

« *SO* » ... *FAIT QUE*... LE FRANÇAIS QUÉBÉCOIS VERNACULAIRE SE DIVERSIFIE-T-IL? UNE ÉTUDE SOCIOLINGUISTIQUE DE L'ALTERNANCE ENTRE *FAIT QUE* ET « *SO* » DANS LE CORPUS « RAPKEB21 »

Emily Brooke Leavitt
Université de Sherbrooke

1. Introduction

Le Québec représente une situation de contact des langues propice à l'étude sociolinguistique de la variation au sein du français en Amérique du Nord. Les études antérieures ont montré le caractère distinct de la variété de français québécoise tant au niveau de son évolution sociohistorique que des normes et contraintes sociolinguistiques en son sein. La présente étude aborde le problème de la diversification du français vernaculaire au Québec et à Montréal en particulier. Cette étude s'inscrit dans la tradition d'études sociolinguistiques montréalaises et dans une veine de recherches traitant des vernaculaires urbains contemporains, comme le « Multicultural Paris French » (Gadet et Guerin 2016) et le « Multicultural London English » (Cheshire et al. 2013).

2. Cadre conceptuel

L'objectif de la présente étude est de vérifier la théorie de Blondeau et Tremblay (Blondeau 2018, 2020; Blondeau et Tremblay 2016, 2022) d'une diversification en cours au sein du français québécois vernaculaire en contexte urbain à Montréal. Nous nous intéressons à la production langagière d'artistes du genre rap québécois, un genre de nature typiquement vernaculaire. Nous proposons une analyse de la variation et des contraintes sociolinguistiques opérant sur l'alternance entre deux variantes vernaculaires, (*ça*) *fait (que)*, du français, et « *so* », de l'anglais, au sein d'un corpus de paroles de chansons rap québécoises, le corpus « RapKeb21 » (Leavitt 2022).

2.1 Question de recherche

Notre question de recherche est si l'alternance entre variantes vernaculaires (*ça*) *fait (que)* et « *so* » est contrainte par des facteurs sociaux de manière à confirmer l'existence de deux pôles normatifs, l'un à Montréal et l'autre délocalisé au sein de la communauté linguistique québécoise hors de Montréal, prévue par la théorie de diversification. Cette variable a été choisie car elle n'a jamais été attestée au Québec lors d'entretiens sociolinguistiques (Blondeau et al. 2019), mais elle a récemment fait surface au sein du sous-corpus québécois francophone de textos *Texto4Science* (Blondeau et Tremblay 2022).

2.2 Hypothèses

Notre hypothèse globale est que les deux variantes vernaculaires (*ça fait (que)* et « *so* » se existent dans la production lyrique des artistes de rap québécois, mais faute de données sur leur alternance, nous ne posons pas d'hypothèse au sujet de leur ampleur relative. Toutefois, nous nous attendons à ce que (*ça fait (que)* soit plus présent chez les artistes de Montréal en raison de la hausse constatée à Montréal de sa forme phonologique /fɛk/ depuis 2012 aux dépens des autres variantes (Blondeau et al. 2019, Blondeau et al. 2021).

Si la variante « *so* » s'avère catégoriquement absente des données d'entretiens sociolinguistiques semi-dirigés menés à Montréal (Blondeau et al. 2019, Blondeau et al. 2021), elle a récemment été attestée par Blondeau et Tremblay (2022) au sein du sous-corpus québécois francophone de textos, Texto4Science. Pour reprendre les mots de ces chercheuses, « cette présence de *so*, dont l'emploi est principalement attesté à l'oral dans des communautés minoritaires bilingues comme en Ontario et en Acadie, tranche avec ce qui a été observé dans les études antérieures sur le français montréalais (Blondeau et al., 2019) » (Blondeau et Tremblay 2022 : 135). Nous estimons que c'est la nature vernaculaire de la variante « *so* » qui explique sa présence dans les textos mais aussi son absence dans les entretiens. Pour cette raison, nous posons l'hypothèse que la nature vernaculaire de la modalité de communication que représente l'expression artistique rap s'avérera être propice à la production de « *so* ». En revanche, cette hypothèse étant basée sur un constat qualitatif de la présence de « *so* » dans le corpus de textos (Blondeau et Tremblay 2022) dont on ignore la fréquence, nous attendons à découvrir si elle se présente dans notre corpus, à quelle fréquence et quels facteurs influencent sa production.

2.3 Le français vernaculaire au Québec et à Montréal

De nombreuses études récentes contribuent à décrire le français vernaculaire émergent à Montréal à travers la comparaison du sous-corpus Hochelaga-Maisonneuve 2012 (Martineau et Seguin 2016) et des corpus Montréal 1971, 1984 et 1995. Entre autres, elles démontrent effectivement de la divergence entre le vernaculaire localisé à Montréal et le vernaculaire délocalisé au Québec hors de Montréal. Certaines variantes typiquement associées au vernaculaire québécois traditionnel sont en régression au sein du français montréalais vernaculaire : la contraction de la préposition *dans* et un déterminant subséquent (Blondeau et al. 2021 : 10–11), les formes pronominales complexes (*nous-autres*, *eux-autres*, *vous-autres*) (Blondeau et al. 2022) et la variante *m'as* (1^e personne du singulier du verbe *aller* comme auxiliaire au futur périphrastique) en faveur des variantes *j'vas* et *j'vais* (toutes deux plus formelles, la seconde davantage) (Blondeau et al. 2016). À l'opposé, le futur synthétique, en recul au Québec, augmente à Montréal (Blondeau 2006).

Le multilinguisme est devenu l'un des caractéristiques les plus saillants du français montréalais vernaculaire. La situation de contact des langues particulièrement intenses à Montréal se manifeste dans la production langagière par des phénomènes comme le *codeswitching* (ou l'alternance codique) (Low et Sarkar 2012) entre le français, l'anglais et de diverses langues d'immigration telles l'arabe, le créole haïtien et l'espagnol, l'emprunt à l'anglais et son intégration au français (Poplack 2018, Friesner 2009), et le *crossing* (Rampton 1995) aux langues dont la personne n'est pas un locuteur (Sarkar 2008). Les répertoires linguistiques de Montréalais(es) étant plurilingues, les identités

linguistiques revendiquées au sein de la communauté linguistique montréalaises sont aussi plurielles, hybrides et fluides. À travers l'indexicalité, les pratiques multilingues portent du sens social et servent d'outils d'expression identitaire nuancés et reliés au réseau social, à l'appartenance à une communauté de pratique et au milieu de socialisation du locuteur.

3. Cadre théorique : le genre rap

Le rap est un genre musical qui a été inventé aux années 1970 dans les rues du Bronx, à New York, aux États-Unis, par des jeunes qui cherchaient un moyen d'échapper au sort trop souvent réservé aux jeunes marginaux, la vie des gangs de rue. À cette époque, de jeunes musiciens se sont réunis autour de la composition musicale de manière collective, partageant du matériel ainsi que leurs compétences techniques, leurs répertoires musicaux et leurs innovations stylistiques. Leur trame de fond était le rythme, qu'ils créaient à travers le découpage et le mixage de segments de chansons axés sur la percussion, sur lequel ils transposaient des paroles rapides, rimées, rythmées et rhétoriquement créatives.

