

Des corpus ciblés, d'hier à aujourd'hui, pour poursuivre la description du français en usage au Québec amorcée dans le dictionnaire *Usito*

Hélène Cajolet-Laganière

 <http://orcid.org/0000-0002-1490-9392>
Helene.Cajolet@USherbrooke.ca

Wim Remysen

 <http://orcid.org/0000-0002-4558-5714>
Wim.Remysen@USherbrooke.ca

Centre de recherche interuniversitaire sur le français en usage au Québec,
Université de Sherbrooke (Canada)

Résumé

Cet article est consacré aux corpus qui alimentent le dictionnaire *Usito*, un dictionnaire général du français au Québec. Nous présentons, dans un premier temps, la Banque de données textuelles de Sherbrooke (BDTS) qui a servi de base à l'élaboration du dictionnaire. Il est question, dans un deuxième temps, de l'apport de nouveaux corpus littéraires intégrés au Fonds de données linguistiques du Québec (FDLQ) qui permettent de poursuivre l'enrichissement de la description. Nous illustrons, dans un troisième temps, les liens informatiques établis entre *Usito* et les corpus versés au Fonds de données linguistiques du Québec de façon à décloisonner la description du dictionnaire et à donner accès à un éventail de documents ayant une valeur culturelle et patrimoniale importante.

Mots-clés : dictionnaire *Usito*, Fonds de données linguistiques du Québec, lexicographie, linguistique de corpus, français en usage au Québec, humanités numériques.

Сажетак

Специјализовани корпуси, стари и нови, у служби описа француског језика у употреби у Квебеку започетог у речнику *Usito*. Овај је чланак посвећен корпусима на основу којих се допуњује речник *Usito*, општи речник француског језика који се употребљава у Квебеку. Најпре се представља *Banque de données textuelles de Sherbrooke* (BDTS) ('Шербручка банка текстуалних података'), која је послужила као основа за израду поменутог речника. Затим се говори о доприносу нових књижевних корпуса похрањених у базу *Fonds de données linguistiques du Québec* (FDLQ) ('Фонд квебекских језичких података'), који омогућују да се обogaћивање речничког описа настави. На крају се приказују информатичке везе успостављене између речника *Usito* и корпуса похрањених у бази *Fonds de données linguistiques du Québec*, путем које се може приступити широком спектру докумената од значајне културне и баштинске вредности.

Кључне речи: речник *Usito*, *Fonds de données linguistiques du Québec* ('Фонд квебекских језичких података'), лексикографија, корпусна лингвистика, квебекски француски језик, дигитална хуманистика.

1. Introduction

Le coffre à outils numérique des lexicographes s'est considérablement enrichi au cours des dernières décennies, grâce notamment à la multiplication d'applications permettant d'assister les linguistes dans la constitution de corpus, dans l'analyse et l'exploitation de faits linguistiques (LEBART *et al.* 2019) ainsi que dans la rédaction d'articles de dictionnaire. Sur le plan de l'accès aux données, il suffit de penser entre autres aux agrégateurs de presse (comme Eureka.cc ou RetroNews) ou encore

aux bibliothèques digitales développées par des organismes publics (BANQ, BNF-Gallica, etc.) et par des entreprises privées (dont les outils de Google). Les lexicographes consultent abondamment ces sources, encore que les limites évidentes de certaines d'entre elles (contenus non lemmatisés, métadonnées succinctes, recherches plein texte limitées, etc.) peuvent parfois rendre ces grands ensembles de données difficilement navigables pour le lexicographe cherchant à observer l'usage, qui doit alors se tourner vers d'autres ressources.

C'est la question des corpus exploités aux fins de la description lexicographique qui sera centrale dans cet article. Celui-ci est consacré plus particulièrement aux divers corpus qui alimentent le dictionnaire *Usito*, un dictionnaire général du français au Québec. Il s'agira premièrement de montrer comment le projet ayant mené à la rédaction de ce dictionnaire a largement contribué, à partir de la fin des années 1990, à développer des corpus informatisés destinés à la description scientifique du français en usage au Québec. Deuxièmement, il sera question d'illustrer comment, à date plus récente, de nouvelles initiatives développées au CRIFUQ (le centre de recherche où le dictionnaire *Usito* a vu le jour) servent aujourd'hui à poursuivre l'enrichissement de la description qu'on trouve dans ce dictionnaire, grâce à l'élaboration de nouveaux corpus. Troisièmement, et enfin, nous illustrerons comment ces nouveaux corpus, regroupés dans une interface informatique baptisée Fonds de données linguistiques du Québec, permettent aux utilisatrices et utilisateurs de consulter eux-mêmes des attestations de l'usage québécois.

Bref, nous bouclerons la boucle : des corpus vers la description des articles du dictionnaire, puis des mots décrits dans le dictionnaire vers les corpus de manière à donner au large public un accès convivial au patrimoine linguistique du Québec, en contexte nord-américain. Il s'agit alors de décroquer le dictionnaire de façon à ne plus le considérer comme un répertoire de mots, mais plutôt comme un genre intertextuel complexe.

2. Le dictionnaire *Usito* et la description globale du français au Québec : objectifs, origines et ligne éditoriale

Le dictionnaire *Usito* est né du désir de combler les lacunes des dictionnaires européens (*Petit Robert*, *Petit Larousse*, *Hachette*, etc.), notamment en ce qui a trait à la description de l'usage du français et à la mise en valeur de la culture et des référents culturels du Québec, en contexte nord-américain (CAJOLET-LAGANIÈRE – D'AMICO 2015). Il s'inscrit dans la riche lignée des dictionnaires réalisés sur le continent américain qui ont été conçus pour donner une description plus « américaine » des langues¹. Ce dictionnaire était réclamé au Québec depuis longtemps par de nombreux intervenants pour combler un vide ressenti par les enseignants, les élèves de tous les cycles d'enseignement, les rédacteurs, les journalistes, les communicateurs, les professionnels, etc. Ils déploraient que nos mots, sens et emplois; nos institutions, notre mode de vie, nos caractéristiques sociales, notre vision du monde; nos textes littéraires, journalistiques et utilitaires, etc., soient absents des ou-

vrages existants. Cette situation avait pour effet le prolongement de l'insécurité linguistique des Québécois et Québécoises (CAJOLET-LAGANIÈRE – MARTEL 1998). Ces souhaits rejoignaient une déclaration, souvent citée, de l'Association québécoise des professeurs et professeuses de français (AQPF) en 1977, qui avait exprimé le souhait suivant :

Que la norme du français dans les écoles du Québec soit le français standard d'ici. Le français standard d'ici est la variété de français socialement valorisée que la majorité des Québécois francophones tendent à utiliser dans les situations de communication formelle (AQPF 1977 : 11).

