une collecte de données effectuée à huit reprises sur le site Google Actualités entre le 3 août et le 11 décembre 2021 dans le but de construire une banque de données et d'en analyser les résultats. Il est donc question d'un audit algorithmique.

La collecte de données est mise en place avec quatre hypothèses de départ découlant de la question générale de recherche :

- Premièrement, les médias locaux et régionaux sont sous-représentés dans les résultats de recherche sur Google Actualités.
- Deuxièmement, les médias locaux et régionaux des grands centres comme Montréal et Québec sont davantage mis de l'avant que les médias locaux des régions éloignées.
- Troisièmement, cette sous-représentation anticipée des médias locaux et régionaux serait le résultat de l'algorithme de Google qui a tendance à promouvoir le contenu jugé « populaire » et

produit par les médias nationaux en ne répondant pas aux besoins informationnels des utilisateurs en région.

- Quatrièmement, considérant la quantité d'information produite, les articles retournés par l'algorithme de Google Actualités sont relativement récents.

CHAPITRE 3

MÉTHODOLOGIE

Cette recherche est purement quantitative. Son objectif est de surveiller le fonctionnement de l'algorithme de Google Actualités et d'établir des tendances. La démarche de collecte de données a été automatisée grâce au langage informatique Python. En utilisant le module Selenium, différents scripts (programmes informatiques¹⁰) ont simulé des recherches sur Google Actualités à partir de différents mots-clés et d'une localisation précise pour recueillir diverses informations dans les résultats de recherche (voir Annexe 1) :

- le nom du média,
- sa position dans les résultats,
- le lien menant à l'article,
- le titre de l'article,
- la date de publication ainsi que
- la différence en nombre de jours entre la date de publication et le jour de la saisie.

Ces informations ont été enregistrées dans un document CSV (*comma separated value*) (voir Annexe 2) dans lequel chaque média enregistré a été classé dans l'une des quatre catégories suivantes :

- média local/régional,
- média national,
- média hors Québec ou
- non-média.

Les critères pour déterminer dans quelle catégorie classer chaque média se basent sur la définition du Local News Research Project :

¹⁰ Les programmes informatiques et les données des résultats de recherche sont disponibles sur Github : https://github.com/NicolasStG/audit_google_news

A local news outlet is defined as a news organization that maintains independence from those it covers, demonstrates a commitment to accuracy/transparency, and is devoted primarily to reporting and publishing timely, originally produced news about local people, places, issues and events in a defined geographic area (Lindgren et Corbett, 2021, p. 2).

Mis simplement, le critère central est la définition géographique de la couverture journalistique. Par exemple, un journal local comme le *Courrier Laval* ou le *Journal Métro* ou un média régional comme *Le Soleil* seront placés dans la catégorie « média local/régional », alors que des organisations médiatiques comme le *Journal de Montréal* et *La Presse* le seront dans la catégorie « média national ». Tout média ou reportage basé en dehors de la province du Québec entrera dans la catégorie « média hors Québec ». Et tous les sites Internet non journalistiques seront classés comme « non-média ».

Il est important, dans notre analyse, de conserver les résultats classés comme « non-média » pour montrer que Google Actualités affiche parfois du contenu qui n'est pas produit par des médias. Bien que cela se produise relativement peu souvent, ces contenus non-journalistiques sont affichés régulièrement par l'agrégateur et ils sont parfois plus nombreux dans les premiers résultats de recherche que dans les derniers, comme on le verra dans la section consacrée à nos résultats. Ultimement, cela peut tromper le public et lui présenter un contenu exempt de démarches journalistiques alors qu'il consulte la plateforme pour lire du contenu provenant de journalistes. La recherche de Fischer *et al.* ne disait pas si certains des résultats de Google Actualités n'étaient pas des médias. Est-ce parce que la version de Google Actualités pour les États-Unis fait un meilleur tri? Difficile de savoir en toute certitude.

Il faut savoir, par ailleurs, que les médias TVA et Radio-Canada pourront compter des résultats dans les catégories « local/régional » (dans les cas où ils font un lien vers un reportage produit par une de leurs stations régionales), « national » (dans les cas où ils font un lien vers un reportage réalisé par la salle de nouvelles nationale du réseau, située à Montréal) et « hors Québec » (dans les cas de reportages réalisés par une station située à l'extérieur du Québec, comme Radio-Canada Ontario, par exemple). La section

[résultats](#) présentera cette classification déterminée par des scripts¹¹ produits par mon directeur de recherche et professeur à l'École des médias, Jean-Hugues Roy.

3.1 Structure des requêtes automatisées

La structure des requêtes envoyées au moteur de recherche de Google Actualités est la suivante : « <mot-clé> près de <ville> ». Chaque requête a trois caractéristiques : un mot-clé, une ville et une localisation.

Premièrement, la majorité des mots-clés utilisés est tirée de l'étude de Fischer *et al.*, mais certains ont été modifiés ou retirés étant donné qu'ils n'étaient pas pertinents dans le contexte québécois de 2021 comme « *corruption* » qui a été remplacé par « racisme », un terme beaucoup plus d'actualité. De même, « *president* » a été remplacé par « Trump OR Biden » dans les requêtes ainsi que par « premier ministre ». La recherche compte donc 26 mots-clés répartis dans trois catégories :

13 termes locaux,

7 termes nationaux et

6 termes « mixtes ».

Cette approche diffère quelque peu de la méthodologie utilisée par Fischer *et al.* (voir Annexe 3, qui compare nos termes avec ceux que Fischer *et al.* ont utilisés), au sens où ces derniers faisaient une distinction nette entre un mot-clé local et un mot-clé national. Toutefois, en dressant la liste des mots-clés utilisés pour ce projet, il est apparu évident que le caractère binaire de la classification des auteurs était limitatif. Des termes comme « construction », « COVID », « santé publique », « élections », « armes à feu » et « député » pouvaient avoir des portées à la fois locales et nationales. Nous avons donc créé la catégorie « mixtes ». En outre, il était inévitable de tenir compte du contexte pandémique d'où l'intégration du mot-clé « COVID », absent de l'étude de Fisher *et al.* Ces termes « mixtes » jouent en quelque sorte le rôle d'un groupe contrôle afin de voir si Google Actualités favorise des sources nationales ou des sources locales ou régionales quand il les rencontre.

11 Les programmes informatiques produits par Jean-Hugues Roy sont accessibles dans le dossier codeJHR: https://github.com/NicolasStG/audit_google_news

Deuxièmement, 104 villes différentes ont été utilisées pour la collecte. Il s'agit des chefs-lieux des 104 municipalités régionales de comté (MRC) et territoires équivalents (agglomérations et villes hors MRC) du Québec.

Troisièmement, la position géographique (latitude et longitude) des 104 villes a été utilisée pour simuler la localisation utilisée par le programme informatique. Chaque requête a été faite dans une nouvelle fenêtre de navigation sur Firefox avec la localisation propre à la ville incluse dans la requête. Par exemple, pour la recherche « accident près de Alma », le programme informatique a simulé une requête lancée exactement à partir d'Alma (latitude : 48,54972 ; longitude : -71,65333). Pour vérifier si la simulation de la localisation fonctionnait, le script ouvrait le site Internet *Find my location*¹² et enregistrait l'adresse ainsi que la latitude et la longitude affichée. Pour la ville de Montréal, la localisation a été utilisée à deux reprises : une recherche utilisait la forme standard avec la mention « près de Montréal » et une autre ne gardait que le mot-clé. (sans la mention « près de Montréal »). L'objectif, ici, était de mesurer le degré de « montréalisation » dans les résultats affichés par Google Actualités.

Pour chacune des requêtes, seuls les 25 premiers résultats ont été enregistrés. Il semble cependant qu'à certains moments, le programme informatique ne soit pas arrivé à récolter 25 résultats. De même, certains mots-clés n'ont généré aucun résultat. Les raisons expliquant cela demeurent nébuleuses, mais il est possible que la page de résultats n'ait pas eu le temps de se télécharger convenablement empêchant le programme informatique de saisir la totalité des résultats.

Les dates pour effectuer les saisies ont été les suivantes :

1. le mardi 3 août 2021,
2. le jeudi 19 août 2021,
3. le samedi 28 août 2021,
4. le samedi 18 septembre 2021,
5. le jeudi 30 septembre 2021,

¹² <https://findmylocation.org/>

6. le mardi 19 octobre 2021,
7. le jeudi 18 novembre 2021 et
8. le samedi 18 décembre 2021.

Pour les six premières, l'objectif était d'étudier, en plus de la question principale de recherche, l'effet des élections fédérales (le 20 septembre 2021) et municipales (le 7 novembre 2021) sur les résultats de recherche, notamment pour le mot-clé « élection OR campagne ». Considérant l'importance des médias locaux pour les démocraties locales, il est pertinent d'évaluer la façon dont Google Actualités les répertorie dans des moments aussi importants. Les deux dernières saisies avaient pour but de voir si certains éléments avaient changé sur le plan de l'affichage des médias locaux dans les résultats de recherche depuis le Community News Summit du Google News Initiative, en août 2021.

Pour résumer, les 26 mots-clés ont été associés à chacune des 104 villes (Montréal est utilisée deux fois, donc 105 requêtes par mot-clé) pour produire un total de 2 730 requêtes générant 68 250 résultats (25 résultats récoltés par requête). Avec huit saisies, nous avons donc un total théorique de 546 000 résultats à analyser. Dans la pratique, ce sont plutôt 488 009 résultats qui ont été récoltés en raison des erreurs possibles mentionnées ci-dessus. Le portrait des résultats est détaillé dans la section « [résultats](#) ».

La recherche de Fischer *et al.* a fait usage du coefficient de Gini pour déterminer les inégalités dans les résultats. Considérant que le mémoire vise à reproduire cette recherche, la même méthodologie a donc été utilisée. Ce faisant, le coefficient de Gini a été utilisé afin d'ajouter un élément de comparaison entre nos résultats et ceux de Fischer *et al.* Ensuite, ce processus a été répété afin de prendre en compte les critiques de Magnusson ([2021](#)) où le coefficient doit être calculé de la même façon que les données ont été récoltées, c'est-à-dire par villes. De plus, des tableaux croisés dynamiques ont été réalisés pour illustrer certains éléments précis et répondre à certaines questions secondaires, dont celle sur les élections.

CHAPITRE 4

RÉSULTATS

À la lumière des résultats obtenus, les médias locaux et régionaux¹³ ne sont pas désavantagés par rapport à leurs pairs nationaux, ce qui ne rejoint pas du tout les conclusions de Fischer *et al.* (2020). En effet, peu importe le type de mot-clé (local, mixte ou national) utilisé lors de la saisie ainsi que le rang du média dans les résultats de recherche, les médias locaux et régionaux apparaissent plus fréquemment (voir figures 1 et 2).

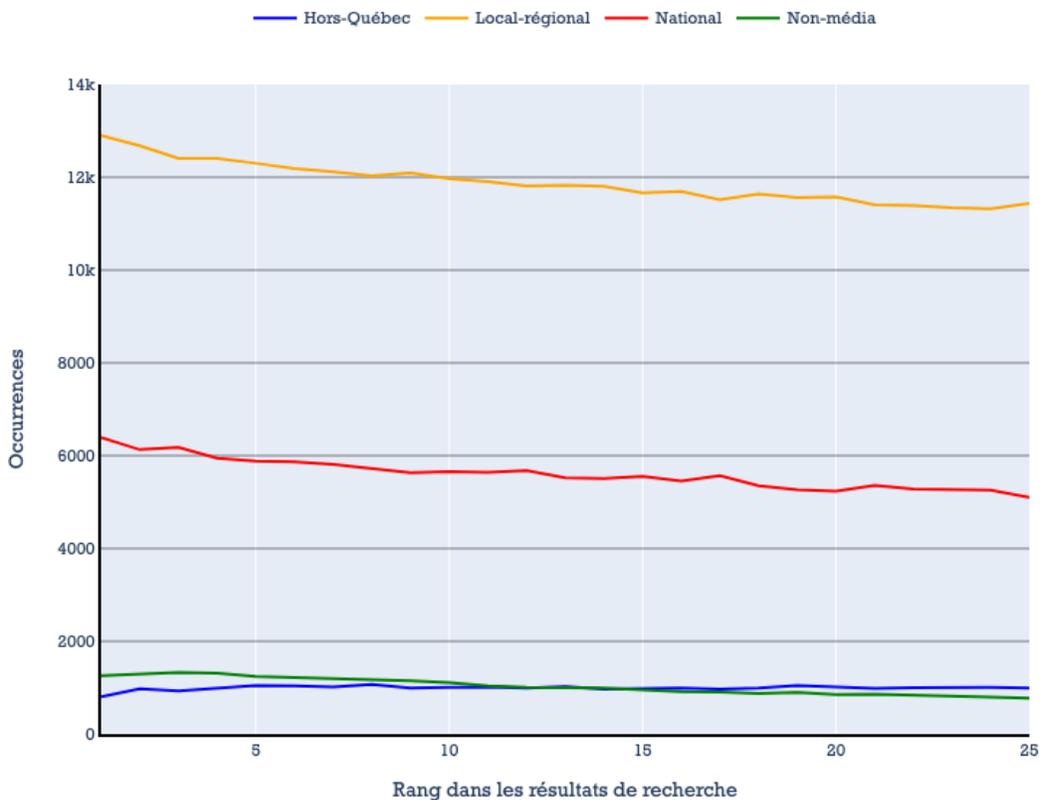


Figure 1. Répartition de l'ensemble des résultats de recherche sur Google Actualités selon le rang de chaque catégorie de média dans ces résultats

13. La liste classant les médias selon les catégories « local/régional », « national », « non-média » et « hors Québec » est disponible ici : https://github.com/NicolasStG/audit_google_news/blob/main/liste-media/liste-media.csv.

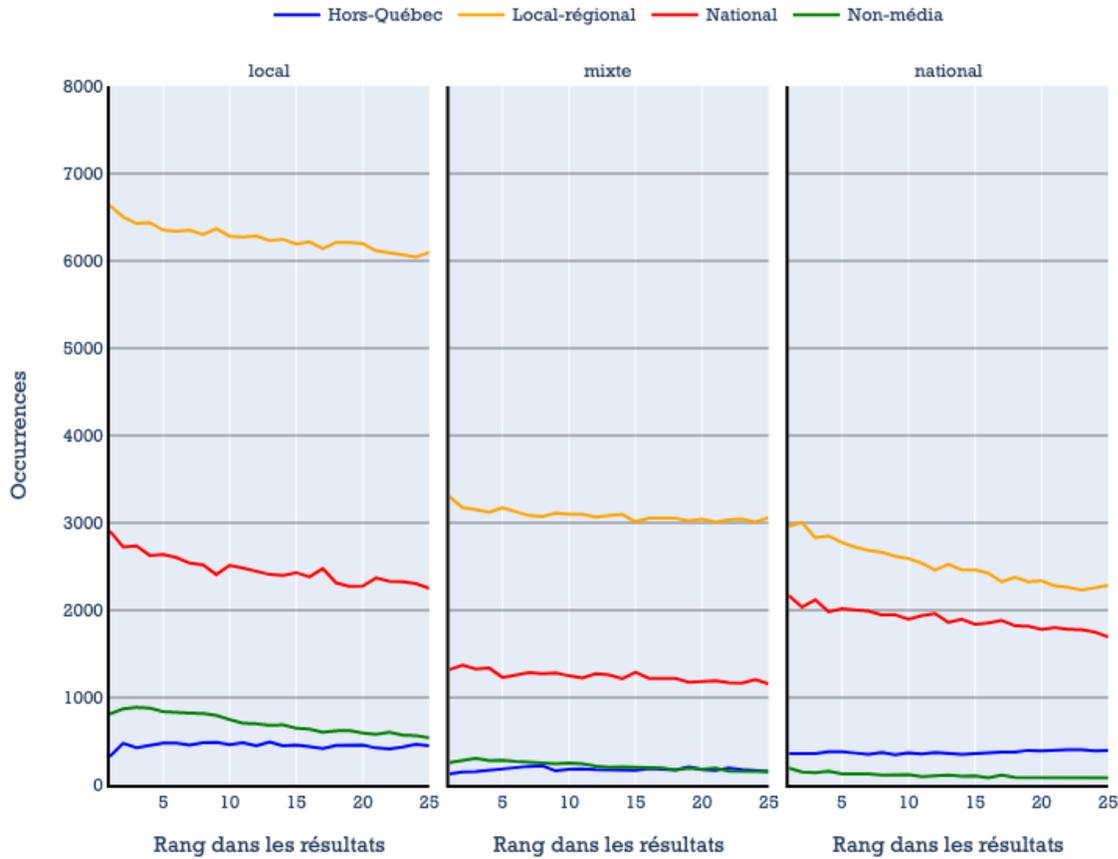


Figure 2. Répartition, par type de mot-clé, des résultats de recherche sur Google Actualités selon le rang de chaque catégorie de média dans les résultats

Les figures 1 et 2 illustrent à quel point les médias locaux et régionaux sont davantage mis de l'avant, et ce, même dans les premiers résultats affichés par Google Actualités. Ces résultats rejoignent les constats de Fischer *et al.*, mais seulement dans le cas des recherches comprenant des mots-clés « local », qui affichent davantage de médias locaux.

En nombres absolus, il y a tout simplement plus de médias locaux et régionaux que de médias nationaux dans les résultats retournés par Google Actualités. En effet, la figure 1 pour l'ensemble des 26 mots-clés, on compte 297 017 résultats provenant de médias locaux et régionaux (60,86 %) contre 140 304 résultats pour les médias nationaux (28,75 %).

En poursuivant l'analyse avec la figure 2, il est possible de remarquer que pour les recherches comportant des mots-clés mixtes, les médias locaux et régionaux sont toujours plus dominants que les médias

nationaux. Cela vient infirmer une des hypothèses de la recherche qui supposait que les résultats seraient équitablement répartis entre médias locaux/régionaux, d'une part, et médias nationaux, d'autre part, pour ce type de mots-clés. Ils sont plutôt de l'ordre de 65,36 % pour les médias locaux et de 26,32 % pour les nationaux. En comparaison, pour les recherches comportant des mots-clés « local », la répartition est plutôt de 63,37 % pour les médias locaux et de 24,96 % pour les médias nationaux. L'écart se rétrécit pour les mots-clés de type national où 51,49 % des résultats proviennent des médias locaux contre 38,72 % pour les médias nationaux.

Indice de Gini pour les types de mots-clés selon le rang dans les résultats de recherche

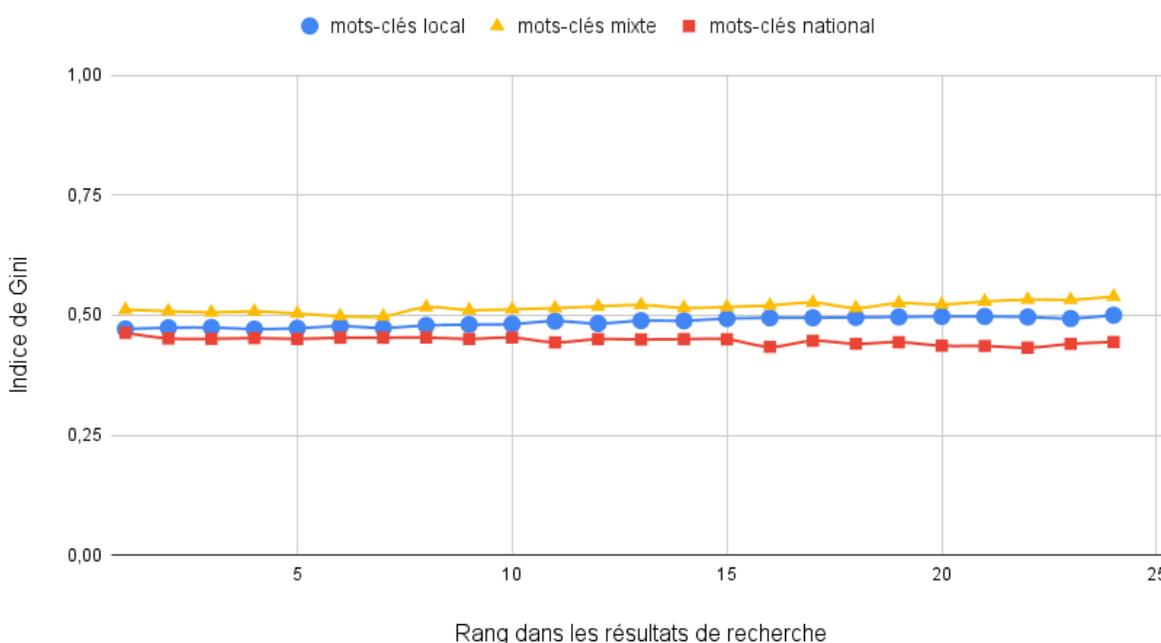


Figure 3. Indice de Gini pour les types de mots-clés local, national et mixte selon la position dans les résultats de recherche.

La figure 3 permet d'illustrer, par l'entremise du coefficient de Gini, le degré d'équité dans la répartition des résultats. Avec un coefficient de Gini à 0,44, le type de mots-clés « national » est celui dont les résultats sont répartis avec la plus grande équité, suivi des mots-clés « local » à 0,48 et des mots-clés « mixte » à 0,51. Il est important de noter que l'équité calculée par le coefficient de Gini prend aussi en compte toutes les catégories de médias, incluant les non-médias et les médias hors Québec qui pèsent moins lourd dans notre base de données.

Malgré tout, ces premiers constats sont loin de ceux obtenus par Fischer *et al.* (2020). Bien que les résultats soient mieux répartis pour les recherches contenant des mots-clés de type national, Google Actualités ne favorise pas, au Québec du moins, les médias nationaux au détriment des médias locaux dans ses résultats de recherche. Au contraire, il semble plutôt que peu importe le type de mots-clés, Google Actualités favorise les médias locaux.

Cela dit, comme il sera montré plus bas, Google Actualités semble davantage favoriser les médias nationaux dans les recherches contenant les mots-clés nationaux sans pour autant que ces derniers soient plus nombreux dans les résultats que les médias locaux et régionaux.

La présente section est séparée en différentes sous-sections. Nous décrivons d'abord notre base de données. Nous répondrons ensuite à nos différentes sous-questions de recherche. Puis, nous reproduirons les analyses de Fischer *et al.* (2020), avec et sans les critiques de Magnusson. Nous approfondirons enfin différents constats observés lors de l'analyse de la base de données, incluant la présence du site Internet necrocanada.ca, un non-média qui figure au dixième rang des domaines les plus souvent retournés par Google Actualités. Évidemment, tout ne peut pas être analysé, c'est pour cela que la base de données demeure ouverte à quiconque qui voudrait y accéder et faire des observations. Une version simplifiée des programmes informatiques est aussi disponible ce qui permettra à quiconque désirant reproduire notre mémoire.

4.1 Explications et détails de la base de données

Avant d'entrer dans les détails de la base de données, il convient de mentionner que l'étude s'attarde à la quantité des résultats obtenus et non pas à leur qualité. Autrement dit, il n'est pas question du contenu des articles que retourne l'algorithme de Google Actualités, ni de se demander s'ils sont pertinents, mais bien simplement du nom des médias qui sont présentés à l'utilisateur ainsi que la date où les contenus ont été publiés. La précédente section offre un premier panorama des données récoltées. Elle vise à aller un peu plus loin et à décortiquer les éléments essentiels.

Le fait que les médias locaux et régionaux ne soient pas sous représentés dans les résultats de recherche sur Google Actualités peut s'expliquer par le contexte médiatique particulier du Québec. Dans un territoire plus petit que les États-Unis et avec une langue différente, la portée des médias québécois est davantage provinciale. Qui plus est, le média qui se retrouve le plus fréquemment dans les résultats, Radio-Canada

— un média financé par l’argent public —, produit beaucoup de contenu local/régional du fait de sa mission et grâce à ses nombreuses antennes aux quatre coins de la province et du pays. Celles-ci ont un poids dans leur région respective et semblent également profiter de la notoriété du média sur Internet, alors que tous les articles redirigent vers le même nom de domaine « <https://ici.radio-canada.ca/> » sans faire de distinction régionale. Cet aspect, où le nom de domaine est le même, peu importe la provenance du reportage, semble affecter positivement le référencement de l’entreprise puisqu’elle n’a qu’un seul domaine à référencer, et non plusieurs, et ce, avec beaucoup de contenus à promouvoir. De ce fait, le classement méritocratique que suggère Google avantage le média où tout partage de contenu donne l’impression à l’agrégateur de provenir du même site, augmentant la notoriété de ce dernier.

