



(Re)donner la parole aux corpus montréalais

Regards rétrospectif et prospectif

Sous la direction de

WIM REMYSEN et HÉLÈNE BLONDEAU

Avec la collaboration de Marty Laforest



Depuis la réalisation des premières grandes collectes de données sociolinguistiques à Montréal, il y a un peu plus de cinquante ans, l'intérêt pour la linguistique de corpus ne s'est pas démenti. Les corpus occupent aujourd'hui une place centrale dans le domaine des sciences du langage et contribuent à une compréhension plus fine des faits de langue. Leur exploitation continue de soulever d'importantes questions théoriques et méthodologiques, qui sont ici regroupées en trois grands thèmes : le changement linguistique, la diachronie et la variation générationnelle ; les récits de vie, la communauté et l'appartenance sociale ; l'instrumentation des corpus.

Cet ouvrage fait apparaître toute la richesse de l'utilisation des corpus oraux dans la recherche sur la diversité et la variation du français, d'abord à Montréal, mais aussi ailleurs dans la francophonie, en particulier dans les grandes villes.

Sociolinguiste, **Wim Remysen** est professeur de linguistique au Département des arts, langues et littératures de l'Université de Sherbrooke. Il dirige le Centre de recherche interuniversitaire sur le français en usage au Québec et il est responsable du Fonds de données linguistiques du Québec.

Hélène Blondeau est professeure à l'Université de Floride où elle enseigne la sociolinguistique et la linguistique française au Département de langues, littératures et cultures et au Département de linguistique.

Avec les contributions de Julie Abbou, Laurence Arrighi, Hélène Blondeau, Claire Bourély, Heather Burnett, Anne-Sophie Calinon, Marie-Hélène Côté, Léa Courdès-Murphy, Anne Dister, Mireille Elchacar, Julien Eychenne, Nancy Gagné, Annette Gerstenberg, Emmanuelle Labeau, Marta Lupica Spagnolo, Angéline Martel, Raymond Mougeon, Katja Ploog, Wim Remysen, Basile Roussel, Marimar Rufino Morales, Hugo Saint-Amant Lamy, Fabio Scetti, Mireille Tremblay et Isabelle Violette.

38,95 \$ • 28 €

Couverture : collage Gianni Caccia

Disponible en libre accès
www.pum.umontreal.ca

ISBN 978-2-7606-5262-0



(RE)DONNER LA PAROLE AUX CORPUS MONTRÉLAIS

Sous la direction de Wim Remysen et Hélène Blondeau
Avec la collaboration de Marty Laforest

**(RE)DONNER LA PAROLE
AUX CORPUS MONTRÉALAIS**

Regards rétrospectif et prospectif

Les Presses de l'Université de Montréal

Catalogage avant publication de Bibliothèque et Archives nationales du Québec et Bibliothèque et Archives Canada

Titre: (Re)donner la parole aux corpus montréalais: regards rétrospectif et prospectif / sous la direction de Wim Remysen et Hélène Blondeau.

Autres titres: Redonner la parole aux corpus montréalais | Donner la parole aux corpus montréalais

Noms: 1971-2021: 50 ans de corpus montréalais (Colloque) (2022: Archives nationales du Québec à Montréal), auteur. | Remysen, Wim, éditeur intellectuel. | Blondeau, Hélène, 1959- éditeur intellectuel.

Collection: PUM.

Description: Mention de collection: PUM | Textes issus du colloque international 1971-2021: 50 ans de corpus montréalais, tenu aux Archives nationales du Québec à Montréal du 15 au 17 septembre 2022. | Comprend des références bibliographiques.

Identifiants: Canadiana (livre imprimé) 20250038196 | Canadiana (livre numérique) 20250038218 | ISBN 9782760652620 | ISBN 9782760652637 (PDF) | ISBN 9782760652644 (EPUB)

Vedettes-matière: RVM: Sociolinguistique—Québec (Province)—Montréal—Congrès. | RVM: Corpus oraux—Québec (Province)—Montréal—Congrès. | RVM: Corpus oraux—Congrès. | RVM: Français (Langue)—Québec (Province)—Montréal—Congrès. | RVM: Français (Langue)—Variation—Congrès. | RVM: Français (Langue)—Aspect social—Congrès. | RVMGF: Actes de congrès.

Classification: LCC P40.45.C3 A15 2025 | CDD 306.4409714/28—dc23

DOI 10.63361/SPGH8927

Mise en pages: Folio infographie

Dépôt légal: 4^e trimestre 2025

Bibliothèque et Archives nationales du Québec

© Les Presses de l'Université de Montréal, 2025

Les Presses de l'Université de Montréal remercient de leur soutien financier le Fonds du livre du Canada, le Conseil des arts du Canada et la Société de développement des entreprises culturelles du Québec (SODEC).

Financé par le gouvernement du Canada
Funded by the Government of Canada

| **Canada**



Conseil des Arts
du Canada



IMPRIMÉ AU CANADA

INTRODUCTION

Un anniversaire remarquable

Hélène Blondeau et Wim Remysen

L'année 2021 marquait le 50^e anniversaire du corpus Montréal 1971, conçu et réalisé à l'Université de Montréal par une équipe de chercheuses et chercheurs dirigée par Gillian Sankoff (UdeM), David Sankoff (UdeM) et Henrietta Cedergren (UQAM). S'il ne s'agissait pas du premier corpus montréalais de langue française – André Dugas avait en effet réalisé, au début des années 1960, une série d'entrevues dans la métropole –, le corpus Sankoff-Cedergren marquait l'émergence de la sociolinguistique moderne (Labov, 2006) et allait propulser la recherche sociolinguistique québécoise et canadienne. Cette initiative mérite, aujourd'hui encore, notre attention. Rappelons d'abord les objectifs qui ont orienté la constitution de ce corpus, entreprise à une époque où la stigmatisation du français québécois demeurait monnaie courante :

Notre but précis était d'obtenir des données sur la nature, l'étendue et la fonction de la diversité linguistique à l'intérieur du français parlé par les Montréalais, afin d'éclaircir la situation d'une population souffrant d'une forte aliénation linguistique. [...] Nous voulions contribuer à une meilleure compréhension du français parlé au Québec en considérant ses aspects propres [...] en tant qu'éléments d'un système cohérent partagé par tous les membres de la communauté. (Sankoff *et al.*, 1976, 88-89)

Outre qu'il constitue un échantillon représentatif des « individus de 15 ans et plus qui étaient à la fois francophones et Montréalais d'origine » (Sankoff *et al.*, 1976, 91, voir aussi Sankoff, 2018), issu d'une sélection rigoureuse des informatrices et informateurs, le corpus Sankoff-Cedergren

s'insère alors de plain-pied dans l'approche variationniste qui connaît à l'époque un développement important, surtout dans le monde anglo-saxon. Si elles reprennent les principes et axiomes de base de cette nouvelle discipline, Gillian Sankoff, Henrietta Cedergren et leur équipe introduisent en même temps de nouvelles questions et méthodes de recherche qui ont inspiré d'autres études sur la variation du français en usage au Québec et ailleurs dans la francophonie, et qui ont donné lieu à des débats en sociolinguistique générale (Thibault, 2001). Ainsi, l'apport du mathématicien David Sankoff à l'équipe a permis de renouveler l'approche variationniste par l'introduction de nouvelles méthodes d'analyse quantitative, fondées sur le recours à des modèles de régression, ou faisant appel à l'informatique (Cedergren et Sankoff, 1974, Rousseau et Sankoff, 1978). Sur le plan théorique, on doit aussi à l'équipe l'introduction d'une nouvelle variable sociale, assise sur le rapport à la langue standard, appelée « participation au marché linguistique » (Sankoff et Laberge, 1978), de même que l'élargissement du domaine d'application des règles variables à des phénomènes relevant de la morphologie, de la morphosyntaxe et de la sémantique (Sankoff et Thibault, 1977, Sankoff, *et al.*, 1978).

Mais c'est peut-être le caractère itératif de l'entreprise qui frappe le plus l'imagination et confirme la pertinence des méthodes variationnistes pour l'étude du changement linguistique (Blondeau, 2020). En 1984, Pierrette Thibault et Diane Vincent (qui avaient fait leur thèse sous la direction de Gillian Sankoff en analysant les données de 1971) entreprennent une nouvelle collecte de données auprès de la moitié des participants du corpus de 1971, soit 60 personnes, et y ajoutent 12 jeunes locuteurs, ouvrant ainsi la possibilité de faire des études comparatives sur le changement linguistique dans la communauté, dans une perspective longitudinale (Thibault et Vincent, 1990). Et une autre dizaine d'années plus tard, en 1995, une nouvelle équipe, composée cette fois-ci de Diane Vincent, Marty Laforest et Guylaine Martel, réalise l'exploit de retracer 14 participants des premières enquêtes (Vincent *et al.*, 1995), donnant ainsi plus de profondeur encore à l'analyse diachronique¹. Dans ces trois corpus, la collecte de données s'appuie sur des entretiens de même type afin de préserver la comparabilité. Par ailleurs, réalisées

1. Les corpus Montréal 1984 et 1995 sont présentés dans des capsules vidéo disponibles sur le site du Fonds de données linguistiques du Québec (<https://fdlq.recherche.usherbrooke.ca/capsules-video>), où il est aussi possible d'interroger les transcriptions (<https://fdlq.crifuq.usherbrooke.ca/>). L'ajout du corpus Montréal 1971 sur cette plateforme est en cours (voir l'article de Remysen et Saint-Amant Lamy, dans ce volume, pour plus d'information).

dans des contextes différents, ces nouvelles collectes de données ont permis d'élargir les recherches initialement entreprises sur le français de Montréal, par la prise en compte de dimensions moins explorées jusque-là, dont l'analyse conversationnelle et la variation discursive, ainsi que par l'introduction de méthodes novatrices, comme l'auto-enregistrement des participants lors d'activités quotidiennes. L'intérêt croissant à l'heure actuelle pour la constitution de corpus dits « écologiques » montre l'originalité de cette méthode déjà proposée par les sociolinguistes québécois dans les années 1990.

Pour rendre hommage aux sociolinguistes québécoises et québécois de la première heure (ainsi qu'aux informateurs et informatrices qui ont participé à leurs collectes de données et ainsi rendu possible la recherche sur le français montréalais), nous avons décidé d'organiser en septembre 2022² un colloque international et de réunir des chercheuses et chercheurs interpellés par la diversité français – en premier lieu à Montréal (et plus largement au Québec et au Canada), mais aussi ailleurs dans la francophonie, en particulier dans d'autres grandes villes –, et par le recours à des corpus de langue orale. Le volume que vous avez entre les mains réunit 13 des contributions sur une trentaine en tout, présentées à l'occasion de ce colloque. Nous les avons regroupées ici en fonction de trois grands thèmes.