Cet avis a par la suite été affirmé par plusieurs autres instances², signe du consensus grandissant entourant les besoins du public québécois en matière d'outils langagiers qui commence à émerger (rendu possible grâce à l'évolution de la conscience linguistique des Québécoises et Québécois, de plus en plus prompts à revendiquer leurs spécificités linguistiques depuis les années 1970; voir à ce sujet REMYSEN – RHEAULT 2023, REMYSEN 2024)³. Ces besoins étaient, et sont toujours, particulièrement criants pour le milieu de l'éducation : comment les élèves peuvent-ils trouver la signification des milliers de mots et d'expressions qui caractérisent la vie sociale, culturelle, littéraire, juridique et politique québécoise ? Comment les enseignants peuvent-ils enseigner la maîtrise du français dans un contexte fonctionnel, en lien avec les besoins de communication de la société québécoise et francophone en général ? Comment peuvent-ils juger de la qualité de la langue orale et écrite des jeunes ? Ils doivent pouvoir se référer à une norme appropriée et contextualisée, à un standard répertorié dans un outil de référence fiable. C'est également fondamental pour les nouveaux arrivants à qui il importe de fournir les mots et les référents culturels pour mieux s'intégrer à leur nouvelle société d'accueil. C'est de même très pertinent pour l'ensemble de la francophonie et tous les francophiles, qui ont accès à notre littérature, à nos journaux, à nos outils de communication, etc.

C'est dans ce contexte, et avec ces besoins en tête, qu'un nouveau projet de dictionnaire a vu le jour à l'Université de Sherbrooke à la fin des années 1990. La concep-

² Voir entre autres l'avis en ce sens du Conseil de la langue française en 1990, la proposition de politique linguistique du gouvernement intitulé *Le français langue commune : promouvoir l'usage et la qualité du français, langue officielle et langue commune du Québec* (1996), l'avis du ministère de l'Éducation, en 1996, qui refuse d'approuver le dictionnaire *Maxi-Débutant* à cause du caractère trop hexagonal de son contenu, etc.

³ D'autres initiatives avaient vu le jour avant, à la fin des années 1980 (*Dictionnaire du français Plus*, 1988) et au début des années 1990 (*Dictionnaire québécois d'aujourd'hui*, 1992 et réédité en 1993). Ces ouvrages, adaptés de dictionnaires français, n'ont pas réussi à s'établir comme ouvrages de référence, notamment en raison du traitement de la variation géolinguistique et de la norme linguistique en usage au Québec.

¹ Que l'on pense notamment aux dictionnaires de l'anglais nord-américain et canadien ou encore aux dictionnaires de l'espagnol (du Mexique) et du portugais (du Brésil).

tion de ce dictionnaire s'inscrivait alors dans le cadre des travaux du Centre d'analyse et de traitement informatique du français en usage au Québec (CATIFQ)⁴ et plus particulièrement du groupe de recherche FRANQUS (FRANÇAIS québécois usage standard) qui a alors été mis sur pied. Il a donné lieu en 2013 à la publication du dictionnaire électronique *Usito*, résultat d'un travail d'équipe qui s'est étalé sur plus de 20 ans et qui se poursuit encore aujourd'hui⁵.

Dès le départ, l'équipe FRANQUS a eu le souci de planifier le projet en fonction des attentes du public québécois. Pour atteindre cet objectif, elle a mené une vaste enquête sous forme de questionnaire écrit dans six régions du Québec auprès de quelque 800 personnes de différents âges et fonctions⁶. Il ressort de cette enquête que les Québécois et Québécoises souhaitent essentiellement un dictionnaire de type normatif, c'est-à-dire qui les informe sur leur bon usage, et qu'ils désirent connaître, d'une part, les emplois reçus ou acceptés selon un certain standard québécois, et d'autre part, les emplois critiqués de même que les emplois qui peuvent varier en fonction de certaines situations de communication. Ils souhaitent en outre un dictionnaire à l'intérieur duquel les mots et les emplois caractéristiques du Québec sont distingués des mots et des emplois caractéristiques de la France ou de la francophonie européenne à l'aide d'une marque quelconque. L'objectif a donc été de répondre aux besoins exprimés par la population québécoise grâce à un système de marques, d'indicateurs et de remarques normatives⁷. Ces demandes ont été et sont encore au cœur de la politique éditoriale, qui est de décrire le français québécois standard (ou le bon usage du français au Québec) tel qu'il est utilisé chez nos meilleurs écrivains, rédacteurs, journalistes, scientifiques, etc. En même temps, il est important de hiérarchiser les usages québécois autour de ce standard afin de refléter la situation linguistique du Québec et d'éviter la fracture sociale; ainsi, les usages de la langue usuelle utilisée en situation informelle, dans la vie de tous les jours, ont également fait l'objet d'une analyse et d'un traitement rigoureux conformément à la politique éditoriale adoptée.

⁴ Ce centre a été renommé Centre de recherche interuniversitaire sur le français en usage au Québec (CRIFUQ) en 2015. Voir la description du centre à l'adresse <<https://www.usherbrooke.ca/crifuq>>.

⁵ Le dictionnaire *Usito* a été officiellement lancé en 2013, d'abord comme produit commercial. En octobre 2019, il a été rendu disponible gratuitement par l'Université de Sherbrooke. Voir le site Web d'*Usito* pour une présentation plus complète de son historique ainsi que de l'ensemble des personnes, établissements et organismes qui ont collaboré au projet depuis ses origines.

⁶ Pour plus de détails à propos de la place centrale accordée à l'utilisateur dans la démarche lexicographique ayant mené à la publication du dictionnaire *Usito*, voir CAJOLET-LAGANIÈRE 2021b.

⁷ Une attention particulière est portée aux emprunts critiqués à l'anglais (voir CAJOLET-LAGANIÈRE – D'AMICO 2014) et aux avis de recommandation d'organismes officiels tant québécois qu'euro-péens.

Pour la première fois, un dictionnaire était élaboré au Québec à partir de zéro, à l'aide de banques de données textuelles représentatives du français (essentiellement écrit) utilisé au Québec et complètement géré informatiquement à partir d'outils originaux dans un univers sans papier⁸. Il ne s'agissait pas d'ajouter des éléments complémentaires à un dictionnaire existant, mais de procéder à une description nouvelle et originale tant pour ce qui est de son contenu, de sa présentation, de sa consultation et de sa diffusion. C'était aussi la première fois qu'un dictionnaire de langue française distingue clairement le tronc commun du français (partagé avec tous les francophones et présenté sans marque géographique) de même que les emplois caractéristiques du Québec (et plus largement du Canada), et de la France (et plus largement de l'Europe francophone), à l'aide de marques et d'indicateurs appropriés. Cette approche de la variation géographique (décrite plus en détail dans MERCIER 2013) fournit aux lecteurs les passerelles dont ils ont besoin pour s'adapter à divers contextes de communication, à l'échelle nationale comme à l'échelle internationale. Le fait de dégager explicitement non pas un seul, mais deux sous-ensembles géographiquement marqués modifie en profondeur la représentation lexicographique de la variation géographique du français, qui n'apparaît plus comme un phénomène périphérique (et donc marginalisant). Le sous-ensemble non marqué ne correspond plus à un centre dominant, mais à une zone commune. Et au-delà de cette zone commune, les deux sous-ensembles *Q/C* (Québec/Canada) et *F/E* (France/Europe) se répondent comme deux manifestations attendues de la variation géographique qui affecte inévitablement toutes les langues que l'histoire a implantées sur divers continents.