Les prochaines sous-sections vont d’abord détailler l’entièreté de la base de données, puis une version réduite ne comptant que les cinq premiers résultats de chaque recherche, pour enfin comparer la base de données complète avec celle de Fischer *et al.* (2020).

4.1.1 L’entièreté de la base de données

<i>Catégorie de média</i>	<i>Nombre de médias affichés dans Google Actualités</i>	<i>Poids dans l’ensemble des médias</i>	<i>Nombre de résultats affichés dans Google Actualités</i>	<i>Poids dans les résultats</i>	<i>Fréquence moyenne par média</i>
<i>hors Québec</i>	938	64,73 %	24 860	5,09 %	26,50
<i>local/régional</i>	105	7,25 %	297 017	60,86 %	2 828,73
<i>national</i>	74	5,11 %	140 304	28,75 %	1 896,00
<i>non-média</i>	332	22,91 %	25 828	5,29 %	77,79
<i>Total</i>	1 449	100 %	488 009	100 %	-

Tableau 2. Répartition de la base de données par catégorie de médias

Le tableau 2 témoigne d’un favoritisme de la part de Google Actualités dans les catégories de médias locaux/régionaux et nationaux, eux qui comptent pour 89,61% des résultats affichés. Néanmoins, ce même tableau montre que la base de données comporte énormément de bruit. Effectivement, la quantité de domaines provenant de médias hors Québec ou de sites non médiatiques est plutôt impressionnante. Néanmoins, le poids de chacun de ceux-ci dans la base de données demeure marginal. Le tableau 2 montre que les médias hors Québec comptent pour 64,73 % des médias répertoriés, mais pour seulement 5,09 % des résultats. Les non-médias, eux, comptent pour 22,91 % des domaines affichés, mais pour 5,29 % des résultats, dont une bonne partie (2,46 %) provient du site Nécrologie Canada. Une section de ce mémoire

est d'ailleurs consacrée à ce site et à l'importance qu'il occupe dans les résultats (il est au dixième rang des domaines les plus fréquemment retournés).

Cela dit, le tableau 2 montre qu'on trouve plusieurs non-médias dans les résultats, mais Google Actualités ne semble pas leur accorder une grande importance. Pour l'ensemble de la base de données, l'on remarque qu'en moyenne chaque non-média n'a été affiché que 77,79 fois par Google Actualités au cours de notre récolte de données. Le chiffre chute à 26,50 pour les articles d'un média hors du Québec. À titre de comparaison, chaque média local/régional a été retourné 2 828,73 fois et chaque média national, 1 896 fois.

Cette « moyenne par média » donne une idée de la fréquence relative de chaque catégorie de média dans les résultats. Elle varie énormément selon le type de mot-clé qui a été utilisé dans les requêtes (voir tableau 3). Effectivement, pour les recherches contenant des mots-clés locaux, chaque média local/régional est retourné en moyenne 1 506 fois contre 963 pour les médias nationaux.

Pour les mots-clés mixtes, la fréquence passe à 756 pour les médias locaux/régionaux et à 501 pour les médias nationaux. Finalement pour les mots-clés nationaux, la fréquence moyenne est de 614 pour chaque média local/régional contre 669 pour les médias nationaux. On remarque aussi que Google Actualités a affiché plus de médias nationaux dans les recherches avec des mots-clés nationaux que dans les autres.

Il semble donc que Google Actualités agisse différemment selon le type de mots-clés qu'on utilise dans les recherches. L'algorithme retourne donc plus souvent chaque média national individuellement qu'un média local/régional, mais comme il y a un moins grand nombre de médias nationaux dans l'ensemble des résultats, leur présence, au final, demeure plus faible que celles des médias locaux et régionaux.

Type de mots-clés	Catégorie de média	Nombre de médias	Nombre de résultats	Poids dans les résultats	Fréquence moyenne par média
local	hors Québec	616	11 215	4,53 %	18,20
	local/régional	104	156 623	63,37 %	1505,99
	national	64	61 680	24,96 %	963,75
	non-média	172	17 629	7,13 %	102,49
	TOTAL	956	247 147	100 %	258,52
mixte	hors Québec	452	4367	3,69 %	9,66
	local/régional	102	77 155	65,36 %	756,42
	national	62	31 077	26,32 %	501,24
	non-média	127	5 454	4,62 %	42,94
	TOTAL	743	118 053	100 %	158,89
national	hors Québec	653	9 278	7,55 %	14,20
	local/régional	103	63 239	51,49 %	613,97
	national	71	47 547	38,72 %	669,68
	non-média	196	2745	2,24 %	14,01
	TOTAL	1 023	122 809	100 %	120,05

Tableau 3. Répartition de la base de données par catégorie de média selon le type de mots-clés

En concentrant notre analyse sur les 20 médias les plus souvent affichés par Google Actualités, on constate que la majorité sont des médias locaux ou régionaux. Le tableau 4 détaille cette distribution selon s'il s'agit d'un média « local/régional » et « national ». On compte également deux domaines « non-média » (necrocanada.com et iheartradio.ca).

Catégorie de média	Nombre de médias affichés dans Google Actualités	Poids dans l'ensemble des médias	Nombre de résultats affichés dans Google Actualités	Poids dans les résultats	Fréquence moyenne par média
<i>local/régional</i>	11	55,00 %	177 371	36,34 %	16 124,64
<i>national</i>	7	35,00 %	126 836	25,99 %	18 119,42
<i>non-média</i>	2	10,00 %	18 879	3,87 %	9 439,5
Total	20	100,00 %	323 086	66,20 %	-

Tableau 4. Répartition des résultats de recherche pour les 20 médias les plus souvent affichés par Google Actualités, selon la catégorie de média. Le poids dans les résultats reflète le poids de chaque catégorie de média dans l'ensemble de la base de données (488 009 résultats).

Ce top 20 représente 66,20 % de la base de données ce qui tend à montrer qu'un petit nombre de médias sont davantage mis de l'avant par Google Actualités. Qui plus est, en moyenne, un média « local » figurant dans cette catégorie est retourné à plus de 16 124,64 reprises. C'est une fréquence 5,7 fois supérieure à la moyenne du tableau 2 qui comprend l'ensemble de la base de données. Pour un média « national », la

moyenne grimpe à 18 119,42 occurrences, plus de 9,55 fois la moyenne du tableau 1. Par ailleurs, les 7 médias nationaux du top 20 comptent pour 90 % de tous les 140 304 résultats affichés pour cette catégorie de média par Google Actualités. De leur côté, les 11 médias locaux et régionaux du top 20 comptent pour 59,72 % des 297 017 résultats pour cette catégorie de média. Ces chiffres semblent indiquer que lorsqu'il est question d'un média national, Google Actualités retourne toujours les mêmes alors que pour les médias locaux et régionaux, il fait davantage preuve de diversité. Plusieurs raisons peuvent expliquer cela. Tout d'abord, la quantité de médias nationaux dans la province réduit le nombre qui peut être retourné par la plateforme alors que pour les médias locaux et régionaux, le fait de cibler un endroit précis dans la requête permet de favoriser un média qui couvre ce territoire au détriment des autres. Il est aussi possible que les médias nationaux retournés dans le top 20 aient un bon référencement.

Type de mots-clés	Catégorie de média	Nombre de médias	Poids dans les médias	Nombre de résultats	Poids dans les résultats	Fréquence moyenne par média
local	local/régional	10	50,00 %	85 506	54,04 %	8 550,6
	national	8	40,00 %	58 065	36,69 %	7 258,13
	non-média	2	10,00 %	14 667	9,27 %	7 333,5
	TOTAL	20	100 %	158 238	100 %	-
mixte	local/régional	12	60,00 %	44 746	58,40 %	3 728,83
	national	6	30,00 %	28 357	37,01 %	4 726,17
	non-média	2	10,00 %	3 513	4,59 %	1 756,50
	TOTAL	20	100 %	76 616	100 %	-
national	hors Québec	1	5,00 %	1 136	1,24 %	1 136,00
	local/régional	12	60,00 %	46 759	50,87 %	3 896,58
	national	7	35,00 %	44 027	47,90 %	6 289,57
	TOTAL	20	100 %	91 922	100 %	-

Tableau 5. Répartition des résultats de recherche pour les 20 médias les plus affichés par Google Actualités selon la catégorie de média et le type de mots-clés

Lorsqu'on répartit les résultats pour le top 20 par type de mots-clés (local/régional, mixte ou national), le portrait se raffine. Pour les recherches comportant des mots-clés locaux, la fréquence moyenne pour les médias locaux et régionaux se chiffre à 8 550,6 contre 7 258,13 pour les médias nationaux. La préférence accordée par Google Actualités aux médias locaux/régionaux est donc moins grande.

Pour les recherches comportant des mots-clés « mixte », les médias nationaux affichent une fréquence moyenne de 4 726,17, une fréquence plus élevée que les médias locaux et régionaux qui s'établit à 3 896,58 par média.

Finalement pour les recherches comportant des mots-clés nationaux, les médias nationaux comptent une moyenne de 6 289,57 occurrences chacun. L'écart avec les médias locaux et régionaux se creuse, ici, ces derniers affichant une fréquence de 3 898,08. Les conclusions que l'on peut tirer de ces chiffres seront abordées plus bas.

L'annexe 5 détaille les médias les plus souvent retournés par Google Actualités ainsi que leur poids en pourcentage dans la base de données complète. Le constat quant au poids de Radio-Canada et de ses antennes est sans équivoque. La portion des contenus « local/régional » de Radio-Canada compte pour 13,36 % de la base de données (65 191 résultats), un chiffre 2,17 fois plus élevé que le deuxième média, *La Presse*, qui pèse pour 6,15 % de la base de données avec 30 034 occurrences. Comme dans l'étude de Nechushtai et Lewis (2019) où une certaine concentration des médias est notée (*The New York Times*, CNN, Politico, *The Washington Post* et HuffPost représentaient 49 % des résultats), dans une moindre mesure au Québec, les cinq médias les plus retournés — Radio-Canada (local/régional), *La Presse*, *Le Journal de Montréal*, *Le Journal de Québec*, et *Le Soleil* — comptent pour 35,01 % des occurrences. Qui plus est, à eux seuls, les trois premiers résultats — Radio-Canada (local/régional), *La Presse*, *Le Journal de Montréal* — équivalent à plus de 25 % de la base de données. En comparaison, le top 3 de Fischer *et al.* (2020) représente 16,4 %. Il semble donc y avoir une concentration plus importante des médias affichés par Google Actualités au Québec qu'aux États-Unis. Les dix médias les plus fréquemment affichés comptent pour 50,01 % de la base de données alors que le top 20 compte pour 66,20 %. C'est donc dire que les 1 429 autres médias répertoriés lors des saisies se partagent environ 164 947 résultats.

Dans les 11 médias locaux et régionaux du tableau 4, on retrouve Radio-Canada, TVA Nouvelles, *Le Soleil*, *Le Nouvelliste*, *La Tribune*, *Le Droit*, *Le Quotidien*, *La Voix de l'Est*, La Nouvelle, CIMT/CHAU, *Info Dimanche*. Alors que dans les sept médias nationaux, on compte *La Presse*, *Le Journal de Montréal*, *Le Journal de Québec*, Radio-Canada, TVA Nouvelles, *Le Devoir* et Météomédia.

Lorsque les données de l'annexe 5 sont réparties par type de mots-clés (local/régional, mixte et national), comme au tableau 5, on constate que les médias qui constituent le top 20 par type de mot-clé varient quelque peu ainsi que le poids relatif de chacun (voir annexe 6). Par exemple, Radio-Canada (local/régional), qui demeure le média le plus fréquemment affiché, peu importe le type de mots-clés, a un poids moindre (11,17 %) lorsque les recherches contiennent des mots-clés nationaux par rapport aux recherches qui contiennent des mots-clés locaux (13,65 %) ou mixtes (15,03 %). Parallèlement, *La Presse*

(8,95 %) gagne en importance dans les recherches nationales par rapport aux recherches locales (4,82 %) et mixtes (6,03 %).

Qui plus est, le poids total des médias du top 20 est passablement plus important dans les recherches contenant des mots-clés nationaux avec 74,85 % de tous les résultats. Dans les recherches de type local/régional et mixte, les pourcentages sont de 64,03 % et 64,90 %, respectivement. C'est donc dire pour les recherches nationales, ce sont les médias nationaux qui sont plutôt favorisés par Google Actualités.

Fait intéressant, si on concentre encore davantage notre analyse en ne retenant que les cinq médias les plus fréquemment retournés on constate qu'à eux seuls, ils représentent 35,03% des occurrences pour l'ensemble des requêtes. Leur poids dans les recherches nationales grimpe à 41,46 %. Ce petit groupe de cinq médias représente 36,15% des résultats pour les recherches mixtes, il passe à 33,53 % dans le cas des recherches locales. Encore une fois, ces données montrent que pour les recherches nationales, Google Actualités se tourne vers certains médias plus que les autres.

Que retenir de tous ces chiffres? Ces données mettent en lumière le fait que Google Actualités favorise une poignée de médias au détriment des autres. Et ces médias sont autant nationaux que locaux/régionaux. La fréquence moyenne par média plus élevée à certains moments pour les médias nationaux ne signifie pas nécessairement que Google Actualités les favorise davantage par rapport aux médias locaux/régionaux, mais plutôt qu'il les favorise par rapport aux autres médias nationaux pour des raisons qui pourraient être liées à la notoriété et/ou au référencement. En quantité et en diversité, des médias locaux et régionaux sont favorisés par l'agrégateur de nouvelles, et ce, même dans le top 20. Ce faisant, les chiffres semblent brosser un portrait plus complet du paysage médiatique de la province. Les calculs selon le type de mots-clés sont pertinents, car ils permettent de montrer comment Google Actualités agit selon le type de recherche, ce qui donne une idée du fonctionnement de son algorithme. Il semble effectivement agir différemment lors des recherches contenant des mots-clés nationaux en favorisant davantage les médias nationaux.

4.1.2 Dans les cinq premiers résultats

Dans le but d'avoir une meilleure idée de l'importance des médias locaux et régionaux dans les résultats affichés par Google Actualités, une autre analyse a été faite, cette fois en ne conservant que les cinq premiers résultats de chaque recherche (au lieu de 25).

Cela donne 104 418 résultats à analyser. Au lieu de compter 1 449 médias uniques, on n'en retrouve plus que 816. En comparant le tableau 6 (ci-bas) avec le tableau 2, on remarque que les données sont plutôt similaires quant au poids de chacune des catégories de médias. C'est donc dire que Google Actualités met de l'avant les contenus provenant des médias locaux/régionaux, que l'on considère uniquement les cinq premiers résultats qu'il présente ou les 25 premiers résultats. Chaque média local ou régional apparaît en moyenne 626,94 fois dans les cinq premiers résultats des requêtes que nous avons faites, contre 469,75 fois pour chacun des médias nationaux.

<i>Catégorie de média</i>	<i>Nombre de médias affichés dans Google Actualités</i>	<i>Poids dans les médias</i>	<i>Nombre de résultats affichés dans Google Actualités</i>	<i>Poids dans les résultats</i>	<i>Fréquence moyenne par média</i>
<i>hors Québec</i>	496	60,78 %	4 751	4,55 %	9,57
<i>local/régional</i>	100	12,25 %	62 694	60,04 %	626,94
<i>national</i>	65	7,97 %	30 534	29,24 %	469,75
<i>non-média</i>	155	19,00 %	6 439	6,17 %	41,54
<i>Total</i>	816	100,00 %	104 418	100,00 %	-

Tableau 6. Répartition des cinq premiers résultats sur Google Actualités par catégorie de média.

Là où l'on remarque une différence entre la base de données complète (qui comprend les 25 premiers résultats) et le sous-ensemble ne comprenant que les cinq premiers résultats, c'est dans le poids du groupe des 20 médias les plus fréquemment retournés. Ils ne représentent que 64,35 % des résultats de ce sous-ensemble, alors qu'ils représentent 66,68 % de l'ensemble des résultats. La diversité des médias présentés par Google Actualités est donc légèrement supérieure si on considère uniquement les cinq premiers résultats de chaque requête que si on considère les 25 premiers.

La même chose se produit si on examine d'autres groupes (top 10, top 5, top 3 ou top 1). Le poids de chacun de ces groupes est moindre dans le sous-ensemble des cinq premiers résultats de recherche que dans la base de données complète. Cela laisse croire que dans les cinq premiers résultats, Google varie davantage les contenus affichés et favorise moins certains médias par rapport à d'autres (tableau 7). Les différences ne sont néanmoins pas si importantes.

	<i>Pourcentage pour les cinq premiers résultats (%)</i>	<i>Pourcentage dans la base de données complète (%)</i>
<i>Top 20</i>	64,35	66,68
<i>Top 10</i>	47,88	50,03
<i>Top 5</i>	33,71	35,05

Top 3	24,60	25,49
Radio-Canada « local/régional »	12,90	13,37

Tableau 7. Comparaison du poids des médias les plus retournés entre la base de données complète et celle comprenant les cinq premiers résultats de recherche sur Google Actualités

4.1.3 Comparaison avec la base de données de Fischer *et al*

Le tableau 8 compare la recherche de Fischer *et al.* (2020) avec ce mémoire. La présente recherche compte moins de résultats, car elle utilise moins de mots-clés et a enregistré moins de résultats par requête. Néanmoins, notre base de données est suffisamment volumineuse pour dégager certains constats permettant de répondre aux différentes hypothèses de recherche.

	Fischer et al. (2020)	La présente recherche
Résultats de recherche (base de données) à analyser	12 290 438	488 009
États/Municipalités	Plus de 3 000 comtés	104 chefs-lieux des MRC
Nombre de médias répertoriés	8740 (local, régional, national, international)	1440 (local/régional, national, hors Québec, non-média)
Mots-clés	32 (Locally-oriented / Nationally-oriented)	26 (Local, mixte, national)
Résultats récoltés par recherche de mots-clés dans un.e État/municipalité	100 premiers résultats	25 premiers résultats

Tableau 8. Comparaison entre la présente recherche et celle de Fischer *et al.*

4.2 Réponses aux hypothèses de recherche

Jusqu'à maintenant, ce chapitre a bien montré, avec quelques nuances, que la question principale de recherche ainsi que l'hypothèse #1 étaient infirmées et que Google, au contraire de ce qui était attendu, ne défavorise pas les médias locaux ou régionaux au détriment des médias nationaux dans ses résultats de recherche et encore moins dans les cinq premiers résultats. Néanmoins, cette recherche comportait une série d'hypothèses primaires et secondaires sur lesquelles nous allons maintenant nous pencher dans cette sous-section.

4.2.1 Hypothèses de recherche primaire

Hypothèse #2 : « Les médias locaux et régionaux des grands centres comme Montréal et Québec sont davantage mis de l'avant que les médias locaux en régions éloignées. »

En regardant simplement les chiffres absolus, disponibles à l'annexe 7, il est possible de remarquer que cette hypothèse est infirmée. À part Québec, qui trône au sommet de la liste avec 30 469 occurrences (*Le Soleil* compte à lui seul 20 709 résultats), Montréal arrive beaucoup plus loin, soit au 16^e rang, avec 5 587 résultats dont 2 376 viennent de *ici.radio-canada.ca*. Trois-Rivières (26 295 résultats, dont 17 967 proviennent du *Nouvelliste*) et Sherbrooke (21 085 résultats, dont 15 809 pour *La Tribune*) se retrouvent aux 2^e et 3^e rang des sièges sociaux des médias locaux et régionaux les plus fréquemment retournés par Google Actualités.

Quand on analyse simplement les médias locaux et régionaux les plus retournés par Google Actualités dans les recherches, on constate que Radio-Canada trône au sommet avec 65 191 résultats. En filtrant par les villes où le contenu radio-canadien a été produit, on remarque que Rouyn-Noranda (10 987 résultats), Matane (10 677 résultats) et Rimouski (9 105 résultats) sont les trois villes qui apparaissent le plus souvent dans les résultats. Montréal figure au dernier rang. C'est donc dire que Radio-Canada met efficacement en valeur ses contenus produits localement, comme sa mission le lui demande. Et Google Actualités les reprend.

En calculant le coefficient de Gini de tous les médias locaux et régionaux répertoriés, on arrive à 0,6868. Cela laisse croire à un manque de diversité dans les médias retournés par Google Actualités. Autrement dit, la plateforme met de l'avant un petit nombre de médias au détriment de plusieurs autres. Ce constat se voit à l'œil alors que les trois médias les plus fréquemment retournés comptent pour 35% de tous les médias locaux et régionaux. En incluant le top 5, le pourcentage grimpe à 44,28%.

En ne considérant que les sièges sociaux, le coefficient de Gini diminue à 0,6228. C'est une baisse significative qui découle du fait que Radio-Canada produit son contenu dans 10 villes différentes. Parallèlement, le poids des médias formant le top 3 et le top 5 est beaucoup plus faible à 18,38% et 26% respectivement.

Malgré tout, les chiffres concernant les médias locaux et régionaux répertoriés par Google Actualités montrent tout de même une certaine concentration. Mais celle-ci n'est pas aussi élevée que ce qui était anticipé. Qui plus est, cette concentration n'est pas du tout l'affaire de Montréal ou de Québec, à l'exception du journal *Le Soleil*. L'annexe 8 permet de consulter la répartition des contenus retournés selon le siège social des médias affichés.

Hypothèse #3 : « La sous-représentation des médias locaux et régionaux est le résultat de l'algorithme de Google Actualités qui a tendance à promouvoir le contenu jugé "populaire" et produit par les médias nationaux en ne portant pas attention aux besoins informationnels des utilisateurs en région. »

Considérant que l'hypothèse générale de recherche ait été infirmée, cette sous-hypothèse ne fait plus de sens. En revanche, comme les médias locaux et régionaux sont finalement plus représentés que ce qui était anticipé, il est intéressant d'analyser le poids des différentes catégories de média et d'observer laquelle est la plus retournée par Google Actualités pour chacune des villes étudiées.

Catégorie de média	Nombre de villes où la catégorie de média est le plus fréquent
local/régional	86
National	16
local/régional et national	1
hors Québec I	1
national et hors Québec	1

Tableau 9. Nombre de villes où chaque catégorie de média est la plus fréquemment retournée.

Dans le tableau 9, il est possible de remarquer que dans 86 des 105 villes analysées, les médias les plus fréquemment retournés par Google Actualités sont des médias locaux ou régionaux. Dans de nombreux cas, c'était avec une majorité qui ne laissait pas de place au doute (voir l'annexe 9 pour les détails). En effet, pour 57 des 86 villes où les articles des médias locaux ou régionaux sont les plus souvent retournés, leur poids dépassait les 60 %. À Coaticook, 92,42 % des résultats provenaient de médias catégorisés comme locaux/régionaux, le pourcentage le plus élevé dans notre base de données.

Catégories de média	Villes
hors Québec	Lachute
local/régional et national	La Malbaie
national	Boischatel
	Chisasibi

	<i>Kuujuuaq</i>
	<i>Laval</i>
	<i>Longueuil</i>
	<i>Mont-Laurier</i>
	<i>Montréal</i>
	<i>Montréal_V2</i>
	<i>Québec</i>
	<i>Repentigny</i>
	<i>Saint-Hyacinthe</i>
	<i>Saint-Pierre-de-l'Île-d'Orléans</i>
	<i>Sainte-Agathe-des-Monts</i>
	<i>Terrebonne</i>
	<i>Vaudreuil-Dorion</i>
	<i>Ville-Marie</i>
<i>national et hors Québec</i>	<i>Mansfield</i>

Tableau 10. Les villes où la catégorie de média la plus fréquente dans les résultats de recherche sur Google Actualités n'est pas « local/régional ».