Thème 1: Changement, diachronie et variation générationnelle

La réalisation du corpus Sankoff-Cedergren a grandement contribué à une meilleure compréhension des multiples variations que l'on peut observer dans le français parlé à Montréal et des changements qui ont marqué cette variété au cours du dernier demi-siècle. Depuis les premiers travaux menés à ce sujet, l'intérêt pour le français montréalais ne s'est pas démenti : les trois corpus montréalais ont donné lieu à plus de 150 études entre 1971 et 2015 et les recherches se poursuivent³. Ce premier axe thématique porte sur le changement et la diachronie et regroupe des recherches portant sur la variété de français parlé à Montréal et celles que l'on trouve dans d'autres centres

2. L'événement était initialement prévu pour 2021, exactement 50 ans après la constitution du corpus Montréal 1971, mais – surprise – la pandémie mondiale de la COVID-19 nous a obligés de reporter d'un an la tenue du colloque.

3. Voir la *Bibliographie des travaux fondés sur les grands corpus sociolinguistiques de Montréal*, préparée par Hélène Blondeau et Marty Laforest et disponible en ligne (<https://fdlq.recherche.usherbrooke.ca/bibliographie-travaux-grands-corpus-sociolinguistiques-mtl.html>).

urbains, ou encore interroge leurs spécificités à la lumière des variétés pratiquées en dehors des villes. Les quatre contributions se rattachant à ce thème ont comme point commun la dimension comparative, puisque chaque étude met en contraste des corpus ou des données recueillies à des dates et/ou dans des lieux différents.

Marie-Hélène Côté examine l'évolution de la liaison depuis les années 1971. L'étude compare la liaison dans le corpus Sankoff-Cedergren de Montréal et le corpus tiré du projet Phonologie du français contemporain (PFC), recueilli au Québec et découpé aux fins de l'analyse en deux zones (Montréal, Québec hors Montréal). Elle compare aussi les données de PFC-Québec aux autres corpus PFC, dans le but de mesurer le phénomène à l'aune du reste de la francophonie. L'étude distingue cinq configurations, définies par les différences entre le corpus de Montréal et PFC : la liaison (quasi) catégorique, la liaison marginale, la liaison plus fréquente dans PFC que dans Montréal 1971, la liaison plus fréquente dans Montréal 1971 que dans PFC et enfin la liaison s'écartant de la norme prescriptive dans Montréal 1971. Les analyses font ressortir l'hétérogénéité du phénomène de la liaison et la nécessité d'une analyse détaillée selon les contextes, pour bien saisir la portée des tendances évolutives.

Hélène Blondeau, Claire Bourély, Raymond Mougéon et Mireille Tremblay explorent l'évolution des variétés du français laurentien, en comparant le français parlé à Montréal et à Welland, en Ontario, dans les années 1970 et 2010. L'étude identifie une tendance à la divergence entre les deux communautés dans l'usage des marqueurs de conséquence (*ça fait que, donc, alors et so*) et dans l'emploi des pronoms simples et complexes du pluriel (*nous, vous, eux/elles* vs *nous autres, vous autres, eux autres*). Malgré le partage de plusieurs traits linguistiques communs à l'origine, cette tendance à la divergence, naissante dans les années 1970, s'accroît dans les années 2010. Selon l'interprétation proposée, la divergence découle de différences liées, d'une part, au faible accès à l'éducation postsecondaire en français et aux médias de langue française en contexte minoritaire à Welland, et, d'autre part, à une exposition accrue à d'autres variétés internationales du français à Montréal.

Mireille Tremblay, Hélène Blondeau, Anne Dister et Emmanuelle Labeau étudient la référence temporelle au futur en Belgique et au Québec et comparent la variation dans des corpus oraux et des corpus de textos. L'étude identifie une différence dialectale puisque le futur synthétique est davantage utilisé en Belgique qu'au Québec, peu importe le support. Cela

dit, les résultats confirment qu'un changement vers les formes analytiques a lieu dans les deux communautés. À l'oral, le changement se déroule à un rythme différent et selon des contraintes linguistiques particulières, puisque les facteurs linguistiques de la polarité et de la distance temporelle ne jouent pas le même rôle dans les deux communautés. Enfin, la comparaison entre les deux supports montre que, dans les deux communautés, il y a beaucoup plus de futur synthétique dans les corpus de textos que dans les corpus oraux, ce qui illustre la position intermédiaire occupée par les messages textes sur le continuum oral-écrit.

Annette Gerstenberg et Marta Lupica Spagnolo proposent une étude de cas à partir du corpus LangAge, recueilli à Orléans, en France, auprès de locuteurs âgés de 70 ans et plus. Cette étude, qui porte sur madame Moreau, une locutrice orléanaise, adopte une méthodologie longitudinale s'échelonnant sur une courte période, caractérisée par des entrevues réalisées à différents moments dans la vie des témoins. Cette approche sert, entre autres, à repérer des marques associées au vieillissement linguistique. L'analyse présentée dans cette contribution est abordée sous les angles de la variation, de l'expressivité et de l'interaction. Elle met en valeur la portée des études longitudinales en sociolinguistique et leur contribution à l'étude de la variation chez l'individu, une voie de recherche qui connaît un développement important depuis quelques années.

Thème 2 : Récits de vie, communauté et appartenance sociale

Les sociolinguistes ont recours à l'entrevue pour solliciter des discours oraux susceptibles de nous renseigner sur les pratiques langagières des locutrices et locuteurs. Leur première préoccupation est de lancer des sujets de conversation jugés propices à « provoquer chez les répondants des commentaires spontanés et prolongés » (Sankoff *et al.*, 1976, 113), mais souvent sans s'intéresser de plus près au contenu des thématiques abordées, dont plusieurs devraient pourtant intéresser les sociolinguistes (que l'on pense par exemple aux questions susceptibles de nous renseigner sur les représentations auxquelles donnent lieu le français et ses différentes variétés, ou encore l'anglais ou d'autres langues parlées par les Montréalais et Montréalaises). En outre, les corpus sociolinguistiques contiennent souvent, en plus des observations sur la vie de quartier ou la situation politique et linguistique montréalaise, des histoires de vie, c'est-à-dire des récits autobiographiques qui mettent au premier plan l'individu, sa trajectoire, ses opinions et ses goûts. La nature

longitudinale des corpus montréalais a d'ailleurs ouvert la voie au suivi de la trajectoire des locuteurs au cours de leur vie au travers de leurs discours et de leurs pratiques langagières. De manière générale, les corpus sociolinguistiques peuvent ainsi devenir des lieux de mémoire de la ville, de la vie dans ses quartiers ou de la perception de l'identité et de la mobilité des personnes. C'est à ce type de narration, qui intéresse de plus en plus les chercheurs en sciences humaines – sociologues, historiens, littéraires et autres (Goodson *et al.*, 2016) – qu'est consacré le deuxième axe thématique du volume, qui comprend cinq contributions.

Claire Bourély signe un chapitre portant sur la gentrification à Hochelaga-Maisonneuve à travers un examen de la toponymie de ce quartier montréalais. Cet examen s'appuie sur une analyse du contenu d'entretiens tirés d'un corpus sociolinguistique recueilli à Hochelaga-Maisonneuve, et d'une comparaison avec le discours des médias. L'autrice arrive ainsi à exposer le processus de gentrification qu'a connu le quartier. Cette contribution illustre bien comment les corpus sociolinguistiques qui contiennent des entretiens peuvent également servir à éclairer des questions de recherches en analyse du discours et en sciences sociales appliquées.

Mireille Elchacar, Angéline Martel et Nancy Gagné présentent une recherche-action en communication et compétences interculturelles à partir du corpus Traits d'union employé comme matériau pour l'analyse des perceptions de la langue française en milieu de travail. La recherche porte sur les perceptions émanant du discours des intervenants de différents services offerts aux citoyens issus de l'immigration. Les résultats indiquent certains parallèles entre, d'une part, les thématiques et perceptions linguistiques des nouveaux arrivants, décrites dans les sources documentaires et, d'autre part, les perceptions recensées chez les intervenants, principalement en ce qui concerne l'importance de la connaissance de la langue pour l'intégration des immigrants. Toutefois, certaines divergences s'observent selon les différents contextes d'intervention. Ces résultats conduisent à une discussion des implications en matière de formation des intervenants dans le but d'améliorer la communication interculturelle.

Fabio Scetti se penche sur la Petite Italie et le Petit Portugal montréalais dans le cadre d'une analyse des récits et des discours circulant au sein des communautés. L'analyse s'appuie sur des observations ethnographiques et des entretiens tirés des corpus multilingues MontPORT (principalement en portugais) et MontreITal (principalement en italien) recueillis lors d'études de terrain de 2011 à 2024. Après une brève présentation des deux commu-

nautés, le chapitre présente l'exploitation des corpus du point de vue linguistique et discursif. En racontant l'histoire de Montréal selon la perspective des membres des deux communautés, l'auteur montre des pratiques langagières empreintes d'une diversité de langues et de dialectes, certains aujourd'hui peu parlés dans leurs localités d'origine, et des représentations révélant l'identité de chacune des communautés en rapport avec ses pratiques langagières et culturelles.

Anne-Sophie Calinon et Katja Ploog explorent la mobilité discursive, c'est-à-dire la capacité de mobiliser des ressources acceptables du point de vue non seulement grammatical, mais aussi sociodiscursif, dans un corpus montréalais, en s'intéressant aux expressions figées. L'étude, qui s'appuie sur un corpus d'entretiens recueillis auprès d'étudiantes et étudiants maghrébins en situation de mobilité et installés au Québec, se penche sur le processus d'appropriation du nouvel espace discursif associé à leur nouvelle expérience québécoise. L'usage des expressions figées sert de mesure à l'appropriation de ce nouvel espace discursif. L'analyse décrit différents usages des expressions figées, certains mettant en scène des cas de créativité, de même que des cas de fusion, qui témoignent d'effets de résonance entre des ressources langagières associées à des indexicalités différentes, en matière de créativité.

Julie Abbou et Heather Burnett présentent le corpus CaFé, un ensemble composé d'entretiens sociolinguistiques avec des militantes et militants féministes et queers, réalisés dans les villes de Paris, Montréal et Marseille. L'étude se focalise sur les portions des entretiens comprenant des récits de vie et illustrant la trajectoire qui a conduit ces personnes à devenir féministes et/ou queers. Malgré plusieurs points communs, l'analyse critique du discours des récits de vie montre des différences d'articulation entre le féminisme et le queer à Paris et à Montréal. Alors qu'à Paris, le queer est considéré comme une orientation politique au sein du féminisme, à Montréal, cette étiquette correspond davantage à une orientation sexuelle ou une catégorie de genre. Il semble donc que la composition démographique des corpus soit intrinsèquement liée à la circulation du sens que prennent les catégories sociodémographiques en contexte, ce qui amène les autrices à considérer la question de la représentativité qualitative.

Thème 3 : Instrumentation des corpus

Nous avons regroupé sous ce thème les quatre contributions portant sur le développement de corpus sociolinguistiques à l'ère des humanités numériques.