3. L'élaboration du dictionnaire *Usito* à partir de la Banque de données textuelles de Sherbrooke (BDTS)

Les deux premières années du projet FRANQUS ont servi essentiellement à la mise en place des bases de l'environnement numérique, incluant la plateforme logicielle lexicographique⁹ et la BDTS qui a servi de base à la constitution de la nomenclature et à la description des articles du dictionnaire. Il importait en effet d'élaborer dans un premier temps un corpus représentatif des usages de la variété de langue à décrire. Les premiers cinq millions de mots de la BDTS, issus des enquêtes sociolinguistiques menées dans l'Estrie, avaient pour

⁸ Le caractère innovateur du dictionnaire *Usito* est donc double puisqu'il concerne non seulement les contenus élaborés, mais aussi le choix de diffuser exclusivement le dictionnaire dans un environnement numérique (CAJOLET-LAGANIÈRE – MERCIER 2012). Lorsque ce choix a été fait, les dictionnaires papier étaient encore largement prédominants.

⁹ Décrite plus en détail dans MASSON *et al.* 2004.

but une meilleure connaissance du français parlé. Ce corpus a donné lieu à la publication d'un ouvrage inédit, soit *Le Dictionnaire de fréquence des mots du français parlé du Québec* (BEAUCHEMIN *et al.* 1992).

Par ailleurs, afin de procéder à la description lexicographique du français standard en usage au Québec, la BDTS devait refléter les différents usages du français au Québec. Elle devait réunir le plus grand nombre possible de thèmes, de discours et de niveaux de langue du français contemporain oral et écrit en usage au Québec. Pour mener à bien le projet, la taille de la BDTS, fixée théoriquement par l'équipe, est de 35 millions de mots-occurrences répartis dans cinq types de discours : 50 % de textes spécialisés; 20 % de textes littéraires 20 % de textes journalistiques; 5 % de textes didactiques et 5 % de transcriptions de l'oral.

Le sous-ensemble de textes spécialisés contient des textes qui fournissent des mots de langue générale et le vocabulaire de base de nombreux domaines spécialisés. Il renferme des textes très diversifiés : monographies et volumes collectifs, mémoires, thèses, rapports, documents administratifs, etc. Le sous-ensemble de textes littéraires comprend des romans, des essais, des poèmes, des nouvelles, des récits, des pièces de théâtre, etc. Le sous-ensemble de textes journalistiques contient des articles tirés de différentes publications québécoises, comme des quotidiens, des magazines ou périodiques spécialisés, etc. Le sous-ensemble de textes didactiques comprend des textes tirés de manuels scolaires (de niveau secondaire, collégial, universitaire), de logiciels informatiques, etc. Enfin, le sous-ensemble de langue orale contient, d'une part, des échantillons de langue parlée spontanée, c'est-à-dire des transcriptions d'enquêtes sociolinguistiques orales réalisées dans différentes régions du Québec, et d'autre part, des échantillons de langue moins spontanée (contes, monologues, téléromans, textes radiophoniques et télévisés, tribunes téléphoniques, témoignages, etc.).

De fait, au début du travail lexicographique, en 2002, la BDTS comptait 37 millions de mots-occurrences (textes spécialisés : 21 625 341 mots-occurrences; textes littéraires : 5 945 361 mots-occurrences; textes journalistiques : 5 292 813 mots-occurrences; textes didactiques : 2 245 341 mots-occurrences; textes oraux : 1 906 787 mots-occurrences). Le traitement informatique et l'indexation de la BDTS ont permis d'élaborer la nomenclature du dictionnaire en dégagant la fréquence d'usage des mots présents de même que leur dispersion dans les divers types de textes et sa relative permanence dans le temps (CAJOLET-LAGANIÈRE 2021a ; voir aussi CAJOLET-LAGANIÈRE *et al.* 2008, MARTEL – CAJOLET-LAGANIÈRE 2004).

Pour les 37 millions de mots-occurrences extraits de la BDTS, quelque 57 000 mots (noms communs) ont été retenus comme devant faire l'objet d'une entrée du dictionnaire. Cette sélection de mots a par la suite été validée par rapport aux nomenclatures des dictionnaires usuels existants de manière à assurer sa complétude (certains mots essentiels à la communication auraient pu être absents de la BDTS) (CAJOLET-LAGANIÈRE *et al.* 2005). Cette comparaison s'est faite sur la base de dictionnaires tant français que québécois, dont 7 dictionnaires scolaires (dont *Larousse des débutants*, *Mon premier dictionnaire français illustré*, *Dictionnaire CEC Jeunesse* et *Robert Junior illustré*, version nord-américaine), 4 dictionnaires usuels (*Dictionnaire québécois d'aujourd'hui*, *Dictionnaire du français Plus*, *Petit Robert électronique* et *Petit Larousse illustré*) ainsi qu'un dictionnaire de difficultés (*Multidictionnaire*).

Il a fallu préalablement procéder à un travail d'uniformisation des nomenclatures de manière à pouvoir les comparer. Sur les quelque 57 000 mots tirés de la BDTS, 40 000 mots sont communs à la BDTS et à l'une ou l'autre nomenclature des dictionnaires (ces mots constituent en d'autres mots un premier noyau lexical commun à la francophonie qu'il a fallu examiner), alors que 17 000 mots sont spécifiques de la BDTS. De même, quelque 15 000 mots n'apparaissent que dans un seul dictionnaire, sans doute dus aux choix des lexicographes. Sur la base de ces résultats, il a été décidé de fixer la nomenclature du futur dictionnaire à quelque 55 000 mots. En parallèle, une autre nomenclature de quelque 2000 emplois pour les emprunts critiqués à l'anglais a été constituée à partir du *Dictionnaire électronique des anglicismes* qui avait été préalablement élaboré (CAJOLET-LAGANIÈRE *et al.* 2001).