Toutefois, dans 16 des 105 villes étudiées, c'est la catégorie « national » qui se retrouve le plus souvent dans les résultats de recherche sur Google Actualités. En analysant davantage les données, on remarque que les grandes villes de la province apparaissent dans la liste. Le tableau 10 détaille les villes où la catégorie de média la plus souvent affichée par Google Actualités n'est pas « local/régional ». Québec, Montréal (deux fois), Laval et Longueuil en font toutes partie. Dans 4 des 16 cas où « national » est la catégorie la plus fréquente, le pourcentage dépasse 60 %. La recherche à Montréal contenant le mot-clé « près de » est celle qui a retourné la plus grande proportion de médias nationaux avec 84,61 %, suivi de Kuujuuaq (77,25 %), de Québec (66,80 %) et de Saint-Pierre-de-l'Île-d'Orléans (62,43 %), une localité de la Capitale-Nationale.

Bien que la recherche pour Montréal ne contenant pas « près de » ait retourné plus de médias nationaux que locaux/régionaux, son pourcentage se chiffre à moins de 50 % des résultats (47,71 %). Parallèlement, le poids de médias « local/régional » y est plus grand (29,73 %) que pour la recherche contenant le « près de » (12,49 %). Cette différence s'explique notamment par le fait que les médias nationaux répertoriés dans la recherche sont pratiquement tous situés à Montréal et couvrent donc davantage les informations de Montréal. Par ailleurs, le « près de » semble accentuer la localisation des contenus affichés à proximité de l'endroit nommé, comme nous l'avons expliqué au moment d'aborder l'hypothèse 1. Retirer l'expression « près de », bien que l'utilisateur soit toujours situé à Montréal géographiquement, ouvre la porte à du contenu ayant été produit en dehors de la ville demandée initialement. Cela crée davantage de diversité dans les résultats. On remarque cela en constatant que le poids de médias « hors Québec »

(17,27 % contre 0,50 %) et « non média » (5,29 % contre 2,40 %) est plus élevé pour la recherche ne contenant pas l'expression « près de ». Cela laisse sous-entendre que sans préciser la localisation dans la recherche, Google Actualités tente davantage de donner un caractère national ou international aux résultats qu'il présente. Il serait pertinent d'analyser en détail les résultats et le contenu retourné par les deux recherches (localisées à Montréal, avec et sans l'expression « près de ») et de les comparer, mais cela sort un peu du cadre de cette recherche.

En conclusion, il serait donc faux de croire, à la vue des résultats présentés, que Google Actualités ne se soucie pas des besoins régionaux. En revanche, il se pourrait, bien que Google retourne suffisamment de contenus locaux et régionaux, que leur fraîcheur laisse à désirer (hypothèse #4). Nous y reviendrons.

4.2.2 Répartition des résultats pour le mot-clé « élection OR campagne »

Ce mémoire se proposait d'observer le comportement du mot-clé « élection OR campagne » et d'analyser l'effet des élections fédérale et municipale sur l'affichage des contenus proposés par Google Actualités. Sur une possibilité théorique de 21 000 résultats (25 résultats par recherche, multiplié par 105 recherches et ensuite par 8 saisies), le programme informatique a récolté 20 266 résultats pour ce mot-clé. Cependant, comme les deux dernières saisies ont eu lieu après les élections, elles ont été retirées. Il y a donc 15 643 résultats à analyser sur une possibilité théorique de 15 750.

La composition de la base de données témoigne d'une plus grande équité dans la distribution en dehors du top 1 et du top 3 des médias les plus souvent retournés. En effet, le tableau 11 montre que le poids des médias les plus retournés s'érode par rapport à la base de données complète plus le nombre de médias dans le palmarès augmente. Cela laisse croire à une plus grande diversité que dans la base de données complète. Néanmoins, les antennes locales/régionales de Radio-Canada demeurent dominantes.

	Poids en pourcentage pour le mot-clé « élection OR campagne » (%)	Poids en pourcentage dans la base de données complète (%)	Écart dans le poids des deux bases de données
Top 20	60,13	66,20	-6,07
Top 10	44,10	50,01	-5,91
Top 5	30,29	35,01	-4,72
Top 3	22,68	25,47	-2,79
Top 1	13,74	13,36	+0,38

Tableau 11. Poids de différents groupes des médias les plus fréquemment retournés par Google Actualités (20 premiers, 10 premiers, cinq premiers, trois premiers et le premier) dans les recherches ne comprenant que le mot-clé « *élection OR campagne* » par rapport à l'ensemble des recherches (base de données complète).

Fait fort intéressant pour ce mot-clé, le poids et le nombre de médias « local/régional » et « national » sont beaucoup plus importants que dans la base de données complète (tableau 1).

Catégorie de média	Nombre de médias	Poids des médias	Nombre de résultats affichés	Poids des résultats affichés
hors Québec	156	48,15 %	612	3,91 %
local/régional	93	28,70 %	11 134	71,18 %
national	37	11,42 %	3428	21,91 %
non-média	38	11,73 %	469	3,00 %
Total	324	100,00 %	15 643	100,00 %

Tableau 12. Répartition des résultats pour le mot-clé « *élection OR campagne* » selon la catégorie de média

Le tableau 12 montre que les médias « local/régional » et « national » comptent pour 40,12 % des différents médias proposés par Google Actualités, contre un maigre 12,36 % dans la base de données complètes. Qui plus est, ils représentent 93,09 % des résultats affichés pour le mot-clé qui nous intéresse, contre 89,61 % dans la base de données complète. Il y a donc moins de médias hors du Québec et de non-médias qui ont été affichés lorsque nous avons fait des requêtes relatives aux élections ou à une campagne, ce qui accentue le poids des médias locaux/régionaux et nationaux dans les résultats. Cela laisse donc croire que Google Actualités fait un bon travail pour mettre de l'avant des contenus d'ici et ainsi réduire le poids des contenus non pertinents en période électorale.

La fréquence relative moyenne des médias locaux et régionaux est d'environ 119 fois contre 92 fois pour les médias nationaux. Cela peut s'expliquer par le fait que le Québec compte un nombre très faible de médias nationaux par rapport au nombre de médias locaux qui ont une voix dans pratiquement toutes les régions et le fait que Radio-Canada dispose d'antennes un peu partout dans la province.

Lorsque les données par mot-clé sont analysées selon la ville, on remarque que dans 92 des 105 villes étudiées, les médias locaux et régionaux ont un poids plus élevé que les médias nationaux et parfois par des marges assez impressionnantes (comme on l'a vu dans la section précédente). Dans 77 des 92 villes, le nombre de résultats provenant de médias locaux et régionaux est au-dessus des 60 %, ce qui inclut 20 villes où le pourcentage dépasse les 90 %.

Il y a seulement huit villes où les médias nationaux sont plus nombreux et une seule où les médias nationaux et hors Québec ont des pourcentages semblables. Les villes sont respectivement : Chisasibi, Kuujuaq, Longueuil, Mansfield, Mont-Laurier, les deux recherches à Montréal, Québec et Lachute. À cinq reprises, le pourcentage de médias nationaux dépasse les 60 %.

Si la recherche à Montréal avec le groupe de mots « près de » est celle qui a retourné le plus de médias nationaux pour nos requêtes électorales (87,33 %), le pourcentage fond à 53,33 % dans la recherche ne comprenant pas le groupe de mots « près de ». Les autres catégories, à l'exception des non-médias, sont plus souvent retournées dans les résultats de recherche : les médias hors Québec passent de 0 à 7,33 %, alors les médias locaux et régionaux grimpent de 12 % à 38,67 %. Un constat qui fait écho à ce qui a été expliqué dans la section précédente. Finalement, dans trois villes, le pourcentage de médias local/régional et national est similaire : Laval, La Malbaie et Saint-Hyacinthe.

On remarque, à la lumière de ces données, qu'à l'exception de Montréal et les villes avoisinantes, ainsi que Québec et quelques territoires plus nordiques comme Chisasibi et Kuujuaq, les médias locaux et régionaux sont davantage mis de l'avant par Google Actualités. Dans un contexte électoral, ces données sont rassurantes.

4.3 Comparaison avec la recherche de Fischer *et al.* ainsi que les critiques de Magnusson

Pour répondre à leur question de recherche Fischer *et al.* (2020) se sont tournés vers le coefficient de Gini dans le but de calculer la disparité dans les résultats affichés par Google Actualités. Bien que leur recherche comptait trois hypothèses de recherche, seules les deux premières seront revisitées dans ce mémoire. Et la première le sera à deux reprises, alors qu'une première version reprendra directement les manœuvres de Fischer *et al.* et la seconde prendra en compte les critiques d'Evan Magnusson, employé de Google, relativement au calcul du coefficient de Gini par ville — de la même manière que les données ont été récoltées.

Voici les deux hypothèses qui seront revisitées à l'aide du coefficient de Gini :

1. « *To systematically audit whether (the decline in) local news' online traffic can be attributed to Google News, we must show that Google News is more likely to feature national news outlets than local news outlets (Hypothesis 1).* »

2. « *Google may stratify results by search term, privileging national outlets for general interest topics, and local outlets for city-specific topics [2, p. 79-82], such as accidents and local governance. Hence, we examine whether Google News prioritizes national news outlets for topics of general interest, and local news outlets for locally-oriented issues (Hypothesis 2).* »

4.3.1 Reproduction de la recherche de Fischer *et al.*

4.3.1.1 Hypothèse 1

Pour l'ensemble de leurs bases de données, Fischer *et al.* calculent un coefficient de Gini à 0,82, une valeur « *[that] is in line with the highest levels of concentration observed in other search audits* » (2020, p. 5). Pour les chercheurs, cela soutient l'hypothèse 1 selon laquelle Google Actualités favorise les médias nationaux aux dépens des médias locaux. Fischer *et al.* disent :

« To test Hypothesis 1, we calculated descriptive statistics about the outlet frequencies across all the news search terms. We also measured the Gini Index for the distributions of local and national news outlets. The Gini Index tells us whether a few outlets dominate search results or whether outlets appear at similar levels. » (2020, p.3)

Malgré cette explication, la façon dont le coefficient a été calculé demeure nébuleuse. Le coefficient de Gini pour la présente recherche a donc été mesuré de quatre façons.

Tout d'abord, notre calcul 1 regroupe tous les 488 009 résultats selon les 1 449 noms de domaine présents dans ces résultats. Nous obtenons un coefficient de Gini de 0,9591 qui témoigne d'un manque flagrant d'équité dans la répartition des résultats. Pour avoir une égalité parfaite, il aurait fallu que chacun des 1 449 noms de domaine aient été affichés environ 325,55 fois par Google Actualités, ce qui n'est pas le cas, parce que la base de données contient plusieurs médias hors Québec dont le poids dans le nombre de résultats affichés est très faible.

Ce haut coefficient de Gini ne signifie pas pour autant que les médias nationaux soient plus fréquemment retournés, mais plutôt que certains médias (autant nationaux que locaux ou régionaux) sont plus représentés que d'autres. À ce sujet, la tableau 2 démontre bien que les médias locaux et régionaux, malgré qu'ils ne représentent que 7,25 % de l'ensemble des médias, sont favorisés par Google dans 6 résultats de recherche sur 10.

Ensuite, notre calcul 2 a été fait en regroupant toutes les données par catégorie (local/régional, national, hors Québec et non média), ce qui est plus représentatif de la constitution de la base de données et de la façon dont la question de recherche est posée. Le coefficient retourné est de 0,4769.

L'écart des coefficients entre les deux méthodes de calcul s'explique par le fait que la quantité et la répartition des données sont différentes. Le calcul 1 donne un coefficient beaucoup plus élevé parce que les données sont réparties dans 1 449 noms de domaines et que les 20 domaines les plus retournés représentent plus du tiers des résultats. La répartition est donc clairement inégale. Dans le calcul 2, les mêmes données sont réparties dans quatre catégories, ce qui réduit l'étalement des résultats. Néanmoins, bien que ce chiffre montre une plus grande équité, ce n'est pas signe que les données sont réparties équitablement. Google Actualités favorise encore les contenus provenant des domaines locaux et régionaux, ce que la figure 2 a permis de montrer.

Les calculs 3 et 4 du coefficient de Gini sont utilisés pour reproduire la figure 1 de la recherche de Fischer *et al.* (2020, p.22) (figure 4 ci-bas). Les mêmes calculs effectués ci-haut sont reproduits, mais ils sont subdivisés par mots-clés. Il y a donc un coefficient de Gini pour chacun des mots-clés utilisés et une moyenne est calculée pour retourner un seul coefficient.

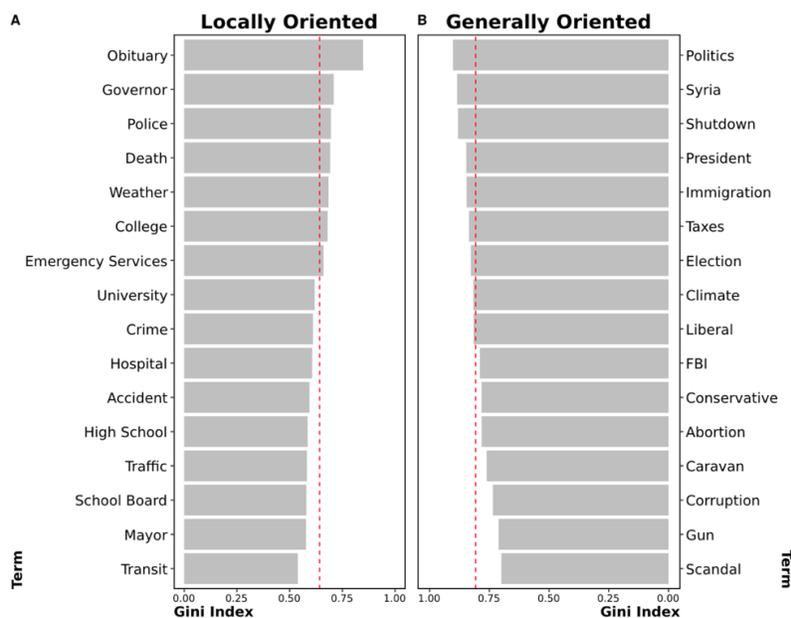


Figure 4. La figure 1 de la recherche Fischer *et al.*

Le calcul 3 reprend la forme du calcul 1, soit pour l'ensemble des domaines retournées par Google Actualités, et retourne un indice de Gini global de 0,8659. Les coefficients de Gini pour chacun des mots-clés sont disponibles aux figures 5 et 6. L'indice de Gini est moins élevé que pour le premier calcul. Cela montre que l'écart entre les domaines est plus faible. Il y a fort probablement moins de domaines contenant qu'un seul résultat réduisant le bruit à l'intérieur des recherches par mots-clés.

La disparité dans les données est aussi plus faible pour les types de mots-clés « local » (0,8369) que pour les mots-clés classés dans les catégories « mixte » (0,8653) et « national » (0,8954). Cela signifie que Google Actualités favorise toujours un petit nombre de médias au détriment des autres, mais pas de la même manière avec tous les mots-clés. Le fait que les distribution des recherches comprenant des mots-clés « local » affiche « plus » d'égalité que les autres rejoint les constatations de Fischer *et al.* (2020) : « *The variation across terms is minimal. The set of locally-oriented terms appears to produce a lower level of inequality, on average, than the set of general terms* » (p. 5). Cela dit, la différence n'est pas énorme.

Reproduction Fischer — Figure 1. Mots-clés orientés mixte et national

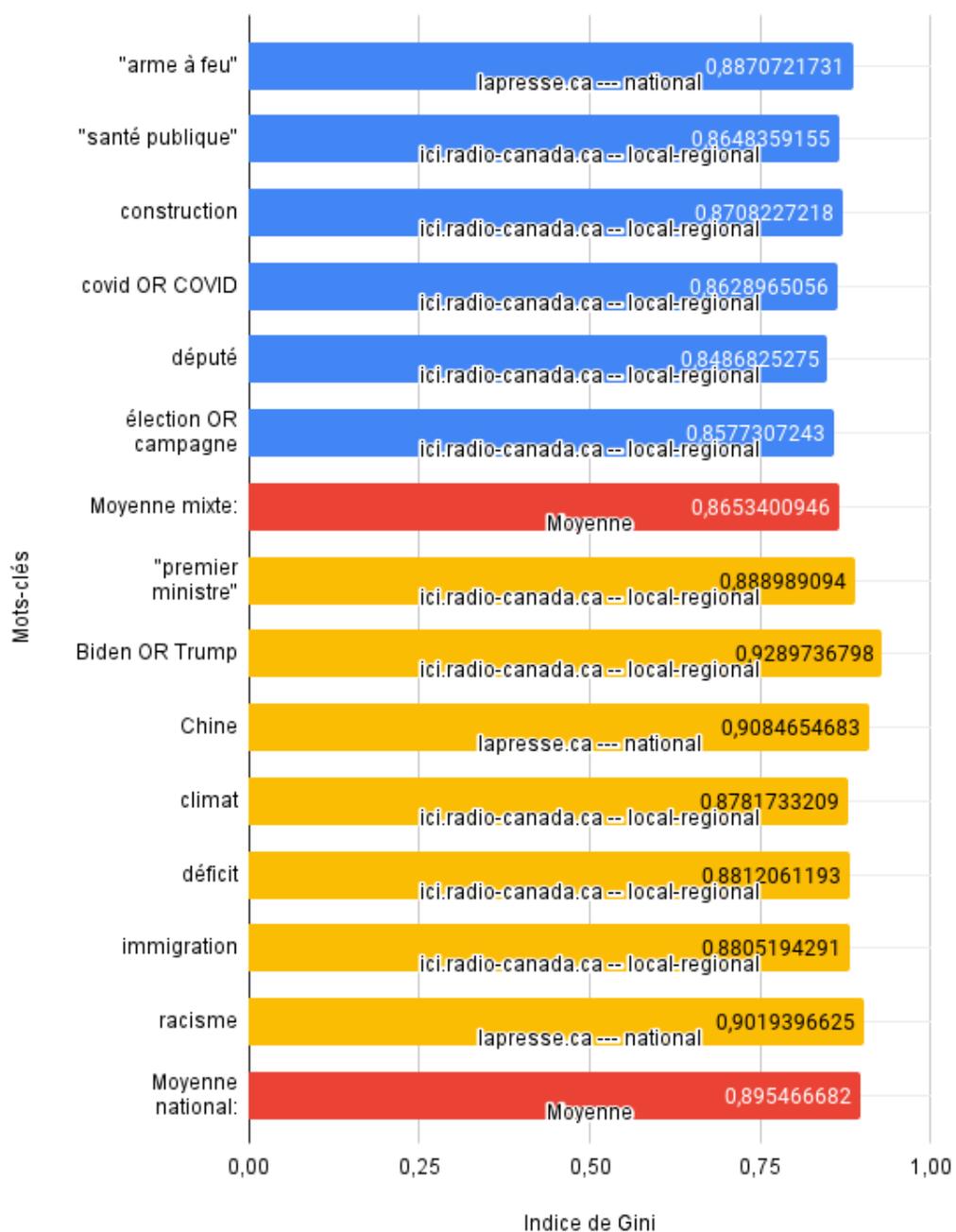


Figure 5. Première reproduction de la figure 1 de Fischer et al. Coefficient de Gini pour les mots-clés orientés national (jaune) et mixte (bleu). Chaque mot-clé est étiqueté avec le média retourné le plus souvent et sa catégorie.

Reproduction Fischer — Figure 1. Mots-clés orientés local

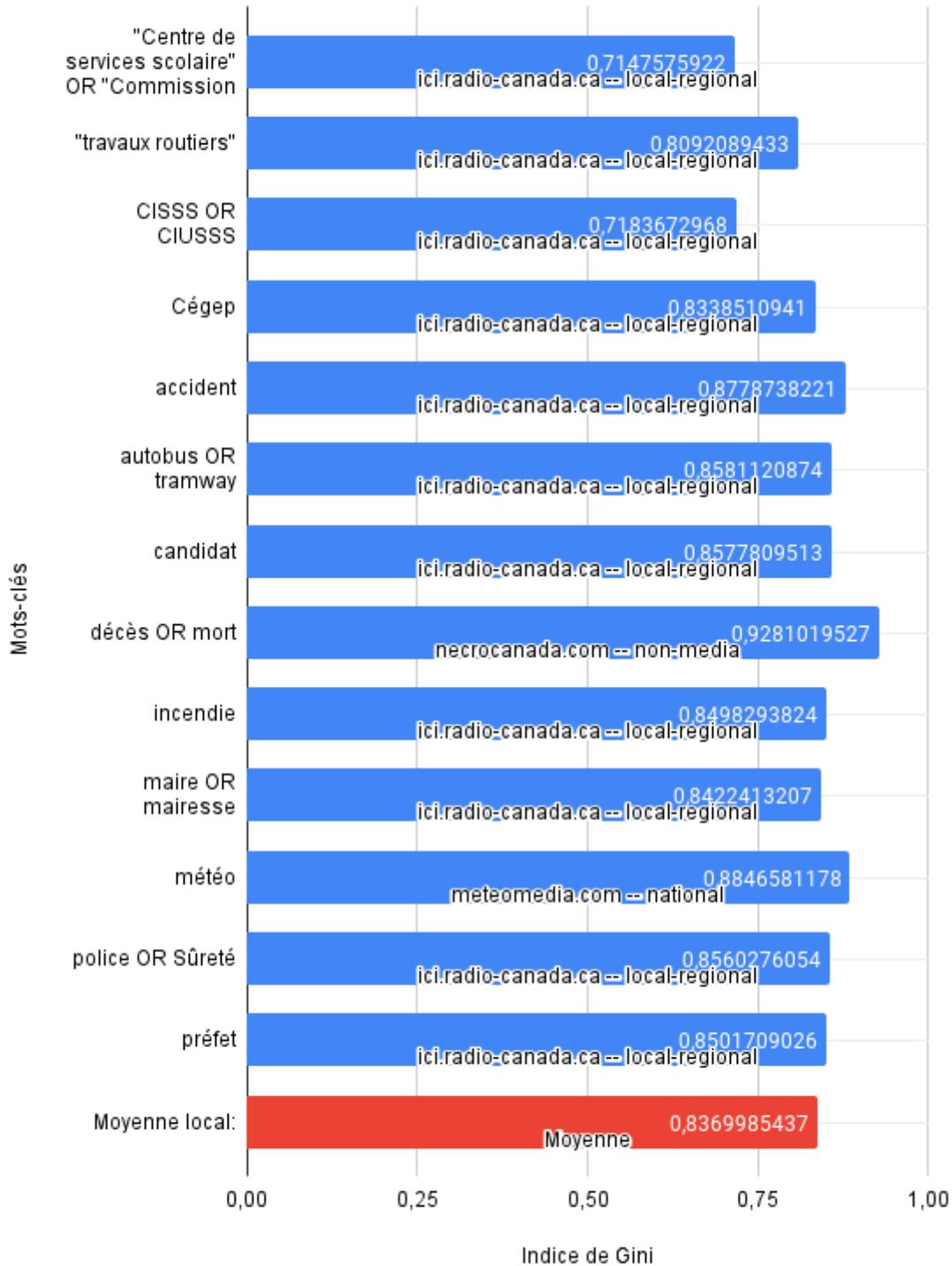


Figure 6. Première reproduction de la figure 1 de Fischer. Coefficient de Gini pour les mots-clés orientés localement. Chaque mot-clé est étiqueté avec le média retourné le plus souvent et sa catégorie.

On remarque à la lecture des figures 5 et 6 que « ici.radio-canada.ca — local-régional » est retourné très fréquemment. Le média figure au premier rang des médias les plus fréquemment retournés pour 21 des 26 mots-clés. *La Presse* suit avec 3 occurrences, alors que necrocanada.com et meteomedia.com comptent une occurrence chacun.

Par ailleurs, parmi les mots-clés orientés « local », « “Centre de services scolaire” OR “Commission scolaire” » et « CIUSSS OR CIUSSS » sont les deux où la disparité dans les résultats est la plus faible avec des coefficients de Gini de 0,71, alors qu’à l’opposé, « décès OR mort » est le plus disparate avec un Gini de 0,92. Le cas de ce dernier mot-clé est fascinant. Le domaine necrocanada.com y compte une énorme partie des résultats avec 8 963 occurrences, ou 43,59 %. C’est plus de cinq fois la fréquence du journaldequebec.com qui arrive au deuxième rang avec 1 577 occurrences.

