

Le Corpus FRAN : réseaux et maillages en Amérique française

Connecting networks of North American French

France Martineau et Marie-Claude Séguin



Édition électronique

URL : <http://journals.openedition.org/corpus/2925>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 15 octobre 2016

ISSN : 1638-9808

Référence électronique

France Martineau et Marie-Claude Séguin, « Le Corpus FRAN : réseaux et maillages en Amérique française », *Corpus* [En ligne], 15 | 2016, mis en ligne le 15 janvier 2017, consulté le 10 novembre 2018.
URL : <http://journals.openedition.org/corpus/2925>

Ce document a été généré automatiquement le 10 novembre 2018.

© Tous droits réservés

Le Corpus FRAN : réseaux et maillages en Amérique française

Connecting networks of North American French

France Martineau et Marie-Claude Séguin

NOTE DE L'AUTEUR

Cet article a reçu l'appui du projet *Le français à la mesure d'un continent : un patrimoine en partage* (dir. F. Martineau, CRSH/GTRC). Nous remercions Jérémie Beauchamp, Mélissa Chiasson et Jade Dumouchel-Trudeau, assistants au projet à l'Université d'Ottawa, pour l'aide au repérage des données sur des variables.

- 1 Notre objectif est d'exposer dans cet article les principes qui ont prévalu à l'élaboration du corpus FRAN (Français d'Amérique du Nord) développé dans le cadre du projet *Le français à la mesure d'un continent : un patrimoine en partage*, basé à l'Université d'Ottawa. Dans un premier temps, nous présentons le projet *Le français à la mesure d'un continent*, l'équipe de nature interdisciplinaire, ses objectifs et ses questions de recherche, en particulier en ce qui a trait au développement des communautés francophones nord-américaines. Puis nous présentons le Corpus FRAN, premier corpus panfrancophone en ligne portant sur l'Amérique française, son architecture et les protocoles de transcription de données. Enfin, nous illustrons les types de recherches qui peuvent être menées, en montrant comment une perspective par réseaux et maillages permet un regard novateur sur les usages, d'abord à partir d'une étude de variable, stratifiée socialement (la première personne du singulier du verbe *aller*, comme auxiliaire du futur périphrastique) sur le terrain montréalais d'Hochelaga-Maisonneuve puis à partir d'un survol d'une deuxième variable, la conséquence (*so, donc, alors, (ça) fait que*), révélatrice de comportements linguistiques différents dans des communautés en contexte minoritaire et en contexte majoritaire.

1. Le projet *Le français à la mesure d'un continent*

- 2 Le projet *Le français à la mesure d'un continent : un patrimoine en partage*¹ est un projet international, subventionné, en 2011, pour une durée de sept ans par le Conseil de recherche en sciences humaines du Canada, dans le cadre du programme des Grands travaux de recherche concertée. Il est constitué de deux grands axes de recherche qui se répondent et s'entrecroisent.

Axe 1 – Histoire sociale et Identités des communautés francophones : les idéologies linguistiques

- 3 Déterminer comment les idéologies, les représentations et les pratiques autour de la langue ont structuré et structurent encore les rapports à l'intérieur des groupes francophones, et entre groupes francophones, anglophones et allophones.

Axe 2 – Le français en Amérique du Nord : variétés du français en contexte majoritaire et minoritaire

- 4 Par l'entremise d'une comparaison panlectale et diachronique, déterminer ce qui caractérise les usages linguistiques des francophones en Amérique du Nord ; évaluer les effets linguistiques du contact ; déterminer quelles sont les modalités d'appropriation du français par les non-francophones.
- 5 L'équipe réunit des chercheurs de plusieurs disciplines (linguistique, histoire, sociologie, anthropologie, ethnologie, géographie, littérature et archivistique), provenant de 44 universités. Le cœur de l'équipe scientifique réunit 13 chercheurs, provenant de France (Françoise Gadet, Université Paris Ouest Nanterre La Défense ; André Thibault, Paris Sorbonne), des États-Unis (Hélène Blondeau, University of Florida ; Sylvie Dubois, Louisiana State University) et du Canada (Mourad Ali-Khodja & Annette Boudreau, Université de Moncton ; Mireille Tremblay, Université de Montréal ; Jean-Pierre Le Glaunec & Wim Remysen, Université de Sherbrooke ; Paul Cohen, University of Toronto ; Raymond Mougeon, York University ; Yves Frenette, Université de Saint-Boniface ; & France Martineau, Université d'Ottawa, directrice), auxquels se joint une coordonnatrice de recherche, Marie-Claude Séguin, de l'Université d'Ottawa.
- 6 Du point de vue des idéologies et représentations, notre démarche consiste à croiser différents terrains d'observation en Amérique du Nord et différentes périodes, de façon à permettre de saisir les éléments que partagent les diverses communautés quant à leurs rapports aux idéologies et aux représentations linguistiques, ainsi que ce qui les distingue en fonction des événements politiques, sociaux, juridiques, économiques, religieux et culturels propres à chacun des contextes. Ainsi, en Amérique du Nord, les discours sur la langue ont fortement contribué à construire chez les locuteurs des représentations ambivalentes à l'égard du français parlé. Les archaïsmes, associés par l'élite à l'authenticité du vernaculaire franco-canadien, ont été reconnus comme légitimes en raison de leur filiation avec le français parlé en France, alors que les anglicismes ont été rejetés, parce qu'ils symbolisaient l'assimilation à la culture dominante (Bouchard, 2002 ; Boudreau, 2009 ; Boudreau & Ali-Khodja, 2009 ; Remysen, 2010). À un niveau macro-

sociétal, ces représentations sont par ailleurs liées à l'idéologie de l'État-nation (Gellner, 1989 ; Hobsbawm, 1992 ; Anderson, 1996). En effet, en tant que formes politico-culturelles et souvent dans une grande proximité au discours religieux, les États-nations ont mobilisé au cours de leur construction à la fois un « travail des langues » et un « travail sur les langues », l'un comme l'autre impliquant qu'à un territoire donné soit associée une langue particulière.

- 7 Du point de vue des usages, le projet s'intéresse à la rencontre de groupes francophones avec d'autres groupes francophones (laurentien, acadien, européen) ou groupes linguistiques (anglophones, amérindiens, notamment) à divers moments et à différentes périodes en Amérique du Nord. Même si tous les linguistes admettent que le contact linguistique joue un rôle dans la variation et le changement linguistique, le poids relatif des facteurs externes et internes reste à être vérifié. Un même phénomène linguistique peut être compris comme provoqué/ facilité par le contact entre groupes linguistiques, ou bien découler de l'effet d'une dynamique interne (Mougeon & Beniak, 1991 ; Mougeon, Nadasdi & Rehner, 2005 ; Poplack & Levey, 2011). Les faits de variation sont partie prenante de *variétés*, même si la fluidité de la langue et du changement qui lui est associé correspondent moins à des communautés définies par des frontières géopolitiques qu'à des réseaux sociaux (en particulier familiaux) à l'intérieur desquels la langue ou les langues du locuteur sont transmises. Notre approche se distingue d'approches plus traditionnelles de la variation en dégageant, dans nos études de terrains, à la fois les réseaux sociaux qui définissent des communautés au sens large et des variables sensibles à la mixité sociale et culturelle qui permettent de mesurer le degré de cohésion sociolinguistique d'une communauté et les changements qui la traversent (Gadet & Martineau, 2012).

2. Le Corpus FRAN

2.1 Structure du Corpus FRAN

- 8 Le Corpus FRAN assure un ancrage empirique commun aux deux axes, favorisant ainsi le dialogue interdisciplinaire et une approche globale de la variation linguistique ainsi que des représentations sur la langue. Le corpus est établi de façon à permettre une comparaison panlectale des variétés de français d'Amérique du Nord dans une perspective à la fois diachronique et synchronique. L'objectif est de comprendre l'évolution de communautés, des réseaux qui se sont tissés et des effets sur la langue et les représentations linguistiques générées. En ce sens, il était important de baser le Corpus FRAN sur un nombre de communautés choisies selon leur ancrage historique et les contacts avec d'autres groupes francophones, anglophones et allophones ; ont été choisies, en Acadie, Moncton et Baie Sainte-Marie ; au Québec, Montréal et en particulier deux quartiers pour les enquêtes modernes, soit Hochelaga-Maisonneuve et Saint-Michel/Montréal-Nord, ainsi que Chicoutimi et Gatineau ; en Ontario, Welland, Windsor et Hearst ; dans l'Ouest canadien, Saint-Boniface au Manitoba ; aux États-Unis, Gardner et Waterville en Nouvelle-Angleterre et en Louisiane, la Nouvelle-Orléans et Lafourche ; en France, Paris et Rouen. À ces terrains se greffe, pour l'éclairage apporté sur les origines du français colonial, le français des Antilles. Trois grands ensembles forment ce corpus et permettent de créer des ponts entre les deux axes du projet, et entre les perspectives diachronique et synchronique.