Par ailleurs, tout au long du projet, la BDTS a fait l'objet d'enrichissements ponctuels commandés par les avancées du travail de rédaction lexicographique afin d'assurer la cohérence du traitement des divers sous-ensembles lexicaux (éducation, droit, faune, flore, alimentation, sport, environnement naturel, aspects sociaux, culturels et politiques, etc.) et de répondre aux exigences éditoriales de rédaction (comme la prise en compte des mots nécessaires au travail de définition). Certains sous-ensembles occupent une place particulièrement importante dans cet ensemble de documents et ont été constitués sur la base de critères de sélection ciblés, notamment le sous-ensemble littéraire. Compte tenu que les citations littéraires constituent un élément majeur du dictionnaire, les œuvres étudiées dans les différents programmes scolaires et les œuvres primées ont été privilégiées. L'inclusion de ces textes a toujours été faite en collaboration avec des écrivains et experts en littérature. De même, tout au long de son élaboration, la BDTS a régulièrement bénéficié d'ajouts de nouvelles

sélections représentatives d'articles journalistiques complémentaires. Ces textes ont été ajoutés au cours du projet de manière à actualiser le corpus journalistique, susceptible de fournir de nouveaux emplois en lien avec l'évolution sociale et politique du Québec.

Au moment de son parachèvement, en 2005, la BDTS comptait quelque 52 millions de mots-occurrences extraits de 15 155 documents différents¹⁰ représentatifs des mêmes cinq types de discours : textes spécialisés (59 %), textes littéraires (16 %), textes journalistiques (14 %), textes didactiques (6 %) et textes oraux (5 %). L'originalité et l'intérêt de la BDTS consistent en sa représentativité : la BDTS comprend une variété de textes qui reflètent la langue générale (orale et écrite) de même que la langue littéraire ainsi que la langue plus spécialisée, utilisée au Québec dans différents domaines du savoir. Grâce à son contenu lexical varié, la BDTS permet d'étudier les mots en contexte (y compris le vocabulaire spécialisé), d'élaborer les définitions de manière précise et adaptée à la réalité nord-américaine, d'extraire des citations et des exemples authentiques en vue d'illustrer les emplois, de relever les cooccurents ou les formes en concurrence et d'analyser l'usage lexical dans divers registres de langue, le tout permettant de hiérarchiser les usages. De fait, la BDTS a un caractère unique par rapport aux autres banques de la francophonie (comme Frantext, Beltext, Suistext ou Québétext), composées essentiellement de textes littéraires.

Aujourd'hui, *Usito* compte 75 564 entrées (et sous-entrées). La nomenclature comprend plus de 5000 particularismes québécois et canadiens et quelque 3000 particularismes français et européens. À cela s'ajoutent plus de 1500 emplois critiqués ou officialisés, quelque 1400 rectifications orthographiques, plus de 6000 tableaux de conjugaison, etc. (voir la liste détaillée de ces emplois dans la page d'accueil d'*Usito*). Pour demeurer pertinent et attirant, le dictionnaire doit constamment faire l'objet de mises à jour. En effet, l'actualité sociale et politique, le développement technologique, l'environnement, l'alimentation, etc. ne manquent pas d'avoir des répercussions sur le vocabulaire, sur les mots que l'on emploie et sur les sens qu'on leur donne. Ces évolutions reflètent les mutations de la société, de l'économie, de l'environnement ainsi que les avancées des sciences et de la technologie.

¹⁰ Quand on constitue un corpus de ce genre, on se demande toujours combien de textes doivent être inclus. Quand peut-on considérer que le matériel est suffisamment représentatif pour réaliser le projet souhaité ? La thèse de doctorat de LABRECQUE (2005) présente l'apport de la BDTS quant à la description lexicographique d'une série de mots significatifs, dans le corpus de 37 millions de mots, comparativement à la BDTS enrichie, de 52 millions de mots. Peu de nouvelles informations ont été notées au regard de la langue usuelle. Aussi, le corpus initial de 37 millions de mots était suffisant et pertinent pour réaliser l'objectif de description lexicographique du français standard (langue générale, usuelle) en usage au Québec.

Sans ces mots et sens nouveaux, on ne peut décoder la société dans laquelle nous évoluons. Aussi, le lexicographe a tout intérêt à surveiller l'apparition de nouveaux mots, sens, expressions en consultant les médias (la presse, la télévision, la radio), le cinéma et la littérature, la publicité, les sites Web, les réseaux sociaux et blogues, les menus de restaurants, l'affichage commercial, les offres d'emploi, etc. Avant d'apparaître dans le dictionnaire, les nouveaux emplois détectés sont validés par l'examen de leur fréquence et de leur dispersion dans les différentes sources et contextes pour extraire les hapax, créations d'auteurs, effets de mode, ce qui permet de s'assurer de leur pérennité et de leur caractère standard. Lorsque l'usage d'un nouveau mot ou d'un nouveau sens est validé et que son inclusion respecte la politique éditoriale, il peut être intégré au dictionnaire. Le recours aux différents dictionnaires ou bases de données (TLFi, Vitrine linguistique de l'OQLF, banques de données terminologiques) et l'examen du discours métalinguistique servent de critères supplémentaires afin d'attester le caractère neutre, standard, d'un emploi ne figurant pas aux dictionnaires, ou de certains mots particuliers figurant dans un seul dictionnaire afin de décrire l'emploi de manière pertinente, avec marque et remarque, le cas échéant. À titre d'exemples, voici une série de mots ou de sens en lien avec de nouvelles réalités dans différents domaines maintenant ancrés dans le lexique qui ont été intégrés au cours des dernières années (*Tableau 1*).

EN LIEN AVEC LA PANDÉMIE	<i>coronavirus, covid, confinement, déconfinement, distanciation sociale, physique, maladie à corona virus 2019, l'acronyme COVID-19, transmission communautaire, etc.</i>
RÉSEAU DE LA SANTÉ QUÉBÉCOIS	<i>aide-soignant, préposé aux bénéficiaires, résidence pour personnes âgées, résidence privée pour aînés (RPA), ressource intermédiaire, ressource de type familial, milieu de vie, âgisme, âgiste, capacitisme, capacitiste (QC), validisme, validiste (FE), bientraitance, téléconsultation, télémedecine, anaplasmose, etc.</i>
ALIMENTATION	<i>véganisme, végane, végétalien, végétalisme, vortex de déchets, vin nature, vin naturel, vin de macération, vin orange, etc.</i>
INFORMATIQUE	<i>infolettre, communauté virtuelle, réseau social (avec adaptation de mots courants, comme contact, ami, relation, abonné, etc.), polluposteur, arroseur, spammeur, etc.</i>
ENVIRONNEMENT	<i>microplastique, continent de plastique (ou rare) septième continent, acceptabilité sociale, écopâturage, écofrais, écotaxe, écoanxiété, etc.</i>
FINANCES	<i>qubit, bit quantique, jeton non fongible, chaîne de blocs, etc.</i>

MÉTÉO	<i>bombe météorologique, bombe météo, capsule temporelle, etc.</i>
MUSIQUE ET CINÉMA	<i>album, minialbum, microalbum, EP, single, monoplage, simple, etc.</i>
ÉDUCATION	<i>présentiel, distanciel, etc.</i>
ASPECTS SOCIAUX	<i>féminicide, burkini, allosexuel, altersexuel, queer, etc.</i>

Tableau 1. Nouveaux articles récemment ajoutés ou modifiés (classés par domaine).