Le calcul 4 reprend la forme du calcul 2. Les figures 7 et 8 présentent le coefficient de Gini pour chacun de nos mots-clés. Il s’agit aussi de la deuxième reproduction de la figure 1 (figure 4) de Fischer *et al.* (2020). Le coefficient de Gini moyen retourné est de 0,5030. La moyenne est plus élevée que dans le deuxième calcul, ce qui laisse croire qu’il y a plus d’inégalité dans la répartition des catégories de médias, lorsqu’on les sépare par mot-clé. Qui plus est, la moyenne de l’indice de Gini pour les mots-clés orientés « local » (0,5229) est plus élevée que la moyenne totale, que la moyenne des mots-clés « mixte » (0,5180) et que la moyenne « national » (0,4532). Cela signifie que pour des recherches contenant des mots-clés de types « local », les résultats de Google Actualités sont moins équitables. Autrement dit, l’agrégateur retourne plus fréquemment certaines catégories de médias au détriment des autres et dans ce cas-ce, ce sont les médias locaux et régionaux qui sont favorisés.

Reproduction Fischer — Figure 1. Mots-clés orientés local par catégorie de médias

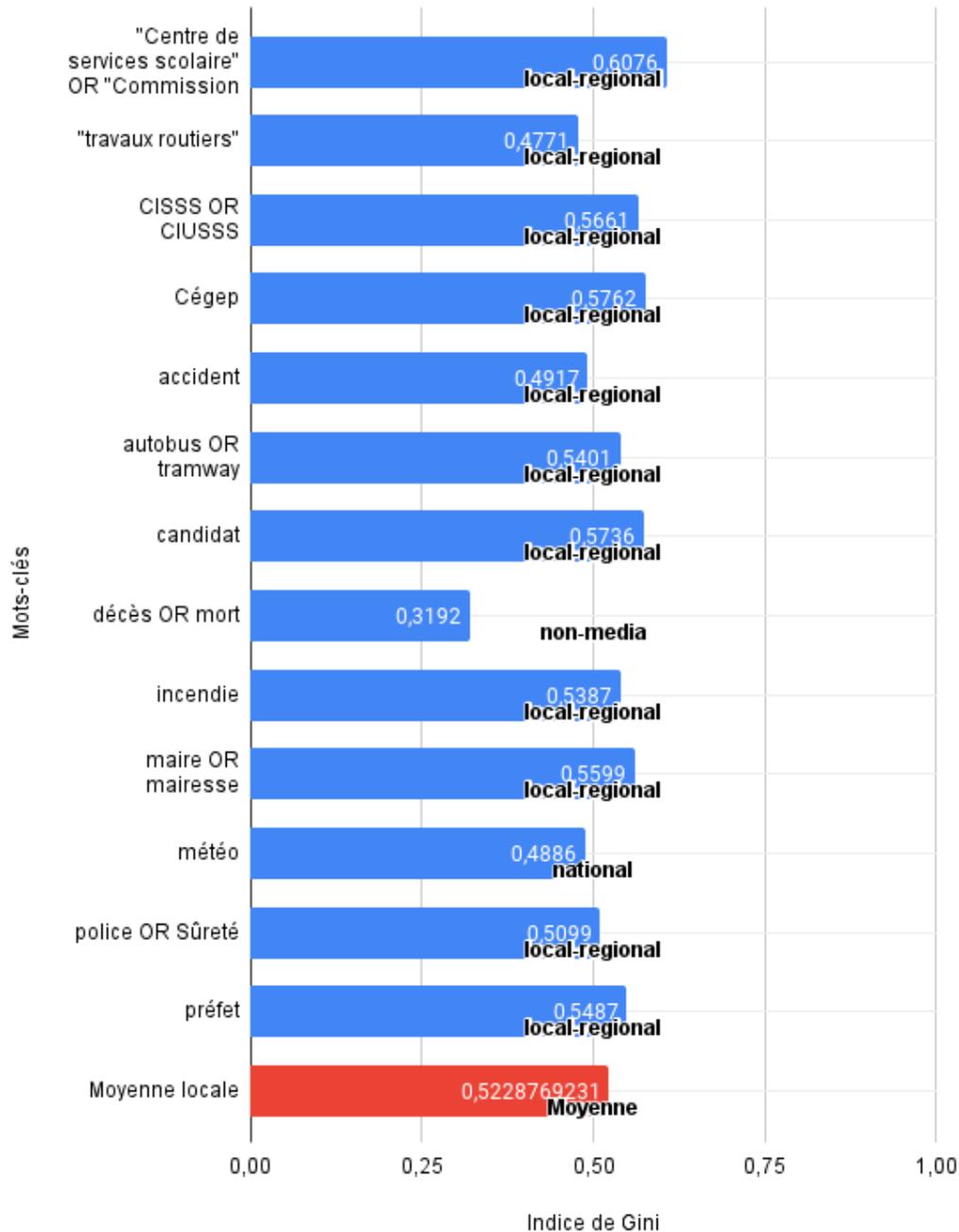


Figure 7. Deuxième reproduction de la figure 1 de Fischer. Coefficient de Gini pour chaque mot-clé orienté localement. Moyenne pour les quatre catégories de média. L'étiquette indique la catégorie de média la plus fréquemment retournée.

Reproduction Fischer — Figure 1. Mots-clés orientés mixte et national par catégorie de médias

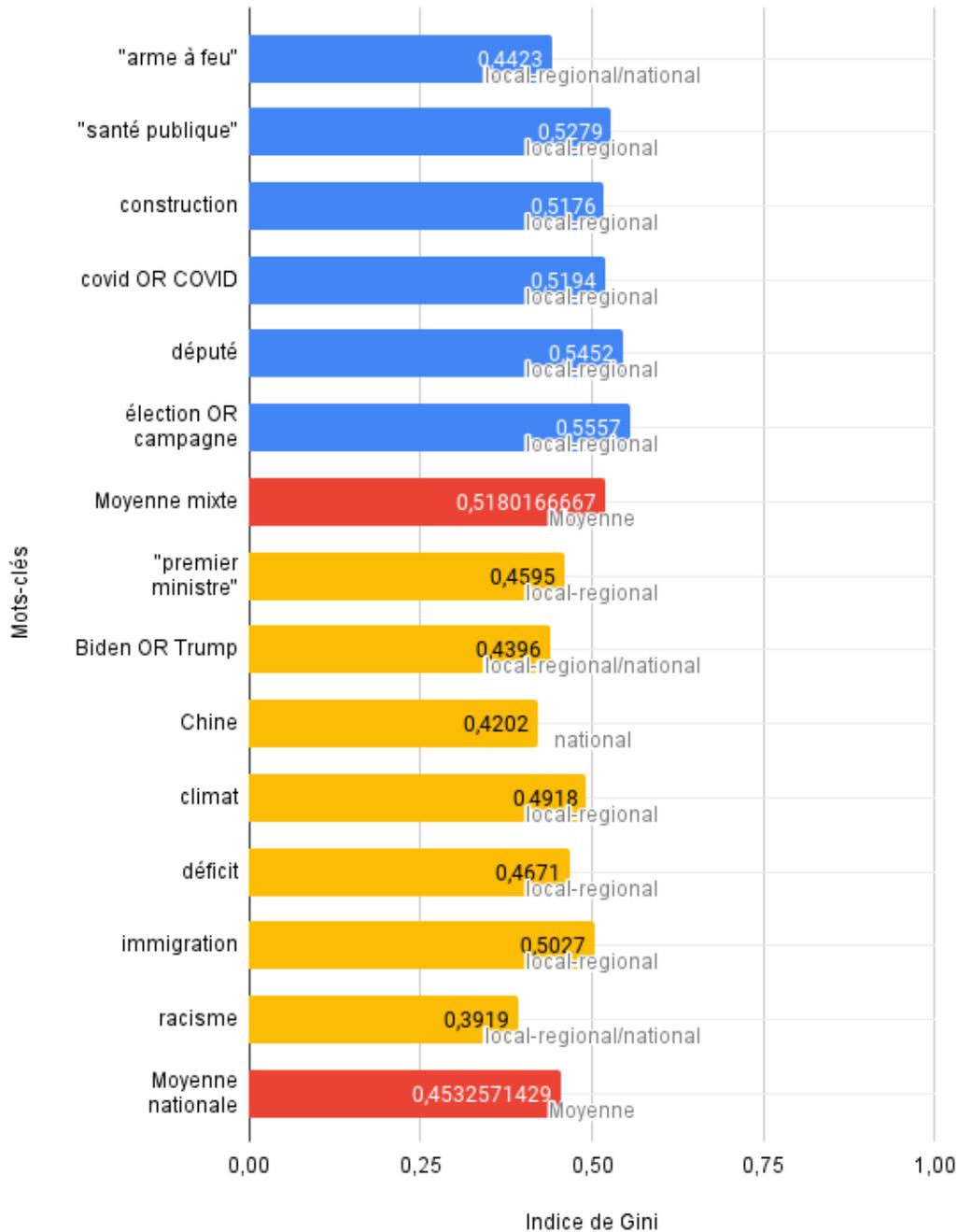


Figure 8. Deuxième reproduction de la figure 1 de Fischer. Coefficient de Gini pour chaque mot-clé orienté mixte (bleu) et national (jaune). Moyenne pour les quatre catégories de média. L'étiquette indique la catégorie de média la plus fréquemment retournée.

En analysant la figure 7, on remarque que « décès OR mort » est le mot-clé où le coefficient de Gini est le plus faible (0,3192), ce qui est à l'opposé de la figure 6. Cela s'explique par le fait que la répartition par catégorie de média (local/régional, national, hors Québec et non médias) est beaucoup plus égale que la distribution des médias. Autrement dit, certains médias se retrouvent plus souvent que d'autres dans les résultats — comme necrocanada.ca (nous y reviendrons). Mais globalement, en les regroupant par catégorie, les résultats sont mieux répartis. En effet, qu'ils proviennent de médias « locaux/régionaux » ou de médias « nationaux », les contenus sont mis de l'avant par Google Actualités avec un degré de prééminence semblable, ce qui tend à démontrer une certaine équité dans les résultats.

Aussi, à la lecture des figures 7 et 8, il est possible de constater que les catégories de médias affichées sont réparties plus équitablement lors des recherches avec les mots-clés « national » que pour les deux autres. En regardant simplement le graphique « national » de la figure 1, au début du chapitre, il est possible de constater que l'écart entre les différentes catégories est plus petit avec des mots-clés de type « national » que pour les autres, mais toujours à l'avantage des médias locaux comme le montre la figure 7 et 8.

Les calculs 1 et 3 du coefficient de Gini montrent que Google Actualités semble favoriser certains médias au détriment des autres, créant une disproportion où les articles de quelques médias sont retournés plus souvent sur la plateforme. Dans la majorité des cas, ce sont les articles provenant des médias locaux et régionaux qui sont favorisés, comme le montre la première partie de ce chapitre.

Les calculs 2 et 4 du coefficient de Gini démontrent, pour leur part, que lorsqu'on regroupe les résultats par catégorie de médias, Google Actualités fait preuve d'une plus grande équité dans la répartition des contenus affichés. Il n'en demeure pas moins que les articles provenant des médias locaux et régionaux sont toujours mis de l'avant, et ce, peu importe que la recherche soit faite, avec des mots-clés nationaux, mixtes ou locaux/régionaux.

En somme, les coefficients de Gini montrent que la distribution des résultats de recherche n'est pas faite de manière équitable sur Google Actualités. Néanmoins, cela ne sous-entend pas que la répartition est déficiente, car comme démontre le tableau 2, bien que la majorité des médias retournés soient des médias hors Québec ou des « non-médias », ceux-ci ont un poids très faible dans l'ensemble des articles de notre base de données. À l'opposé, les contenus provenant des médias « local/régional », ainsi que « national »,

bien que moins nombreux, ont un poids plus important. Il ne faut pas non plus oublier les particularités du Québec où le média le plus proposé par Google Actualités (Radio-Canada) a le mandat de couvrir les événements dans les régions où il compte de nombreuses antennes et que les années 2010 ont vu naître de nombreux nouveaux médias locaux.

4.3.1.2 Reproduction de l'hypothèse 1 avec les critiques de Magnusson

Dans sa critique formulée à Fischer *et al.* (2020), l'économiste Evan Magnusson, de Google, reproche aux chercheurs d'avoir mal calculé leur coefficient de Gini dans leur figure 1 (que l'on peut observer à la figure 4).

They compute the Gini coefficient over the full set of observed news outlets in these county-by-county searches, based on how often each outlet is returned in these searches. However, by doing so, they ignore a key part of how they obtained their data: including location specific context in each query. (Magnusson, 2021, p. 5)

Autrement dit, Fischer *et al.* ont calculé le coefficient de Gini par mot-clé seulement, sans prendre en compte le comté ce qui contrevient à la manière dont les données ont été récoltées. Cette façon de faire, selon Magnusson, « *conflate[s] differences in relevance with inequality* » (*Ibid.* p. 6). Magnusson a donc recalculé les coefficients de Gini pour chacun des mots-clés en les séparant par comté pour ensuite calculer une moyenne. Globalement, il obtient un coefficient de Gini plus faible de 16 % passant de 0,825 à 0,69. Pour chacun des mots-clés, il remarque une baisse dans le coefficient, variant entre 29 % et 80 %. Les figures 9 et 10 montrent les deux tableaux de Magnusson.

Table 1: Gini Coefficients for Local Topic Search Terms

Search Term	Fischer et al. (2020) Gini	Average County Gini	Percent Difference
Accident	0.625	0.227	-63.68%
College	0.700	0.264	-62.34%
Crime	0.668	0.277	-58.59%
Death	0.738	0.370	-49.86%
Emergency Services	0.691	0.386	-44.21%
Governor	0.745	0.336	-54.97%
High School	0.615	0.172	-72.12%
Hospital	0.635	0.246	-61.28%
Mayor	0.560	0.109	-80.50%
Obituary	0.845	0.599	-29.16%
Police	0.744	0.386	-48.17%
School Board	0.584	0.148	-74.56%
Traffic	0.592	0.195	-67.08%
Transit	0.556	0.236	-57.49%
University	0.684	0.304	-55.65%
Weather	0.714	0.366	-48.72%

Figure 9. Écart des coefficients de Gini calculés par Magnusson et par Fischer et al. pour les recherches avec les mots-clés locaux.

Table 2: Gini Coefficients for General-Interest Topic Search Terms

Search Term	Fischer et al. (2020) Gini	Average County Gini	Percent Difference
Abortion	0.782	0.393	-49.72%
Caravan	0.761	0.413	-45.77%
Climate	0.817	0.470	-42.47%
Conservative	0.783	0.456	-41.69%
Corruption	0.735	0.345	-53.06%
Election	0.827	0.475	-42.56%
FBI	0.790	0.449	-43.20%
Gun	0.711	0.283	-60.16%
Immigration	0.845	0.508	-39.89%
Liberal	0.816	0.476	-41.46%
Politics	0.902	0.629	-30.25%
President	0.847	0.524	-38.05%
Scandal	0.699	0.222	-68.27%
Shutdown	0.881	0.593	-32.69%
Syria	0.885	0.587	-33.63%
Taxes	0.835	0.465	-44.28%

Figure 10. Écart des coefficients de Gini calculés par Magnusson et par Fischer et al. pour les recherches avec les mots-clés nationaux.

Pour les besoins de la recherche, cette critique a été prise en compte pour recalculer l'indice de Gini de chacun de mots-clés en reprenant le calcul 4 de la précédente section, soit par catégorie de média. Il est important de noter que Magnusson, ainsi que Fischer *et al.*, ont fait leur calcul par nom de domaine (par média). Mais comme nous l'avons vu précédemment, le calcul par catégorie est davantage représentatif de l'objectif de la recherche qui est d'observer la part des médias locaux et régionaux dans les résultats fournis par Google Actualités, et non la part de chaque média dans ces mêmes résultats. Le recalcul a donné le tableau 13 ci-bas.

Mot-clé	Coeff. de Gini, méthode de Fischer <i>et al.</i>	Coeff. de Gini, méthode de Magnusson	Écart (%)
« Centre de services scolaires » OR « Commission scolaire »	0,6076	0,6186	1,81
« travaux routiers »	0,4771	0,5667	18,78
CISSS OR CIUSSS	0,5661	0,5955	5,20
Cégep	0,5762	0,5979	3,77
accident	0,4917	0,5658	15,07
autobus OR tramway	0,5401	0,5636	4,36
candidat	0,5736	0,6060	5,65
décès OR mort	0,3192	0,4871	52,60
incendie	0,5387	0,5851	8,62
maire OR mairesse	0,5599	0,5961	6,47
météo	0,4886	0,5760	17,89
police OR Sûreté	0,5099	0,5582	9,48
préfet	0,5487	0,6167	12,39
Moyenne locale	0,5229	0,5795	10,83
« arme à feu »	0,4423	0,5445	23,12
« santé publique »	0,5279	0,5684	7,67
construction	0,5176	0,5608	8,35
covid OR COVID	0,5194	0,5766	11,02
député	0,5452	0,5781	6,03
élection OR campagne	0,5557	0,5898	6,14
Moyenne mixte	0,5180	0,5697	9,98
« premier ministre »	0,4595	0,5355	16,54
Biden OR Trump	0,4396	0,5594	27,26
Chine	0,4202	0,5225	24,34
climat	0,4918	0,5423	10,26
déficit	0,4671	0,5268	12,78
immigration	0,5027	0,5511	9,62
racisme	0,3919	0,5166	31,81
Moyenne nationale	0,4533	0,5318	18,32

Tableau 13. Reproduction du calcul de l'indice de Gini selon le mot-clé en filtrant par villes.

L'élément le plus frappant de ce recalcul est que pour tous les mots-clés, le coefficient de Gini calculé selon la méthode de Magnusson est plus élevé, ce qui témoigne d'une augmentation de l'inégalité dans la répartition des catégories de médias lorsqu'on effectue le calcul par ville. Le Gini moyen a grimpé de 12,44 %

passant de 0,5030 à 0,5656. Parmi les hausses les plus élevées, il y a celle du mot-clé « décès OR mort » qui augmente de 52 % passant de 0,31 à 0,48 et celle du mot-clé « racisme » où le coefficient prend 31 % passant de 0,39 à 0,51. Ces changements dans le coefficient de Gini montrent qu'à l'intérieur des villes étudiées, l'écart entre la répartition des résultats selon le type de mots-clés est davantage inégal.

4.3.1.3 Hypothèse 2

Fischer *et al.* (2020) estiment que Google Actualités pourrait stratifier les résultats selon le mot-clé utilisé, favorisant ainsi plus de contenus locaux lorsque le mot-clé l'est et plus de contenus nationaux lors des recherches avec des mots-clés du même genre. Les chercheurs mentionnent également que les utilisateurs ne vont pas plus loin que les premiers résultats, s'arrêtant souvent à la première page et se limitant même aux 10 premiers. Ils ont donc construit une figure (voir figure 11) comprenant différents graphiques linéaires qui montrent la fréquence relative des différentes catégories de médias par mot-clé selon leur rang dans les résultats de recherche (de 1 à 100).

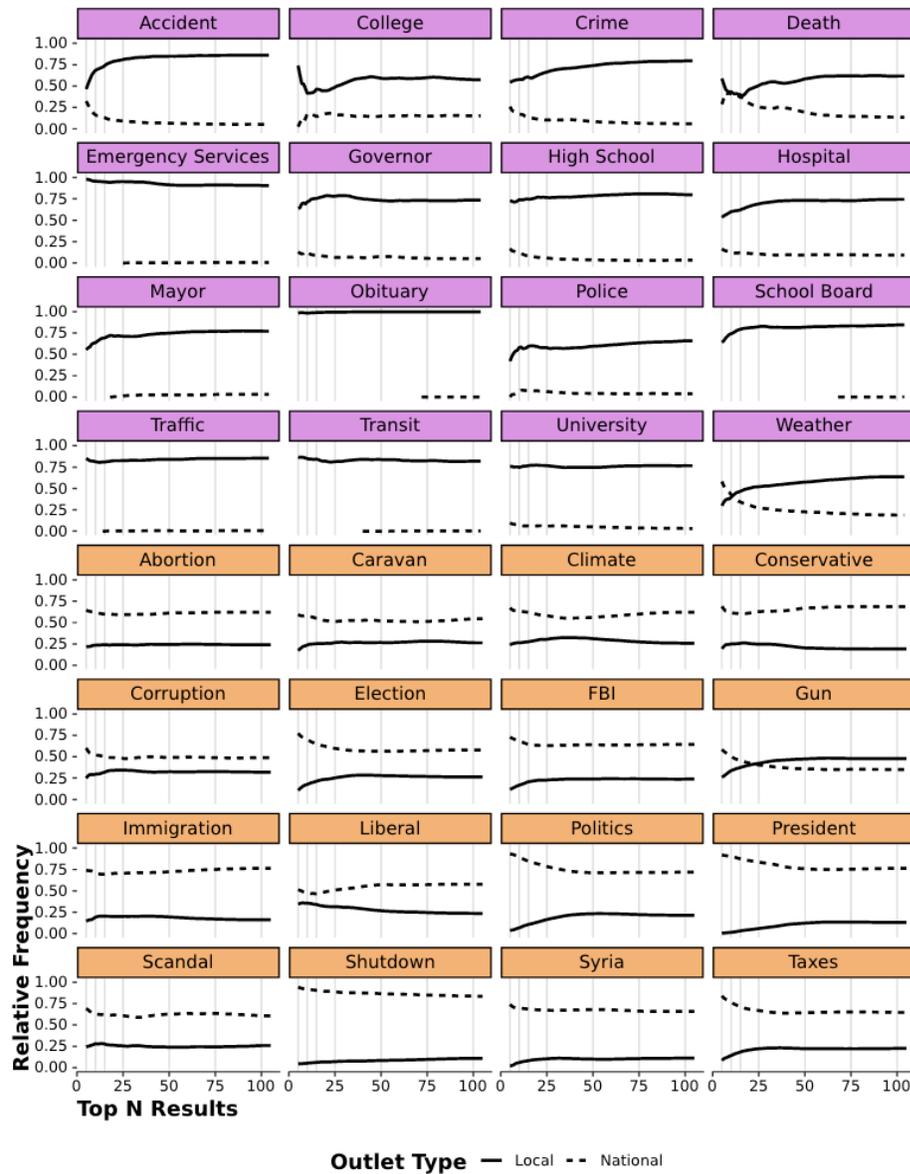


Figure 11. « Figure 2 : The relative frequency of local and national outlets for the top n results of each search query. The first sixteen purple-labeled terms are locally-oriented terms. The last sixteen orange-labeled terms are general terms. » (Fischer et al., 2020, p. 23)

La même approche a été utilisée pour produire les figures 12, 13 et 14. Elles présentent comment Google Actualités a réparti nos mots-clés selon la catégorie de média et leur rang dans les résultats de recherche.