Les questions théoriques et méthodologiques soulevées au moment de la constitution du premier corpus montréalais demeurent en grande partie d'actualité pour la sociolinguistique contemporaine. Mais aux questions posées dès le lancement de ce premier grand chantier concernant la représentativité et l'authenticité des corpus, indispensables pour l'étude de la variation linguistique et des caractéristiques de l'oral spontané, s'ajoutent désormais, à l'ère des humanités numériques, une série de préoccupations nouvelles. Celles-ci concernent, entre autres, mais pas exclusivement, le recours à des méthodes de traitement automatique des données, la conservation et la pérennisation des corpus patrimoniaux et les enjeux (éthiques, par exemple) qui en découlent, ou encore la gestion des données de masse (« *big data* ») et leur exploitation à des fins sociolinguistiques. S'y ajoutent aussi les questions relatives à la comparaison des données provenant de différents corpus et celles associées à la transcription. Cet axe propose de réfléchir à ces enjeux de constitution, d'outillage et d'analyse des corpus sociolinguistiques à partir d'exemples de collectes de données réalisées dans des contextes variés.

Wim Remysen et Hugo Saint-Amant-Lamy offrent un tour d'horizon de la plateforme du Fonds de données linguistiques du Québec (FDLQ) et se penchent sur son rôle dans la préservation des données et leur accessibilité. Une première section présente la plateforme numérique destinée à la mise en valeur des corpus oraux et écrits recueillis aux fins de description et d'étude du français du Québec, colligés depuis les années, 1960. La contribution, qui se focalise ensuite sur les corpus montréalais, fournit des informations sur les différentes étapes du traitement des corpus. Un exposé sur l'exploitation des données fournit une illustration à partir des mots en /ar#/ afin d'illustrer le phénomène de la neutralisation des voyelles basses antérieure et postérieure devant un /R/ final. Le chapitre expose comment le FDLQ a le mérite de pérenniser des corpus d'une valeur patrimoniale cruciale pour la linguistique du français québécois, ce qui constitue une avancée majeure pour les humanités numériques au Québec.

Laurence Arrighi, Basile Roussel et Isabelle Violette rendent compte d'un projet de description du français parlé à Moncton, qui contribuerait non seulement aux études sociolinguistiques sur le français acadien, mais également à celles sur l'urbanité dans les villes de taille moyenne. La contribution présente l'état des lieux de la recherche portant sur le fait français à Moncton, pour ensuite émettre des hypothèses sur les effets linguistiques de la densification et la complexification de cette ville au fil du temps. Le chapitre montre que certains éléments auparavant négligés nécessitent une

prise en compte, notamment la contribution des Néo-Monctoniens à la dynamique du français parlé dans la ville. De surcroît, il discute le problème du discours homogénéisant du français parlé à Moncton, qui représente trop souvent les pratiques langagières associées à l'anglicisation ou à l'hybridation du français.

Julien Eychenne et Léa Courdès-Murphy présentent la plateforme logicielle Phonometrica, un outil pour l'annotation et l'analyse de corpus alignés sur le signal sonore. Ce logiciel vise à faciliter un certain nombre de tâches pour le codage et l'analyse de variables en (socio)phonologie et en (socio)phonétique. Le chapitre fournit des informations détaillées sur les principales fonctionnalités du logiciel en ce qui concerne le traitement des données de corpus et l'organisation des différentes requêtes. Certains enjeux importants liés à la programmabilité et à l'extensibilité de la plateforme sont également discutés. Le chapitre se termine par une présentation des développements visant à améliorer cette plateforme en constante évolution.

Marimar Rufino Morales se penche sur la question de la transcription des corpus linguistiques à l'aide de la technologie intelligente. Elle aborde cette question à partir des enjeux reliés à la transcription du Corpus oral de la langue espagnole à Montréal (COLEM) et l'exploitation de la technique de la redite, qui consiste à écouter un son original et à le redire à un logiciel de reconnaissance automatique de la parole (RAP). Après une revue des différentes techniques de transcription, le chapitre présente les résultats d'une étude comparative expérimentale de la transcription du COLEM. L'analyse montre que les résultats obtenus avec la technique de transcription de la redite sont supérieurs à ceux de la transcription manuelle. L'autrice en conclut que l'optimisation de la transcription redite repose sur la collaboration humain-machine et exige un travail en amont sur la variabilité de la langue et d'autres éléments liés aux caractéristiques des témoins et aux bruits environnants.

* * *

En terminant, nous tenons à remercier toutes les personnes qui ont participé au colloque de 2022, à commencer par les sociolinguistes à qui nous avons voulu rendre hommage, et tous nos collègues qui ont rédigé un chapitre pour contribuer à ce volume. Ce livre n'aurait pas pu voir le jour sans la précieuse collaboration de quelques personnes que nous remercions très sincèrement : Pamela Vachon (Université de Sherbrooke), pour son appui à la coordination,

Marty Laforest (UQTR), qui a été très présente tout au long de la préparation du volume, Guy Champagne et Marie-Chantal Plante (Les Presses de l'Université de Montréal) ainsi que les collègues qui en ont évalué les contenus. La publication de ce livre a bénéficié du soutien financier du Fonds de recherche du Québec – Société et culture, du Centre de recherche interuniversitaire sur le français en usage au Québec (CRIFUQ), du Fonds de données linguistiques du Québec (FDLQ) ainsi que de l'Université de Sherbrooke et de la Faculté des lettres et des sciences humaines de l'Université de Sherbrooke (dans le cadre du programme d'appui aux activités de création et d'édition savantes). Et merci au graphiste Gianni Caccia qui, sans le savoir, a inséré sur la couverture de ce livre un beau clin d'œil à l'ancien Laboratoire d'ethnolinguistique du Département d'anthropologie de l'Université de Montréal où a été élaboré le corpus de Montréal 1971. Sur les murs de ce labo était en effet installée une carte de la ville indiquant les lieux de résidence des locutrices et locuteurs rencontrés pour la constitution de ce corpus⁴.

Références

- BLONDEAU, Hélène (2020), « Du cheminement sociolinguistique des individus: l'apport des études longitudinales à l'étude de la variation et du changement en français parlé », dans Wim REMYSEN et Sandrine TAILLEUR (dir.), *L'individu et sa langue: hommages à France Martineau*, Québec, Presses de l'Université Laval, 13-46.
- CEDERGREN, Henrietta et David SANKOFF (1974), « Variable rules: performance as a statistical reflexion of competence », *Language*, 50(2), 333-355.
- GOODSON, Ivor, Ari ANTIKAINEN, Pat SIKES et Molly ANDREWS (dir.) (2016), *The Routledge international handbook on narrative and life history*, Londres, Routledge.
- LABOV, William (2006 [1966]), *The social stratification of English in New York City*, Cambridge, Cambridge University Press.
- ROUSSEAU, Pascale et David SANKOFF (1978), « Advances in variable rule methodology », dans David SANKOFF (dir.), *Linguistic variation: models and methods*, New York, Academic Press, 57-69.
- SANKOFF, Gillian (2018), « Before there were corpora: the evolution of the Montreal project as a longitudinal study », dans Susanne EVANS WAGNER et Isabelle BUCHSTALLER (dir.), *Panel studies of variation and change*, New York/Londres, Routledge, 21-52.
- SANKOFF, Gillian et Pierre THIBAUT (1977), « L'alternance entre les auxiliaires *avoir* et *être* en français parlé à Montréal », *Langue française*, 34, 81-108.

4. Une visite virtuelle de ce laboratoire est disponible à l'adresse https://youtu.be/VjgJlvNtj6I?si=eaidyZW_TI5Jh7Yb (vidéo préparée en 2012 par la sociolinguiste Pierrette Thibault, coauteure du corpus Montréal 1984).

- SANKOFF, David et Suzanne LABERGE (1978), « The linguistic market and the statistical explanation of variability », dans David SANKOFF (dir.), *Linguistic variation: models and methods*, New York, Academic Press, 239-250.
- SANKOFF, David, Gillian SANKOFF, Suzanne LABERGE et Marjorie TOPHAM (1976), « Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale », *Cahiers de linguistique*, 6, 85-125.
- SANKOFF, David, Pierrette THIBAUT et Hélène BÉRUBÉ (1978), « Semantic field variability », dans David Sankoff (dir.), *Linguistic variation: models and methods*, New York, Academic Press, 23-43.
- THIBAUT, Pierrette (2001), « Regard rétrospectif sur la sociolinguistique québécoise et canadienne », *Revue québécoise de linguistique*, 30(1), 19-42.
- THIBAUT, Pierrette et Diane VINCENT (1990), *Un corpus de français parlé: Montréal 84. Historique, méthodes et perspectives de recherche*, Québec, Département de langues et linguistique.
- VINCENT, Diane, Marty LAFOREST et Guylaine MARTEL (1995), « Le corpus de Montréal, 1995: adaptation de la méthode d'enquête sociolinguistique pour l'analyse conversationnelle », *Dialangue*, 6, 29-45.

Changement, diachronie et variation générationnelle

CHAPITRE 1

Cinquante ans de liaison à Montréal éclairée par le reste du monde

Marie-Hélène Côté

Deux phénomènes omniprésents et incontournables – le schwa et la liaison – marquent la prononciation du français. Par la variabilité qui les caractérise et la multiplicité des facteurs qui en déterminent le comportement, ces phénomènes se sont imposés depuis des décennies comme des classiques de la phonologie et de la sociolinguistique. Dans les deux cas, la dimension diaphasique de la variation ne fait aucun doute : les consonnes de liaison et le schwa sont plus souvent réalisés dans des situations formelles. Le schwa et la liaison se distinguent cependant par le rôle traditionnellement accordé à la variation diatopique. Pour le schwa, on oppose immédiatement deux types de systèmes, associés à des variétés définies géographiquement : le type septentrional, auquel appartient la variété de référence (Lyche, 2010 ; Detey *et al.*, 2016), et le type méridional. Mais des travaux font régulièrement état des pressions du système septentrional sur le méridional (Müller et Wiedner, 2022). Pour la liaison, les travaux sur la variété montréalaise des années 1970 (Ameringen, 1977 ; Ameringen et Cedergren, 1981 ; Tousignant, 1978 ; de Jong, 1993) ont mis en lumière les spécificités de cette variété, mais ces résultats n'ont pas été intégrés aux discussions plus générales sur la liaison en français, les particularités nord-américaines étant généralement traitées de façon plus autonome, du moins pour ce qui relève de la composante sonore du langage, par rapport aux variétés parlées en Europe.

Mis sur pied au tournant des années 2000 et toujours en développement, le projet Phonologie du français contemporain (PFC ; Durand *et al.*, 2002 ;

<http://www.projet-pfc.net>) a permis un renouvellement de la dimension diatopique en intégrant toutes les variétés dans un cadre commun. Le protocole prévoit en outre un schéma d'annotation du schwa et de la liaison. Pour cette dernière, nous avons récemment montré (Côté, 2017) que la variation diatopique surpassait en fait la variation diaphasique. Une des fractures oppose les variétés nord-américaines aux variétés européennes et africaines par la régularité de règles de liaison ne correspondant pas aux normes prescriptives établies. Les résultats du programme PFC sur la liaison redonnent ainsi un nouveau souffle aux données montréalaises des années 1970, en leur offrant un cadre d'interprétation plus général et en ajoutant une dimension diachronique par la comparaison avec les données PFC recueillies récemment au Québec.