L'enrichissement du dictionnaire concerne aussi l'ajout de mots ou d'expressions relevant de la langue familière ou encore de nouvelles francisations proposées pour éviter des emprunts à l'anglais. Certains articles sont modifiés lorsque la norme évolue et que de nouveaux usages sont acceptés dans le registre standard (Tableau 2).

EXPRESSIONS RELEVANT DE LA LANGUE FAMILIÈRE	<i>être assis sur son steak, parler à qqn dans le casque, etc.</i>
EMPRUNTS À TRADUIRE	<i>black Friday, fake news, fatbike, etc.</i>
CHANGEMENTS QUANT À LA NORME	<i>après que</i> (utilisation du subjonctif), <i>être familier avec qqn, condominium, panel, paneliste, etc.</i>

Tableau 2. Nouveaux articles récemment ajoutés ou modifiés (autres aspects).

Il arrive aussi que certains emplois soient ajoutés pour répondre à des demandes d'ajouts des usagers, reçues par courrier électronique. Également, l'examen des recherches infructueuses des usagers peut compléter le travail d'enrichissement. Par ailleurs, depuis quelques années, un effort a été fait pour donner encore plus de visibilité au patrimoine linguistique et culturel du Québec dans le dictionnaire. Ce travail est rendu possible par la constitution d'un nouveau Fonds de données linguistiques du Québec.

4. L'enrichissement du dictionnaire *Usito* et l'apport du Fonds de données linguistiques du Québec (FDLQ) sur le plan patrimonial

Le Fonds de données linguistiques du Québec est une plateforme numérique destinée à rendre accessibles des données de nature linguistique, tant à la communauté scientifique qu'au grand public. Elle réunit plus particulièrement un ensemble de corpus, contemporains et patrimoniaux, qui sont représentatifs des usages du français en contexte québécois, à l'écrit comme à l'oral (voir REMYSEN – ST-AMANT LAMY 2025 pour une présentation plus détaillée). Depuis sa mise en ligne en 2022, le Fonds poursuit trois objectifs : 1^o pérenniser, soit préserver les corpus linguistiques constitués au Québec depuis les années 1960, plusieurs d'entre eux n'ayant

jamais été transférés vers un format numérique permettant leur exploitation avec les outils technologiques actuels; 2^o diffuser, soit rendre accessibles les données contenues dans les divers corpus et permettre d'interroger les contenus en question par l'intermédiaire d'une interface de recherche; 3^o vulgariser, soit diffuser auprès du grand public diverses connaissances concernant sa langue, grâce aux nombreux usages linguistiques que le Fonds permet d'illustrer. La plateforme a ainsi le potentiel de contribuer à valoriser le français du Québec et de contrer certaines idées reçues qui circulent à son sujet.

Le Fonds est aujourd'hui constitué de 26 corpus, ce qui représente un peu plus de 34 750 documents (96,3 millions de mots)¹¹, mais il est constamment enrichi par l'ajout de nouveaux contenus. La plateforme numérique permet des fouilles dans l'ensemble des corpus en fonction de plusieurs critères et métadonnées (certaines données dont l'accès doit être limité pour des raisons de droits d'auteur ou de confidentialité peuvent faire l'objet de recherches dans une interface d'accès limité). Par sa composition, le Fonds a une valeur patrimoniale indéniable, d'une part parce qu'il garde les traces de nombreuses pratiques langagières qui ont forgé le français du Québec et qui lui ont donné son originalité et son caractère unique, d'autre part parce qu'il facilite la découverte et la consultation de documents qui occupent une place de choix dans l'histoire linguistique et culturelle du Québec. En même temps, depuis sa création, le Fonds joue un rôle de catalyseur en ce qu'il a favorisé la constitution de nouveaux corpus qui permettront de garder des traces du patrimoine linguistique de demain.

Même s'il s'agit de deux projets indépendants, les liens entre le Fonds de données linguistiques du Québec et le dictionnaire *Usito* sont multiples. Le Fonds permet notamment d'enrichir les contenus lexicographiques du dictionnaire. Il inclut actuellement cinq corpus textuels, dont une restructuration de la Banque de données textuelles de Sherbrooke ainsi que trois corpus littéraires, soit le Corpus de littérature québécoise (CLIQ; un corpus dont la constitution rappelle le corpus Frantext pour la littérature française), le Corpus de littérature jeunesse québécoise (CLIQ-Jeunesse) et le Corpus Ébullition (composé de bandes dessinées québécoises)¹². Les corpus littéraires s'adressent à un public varié, mais sont d'un intérêt particulièrement marqué pour le milieu de l'éducation, où la littérature est souvent utilisée à des fins didactiques.

¹¹ Chiffres en date du 19 février 2026.

¹² Pour une présentation plus détaillée, voir les fiches de présentation des corpus disponibles sur le site Web du Fonds, à l'adresse <<https://fdlq.recherche.usherbrooke.ca>> (onglet « Présentation des corpus »). Pour plus de détails méthodologiques concernant la constitution des corpus, voir aussi GIAUFRET *et al.* 2022, PAQUET-GAUTHIER – REMYSEN 2024.

Les divers corpus textuels réunis au Fonds sont tout à fait complémentaires. Si la BDTs se concentre essentiellement sur des textes parus entre les années 1960 et les années 2000, le corpus CLIQ complète la composante littéraire de la BDTs par l'ajout d'œuvres patrimoniales (parues avant les années 1960) et contemporaines (publiées à date plus récente, depuis 2000). Les corpus CLIQ-Jeunesse et Ébullition permettent pour leur part de diversifier les genres textuels représentés dans la BDTs, qui n'inclut pas de littérature jeunesse ou de bandes dessinées. Or ces deux formes littéraires ont connu dans les dernières années un engouement particulièrement important au Québec, comme en fait foi la multiplication de publications, et il était donc important de pouvoir tenir compte de ces manifestations culturelles dans la description de la langue. À l'heure actuelle, plus de 1000 œuvres littéraires appartenant à ces trois corpus ont déjà été publiées sur la plateforme et le développement des corpus se poursuivra dans les prochaines années. Ces corpus donnent accès à des milliers de citations littéraires de tous genres. C'est ainsi que dès 2021, l'équipe d'*Usito* a procédé à l'enrichissement de l'exemplification, par l'ajout de citations littéraires, dans le dictionnaire de façon à assurer une plus grande représentativité de la littérature patrimoniale et contemporaine.