La fréquence relative des mots-clés locaux selon son rang dans les résultats de recherche

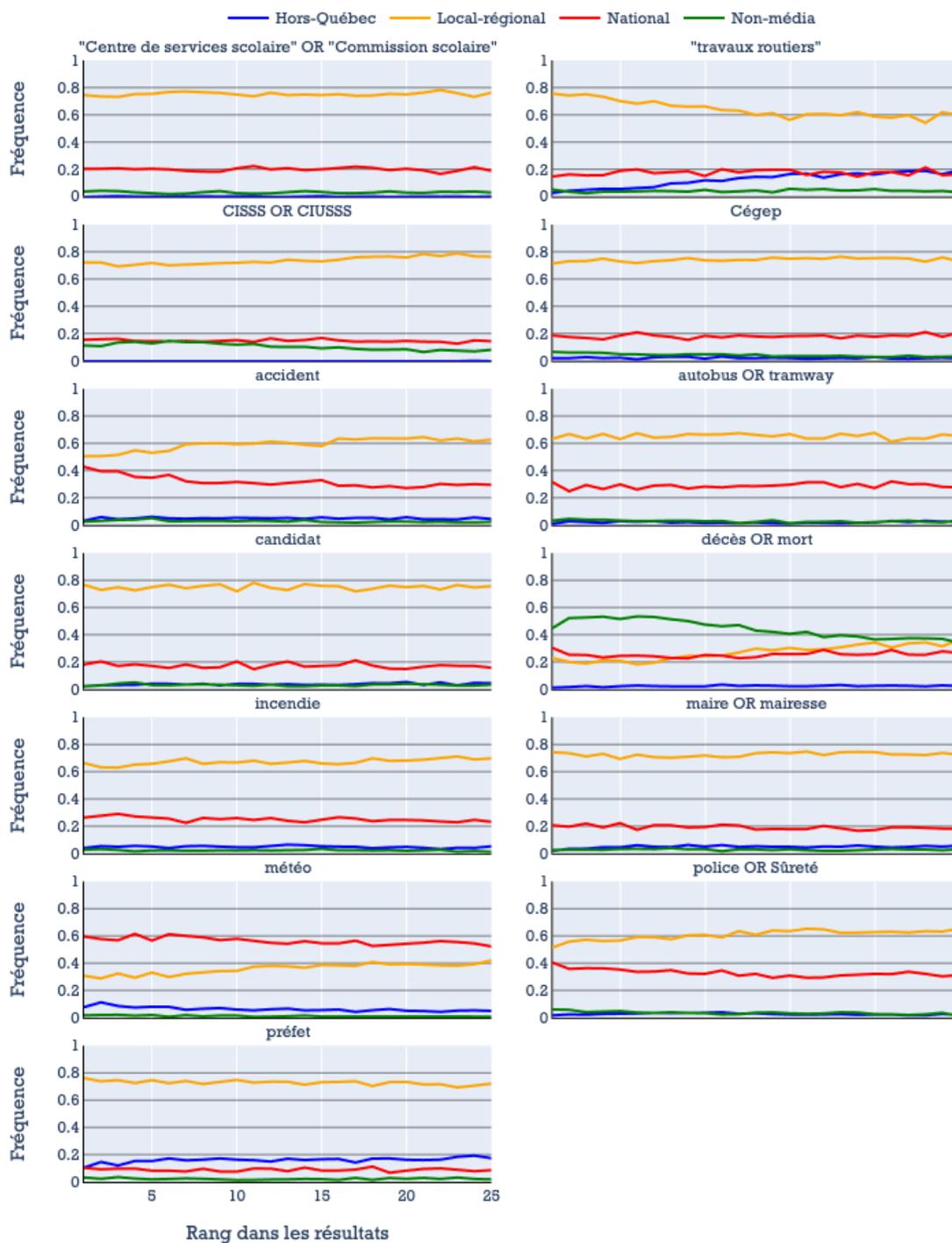


Figure 12. Fréquence relative des résultats pour les requêtes contenant les 13 mots-clés locaux

L'analyse des recherches contenant des mots-clés locaux montre clairement ce qui est illustré par la figure 1 et qui émerge depuis le début de ce chapitre : les médias locaux et régionaux sont la plupart du temps mis de l'avant par Google Actualités. Ils le sont encore davantage pour des recherches contenant des mots-clés locaux. Dans 11 cas sur 13, les médias locaux et régionaux sont retournés à des fréquences variant entre 5 et 8 fois sur 10. Dans un seul cas, pour le mot-clé « décès OR mort », les non-médias sont plus importants, mais cela vient de la présence du site Nécrologie Canada, une aberration dans nos résultats.

Dans un seul cas, pour le mot-clé « météo », les médias nationaux sont plus fréquemment retournés. On remarque même que 6 fois sur 10, dans les 5 premiers résultats, que les médias nationaux sont plus souvent retournés alors que les médias locaux/régionaux sont retournés 3 fois sur 10. L'écart se rétrécit au fur et à mesure qu'on descend dans le rang des résultats. Il est important de noter, comme le montrait la figure 6, que pour ce mot-clé, le média météo.com était le plus fréquemment affiché avec 4 121 occurrences, soit une fois et demie plus souvent que le *Journal de Québec* qui se trouve au deuxième rang avec 2 708 occurrences. Finalement, en observant les graphiques, on remarque, dans la majorité des cas, que la fréquence relative des non-médias et des médias hors Québec est très faible, variant généralement entre 0 et 0,1.

La fréquence relative des mots-clés mixtes selon son rang dans les résultats de recherche

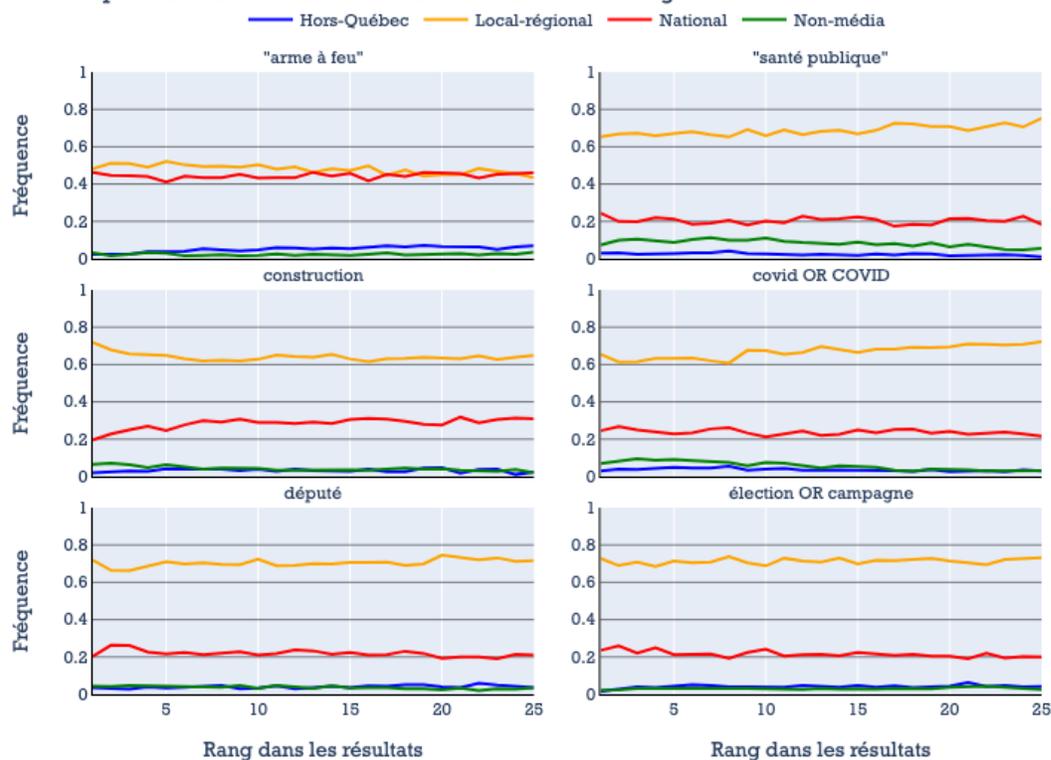


Figure 13. Fréquence des résultats pour les requêtes contenant les six mots-clés mixtes

L'analyse des résultats pour les recherches contenant les mots-clés mixtes montre qu'à l'exception du mot-clé « arme à feu », les médias locaux et régionaux sont davantage mis de l'avant que les autres catégories par Google Actualités, à une fréquence de plus de six fois sur 10, et ce, peu importe la position du résultat dans la recherche. « Arme à feu » est le seul mot-clé qui rejoint l'hypothèse affirmant que les recherches contenant les mots-clés mixtes allaient être près de la parité dans la répartition.

De plus, la fréquence des non-médias est pratiquement nulle dans les recherches, à l'exception du mot-clé « santé publique » ce qui peut s'expliquer par la COVID-19 et la diffusion constante d'informations relatives à la santé publique par les gouvernements. Justement parmi les 38 domaines classés comme non-média, quelques-uns réfèrent à des sites gouvernementaux comme mobile.inspq.qc.ca avec 50

occurrences et www.indexsante.ca avec 31 occurrences. Néanmoins, la fréquence de ces résultats aberrants demeure sous les 2 fois sur 10.

Qui plus est, la fréquence relative des médias hors Québec est superficielle. Cette figure montre aussi clairement la répartition des résultats pour le mot-clé « élection OR campagne », où les médias locaux et régionaux sont beaucoup bien davantage mis de l'avant que les médias nationaux, et ce, malgré une campagne électorale fédérale qui a eu lieu durant la saisie, ce qui ultimement aurait pu hausser le nombre de contenus hors Québec.

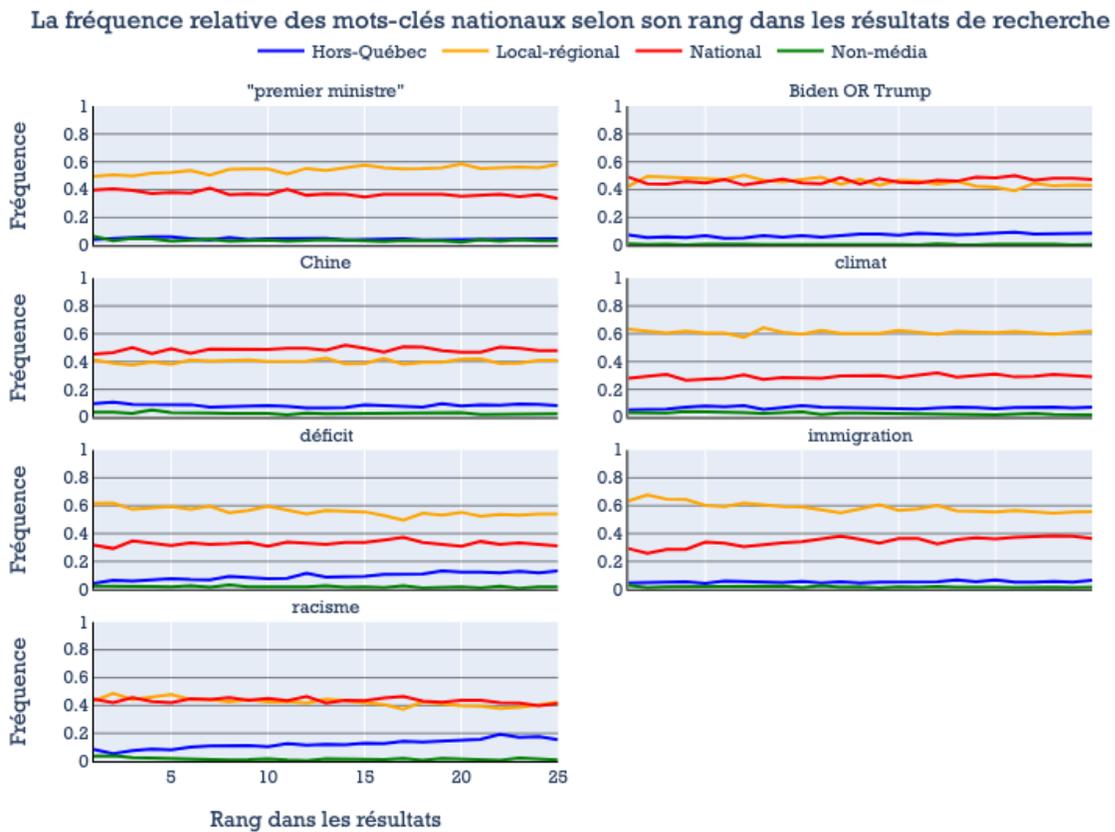


Figure 14. Fréquence des résultats pour les requêtes contenant les sept mots-clés nationaux

Pour ce qui est des recherches comprenant des mots-clés nationaux, comme nous l'avons relevé à plusieurs reprises dans ce mémoire, l'écart entre les médias locaux/régionaux et nationaux est moins

grand pour ce type de mots-clés que pour les précédents. Cela se constate simplement en regardant la figure 14. Malgré tout, les médias locaux/régionaux sont encore soit plus nombreux ou à égalité avec les médias nationaux dans leur fréquence relative. Le seul mot-clé où les médias nationaux sont proéminents est « Chine » et encore là, l'écart est très mince.

Le cas du terme « racisme » est plutôt intéressant : les deux principales catégories de médias (local/régional et national) s'échangent le titre du plus retourné selon le rang dans les résultats. Pour ce mot-clé, le nombre de contenus provenant des médias hors Québec semble aussi progresser plus l'on s'éloigne des premiers résultats montant jusqu'à une fréquence de 2 fois sur 10. Du lot, on note radio-canada – hors-Québec (avec 302 occurrences) ainsi que www.ouest-france.fr (avec 140 occurrences). En analysant les données plus précisément, il ne semble pas y avoir un média qui est retourné plus qu'un autre entre la 15^e et 25^e position. Cela semble plutôt être une accumulation de tous les médias de cette catégorie.

4.4 Le cas de Nécrologie Canada

Le nom de domaine necrocanada.com se classe au dixième rang des plus retournés avec un poids de 2,45 % ou 12 001 occurrences dans la base de données complète (Annexe 5). La présence de ce domaine dans nos résultats peut étonner.

On retrouve ce domaine majoritairement dans les recherches comprenant le mot-clé « décès OR mort », à hauteur de 74 %. Cette expression a été incluse dans les mots-clés régionaux pour repérer des articles sur des faits divers dans les résultats de recherche. Le domaine necrocanada.com ressort également, dans une moindre mesure, dans les recherches avec les expressions « CISSS or CIUSSS » (1 185 occurrences ou 9,87 %), « santé publique » (1054 occurrences ou 8,78 %) et « covid OR COVID » (527 occurrences ou 4,39 %). Ce sont tous des domaines liés à la santé et, indirectement, à la mort. La présence de ce nom de domaine dans les autres mots-clés est anecdotique.

Le site ne devrait pas se retrouver dans les résultats de recherche de Google Actualités. Bien qu'il « informe sur les avis de décès » (Nécrologie Canada, s. d.), cela ne s'inscrit pas comme un contenu journalistique à proprement parler. De plus, le contenu des avis de décès est une simple reprise des informations publiées par les maisons funéraires. Il s'agit donc plutôt d'un site d'agrégation ou « d'archivage » de contenu. De plus, il n'y a pas de mention sur le site de Nécrologie Canada qui fasse référence à un partenariat

quelconque avec Google Actualités, comme on en retrouve sur certains sites. La présence de ce site dans les résultats de Google Actualités semble donc relever d'une erreur de l'algorithme.

4.5 La temporalité des contenus affichés

À chacune des saisies effectuées pour cette recherche, la date de publication des articles affichés par Google Actualités a été enregistrée lorsque cela était possible. Une formule mathématique a calculé automatiquement la différence en jours entre la date de publication du reportage et la date de saisie. L'objectif était de mesurer la « fraîcheur » des contenus proposés par Google Actualités en lien avec l'hypothèse 4 de la recherche.

Hypothèse #4 : « Considérant la quantité d'information produite, les articles retournés par l'algorithme de Google Actualités sont relativement récents. »

Cet élément n'a pas été initialement intégré dans les hypothèses de recherche, car il était supposé que les contenus seraient frais, ce qui allait de soi considérant la quantité d'articles générés par les médias. Toutefois, en analysant les contenus pour répondre aux précédentes sections, nous avons constaté que l'algorithme de Google Actualités ne semblait pas faire un bon travail pour afficher des contenus frais. Qui plus est, cet ajout au mémoire permet d'aller plus loin que la recherche initiale sur laquelle se base cette recherche, soit celle de Fischer *et al.* qui n'ont pas tenu compte de cette dimension dans leur article.

Sur l'ensemble de la base de données, l'âge moyen des reportages retournés par l'agrégateur est de 128 jours toutes catégories de médias confondues, soit un peu plus de 4 mois. Ce chiffre est très élevé considérant que les médias ont tendance à produire beaucoup de contenus, et ce, quotidiennement. Séparée par catégorie, la moyenne s'élève à 134 jours pour les médias locaux et régionaux et à 140 jours pour les médias nationaux. C'est donc dire que ce sont les médias hors Québec (69 jours) et les non-médias (51 jours) qui ont fait descendre la moyenne. Parmi les noms de domaine provenant de ces deux dernières catégories, on retrouve :

- ciussmcq.ca (367 occurrences [non média]),
- news.ontario.ca (81 occurrences [non média]),
- www.ville.quebec.qc.ca (50 occurrences [non média]),
- www.ouest-france.fr (1933 occurrences [hors Québec]),

- www.lemonde.fr (599 occurrences [hors Québec]) et
- www.lefigaro.fr (372 occurrences [hors Québec]).

Cela nous amène à nous demander si Google Actualités filtre mieux la fraîcheur des contenus non-média, des contenus qui rappelés-le, ne devrait pas se retrouver sur la plateforme.

Lorsqu'on filtre par villes de saisie, Google Actualités affiche une moyenne de fraîcheur de presque un an dans trois villes, soit Kuujuaq (347 jours), Blanc-Sablon (349 jours) et Forestville (364 jours). À l'opposé, lorsque les recherches prenaient place à Montréal, sans le groupe de mots « près de », l'âge moyen des contenus retournés chutait à 9 jours, alors que pour celles contenant le « près de », il passait à 26 jours. Québec figure en deuxième place avec une fréquence de 22 jours.

Plusieurs raisons peuvent expliquer ces écarts dans la « fraîcheur » des reportages selon la ville. Cela pourrait être lié à l'algorithme de Google qui donnerait moins d'importance à la date de publication qu'à la pertinence des contenus dans son affichage. Cela peut aussi être lié au fait que les contenus sont moins nombreux dans certaines régions, ce qui refléterait l'appauvrissement de l'information mentionné précédemment.

Bien que l'échantillon soit mince, avec une seule recherche qui excluait le « près de », il semble que l'ajout de cette mention dans la recherche force Google Actualités à sélectionner du contenu ciblé à la ville, un contenu qu'il juge « le plus pertinent » sans prendre en considération la date de publication. Les figures 15 et 16 suivantes permettent d'observer dans chacune des villes étudiées, la fraîcheur du contenu selon le rang dans les résultats de recherche de l'article de 0 à 25.

L'âge moyen des articles selon la ville et son rang dans les résultats de recherche

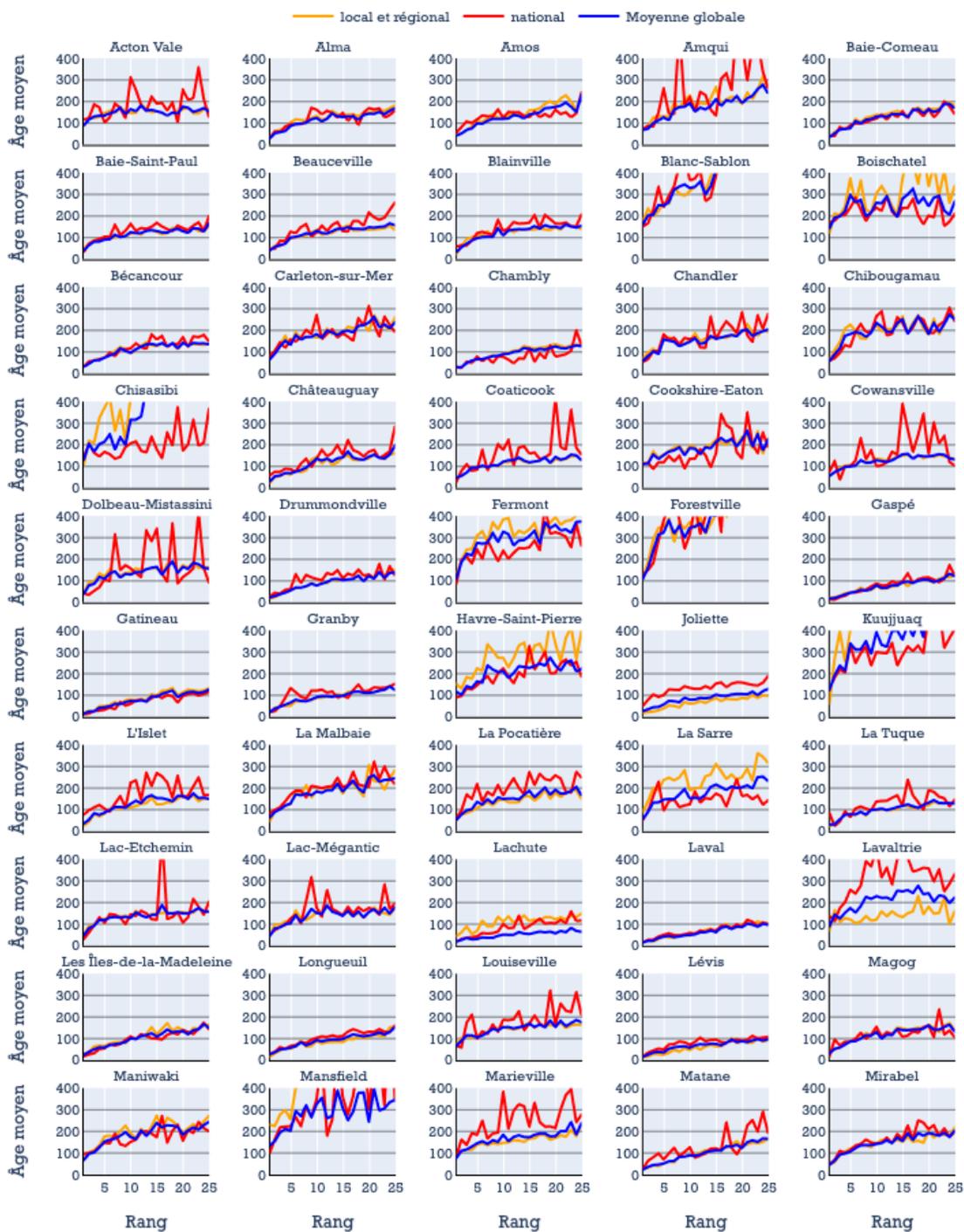


Figure 15. La fraîcheur en jours des articles selon la catégorie pour les 50 premières villes étudiées

L'âge moyen des articles selon la ville et son rang dans les résultats de recherche

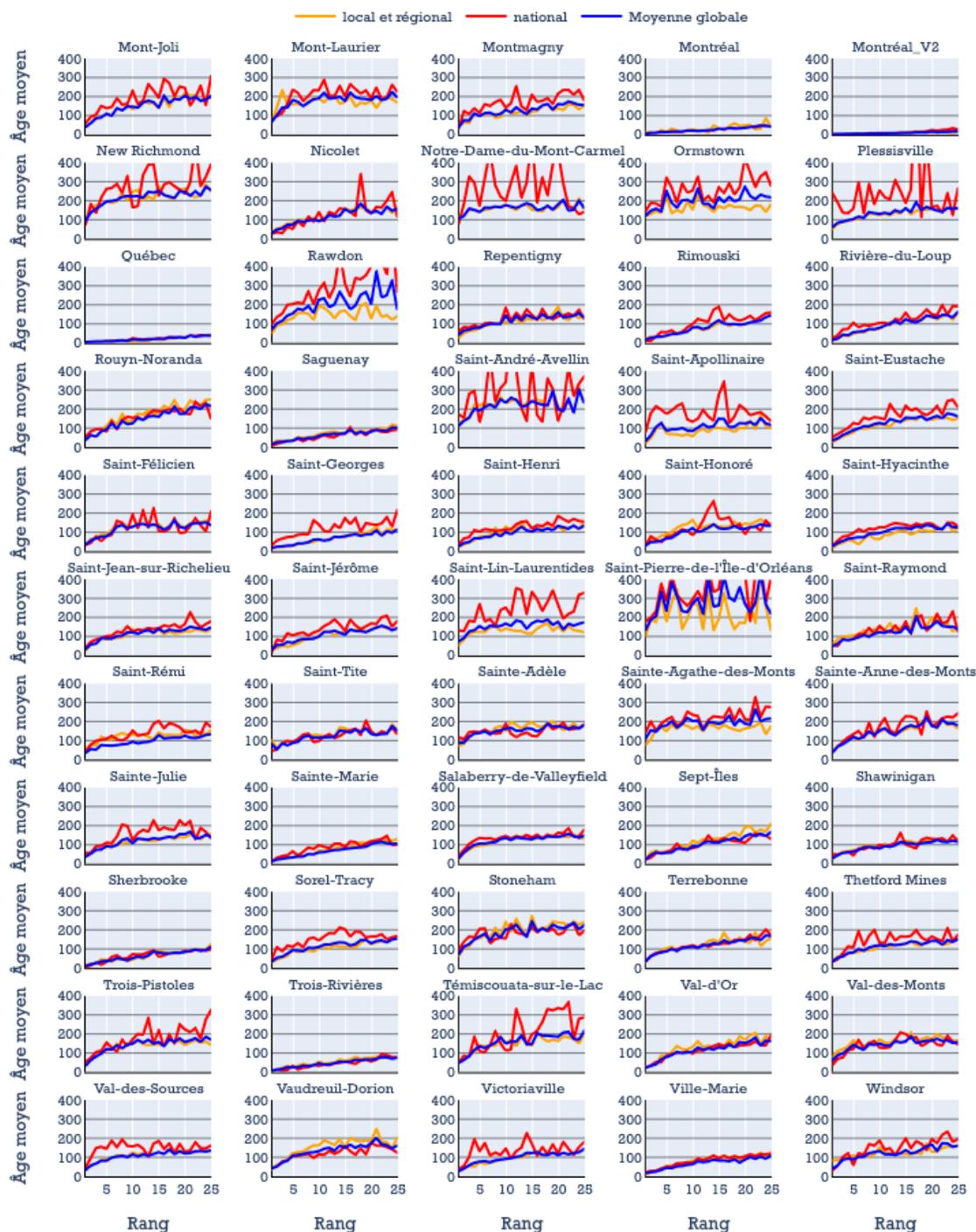


Figure 16. La fraîcheur en jours des articles selon la catégorie pour les 55 dernières villes étudiées

Dans l'ensemble, les 105 graphiques montrent que plus les articles sont loin dans les résultats de recherche, moins les contenus sont frais. Dans tous les cas, les premiers résultats présentent les résultats les plus récents. Par contre, certaines villes affichent des variations inconstantes comme Boischatel, Mansfield, Ormstown et Saint-Pierre-de-l'Île-d'Orléans. Dans peu de villes, les résultats se maintiennent sous la barre des 100 jours, peu importe leur rang dans les résultats. Des villes comme Laval, Longueuil, Gatineau, les deux Montréal, Sherbrooke et Québec font partie du lot.