La composition lyrique rap est largement destinée à faire entendre des voix autrement marginalisées à travers la représentation de réalités interdites, négligées ou sous-représentées dans d'autres contextes sociaux. Certains thèmes sont mis en évidence, notamment les injustices sociales, les relations de famille, les divers aspects de la vie de la rue, la jeunesse, l'adolescence et l'âge adulte, les perspectives d'avenir, la construction et l'expression identitaire particulièrement par la mise en relation et la négociation de relations entre stéréotypes et la vantardise des compétences en matière de rap. Dans cette optique, la représentation de la vie réelle et des réalités modernes supplante les conventions sociétales de comportements langagiers et stylistiques normatifs, ce qui signifie que toute infraction à ces normes sociétales est autorisée. Pour cette raison, le rap se caractérise par un langage rude et vernaculaire, fertile aux usages non standard tels l'argot, les jurons et la vulgarité, parmi d'autres. C'est effectivement la richesse de ces pratiques linguistiques non standard qui explique l'intérêt des chercheurs en linguistique pour l'étude du rap.

3.1 Le rap québécois

Tout comme le genre rap en général, le rap québécois jouit d'une popularité considérable. Cette popularité se reflète par l'engouement des médias concernant l'ascension fulgurante vers la célébrité de certains artistes tels Lost, Tizzo x Shreez et Yes McCan (Jbeili 2023, Leblanc 2023b), l'incarcération d'autres comme VT et White-B (Perron 2022, Bordeleau 2022), et le langage de certains groupes en particulier tels Dead Obies, Maky Lavender et Muzion (Renaud 2021). Elle est aussi palpable sur les ondes radios où sont diffusées des chansons de certains artistes tels Loud, FouKi et Koriass ainsi que sur les ondes télévisées, notamment de TéléQuébec, qui diffusent *La fin des faibles* – un concours de rappeurs en herbe dont les juges sont Koriass, Souldia et Sarahmée – et la remise annuelle des palmarès ADISQ. La popularité du rap québécois se manifeste également sur la scène par des artistes qui font des tournées comme Loud, Monk.E et Sarahmée (Agence QMI 2022, Alexy 2022) et par ceux qui jouent aux festivals musicaux, notamment Les Francos de Montréal.

3.2 L'étude sociolinguistique du rap québécois

Les premières études sociolinguistiques menées sur le rap québécois sont issues du projet « Hip Hop Language and Youth Culture in Quebec: Multilingual Identity Practices In and Out of Classrooms » de Mela Sarkar et ses collègues (Low et Sarkar 2012, 2014; Low et al. 2009; Sarkar 2008, 2009; Sarkar et Allen 2007; Sarkar et Winer 2006). Ces travaux représentent un ensemble d'analyses de discours de chansons rap québécoises et d'observation ethnographique portant sur 12 albums parus de 1997 à 2004, 10 entrevues d'artistes et six entrevues de jeunes amateurs de rap. Les résultats mettent en avant « une grande aisance à mélanger plusieurs influences linguistiques, autant dans les paroles composées (donc préalablement réfléchies et assidûment retravaillées avant de paraître sur disque) que lors de conversations spontanées » (Low et Sarkar 2012 : 31). Les variétés de langue concernées sont le français québécois, l'anglais américain, l'anglais vernaculaire Africain-Américain (AAVE), l'anglais rap mondial (un champ lexical associé au rap), le créole haïtien, le créole jamaïcain, le créole trinitadien, l'espagnol et l'arabe (Sarkar 2009).

Low et Sarkar (2012) constatent chez des adolescents du *crossing*¹ (Rampton 1995) en créoles haïtien et jamaïcain. Toutefois, chez certains artistes pour qui ces langues font véritablement partie du répertoire, il ne s'agit pas de *crossing*, plutôt d'alternance. Voici un extrait de la rappeuse J.Kyll cité par Low et Sarkar (2012 : 32)² : « Show me respect, j'suis la true ill nana / Pourquoi t'es venu, si tu *front sou kote*? / Fais pas ton **mean**, j'vois ton *bounda* sauter / Si tu sais pas danser, qu'est-ce que t'as à te moquer? / Dis-le, t'aimes mes **moves**, pas vrai, t'es choqué? / Hey! Mais qu'est-ce ta main fait là? Go away! / **Neg pa lave, pafume...** No way! » (Muzion 1999). Les auteures concluent que « la production artistique est ici un baromètre de transformation sociale [...] qui témoigne d'un changement d'attitude et du développement de nouvelles compétences qui concernent le plurilinguisme » (Low et Sarkar 2012 : 35). Ce plurilinguisme remplit, selon elles, plusieurs fonctions socioculturelles. La principale est la promotion de la diversité d'identités revendiquées surtout par la nouvelle génération de jeunes multilingues au Québec issus de la migration, la mondialisation et l'urbanisation (Sarkar et Allen 2007). Les autres comprennent la résistance, la revendication de droits linguistiques, la déstabilisation de la hiérarchie des langues dominantes et le questionnement de ce qui constitue aujourd'hui le fait d'être Québécois(e) (Sarkar et al. 2007, 2008, 2009).

Le relais a été pris par l'auteure de la présente étude qui a mené une analyse de la prosodie dans deux chansons du rap québécois (Leavitt 2021). Les résultats révèlent que les schémas d'accent et d'intonation dans le rap suivent largement les mêmes tendances que l'on constate en français québécois oral. De plus, le rap s'avère original dans son utilisation de procédés rythmiques tels le trémolo et la symétrie mélodique entre groupes rythmiques ainsi que dans la manière intentionnelle dont s'effectue l'alternance entre le français et l'anglais pour aboutir à la perturbation, la satisfaction ou l'amplification de l'effet prosodique des accents tonique, initial et expressif (Leavitt 2021).

La prochaine étude de Leavitt (2022) fut son mémoire de maîtrise, une analyse sociolinguistique variationniste de trois variables : l'alternance entre *alors, donc, (ça) fait*

¹ *Crossing* désigne le fait d'alterner à une langue dont on ne maîtrise pas mais à laquelle on est exposé(e).

² Ici, la mise en forme indique la langue source : le français québécois standard (romain maigre), le français québécois non standard (romain gras), le français européen (romain gras et souligné), l'anglais nord-américain standard (romain souligné), l'AAVE (italique maigre) et le créole haïtien (italique gras).