Pour mieux encadrer ce travail, l'équipe a d'abord constitué une liste de québécismes (linguistiques et référentiels) décrits dans *Usito*, mais sans citation littéraire pour illustrer leur emploi. Ces emplois, en grande partie des québécismes portant la marque *littéraire*, *familier*, *vieilli* ou *anciennement*, et plus largement des emplois ayant une valeur culturelle et encyclopédique importante, ont été priorisés. Cela dit, d'autres articles qui ne contiennent pas de québécismes ont également bénéficié de l'ajout de citations. En tout, près de 850 citations ont déjà été ajoutées et publiées dans le dictionnaire. Ces citations ont permis d'enrichir des articles portant sur des mots appartenant à des domaines très variés, notamment la faune et la flore, les sports, l'histoire et le territoire, la cuisine, l'éducation, les arts et métiers ainsi que la vie quotidienne. Voici une série d'exemples de citations anciennes (*Tableau 3*) et contemporaines (*Tableau 4*) tirées du corpus CLIQ qui viennent bonifier les articles du dictionnaire *Usito*.

LOUP-MARIN	<i>La fenêtre d'une sorte d'appentis, [...] s'ouvrit doucement, et une tête brune, coiffée d'une casquette de loup-marin, surgit.</i> (W.-E. Dick, 1897)
CABANE À SUCRE	<i>Quand les goudrelles sont préparées, on chausse les raquettes pour aller visiter et mettre en ordre la cabane à sucre dans laquelle ont été emmagasinés, à la fin de la saison précédente, les baquets, les cassots, les chaudières, les moules et autres ustensiles du sucrier.</i> (N. Legendre, 1898)

SENELIER	<i>Au fond se trouvait une belle rangée de hauts arbres fruitiers, et au sud, d'autres arbres à tiges moins élevées, tels que senelliers, gadeliers, groseilliers, framboisiers, etc.</i> (A. Gérin-Lajoie, 1876)
PASSER AU FEU	<i>J'ai appris dernièrement que leur maison avait passé au feu et était entièrement détruite.</i> (A. Buies, 1890)
ÉCORNIFLEUR	<i>Du cadre de la porte, les petits écornifleurs disparurent à toutes jambes, allant porter à tout le bas du rang la nouvelle de l'arrivée d'un petit monsieur de la ville.</i> (Frère Marie-Victorin, 1919)
SOULIERS DE BŒUF	<i>Ils allaient bon train depuis quelques instants, leurs petits souliers de bœuf faisant craquer les mottes de terre sèche.</i> (J. A. Loranger, 1925)
TARTE À LA FERLOUCHE	<i>Les tartes à la ferlouche étaient et sont encore un dessert bien populaire dans les campagnes. Dans ces tartes les confitures étaient remplacées par un mélange de mélasse et de farine.</i> (H. Berthelot, 1924)
PRÔNE	<i>Le dimanche précédent, le curé avait rappelé au prône l'obligation pour tous les fidèles de s'approcher du tribunal de la pénitence et de recevoir la sainte communion.</i> (A. Laberge, 1942)

Tableau 3. Exemples de citations anciennes tirées du corpus CLIQ.

COMME DU MONDE	<i>Ma mère m'avait appris comment plier des t-shirts comme du monde en ramenant les manches dans le milieu avant de replier le haut sur le bas.</i> (G. Pettersen, 2014)
ONCE	<i>Elle est d'une telle finesse, pas une once de vulgarité dans ses remarques.</i> (M.-C. Agnant, 2015)
DEC	<i>Elles sont en dernière session de DEC, l'une va étudier en littérature l'automne prochain, l'autre s'est inscrite en droit.</i> (A. Morin, 2018)
CENTRE COMMUNAUTAIRE	<i>Les répétitions de la chorale avaient lieu les lundis soir au centre communautaire.</i> (F. Britt, 2020)
AVOIR LA MÈCHE COURTE	<i>Il était ponctuel, n'avait pas d'antécédents criminels; qu'est-ce que ça pouvait faire qu'il ait la mèche courte ?</i> (C. Dawson, 2020)
POUCEUX	<i>Le pouceux, lui, raconte son voyage dans l'Ouest canadien, persuadé que ça le rend spécial.</i> (M. Andraos, 2021)
POMME DE TERRE GRELOT	<i>Carottes nantaises, pommes de terre grelots, topinambours, petits oignons, haricots jaunes, garnis de persil haché; un filet d'omble de l'Arctique.</i> (A. Farah, 2021)
CALLER	<i>Quand il callait l'original, c'est tout un troupeau de mâles en rut qui débarquait telle une harde de caribous.</i> (L.-K. Picard-Siouli, 2022)

Tableau 4. Exemples de citations contemporaines tirées du corpus CLIQ.

5. Des corpus aux dictionnaires, des dictionnaires aux corpus : tirer profit de toutes les ressources disponibles

La pratique lexicographique menant à la rédaction des articles du dictionnaire *Usito* relève de la linguistique de corpus. En effet, toute prise de décision concernant l'ajout d'un mot (ou d'un sens) à la nomenclature et le traitement du mot (ou du sens) en question est ancrée dans la consultation de corpus contenant des données authentiques et contextualisées. Cette approche assure une certaine objectivité à la démarche et permet de rendre compte des usages de la langue en tenant compte de leur représentation collective. Habituellement, l'utilisateur n'a pas accès à ces données et il doit se contenter des attestations citées dans les articles de dictionnaire qu'il consulte pour avoir des illustrations concrètes d'un mot, d'une expression ou d'un sens. Selon la tradition lexicographique, ces attestations relèvent généralement de corpus littéraires, les écrivaines et écrivains étant généralement considérés comme des modèles en matière de bon usage (encore que les œuvres littéraires puissent donner accès à la langue familière et populaire). Malgré leur intérêt, ces données demeurent présélectionnées par le lexicographe, qui peut décider de retenir telle ou telle citation littéraire pour illustrer le sens d'un mot ou pour fournir une information encyclopédique complémentaire à la définition.

Or, la mise à disposition des corpus par l'intermédiaire du Fonds de données linguistiques du Québec permet de changer la donne et de décroisonner la description lexicographique, en l'ancrant dans un univers linguistico-textuel et culturel dont il fait lui-même partie. Une telle intertextualité numérique est au cœur des humanités numériques dites « 2.0 » (CITTON 2015). Ainsi, le FDLQ permet aux personnes intéressées de trouver des attestations authentiques, tant dans la langue parlée que dans la langue écrite, des emplois répertoriés dans le dictionnaire, tout en contrôlant la qualité des données qui peuvent être interrogées¹³. Il permet, entre autres recherches, d'interroger des œuvres littéraires parues au Québec depuis la fin du 19^e siècle jusqu'à nos jours et d'y puiser des citations de nos auteurs et autrices. Tous les corpus diffusés sur la plateforme FDLQ sont dotés de données de référence (métadonnées) riches, structurées et pérennes qui facilitent le repérage des contenus en fonction de divers critères. La disponibilité de ces informations sur la plateforme du Fonds est extrême-

ment importante : puisque les données versées au Fonds sont moissonnées par les moteurs de recherche, elles ont le potentiel d'augmenter la découvrabilité des documents qu'il contient et, partant, de rendre plus visible le patrimoine linguistique et culturel du Québec. Il s'agit d'un enjeu particulièrement important à l'heure actuelle, où les gouvernements sont de plus en plus nombreux à s'interroger sur les meilleures façons de rendre leur culture visible et disponible. Lorsque les droits d'auteur le permettent, le Fonds permet en outre de consulter les documents dans leur intégralité¹⁴.