Notre objectif est de procéder à une comparaison systématique de la liaison dans deux corpus : celui de Montréal 1971 (Mtl71), dont la liaison a été décrite par Ameringen (1977), et le corpus PFC, découpé en trois zones (Montréal, le Québec hors Montréal et le reste de la francophonie). Cette dernière catégorie est représentée par les points d'enquête effectués dans différentes régions du monde francophone et dont les données sont déposées sur le site PFC. Ces points se trouvent surtout en France, mais également dans le reste de l'Europe francophone et en Afrique du Nord et subsaharienne. Ce découpage en zones permet d'aborder trois questions spécifiques :

- Comment la liaison a-t-elle évolué à Montréal depuis 1971 ?
- Comment se comporte la liaison à Montréal par rapport aux autres régions du Québec ?
- Comment la liaison montréalaise/québécoise se positionne-t-elle par rapport au reste de la francophonie ?

Données

Ameringen (1977) a analysé la liaison dans le corpus Mtl71 en ciblant trois types de contextes :

- après des morphèmes monosyllabiques grammaticaux, à savoir,
 - déterminants (ex. *un* [n] *ami*)
 - pronoms prétéverbaux (ex. *je les* [z] *aime*)
 - adverbes *plus* et *moins* (ex. *plus* [z] *important*)
 - prépositions *en*, *dans*, *chez*, *sans* (ex. *en* [n] *arrière*)

- verbes *être* et *avoir* (ex. *c'est* [t] *ici*)
- verbes modaux *pouvoir*, *aller*, *vouloir*, *devoir*, *falloir* (ex. *il faut* [t] *aller*)
- après des adjectifs prénominaux (ex. *grand* [t] *ami*, *vraies* [z] *affaires*)
- après des formes bisyllabiques à l'imparfait de *être* et *avoir* (ex. *c'était* [t] *arrivé*)

Les résultats d'Ameringen (1977), repris dans Ameringen et Cedergren (1981), s'appuient sur les données de 27 informateurs et informatrices, âgés entre 20 et 76 ans et répartis dans trois groupes distincts, définis dans Ameringen (1977) par la profession et la scolarité et dans Ameringen et Cedergren (1981) par une cote d'intégration au marché linguistique, qui indique dans quelle mesure la maîtrise de la langue légitime est considérée comme nécessaire dans la vie économique des individus. Le groupe I (15 individus) est le moins intégré au marché linguistique, le groupe II (9 individus) est d'intégration moyenne, alors que le groupe III (3 individus) présente une intégration maximale. Pour les contextes de liaison les plus fréquents, on n'a considéré que des sous-ensembles de 7 et 4 individus dans les groupes I et II; nous parlerons des groupes I et II élargis et restreints. Le groupe III, lui, n'a pas été analysé pour les contextes où les membres des groupes I et II (élargis ou restreints) réalisent déjà la liaison de façon catégorique ou quasi catégorique et correspondant à la norme prescriptive.

Ameringen (1977) a ainsi déterminé l'absence ou la présence de la consonne de liaison dans un total de 2 123 occurrences. Le nombre d'occurrences varie grandement selon les contextes: le plus fréquent est *on* (499 occurrences dans les groupes I et II restreints), alors qu'on ne trouve que 15 cas de liaison potentielle après des adjectifs prénominaux au singulier dans les groupes I et II élargis. Pour chaque contexte, on établit un taux de réalisation, qui s'échelonne de 0 % à 100 %, avec différentes valeurs intermédiaires.

La liaison dans l'ensemble du corpus PFC est décrite et analysée dans Durand et Lyche (2008), Durand *et al.*, (2011), Mallet (2008) et Barreca (2015). Ces études s'appuient sur l'annotation de la liaison prévue dans le protocole du projet. Une occurrence de liaison potentielle est définie par la présence d'une consonne finale latente suivie d'un mot à initiale vocalique, dans les contextes énumérés par Delattre (1951); Côté (2025) souligne les problèmes que pose cette définition d'un contexte de liaison, mais ceux-ci n'affectent pas les résultats de la présente étude. Chaque occurrence de liaison potentielle, de même que toute liaison réalisée dans un contexte non prévu (p. ex.,

il va [t] *aller*), fait l'objet d'un codage alphanumérique qui indique : le nombre de syllabes du mot précédent (mono vs polysyllabique), la réalisation (ou non) de la consonne de liaison et, le cas échéant, l'identité de la consonne réalisée et son enchaînement (ou non) à la voyelle suivante. La présence d'une pause ou hésitation est également mentionnée, de même que la nature orale ou nasale de la voyelle précédent un [n] de liaison (p. ex., voyelle orale dans *bon* [bɔn] *emploi* mais nasale dans *mon* [mɔ̃n] *emploi*).

Les données québécoises, recueillies plus tard, sont cependant très peu représentées dans ces analyses. Entre 2010 et 2017, nous avons constitué le corpus PFC-Québec, qui comprend 440 locuteurs et locutrices provenant de 32 points d'enquête répartis sur toute la province⁵ (Côté, 2014; Côté et Saint-Amant Lamy, 2023). Chaque point d'enquête comprend une douzaine de locuteurs ventilés en trois générations et comptant un nombre égal d'hommes et de femmes. Chaque participant à une enquête a été enregistré dans différentes tâches, conformément au protocole PFC, dont deux conversations. S'agissant de la comparaison entre Mtl71 et PFC, le codage liaison prévu dans le protocole de traitement PFC a été appliqué aux conversations d'un sous-ensemble de 67 locuteurs et locutrices de PFC-Québec, divisé en deux zones (Montréal et hors Montréal) et en trois groupes d'âge définis par l'année de naissance (voir le tableau 1.1).

Pour Saguenay et Trois-Rivières, le codage liaison a été appliqué à cinq minutes de chacune des conversations, comme le prévoit le protocole PFC. Cette durée s'avère pourtant insuffisante pour l'analyse de contextes de liaison peu fréquents, raison pour laquelle nous avons annoté l'entièreté des conversations pour les autres localités. Nous avons ainsi annoté un total de 25 158 contextes de liaison. Ces derniers couvrent cependant un ensemble de contextes plus large que celui analysé par Ameringen (1977). Afin d'assurer la comparabilité de nos données avec celles du corpus Mtl71, nous ne conserverons ici que les occurrences correspondant aux contextes retenus par Ameringen (1977). Nous écarterons également la liaison après les adverbes *plus* et *moins*, pour lesquels le codage effectué par Ameringen (n=43) n'est pas suffisamment explicite pour une comparaison fiable avec les données PFC. Finalement, 12 067 contextes de liaison seront étudiés dans PFC-Québec : 6 848 à Montréal et 5 219 hors Montréal.

Ces données seront systématiquement mises en regard des données PFC obtenues dans le reste de la francophonie. Au moment de la rédaction de ce

5. Dont un point d'enquête (Saguenay) effectué sous la direction de Luc Baronian en 2009.

TABLEAU 1.1

Échantillon de PFC-Québec analysé

		Groupe âgé (1921-1950)	Groupe moyen (1952-1979)	Groupe jeune (1981-1997)
Montréal (n=29)		10	9	10
Hors Montréal (n=38)	St-Éphrem-de-Beauce (12)	5	3	4
	Saguenay (11)	2	5	4
	Trois-Rivières (12)	4	2	6
	Sainte-Adèle (2)	2		
	Québec (1)			1
TOTAL (n=67)		22	19	26

chapitre, les annotations de liaison étaient disponibles sur le site PFC pour 39 points d'enquête, répartis comme suit : France métropolitaine (20), France d'outre-mer (1), Belgique (3), Suisse (3), Afrique subsaharienne (6), Algérie (2), Canada (4)⁶. Grâce au moteur de recherche disponible sur le site PFC, nous avons recueilli pour chaque contexte le nombre d'occurrences avec et sans liaison, en retirant les données des enquêtes canadiennes. Seules les conversations guidées et libres ont été prises en considération.

Résultats

Pour les contextes de liaison examinés, nous présenterons les résultats en distinguant cinq configurations distinctes, définies par la fréquence relative de la liaison entre Mtl71 et le corpus PFC général. La liaison peut être quasi catégorique ou quasi absente dans les deux corpus ou plus fréquente dans l'un que dans l'autre ; pour les contextes de liaison plus fréquents dans Mtl71, nous examinerons séparément les cas où la liaison montréalaise s'écarte de la norme prescriptive.

Liaison (quasi) catégorique dans Mtl71 et PFC

Cette catégorie comprend trois contextes, pour lesquels le tableau 1.2 présente les taux de réalisation (%) et le nombre d'occurrences avec et sans liaison (n)

6. Deux des enquêtes canadiennes sont incluses dans PFC-Québec (Saguenay et Trois-Rivières). Les deux autres sont l'enquête pilote de Jacques Durand à l'Université Laval et l'enquête de Peace River (Alberta).

TABLEAU 1.2

Contextes de liaison (quasi) catégorique dans Mtl71 et PFC-Québec

	Mtl71		PFC-Québec		PFC-autre	
	%	n	%	n	%	n
Déterminants singuliers	97,5 %	83 (restreints)	97,5 %	903	99,4 %	1027
Déterminants pluriels, sauf <i>les</i>	99,5 %	188 (restreints)	96,4 %	943	99,0 %	1601
Déterminant <i>les</i>			95,8 %	644	99,0 %	1435
Pronom <i>les</i>	94,4 %	18' (élargis)	99,2 %	119		
Pronoms <i>nous/vous</i>	92,9 %	42 (élargis)	100 %	282	99,5 %	1017
Pronom <i>en</i>	98,6 %	221 (restreints)	99,6 %	826	98,3 %	1242
Préposition <i>en</i>	90,1 %	44 (restreints)	90,3 %	629		

1. L'exemple (20) *je les... je les achète pas* [ʒəle // ʒlaʃɛtpa] (Ameringen, 1977, p. 16) a été retiré du compte car, comme le note l'auteur, il s'agit sans doute d'une reformulation du pluriel en un singulier (*je l'achète pas*).

dans Mtl71, PFC-Québec (sans distinction entre Montréal et hors Montréal) et le reste de PFC sans le Canada⁷. Les taux pour Mtl71 sont obtenus à partir des groupes I et II seulement, restreints ou élargis⁸. Pour PFCautre, nous n'avons pas distingué « *les* » déterminant versus pronom et *en* pronom versus

7. Pour PFC-autre, nous avons pris soin de retirer du compte des occurrences sans liaison les cas où le mot cible n'est en fait pas suivi d'un mot avec lequel la liaison est possible, comme dans *on va donc les*, *on va s/ tâcher de se rencontrer* (sgamc1), où le locuteur change de structure après le pronom *les*, ou *il y en a un à Saint-Denis* (75ccr1), où *un* est pronom et non déterminant. Ce type de cas n'a pas été annoté dans PFC-Québec. Les mêmes exclusions ont été appliquées aux contextes du tableau 1.4, où la non-liaison est rare dans PFC-autre.