(*que*) et « *so* »; l’alternance entre *j’veais*, *j’vas* et *m’as*; la neutralisation de *tout* et *tous* en [tʊt]) au sein du mégacorpus de paroles de chansons rap québécoises, « RapKeb21 » (Leavitt 2022). Ses résultats montrent une divergence significative entre la production d’artistes « ayant pour modèle normatif une variété de français montréalaise distincte » qui privilégient la variante empruntée de l’anglais « *so* », et ceux « ayant pour modèle normatif le français québécois » (Leavitt 2022 : iii) qui favorisent (*ça*) *fait (que)* et la variante *j’vas* pour la deuxième variable, ces dernières étant les variantes les plus typiquement associées au français québécois traditionnel. Par ailleurs, les résultats de Leavitt (2022) révèlent une systématicité des contraintes linguistiques sur la production rap semblable à ce que l’on atteste dans les études antérieures sur les pratiques langagières (hors du rap) au Québec.

4. Méthodologie

Nous avons effectué une analyse variationniste de l’alternance entre (*ça*) *fait (que)* et « *so* » comme variable dépendante et des facteurs sociaux et linguistiques qui la conditionnent au sein du corpus « RapKeb21 » (Leavitt 2022). Celui-ci représente les œuvres complètes d’un échantillon de 80 artistes de rap québécois, dont 37 artistes solitaires et 43 artistes appartenant à l’un des 13 groupes de deux membres ou plus. Les artistes de l’échantillon ont été sélectionnés en fonction de critères d’appartenance à la communauté rap québécoise et de critères basés sur la popularité et la visibilité de l’artiste, la prééminence de l’artiste au sein du mouvement rap local et de la culture populaire québécoise, et la représentativité de la communauté québécoise plus large (pour plus de détails, voir Leavitt 2022 : 71–72).

4.1 Données

Les données du corpus correspondent à 2965 chansons, chacune représentée sous trois formes : un fichier audio *.wav*, un fichier texte *.txt* de ses paroles, et un fichier *.TextGrid* composé des horodatages qui signalent l’emplacement temporel dans l’audio de chacun des mots de paroles. Ainsi, le corpus se compose de 8895 fichiers uniques. La création de chaque type de fichier est détaillée dans Leavitt (2022 : 74). La série textuelle qui représente la version orthographique des paroles compte un total de 1 454 425 mots. Nous avons effectué une recherche par logiciel concordancier et identifié 300 occurrences de (*ça*) *fait (que)* (sous toutes ses formes, *ça fait que*, *ça fait* et *fait que*) et 927 occurrences de « *so* » au sein du corpus (N=1227). Nous avons extrait les occurrences et leur contexte précédant et suivant au sein des paroles et, ensuite, codé chacune en fonction de l’artiste l’ayant produite, de quatre facteurs sociaux (âge, genre, lieu de naissance et origine géographique) et de deux facteurs sociopsychologiques (identité linguistique et identité culturelle) concernant l’artiste, ainsi qu’en fonction de l’année de parution de la chanson et de la fonction grammaticale de l’occurrence (marqueur de conséquence ou de discours).

4.2 Variables indépendantes

Nous avons tenu compte de cinq facteurs sociaux : âge, genre, lieu de naissance et origine géographique de l’artiste et année de parution de la chanson. Nous avons aussi tenu compte

de l'artiste comme variable aléatoire afin d'isoler l'effet de la variation individuelle. Les catégories d'âge et d'année ont été codé de manière continue, de 16 à 59 pour la catégorie d'âge et de 1997 jusqu'à 2021 pour la catégorie d'année. La catégorie de genre³ tient compte d'artistes de genre masculin, de genre féminin et de genre non binaire. Les catégories de lieu de naissance et d'origine géographique correspondent soit à la ville et la province pour des lieux au Canada, soit au pays pour des lieux hors du Canada. Le lieu de naissance signifie l'endroit précis où l'artiste est né(e) et son origine géographique signifie l'endroit précis d'où sont nés ses parents. Si l'un des parents est né au Canada et l'autre hors du Canada, c'est l'origine géographique du dernier qui est prise en compte.

Nous avons tenu compte de deux facteurs sociopsychologiques : identité linguistique et identité culturelle. L'identité linguistique se divise entre « francophone », « anglophone » et « bilingue ». La catégorie d'identité culturelle renvoie aux différentes combinaisons de quatre désignations d'identité civique locale revendiquées par l'artiste : *Montréalais(e)* (M), *Québécois(e)* (Q), *Canadien(ne)* (C) et *Américain(e)* (A). Pour chacun des artistes, une recherche par concordancier a été effectuée sur l'ensemble des paroles de leurs œuvres complètes afin d'identifier le nombre de chacune des désignations en français et en anglais revendiquée par l'artiste. Nous avons retenu la désignation majoritaire pour l'identité linguistique et la combinaison de désignations pour l'identité culturelle.

Nous avons pris en compte un seul facteur interne, la fonction grammaticale, soit marqueur de conséquence (MC), soit marqueur de discours (MD). Voici un exemple des variantes « *so* » et (*ça fait (que)*) servant chacune des deux fonctions.

- (1) *All the places I've been, fait que* [MC] *mes shoes s'abîment* (Yes 2016).
- (2) *Jamais d'mic check, ça fait que* [MD] *yes, respect Bas-Canada* (Alaclair 2013).
- (3) *Jamais égalé, so* [MC] *pourquoi tu imites, huh?* (Izzy-S 2018).
- (4) *Comme le silence pour faire durer l'suspense, so...* [MD] (Yes 2016).

4.3 Analyses

À partir des dix codes attribués aux occurrences, nous avons créé une trame de données de 1227 rangées sur 10 colonnes que nous avons importé dans RStudio (RStudio 2020), un environnement de programmation en langage *R* (R Core Team 2021). En utilisant la trousse *lme4* (Bates et al. 2015), nous avons soumis les données à une modélisation par régression logistique⁴ de la variable dépendante en fonction des facteurs sociaux, sociopsychologiques et interne à effet fixe et de l'individu à effet aléatoire. Dans ce modèle, les variables catégorielles sont traitées à travers une technique de contrastes binaires dite de « traitement » entre la première sous-catégorie qui se voit jouer le rôle de référence et chacune des autres sous-catégories. La variable dépendante est traitée de manière binaire et les effets de facteurs sortis du modèle sont sur l'échelle logarithmique. Ainsi, la valeur