Des hyperliens ajoutés sur la plateforme FDLQ et dans le dictionnaire *Usito* permettent à l'utilisateur de découvrir les possibilités d'exploitation dont il dispose et de naviguer entre les deux ressources. D'une part, la plateforme FDLQ permet à l'utilisateur qui consulte des attestations dans les corpus de chercher la définition d'un mot (ou encore d'une expression qui contient un mot donné) dans le dictionnaire *Usito*, et ce, directement à partir des résultats d'une recherche en cliquant sur l'hyperlien prévu à cet effet. D'autre part, un hyperlien ajouté dans le dictionnaire *Usito* invite l'utilisateur à faire des recherches associées dans le Fonds (au complet) ou dans certains corpus littéraires pour découvrir des exemples des mots consultés. Les liens réciproques ainsi établis entre le dictionnaire et les corpus permettent de décroisonner les deux. Dans cette vision des choses en effet, le dictionnaire n'est pas considéré comme un simple répertoire de mots, mais plutôt comme un genre intertextuel complexe qui entretient des liens avec d'autres productions écrites. Le dictionnaire est plutôt conçu comme une représentation de la langue (une description suivant une tradition bien établie), soit une synthèse d'une somme de connaissances, qui peut à son tour donner accès à un ensemble de ressources linguistico-culturelles (contextualisation des données). Le dictionnaire n'est plus seulement un outil de recherche (servant à vérifier une information de nature linguistique : orthographe, sens, appartenance à un registre, statut normatif), mais aussi un outil d'exploration et un véritable levier en vue d'optimiser la découvrabilité et le référencement. Cette réciprocité accrue entre les corpus (données premières) et les outils lexicographiques disponibles en ligne permet aux uns de découvrir les autres (TASOVAC 2010).

¹³ Un tel contrôle apparaît essentiel. Certains outils lexicographiques, par le recours à des algorithmes, permettent de moissonner des données disponibles sur le Web, et de trouver des attestations d'emplois, mais sans aucune vérification concernant les sources utilisées. Par ailleurs, pour que les données authentiques mises à la disposition des utilisateurs puissent être correctement utilisées, un minimum de traitement est nécessaire en termes de lemmatisation et d'étiquetage morphosyntaxique. De telles précautions ne sont généralement pas prises par les outils qui fournissent des exemples *in situ*.

¹⁴ À titre d'exemple, le roman *L'influence d'un livre* de Philippe Aubert de Gaspé (fils), souvent considéré comme une des œuvres fondatrices de la littérature québécoise, peut être consulté sur la plateforme du Web, où on peut lire la transcription intégrale du document ou encore consulter une version numérique (au format PDF) de l'original publié en 1837. Voir l'URL <<https://fdlq.criфу.usherbrooke.ca/corpus/cf09e093-ad65-4978-8095-e708c410fd8d/documents/6f6110cb-b57d-4932-a50b-a11c76d5f282#texteIntegral>>.

6. Conclusion

Depuis son lancement, *Usito* constitue l'ouvrage de référence le plus complet du français standard utilisé au Québec et son succès ne se dément pas, comme en témoigne le nombre de consultations par année. En 2023, à peine quatre ans après l'annonce de la gratuité du dictionnaire par l'Université de Sherbrooke, *Usito* a franchi le cap de 10 millions d'utilisateurs et d'utilisatrices. En 2025 seulement, le dictionnaire a attiré 8,6 millions d'usagères et d'usagers individuels (en tout, 69 millions de pages vues).

Quant au Fonds de données linguistiques, depuis son lancement en 2022, la plateforme a attiré près de 100 000 utilisatrices et utilisateurs différents et ce nombre ne cesse d'augmenter. De même qu'*Usito*, le Fonds est principalement consulté au Québec et au Canada, mais rejoint aussi un public intéressé dans les autres pays de la francophonie (avant tout la France, la Belgique, la Suisse et les pays du Maghreb) ainsi qu'aux États-Unis. Grâce à l'existence des deux ressources, la description du français tel qu'il est utilisé au Québec s'en trouve enrichie.

En plus, grâce à la complémentarité entre les outils par l'ajout d'hyperliens permettant de naviguer de l'un à l'autre, les ressources développées au CRIFUQ fournissent des clés d'interprétation fiables pour les utilisatrices et utilisateurs désireux d'en savoir plus sur le français du Québec, sur sa culture ainsi que sur l'environnement dans lequel il évolue.