2.1.1 Corpus historiques

- 9 Les corpus historiques sont constitués de correspondance privée (lettres, journaux personnels) qui permettent de retracer des marques de la langue parlée chez des scripteurs malhabiles (Schneider, 2002 ; Martineau, 2007 ; Ernst, 2010 ; van der Wal, Rutten & Simons, 2012) et qui font ainsi le pont avec des corpus oraux modernes pour l'étude de la variation linguistique. Nous avons aussi examiné la correspondance du clergé et de la presse de façon à cerner les idéologies linguistiques qui s'y déploient ; ce corpus répond ainsi à ceux constitués par les entrevues modernes des personnalités (voir en 2.1.3). À ces corpus nous avons ajouté des enquêtes de nature dialectologique et sociolinguistique sur les langues créoles et le français des Antilles qui constituent un apport précieux à la connaissance du français parlé au début de la colonisation, dans la suite de travaux de ce type (Poirier, 1979 ; Chaudenson, Mougeon & Béniak 1993 ; Thibault, 2008).

2.1.2 Corpus patrimoniaux

- 10 L'établissement d'ententes avec des chercheurs a permis l'apport de corpus patrimoniaux recueillis dans le dernier quart du XX^e siècle (corpus Lefebvre-Drapeau sur le quartier Centre-Sud à Montréal ; corpus Fox-Smith sur la Nouvelle-Angleterre ; corpus S. Dubois sur la Louisiane ; corpus Boudreau-Dubois & Marie-Marthe Roy sur Moncton ; corpus Mougeon sur Welland 1975). Cette avancée unique dans le partage de corpus a enrichi le corpus FRAN et permet la préservation de ces corpus et leur diffusion plus large au sein de la communauté scientifique. Ces corpus permettent des comparaisons en temps réel avec nos corpus modernes sur plusieurs décennies (1970-2014) pour suivre l'évolution de la langue.

2.1.3 Corpus modernes

- 11 Pour sonder l'ensemble des pratiques linguistiques, les nouveaux corpus sont de trois types. Tout d'abord, des entrevues auprès de personnalités (des *leaders* dans la communauté) ont été effectuées dans des communautés en contexte minoritaire, là où l'insécurité linguistique est souvent aiguë (Francard, 1994). Ces entrevues nous permettent de cerner les représentations et les pratiques linguistiques. Ensuite, des entrevues de nature variationniste ont aussi été effectuées de façon à cartographier les usages linguistiques selon des paramètres sociolinguistiques (âge, classe sociale, éducation, etc.). Le corpus recueilli dans la métropole montréalaise permet une comparaison avec la situation linguistique à Paris, autre métropole francophone. Le choix de Welland et de Montréal nous permet également de jeter les bases d'une comparaison en temps réel avec nos corpus patrimoniaux des années 1970 versés au Corpus FRAN. Enfin, nous avons aussi effectué des entrevues de nature écologique, c'est-à-dire des enregistrements recueillis dans des situations naturelles, sans la présence d'un enquêteur (p. ex. autour d'un repas de famille ou entre amis), de façon à examiner toute la palette variationnelle de locuteurs pour qui le français alterne souvent avec d'autres langues (anglais surtout, et aussi langues de migration) (Gadet, 2013).
- 12 Le choix de mener des entrevues auprès de membres des élites (« personnalités ») œuvrant dans plusieurs sphères d'activité et d'intégrer dans nos questionnaires

d'enquêtes variationnistes un module sur les attitudes linguistiques a permis de cerner au présent l'état de ces représentations linguistiques, tout en reconstituant les parcours sociobiographiques de ces personnes, leurs conditions de socialisation, leurs rapports au français et à l'anglais, ainsi que leurs pratiques dans leur domaine d'activités respectif. Ces entretiens sont riches d'enseignement ; on constate ainsi, en comparant des entretiens de personnalités de Welland 1975 à celles effectuées en 2013, que la communauté n'est plus divisée sur la question de la scolarisation des enfants en français et de la scolarisation dans les deux langues. Le principe de l'instruction totalement en français, qui n'était soutenu que par une minorité en 1975 (minorité issue principalement de l'élite), est maintenant accepté par l'ensemble des parents (Mougeon, 2012).

- 13 Le Corpus FRAN est le premier corpus librement accessible en ligne documentant le français sur plusieurs terrains en Amérique du Nord, à partir de ressources orales et textuelles et un profil sociohistorique. C'est aussi un outil précieux pour la francophonie en dehors de l'Amérique du Nord par les réponses qu'il permet d'apporter sur des enjeux partagés, comme le contact de langues, la palette variationnelle des locuteurs et la mobilité sociale et géographique. Au Corpus FRAN s'ajoutent le corpus MCVF (Martineau, 2010), qui porte sur la période médiévale au français classique, également accessible en ligne, et le Corpus LFFA (Martineau, 2005-), qui porte sur la correspondance familiale de scripteurs lettrés et peu lettrés du français classique au début du XX^e siècle, ainsi que sur des entretiens de nature ethnologique effectués en Amérique du Nord, corpus en cours de mise en ligne ; ces trois corpus réunis permettent une interrogation depuis l'ancien français jusqu'à la période contemporaine, avec stratification sociale et régionale.

2.2 Transcription, alignement texte/son et protocole

- 14 Les entretiens de personnalités et les entretiens patrimoniaux sont transcrits avec MS Word, mais nous avons opté, pour les nouvelles entretiens variationnistes et les entretiens écologiques, pour un système de transcription d'alignement texte/son avec Praat (Boersma & Weenink, 2001-) et ELAN (Wittenburg, Brugman, Russel, Klassmann & Sloetjies, 2006), en harmonie avec les développements récents en sciences du langage.

2.2.1 Les avantages de l'alignement texte/son

- 15 L'alignement texte/son à l'aide de Praat ou ELAN² présente plusieurs avantages par rapport à la simple transcription séquentielle dans un programme de traitement de texte, surtout en ce qui a trait à l'efficacité de la production des transcriptions, à l'analyse ultérieure des textes et aux liens directs que l'on peut établir entre le sonore et l'écrit.
- 16 En effet, l'alignement texte/son favorise la production rapide et efficace de la transcription d'enregistrements, en permettant la segmentation en intervalles de quelques énoncés tout au plus, sans que cela affecte pour autant la piste sonore originale. Le travail du transcripateur/vérificateur est facilité du fait qu'il peut réécouter un intervalle en boucle, tout en se concentrant sur une quantité de texte très restreinte. L'annotation à niveaux multiples (*multi-tier annotation*), c'est-à-dire la fonction permettant la superposition de plusieurs niveaux de texte, accélère davantage le processus de transcription en ce qu'elle permet d'utiliser une seule piste de texte par participant, ce qui est particulièrement utile dans le cas d'entretiens écologiques où il y a présence de multiples participants. De plus, l'annotation à niveaux multiples peut être

mise à contribution afin de promouvoir la communication entre les différents acteurs intervenant dans la création d'une transcription. Par exemple, un transcripteur qui aurait confirmé l'orthographe particulière d'un nom propre peut indiquer, sur une piste dédiée à cet effet, la source d'information qu'il a utilisée. Finalement, la contiguïté de l'affichage visuel de la piste sonore avec sa transcription peut éventuellement contribuer à désambiguïser un son ou des paroles, quand le transcripteur possède des aptitudes lui permettant d'analyser les oscillogrammes et les spectrogrammes.

- 17 En sus de l'exécution expédiente des transcriptions, l'annotation à niveaux multiples permet de préparer le document pour les analyses linguistiques. Notamment, les niveaux d'annotation peuvent être employés pour le repérage de variables ou d'unités linguistiques (par exemple, des anglicismes, des formes verbales ou des phénomènes, comme la liaison) ou encore pour préparer le texte à l'indexation morphosyntaxique ou lexicale.
- 18 Finalement, l'alignement du texte et du son permet d'établir des marqueurs temporels dans la transcription, ce qui peut servir, d'une part, à documenter les phénomènes extralinguistiques ayant une portée sur l'enregistrement ou la transcription et, d'autre part, à l'ajout éventuel d'un lecteur audio intégré à une page web qui permettrait d'entendre des sections d'enregistrement.

2.2.2 Les protocoles de transcription : pour qui et pour quoi ?

2.2.2.1 Le protocole de transcription des enregistrements

- 19 À l'instar des protocoles de transcription du VALIBEL³ (Dister, Francard, Feron, Giroul, Hambye, Simon & Wilmet, 2006) et du PFC (Durand, Laks & Lyche, 2009), le protocole élaboré par le projet (Martineau, 2011-) adopte une orthographe traditionnelle, sans « aménagement graphique »⁴ qui ferait de la transcription qu'elle ne tiendrait « ni tout à fait de l'oral, ni tout à fait de l'écrit » (Dister & Simon, 2007).
- 20 Le protocole du projet a été conçu de façon à promouvoir une interrogation rapide et ergonomique du corpus. Il a comme visée de répondre aux besoins d'utilisateurs intéressés à la morphosyntaxe, au lexique et, de façon plus générale, aux variétés de français nord-américaines. Les seuls éléments phonétiques pris en considération lors de la transcription sont les liaisons non standard (p. ex. : Je suis Tun homme riche) et la prononciation des consonnes finales (p. ex. : bouT, faiT, genS), que nous considérons comme des phénomènes morphophonétiques lexicaux. Autrement, la représentativité morphosyntaxique est respectée dans son intégralité, dans le sens où les paroles des locuteurs sont reproduites fidèlement et ce, indépendamment de leur impropreté perçue par les prescriptivistes. À titre d'exemple, les formes verbales non standard comme « ils jouaient » ('ils jouaient') et « ils sondaient » ('ils étaient') sont transcrites telles quelles ; l'omission de « ne » et de « que » est respectée ; des lexèmes récurrents dans certaines variétés sont transcrits sans normalisation, par exemple, « astheure » en français laurentien (et non 'à cette heure'), « marabouse » (féminin de 'marabout', qui est invariable) en français acadien, « ployer » (pour 'plier') en français cajun.
- 21 Les protocoles de transcription énumérés en début de section et celui du projet *Le français à la mesure d'un continent* ont également ceci en commun qu'ils incluent des mesures pour noter les amorces, les répétitions, les pauses, et les chevauchements, quoique la notation soit différente. Ils divergent cependant en ce qui a trait à la démarcation des énoncés pour former ce qu'on appelle, à défaut d'un terme plus approprié, une « phrase ». Nous

avons opté pour accepter une certaine délimitation sommaire des éléments phrastiques, mais avons toutefois limité les éléments de ponctuation au point et au point d'interrogation.