8. Ameringen (1977) fournit normalement le nombre d'occurrences et le pourcentage de réalisation de la liaison. Les données pour les groupes I et II sont parfois regroupées, parfois mentionnées séparément; lorsque le groupe III est inclus, les résultats sont toujours livrés séparément. À partir des occurrences des différents groupes, nous avons pu compiler les résultats pour l'ensemble du corpus Mtl71, tels qu'ils apparaissent dans les tableaux 1.2 à 1.4 et 1.7. Par ailleurs, nous avons systématiquement recalculé les pourcentages de réalisation de la liaison, parfois légèrement erronés dans Ameringen (1977). Dans ces cas, nous avons indiqué la valeur corrigée et celle d'Ameringen entre crochets, par exemple, 40,6 % [41,8]. Les écarts constatés ne modifient jamais la nature des résultats.

préposition ; dans Mtl71, « *les* » déterminant est intégré au groupe des déterminants pluriels.

Ces cas correspondent à ce que nous avons appelé le noyau dur de la liaison catégorique (Côté, 2017). Durand et Lyche (2008) ont montré que les six premiers éléments du tableau déclenchent systématiquement la liaison dans toutes les variétés, ce que confirment les données de Mallet (2008). Si les taux n'atteignent généralement pas 100 %, c'est que la liaison peut ne pas s'appliquer dans des contextes de disjonction, qui bloquent différents processus de sandhi aux frontières de mots. Ces contextes se présentent notamment devant des mots dits à « h aspiré » (ex. *en haut*), devant certaines catégories de mots, comme les chiffres, les lettres et les noms propres (ex. *trois E, deux Émile*) et dans les usages autonymiques des Mots2 (ex. *dans « inexact », on trouve le suffixe « in »*) (pour une analyse unifiée de la disjonction, voir Abouda, Dugua et Enguehard, 2020). Le codage PFC de la liaison ne prévoit pas d'exclure les contextes de disjonction de l'annotation, en raison de la grande variabilité qu'ils peuvent manifester d'une variété, d'un individu ou d'un contexte à l'autre. Nous avons d'ailleurs noté que le nom des lettres, par exemple dans *les R*, bloquait systématiquement la liaison au Québec, mais pas dans PFC-autre ; il en est de même pour un emprunt comme *hamburger*. C'est l'analyse des cas de disjonction qui nous a permis d'ajouter la préposition *en* au noyau dur de la liaison catégorique (Côté, 2017), son taux de réalisation étant un peu moins élevé en raison du fait que les cas de disjonction y sont plus nombreux. Nous ne discuterons pas davantage des environnements du tableau 1.2, qui ne présentent pas d'intérêt du point de vue de la variation.

Liaison marginale dans Mtl71 et PFC

Cette catégorie comprend, à l'inverse de la précédente, quatre contextes, tous verbaux, avec des taux de réalisation très faibles dans l'ensemble des corpus (voir tableau 1.3). Les données de Mtl71 incluent ici les trois groupes, en mentionnant s'il s'agit des versions restreintes ou élargies des groupes I et II. Pour PFC-Québec, le tableau distingue cette fois-ci Montréal (PFC-Mtl) du reste du Québec (PFC-Qc/Mtl). Les formes considérées sont les mêmes que celles indiquées par Ameringen (1977) : *avais, avait, avaient* et *étais, était, étaient* pour l'imparfait de *avoir* et de *être*, *ont* pour le présent de *avoir* et *dois, doit, doivent, veux, veut, veulent, peux, peut, peuvent, faut, vont* pour les verbes modaux.

Deux précisions sur ces données s'imposent. Premièrement, dans tous les cas de liaison réalisée après un verbe modal au Québec, le verbe est suivi

TABEAU 1.3

Contextes de liaison marginale dans Mtl71 et PFC

	Mtl71		PFC-Mtl		PFC-Qc/Mtl		PFC-autre	
	%	n	%	n	%	n	%	n
Imparfait de <i>être</i>	7,5 %	159 (restreints)	6,4 %	514	1,5 %	271	5,2 %	1487
Imparfait de <i>avoir</i>	0 %	50 (restreints)	0 %	280	0 %	252	2,0 %	599
<i>ont</i>	1,7 %	58 (élargis)	1,1 %	91	0 %	73	4,2 %	238
Verbes modaux	13,3 %	60 (élargis)	2,6 %	155	1,5 %	132	2,9 %	550

de l'infinitif *être* (ex. *ça doit* [t] *être*). En fait, on observe aussi des [t] de liaison devant *être* dans des contextes où le mot précédent ne permet normalement pas la liaison, comme dans *ça va* [t] *être*. De tels exemples ont été retirés des données analysées ici (car non considérés par Ameringen, 1977), mais ils nous permettent de conclure que la liaison est déclenchée par *être* et qu'elle n'est autrement pas réalisée après les verbes modaux. Le taux de 13,3 % indiqué pour Mtl71 ne représente donc pas la productivité de la liaison après un verbe modal.

Les résultats concernant les formes à l'imparfait de *être* présentent une configuration intéressante. Dans Mtl71, les 12 cas de liaison sont le fait d'un seul locuteur, professeur de français à l'université et appartenant au groupe III. Dans PFC-Québec, les 37 occurrences de liaison proviennent de 5 personnes (une de Sainte-Adèle, quatre de Montréal), nées entre 1927 et 1938 et pour qui le taux de réalisation de la liaison dans ce contexte se situe entre 30 % et 59 %. Il semble donc que la liaison à l'imparfait de *être* ait déjà été productive pour une minorité de locuteurs québécois, mais qu'elle ait essentiellement disparu. Qui plus est, il ne s'agit pas d'une liaison nécessairement formelle, même si le locuteur qui l'utilise dans Mtl71 appartient au groupe le plus intégré au marché linguistique⁹. D'une part, la liaison est dans tous les cas [t], y compris après *étais*, et la forme (c')*était* avec [t] de liaison est souvent réduite à [stet]. D'autre part, le bagage éducatif et le statut socio-économique des utilisateurs de cette liaison dans PFC-Québec sont variables. Enfin, le fait que nous n'ayons trouvé aucune occurrence à Saint-Éphrem ou

9. Ce locuteur a aussi produit la liaison après *ont* dans le tableau 1.3.

Saguenay peut suggérer que cette liaison est plutôt caractéristique de l'ouest du Québec ou de la (très) grande région montréalaise; il faudra analyser les autres points d'enquête de PFC-Québec pour répondre à cette question.

Dans l'ensemble, on constate que la liaison après des verbes autres que *être* au présent est très marginale au Québec et qu'elle n'a aucunement progressé depuis les années 1970. Dans PFC-autre, la liaison est également rare dans tous ces contextes verbaux¹⁰. Elle est aussi largement concentrée lexicalement ou géographiquement. Ainsi, avec les formes de *être* et *avoir* à la troisième personne (singulier et pluriel), la liaison s'avère largement localisée dans le midi de la France, alors que la liaison après *étais* apparaît dans 13 cas sur 16 dans l'enquête de Bangui et dans la séquence *j'étais* [z] *allé(e)*. Comme au Québec, la liaison après un verbe modal précède le plus souvent l'infinitif *être* (13 cas sur 16 également).

Liaison plus fréquente dans PFC que dans Mtl71

Il s'agit de trois éléments : *ils*, *on* et les prépositions *dans*, *chez* et *sans*, toujours regroupées dans les résultats de Ameringen (1977) (voir tableau 1.4). Les données de Mtl71 incluent les trois groupes de locuteurs. Pour les prépositions, Ameringen a écarté de son codage les occurrences devant un nom propre (ex. *dans Ahuntsic*, *chez Adèle*), qui excluent la liaison. Nous avons fait de même pour les données PFC.

TABLEAU 1.4

Contextes où la liaison est plus fréquente dans PFC que dans Mtl71

	Mtl71		PFC-Mtl		PFC-Qc/Mtl		PFC-autre	
	%	n	%	n	%	n	%	n
<i>ils</i>	5,2 %	286 (restreints)	11,4 %	466	3,2 %	467	98,3 %	1210
<i>on</i>	57,1 %	499 (restreints)	96,5 %	1323	84,7 %	1103	99,9 %	3712
<i>dans, sans, chez</i>	51,2 %	82 (élargis)	88,3 %	196	61,7 %	133	92,4 %	602

10. Le taux pour *ont* dans le tableau 1.3 est inférieur à celui de Mallet (2008) car six des seize cas rapportés avec liaison avaient été mal codés. Pour les verbes modaux, nous avons aussi retiré des cas avec liaison les 26 occurrences de *peut-être* annotées avec liaison (qu'on peut considérer comme complètement lexicalisée ici).

Dans les trois cas, on observe une hausse de la liaison entre Mtl71 et PFC-Québec, où Montréal est toujours en avance sur le reste du Québec. *Ils* offre l'une des différences les plus spectaculaires entre la francophonie européenne et africaine, où la liaison est presque catégorique¹¹, et le Québec, qui a conservé la forme historique sans liaison. *Ils* y est donc normalement prononcé [j] devant voyelle (et [i] devant consonne, comme dans d'autres variétés). Le taux de liaison varie cependant de façon significative entre les trois corpus québécois ($\chi^2(2) = 25,93$, $p < 0,0001$) et la valeur des résidus standardisés indique une liaison moins fréquente dans PFC-Québec hors Montréal (-2,98) et plus fréquente dans PFC-Montréal (+3,78), avec respectivement des taux de 3,2 % et 11,4 %. Ce dernier taux reste cependant encore très loin du taux observé dans le reste de la francophonie. Mtl71 présente un taux intermédiaire de 5,2 %. Mais les taux globaux masquent des comportements plus subtils. Dans Mtl71, les membres des groupes I et II ne réalisent la liaison que dans 1,8 % des cas, alors que ce taux grimpe à 59 % pour les trois locuteurs du groupe III. La liaison semble donc être l'apanage de personnes très favorisées. Dans PFC-Montréal, pas moins de 11 des 29 participants, répartis également entre les groupes d'âge, présentent des taux de liaison assez élevés, entre 17 % et 50 %, mais pas autant que les trois locuteurs du groupe III de Mtl71. La liaison avec *ils* semble donc s'être quelque peu démocratisée, même si elle reste minoritaire. Hors Montréal, quatre locuteurs sur 36, tous âgés, ont un taux situé entre 18 % et 71 %. On n'observe donc pas de hausse chez les jeunes.