En analysant les figures 15 et 16, à vue d'œil, on remarque une plus grande fluctuation dans l'âge moyen des médias nationaux (rouge), se détachant souvent vers le haut de la moyenne (bleu). Qui plus est, l'âge moyen des articles retournés dépasse souvent les 200 jours dans les 10 premiers résultats comme à Notre-Dame-du-Mont-Carmel, à New Richmond ou à Amqui.

De leur côté, les médias locaux et régionaux semblent la plupart du temps être près de la moyenne ou en bas à l'exception des villes suivantes qui ont déjà majoritairement des moyennes globales très élevées (supérieures à 300 jours) : Boischatel, Chisasibi, Fermont, Havre-Saint-Pierre, Kuujuaq, La Sarre, Mansfield, Lachute et Vaudreuil-Dorion.

Lorsque ces mêmes données sont séparées selon le mot-clé, on constate que « décès OR mort » est celui qui affiche des contenus les plus à jour avec une moyenne tournant autour de 66 jours alors que « travaux routiers » est le plus élevé avec une moyenne de 220 jours. Un peu comme pour les villes, les raisons qui peuvent expliquer ces écarts sont multiples, mais encore une fois, le nombre d'articles produit pour chacun des mots-clés peut influencer l'âge moyen. Comme notre recherche n'évalue pas la qualité du contenu, il est difficile d'établir si Google Actualités fait un bon travail de curation. Les figures 17 et 18 et 19 permettent d'observer la fraîcheur des articles en moyenne par mot-clé selon le rang dans les résultats de recherche du contenu récolté.

L'âge moyen en jours des articles selon le mot-clé local et son rang dans les résultats de recherche

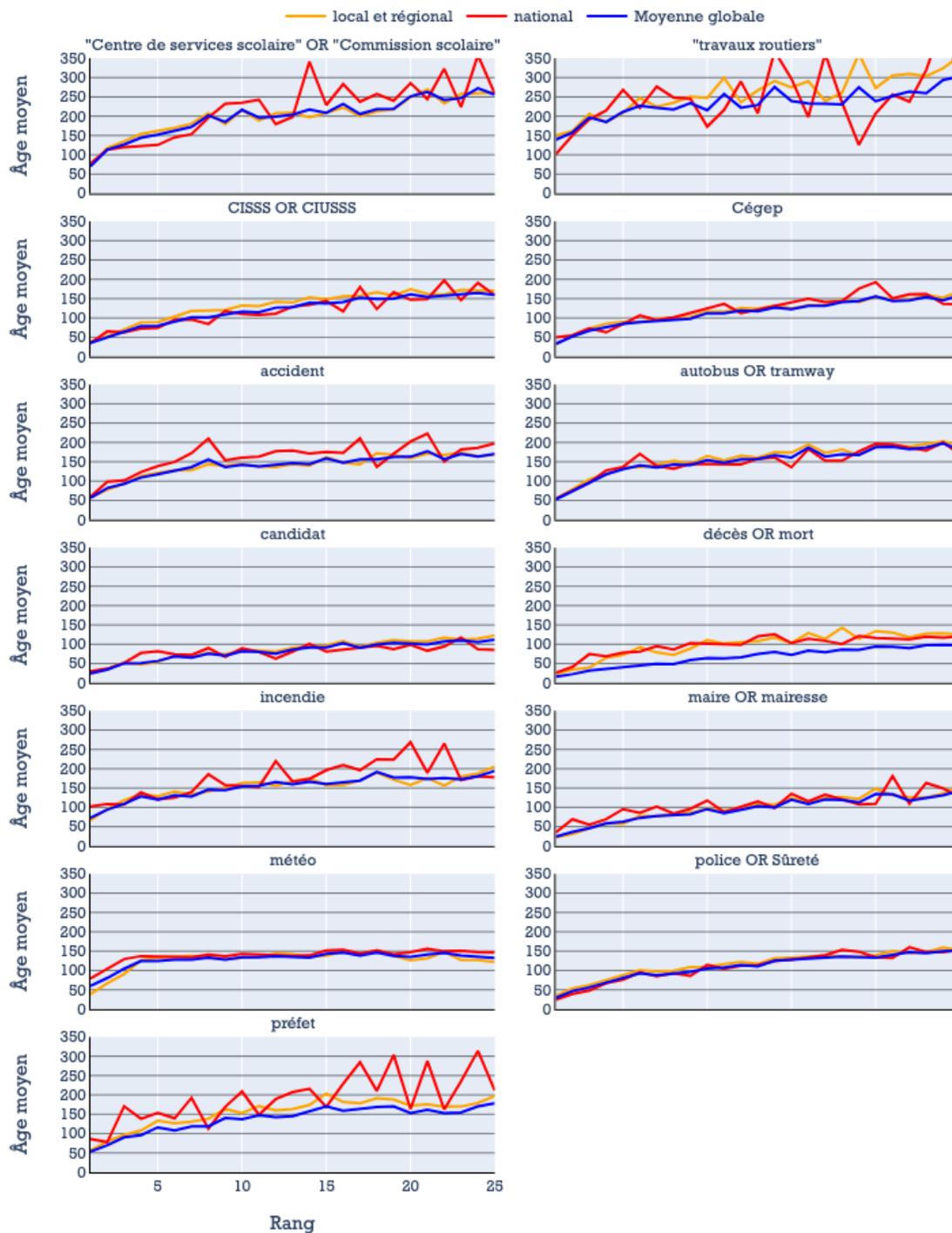


Figure 17. La fraîcheur en jours des articles, par catégorie, pour les mots-clés locaux.

L'âge moyen en jours des articles selon le mot-clé national et son rang dans les résultats de recherche

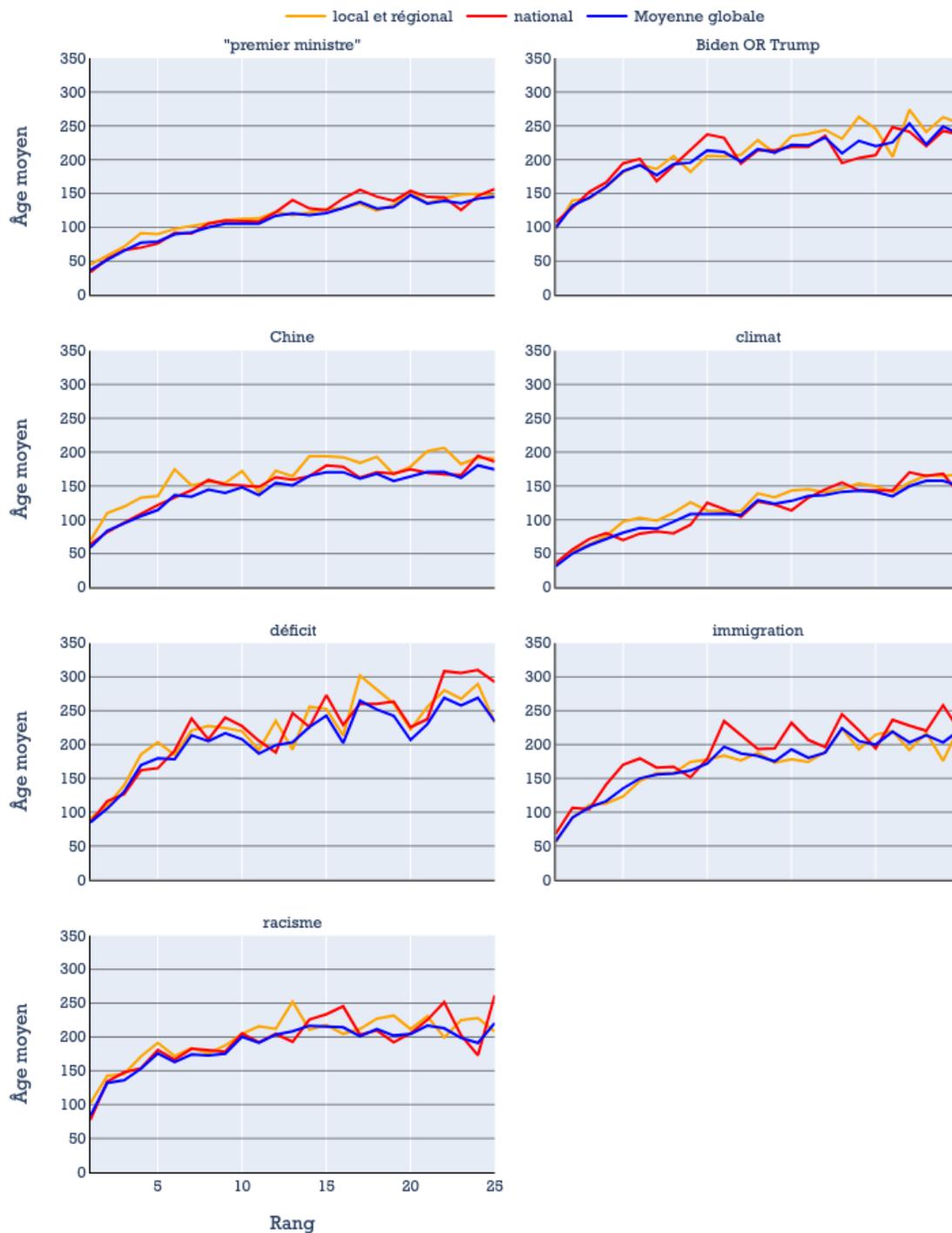


Figure 18. La fraîcheur en jours des articles, par catégorie, pour les mots-clés nationaux.

L'âge moyen en jours des articles selon le mot-clé mixte et son rang dans les résultats de recherche

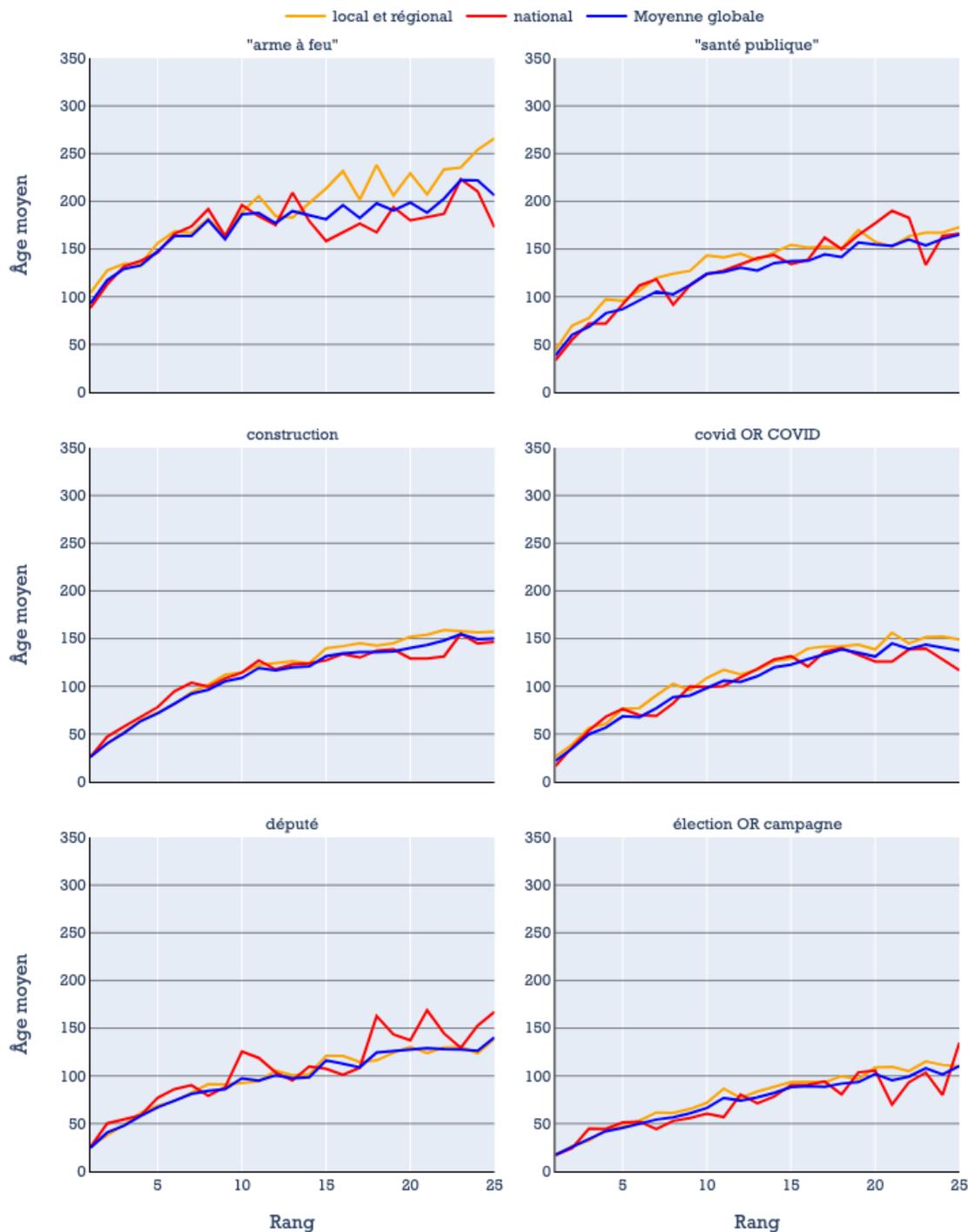


Figure 19. La fraîcheur en jours des articles, par catégorie, pour les mots-clés mixtes

À la lecture de ces trois figures, on remarque que les moyennes par mot-clé sont très élevées et seuls trois se maintiennent en bas de 100 jours, peu importe la position du contenu : « élections OR campagne », « décès OR mort » et « candidat ». Cela dit, les graphiques affichent tous des courbes en progression, avec une corrélation directe entre l'âge des contenus et le rang dans les résultats de recherche. Qui plus est, il ne semble pas y avoir de distinction entre les médias locaux et régionaux, d'une part, et les médias nationaux, d'autre part, du moins dans les premiers résultats. Par ailleurs, il est possible de remarquer que la fraîcheur des médias nationaux fluctue davantage que celle des médias locaux et régionaux.

Bref, un approfondissement quant à la pertinence des contenus affichés ainsi que l'effet du terme « près de » sur la fraîcheur des contenus serait intéressant. Car justement, à première vue, la recherche de Montréal sans le « près de » montre une fraîcheur beaucoup plus basse par mot-clé que la moyenne totale et celle de Montréal sans le « près de ». Le tableau 14 permet de les comparer.

Mot-clé	Montréal sans le « près de »	Montréal avec le « près de »	Moyenne pour toutes les villes
"arme à feu"	10,41	12,06	176,06
"Centre de services scolaires" OR "Commission scolaire"	13,38	43,15	199,06
"premier ministre"	6,75	5,01	110,19
"santé publique"	5,75	7,60	122,44
"travaux routiers"	27,06	121,45	232,43
accident	5,54	30,81	140,44
autobus OR tramway	10,35	12,65	151,51
Biden OR Trump	50,61	60,20	201,80
candidat	4,80	14,96	81,59
Cégep	5,94	15,73	114,10
Chine	7,37	26,10	144,62
CISSS OR CIUSSS	9,21	11,72	121,01
climat	10,28	21,47	113,11
construction	8,53	17,09	110,91
covid OR COVID	1,33	7,98	103,62
décès OR mort	2,47	10,49	67,19
déficit	8,34	56,48	207,11
député	6,88	26,67	96,76
élection OR campagne	5,90	8,49	73,23
immigration	11,28	15,72	171,83
incendie	5,45	26,33	152,52
maire OR mairesse	6,67	8,82	95,31
météo	4,34	71,34	129,04
police OR Sûreté	6,91	8,54	109,56
préfet	4,94	38,82	137,77
racisme	9,25	34,04	187,02
Moyenne totale	9,60	27,45	136,55

Tableau 14. Fraîcheur en jours des articles retournés par Google Actualités selon le mot-clé pour les deux recherches effectuées à partir de Montréal, ainsi que pour l'ensemble de la base de données.

CONCLUSION

D'un côté, les données brutes montrent que Google Actualités ne favorise pas les médias nationaux au détriment des médias locaux ou régionaux. Au contraire, elles indiquent que l'agrégateur de nouvelles semble offrir à ces derniers une part plus que suffisante de couverture, et ce, surtout en dehors des grands centres comme Montréal et Québec. Par contre, l'âge moyen des articles affichés dans les résultats de recherche, toutes catégories de médias confondues, remet en question la pertinence des contenus retournés. Google Actualités affiche de vieux contenus généralement publiés plus de 3 mois avant la date de la requête. Ainsi, bien que les contenus des médias locaux et régionaux soient efficacement mis de l'avant, la qualité des résultats, elle, vient mitiger le constat initial.

Qui plus est, une bonne partie de ce constat quant à la quantité de médias retournés par Google Actualités repose sur le fait que Radio-Canada a une grande notoriété, compte plusieurs antennes partout au Québec et offre, par sa mission, du contenu local et régional. Ce contenu est énormément mis de l'avant par Google dans les régions.

Cependant, Google Actualités semble aussi favoriser certains domaines plus que d'autres. L'importance des 20 médias les plus fréquemment retournés témoigne d'une certaine iniquité dans la diffusion des contenus par l'algorithme, une tendance également observée aux États-Unis. Qui plus est, parmi les 74 médias nationaux répertoriés, l'algorithme en favorise surtout sept. Ce sont ceux qui forment le top 20 présenté dans l'annexe 5. Ceux-ci comptent pour plus de 90 % des résultats totaux pour cette catégorie.

En revanche, Google Actualités semble mieux répartir les résultats provenant des 105 médias locaux et régionaux répertoriés, alors que les 11 médias locaux/régionaux du top 20 ne comptent que pour 59,72 % des 297 095 résultats pour cette catégorie. Cela peut s'expliquer par le fait que ces médias ont une importance strictement régionale et qu'ils rayonnent peu à l'extérieur de leur territoire.

La mention « près de » ainsi que le nom de la ville étudiée utilisée pour les saisies semblent affecter positivement l'affichage de contenu local et régional alors que pour la majorité des villes, le contenu affiché provient de médias à proximité. Il est difficile, considérant qu'une seule de nos requêtes excluait l'expression « près de », d'établir des constats, mais une tendance semble ressortir lorsque les deux recherches à Montréal sont comparées. On remarque effectivement une plus grande internationalisation

des résultats lorsque le « près de » n'est pas inclus dans la requêtes, alors que celle qui l'inclut semble tenter le plus possible de cibler des médias dans la région étudiée.

Finalement, les constats de cette recherche ne permettent pas de rejoindre totalement ceux de Fischer *et al.* (2020), mais ils offrent quand même un bon premier panorama du fonctionnement de Google Actualités au Québec. Ce premier audit algorithmique permet de distinguer les réalités propres au Québec, soit sa langue, son territoire et son grand nombre de médias locaux et régionaux. Il permet aussi de fournir un premier examen du fonctionnement de l'algorithme de Google Actualités et des décisions qu'il prend.

Cependant, cette recherche n'est que quantitative et non qualitative. Bien que nous ayons relevé une erreur de l'algorithme concernant la présence du site Necrologie Canada et qu'elle ait souligné l'âge considérable des résultats proposés par Google Actualités dans certains cas, aucune partie de ce mémoire ne s'est intéressé de manière approfondie à la qualité des articles mis de l'avant par l'algorithme. Un examen de cette qualité constituerait une excellente façon, pour un autre projet de recherche, de s'appuyer sur les conclusions de ce mémoire.

ANNEXE 1

Exemple de résultats de recherche à partir du mot-clé « accident » et de la localisation « près de Laval ».

The screenshot shows a Google News search interface. The search bar contains the text "accident près de laval". The results are displayed as a list of six items, each with a red rank number on the left, a title, a source and date, and a thumbnail image. A search filter sidebar is visible on the right.

Rang	Titre de l'article	Source et date	Localisation
1	Rage au volant et accident spectaculaire sur l'A-13	Courrier Laval · 26 mars	Laval
2	Loiron : trois blessés légers dans un accident de camping-car	France Bleu · 22 mars	Loiron
3	[8 avril 2021] Accident avec blessés à Terrebonne Accueil	Mon Joliette · il y a 7 jours	Terrebonne
4	Près de Vitré, deux blessés dans un accident de la route	Le Journal de Vitré · 30 mars	Vitré
5	Axe Saint-Nazaire - Laval : un camion couché sur la route près de Nozay, la circulation perturbée	actu.fr · 18 mars	Saint-Nazaire, Laval, Nozay
6	Laval. Quatre blessés dont un grave dans un accident près de l'Espace Mayenne	Ouest-France · 26 oct.	Laval, Mayenne

Les chiffres montrent le rang de chacun des résultats. Les points jaunes montrent les informations qui sont enregistrées à chaque saisie : le nom du média, le titre de l'article, le mot-clé, la localisation et la date de publication. L'URL menant à l'article est aussi enregistré.

ANNEXE 2

Exemple de document CSV contenant les résultats de requêtes effectuées le 18 décembre 2021.