- 22 Étant donné qu'un des axes de recherche du projet porte sur les variétés de français en contexte majoritaire/minoritaire et les effets linguistiques du contact entre les communautés, certains éléments susceptibles d'intéresser les usagers du Corpus FRAN ont été étiquetés directement dans le texte. C'est le cas des mots anglais et des anglicismes (ceints de guillemets français), des calques et des extensions sémantiques (ceints de chevrons) et des mots étrangers (ceints d'accolades). Il est à noter que ce repérage initial ne remplace pas la recherche exhaustive que doit effectuer le chercheur chevronné ; son utilité consiste à attirer l'attention sur ces phénomènes ainsi qu'à permettre à notre moteur de recherche, *PhiloLogic*, de dégager une liste de toutes ces occurrences en entrant l'un ou l'autre de ces codes dans le champ de recherche des mots.

2.2.2.2 Le protocole de retranscription de manuscrits

- 23 La retranscription des manuscrits (lettres, journaux) s'effectue selon un protocole dont la directive première est de retranscrire le texte exactement tel qu'il apparaît. Ainsi, l'orthographe, même si elle dévie des normes modernes ou de l'époque, la grammaire, la ponctuation, les coquilles, l'agglutination des mots et autres phénomènes propres à l'écriture sont-ils inchangés par rapport à l'original. Ceci inclut l'utilisation du caractère < l > (s long) et des majuscules considérées « inopportunes » selon les standards actuels. Les caractères raturés, soulignés ou supérieurs sont également représentés tels quels sur l'écran.
- 24 Quant aux éléments hors teneur de la lettre, dont sa mise en page, ils ne sont pas représentés à l'écran par *PhiloLogic*. Les alinéas, les marques de sceau, l'écriture perpendiculaire dans la marge et autres éléments similaires sont toutefois attestés par le biais d'une note du transcripteur placée entre crochets.

2.3 L'interface PhiloLogic

- 25 Afin de rendre accessible le corpus FRAN à la communauté internationale, nous avons implanté sur notre serveur le logiciel *PhiloLogic* et développé une interface qui tient compte de paramètres sociohistoriques⁵. Les corpus ont été anonymisés et sont accessibles par mode d'interrogation de type concordance ; les corpus des personnalités ne sont pas accessibles en ligne pour des raisons éthiques, l'anonymat des locuteurs étant difficile à préserver en raison de la nature des entrevues, mais peuvent être consultés au laboratoire *Polyphonies du français* (directrice F. Martineau), à l'Université d'Ottawa. Des corpus patrimoniaux, notamment ceux pour lesquels nous n'avons pas l'autorisation de mettre en ligne, et des bases de données sont aussi accessibles au laboratoire.

2.3.1 Liste des textes disponibles et accès au Corpus FRAN

- 26 Une liste de tous les documents disponibles pour interrogation est disponible sur le site web. Elle inclut le titre du document et sa forme (type entrevue, lettre, etc.), ainsi que le lieu et la date de sa création. L'internaute désirant avoir accès à ces documents doit faire une demande d'accès au Corpus FRAN et s'engager à respecter les conditions d'utilisation.

2.3.2 PhiloLogic et son interface

- 27 *PhiloLogic* est un moteur de recherche développé à l'Université de Chicago, dans le cadre du projet ARTFL⁶ (American Research on the Treasury of French Language) et du DLDC⁷ (Digital Library Development Center). Le projet *Le français à la mesure d'un continent* a reçu une subvention de la Fondation canadienne pour l'innovation (FCI) destinée à adapter le logiciel et son interface aux spécifications particulières que requiert l'interrogation de notre corpus.
- 28 L'interface du moteur de recherche PhiloLogic permet la recherche de mots ou d'expressions. Il est possible de raffiner la recherche en choisissant des filtres qui s'appliquent soit au document, soit à l'individu qui est à la source du document. Les critères de recherche relatifs aux documents sont par exemple ceux du sous-corpus (voir Figure 1) auquel il appartient, de son type (voir Figure 2) et de son origine géo-temporelle (voir Figure 3).

Figure 1. Filtres de recherche pour les documents : critère sous-corpus

Document

Titre :

Corpus :

Sous corpus : HOMA Martineau Moderne Variationniste ✓

Type 1 :

Type 2 :

Date de création :

Siècle :

Pays de création :

Province de création :

Ville de création :

Région de création :

- Gardner Fox Martineau Patrimonial
- Gatineau Martineau Moderne Écologique
- HOMA Blondeau Tremblay Moderne Variationniste
- HOMA Martineau Moderne Variationniste
- Louisiane Lafourche Dubois Martineau Patrimonial
- Moncton Roy Martineau Patrimonial
- Montréal Nord Blondeau Tremblay Moderne Variationniste
- Montréal-fonds_Papineau Martineau Historique Lettres
- Normandie Gadet Moderne Écologique
- Welland Mougeon Martineau Moderne Variationniste

Figure 2. Filtres de recherche pour les documents : critère type de document

Document

Titre :

Corpus :

Sous corpus :

Type 1 :

Type 2 :

Date de création :

Siècle :

Pays de création :

Province de création :

Ville de création :

Région de création :

- Entrevue variationniste ✓
- Enquête écologique
- Enquête auprès des personnalités
- Correspondance personnelle
- Correspondance publique ou semi-publique
- Journaux personnels
- Presse
- Chroniques de langue

Figure 3. Filtres de recherche pour les documents : critère province de création

Document

Titre :

Corpus :

Sous corpus :

Type 1 :

Type 2 :

Date de création :

Siècle :

Pays de création :

Province de création :

Ville de création :

Région de création :

Source publiée :

Source archivistique :

✓ Louisiane

Massachusetts

Nouveau-Brunswick

Ontario

Québec

Seine-Maritime

- 29 Les critères de recherche relatifs aux locuteurs et aux scripteurs, quant à eux, sont de nature socio-économique et permettent, en sus des critères biographiques de base comme le sexe et le lieu et la date de naissance, l'ajout de filtres comme l'occupation, les langues parlées, lues et écrites, ainsi que le niveau d'éducation le plus élevé. Il est possible de jumeler certains critères, par exemple l'occupation et le niveau d'éducation, afin de situer socialement les documents issus des locuteurs/ scripteurs que l'on veut interroger.

2.3.3 La recherche avec le moteur PhiloLogic

- 30 Il est possible d'effectuer la recherche d'un mot ou d'une expression avec ou sans filtre. La fonction *recherche de mots similaires* permet de repérer à la fois les formes standards des mots et celles dont l'orthographe n'est pas moderne (p. ex. : « françois » pour 'français') ou non régulière (p. ex. dans le cas des écrits des peu lettrés) (voir Figures 4 et 5).

Figure 4. Recherche de(s) mot(s) : Recherche de mots similaires

Recherche de(s) mot(s)

Affichage: Contexte KWIC Recherche de mots similaires

Figure 5. Recherche de mots similaires à « français » : résultats

5	<input type="checkbox"/>	façons
1	<input type="checkbox"/>	francois
2	<input type="checkbox"/>	français
2	<input type="checkbox"/>	franç
553	<input type="checkbox"/>	français
23	<input type="checkbox"/>	française
10	<input type="checkbox"/>	françaises
2	<input type="checkbox"/>	françois
3	<input type="checkbox"/>	reçois
1	<input type="checkbox"/>	canadien-français
1	<input type="checkbox"/>	canadiens-français
54	<input type="checkbox"/>	français
1	<input type="checkbox"/>	française
1	<input type="checkbox"/>	françois
3	<input type="checkbox"/>	saint-françois
1	<input type="checkbox"/>	saint-françois-xavier

- 31 De plus, différents opérateurs de recherche sont disponibles afin d'élargir le champ de recherche ou de le restreindre. Par exemple, si on cherche toutes les déclinaisons d'un verbe, comme le verbe « aimer », on écrira « aim* » dans le champ de recherche pour obtenir « aime, aimes, aimons, aimez, aiment, etc. » (voir Figure 6). Cette recherche ne remplace pas une réelle lemmatisation (qui permettrait par exemple de trouver toutes les formes irrégulières du verbe « aller ») mais permet néanmoins de ratisser assez large.