Pour *on*, la situation est plus nette. Globalement, l'absence de liaison après *on* a beaucoup régressé entre Mtl71 (57,1 % de liaison) et PFC-Montréal (96,5 %), le reste du Québec présentant un taux intermédiaire de 84,7 % ($\chi^2(2) = 450,23$, $p < 0,0001$, la valeur des résidus standardisés confirmant que la liaison est significativement plus fréquente dans PFC-Montréal, et moins fréquente dans Mtl71). À l'intérieur de Mtl71, la liaison après *on* varie beaucoup entre les trois groupes, passant du simple au double entre les groupes I et III (voir tableau 1.5). Ameringen (1977) observe aussi que les jeunes de son échantillon (moyenne d'âge de 21,6 ans) font plus souvent la liaison que les personnes âgées (moyenne de 71 ans) dans les groupes I et II confondus (40,6 % [41,8] contre 31,8 % [32,0]), mais cette différence est moins importante que celle entre les groupes. Dans PFC, la progression de la liaison est aussi plus avancée chez les jeunes, tant à Montréal que hors Montréal, ce qui

11. Treize des 21 cas sans liaison proviennent du point d'enquête du Domfrontais, en Normandie, où la non-liaison historique après *ils* s'est conservée, ce qui rappelle la situation québécoise. À la différence du Québec, *ils* sans liaison devant voyelle se prononce [il] (et non [j]) et cette forme n'est produite que par des locuteurs de plus de 50 ans.

TABLEAU 1.5

Réalisation de la liaison après *on* dans Mtl71 et PFC-Québec

	%	n	%	n	%	n
Mtl71	Groupe I		Groupe II		Groupe III	
	40,3%	258	70,9% [71,0]	148	81,7% [81,5]	93
PFC	Âgés		Moyens		Jeunes	
Montréal	93,8%	465	97,2%	528	99,4%	330
hors Montréal	80,5%	507	82,0%	334	96,2%	262

appuie l'idée d'un changement en cours vers la catégoricité. Les Québécois rejoignent donc sur ce point le reste de la francophonie.

Les prépositions *dans*, *chez* et *sans* offrent une situation comparable avec la hiérarchie PFC-Montréal > PFC-hors Montréal > Mtl71 ($\chi^2(2) = 50,63$, $p < 0,0001$, la valeur des résidus standardisés confirmant que la liaison est significativement plus fréquente dans PFC-Montréal et moins fréquente dans Mtl71). Dans Mtl71, la distinction entre les groupes est massive, passant de l'absence quasi catégorique de la liaison dans le groupe I à sa présence systématique dans le groupe III (voir tableau 1.6). Les données de PFC-Québec montrent une progression considérable de la liaison à Montréal (88,3 % pour l'ensemble des locuteurs), le reste du Québec accusant un certain retard (61,7 %); les jeunes font un peu plus souvent la liaison que les autres, mais l'écart entre les groupes d'âge reste faible.

Chaque préposition présente cependant une configuration un peu différente. Ces résultats sont dominés par *dans*, qui est six fois plus fréquent que *chez* et treize fois plus que *sans* dans PFC-Québec. Ils masquent donc le

TABLEAU 1.6

Réalisation de la liaison après *dans*, *chez*, *sans* dans Mtl71 et PFC-Québec

	%	n	%	n	%	n
Mtl71	Groupe I		Groupe II		Groupe III	
	10,3 %	39	48,3 % [50,0]	29	100 %	14
PFC	Âgés		Moyens		Jeunes	
Montréal	88,9%	63	85,5%	76	91,2%	57
hors Montréal	61,9%	42	59,4%	32	62,7%	59

comportement particulier des deux dernières prépositions. Dans PFC-Québec, on constate une distinction catégorique entre les formes lexicalisées *chez eux/elle(s)*, où la liaison est toujours réalisée (25/25), et *chez un(e)*, où elle n'est produite qu'une fois sur 10. Sur ce point, les données québécoises se démarquent de l'usage européen, où la liaison est nettement majoritaire dans le contexte *chez un(e)*. Pour *sans*, le taux de liaison dans PFC-autre est de 87,5 % (35/40) (mais quatre des cinq réalisations sans liaison proviennent d'un même locuteur de l'enquête du Domfrontais). Dans PFC-Québec, une autre distinction s'opère entre *sans*+Nom (ex. *sans arrêt*), avec liaison catégorique (11/11), et *sans*+Verbe (ex. *sans être*), où la liaison est très minoritaire (2/9)¹²; l'explication de cette distinction nous échappe. On peut conclure que *chez a* en fait un comportement catégorique selon le contexte et que la liaison après *sans* n'est variable que devant un verbe. Pour ce qui est de *dans*, la liaison est certainement variable dans Mtl71; elle l'est également dans PFC-Québec hors Montréal (59,4 % de liaison, 60/101), mais devient quasi catégorique dans PFC-Montréal (90,2 % de liaison, 156/173, le taux passant de 88,2 % dans le groupe âgé à 92,5 % chez les jeunes).

Liaison plus fréquente dans Mtl71 que dans PFC

Dans ces contextes, la liaison est variable partout, mais plus fréquente au Québec que dans le reste de la francophonie. Il s'agit des formes *est* et *sont* et des adjectifs prénominaux. Dans le tableau 1.7, les données de Mtl71 incluent les trois groupes pour *est* et *sont*, mais uniquement les groupes I et

TABLEAU 1.7

Contextes où la liaison est plus fréquente dans Mtl71 que dans PFC

	Mtl71		PFC-Mtl		PFC-Qc/Mtl		PFC-autre	
	%	n	%	n	%	n	%	n
adj. singulier	93,3 %	15 (élargis)	90,2 %	41	84 %	25	82 %	255
adj. pluriel	89,7 %	29 (élargis)	88,5 %	113	70,4 %	54		
<i>est</i>	95,9 %	170 (restreints)	75,3 %	860	81,4 %	645	27,8 %	2507
<i>sont</i>	100 %	24 (restreints)	70,3 %	64	67,3 %	49	16,8 %	286

12. Cette distribution se confirme dans une partie de PFC-Québec élargie à 206 individus : 6/28 liaisons réalisées dans *sans*+Verbe et 31/32 dans *sans*+Nom.

II élargis pour les adjectifs. Les données de PFC-autre pour les adjectifs proviennent de Barreca (2015), qui a enrichi le codage liaison d'annotations morphosyntaxiques permettant d'identifier les séquences Adj+Nom. Ses données ne distinguent cependant pas les adjectifs au singulier et au pluriel, et les 38 points d'enquête considérés incluent trois points d'enquête canadiens qu'il n'est pas possible d'isoler du reste des données.

La différence entre le Québec et PFC-autre est moins flagrante pour les adjectifs prénominaux que pour *est* et *sont*. Dans le tableau 1.7, on constate, pour les adjectifs, une légère baisse entre Mtl71 et PFC-Québec, les taux de liaison se rapprochant de ceux de PFC-autre. Pour les adjectifs, les quatre corpus ne se distinguent cependant pas statistiquement au seuil de 0,01 ($\chi^2(3) = 10,05$, $p = 0,0181$), la valeur des résidus standardisés indiquant que la liaison n'est significativement différente dans aucun des corpus. Les adjectifs prénominaux forment cependant une classe très hétérogène, avec des comportements lexicalement spécifiques, et les taux globaux masquent des régularités plus fines. Au singulier, la liaison est régulière avec *petit* (8/9 dans PFC-Québec, hors trois cas de disjonction), mais passablement moins avec d'autres adjectifs, comme le montre Sampson (2001) pour la liaison en [n]. Malheureusement, les corpus oraux à disposition sont de taille trop réduite pour permettre une analyse par adjectif, en raison de la faible fréquence de la plupart d'entre eux en contexte de liaison. Au pluriel, 16 des 29 occurrences de non-liaison dans PFC-Québec touchent deux contextes particuliers (en plus de trois cas de disjonction) :

- les cinq occurrences de *mauvaises herbes*, où la non-liaison résulte vraisemblablement de la lexicalisation d'une haplogogie (la consonne finale de *mauvaises* étant identique à la consonne de liaison attendue) ;
- onze occurrences de *autres* (ex. *d'autres émissions*), dont le taux de réalisation de la liaison est de 52 % dans PFC-Québec, proche de celui observé dans le reste du corpus PFC (60,6 %, 20/33).

Au-delà des taux globaux, quelques faits particuliers suggèrent que le contexte Adj+Nom favorise davantage la liaison au Québec. D'abord, PFC-Québec présente des taux de réalisation de 100 % après *grand* et *gros*, contre 67 % pour *grand*+Nom (Barreca, 2015, p. 87) et 88 % pour *gros* (Durand et Lyche, 2008) dans PFC général. On observe aussi dans PFC-Québec des cas d'insertion de consonnes après des adjectifs qui, selon la norme, excluent la liaison ou commandent une consonne différente, comme *vrai* [t] *investissement*, *vrai* [t] *effort*, *gros* [t] *hôtel* et *super* [z] *emplois*.

Est et *sont* présentent l’une des situations les plus intéressantes. Alors que la liaison est presque catégorique dans Mtl71, elle est très minoritaire dans PFC hors Canada, avec des taux de 27,8 % pour *est* (après retrait des liaisons annotées dans l’expression *c’est-à-dire*, où le [t] est lexicalisé) et 16,8 % pour *sont*. PFC-Québec occupe une position intermédiaire et, comme pour *ils*, *on* et les prépositions, PFC-Montréal s’éloigne davantage de Mtl71 que le reste du Québec ($\chi^2(2) = 45,54$, $p < 0,0001$, la valeur des résidus standardisés indiquant que la liaison est significativement plus fréquente dans Mtl71 et la non-liaison plus fréquente dans PFC-Mtl). On peut de nouveau faire l’hypothèse que les taux de liaison au Québec tendent vers ceux en usage en Europe, mais cette fois-ci dans le sens d’une diminution plutôt que d’une augmentation de la liaison. Le tableau 1.8 éclaire la situation en séparant PFC-Québec en groupes d’âge et en distinguant le présentatif *c’est* des autres usages de *est* (ex. *on/il/qui est*), rassemblés sous l’étiquette *(*c’est)*.

Deux tendances ressortent du tableau. Alors que la réalisation de la liaison après *(*c’est)* est stable autour de 90 % tant à Montréal que hors

TABLEAU 1.8
Réalisation de la liaison après *est* dans PFC-Québec

	PFC-Montréal						PFC-Québec hors Montréal					
	Âgés		Moyens		Jeunes		Âgés		Moyens		Jeunes	
	%	n	%	n	%	n	%	n	%	n	%	n
<i>(*c’est)</i>	96,6 %	118	92,8 %	125	87,2 %	78	88,6 %	88	92,0 %	100	88,5 %	87
<i>c’est</i>	70,4 %	152	62,1 %	240	49,5 %	194	83,2 %	107	67,9 %	106	64,6 %	189
<i>sont</i>	93,8 %	16	72,0 %	25	52,2 %	23	75,0 %	8	87,5 %	16	52,0 %	25

Montréal (même si on observe une légère augmentation avec l’âge à Montréal), le taux de liaison après *c’est* augmente nettement avec l’âge et il est dans chaque tranche d’âge plus élevé hors Montréal qu’à Montréal. Dans un groupe de 21 points d’enquête du corpus PFC, Mallet (2008) rapporte également un taux plus élevé après *(*c’est)* (43,9 %) qu’après *c’est* (28,1 %). Comme pour *c’est*, la liaison après *sont* augmente légèrement avec l’âge dans les deux zones. Cela suggère que la chute de la liaison après *est* et *sont* correspond en effet à un changement en cours dans le sens de l’usage européen et que cette tendance est plus avancée dans la métropole montréalaise. La chute après *est* est essentiellement due à celle de *c’est*.