Références bibliographiques

- ASSOCIATION QUÉBÉCOISE DES PROFESSEURS DE FRANÇAIS (AQPF) (1977). Les résolutions de l'Assemblée générale, *Québec français*, 28, 11.
- BEAUCHEMIN, Normand, MARTEL, Pierre, THÉORET, Michel (1992). *Dictionnaire de fréquence des mots du français parlé au Québec : fréquence, dispersion, usage et écart réduit*, New York : Peter Lang.
- CAJOLET-LAGANIÈRE, Hélène (2021a). L'essor d'une norme endogène au Québec : l'exemple du dictionnaire *Usito*, *Gragoatá*, 26(54), 105–138.
- CAJOLET-LAGANIÈRE, Hélène (2021b). L'utilisateur au cœur de la démarche lexicographique d'*Usito*, *Éla : études de linguistique appliquée*, 201(1), 105–119.
- CAJOLET-LAGANIÈRE, Hélène, D'AMICO, Serge (2014). Le traitement des emprunts critiqués à l'anglais dans le *Dictionnaire de la langue française : le français vu du Québec (FVQ)*. In : Wim Remysen (dir.), *Les français d'ici : du discours d'autorité à la description des normes et des usages* (pp. 141–162). Québec : Presses de l'Université Laval.
- CAJOLET-LAGANIÈRE, Hélène, D'AMICO, Serge (2015). Pour une approche plus ouverte du français et de sa variation géographique, pour une meilleure prise en compte des contextes québécois, canadien et nord-américain. In : Katja Sarkowsky, Rainer-Olaf Schultze et Sabine Schwarze (dir.), *Migration, Regionalization, Citizenship Comparing Canada and Europe* (pp. 209–229). Wiesbaden : Springer.
- CAJOLET-LAGANIÈRE, Hélène, LABRECQUE, Geneviève, MARTEL, Pierre, MASSON, Chantal-Édith, MERCIER, Louis, THÉORET, Michel (2008). *Dictionnaires usuels du français et Banque de données textuelles de Sherbrooke (BDTS) : convergence et divergence des nomenclatures*. In : Brigitte Horiot et Chiari Bignamini-Verhoeven (dir.), *Français du Canada – français de France VII. Actes du septième Colloque international de Lyon, du 16 au 18 juin 2003* (pp. 9–28). Tübingen : Max Niemeyer.
- CAJOLET-LAGANIÈRE, Hélène, MARTEL, Pierre (1998). La reconnaissance d'un usage standard propre au Québec ou le moyen de vaincre notre insécurité linguistique. In : Annette Boudreau et Lise Dubois (dir.), *Le français, langue maternelle, dans les collèges et universités en milieu minoritaire* (pp. 125–140). Moncton : Éditions de l'Acadie.
- CAJOLET-LAGANIÈRE, Hélène, MARTEL, Pierre, LANGLOIS, Marie-France (2001). Les textes journalistiques sont-ils « envahis » par les emprunts critiqués à l'anglais ? *Terminogramme*, 97-98, 47–72.
- CAJOLET-LAGANIÈRE, Hélène, MARTEL, Pierre, MASSON, Chantal-Édith (2005). Accroissement et particularités des nomenclatures propres à une série de dictionnaires usuels du français. In : Brigitte Horiot, Elmat Schafroth et Marie-Rose Simoni-Aurembou (dir.), *Mélanges offerts au professeur Lothar Wolf : « Je parle, donc je suis... de quelque part »* (pp. 477–497). Lyon : Université Lyon III Jean Moulin, Centre d'études linguistiques J. Goudet.
- CAJOLET-LAGANIÈRE, Hélène, MARTEL, Pierre, MASSON, Chantal-Édith, MERCIER, Louis (dir.), *Usito*, Université de Sherbrooke. <<https://usito.usherbrooke.ca/>>. [19/02/2026].
- CAJOLET-LAGANIÈRE, Hélène, MERCIER, Louis (2012). En quoi l'environnement numérique contribue-t-il au renouvellement de la pratique et du discours lexicographiques ? Conférence présentée au colloque international, 4^e Journée québécoise des dictionnaires. *Du papier au numérique : la mutation des dictionnaires*, 4 octobre 2012, Montréal.
- CITTON, Yves (2015). Humanités numériques : une médiapolitique des savoirs encore à inventer. *Multitudes*, 59, 169–180.
- Fonds de données linguistiques du Québec. Centre de recherche interuniversitaire sur le français en usage au Québec, Université de Sherbrooke. <<https://fdlq.recherche.usherbrooke.ca/>>. [19/02/2026].
- GIAUFRET, Anna, REMYSEN, Wim, RIOUX, Philippe (2022). Le corpus Ébullition : un outil pour l'analyse de la représentation de l'espace et du paysage linguistique dans la BD québécoise, *Publifarum*, 38(2), 63–91.
- LABRECQUE, Geneviève (2005). *Les apports et les limites de la Banque de données textuelles de Sherbrooke au regard de la description lexicographique du français en usage au Québec*, [thèse de doctorat, Université de Sherbrooke]. USavoirs. <<http://savoires.usherbrooke.ca/handle/11143/2748>>. [19/02/2026].
- LEBART, Ludovic, PINCEMIN, Bénédicte, POUDAT, Céline (2019). *Analyse des données textuelles*, Québec : Presses de l'Université du Québec.
- MARTEL, Pierre, CAJOLET-LAGANIÈRE, Hélène (2004). L'apport de la Banque de données textuelles de Sherbrooke (BDTS) : des nomenclatures enrichies. In : Louis Mercier et Hélène Cajolet-Laganière (dir.), *Français du Canada – français de France. Actes du sixième Colloque international d'Orford, Québec, du 26 au 29 septembre 2000* (pp. 263–278). Tübingen : Max Niemeyer Verlag.
- MASSON, Chantal-Édith, CAJOLET-LAGANIÈRE, Hélène, MARTEL, Pierre (2004). La BDTS-concordances : un outil d'enrichissement de la pratique lexicographique. In : Gérald Purnelle, Cédric Faron et Anne Dister (dir.), *Le poids des mots. Actes des 7^{es} Journées internationales d'Analyse statistique des Données Textuelles* (volume II) (pp. 764). Louvain-la-Neuve : Presses universitaires de Louvain.
- MERCIER, Louis (2013). Un nouveau dictionnaire général de la langue française qui vient du Québec, mais pourquoi donc ? *Repères-Dorif autour du français : langues, cultures et plurilinguisme*, 2(2). <<https://www.dorif.it/reperes/louis-mercier-un-nouveau-dictionnaire-general-de-la-langue-francaise-qui-vient-du-quebec-mais-pourquoi-donc/>>. [19/02/2026].

- PAQUET-GAUTHIER, Myriam, REMYSEN, Wim (2024). Les usages graphiques dans les œuvres littéraires québécoises d'avant 1940. *Le français moderne*, 92(2), 195–211.
- REMYSEN, Wim (2024). Comment les Québécois perçoivent-ils le français ? Bilan de 60 ans de recherches menées sur les représentations linguistiques au Québec. In : Aurore Dumont et Dan Van Raemdonck (dir.), *La langue française en représentation(s). Actes du colloque de Mons, 2022* (pp. 98–113). Bruxelles : Fédération Wallonie-Bruxelles/Direction de la Langue française.
- REMYSEN, Wim, RHEAULT, Amélie-Hélène (2023). La linguistique populaire dans la Romania : le français au Québec. In : Lidia Becker, Sandra Herling et Holger Wochele (dir.), *Manuel de linguistique populaire* (pp. 423–447), Berlin/Boston : De Gruyter.
- REMYSEN, Wim, SAINT-AMANT LAMY, Hugo (2025). Les corpus oraux montréalais à l'ère numérique : des bobines magnétiques à la diffusion sur la plateforme du Fonds de données linguistiques du Québec. In : Wim Remysen et Hélène Blondeau (dir.), avec la collaboration de Marty Laforest, *(Re)donner la parole aux corpus montréalais : regards rétrospectif et prospectif* (pp. 209–234). Montréal : Presses de l'Université de Montréal.
- TASOVAC, Toma (2010). Reimagining the dictionary, of why lexicography needs digital humanities. Conférence présentée à *Digital Humanities 2010*, du 7 au 10 juillet, Londres : King's College.

Hélène Cajolet-Laganière — Wim Remysen

Describing the French Language Used in Québec: Which Are the Corpora Used to Enrich the Dictionary *Usito*?

(Summary)

This paper provides an overview of the corpora at the base of *Usito*, a general dictionary of French used in Québec. We will first present the *Banque de données textuelles de Sherbrooke* (BDTS), which served as the foundation for the dictionary. In the second part, we will discuss the use of new literary corpora integrated into the *Fonds de données linguistiques du Québec* (FDLQ), which allow for the continued enrichment of the description. Finally, in the third part, we will illustrate the hyperlinks established between *Usito* and the corpora deposited in the *Fonds de données linguistiques du Québec*, thus providing access to a range of documents that hold significant cultural and heritage value.

Keywords : *Usito* dictionary, *Fonds de données linguistiques du Québec*, lexicography, corpus linguistics, French in Québec, digital humanities.