Figure 6. Trois premiers résultats de la recherche avec l'opérateur * (aim*)

1. Papmt1_PapVi18130302_VR_MA-F (bib:p.1) et croyions aimons [Les deux mots sont superposés] nous, j'e
 2. Papmt1_PapVi18160212_VR_MA-F (bib:p.1)lait pas, on aime quelque fois la folitude, pour se delasser
 3. Papmt1_PapVi18160212_VR_MA-F (bib:p.2)que dieu fut aimé, fervi et adoré, dans tous les Coins de L

- 32 Une liste de ces opérateurs (caractères *wildcard*) se trouve en fenêtre contextuelle dans l'interface. Il existe également une fenêtre contextuelle pour afficher une version abrégée du protocole de transcription.

2.3.4 Les résultats de recherche

- 33 L'interface de recherche permet l'affichage des résultats de deux manières. Par défaut, on obtient l'occurrence recherchée en concordance, c'est-à-dire accompagnée des lignes de textes immédiatement adjacentes. *PhiloLogic* fournit aussi la référence du texte et, au bas de la page, la bibliographie (voir Figure 7).

Figure 7. Affichage des résultats en mode concordance

1. 1804-8-9. Marie-Rosalie Cherrier écrit à son fils Denis-Benjamin à Montréal, en 1804. [Page 2 |]

de toille de rusci chaquun 1 ou 2 perre
de Culot Corde rois un veux Capo pour Rouler
Et votre meilleur pour Le dimanche avec une
veste Et une **culotte** propre ayez bien foin
de votre butin afin de nerien perdre
je vous Envois par votre papa Cequill
vous faut Enparti Et votre papa
arangeras vous pour voiras dureste

2. 1804-8-9. Marie-Rosalie Cherrier écrit à son fils Denis-Benjamin à Montréal, en 1804. [Page 3 |]

deplus faits mois dans tes vacance un
memoire de Ce quil te faudres pour ton hivers
apar, Capo, Casque, finture, pe de **culotte**
de Cor de rois qui Est Che m'r Emon Ele reste
je vous Lan vairez dici que Ce memoire
dise aussi Ce

Bibliographie de résultats

[Cherrier, Marie-Rosalie] [Corpus FRAN (Le français à la mesure d'un continent, dir. F. Martineau, CRSH/GTRC)] [Montréal-fonds_Papineau Martineau Historique Lettres]FOCMTL-H-LET_Paprim_CheMR18040809_VR_MA_F_phioc [nombre de mots]

- 34 Il est aussi possible de faire une requête pour les résultats en affichage KWIC (*Key Word in Context*), comme démontré plus haut dans la Figure 6.

3. Variation et changement : deux études de cas

3.1 La variable du futur périphrastique à Hochelaga-Maisonneuve, Montréal⁸

- 35 Les locuteurs francophones de l'est de Montréal, majoritaires depuis l'urbanisation du territoire à la fin du XIX^e siècle (Linteau, 2012), ont créé un milieu qui s'est transformé au fil du temps en une zone de mixité sociale où les pratiques culturelles et langagières s'entremêlent, et les contacts avec les locuteurs anglophones ne sont pas absents. Nous présentons les premiers résultats d'une étude qui s'est intéressée aux trajectoires géographiques, familiales et sociales d'un groupe de douze locuteurs âgés de Hochelaga-Maisonneuve (HOMA), quartier ouvrier de l'est de Montréal, qui a connu dans la dernière décennie une gentrification (Martineau, Blondeau & Frenette, 2014 ; Gadet & Martineau, 2014). La variable de la 1^{re} personne de l'auxiliaire *aller* employé comme auxiliaire du futur périphrastique (*je vais/ je vas/m'as partir*) est un bon analyseur sociolinguistique de la variation diastratique et diatopique (voir entre autres Martineau & Mougeon, 2005 ; Martineau, 2009, 2012). En français canadien, trois variantes coexistent pour marquer la première personne du futur périphrastique : *je vais*, associé au français soutenu (1a), *je vas*, associé à un emploi familier et même neutre (1b), et *m'as*, associé au français populaire (1c). En français continental, la variante *m'as* est pour ainsi dire inexistante, ce qui a pour effet que la variante *je vas* est généralement associée à un français familier ou populaire/régional⁹.

(1) a. *Je vais* partir.

b. *Je vas* partir.

c. *M'as* partir.

- 36 À partir de l'analyse de cette variable, en particulier de l'emploi de *m'as*, nous montrons comment les réseaux qu'ont tissés les locuteurs permettent de comprendre les continuités/ ruptures linguistiques et les alliances communicationnelles entre locuteurs dans un quartier dit culturellement homogène.

- 37 La Figure 8 montre la fréquence d'emploi des trois variantes chez les locuteurs âgés interrogés.

Figure 8. Fréquence d'emploi de trois variantes de la première personne du singulier du verbe *aller*, auxiliaire du futur périphrastique chez douze locuteurs âgés d'Hochelaga-Maisonneuve à Montréal

<i>Je vais</i>	<i>Je vas</i>	<i>M'as</i>	Total
18,8 % (39/207)	70 % (145/207)	11,2 % (23/207)	100 % (207/207)

- 38 Comme le montre la Figure 8, les locuteurs âgés que nous avons interrogés affichent une fréquence élevée de l'emploi de la variable *je vas* (70 %) ; rien d'étonnant donc puisqu'au Québec – et plus généralement dans le parler laurentien – c'est la variante par défaut à l'oral. La variante *m'as* est la moins fréquente, et ce, même dans un quartier populaire. Ces données, lorsqu'on les compare à des locuteurs nés à la fin du XIX^e siècle, montrent que la variante *m'as* a subi un déclin progressif (*je vais* 0,3 % ; *je vas* 60,1 % ; *m'as* 39,6 %, Martineau, 2014). Cette tendance vers une régression de l'emploi de la variante la plus associée au parler populaire est confirmée par l'étude de Sankoff & Thibault (2011), qui comparent le corpus de Montréal 1971 à celui de 1984. Dans le premier, le pourcentage d'emploi de *m'as* est de 30,7 % (177/576), alors que dans le second, il n'est plus que de 15,9 % (217/1368). Dans leur corpus recueilli en Estrie à Stanstead en 2001, les auteures montrent que le pourcentage est encore plus bas (12,2 % 19/156), résultat qui se rapproche des chiffres révélés par la Figure 8. Lorsqu'on ne considère que le groupe des jeunes, comme l'ont fait Martineau, Blondeau & Frenette (2014) à partir du corpus d'Hochelaga-Maisonneuve, l'emploi de *m'as* est encore plus bas, soit à peine 1 % des occurrences, sans doute en partie en raison d'un effet de la gentrification du quartier (voir aussi, pour des résultats similaires chez les jeunes, Martineau & Dumouchel-Trudeau, 2013, pour la ville de Gatineau, située au Québec à la frontière avec Ottawa, en Ontario). Ces résultats doivent évidemment tenir compte d'autres facteurs, comme la classe sociale ou le sexe, mais il n'empêche que cette variante *m'as* semble avoir connu une régression depuis le début du XX^e siècle, et ce, même dans des quartiers à caractère ouvrier comme Hochelaga-Maisonneuve, comme le démontrent nos résultats à la Figure 8. Mais ce tableau cache une grande hétérogénéité entre locuteurs qui ne peut se comprendre que par une étude du parcours sociobiographique des locuteurs et de leur position dans le réseau social de leur communauté. Nous nous arrêterons sur trois locuteurs, habitant à quelques rues les uns des autres, mais dont les réseaux, les parcours et les aspirations sont bien différents.
- 39 Le premier locuteur, Gérard, est retraité et a été contremaître dans une usine située dans le quartier. Il a vécu sa jeunesse dans un quartier ouvrier en périphérie de Montréal, puis a déménagé dans le quartier Hochelaga-Maisonneuve lorsqu'il s'est marié. En cela, il correspond au profil ouvrier de Hochelaga-Maisonneuve, bien que le poste qu'il a occupé ait impliqué des responsabilités importantes. La Figure 9 montre les résultats pour la variable à l'étude.

Figure 9. Fréquence d'emploi de trois variantes de la première personne du singulier du verbe *aller*, auxiliaire du futur périphrastique chez le locuteur Gérard, d'Hochelaga-Maisonneuve

<i>Je vais</i>	<i>Je vas</i>	<i>M'as</i>	Total
7,4 % (2/27)	77,7 % (21/27)	14,9 % (4/27)	100 % (27/27)

- 40 Les usages linguistiques de Gérard montrent un emploi fréquent de la variante neutre *je vas*, en accord avec les fréquences moyennes du quartier. Ce serait donc le locuteur type du quartier, ce qui est appuyé par le fait qu'il est d'ailleurs très bien réseauté dans le quartier et membre de plusieurs organismes bénévoles. En accord avec ce milieu ouvrier qui connaît des racines militantes à la cause souverainiste et à la protection du français dans la sphère publique, Gérard est sensible au statut du français au Québec. Il dira ainsi : « On appelle ça une shop en anglais, mais c'est modèlerie en français ».
- 41 Le second locuteur, Jacques, connaît Gérard et habite à quelques rues de chez lui. Mais, contrairement à Gérard, pour Jacques, son usage le plus fréquent est *je vais*, en rupture avec les 'normes' de son quartier, comme le montre la Figure 10.