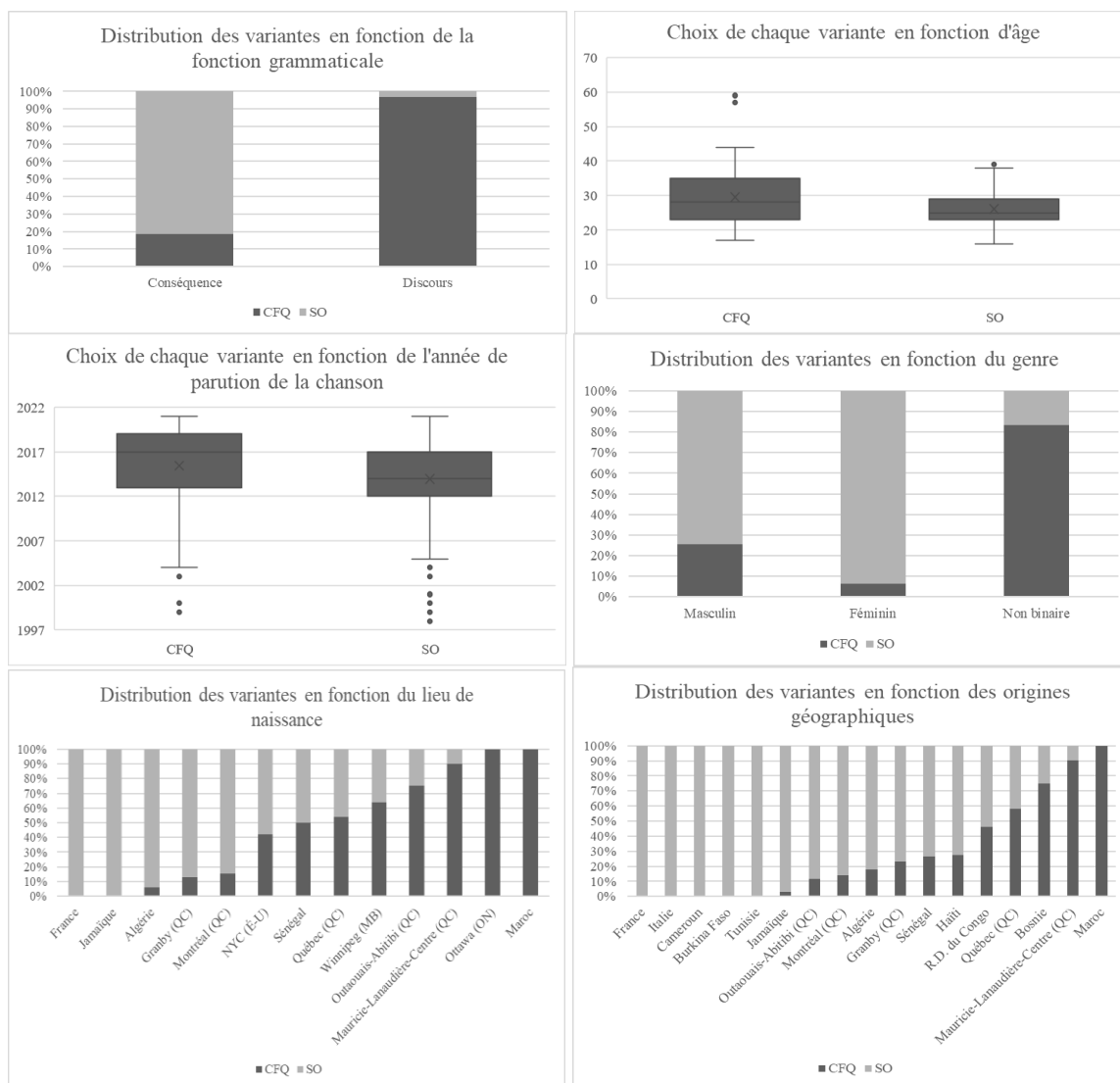
³ Le choix de traiter du genre au lieu du sexe représente une prise de distance par rapport à la tradition destinée à nous permettre de tenir compte des réalités modernes en traitant d'une catégorie basée principalement sur des inférences sociales faute d'informations biographiques disponibles.

⁴ Voici la formule ultime utilisée lors de la modélisation : `modèle=lmer(VD ~ Âge + Genre + Nais + Orig + IdLing + IdCult + Année + Fonction + (1|Artiste), data=données_rapkeb21)`.

de zéro sert de valeur pivot, c'est-à-dire que les valeurs positives indiquent une tendance vers « so » et les valeurs négatives indiquent une tendance vers (*ça*) fait (*que*).

5. Résultats

La distribution globale des occurrences (N=1227) est de 24,45% (n=300) de (*ça*) fait (*que*) et de 75,55% (n=927) de « so » (désormais 'CFQ' et 'SO'). Nous présentons ci-dessous la distribution des variantes en fonction des variables indépendantes dont nous discuterons afin de partir d'un portrait global avant de procéder aux résultats d'analyses statistiques.



Figures 1 à 6. Distribution des variantes CFQ et SO en fonction de la fonction grammaticale, de l'âge, de l'année de parution, du lieu de naissance et des origines géographiques.

D'abord, en fonction du seul facteur interne, la fonction grammaticale, les deux variantes sont distribuées de manière largement inverse. CFQ est majoritaire pour la fonction de marqueur de discours, et SO est majoritaire en tant que marqueur de conséquence, bien que ces tendances ne soient pas catégoriques.

Pour les deux facteurs sociaux de type continu, la distribution des variantes est présentée sous la forme de diagramme de quartiles. Si pour chacune des deux variables l'âge moyen des artistes l'ayant produite est dans la vingtaine, l'âge moyen des artistes ayant produit CFQ est légèrement plus élevé et l'éventail d'âges est plus important que ceux correspondant à SO. Aucun artiste d'âge supérieur à 40 ans n'a produit SO.

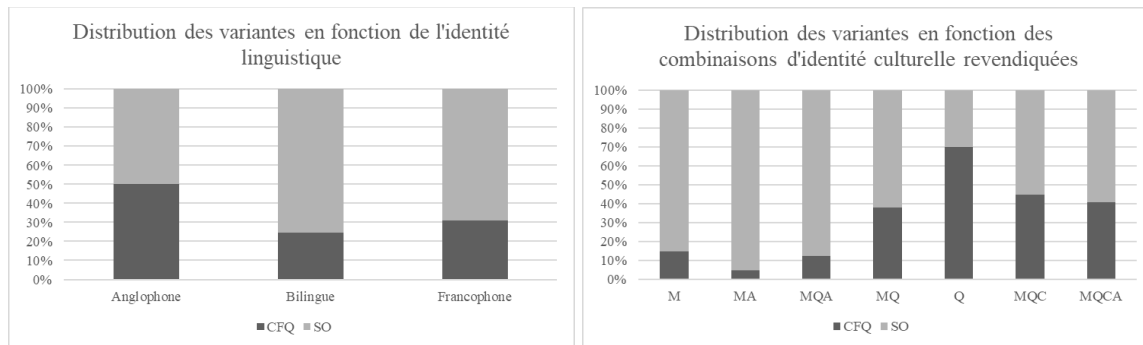
Quant à l'année de parution de la chanson, les deux variantes sont distribuées différemment. Le premier quartile des occurrences de chacune des deux variantes a été produit à travers une période relativement plus longue que celle du prochain quartile. Cette accélération est plus marquée pour la variante SO, le premier quartile ayant été produit sur une période de 7 ans et le second sur une période de 2 ans (2012-2014). À partir de 2014, l'accélération de SO s'est atténuée, pendant que la production de CFQ s'est accélérée, la moitié de ses occurrences ayant été produite entre 2017 et 2021.