A	B	C	D	E	F	G	H	I	J
Villes	Mots-clés	position du média	Nom du média	Titre de l'article	Date de publication	Journée de différence	URL	date et heure saisie	type mot-clés
Acton Vale	"premier ministre"	1	VINGT 55 Média web	«hefs des partis politique	2021-08-26	114	https://vingt55.ca/e	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	2	VINGT 55 Média web	«c ajoutent 94 M\$ pou	2021-07-21	150	https://vingt55.ca/g	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	3	La Tribune	«ouvernement Legault c	2021-10-27	52	https://www.latribun	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	4	La Voix de l'est	« Stefan Lofven démis	2021-11-10	38	https://www.lavoixd	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	5	VINGT 55 Média web	«ion ambitieuse et inn	2021-11-12	36	https://www.lavoixd	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	6	Tribune - Groupe Capitales Méd	«nent les devants 2-1 f	2021-09-07	102	https://vingt55.ca/f	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	7	VINGT 55 Média web	«ne commission parle	2021-08-17	123	https://www.lavoixd	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	8	Le Droit	«érait un «choc écon	2021-10-27	52	https://vingt55.ca/f	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	9	VINGT 55 Média web	«iderons du prochain g	2021-08-15	125	https://vingt55.ca/e	2021-12-18 17:26:1	national
Acton Vale	"premier ministre"	10	La Voix de l'est	« conseiller de Biden	2021-11-03	45	https://www.lavoixd	2021-12-18 17:26:1	national
Acton Vale	climat	1	La Voix de l'est	«ts à contribuer davan	2021-11-07	41	https://www.lavoixd	2021-12-18 17:26:4	national
Acton Vale	climat	2	Le Droit	«mber certaines prom	2021-11-17	31	https://www.ledroit	2021-12-18 17:26:4	national
Acton Vale	climat	3	La Voix de l'est	« rôle dans la lutte aux	2021-11-08	40	https://www.lavoixd	2021-12-18 17:26:4	national
Acton Vale	climat	4	Le Nouvelliste	«léaire pour combatr	2021-11-04	44	https://www.lenouv	2021-12-18 17:26:4	national
Acton Vale	immigration	1	VINGT 55 Média web	«usiciens issus de l'im	2021-10-14	65	https://vingt55.ca/f	2021-12-18 17:26:5	national
Acton Vale	immigration	2	VINGT 55 Média web	« cap sur ses cibles d'	2021-10-29	50	https://vingt55.ca/g	2021-12-18 17:26:5	national
Acton Vale	immigration	3	Le Droit	«mber certaines prom	2021-11-17	31	https://www.ledroit	2021-12-18 17:26:5	national
Acton Vale	immigration	4	VINGT 55 Média web	«hefs des partis politici	2021-08-26	114	https://vingt55.ca/e	2021-12-18 17:26:5	national
Acton Vale	immigration	5	VINGT 55 Média web	«e dose de vaccin dès	2021-05-31	201	https://vingt55.ca/d	2021-12-18 17:26:5	national
Acton Vale	immigration	6	VINGT 55 Média web	«iser la venue et l'em	2021-08-06	134	https://vingt55.ca/e	2021-12-18 17:26:5	national
Acton Vale	immigration	7	Le Droit	«re: les employeurs ap	2021-11-12	36	https://www.ledroit	2021-12-18 17:26:5	national
Acton Vale	immigration	8	uvelliste - Groupe Capitales Méd	«nent mieux intégrer de	2021-04-17	245	https://www.lenouv	2021-12-18 17:26:5	national
Acton Vale	immigration	9	VINGT 55 Média web	«éral dans Drummond	2021-09-16	93	https://vingt55.ca/m	2021-12-18 17:26:5	national
Acton Vale	immigration	10	VINGT 55 Média web	«citera pas de deuxiè	2021-06-28	173	https://vingt55.ca/d	2021-12-18 17:26:5	national
Acton Vale	immigration	11	otidien - Groupe Capitales Méd	« au recrutement de m	2021-09-01	108	https://www.leguot	2021-12-18 17:26:5	national
Acton Vale	immigration	12	VINGT 55 Média web	« cœur des préoccup	2021-08-30	110	https://vingt55.ca/e	2021-12-18 17:26:5	national
Acton Vale	immigration	13	VINGT 55 Média web	« mondiale des réfugié	2021-06-18	183	https://vingt55.ca/f	2021-12-18 17:26:5	national
Acton Vale	immigration	14	Tribune - Groupe Capitales Méd	«quitter pour se recon	2021-07-04	167	https://www.latribu	2021-12-18 17:26:5	national
Acton Vale	immigration	15	Le Droit	«cheuses craignent un	2021-10-23	56	https://www.ledroit	2021-12-18 17:26:5	national
Acton Vale	immigration	16	ix de l'est - Groupe Capitales Méd	«oeuvre: des renforts	2021-08-11	129	https://www.lavoixd	2021-12-18 17:26:5	national
Acton Vale	immigration	17	uvelliste - Groupe Capitales Méd	«rie de main-d'oeuvre	2021-09-04	105	https://www.lenouv	2021-12-18 17:26:5	national
Acton Vale	immigration	18	uvelliste - Groupe Capitales Méd	«anada: «On a été cap	2021-05-06	226	https://www.lenouv	2021-12-18 17:26:5	national
Acton Vale	immigration	19	Le Nouvelliste	« l'honneur au congr	2021-11-13	35	https://www.lenouv	2021-12-18 17:26:5	national
Acton Vale	racisme	1	Tribune - Groupe Capitales Méd	«mande que Québec r	2021-10-01	78	https://www.latribu	2021-12-18 17:27:0	national
Acton Vale	Chine	1	Le Nouvelliste	«ent: est-il temps de	2021-10-27	52	https://www.lenouv	2021-12-18 17:27:0	national
Acton Vale	Chine	2	ix de l'est - Groupe Capitales Méd	«débarquent au Club	2021-05-30	202	https://www.lavoixd	2021-12-18 17:27:0	national

ANNEXE 3

Les 26 mots-clés utilisés pour la présente recherche ainsi que les 32 utilisés par Fischer *et al.*,
par type de mot-clé

Type de mot-clé	Présente recherche	Fischer <i>et al.</i>
Local	<ol style="list-style-type: none"> 1. accident 2. Cégep 3. police OR Sûreté 4. décès OR mort 5. incendie 6. préfet 7. « Centre de services scolaires » OR « Commission scolaire » 8. CISSS OR CIUSSS 9. maire OR mairesse 10. candidat 11. « travaux routiers » 12. autobus OR tramway 13. météo 	<ol style="list-style-type: none"> 1. Accident 2. College 3. Crime 4. Death 5. Emergency Services 6. Governor 7. High School 8. Hospital 9. Mayor 10. Obituary 11. Police 12. School board 13. Traffic 14. Transit 15. University 16. Weather
Mixte	<ol style="list-style-type: none"> 1. Construction 2. élection OR campagne 3. député 4. « arme à feu » 5. « santé publique » 6. « covid OR COVID » 	Aucun
National	<ol style="list-style-type: none"> 1. immigration 2. climat 3. racisme 4. « premier ministre » 5. Chine 6. Biden OR Trump 7. déficit 	<ol style="list-style-type: none"> 1. Abortion 2. Caravan 3. Climat 4. Conservative 5. Corruption 6. Election 7. FBI 8. Gun 9. Immigration 10. Liberal 11. Politics 12. President 13. Scandal 14. Shutdown 15. Syria 16. Taxes

ANNEXE 4

Listes des 104 MRC ou territoires équivalents et leurs chefs-lieux ou villes les plus peuplées.

Ville	MRC ou territoire équivalent
Acton Vale	Acton
Alma	Lac-Saint-Jean-Est
Amos	Abitibi
Amqui	La Matapédia
Baie-Comeau	Manicouagan
Baie-Saint-Paul	Charlevoix
Beauceville	Robert-Cliche
Bécancour	Bécancour
Blainville	Thérèse-De Blainville
Blanc-Sablon	Le Golfe-du-Saint-Laurent
Boischatel	La Côte-de-Beaupré
Carleton-sur-Mer	Avignon
Chambly	La Vallée-du-Richelieu
Chandler	Le Rocher-Percé
Châteauguay	Roussillon
Chibougamau	Jamésie
Chisasibi	Gouvernement régional d'Eeyou Istchee Baie-James
Coaticook	Coaticook
Cookshire-Eaton	Le Haut-Saint-François
Cowansville	Brome-Missisquoi
Dolbeau-Mistassini	Maria-Chapdelaine
Drummondville	Drummond
Fermont	Caniapiscau
Forestville	La Haute-Côte-Nord
Gaspé	La Côte-de-Gaspé
Gatineau	Gatineau
Granby	La Haute-Yamaska
Havre-Saint-Pierre	Minganie
Joliette	Joliette
Kuuujuaq	Administration régionale Kativik
L'Islet	L'Islet
La Malbaie	Charlevoix-Est
La Pocatière	Kamouraska
La Sarre	Abitibi-Ouest
La Tuque	Agglomération de La Tuque
Lac-Etchemin	Les Etchemins
Lac-Mégantic	Le Granit
Lachute	Argenteuil
Laval	Laval
Lavaltrie	D'Autray
Les Îles-de-la-Madeleine	Communauté maritime des Îles-de-la-Madeleine
Lévis	Lévis
Longueuil	Agglomération de Longueuil
Louiseville	Maskinongé

Magog	Memphrémagog
Maniwaki	La Vallée-de-la-Gatineau
Mansfield-et-Pontefract	Pontiac
Mariville	Rouville
Matane	La Matanie
Mirabel	Mirabel
Mont-Joli	La Mitis
Mont-Laurier	Antoine-Labelle
Montmagny	Montmagny
Montréal	Agglomération de Montréal
New Richmond	Bonaventure
Nicolet	Nicolet-Yamaska
Notre-Dame-du-Mont-Carmel	Les Chenaux
Ormstown	Le Haut-Saint-Laurent
Plessisville	L'Érable
Québec	Agglomération de Québec
Rawdon	Matawinie
Repentigny	L'Assomption
Rimouski	Rimouski-Neigette
Rivière-du-Loup	Rivière-du-Loup
Rouyn-Noranda	Rouyn-Noranda
Saguenay	Saguenay
Saint-André-Avellin	Papineau
Saint-Apollinaire	Lotbinière
Saint-Eustache	Deux-Montagnes
Saint-Félicien	Le Domaine-du-Roy
Saint-Georges	Beauce-Sartigan
Saint-Henri	Bellechasse
Saint-Honoré	Le Fjord-du-Saguenay
Saint-Hyacinthe	Les Maskoutains
Saint-Jean-sur-Richelieu	Le Haut-Richelieu
Saint-Jérôme	La Rivière-du-Nord
Saint-Lin--Laurentides	Montcalm
Saint-Pierre-de-l'Île-d'Orléans	L'Île-d'Orléans
Saint-Raymond	Portneuf
Saint-Rémi	Les Jardins-de-Napierville
Saint-Tite	Mékinac
Sainte-Adèle	Les Pays-d'en-Haut
Sainte-Agathe-des-Monts	Les Laurentides
Sainte-Anne-des-Monts	La Haute-Gaspésie
Sainte-Julie	Marguerite-D'Youville
Sainte-Marie	La Nouvelle-Beauce
Salaberry-de-Valleyfield	Beauharnois-Salaberry
Sept-Îles	Sept-Rivières
Shawinigan	Shawinigan
Sherbrooke	Sherbrooke
Sorel-Tracy	Pierre-De Saurel
Stoneham-et-Tewkesbury	La Jacques-Cartier
Témiscouata-sur-le-Lac	Témiscouata
Terrebonne	Les Moulins

Thetford Mines	Les Appalaches
Trois-Pistoles	Les Basques
Trois-Rivières	Trois-Rivières
Val-d'Or	La Vallée-de-l'Or
Val-des-Monts	Les Collines-de-l'Outaouais
Val-des-Sources	Les Sources
Vaudreuil-Dorion	Vaudreuil-Soulanges
Victoriaville	Arthabaska
Ville-Marie	Témiscamingue
Windsor	Le Val-Saint-François

ANNEXE 5

Liste des 20 médias, filtrés par catégorie, les plus souvent affichés par Google Actualités dans la base de données complète.

Domaines affichés par Google Actualités	Catégorie	Nombre d'occurrences	Poids dans la base de données complète
ici.radio-canada.ca	local/régional	65 191	13,36 %
lapresse.ca	national	30 034	6,15 %
journaldemontreal.com	national	29 077	5,96 %
journaldequebec.com	national	25 843	5,30 %
lesoleil.com	local/régional	20 709	4,24 %
lenouvelliste.ca	local/régional	17 967	3,68 %
latribune.ca	local/régional	15 809	3,24 %
ici.radio-canada.ca	national	14 744	3,02 %
tvanouvelles.ca	national	12 701	2,60 %
necrocanada.com	non-média	12 001	2,46 %
ledroit.com	local/régional	11 610	2,38 %
lequotidien.com	local/régional	10 018	2,05 %
ledevoir.com	national	9 809	2,01 %
lavoixdelest.ca	local/régional	9 150	1,87 %
lanouvelle.net	local/régional	8 477	1,74 %
cimtchau.ca	local/régional	7 762	1,59 %
www.iheartradio.ca	non-média	6 878	1,41 %
infodimanche.com	local/régional	6 628	1,36 %
meteomedia.com	national	4 628	0,95 %
tvanouvelles.ca	local/régional	4 050	0,83 %

Radio-Canada et TVA Nouvelles figurent à deux reprises, car ceux-ci diffusent des contenus dans les catégories local/régional et national.

ANNEXE 6

Liste des 20 médias les plus fréquemment affichés par Google Actualités dans la base de données complète, selon le type de mots-clés.

Type de mots-clés	Domaines affichés par Google Actualités	Catégorie de média	Nombre d'occurrences	Poids dans la base de données
local	ici.radio-canada.ca	local/régional	33 724	13,65 %
	journaldequebec.com	national	14 059	5,69 %
	journaldemontreal.com	national	12 770	5,17 %
	lapresse.ca	national	11 918	4,82 %
	necrocanada.com	non-média	10 376	4,20 %
	lenouvelliste.ca	local/régional	7 866	3,18 %
	lesoleil.com	local/régional	7 768	3,14 %
	latribune.ca	local/régional	7 672	3,10 %
	tvanouvelles.ca	national	5 909	2,39 %
	lequotidien.com	local/régional	5 361	2,17 %
	cimtchau.ca	local/régional	5 170	2,09 %
	lanouvelle.net	local/régional	4 945	2,00 %
	ledroit.com	local/régional	4 707	1,90 %
	lavoixdelest.ca	local/régional	4 534	1,83 %
	www.iheartradio.ca	non-média	4 291	1,74 %
	meteomedia.com	national	4 168	1,69 %
	infodimanche.com	local/régional	3 759	1,52 %
	ici.radio-canada.ca	national	3 647	1,48 %
	zone911.com	national	2 951	1,19 %
	ledevoir.com	national	2 643	1,07 %
	Total		158 238 / 247 147	64,03 %
mixte	ici.radio-canada.ca	local/régional	17 744	15,03 %
	lapresse.ca	national	7 121	6,03 %
	journaldequebec.com	national	6 613	5,60 %
	journaldemontreal.com	national	6 591	5,58 %
	lesoleil.com	local/régional	4 612	3,91 %
	lenouvelliste.ca	local/régional	4 065	3,44 %
	latribune.ca	local/régional	3 333	2,82 %
	ledroit.com	local/régional	3 124	2,65 %
	ici.radio-canada.ca	national	3 041	2,58 %
	tvanouvelles.ca	national	2 993	2,54 %
	lanouvelle.net	local/régional	2 206	1,87 %
	lequotidien.com	local/régional	2 091	1,77 %
	ledevoir.com	national	1 998	1,69 %
	lavoixdelest.ca	local/régional	1 928	1,63 %
	www.iheartradio.ca	non-média	1 897	1,61 %
	infodimanche.com	local/régional	1 723	1,46 %
	cimtchau.ca	local/régional	1 719	1,46 %
	necrocanada.com	non-média	1 616	1,37 %
	enbeauce.com	local/régional	1 115	0,94 %

	journaldelevis.com	local/régional	1 086	0,92 %
		Total	76 616 / 118 053	64,90 %
national	ici.radio-canada.ca	local/régional	13 723	11,17 %
	lapresse.ca	national	10 995	8,95 %
	journaldemontreal.com	national	9 716	7,91 %
	lesoleil.com	local/régional	8 329	6,78 %
	ici.radio-canada.ca	national	8 056	6,56 %
	lenouvelliste.ca	local/régional	6 036	4,91 %
	journaldequebec.com	national	5 171	4,21 %
	ledevoir.com	national	5 168	4,21 %
	latribune.ca	local/régional	4 804	3,91 %
	tvanouvelles.ca	national	3 799	3,09 %
	ledroit.com	local/régional	3 779	3,08 %
	lavoixdelest.ca	local/régional	2 688	2,19 %
	lequotidien.com	local/régional	2 566	2,09 %
	lanouvelle.net	local/régional	1 326	1,08 %
	infodimanche.com	local/régional	1 146	0,93 %
	ici.radio-canada.ca	hors Québec	1 136	0,93 %
	droit-inc.com	national	1 122	0,91 %
	cimtchau.ca	local/régional	873	0,71 %
	tvanouvelles.ca	local/régional	789	0,64 %
	journalmetro.com	local/régional	700	0,57 %
		Total	91 922 / 122 809	74,85 %

En jaune se trouvent les domaines qui sont soit des non-médias ou des médias hors Québec. À la lecture de ce tableau, on remarque que le poids de Radio-Canada, local/régional, est moins important dans les recherches contenant des mots-clés « national ». De plus, le poids des médias formant le top 20 des mots-clés « national » est plutôt costaud, dépassant de plus de 10 % le poids des mots-clés « local » et « mixte ».

ANNEXE 7

Liste des 62 villes où se trouvent les sièges sociaux des médias locaux et régionaux répertoriés par Google Actualités.

Villes des sièges sociaux pour les médias locaux et régionaux répertoriés par Google Actualités	Fréquence, dans les résultats, des médias basés dans cette ville
Québec	30 469
Trois-Rivières	26 295
Sherbrooke	21 085
Saguenay	16 477
Matane	15 326
Rimouski	13 226
Gatineau	11 727
Rouyn-Noranda	11 658
Granby	11 027
Saint-Georges	9 361
Sept-Îles	9 233
Victoriaville	8 710
Rivière-du-Loup; Carleton-sur-Mer	7 762
Rivière-du-Loup	6 628
Joliette	5 977
Montréal	5 587
Salaberry-de-Valleyfield	5 176
Lévis	4 875
Chambly	4 699
Saint-Sauveur	4 558
Ottawa-Gatineau	4 180
Drummondville	3 991
Baie-Comeau	3 724
Sainte-Thérèse	3 474
Lac-Etchemin	3 424
Sorel-Tracy	3 212
Saint-Eustache	3 178
Thetford Mines	2 816
Boucherville	2 594
Châteauguay	2 305
La Tuque	2 269
Coaticook	2 195
Saint-Lin-Laurentides	2 065
Cowansville	2 004
Magog	1 934
Saint-Jacques-le-Mineur	1 677
Longueuil	1 597
Shawinigan	1 498
Laval	1 474
Saint-Pascal	1 415
Saint-Jérôme	1 375

Repentigny	1 341
Saint-Jean-sur-Richelieu	1 245
Sainte-Marie	1 169
Alma	1 095
Dolbeau-Mistassini	1 038
Roberval	1 008
Amqui	1 001
Chibougamau	984
Val-d'Or	984
Sainte-Julienne	953
Saint-Charles-Borromée	904
Gaspé	747
Candiac	559
Trois-Pistoles	538
Amos	271
Beloeil	259
Saint-Hyacinthe	50
New Richmond	39
Saint-Bruno-de-Montarville	20
Chandler	4
Chénéville	2
TOTAL	296 468

À noter : bien qu'il ait 297 017 résultats dans la catégorie « local/régional », il n'y a que 296 468 résultats (soit une différence de 549) où la ville du siège social du média a pu être retrouvée.

ANNEXE 8

Liste des 122 médias locaux et régionaux les plus retournés par Google Actualités, avec leurs sièges sociaux.

Domaines affichés par Google Actualités	Ville du siège social	Nombre d'occurrences
lesoleil.com	Québec	20 709
lenouvelliste.ca	Trois-Rivières	17 967
latribune.ca	Sherbrooke	15 809
ledroit.com	Gatineau	11 610
ici.radio-canada.ca	Rouyn-Noranda	10 987
ici.radio-canada.ca	Matane	10 677
lequotidien.com	Saguenay	10 018
lavoixdelest.ca	Granby	9 150
ici.radio-canada.ca	Rimouski	9 105
lanouvelle.net	Victoriaville	8 477
ici.radio-canada.ca	Québec	8 306
cimtchau.ca	Rivière-du-Loup; Carleton-sur-Mer	7 762
infodimanche.com	Rivière-du-Loup	6 628
ici.radio-canada.ca	Saguenay	5 803
ici.radio-canada.ca	Sept-Îles	5 394
ici.radio-canada.ca	Trois-Rivières	4 329
ici.radio-canada.ca	Ottawa-Gatineau	4 180
ici.radio-canada.ca	Sherbrooke	4 034
lemanic.ca	Baie-Comeau	3 724
journalaccs.ca	Saint-Sauveur	3 679
enbeauce.com	Saint-Georges	3 669
nordinfo.com	Sainte-Thérèse	3 474
lavoixdusud.com	Lac-Etchemin	3 424
macotenord.com	Sept-Îles	3 347
leveil.com	Saint-Eustache	3 178
journaldelevis.com	Lévis	3 100
journaldechambly.com	Chambly	3 046
monmatane.com	Matane	3 021
leclairleurprogres.ca	Saint-Georges	2 982
laction.com	Joliette	2 805
infosuroit.com	Salaberry-de-Valleyfield	2 632
lareleve.qc.ca	Boucherville	2 594
courrierfrontenac.qc.ca	Thetford Mines	2 565
journalssaint-francois.ca	Salaberry-de-Valleyfield	2 544
ici.radio-canada.ca	Montréal	2 376
journalexpress.ca	Drummondville	2 354
cybersoleil.com	Châteauguay	2 305
leprogres.net	Coaticook	2 195
monjoliette.com	Joliette	2 165
l'expressmontcalm.com	Saint-Lin-Laurentides	2 065

journalmetro.com	Montréal	1 997
lerefletdulac.com	Magog	1 934
les2rives.com	Sorel-Tracy	1 913
lavantage.qc.ca	Rimouski	1 895
lepeuplelotbiniere.ca	Lévis	1 775
lechodelatuque.com	La Tuque	1 687
journallesoir.ca	Rimouski	1 683
coupdoeil.info	Saint-Jacques-le-Mineur	1 677
chamblymatin.com	Chambly	1 653
vingt55.ca	Drummondville	1 637
lavantagegaspesien.com	Matane	1 628
lecourriersud.com	Trois-Rivières	1 623
granbyexpress.com	Granby	1 603
journalleguide.com	Cowansville	1 516
lhebdojustmaurice.com	Shawinigan	1 498
lejournaldejoliette.ca	Saint-Georges	1 457
leplacoteux.com	Saint-Pascal	1 415
topolocal.ca	Saint-Jérôme	1 375
hebdoorivenord.com	Repentigny	1 341
courrierlaval.com	Laval	1 315
soreltracy.com	Sorel-Tracy	1 299
canadafrancais.com	Saint-Jean-sur-Richelieu	1 245
lechodemaskinonge.com	Trois-Rivières	1 198
beaucemedia.ca	Sainte-Marie	1 169
tvanouvelles.ca	Montréal	1 111
lecourrierdusud.ca	Longueuil	1 105
lelacstjean.com	Alma	1 095
nouvelleshebdo.com	Dolbeau-Mistassini	1 038
letoiledulac.com	Roberval	1 008
lactiondautray.com	Joliette	1 007
lavantposte.ca	Amqui	1 001
lasentinelle.ca	Chibougamau	984
lecitoyenvaldoramoss.com	Val-d'Or	984
plumelibre.ca	Sainte-Julienne	953
lanauweb.info	Saint-Charles-Borromée	904
journalnord.com	Saint-Sauveur	879
estriplus.com	Sherbrooke	813
coolfm.biz	Saint-Georges	797
quebechebdo.com	Québec	714
lhebdojournal.com	Trois-Rivières	712
lecitoyenrouynlasarre.com	Rouyn-Noranda	623
monlatuque.com	La Tuque	582
lereflet.qc.ca	Candiac	559
tvanouvelles.ca	Rimouski	543
journalhorizon.com	Trois-Pistoles	538
fm1033.ca	Longueuil	492
tvanouvelles.ca	Sept-Îles	492
laveniretdesrivieres.com	Cowansville	488
radiogaspesie.ca	Gaspé	486
tvanouvelles.ca	Québec	468

tvanouvelles.ca	Trois-Rivières	466
mabeauce.com	Saint-Georges	456
tvanouvelles.ca	Sherbrooke	429
tvanouvelles.ca	Saguenay	376
informeaffaires.com	Saguenay	280
m105.ca	Granby	274
mediat.ca	Amos	271
gaspesienouvelles.com	Gaspé	261
oeilregional.com	Beloil	259
monthetford.com	Thetford Mines	251
monvicto.com	Victoriaville	233
lechodelaval.ca	Laval	159
tvanouvelles.ca	Gatineau	117
monlimoilou.com	Québec	95
lecourrier.qc.ca	Saint-Hyacinthe	50
tvanouvelles.ca	Rouyn-Noranda	48
journaldesvoisins.com	Montréal	46
monquartier.quebec	Québec	44
chaleurnouvelles.com	New Richmond	39
zone911.com	Québec	35
monsaintsauveur.com	Québec	33
monmontcalm.com	Québec	27
delitfrancais.com	Montréal	26
impactcampus.ca	Québec	26
linitiative.ca	Montréal	26
lemontarvillois.com	Saint-Bruno-de-Montarville	20
monsaintroch.com	Québec	12
journallehavre.ca	Chandler	4
montrealcampus.ca	Montréal	3
lapetitenation.com	Chénéville	2
montreal.citycrunch.ca	Montréal	2