Figure 10. Fréquence d'emploi de trois variantes de la première personne du singulier du verbe *aller*, auxiliaire du futur périphrastique chez le locuteur Jacques, d'Hochelaga-Maisonneuve

<i>Je vais</i>	<i>Je vas</i>	<i>M'as</i>	Total
53,3 % (8/15)	46,7 % (7/15)	0 % (0/15)	100 % (15/15)

- 42 La position sociale de Jacques est assez semblable à celle de Gérard, en ce sens qu'il a occupé au cours de sa vie un poste à responsabilité, celui de commis pharmacien. Il n'appartient toutefois pas au même monde ouvrier que Gérard. Même s'il a vécu, à partir de sa prime enfance, dans le quartier de Hochelaga-Maisonneuve, contrairement à Gérard, qui s'identifie à l'élément ouvrier et œuvre dans des groupes communautaires, Jacques est très conscient de sa position sociale. Il se dit pharmacien alors que, même s'il a sans doute travaillé très fort à la pharmacie, il n'en a pas le titre et a été commis à la pharmacie. Il a le sentiment d'être le gardien du passé glorieux du quartier et réagit fortement lorsqu'il y a migration de quartiers plus pauvres vers son quartier.

(2) Pis d'autant plus on euh je me souviens parce qu'à partir des années soixante et dix euh moi j'avais pas vu ça dans le quartier ici en pharmacie des poux. <I1 : Ah oui.> Mais c'est drôle quand même c'était une coïncidence peut-être mais on a vu une épidémie de poux arriver en même temps que ces euh (phrase non terminée) <I1 : Oui oui oui.> Ça c'est une affaire qui m'avait marqué à l'époque on en parlait souvent. L'épidémie de poux là quand ça a commencé dans ces <I1 : Oui oui oui.> on appelait ça euh oui ça a changé de façon radicale le quartier.

- 43 On ne peut pas associer Jacques à la nouvelle gentrification du quartier. C'est essentiellement un petit ouvrier dont les aspirations sont à la fois celles de son quartier, auquel il est très attaché, mais qu'il idéalise. Comme Gérard, il présente une très grande sensibilité au statut du français dans la métropole montréalaise et il exige de se faire servir en français, partout où il va.

- 44 Contrairement à Gérard et à Jacques, le troisième locuteur, Lucien, n'a pas de racines aussi profondes dans le quartier. C'est un enfant de migrant. Son père est né en France, à Carnières, département du Nord de la France et arrive à Montréal à l'âge adulte au début du XX^e siècle, peut-être en rupture sociale avec une famille de notaires bien établie dont il était issu. Sa mère est de Québec. Lucien a habité Montréal toute sa vie, surtout dans l'est ouvrier, mais il n'a pas vécu toute sa vie dans Hochelaga-Maisonneuve. Il représente le groupe d'invasion d'habitants des quartiers pauvres vers Hochelaga-Maisonneuve qu'a décrit Jacques.
- 45 Dans son entrevue, Lucien marque clairement sa rupture avec le passé français de son père, comme l'illustre le passage suivant :
- (3) Pis euh les « chum »s à mon père ils venaient chez nous pis ils me disaient « Comment ça fait tu parles pas français ? » J'ai dit « Moi si je parle français icitte m'as me faire tuer. » Parce que les/ les/ les gars je me tiens avec là c'est pas des/ c'est pas des gars qui/ c'est des gars qui sacraient pis des gars qui parlaient mal pis en tout cas. Pis si je parle français ils vont prend/ me prendre pour une tapette.
- 46 Lucien ne participe pas activement aux activités du quartier. C'est un marginal, côtoyant parfois la violence. Son emploi de la variante *m'as*, comme le montre la Figure 11, beaucoup plus élevé que dans le quartier Hochelaga-Maisonneuve, ou qu'au Québec, reflète cette marginalité, sa rupture avec le passé français de son père, mais aussi sa rupture avec le quartier Hochelaga-Maisonneuve.

Figure 11. Fréquence d'emploi de trois variantes de la première personne du singulier du futur périphrastique chez le locuteur Lucien, d'Hochelaga-Maisonneuve

<i>Je vais</i>	<i>Je vas</i>	<i>M'as</i>	Total
0 % (0/13)	15,3 % (2/13)	84,7 % (11/13)	100 % (13/13)

- 47 Cette étude de la variable de la première personne du singulier du futur périphrastique, en particulier de *m'as*, montre qu'en dehors des questions de classes sociales ou de territoires (que ce soit quartier, ville ou région), il est nécessaire de tenir compte des réseaux, auxquels se sentent liés ou pas les individus. Dans Hochelaga-Maisonneuve, l'emploi de *m'as* correspond à ce que l'on trouve généralement ailleurs dans les parlers laurentiens, mais lorsqu'on examine de plus près les données, il devient évident que des locuteurs habitant la même rue, qui ont le même âge, qui sont issus de mêmes conditions sociales, et qui se côtoient sans doute dans les dépanneurs du coin, peuvent avoir des usages très différents selon leur parcours de vie et leurs réseaux sociaux. De même, les jeunes, qu'ils soient de Montréal ou de Gatineau, ont plus en commun du fait d'appartenir à la même génération, ouverte à une certaine mondialisation et à la standardisation du français, ce qui explique à la fois leur fréquence très basse d'un trait comme *m'as*, non seulement associé au parler populaire mais aussi au parler laurentien, et leur intégration de termes anglais, même lorsqu'ils vivent dans des contextes où le français est majoritaire et a un statut officiel, comme c'est le cas au Québec.

3.2 La variable de conséquence (so, ça fait que, donc, alors) : français en contexte minoritaire et majoritaire

- 48 Le Corpus FRAN, en plus de permettre des études sur des facteurs externes comme le sexe, l'âge, la classe professionnelle, l'éducation, permet également de comparer des communautés en contexte minoritaire et majoritaire, dans des situations linguistiques à différents pôles de la restriction linguistique. Par exemple, à Welland et à Moncton, le français est en contexte minoritaire. Mais tandis qu'à Moncton il se trouve dans une province officiellement bilingue (le Nouveau-Brunswick), ce n'est pas le cas de Welland, qui se trouve dans une province officiellement unilingue anglophone (l'Ontario). Dans une métropole comme Montréal, le français est en contexte majoritaire, autant dans des quartiers à forte densité francophone comme Hochelaga-Maisonneuve que dans des quartiers multiculturels comme Montréal-Nord, où il coexiste avec des langues issues de l'immigration ; à Gatineau (Québec), ville frontalière avec Ottawa, en Ontario, le français est en contact étroit avec l'anglais. Aux États-Unis, comme en Louisiane ou en Nouvelle-Angleterre, il est en étiolement, et parlé presque exclusivement par la génération la plus âgée.
- 49 On peut faire l'hypothèse que ces contextes variés auront une influence sur les usages, qu'ils soient associés au français canadien traditionnel, ou au contact étroit avec l'anglais. La variable de la conséquence, en emploi interphrastique, est particulièrement intéressante, car elle permet de tester notamment la présence d'innovations comme *so* (4a), dans des communautés en contact étroit avec l'anglais, innovations qui pourraient être absentes en français québécois, ainsi que la présence de variantes vernaculaires comme *(ça) fait que* (4b) ou de variantes plus formelles attestées en québécois (*donc* et *alors*, 4c, d) (les exemples sont tirés de Mougeon, Nadasdi & Rehner, 2009).
- (4) a. je suis pas une jaseuse *so* c'est un peu difficile
 b. je trouve ça pas mal théorique à Ottawa *(ça) fait que* j'irais plutôt faire mon bac en psychologie
 c. je suis pas une personne gênée *alors* j'y réponds t'sais
 d. elle a trois mois de convalescence *donc* j'ai eu beaucoup d'ouvrage
- 50 Les études de Dessureault-Dober (1974) et Thibault & Daveluy (1989) sur le français montréalais montrent a) l'absence de la variante *so* dans cette variété de français b) l'emploi plus marqué de *(ça) fait que* dans la classe ouvrière et dans un registre informel c) l'emploi d'*alors* par les classes socialement plus élevées d) le caractère très marginal de *donc*, renvoyé au style hyperformel (voir Mougeon, Nadasdi & Rehner, 2009). La comparaison de ces résultats en contexte majoritaire avec différentes études de Mougeon (Mougeon & Béniak, 1991, Mougeon, 2006) en contexte franco-ontarien, montre que a) *(ça) fait que* demeure associé à la classe ouvrière et est un trait de locuteurs pour qui le français est d'usage courant (les locuteurs non restreints) (Mougeon & Béniak, 1991, Mougeon, 2006) b) *alors* que *so* est absent des études de Dessureault-Dober et de Thibault & Daveluy sur le français québécois, cette variante est surtout présente chez les locuteurs franco-ontariens de la classe ouvrière (Mougeon & Beniak, 1991 ; Golembeski, 1998 ; Mougeon, 2006) et chez les locuteurs semi-restreints, c'est-à-dire ceux qui font un emploi presque égal du français et de l'anglais, et les locuteurs restreints, c'est-à-dire ceux qui sont anglo-dominants (Mougeon & Beniak, 1991 ; Golembeski, 1998 ; Mougeon, 2006) c) *alors* et *donc* sont employés par les locuteurs socialement plus élevés (Mougeon & Beniak, 1991 ; Mougeon, 2006) mais aussi par des locuteurs restreints (Mougeon, 2006). Dans ce