Liaison s'écartant de la norme prescriptive dans Mtl71

La dernière catégorie de résultats concerne la liaison en [t] après les formes du verbe *être suis* et *es*. Cette liaison tout à fait régulière au Québec s'écarte de la norme prescriptive et on ne l'observe pas dans PFC hors Canada. Elle est donc susceptible de subir deux types de pression : celle de la norme, qui commanderait plutôt une liaison en [z], et celle de l'usage dominant en francophonie, qui exclut la liaison dans 89,1 % des cas dans PFC-autre. Le tableau 1.9 est organisé un peu différemment des précédents, puisque trois variantes sont en concurrence. Il fait émerger plusieurs tendances dans un portrait complexe. Ici les taux globaux pour les trois corpus québécois ne sont guère informatifs et il faut distinguer d'emblée les différents groupes à l'intérieur de chaque corpus. Dans Mtl71, la liaison en [t] est presque catégorique pour les groupes I et II, mais inexistante pour le groupe III, qui alterne entre la liaison en [z] et l'absence de liaison ; il faut toutefois noter que les quatre occurrences de [z] sont produites par le même locuteur, professeur de français à l'université, qui a réalisé les liaisons après *ont* et les formes à l'imparfait de *être* (voir la section « Liaison marginale dans Mtl71 et PFC »). Dans PFC-Québec, la liaison en [z], globalement marginale, est essentiellement le fait des personnes âgées et les jeunes ne l'utilisent pas du tout : dans l'ensemble de la province, le taux de liaison en [z] passe de 10,9 % chez les plus âgés à 1,8 % pour les personnes d'âge moyen et 0,5 % pour les jeunes. Fait plus surprenant, on observe pour la liaison en [t] des trajectoires opposées à Montréal et dans le reste du Québec. À Montréal, la liaison en [t] est très minoritaire dans le groupe âgé, mais elle augmente chez les personnes d'âge moyen et les jeunes. Hors Montréal, elle est majoritaire dans les deux groupes les plus âgés, puis elle diminue chez les jeunes. Les jeunes des deux zones convergent donc vers des taux de liaison similaires proches de 50 %, très inférieurs au taux de 90,2 % observé dans les groupes I-II de Mtl71.

TABLEAU 1.9

Réalisation de la liaison après *suis* et *es* dans Mtl71 et PFC

	Mtl71 n=49		PFC-Montréal n=251			PFC-Qc/ Montréal n=145			PFC- autre n=725
	gr. I+II n=41	gr. III n=8	Âgés n=87	Moyens n=83	Jeunes n=81	Âgés n=23	Moyens n=26	Jeunes n=96	
[z]	0 %	50,0 %	11,5 %	1,2 %	1,2 %	8,7 %	3,8 %	0 %	10,9 %
[t]	90,2 %	0 %	14,9 %	27,7 %	46,9 %	60,8 %	65,4 %	49,0 %	0 %
Ø	9,8 %	50,0 %	73,6 %	71,1 %	51,9 %	30,4 %	30,8 %	51,0 %	89,1 %

Discussion

Cette comparaison entre Mtl71 et PFC est nécessairement limitée par les contextes considérés pour les analyses du corpus de 1971, qui ne bénéficiaient évidemment pas des avancées récentes sur la liaison. Un bel éventail de résultats émerge tout de même. On note d'abord le fait que les taux de réalisation de la liaison se concentrent nettement aux deux extrémités de l'échelle : 90 % et plus ou 10 % et moins. Seuls quatre contextes présentent des taux intermédiaires dans l'un ou l'autre des corpus, en distinguant Mtl71 (groupes I+II), PFC-Montréal et PFC-Québec hors Montréal. Il s'agit de¹³ :

- *ils* : la non-liaison est quasi catégorique en 1971 et dans PFC hors Montréal, mais elle présente un début de variabilité (11 %) dans PFC-Montréal ;
- *on* et *dans*¹⁴ : la liaison est variable dans Mtl71 et dans PFC-Québec hors Montréal, mais elle devient catégorique dans PFC-Montréal ;
- présent de *être* : alors que toutes les formes font catégoriquement la liaison en [t] dans Mtl71, la liaison devient variable dans les deux zones de PFC-Québec.

Dans tous ces cas, la tendance est à la convergence vers les usages européens : diminution de la liaison après les formes de *être* au présent, augmentation dans les autres cas. Montréal précède aussi le reste du Québec dans ces évolutions, ce qui correspond à la tendance générale pour la diffusion des traits linguistiques au Québec (Friesner, 2010). Sauf pour *ils* et *suis* dans PFC-Montréal, les jeunes sont en avance. La liaison après *ils* semble relever d'un comportement plus individuel, puisque seules certaines personnes y ont recours de façon plus ou moins régulière. Dans PFC-Montréal, il s'agit de 11 personnes de tous âges sur les 29 de l'échantillon. Assiste-t-on à une forme de démocratisation de la liaison avec *ils*, disponible comme variante pour une plus grande variété d'individus ? Il faudra aussi vérifier avec d'autres données si la liaison avec *ils* chez certains jeunes Montréalais annonce une montée plus générale.

13. Nous ne tenons pas compte de la liaison après des verbes modaux devant *être* (voir « Liaison marginale dans Mtl71 et PFC »).

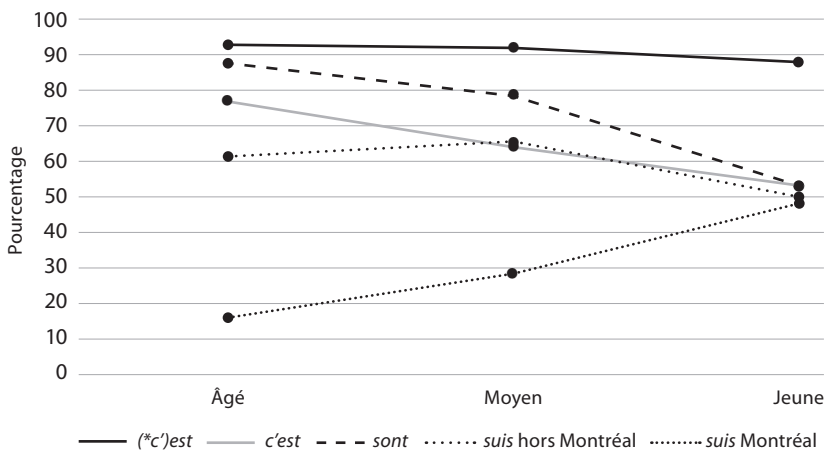
14. La liaison dans *sans*+Verbe est également variable, mais les données détaillées pour ce contexte dans Mtl71 ne sont pas disponibles et le nombre d'occurrences est de toute façon réduit.

Il y a pourtant un contexte de liaison en changement rapide au Québec qui ne s'aligne pas avec l'usage européen. Il s'agit de la liaison après *quand*, qui, même si elle est suffisamment fréquente, n'a pas été examinée par Ameringen (1977). *Quand* a la particularité d'alterner entre [kâ] et [kât] devant voyelle et devant consonne (ex. *quand* [t] *c'est devenu*). Morin (1990) conclut qu'il s'agit d'un cas de supplétion /kâ, kât/ et non de liaison. Devant voyelle et pour les 44 locuteurs dont l'entièreté des conversations a été annotée, la forme [kât] passe de 70,5 % (117/170) dans le groupe âgé à 45,1 % (65/144) chez les personnes d'âge moyen et à 5,6 % (5/89) chez les jeunes. Cette chute vertigineuse contraste avec le taux élevé de liaison après *quand* dans PFC hors Canada (75,9 %, 616/812).

Pour les formes au présent de *être*, la diminution globale de la liaison en [t] depuis un demi-siècle s'intègre à un portrait plus complexe qui distingue quatre formes, dont la tendance dans les données PFC-Québec apparaît dans la figure 1.1. Alors que la liaison après (*c')est se maintient à un niveau très élevé, celle après *c'est* et *sont* a régressé. *Je suis* présente chez les locuteurs âgés des taux plus faibles qu'après les autres formes, mais des tendances géographiques opposées : en progression à Montréal, les Montréalais plus âgés évitant clairement la liaison en [t], et en légère diminution hors Montréal. En fin de compte, les jeunes Québécois ont des taux de liaison comparables après *c'est*, *sont* et *suis*.

FIGURE 1.1

Pourcentage de réalisation de la liaison au présent de *être* dans PFC-Québec



Plusieurs facteurs expliquent ces tendances. L'alignement sur les usages dominants en francophonie intervient sans doute dans la diminution générale de la liaison après les formes de *être*. On note une chute plus importante de la liaison après *sont* qu'après *(*c')est*, deux formes de troisième personne qui apparaissent dans les mêmes environnements. Cela pourrait s'expliquer par la corrélation déjà établie entre la fréquence du mot déclencheur et la réalisation de la liaison. Dans un corpus orléanais, de Jong (1994, p. 109) observe par exemple pour *est* (69 %, n=1692), *sont* (46 %, n=200) et *soit* (11 %, n=37) la même hiérarchie de fréquence et de taux de réalisation.

Le poids de la norme prescriptive explique certainement le comportement particulier de la liaison en [t] après *suis*, globalement moins fréquente qu'après les autres formes. L'évitement est plus marqué chez les personnes âgées, surtout à Montréal; chez les jeunes, on observe une convergence entre toutes les formes sauf *(*c')est*. Il est possible que la pression normative qui s'oppose à la liaison en [t] ait été plus forte à Montréal, mais que les jeunes s'en soient partiellement détachés dans le contexte d'une réduction de l'insécurité linguistique dans les dernières décennies (Chalier, 2021). Mais il resterait à confirmer dans quelle mesure il s'agit d'un changement en cours ou d'un effet d'âge. L'absence d'augmentation de la liaison après *suis* chez les jeunes hors Montréal par rapport à leurs aînés peut suggérer qu'il ne s'agit pas uniquement d'un effet d'âge. Chose certaine, la liaison en [z] ne progresse pas.