ANNEXE 9

Poids relatif des différentes catégories de médias et catégorie la plus souvent retournée par Google Actualités, par ville

Villes étudiées	Poids des médias « national »	Poids des médias « local/régional »	Poids des médias « hors Québec »	Poids des médias « non-média »	Catégorie la plus importante
Acton Vale	15,62 %	78,97 %	0,32 %	5,09 %	local/régional
Alma	14,61 %	73,12 %	8,74 %	3,54 %	local/régional
Amos	20,67 %	64,48 %	6,35 %	8,49 %	local/régional
Amqui	14,23 %	69,04 %	0,14 %	16,59 %	local/régional
Baie-Comeau	22,01 %	73,52 %	0,74 %	3,74 %	local/régional
Baie-Saint-Paul	30,04 %	61,30 %	4,37 %	4,29 %	local/régional
Beauceville	18,26 %	79,88 %	0,25 %	1,61 %	local/régional
Blainville	18,64 %	77,72 %	0,16 %	3,48 %	local/régional
Blanc-Sablon	35,94 %	49,98 %	9,74 %	4,34 %	local/régional
Boischatel	41,03 %	50,48 %	7,69 %	0,81 %	local/régional
Bécancour	51,79 %	45,27 %	0,06 %	2,89 %	national
Carleton-sur-Mer	22,70 %	71,60 %	1,99 %	3,71 %	local/régional
Chambly	9,63 %	68,53 %	17,22 %	4,62 %	local/régional
Chandler	19,59 %	70,09 %	7,68 %	2,64 %	local/régional
Chibougamau	28,36 %	67,11 %	0,06 %	4,46 %	local/régional
Chisasibi	37,54 %	59,00 %	0,37 %	3,09 %	local/régional
Châteauguay	55,78 %	40,27 %	0,89 %	3,06 %	national
Coaticook	6,20 %	92,42 %	0,06 %	1,31 %	local/régional
Cookshire-Eaton	14,55 %	83,57 %	0,12 %	1,77 %	local/régional
Cowansville	7,52 %	88,15 %	0,06 %	4,27 %	local/régional
Dolbeau-Mistassini	9,14 %	83,76 %	0,26 %	6,84 %	local/régional
Drummondville	16,67 %	67,85 %	0,14 %	15,34 %	local/régional
Fermont	39,84 %	55,52 %	1,89 %	2,75 %	local/régional
Forestville	23,32 %	69,72 %	0,94 %	6,02 %	local/régional
Gaspé	33,84 %	62,75 %	0,96 %	2,45 %	local/régional
Gatineau	18,20 %	73,77 %	1,07 %	6,95 %	local/régional
Granby	15,98 %	80,88 %	0,08 %	3,07 %	local/régional
Havre-Saint-Pierre	26,77 %	45,82 %	22,71 %	4,71 %	local/régional
Joliette	31,60 %	64,12 %	1,01 %	3,27 %	local/régional
Kuuujuaq	77,25 %	18,83 %	2,16 %	1,75 %	national
L'Islet	20,67 %	73,17 %	2,64 %	3,53 %	local/régional
La Malbaie	45,57 %	45,68 %	2,71 %	6,04 %	local/régional et national
La Pocatière	22,96 %	74,78 %	0,67 %	1,60 %	local/régional
La Sarre	11,76 %	52,44 %	26,04 %	9,76 %	local/régional
La Tuque	11,46 %	86,37 %	0,89 %	1,28 %	local/régional
Lac-Etchemin	10,73 %	86,06 %	0,33 %	2,88 %	local/régional
Lac-Mégantic	30,15 %	62,84 %	0,53 %	6,48 %	local/régional

Lachute	30,91 %	17,97 %	48,56 %	2,56 %	hors Québec
Laval	49,39 %	34,54 %	11,51 %	4,57 %	national
Lavaltrie	41,99 %	55,25 %	0,05 %	2,71 %	local/régional
Les Îles-de-la-Madeleine	38,10 %	55,08 %	3,90 %	2,92 %	local/régional
Longueuil	43,87 %	51,44 %	1,33 %	3,36 %	local/régional
Louiseville	53,67 %	38,90 %	0,29 %	7,14 %	national
Lévis	14,45 %	80,54 %	0,81 %	4,19 %	local/régional
Magog	12,43 %	83,36 %	0,26 %	3,95 %	local/régional
Maniwaki	22,49 %	68,59 %	1,08 %	7,85 %	local/régional
Mansfield	32,02 %	26,64 %	31,12 %	10,23 %	national et hors Québec
Marieville	17,58 %	74,89 %	0,66 %	6,87 %	local/régional
Matane	16,24 %	75,48 %	0,47 %	7,81 %	local/régional
Mirabel	40,77 %	50,53 %	4,75 %	3,94 %	local/régional
Mont-Joli	18,05 %	71,58 %	4,75 %	5,63 %	local/régional
Mont-Laurier	51,08 %	39,79 %	0,66 %	8,47 %	national
Montmagny	27,27 %	64,75 %	3,33 %	4,66 %	local/régional
Montréal	84,61 %	12,49 %	0,50 %	2,40 %	national
Montréal_V2	47,71 %	29,73 %	17,27 %	5,29 %	national
New Richmond	19,27 %	71,53 %	5,15 %	4,05 %	local/régional
Nicolet	13,67 %	80,12 %	1,81 %	4,40 %	local/régional
Notre-Dame-du-Mont-Carmel	6,58 %	85,26 %	4,98 %	3,18 %	local/régional
Ormstown	40,42 %	56,07 %	0,18 %	3,32 %	local/régional
Plessisville	9,62 %	86,13 %	0,13 %	4,12 %	local/régional
Québec	66,80 %	29,66 %	0,81 %	2,74 %	national
Rawdon	45,01 %	50,76 %	1,18 %	3,05 %	local/régional
Repentigny	53,27 %	42,07 %	0,80 %	3,86 %	national
Rimouski	21,68 %	68,18 %	0,35 %	9,79 %	local/régional
Rivière-du-Loup	17,39 %	79,50 %	1,06 %	2,05 %	local/régional
Rouyn-Noranda	21,69 %	62,78 %	0,86 %	14,66 %	local/régional
Saguenay	38,35 %	56,98 %	0,31 %	4,36 %	local/régional
Saint-André-Avellin	10,18 %	80,76 %	0,84 %	8,22 %	local/régional
Saint-Apollinaire	19,41 %	50,04 %	26,31 %	4,23 %	local/régional
Saint-Eustache	33,89 %	55,41 %	4,76 %	5,93 %	local/régional
Saint-Félicien	13,55 %	79,36 %	2,13 %	4,96 %	local/régional
Saint-Georges	18,10 %	58,82 %	19,26 %	3,82 %	local/régional
Saint-Henri	37,53 %	42,94 %	13,68 %	5,85 %	local/régional
Saint-Honoré	14,25 %	61,47 %	17,00 %	7,29 %	local/régional
Saint-Hyacinthe	51,31 %	25,75 %	2,00 %	20,94 %	national
Saint-Jean-sur-Richelieu	29,08 %	53,06 %	0,41 %	17,45 %	local/régional
Saint-Jérôme	40,95 %	45,49 %	6,83 %	6,73 %	local/régional
Saint-Lin-Laurentides	28,40 %	67,78 %	0,67 %	3,15 %	local/régional
Saint-Pierre-de-l'Île-d'Orléans	62,43 %	33,24 %	1,47 %	2,85 %	national
Saint-Raymond	33,48 %	50,89 %	8,12 %	7,51 %	local/régional
Saint-Rémi	20,10 %	43,24 %	30,29 %	6,38 %	local/régional

Saint-Tite	21,07 %	70,62 %	3,37 %	4,95 %	local/régional
Sainte-Adèle	43,33 %	51,83 %	3,05 %	1,80 %	local/régional
Sainte-Agathe-des-Monts	55,45 %	35,11 %	6,51 %	2,93 %	national
Sainte-Anne-des-Monts	20,31 %	75,76 %	2,21 %	1,72 %	local/régional
Sainte-Julie	30,38 %	57,56 %	4,82 %	7,25 %	local/régional
Sainte-Marie	25,72 %	47,35 %	20,87 %	6,07 %	local/régional
Salaberry-de-Valleyfield	38,38 %	57,03 %	0,11 %	4,48 %	local/régional
Sept-Îles	33,75 %	55,58 %	6,31 %	4,36 %	local/régional
Shawinigan	16,23 %	75,72 %	0,32 %	7,73 %	local/régional
Sherbrooke	29,36 %	67,90 %	0,22 %	2,52 %	local/régional
Sorel-Tracy	26,72 %	67,62 %	0,27 %	5,40 %	local/régional
Stoneham	40,24 %	56,17 %	1,24 %	2,35 %	local/régional
Terrebonne	16,34 %	80,70 %	0,47 %	2,49 %	local/régional
Thetford Mines	56,34 %	38,59 %	0,15 %	4,93 %	national
Trois-Pistoles	21,41 %	73,87 %	0,27 %	4,46 %	local/régional
Trois-Rivières	19,40 %	76,36 %	0,84 %	3,41 %	local/régional
Témiscouata-sur-le-Lac	26,71 %	65,20 %	3,65 %	4,44 %	local/régional
Val-d'Or	28,13 %	55,04 %	0,83 %	15,99 %	local/régional
Val-des-Monts	18,94 %	67,56 %	8,96 %	4,53 %	local/régional
Val-des-Sources	16,71 %	65,37 %	13,08 %	4,84 %	local/régional
Vaudreuil-Dorion	50,94 %	45,39 %	0,08 %	3,58 %	national
Victoriaville	15,03 %	78,25 %	0,53 %	6,20 %	local/régional
Ville-Marie	43,58 %	38,77 %	8,98 %	8,67 %	national
Windsor	22,03 %	41,84 %	29,79 %	6,35 %	local/régional

RÉFÉRENCES

Académie de la transformation numérique. (2020). *NETendances 2020 - Portrait numérique des foyers québécois* (3). Université Laval. <https://transformation-numerique.ulaval.ca/enquetes-et-mesures/netendances/netendances-2020-portrait-numerique-des-foyers-quebecois/>

Académie de la transformation numérique. (2021). *NETendances 2021 - Portrait numérique des foyers québécois* (4). <https://transformation-numerique.ulaval.ca/wp-content/uploads/2020/09/netendances-2018-usage-medias-sociaux.pdf>

Agence QMI. (2020, 26 août). Crise des médias: 200 stations de radio et télé pourraient bientôt disparaître. *Le Journal de Montréal*. <https://www.journaldemontreal.com/2020/08/26/crise-des-medias-200-stations-de-radio-et-tele-pourraient-bientot-disparaitre>

Angwin, J., Larson, J., Mattu, S. et Kirchner, L. (2016). Machine Bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=gGXmoojk8wQhWTulwj4V5m6T87yQfqIJ>

Bandy, J. et Diakopoulos, N. (2020). Auditing News Curation Systems: A Case Study Examining Algorithmic and Editorial Logic in Apple News. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, 36-47.

Bizimana, A.-J. et Kane, O. (2019). La presse au défi du numérique : une économie politique des médias régionaux au Québec. *Les Cahiers du journalisme - Recherches*, 2(3). <http://cahiersdujournalisme.org/V2N3/CaJ-2.3-R141.html>

Bro, P. (2019). Gatekeeping Theory. Dans *The International Encyclopedia of Journalism Studies* (p. 1-6). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118841570.iejs0022>

Broussard, M. (2019). *Artificial unintelligence: how computers misunderstand the world* (The MIT Press).

Cardon, D. (2015). *A quoi rêvent les algorithmes: nos vies à l'heure des big data*. La République des idées : Seuil.

Carignan, M.-È. (2011). Perception de la situation médiatique locale et régionale au Québec : l'avis du public. Dans *Les mutations de l'information et des médias locaux et régionaux: économie, contenus, usages et pratiques professionnelles* (p. 119-134). Éditions du Lerass.

Carroll, N. (2018). Search Engine Optimization. Dans *Encyclopedia of Library and Information Sciences* (4^e éd.). CRC Press.

Castets-Renard, C. (2018, 20 septembre). *Quels impacts de l'intelligence artificielle sur les métiers du droit et du journalisme ? (What Are the Impacts of Artificial Intelligence on the Legal and Journalistic Professions?)* (ID 3391272) [SSRN Scholarly Paper]. Social Science Research Network. <https://papers.ssrn.com/abstract=3391272>

Castets-Renard, C. (2019). Intelligence artificielle : combattre les biais des algorithmes. *The Conversation*. <http://theconversation.com/intelligence-artificielle-combattre-les-biais-des-algorithmes-125004>

CEFRIO. (2018). *L'usage des médias sociaux au Québec — NETendances* (5). <https://transformation-numerique.ulaval.ca/wp-content/uploads/2020/09/netendances-2018-usage-medias-sociaux.pdf>

CEM. (2020). *Publicité*. Centre d'études sur les médias. <https://www.cem.ulaval.ca/economie/donnees-financieres/publicite/>

Charlton, S., Giroux, D. et Lemieux, M. (2013). *Comment les Québécois s'informent-ils?* Centre d'études sur les médias.

Chong, P. (2019). Valuing subjectivity in journalism: Bias, emotions, and self-interest as tools in arts reporting. *Journalism*, 20(3), 427-443. <https://doi.org/10.1177/1464884917722453>

Collard, N. (2012, 8 avril). Quebecor c. Transcontinental: la guerre des hebdomadaires. *La Presse, Arts*. <https://www.lapresse.ca/arts/medias/201204/08/01-4513468-quebecor-c-transcontinental-la-guerre-des-hebdos.php>

Conseil de la radiodiffusion et des télécommunications canadiennes. (2016). *Politique réglementaire de radiodiffusion CRTC 2016-224*, 65.

Dean, B. (2021, 10 octobre). *Google's 200 Ranking Factors: The Complete List (2022)*. Backlinko. <https://backlinko.com/google-ranking-factors>

Delporte, C. (2005). Jules Verne et le journaliste. Imaginer l'information du xxe siècle. *Le Temps des médias*, n° 4(1), 201-213.

Diakopoulos, N. (2013, 2 août). Sex, Violence, and Autocomplete Algorithms. *Slate*. <https://slate.com/technology/2013/08/words-banned-from-bing-and-googles-autocomplete-algorithms.html>

Diakopoulos, N. (2014a). Algorithmic Accountability. *Digital Journalism*, 3(3), 398-415. <https://doi.org/10.1080/21670811.2014.976411>

Diakopoulos, N. (2014b). Algorithmic Accountability Reporting: On the Investigation of Black Boxes. *Tow Center for Digital Journalism, Columbia University*. <https://doi.org/10.7916/D8ZK5TW2>

Dubois, J. (2021). *Journalisme, médias sociaux et intérêt public: enquête auprès de 393 journalistes québécois*. <http://www.deslibris.ca/ID/460402>

Fédération professionnelle des journalistes du Québec. Guide de déontologie des journalistes du Québec. 24 novembre 1996.

Fischer, S., Jaidka, K. et Lelkes, Y. (2020). Auditing local news presence on Google News. *Nature Human Behaviour*, 4(12), 1236-1244. <https://doi.org/10.1038/s41562-020-00954-0>

Giroux, D. (2022). *Les médias québécois d'information : États des lieux* (Centre d'études sur les médias).

Google News Initiative. (2021, 3 septembre). *Community News Summit 2021*. <https://www.youtube.com/watch?v=EhdWsJOuBtY>

Gouritin, T. (2018, février). *Rencontre avec Antoinette Rouvroy: la data et les algorithmes; gouvernementalité algorithmique et idéologie des big data*. <https://april.org/data-et-algorithmes-gouvernementalite-algorithmique-et-ideologie-des-big-data-antoinette-rouvroy>

Hagey, K. et Horwitz, J. (2021, 15 septembre). Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. *Wall Street Journal*, Tech. <https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215>

Hoffman, J. (2020, 30 novembre). Canada's local news deserts : Local newsrooms are fast disappearing. Do they have a future? *The Signal*. <https://signalhfx.ca/canadas-local-news-deserts/>

Infopresse. (2018, 23 février). *Les médias et leur incidence sur la démocratie*. Infopresse. <https://web.archive.org/web/20180223135907/https://www.infopresse.com/article/2018/2/23/les-medias-et-leur-incidence-sur-la-democratie>

Jenkins, J. et Nielsen, R. K. (2018). The Digital Transition of Local News. *Reuters institute for the study of Journalism*, 50.

Jennings, J. et Rubado, M. (2019). *Newspaper Decline and the Effect on Local Government Coverage*, 14.

Lindgren, A. et Corbett, J. (2021, 1^{er} décembre). *Local News Map Data*. <https://s35582.pcdn.co/wp-content/uploads/2021/08/LocalNewsMapDataAugust2021.pdf>

Magnusson, E. (2021, 25 février). *Local news in Google News: a response to Fischer et al. (2020)* [MPRA Paper]. <https://mpra.ub.uni-muenchen.de/106431/>

Martel, M.-È. (2018). *Extinction de voix: plaidoyer pour la sauvegarde de l'information régionale*. Éditions Somme toute.

Mosco, V. (2017). *Becoming digital: towards a post-Internet society* (First edition). Emerald Publishing Limited.

Moz. (s. d.). *Google Algorithm Update History*. Moz. Récupéré le 11 janvier 2022 de <https://moz.com/google-algorithm-change>

Nechushtai, E. et Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in*

Human Behavior, 90, 298-307. <https://doi.org/10.1016/j.chb.2018.07.043>

Nécrologie Canada. (s. d.). *Avis de décès*. Nécrologie Canada. Récupéré le 18 décembre 2022 de <https://necrocanada.com/>

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K. et Nielsen, R. K. (2022). *Reuters Institute Digital News Report 2022*, 164.

Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T. et Nielsen, R. K. (2021). Reuters Institute Digital News Report — 10TH EDITION. *Digital News Report*, 164.

Newman, N., Richard Fletcher, Anne Schulz, Simge Andi et Rasmus Kleis Nielsen. (2020). *Reuters Institute Digital News Report*. Reuter Institute and University of Oxford. <https://www.digitalnewsreport.org/>

Nisbett, R. et Ross, L. (1983). Human Inference: Strategies and Shortcomings of Social Judgment. *The Philosophical Review*, 92(3), 462. <https://doi.org/10.2307/2184495>

Nocetti, J. (2017). *Comment l'information recompose les relations internationales. La guerre de l'information aura-t-elle lieu ?* (p. 138-143). pp. <http://www.cairn.info/la-guerre-de-l-information-aura-t-elle-lieu--9782100759729-page-138.htm>

Ouellet, M., Ménard, M., Bonenfant, M. et Mondoux, A. (2015). Big Data et quantification de soi : La gouvernementalité algorithmique dans le monde numériquement administré. *Canadian Journal of Communication*, 40(4), 597-613. <https://doi.org/10.22230/cjc.2015v40n4a2869>

Proulx, S. (2020, décembre). *Comment les médias d'information peuvent-ils faire le poids face aux GAFAs ?* Virtual Artifacts. https://www.virtualartifacts.com/publications.html?lang=fr&utm_source=lesecrans&utm_medium=inlin e&utm_campaign=hibe-rapport-gafa

Rashidian, N., Brown, P. D., Hansen, E., Bell, E. J. et Albright, J. R. (2019). *Friend and Foe: The Platform Press at the Heart of Journalism*. <https://doi.org/10.7916/d8-15pq-x415>

Rashidian, N., Tsiveriotis, G., Brown, P. D., Bell, E. J. et Hartstone, A. (2020). *Platforms and Publishers: The End of an Era*. <https://doi.org/10.7916/d8-sc1s-2j58>

Rebillard, F. et Smyrnaiois, N. (2019). Quelle « plateformisation » de l'information ? Collusion socioéconomique et dilution éditoriale entre les entreprises médiatiques et les infomédiaires de l'Internet. *tic&société*, (Vol. 13, N° 1-2), 247-293. <https://doi.org/10.4000/ticetsociete.4080>

Rouvroy, A. et Berns, T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation: Le disparate comme condition d'individuation par la relation ? *Réseaux*, n° 177(1), 163-196. <https://doi.org/10.3917/res.177.0163>

Roy, J.-H. (2015). *Le grand dérangement numérique : plaidoyer pour un journalisme hacker*, 125-142.

Roy, J.-H. (2020, 6 novembre). Google et Facebook seuls ne sauveront pas les médias de leur marasme financier. *The Conversation*. <http://theconversation.com/google-et-facebook-seuls-ne-sauveront-pas-les-medias-de-leur-marasme-financier-149054>

Roy, J.-H. (2021, 9 décembre). Les Coops de l'information : une expérience unique au monde. *The Conversation*. <http://theconversation.com/les-coops-de-linformation-une-experience-unique-au-monde-171636>

Ryerson University. (2021, 26 février). *Ryerson Journalism Professor April Lindgren Sounds Alarm on Loss of Local Journalism and Rising News Poverty*. City Building Ryerson University. <https://www.ryerson.ca/city-building/news-research/2021/02/april-lindgren-research/>

Sandvig, C., Hamilton, K., Karahalios, K. et Langbort, C. (2014). *Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms*, 23.

Shafer, J. (2021, 12 juin). Opinion | Why Has Local News Collapsed? Blame Readers. *POLITICO*. <https://www.politico.com/news/magazine/2021/06/12/local-news-demand-493495>

Shoemaker, P. J. et Vos, T. P. (2009). *Gatekeeping theory*. Routledge.

Smyrnaioi, N. et Rebillard, F. (2011). Entre coopération et concurrence : Les relations entre infomédiaires et éditeurs de contenus d'actualité. *Colloque - Concurrences*, (3), 46.

Soltani, J. V.-D., Jeremy Singer-Vine and Ashkan. (2012, 24 décembre). Websites Vary Prices, Deals Based on Users' Information. *Wall Street Journal*, Tech. <https://online.wsj.com/article/SB10001424127887323777204578189391813881534.html>

Sullivan, M. (2020). *Ghosting the news: local journalism and the crisis of American democracy*. Columbia Global Reports.

The Globe and Mail Inc. (2020, 20 octobre). *Sophi.io, du Globe and Mail, remporte un prestigieux prix de journalisme en ligne*. Cision. <https://www.prnewswire.com/news-releases/sophi-io-du-globe-and-mail-remporte-un-prestigieux-prix-de-journalisme-en-ligne-866949468.html>

Transcontinental. (2017, 18 avril). *Transcontinental inc. met en place un processus visant la vente de ses journaux locaux et régionaux du Québec et de l'Ontario*. TC Transcontinental. <https://tctranscontinental.com/fr-ca/apercu/salle-de-presse/communiqués-de-presse/transcontinental-inc-met-en-place-un-processus-visant>

Trielli, D. et Diakopoulos, N. (2019). Search as News Curator: The Role of Google in Shaping Attention to News Information. Dans *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 1-15). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300683>