dernier cas, l'emploi de variantes formelles s'explique par le fait que les locuteurs s'expriment généralement en français dans un contexte scolaire. Comme le notent Mougeon et collaborateurs (2009), « l'emploi de *so* par les adolescents franco-ontariens, et en particulier par les locuteurs semi-restreints et restreints, symboliserait leur identité bilingue et l'importance de l'anglais dans leur vie quotidienne. » L'étude de Mougeon et collaborateurs (2009) sur le parler d'adolescents de la communauté francophone dominante, comme Hawkesbury, montre toutefois une progression de la variante *so* et une raréfaction de la variante *alors*, deux aspects qui distinguent les usages des jeunes franco-ontariens de cette communauté de Québécois, si on s'en tient aux études faites sur les corpus montréalais de 1971 et 1984. La variable de la conséquence a également été examinée dans la variété acadienne par Wiesmath (2006), qui montre que les variantes (*ça fait que*, *so* et *alors*) sont utilisées tandis que *donc* est rare. Toutefois, dans le corpus de chiac, « *so* y est généralisé et n'alterne avec aucun des équivalents français » (p. 100).

- 51 L'intérêt d'examiner cette variable de façon panlectale dans des corpus francophones nord-américains est multiple ; a) à la fois pour situer la variante *so*, associée aux contextes où le français est minoritaire, et pour vérifier si cette variante, qui semble jouir d'un prestige voilé chez les francophones en milieu minoritaire, a progressé en contexte majoritaire, dans un contexte où la mondialisation et les échanges sont plus importants ; b) pour mesurer la compétition entre *alors* et *donc*, la seconde variante étant très formelle, et vérifier si, en contexte de standardisation, non seulement en milieu minoritaire où cet effet est lié à des locuteurs restreints ou non restreints parlant le français à l'école, mais aussi en contexte majoritaire où le français se standardise, la variante *alors* régresse au profit de *donc* ; c) mesurer l'emploi de (*ça fait que*), lié au milieu ouvrier, et sa diffusion dans les communautés en milieu minoritaire et majoritaire.
- 52 Nous avons donc relevé la variable de conséquence en contexte interphrastique dans différents sous-corpus du Corpus FRAN, à la fois en contexte majoritaire et minoritaire, de façon à dégager certaines tendances. La recherche est à l'étape exploratoire étant donné que le Corpus FRAN est encore en cours de construction et que toutes les entrevues ne sont pas recueillies ou transcrites. Les différents emplois de cette variable sont un indice de la vitalité du français vernaculaire ((*ça fait que*), des emprunts intersystémiques et de la perception qui peuvent y être associées (comme *so*), ainsi que de tendances vers la standardisation (*alors*, *donc*). Les résultats sont présentés dans la Figure 12.

Figure 12. Fréquence d'emploi de quatre variantes de la conséquence dans différentes communautés nord-américaines dans le corpus FRAN¹⁰

Terrain	<i>Ça fait que/fait que</i> (%)	<i>So</i> (%)	<i>Donc</i> (%)	<i>Alors</i> (%)	Total (%)
Minoritaire	16,06 (333/ 2074)	51,93 (1077/ 2074)	1,30 (27/ 2074)	30,71 (637/ 2074)	100 (2074)
Gardner, Massachusetts (22 locuteurs)	1,62 (11/679)	77,91 (529/ 679)	0,29 (2/679)	20,18 (137/ 679)	100 (679)

Lafourche, Louisiane (9 locuteurs)	9,89 (26/263)	73,76 (194/ 263)	2,28 (6/263)	% (37/ 263)	14,07 (263)	100 (263)
Moncton, Nouveau-Brunswick (6 locuteurs)	2,41 (2/83)	97,59 (81/83)	0,00 (0/83)	0,00 (0/83)	0,00 (83)	100 (83)
Welland, Ontario (23 locuteurs)	28,03 (294/1049)	26,02 (273/ 1049)	1,81 (19/ 1049)	44,14 (463/ 1049)	100 (1049)	100 (1049)
Majoritaire	67,72 (1114/ 1645)	0,12 (2/ 1645)	16,47 (271/ 1645)	15,68 (258/ 1645)	100 (1645)	100 (1645)
Hochelaga-Maisonneuve, Montréal, Québec (22 locuteurs)	62,04 (773/1246)	0,08 (1/ 1246)	19,58 (244/ 1246)	18,30 (228/ 1246)	100 (1246)	100 (1246)
Montréal, Québec (4 locuteurs)	98,58 (139/141)	0,00 (0/141)	1,42 (2/141)	0,00 (0/141)	100 (141)	100 (141)
Montréal-Nord, Québec (2 locuteurs)	94,81 (146/154)	0,65 (1/154)	2,60 (4/154)	1,95 (3/154)	100 (154)	100 (154)
Gatineau, Québec (10 locuteurs)	53,85 (56/104)	0,00 (0/104)	20,19 (21/ 104)	25,96 (27/ 104)	100 (104)	100 (104)

- 53 On peut, à partir de la Figure 12, faire une première distinction entre les milieux où le français est minoritaire (hors Québec) et où il est en contexte majoritaire (au Québec) ; à Montréal et à Gatineau, situés au Québec, il y a quasi-absence d'emploi de *so* dans les entrevues et la variante *ça fait que/fait que* est la plus courante ; nos locuteurs sont de tous groupes d'âge, et la présence de *ça fait que* est ainsi répartie dans tous les groupes d'âge. C'est donc dire que les locuteurs de notre corpus de Gatineau, même s'ils habitent à la frontière avec Ottawa, en Ontario, où le français est en contexte minoritaire, se comportent comme des locuteurs en contexte majoritaire ; géographiquement, ils sont plus près d'Ottawa, mais linguistiquement, ils appartiennent au couloir Gatineau-Montréal. On remarque aussi un emploi relativement important de la variante *alors* dans certaines communautés minoritaires et majoritaires (Gardner, Lafourche, Welland, Gatineau, Montréal à Hochelaga-Maisonneuve) tandis que la variante *donc* semble plutôt limitée aux communautés majoritaires. Ce résultat est à contraster avec ceux obtenus par Mougeon et collaborateurs (2009) qui montrent que dans le parler d'adolescents franco-ontariens, là où le français est majoritaire, comme à Hawkesbury, *donc* est une variante relativement peu employée (14 % en 1975 ; 8 % en 2005) tandis que *alors* est pour ainsi dire inexistant (p. 157 ; p. 162-163), données qui vont dans le même sens que les nôtres auprès de communautés majoritaires adultes. Par contre, en milieu minoritaire, *donc* serait

moins bien installé que *alors*, comme nos données révèlent pour Gardner, Lafourche et Welland ; Mougeon et ses collaborateurs (2009) montrent toutefois à Pembroke, dans leur corpus adolescents de 2005, une remontée de *donc*, peut-être due à l'effet de l'école chez de jeunes adolescents pour qui il s'agit du principal lieu de socialisation en français.