Pour les trois facteurs sociaux de type catégoriel, la distribution des variantes est présentée sous la forme de diagrammes à barres empilés. Si la variante SO est majoritaire chez les artistes de genre féminin et masculin, elle l'est davantage pour ces premiers. En revanche, parmi ces artistes, c'est chez les artistes de genre masculin que CFQ est plus présent, mais il est exclusivement majoritaire chez les artistes de genre non binaire. Les distributions des variantes en fonction du lieu de naissance et de l'origine géographique présentent certaines similitudes, mais aussi des différences subtiles et plus marquées. Les utilisateurs catégoriques de SO sont nés en France ou en Jamaïque, tandis que les utilisateurs catégoriques de CFQ sont nés à Ottawa en Ontario ou au Maroc. Il existe de la variation par rapport à tous les autres lieux de naissance. Les artistes ayant une production variable mais majoritaire de la variante SO sont nés en Algérie, à Granby, à Montréal ou à la ville de New York aux États-Unis. Les artistes nés au Sénégal n'ont tendance à produire ni SO, ni CFQ. Ceux faisant preuve d'une production variable mais majoritaire de CFQ sont nés à la ville de Québec, à Winnipeg au Manitoba, dans les régions d'Outaouais-Abitibi et de Mauricie-Lanaudière-Centre du Québec, ou à Ottawa en Ontario.

En fonction de l'origine géographique, la distribution des variantes révèle plus d'utilisateurs catégoriques de « so » que dans le cas précédent. À l'instar des artistes nés en France, ceux d'origine française, mais également les artistes originaires de l'Italie, du Cameroun, du Burkina Faso et de la Tunisie, produisent exclusivement SO. À l'opposé, les seuls utilisateurs catégoriques de CFQ sont d'origines marocaine tout comme les artistes nés au Maroc. Comme les artistes nés dans les mêmes lieux, les artistes d'origine jamaïcaine, montréalaise, algérienne, granbyenne et sénégalaise font un usage majoritaire de SO. Dans ce groupe figurent ceux ayant des origines d'Outaouais-Abitibi au Québec, haïtiennes ou congolaises. En revanche, les artistes produisant de manière variable mais quand même majoritaire de CFQ ont des origines de la ville de Québec, de la Bosnie ou de la région Mauricie-Lanaudière-Centre du Québec. Il existe une grande cohérence entre ces deux facteurs sociaux qui n'est altérée, si l'on exclut les groupes qui n'ont pas d'homologue dans l'une des deux variables, que par les artistes nés au Sénégal ou en région d'Outaouais-

Abitibi et les artistes ayant des origines géographiques sénégalaises ou de région d'Outaouais-Abitibi. Ceux nés au Sénégal n'ont aucune préférence entre les deux variantes, tandis que ceux d'origines sénégalaises produisent plus de SO que de CFQ. Pendant que ceux nés en région d'Outaouais-Abitibi produisent plus de CFQ que de SO, ceux ayant que des origines de région Outaouais-Abitibi produisent davantage de SO.

Les distributions en fonction des facteurs sociopsychologiques, l'identité linguistique et l'identité culturelle, sont présentées sous forme de diagramme à barres empilés. Ces données ne renvoient pas à l'échantillon au complet (N=1084 et N=1179, respectivement), certains artistes n'ayant revendiqué aucune des désignations recherchées. Il est très intéressant à constater que les artistes d'identité *anglophone* présentent des proportions égales des deux variantes, pendant que les artistes d'identité *francophone* ou *bilingue* produisent davantage de la variante SO de l'anglais que de la variante CFQ du français. Cette dernière tendance est légèrement plus marquée pour les artistes *bilingues*.



Figures 7 et 8. Distribution de CFQ, SO en fonction de l'identité linguistique et culturelle.

La distribution des variantes en fonction de l'identité culturelle s'avère être plus complexes en raison du choix de tenir compte des différentes combinaisons de désignations *Montréalais(e)* (M), *Québécois(e)* (Q), *Canadien(ne)* (C) et *Américain(e)* (A) (ainsi que leurs équivalents en anglais). Le seul groupe qui fait un usage majoritaire de CFQ comprend les artistes qui revendiquent exclusivement une identité québécoise. À l'opposé, ceux revendiquant une identité exclusivement montréalaise produisent SO comme variante majoritaire. Hormis ce groupe, les seuls à ne revendiquer d'identité partiellement québécoise se disent *Montréalais(es)* et *Américain(es)* et sont les utilisateurs les plus exclusifs de SO. Les autres combinaisons comprennent toutes une dimension montréalaise et québécoise. Les artistes qui en rajoutent une dimension américaine produisent davantage de SO que ceux se disant *Montréalais(es)* et *Québécois(es)*. Ceux qui en rajoutent une dimension canadienne produisent relativement plus de CFQ que ces derniers, autant plus pour ceux qui n'en rajoutent pas aussi une dimension américaine.

Tournons vers les résultats statistiques d'analyse des effets de variables indépendantes obtenus par régression logistique à effets mixtes. Rappelons que le modèle tient compte des effets fixes des variables d'âge, genre, lieu de naissance, origines géographiques, année de parution, identité linguistique, identité culturelle et fonction grammaticale, ainsi que de l'effet aléatoire de l'individu. La valeur des variantes SO et

CFQ a été définie à '1' et à '-1', respectivement. La variance expliquée par le modèle est de 0,486. La statistique sommaire du modèle est de : $F(44, 1182) = 25.42$, $p < 2,2e-16$. Le tableau ci-dessous présente la valeur estimée des coefficients pour l'effet des catégories d'âge, année, fonction, genre, identité linguistique et identité culturelle.

Tableau 1. Résultats de régression logistique – facteurs conditionnant SO (1) et CFQ (-1).

	Valeur	Écart-type	Z	p	
Ordonnée à l'origine (« intercept »)	-5.006	6.904	-0.725	0.469	
Âge	-0.002	0.003	-0.531	0.596	
Année	0.003	0.003	0.877	0.381	
Fonction - Marqueur de discours	-0.747	0.040	-18.719	<2E-16	***
Genre - Masculin	-0.169	0.070	-2.423	0.016	*
Genre - Non binaire	-0.939	0.223	-4.216	2.68E-5	***
Identité linguistique - Bilingue	-0.029	0.346	-0.083	0.934	
Identité linguistique - Francophone	-0.003	0.352	-0.008	0.993	
Identité culturelle - MA	0.142	0.093	1.528	0.127	
Identité culturelle - MQ	0.109	0.105	1.037	0.300	
Identité culturelle - MQA	0.156	0.063	2.475	0.013	*
Identité culturelle - MQC	0.010	0.069	0.148	0.882	
Identité culturelle - MQCA	0.151	0.078	1.945	0.052	.
Identité culturelle - Q	0.033	0.122	0.273	0.785	

Nous constatons tout d'abord que les effets des variables continues d'âge et d'année ainsi que celui de la variable catégorielle d'identité linguistique ne se sont pas avérés statistiquement significatifs. Selon nos analyses, l'effet le plus important sur l'alternance entre SO et CFQ est celui de la fonction grammaticale. Le coefficient de la fonction de marqueur de discours tend plus vers le négatif que celle de l'ordonnée à l'origine qui englobe la fonction de marqueur de conséquence, ce qui signifie un effet favorable vers la variante CFQ en tant que marqueur de discours. Cette différence est hautement significative ($p < 2E^{-16}$). Le prochain effet le plus important est celui de la catégorie de genre. Le coefficient de la sous-catégorie non binaire tend le plus vers le négatif et celui de la catégorie masculine tend aussi plus vers le négatif, bien que de manière moins prononcée, par rapport au coefficient de la sous-catégorie féminine représentée à l'origine. Ces différences sont significatives, la première d'autant plus ($p = 2.68E^{-5}$) que la seconde ($p = 0.016$). Le dernier effet présenté dans ce tableau est celui de l'identité culturelle. Quant à cet effet, seul deux sous-catégories font preuve d'une différence significative par rapport aux autres. Ce sont la combinaison de dimensions identitaires montréalaise, québécoises et américaines qui produit le coefficient le plus élevé, ce qui indique un effet favorable vers la variante SO. Le prochain coefficient le plus élevé est celui de la combinaison des quatre dimensions montréalaise, québécoise, canadienne et américaine qui montre aussi un effet favorable vers la variante SO. Si l'on compare ces deux combinaisons (MQA et MQCA), nous constatons que la différence la plus remarquable entre elles et les autres combinaisons

de désignations identitaires est la présence de la dimension américaine. Bien que ce ne surpasse pas le seuil de signification, l'autre combinaison comprenant la dimension américaine (MA), a le troisième coefficient le plus important.

Les derniers résultats dont nous discutons sont ceux correspondant aux facteurs sociaux, lieu de naissance et origines géographiques. Nous avons exclu du tableau suivant les sous-catégories dont la différence par rapport aux autres n'est pas significative, sauf celles de Montréal et de Québec, pour faciliter la comparaison. Par rapport au lieu de naissance, nous constatons cinq différences significatives au sein des données. Il s'agit dans chaque cas d'une tendance vers le négatif qui signalent une préférence pour la variante CFQ. Cet effet est le plus prononcé chez les artistes nés en région d'Outaouais-Abitibi, suivis de près par les artistes nés au Maroc, à Ottawa et en région de Mauricie-Lanaudière-Centre du Québec. Les artistes nés à Winnipeg au Manitoba font également preuve d'une préférence pour la variante CFQ, bien que moins prononcée.

Tableau 2. Résultats de régression logistique – facteurs conditionnant SO (1) et CFQ (-1).

	Valeur	Écart-type	Z	<i>p</i>	
Ordonnée à l'origine ("intercept")	-5.01	6.90	-0.73	0.47	
Naissance - Montréal (QC)	-0.28	0.19	-1.50	0.13	
Naissance - Québec (QC)	-0.23	0.23	-1.02	0.31	
Naissance - Mauricie-Lanaudière	-0.75	0.13	-5.85	6.3 ^E -9	***
Naissance - Winnipeg (MB)	-0.51	0.24	-2.13	0.03	*
Naissance - Maroc	-0.87	0.48	-1.82	0.07	.
Naissance - Ottawa (ON)	-0.79	0.39	-2.01	0.04	*
Naissance - Outaouais-Abitibi (QC)	-1.13	0.34	-3.34	0.001	***
Origines - Montréal (QC)	0.23	0.16	1.42	0.16	
Origines - Québec (QC)	-0.15	0.21	-0.72	0.47	
Origines - Liban	-0.60	0.36	-1.68	0.09	.
Origines - Cameroun	0.41	0.21	1.95	0.05	.
Origines - Bosnie	-0.43	0.21	-2.03	0.04	*

Enfin, trois sous-catégories d'origines géographiques s'avèrent présenter une différence significative par rapport aux autres. Dans ce cas, les coefficients correspondant aux artistes d'origines bosniennes et libanaises tendent vers le négatif, soulignant ainsi un effet favorable pour la variante CFQ, tandis que celui des artistes d'origines camerounaises tend vers le positif et signale ainsi un effet favorable pour la variante SO.

6. Discussion

Tout d'abord, il faut apprécier la forte présence de « *so* » (75,5%) dans notre corpus. Cette variante n'a que très récemment été attestée au Québec et, ce, uniquement au sein d'un corpus de textos d'adolescents (Blondeau et Tremblay 2022). Il est aussi intéressant à constater que « *so* » n'est guère en concurrence avec (*ça*) *fait* (*que*) en tant que marqueur de discours. Les deux variantes ne sont concurrentes que sous forme de marqueur de

conséquence, leur distribution au sein de cette catégorie étant semblable à leur distribution globale au sein du corpus. Ce conditionnement interne l'emporte largement sur les effets de facteurs sociaux. Nous nous demandons s'il ne s'agit de cas de variation que pour la fonction grammaticale de marquer une relation de conséquence. En revenant sur les données brutes, nous n'observons que trois occurrences de « *so* » en tant que marqueur de discours, un nombre si faible qu'il pourrait s'agir de valeurs aberrantes. Si l'on considère les trois occurrences, reproduites ci-dessous, deux ont été produites dans la même chanson de Dead Obies et l'autre dans une chanson de One Nessa. Il s'avère fort probable dans ces deux emplois différents que « *so* » apparaît en position finale d'énoncé afin de servir effectivement à la rime.

- (1) *Comme le silence pour faire durer l'suspense, so...more shows, no-doz* (Dead Obies 2016).
- (2) *I know you fine but you ain't mine so that's why we on then we off* (One Nessa 2015).

Ainsi, il semble juste d'affirmer que l'emploi de « *so* » comme marqueur de discours n'est pas conditionné par des facteurs sociaux, plutôt par la rime. Les effets sociaux sur la production de « *so* » et (*ça*) *fait (que)* ne se produisent qu'en fonction de marqueur de conséquence. Cela implique que la distribution latente de ces deux variantes lorsqu'il est question de variation est encore plus asymétrique qu'il n'y paraît à première vue.

Si l'on exclut les occurrences de la catégorie de marqueurs de discours, il s'avère que « *so* » prédomine et représente 81,41% (n=924) des occurrences pendant que (*ça*) *fait (que)* n'en représente que de 18,59% (n=211). Car cette nouvelle répartition permet d'éclairer la situation de conditionnement social de la variable, nous avons effectué une nouvelle analyse par régression à effets mixtes en tenant uniquement en compte les 1135 occurrences produites comme marqueurs de conséquence. Nous avons vérifié que les effets sur la variation sont quasiment identiques à ce que nous avons présenté dans la section précédente, à la seule exception de l'effet des origines géographiques. Il est très intéressant à constater cette fois une différence significative chez les artistes ayant des origines de la ville de Québec. La différence est négative et signale que ces artistes tendent à utiliser (*ça*) *fait (que)* plus des autres artistes sauf ceux d'origines libanaises, bosniennes et congolaises qui tendent, eux aussi, à produire (*ça*) *fait (que)*. Si nous avons observé auparavant une différence significative, bien que faible, chez les artistes ayant des origines camerounaises, cet effet disparaît lors de cette nouvelle analyse. Tous les autres résultats étant les mêmes que ceux déjà rapportés, continuons notre discussion des autres variables indépendantes.