- 54 Ce que nos résultats montrent également, c'est la différence entre les milieux en contexte minoritaire. Dans deux milieux où le français est en étiolement et parlé par des locuteurs âgés, à Garner et à Lafourche, c'est la variante anglaise *so* qui domine, suivie de la variante *alors* associée à la formalité ou au code écrit. La variante la plus courante du français laurentien, *ça fait que/fait que*, est peu utilisée, bien qu'elle se maintienne à Lafourche. Ces résultats vont dans le sens des résultats des études de Mougeon et ses collaborateurs (2009), qui ont montré que *ça fait que* est un trait des locuteurs pour qui le français est d'usage courant, ce qui n'est le cas ni en Louisiane ni en Nouvelle-Angleterre. En ce sens, Moncton se distingue de la Nouvelle-Angleterre et de la Louisiane, puisque l'emploi de *so* est presque catégorique, avec un emploi marginal de *ça fait que* mais aussi un emploi inexistant de *alors* ou de *donc*, contrairement à Gardner, Welland et Lafourche, qui présentent encore un emploi variable de ces variantes. En ce sens, nos données de Moncton ressemblent à celles de Wiesmath (2006) pour le chiac où la variante *so* a supplanté toutes les autres. Welland se détache des autres communautés minoritaires examinées. En effet, contrairement à ces dernières, *so* est en forte compétition avec la variante traditionnelle (*ça*) *fait que* à Welland, qui est presque autant utilisée ; les deux variantes vernaculaires se font compétition, avec un pourcentage plus important d'emploi de *so* chez les plus jeunes (47,1 % vs 22,3 % chez les plus âgés) et de (*ça*) *fait que* chez les plus âgés (32,3 % chez les plus âgés vs 3,2 % chez les plus jeunes). Cette tendance selon laquelle l'emploi de *so* serait en progression chez les locuteurs les plus jeunes va dans le sens des données relevées pour Hawkesbury, selon les études de Mougeon et ses collaborateurs (2009). La variante *alors* est la variante la plus utilisée, contrairement aux autres communautés minoritaires où *so* domine. On peut s'interroger sur cette robustesse de *alors* à Welland tandis que cette variante est en régression dans les milieux majoritaires, mais aussi, comme l'ont montré Mougeon et ses collaborateurs (2009) pour Pembroke, dans certains milieux minoritaires ; ces résultats devraient être explorés en examinant notamment des facteurs de formalité et de statut des locuteurs en fonction de leur restriction linguistique. Il y aurait donc des configurations linguistiques différentes du français en milieu minoritaire, notamment entre des milieux comme Moncton et Welland, cette dernière communauté présentant encore une robustesse de la variante traditionnelle.
- 55 Un dernier commentaire sur le quartier multiculturel de Montréal-Nord. Les résultats laissent entrevoir la richesse des corpus provenant de locuteurs issus de l'immigration, dans les grandes métropoles francophones comme Montréal ou Paris. Les locuteurs interrogés, tous deux jeunes, présentent des traits semblables à ceux des jeunes Québécois, c'est-à-dire avec une nette préférence pour l'emploi de la variante (*ça*) *fait que/fait que*. Il faut dire que l'un des deux jeunes est une locutrice née à Montréal d'un père d'origine libanaise avec qui les liens n'ont pas été étroits alors que sa mère est d'origine québécoise. Le deuxième locuteur est sans doute plus typique du milieu multiculturel ; d'origine haïtienne, il présente vraiment un réseau d'amis québécois et haïtiens. On sait que la communauté haïtienne est très vivante à Montréal et bien intégrée à la communauté francophone montréalaise ; la fréquence élevée d'un trait associé au français laurentien comme la variante (*ça*) *fait que* signale cette intégration linguistique.

4. Conclusion

- 56 L'avancée des technologies numériques permet de croiser des données sur de grands ensembles linguistiques, de façon à faire émerger les convergences et les divergences entre les communautés et à mettre en évidence la variation linguistique. L'Amérique française est un véritable laboratoire linguistique par sa longue histoire (plus de quatre siècles), par l'étendue de son territoire dont les frontières ont été remodelées au fil du temps, par le statut du français qui y est différent selon les pays, les provinces et les communautés. Le Corpus FRAN est le premier corpus en ligne à englober le fait français nord-américain de façon aussi vaste, à la fois de façon synchronique et diachronique, en s'intéressant à la fois aux usages et aux représentations de la langue. Les résultats préliminaires présentés ici montrent comment ce corpus permet des études sur les réseaux et les communautés et dégage des pistes de réflexion sur ce qui définit une variété linguistique, au-delà des étiquettes commodes de variété acadienne ou laurentienne, de variété en contexte minoritaire ou majoritaire.

BIBLIOGRAPHIE

- Anderson B. (1996). *L'Imaginaire national. Réflexions sur l'origine et l'essor du nationalisme*. Traduction de l'anglais par Pierre Emmanuel Dauzat. Paris : La Découverte.
- Boersma P. & Weenink D. (2001-). *Praat : doing phonetics by computer* [Programme d'ordinateur]. Dernière mise à jour : 7 septembre 2015. <http://www.praat.org/>.
- Bouchard Ch. (2002). *La Langue et le nombril : histoire d'une obsession québécoise*. Montréal : Fides, coll. « Nouvelles études québécoises ».
- Boudreau A. (2009). « La construction des représentations linguistiques : le cas de l'Acadie », in F. Martineau, R. Mougeon, T. Nadasdi & M. Tremblay (éd.) *Revue canadienne de linguistique* 54, 3 : 439-459.
- Boudreau A. & Ali-Khodja M. (éd.) (2009). « Le français en milieu minoritaire », *Langage et société* 129 : 3.
- Chaudenson R., Mougeon R. & Beniak E. (1993). *Vers une approche panlectale de la variation du français*. Paris : Didier Érudition.
- Dessureault-Dober. D. (1974). *Étude sociolinguistique de (ça) fait que : « coordonnant logique » et « marqueur d'interaction »*, Thèse de doctorat, Université du Québec à Montréal.
- Dister A. & Simon A. C. (2007). « La transcription synchronisée des corpus oraux. Un aller-retour entre théorie, méthodologie et traitement informatisé », *Arena Romanistica* 1, 2 : 54-79.
- Dister A., Francard M., Geron G., Giroul V., Hambye Ph., Simon A. C. & Wilmet R. (2006). « Conventions de transcription régissant les corpus de la banque de données VALIBEL », <http://valibel.fltr.ucl.ac.be>.

- Durand J., Laks B. & Lyche Ch. (2009). « Le projet PFC : une source de données primaires structurées », in J. Durand, B. Laks et Ch. Lyche (éd.) *Phonologie, variation et accents du français*. Paris : Hermès, 19-61.
- Ernst G. (2010). « ‘qu’il n’y a orthographe ny virgule encorre moins devoielle deconsol et pleinne delacunne’: la norme des personnes peu lettrées (XVII^e et XVIII^e siècles) », in M. Iliescu, H. Siller-Runggaldier & P. Danler (éd.) *Actes du XXV^e Congrès international de linguistique et de philologie romanes*, Innsbruck 2007, vol. 3. Berlin : De Gruyter, 543-551.
- Francard, M. (éd.) (1994). *L’Insécurité linguistique dans les communautés francophones périphériques*, avec la collaboration de Geneviève Géron et Régine Wilmet, vol. 1, *Cahiers de l’Institut de linguistique de Louvain* 19, 3-4, 1993 [paru en 1994] ; vol. 2, *Cahiers de l’Institut de linguistique de Louvain* 20, 1-2.
- Gadet F. (2013). « Les lieux du style en français oral contemporain », in Sld M.-G. Boutier, P. Hadermann & M. Van Acker Helsinki (éd.), *La Variation et le changement en langue (langues romanes)*, Helsinki : Société Néophilologique, 7-20.
- Gadet F. (2003). *La Variation sociale en français*. Paris : Ophrys.
- Gadet F. & Martineau F. (2014). « Le maillage du français en Amérique du Nord, dans un cadre de francophonie », in 10^e Colloque international « Français du Canada – Français de France » L’Amérique francophone – Carrefour culturel et linguistique, Trèves, 19-22 juin 2014.
- Gadet F. & Martineau F. (2012). « Le français panfrancophone saisi à travers un maillage de réseaux », *Cahiers de linguistique, Construction des connaissances sociolinguistiques. Du terrain au positionnement théorique* 38, 2 : 63-88.
- Gellner E. (1989). *Nations et nationalisme*. Traduction de l’anglais par Bénédicte Pineau, Paris : Payot.
- Golembeski D. (1998). *French Language Maintenance in Ontario, Canada : A sociolinguistic portrait of the community of Hearst*. Thèse de doctorat, Indiana University.
- Hobsbawm E. (1992). *Nations et nationalisme depuis 1780*. Traduction de l’anglais par Dominique Peters. Paris : Gallimard.
- Linteau P.-A. (2012). « The francophone reconquest of Montreal : the early years », atelier Urban Francophone Language Practices in North America : A Comparative Perspective, *Sociolinguistics Symposium 19*, Berlin, août.
- Martineau F. (2014) « Le français des pionniers de la Saskatchewan : quelques pistes de réflexion », in S. Hallion & R. Papen (éd.) *À l’ouest des Grands Lacs : communautés francophones et variétés de français dans les Prairies et en Colombie-Britannique*. Québec : Les Presses de l’Université Laval, coll. « Voies du français », 155-188.
- Martineau F. (2012). « Les voix silencieuses de la sociolinguistique historique », *Cahiers de linguistique. Construction des connaissances sociolinguistiques. Variation et contexte social*, 38, 1 : 111-135.
- Martineau F. (dir.) (2011-). Corpus FRAN Corpus du français d’Amérique du Nord, élaboré dans le cadre du projet *Le français à la mesure d’un continent : un patrimoine en partage*. www.continent.uottawa.ca.
- Martineau F. (dir.) (2010). *Le Corpus MCVF (Modéliser le changement : les voies du français)*. www.voies.uottawa.ca.