Tableau 3. Résultats de régression logistique – l'effet d'origine sur SO et CFQ.

	Valeur	Écart-type	Z	p	
Ordonnée à l'origine ("intercept")	-6.030	7.823	-0.771	0.441	
Origines - Montréal (QC)	0.055	0.126	0.435	0.664	
Origines - Québec (QC)	-0.454	0.201	-2.257	0.024	*
Origines - Liban	-0.741	0.352	-2.104	0.036	*
Origines - Bosnie	-0.753	0.220	-3.428	6.0 ^E -4	***
Origines - R.D. du Congo	-0.466	0.211	-2.206	0.028	*

En fonction du genre, il semble bien que les artistes de genre féminin se démarquent des autres artistes par une forte préférence pour la variante « *so* », pendant que les artistes de genre non binaire se démarquent tout à fait dans l'autre sens en privilégiant (*ça*) *fait* (*que*). Bien que l'effet d'âge ne surpasse pas le seuil de signification statistique, la distribution des variantes en fonction d'âge révèle que ce sont les artistes les plus jeunes qui produisent « *so* ». La combinaison de ces deux effets pourrait éventuellement signaler un changement en cours concernant la progression de « *so* » chez les artistes rap.

Il reste encore à considérer l'effet significatif du lieu de naissance de l'artiste et de son choix de désignations identitaires culturelles. Les artistes nés à l'ouest du Québec, en région Mauricie-Lanaudière-Centre du Québec et dans d'autres provinces produisent tous plus la variante (*ça*) *fait* (*que*) par rapport à l'échantillon global d'artistes.

Concernant l'identité culturelle, la seule tendance qui atteint le seuil de signification statistique semble surtout être liée à la dimension américaine lorsqu'il est question d'identités plurielles. Les artistes qui se disent être Américain(e)s sont plus enclins à produire la variante « *so* ». Faute d'informations supplémentaires, nous nous contentons de soumettre la possibilité, qui devra être explorée, que l'aspect multilingue de l'utilisation de cette variante soit motivé par le désir d'exprimer une identité dépassant le cadre local. Dans un contexte moderne de mondialisation, il semble possible que certains artistes fassent usage de certaines variantes linguistiques indexant des associations identitaires non seulement locales mais aussi plus larges. Cela pourrait être motivé par un désir latent de se connecter à la nation rap mondiale en favorisant une variante rattachée au multilinguisme.

7. Conclusion

En résumé, l'objectif de notre étude était de vérifier la théorie de Blondeau et Tremblay (Blondeau 2018, 2020; Blondeau et Tremblay 2016, 2022) de la diversification du français vernaculaire contemporain au Québec et en contexte urbain à Montréal. Notre question de recherche concernant plus particulièrement l'alternance entre (*ça*) *fait* (*que*) et « *so* » était d'évaluer si cette dernière est contrainte par des facteurs sociaux d'une manière à suggérer l'existence de deux pôles normatifs quant au vernaculaire au Québec. Nous avons posé l'hypothèse qu'il existe un pôle centré dans la métropole montréalaise et un autre qui serait plutôt délocalisé sur l'ensemble de la province. Nous avons effectivement constaté des tendances significatives qui signalent la progression de la variante « *so* » en tant que marqueur de conséquence dans la production lyrique des 80 artistes de l'échantillon représenté dans notre corpus « RapKeb21 » (Leavitt 2022). Ce sont particulièrement les artistes de genre féminin et de plus jeune âge qui produisent davantage cette variante. Il semble que malgré la hausse de « *so* » au sein de la communauté rap québécoise l'utilisation de la variante (*ça*) *fait* (*que*) indexe dans une certaine mesure une association aux origines québécoises. Notre analyse semble également mettre en évidence la possibilité que la variante « *so* » fait aussi l'objet d'une indexicalisation renvoyant, elle, à une identité plurielle et complexe liée à la fois aux pratiques langagières locales et mondiales. Ces constats sont généralisables pour leur part à la communauté rap québécoise et ne peuvent être considérés comme représentatifs des pratiques linguistiques de la population québécoise et de ses différentes communautés linguistiques. Il convient toutefois

d'approfondir la question pour déterminer si les mêmes schémas peuvent être observés dans certaines communautés linguistiques, telles que celle des jeunes Montréalais(es), dont il a été démontré lors des études antérieures qu'ils partagent certaines pratiques langagières avec les artistes rap québécois avec qui ils interagissent par le biais de leur participation à la culture populaire ainsi qu'au public du rap québécois.

Œuvres littéraires citées

- Agence QMI. 2022, 29 avril. Loud reprend la tournée. *Le Journal de Montréal*. <https://www.journaldemontreal.com/2022/04/29/loud-reprend-la-tournee-1>
- Alaclair Ensemble. 2013. *Les maigres blancs de l'Amérique du Noir* [Album]. Disques 7ième Ciel.
- Alexy, Kalam. 2022, 19 avril. Après une tournée en Afrique, Monk.E revient au Québec avec du nouveau matériel. *Journal du hip hop*. <https://journalduhiphop.ca/article/apres-un-tournee-en-afrique-monk-e-revient-au-quebec-avec-du-nouveau-materiel/>
- Bordeleau, Jean-Louis. 2022, 28 avril. Le rappeur White-B écope d'une peine de prison de deux ans moins un jour. *Le Devoir*. <https://www.ledevoir.com/culture/musique/704508/justice-le-rappeur-white-b-passera-pres-de-3-ans-et-demi-en-prison>
- Dead Obies. 2016. DO 2 Get [Chanson]. Sur *Gesamtkunstwerk* [Album]. Bonsound.
- Izzy-S. 2018. Le sablier [Chanson]. Sur *Le sablier* [Album]. Explicit Nation.
- Jbeili, Christophe. 2023, 12 juin. New in Rap Queb: Lost is Quebec's biggest rapper right now. *Complex Canada*. <https://www.complex.com/music/a/christophe-jbeili/new-in-rap-queb-lost-is-quebecs-biggest-rapper-right-now>
- Leblanc, Pascal. 2023a, 9 juin. Lost en quatre temps. *La Presse*. <https://www.lapresse.ca/arts/musique/2023-06-09/francos-de-montreal/lost-en-quatre-temps.php>
- Leblanc, Pascal. 2023b, 19 juin. Yes Mccan change de nom et annonce un nouvel album. *La Presse*. <https://www.lapresse.ca/arts/musique/2023-06-19/yes-mccan-change-de-nom-et-annonce-un-nouvel-album.php>
- Muzion. 1999. Lounge with us [Chanson]. Sur *Mentalité moune morne, ils n'ont pas compris* [Album].
- One Nessa. 2015. The Archer [Chanson]. Sur *The Archer* [Album]. Melomayne; Ravalian.
- Perron, Louis-Samuel. 2022, 9 juin. Un rappeur condamné à trois ans de prison soupçonné d'avoir menacé une avocate. *La Presse*. <https://www.lapresse.ca/actualites/justice-et-faits-divers/2022-06-09/armes-a-feu-exhibees-dans-un-vidéo-clip/un-rappeur-condamne-a-trois-ans-de-prison-soupconne-d-avoir-menace-une-avocate.php>
- Renaud, Philippe. 2021, 29 mars. Les leçons de la polémique du français dans le rap québécois. *Le Devoir*. <https://www.ledevoir.com/culture/597799/serie-devoirs-de-francais-les-lecons-de-la-polemique-du-francais>
- Yes McCan. 2016. Yes McCan vs Hermano Salvatore [Chanson]. Sur *WordUP! #TREIZE* [Album].

Références

- Bates, Douglas, Martin Mächler, Ben Bolker et Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1) : 1–48.
- Blondeau, Hélène. 2006. La trajectoire de l'emploi du futur chez une cohorte de Montréalais francophones entre 1971 et 1995. *Revue de l'Université de Moncton* 37(2) : 73–98.
- Blondeau, Hélène. 2018. Pratiques langagières à Montréal et à Bruxelles : diversité et dualité linguistique. *Revue française de linguistique appliquée* 23(1) : 73–88.
- Blondeau, Hélène. 2020. Pratiques langagières et diversité culturelle chez de jeunes Montréalais : le français dans la métropole. Dans *Attribuer un sens. La diversité des pratiques langagières et les représentations sociales*, sous la direction de Kristin Reinke, 151–175. Québec : Presses de l'Université Laval.
- Blondeau, Hélène, France Martineau et Yves Frenette. 2016. Francophonie montréalaise et globalisation : évolution des pratiques langagières en contexte à Hochelaga-Maisonneuve. *Cahiers internationaux de sociolinguistique* 10(1) : 159–182.

- Blondeau, H  l  ne, Raymond Mougeon et Mireille Tremblay. 2019. Analyse comparative de   a fait que, alors, donc et so    Montr  al et    Welland: mutations sociales, convergences, divergences en fran  ais laurentien. *Journal of French Language Studies* 29(1) : 35–65.
- Blondeau, H  l  ne et Mireille Tremblay. 2016. Le traditionnel et l’  mergent : l’apport de jeunes Montr  alais issus de l’immigration au fran  ais vernaculaire. *Cahiers internationaux de sociolinguistique* 10 : 19–45.
- Blondeau, H  l  ne et Mireille Tremblay. 2022.   crire son vernaculaire: variation et normes communautaires dans les messages textes en fran  ais qu  b  cois. *Journal of French Language Studies* 32(2) : 120–144.
- Blondeau, H  l  ne, Mireille Tremblay, Anne Bertrand et Elizabeth Michel. 2021. A new milestone for the study of variation in Montr  al French: The Hochelaga-Maisonneuve sociolinguistic survey. *Corpus* 22 : 16 p. <https://journals.openedition.org/corpus/6221>
- Blondeau, H  l  ne, Mireille Tremblay et Claire Bour  ly. 2022. La dynamique de la variation pronominale : micro-diachronie des formes simples et complexes en fran  ais montr  alais. *Langages* 226 : 69–82.
- Cheshire, Jenny, Susan Fox, Paul Kerswill et Eivind Torgersen. 2013. Language contact and language change in the multicultural metropolis. *Revue fran  aise de linguistique appliqu  e* 18 : 63–76.
- Friesner, Michael Lee. 2009. The social and linguistic predictors of the outcomes of borrowing in the speech community of Montr  al. Th  se de doctorat, University of Pennsylvania.
- Gadet, Fran  oise et Emmanuelle Guerin. 2016. Construire un corpus pour des fa  ons de parler non standard :    Multicultural Paris French   . *Corpus* 15, 16p. <https://journals.openedition.org/corpus/3049>
- Leavitt, Emily Brooke. 2021. A prosodic comparison of accentuation, intonation, and rhythm between Quebec French rap production and normative Quebec French discourse. *Midwest Journal of Undergraduate Research* 12 : 118–136.
- Leavitt, Emily Brooke. 2022. De la variation linguistique dans le rap qu  b  cois. Une   tude sociolinguistique. M  moire de ma  trise, Universit   Concordia.
- Low, Bronwen et Mela Sarkar. 2012. Le plurilinguisme dans les arts populaires, un terrain inexplor   ? L’  tude du langage mixte du rap montr  alais en guise d’exemple. *Kin  phanos. Revue d’  tudes des m  dias et de culture populaire* 3 : 20–47.
- Low, Bronwen et Mela Sarkar. 2014. Translanguaging in the multilingual Montreal hip-hop community: Everyday poetics as counter to the myths of the monolingual classroom. Dans *Educational Linguistics 20: Heteroglossia as practice and pedagogy*, sous la direction de Adrian Blackledge et Angela Creese, 99–118. Dordrecht: Springer.
- Low, Bronwen, Mela Sarkar et Lise Winer. 2009. “Chus mon propre Bescherelle”: Challenges from the hip-hop nation to the Quebec nation. *Journal of Sociolinguistics* 13 : 59–82.
- Martineau, France et Marie-Claude S  guin. 2016. Le Corpus FRAN : r  seaux et maillages en Am  rique fran  aise. *Corpus de fran  ais parl   et fran  ais parl   des corpus* 15 : 1–24.
- Poplack, Shana. 2018. *Borrowing: Loanwords in the Speech Community and in the Grammar*. New York, NY: Oxford University Press.
- Rampton, Ben. 1995. *Crossing: Language and ethnicity among adolescents*. London: Longman.
- R Core Team. 2021. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria.
- RStudio Team. 2020. RStudio: Integrated Development for R. *RStudio*. Boston, MA: PBC.
- Sarkar, Mela. 2008.    Ousqu’on chill    soir ?    : pratiques multilingues comme strat  gies identitaires dans la communaut   hip-hop montr  alaise. *Diversit   urbaine : plurilinguisme et identit  s au Canada* 1 : 27–44.
- Sarkar, Mela. 2009. “Still reppin’ pour mi gente”. The transformative power of language mixing in Quebec hip hop. Dans *Global linguistic flows: Hip hop cultures, youth identities, and the politics of language*, sous la direction de H. Samy Alim, Awad Ibrahim et Alastair Pennycook, 139–158. New York, NY: Routledge.
- Sarkar, Mela et Dawn Allen. 2007. Hybrid identities in Quebec hip-hop: Language, territory and ethnicity in the mix. *Journal of Language, Identity and Education* 6(2): 117–130.
- Sarkar, Mela et Lise Winer. 2006. Multilingual codeswitching in Quebec rap: Poetry, pragmatics and performativity. *International Journal of Multilingualism* 3(3): 173–192.