- Martineau F. (2009). « Vers l'Ouest : les variétés laurentiennes », in L. Baronian & F. Martineau (éd.) *Le français, d'un continent à l'autre*. Québec : Presses de l'Université Laval, coll. « Voies du français », 291-325.
- Martineau F. (2007). « Variation in Canadian French Usage from the 18th to the 19th Century », *Multilingua* 26, 2-3 : 203-227.
- Martineau F. (dir.) (2005-). *Le Corpus LFFA (laboratoire de français familial ancien)*.
www.polyphonies.uottawa.ca.
- Martineau F., Blondeau H. & Frenette Y. (2014). « Francophonie montréalaise : évolution des pratiques linguistiques en contexte », *Les Métropoles francophones en temps de globalisation*, 5-7 juin 2014.
- Martineau F. (dir.) (2014) en collaboration avec M.-Cl. Séguin, A. Bertrand, J. Dumouchel-Trudeau, R. Mougeon & D. Thomas. « Protocole de transcription du projet GTRC *Le français à la mesure d'un continent : un patrimoine en partage* », version 3.0.
- Martineau F. & Dumouchel-Trudeau J. (2013). « Enquête écologique à Gatineau, le français autour de la table », Panel *Pratiques et idéologies linguistiques en Amérique du Nord : Des réalités en tension ?*, ACFAS, Québec, 7-8 mai 2013.
- Martineau F. & Mougeon R. (2005). « Vais, vas, m'as in spoken French : a diachronic and dialectal perspective », *Linguistic Symposium on Romance Languages*, Austin, février 2005.
- Mougeon R. (2012). « La communauté francophone de Welland durant les années 1970 : le début de l'érosion linguistique et culturelle ? », Communication au Centre de recherches en civilisation canadienne-française, Université d'Ottawa, 27 mars.
- Mougeon R. (2006). « Diversification du parler des adolescents franco-ontariens : le cas des conjonctions et locutions de conséquence », *Cahiers de la Société Charlevoix* 7 : 231-276.
- Mougeon R., Nadasdi T. & Rehner K. (2009). « Les conjonctions et locutions de conséquence dans le parler des adolescents franco-ontariens de Hawkesbury : variation sociostylistique et changement linguistique (1978-2005) », in F. Martineau, R. Mougeon, T. Nadasdi & M. Tremblay (éd.) *Le français d'ici : études linguistiques et sociolinguistiques sur la variation du français au Québec et en Ontario*. Toronto : GREF, 145-184.
- Mougeon R., Nadasdi T. & Rehner K. (2005). « Contact-induced linguistic innovations on the continuum of language use : The case of French in Ontario », *Bilingualism : Language and Cognition* 8, 2 : 99-115.
- Mougeon R. & Beniak É. (1991). *Linguistic Consequences of language contact and restriction : The case of French in Ontario, Canada*. Oxford : Oxford University Press.
- Poplack Sh. & Levey S. (2011). « Variabilité et changement dans les grammaires en contact », in F. Martineau & T. Nadasdi (éd.) *Le français en contact*. Québec : Presses de l'Université Laval, coll. « Voies du français », 247-280.
- Remysen W. (2010). « L'évaluation des emplois canadiens à l'aune de leurs origines françaises : le point de vue des chroniqueurs de langage », in C. LeBlanc, F. Martineau & Y. Frenette (éd.) *Vues sur les français du Canada*. Québec : Presses de l'Université Laval, coll. « Voies du français », 241-266.
- Poirier Cl. (1979). « Créoles à base française, français régionaux et français québécois : éclairages réciproques », *Revue de linguistique romane* 43 : 400-425.

Sankoff G. & Thibault P. (2011). « Sur les traces de *m'as* en français québécois de 1971 à 2001 », in F. Martineau & T. Nadasdi (dir.) *Le français en contact : Hommages à Raymond Mougeon*. Québec : Presses de l'Université Laval, coll. « Voies du français », 351-354.

Schneider E. (2002). « Investigating variation and change in written documents », in J. Chambers, P. Trudgill & N. Schilling-Estes (éd.) *The Handbook of Language Variation and Change*. Oxford/Cambridge : Blackwell, 67-96.

Thibault A. (2008). « Français des Antilles et français d'Amérique : les diatopismes de Joseph Zobel, auteur martiniquais », *Revue de linguistique romane* 72 : 115-156.

Thibault P. & Daveluy M. (1989). « Quelques traces du passage du temps dans le parler des Montréalais, 1971-1984 », *Language Variation and Change* 1, 1 : 19-45.

Van der Wal M., Rutten G. & Simons T. (2012). « Letters as loot : Confiscated letters filling major gaps in the history of Dutch », in M. Dossena & G. Del Lungo Camiciotti (éd.) *Letter Writing in Late Modern Europe*. Amsterdam : John Benjamins, 139-161.

Wiesmath R. (2006). *Le français acadien. Analyse syntaxique d'un corpus oral recueilli au Nouveau-Brunswick/Canada*. Paris : L'Harmattan.

Wittenburg P., Brugman H., Russel A., Klassmann A. & Sloetjies H. (2006). *ELAN : a Professional Framework for Multimodality Research*, Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen. [programme d'ordinateur]. Dernière mise à jour : version 4.7.3. <http://tla.mpi.nl/tools/tla-tools/elan/>.

NOTES

1. www.continent.uottawa.ca
2. Le choix du logiciel est laissé au responsable de terrain d'enquête, puisque ELAN permet la conversion des formats de transcription depuis ou vers Praat.
3. Variétés linguistiques du français en Belgique.
4. Terme utilisé par Gadet (2003) et qui désigne les mécanismes servant à représenter certains aspects phonétiques de la parole dans l'écrit, p. ex. une apostrophe correspondant à la syncope d'un schwa.
5. <http://continent.uottawa.ca/fr/corpus-et-ressources-electroniques/corpus/>
6. <https://artfl-project.uchicago.edu/> Nous tenons à remercier Mark Olsen pour son appui dans le développement de PhiloLogic pour les besoins du projet.
7. <http://dldc.lib.uchicago.edu/>
8. Cette section provient en partie de deux communications, l'une présentée à Paris au colloque *Les métropoles francophones en temps de globalisation* (Martineau, Blondeau & Frenette, 2014) et l'autre à Trêves au colloque international « Français du Canada – Français de France » *L'Amérique francophone – Carrefour culturel et linguistique*, (Gadet & Martineau, 2014).
9. Nous avons exclu de notre étude la variable de *aller*, verbe de mouvement (*je vas/je vais*), qui ne fait pas intervenir la variante *m'as*.
10. Les corpus font tous partie du Corpus FRAN (dir. F. Martineau). Ils sont constitués des sous-corpus suivants : Gardner : *Corpus Jane Smith-Cynthia Fox 2005 enrichi* (Fox, Smith & Martineau, 2013) ; Lafourche : *Corpus Sylvie Dubois de la paroisse Lafourche 1997 enrichi* (Dubois & Martineau, 2013) ; Moncton : *Corpus Marie-Marthe Roy 1976 enrichi* (Roy & Martineau, 2012) ; Welland : *Corpus Welland France Martineau-Raymond Mougeon, 2011* ; Hochelaga-Maisonnette : *Corpus Hochelaga-Maisonnette Hélène Blondeau-France Martineau-Mireille Tremblay, 2012* ; Montréal : *Corpus France Martineau, 2012 CIEL-Québec*, corpus CIEL international sous la conduite de Françoise Gadet ;

Montréal-Nord : *Corpus MOMU Hélène Blondeau-Mireille Tremblay 2013* ; Gatineau : *Corpus Martineau, 2012, CIEL-Québec*, corpus CIEL international sous la conduite de Françoise Gadet. Nos résultats sont présentés à titre indicatif ; le dépouillement présente un nombre important d'occurrences, qui permettent de brosser un premier tableau de pourcentages. Une étude de nature statistique pourrait venir valider les différences significatives entre communautés.

RÉSUMÉS

Cet article présente le Corpus FRAN, premier corpus panfrancophone en ligne sur les variétés de français nord-américaines, élaboré dans le cadre du projet international *Le français à la mesure d'un continent* (dir. F. Martineau). Il présente d'abord les grandes questions théoriques qui sous-tendent le projet et l'élaboration du Corpus FRAN, puis discute de l'architecture du Corpus FRAN ainsi que de l'interface élaborée pour son interrogation et du protocole de transcription. La configuration du Corpus FRAN, couvrant plusieurs siècles et plusieurs communautés, permet des recherches croisées qui sont susceptibles de mettre en évidence les convergences et divergences entre ces communautés et d'examiner le parcours particulier des locuteurs et scripteurs. Nous illustrons les perspectives qu'ouvre le Corpus FRAN sur la variation et le changement linguistiques par l'étude de deux traits typiques du français nord-américain : la variante *m'as* (et les variantes associées *je vas* et *je vais*) et les variantes de la conséquence *ça fait que* et *so* (et les variantes associées *alors* et *donc*).

This article aims to introduce Corpus FRAN, the first online pan-francophone corpus pertaining to North American varieties of French, developed as part of the international project *Le français à la mesure d'un continent* (F. Martineau, dir.). We begin by stating the general theoretical questions underlying the project as a whole, and the development of Corpus FRAN in particular. We continue with a discussion of the architecture of the corpus, its transcription protocols, and the user interface. Corpus FRAN was designed so as to cover several centuries and different communities. It features a mode of crossed interrogation allowing the study of converging and diverging trends within and between communities, as well as insights into the life paths of speakers and writers. We illustrate some of the research possibilities afforded by the corpus with the study of two variants that are typical of North American French: *m'as* (and the related variants *je vas* and *je vais*) and the variants of consequence *ça fait que* and *so* (and the related *alors* and *donc*).

INDEX

Keywords : corpus linguistics, digital humanities, varieties of French, sociolinguistic variation, consequence discourse markers, semi-modal “aller”

Mots-clés : linguistique de corpus, humanités numériques, variétés de français, variation sociolinguistique, la variable de conséquence, « aller » semi-auxiliaire

AUTEURS

FRANCE MARTINEAU

Université d'Ottawa

MARIE-CLAUDE SÉGUIN

Université d'Ottawa