Je suis et *c'est* impliquent aussi une alternance entre trois réalisations devant un mot à initiale vocalique, plutôt que deux pour les autres mots déclencheurs de liaison: [ʃ(s)qi]~[ʃy] et [se] sans liaison, [ʃ(s)qit]~[ʃyt] et [set] avec liaison, [ʃt] et [st] avec consonne de liaison et absence de la voyelle précédente. *C'est assez* est ainsi réalisé [sease], [setase] et [stase]. Nous avons analysé la distribution des trois variantes pour les 44 locuteurs dont l'entièreté des conversations a été annotée, dont 29 Montréalais. Pour *je suis*, les formes [ʃ(s)qit]~[ʃyt] sont complètement absentes à Montréal et disparaissent progressivement hors Montréal: 21 % (4/19) dans le groupe âgé, 5 % (1/19) chez les personnes d'âge moyen et 0 % (0/82) chez les jeunes. *C'est* connaît la même évolution, mais avec un peu de retard, et sans distinction entre les deux zones géographiques: la variante [set] passe de 35 % (82/235) dans le groupe âgé à 14 % (41/291) chez les personnes d'âge moyen et à 5 % (14/297) chez les jeunes. *Je suis* et *c'est* tendent donc à n'avoir plus que deux variantes: [ʃ(s)qi]~[ʃy] et [ʃt], [se] et [st]. La diminution de la liaison après *c'est*, donc la progression de la variante [se] (de 24 % dans le groupe âgé à

48 % chez les jeunes) se fait ainsi entièrement au détriment de [set] ; la variante [st] n'est pas affectée (âgés, 41 % ; moyens, 49 % ; jeunes, 48 %).

Plusieurs questions se posent : 1) comment ont émergé les variantes sans voyelle ? 2) comment analyser formellement ces variantes ? 3) pourquoi les variantes avec voyelle et consonne de liaison disparaissent-elles ? 4) quel impact l'évolution des variantes peut-elle avoir sur le comportement de la liaison ?

Les formes réduites sans voyelle sont caractéristiques de formes hyperfréquentes, comme *voilà* [vla] ou *peut-être* [ptɛtʁ]. L'élision de la voyelle dans *c'est* est courante en Europe dans l'expression *c'est-à-dire*, et elle est attestée en contexte de liaison : sur les 26 cas de liaison après *c'est*, relevés dans 6 points d'enquête de France septentrionale par le moteur de recherche du site PFC, nous avons noté un cas clair d'élision. Dans le cas de *je suis*, la variante [ʃt] au Québec résulte aussi de l'élision régulière des voyelles fermées en syllabe inaccentuée (Cedergren et Simoneau, 1985), qui plus est favorisée ici par leur dévoisement entre deux obstruantes sourdes. L'établissement de la variante [ʃt] par un processus de réduction phonétique régulier a alors pu soutenir, par analogie, celui de [st], nettement plus fréquent en français laurentien que dans d'autres variétés de français.

Quelle que soit leur origine, ces variantes sans voyelle rappellent une distinction, mieux connue dans les travaux sur la liaison, entre la liaison avec et sans dénasalisation de la voyelle. On oppose ainsi deux classes de mots : ceux comme *mon*, qui alternent entre [mɔ̃] devant consonne (ex. *mon voisin*) et [mɔ̃n] devant voyelle (ex. *mon ami*) et ceux, comme *bon*, qui alternent entre [bɔ̃] (ex. *bon voisin*) et [bɔn] (ex. *bon ami*). Dans le premier cas, seule la présence de la consonne de liaison distingue les formes préconsonantique et prévocalique ; dans le second cas, la qualité de la voyelle est également modifiée. Nous avons déjà proposé (Côté, 2005) que seuls les mots du type *mon* impliquent la liaison au sens strict, qui correspond à l'ajout d'une consonne. Les mots du type *bon* correspondent plutôt à un phénomène de supplétion entre deux formes lexicalisées /bɔ̃, bɔn/. La même logique s'appliquerait à *je suis* et *c'est* : les variantes sans voyelle sont dérivées par supplétion (/ʃsqi, ʃt/, /se, st/), les formes intermédiaires [ʃ(s)qit]~[ʃyt] et [set] par liaison à partir de /ʃsqi/ et /se/.

La marginalisation des variantes de liaison [ʃ(s)qit]~[ʃyt] et [set] pourrait relever d'une tendance générale à favoriser les variantes lexicalisées, au détriment de variantes intermédiaires dérivées, ou les variantes extrêmes lorsque plus de deux variantes forment un continuum, comme c'est le cas

ici¹⁵. La vulnérabilité de la variante de liaison de *c'est*, par opposition à celle de *(*c')est*, peut expliquer au moins en partie pourquoi la chute de la liaison a d'abord ciblé essentiellement *c'est* et épargné *(*c')est*. C'est ainsi qu'a émergé la distinction entre *c'est* et *(*c')est*, absente dans Mtl71 puisque la liaison y était quasi catégorique dans les deux contextes, mais déjà observée dans PFC général. Les données de PFC-Montréal au tableau 1.8 suggèrent que la perte de la liaison commence maintenant à toucher *(*c')est*. Les usages québécois se rapprochent donc des usages européens à la fois par la régression de la liaison et par l'émergence d'un contraste entre *(*c')est* et *c'est*. Enfin, la lexicalisation des variantes [ʃt] et [st] est susceptible de favoriser leur autonomie par rapport à [ʃ(s)qi]~[ʃy] et [se] et peut avoir contribué à la montée récente de [ʃt] à Montréal.

* * *

La complexité des données fait ressortir l'hétérogénéité du phénomène. La liaison ne se laisse pas embouteiller dans quelques grands contextes définis par des catégories lexicales comme adjectif ou verbe. Chaque forme est susceptible d'avoir son histoire, déterminée par les caractéristiques spécifiques des contextes dans lesquels elle apparaît. Pour plusieurs catégories, nous avons ainsi établi des distinctions, entre *c'est* et *(*c')est*, entre *être* et d'autres infinitifs après des verbes modaux, entre *chez eux/elle(s)* et *chez un(e)*, entre *sans*+Nom et *sans*+Verbe; d'autres émergeraient certainement d'une analyse plus poussée des données. Ces distinctions sont essentielles à la compréhension du phénomène et, de façon générale, elles ont comme effet de confiner la variabilité à des contextes plus précis. Ainsi, la liaison après un verbe modal n'est plus variable après le retrait des séquences modal+*être*. De même, le taux global de 63 % observé pour la préposition *sans* dans PFC-Québec se révèle trompeur lorsqu'on constate une différence de comportement marquée devant un nom et un verbe. C'est possiblement notre conception, bien établie, de la liaison variable qui est en jeu ici.

15. De telles tendances pourraient par exemple rendre compte du fait, observé dans PFC-Québec, qu'une séquence préposition+déterminant comme *sur la* alterne essentiellement entre les formes lexicalisées [syrla] et [sa:] (Baronian, 2006), alors que les formes intermédiaires [syla] et [sq̣a] sont beaucoup moins fréquentes.

Références

- ABOUDA, Lotfi, Céline DUGUA et Guillaume ENGUEHARD (2020), « À propos de quelques exceptions aux règles de la liaison et de l'élision », *7^e Congrès mondial de linguistique française, SHS Web Conf.*, 78, <https://doi.org/10.1051/shsconf/20207809010>.
- AMERINGEN, Arie van (1977), *La liaison en français de Montréal*, mémoire de maîtrise, Université du Québec à Montréal.
- AMERINGEN, Arie van et Henrietta J. CEDERGREN (1981), « Observations sur la liaison en français de Montréal », dans David SANKOFF et Henrietta J. CEDERGREN (dir.), *Variation omnibus*, Linguistic Research, Edmonton, 141-149.
- BARONIAN, Luc (2006), « Preposition contractions in Quebec French », dans Patrick SAINT-DIZIER (dir.), *Syntax and Semantics of Prepositions*, Springer, Dordrecht, 27-42.
- BARRECA, Giulia (2015), *L'acquisition de la liaison chez des apprenants italophones : des atouts d'un corpus de natifs pour l'étude de la liaison en français langue étrangère (FLE)*, thèse de doctorat, Università cattolica del Sacro Cuore di Milano et Université Paris Ouest Nanterre La Défense.
- CEDERGREN, Henrietta et Louise SIMONEAU (1985), « La chute des voyelles hautes en français de Montréal : "As-tu entendu la belle syncope?" », dans Monique LEMIEUX et al. (dir.), *Les tendances dynamiques du français parlé à Montréal*, Office québécois de la langue française, Montréal, tome 1, 57-144.
- CHALIER, Marc (2021), *Les normes de prononciation du français : une étude perceptive panfrancophone*, De Gruyter, Berlin.
- CÔTÉ, Marie-Hélène (2005), « Le statut lexical des consonnes de liaison », *Langages*, 158, 66-78.
- CÔTÉ, Marie-Hélène (2014), « Le projet PFC et la géophonologie du français laurentien », dans Jacques DURAND et al. (dir.), *La phonologie du français : normes, périphéries, modélisation*, Presses universitaires de Paris Ouest, Nanterre, 175-198.
- CÔTÉ, Marie-Hélène (2017), « La liaison en diatopie : esquisse d'une typologie », *Journal of French Language Studies*, 27, 13-25.
- CÔTÉ, Marie-Hélène (2025), « Annoter ou ne pas annoter : vers une définition des contextes de liaison », *Langages*, 238, 27-44.
- CÔTÉ, Marie-Hélène et Hugo Saint-Amant Lamy (2023), « The "Phonologie du français contemporain" project in Quebec: methodological and dialectometric considerations », dans Elissa PUSTKA, Carmen Quijada Van den BERGHE et Verena WEILAND (dir.), *Corpus Dialectology*, John Benjamins, Amsterdam/Philadelphie, 60-83.
- DE JONG, Daan (1993), « Sociophonological aspects of Montreal French liaison », dans William J. ASHBY et al. (dir.), *Linguistic Perspectives on the Romance Languages. Selected papers from LSRL XXI*, John Benjamins, Amsterdam/Philadelphie, 127-137.
- DE JONG, Daan (1994), « La sociophonologie de la liaison orléanaise », dans Chantal LYCHE (dir.), *French Generative Phonology: Retrospective and Perspectives*, ESRI, Salford, 95-130.
- DELATTRE, Pierre (1951), *Principes de phonétique française à l'usage des étudiants anglo-américains*, Middlebury College, Middlebury, VT.

- DURAND, Jacques et Chantal LYCHE (2008), « French liaison in the light of corpus data », *Journal of French Language Studies*, 18, 33-66.
- DURAND, Jacques *et al.* (2011), « Que savons-nous de la liaison aujourd'hui ? », *Langue française*, 169, 103-135.
- DURAND, Jacques, Bernard LAKS et Chantal LYCHE (2002), « La phonologie du français contemporain: usages, variétés et structure », dans Claus D. PUSCH et Wolfgang RAIBLE (dir.), *Romanistische Korpuslinguistik. Korpora und gesprochene Sprache/Romance Corpus Linguistics. Corpora and Spoken Language*, Narr, Tübingen, 93-106.
- FRIESNER, Michael (2010), « Une prononciation "tsipéquement québécoise" ? : la diffusion de deux aspects stéréotypés du français canadien », *Revue canadienne de linguistique*, 55, 27-53.
- MALLET, Géraldine (2008), *La liaison en français: descriptions et analyses dans le corpus PFC*, thèse de doctorat, Université Paris Ouest Nanterre La Défense.
- MORIN, Yves-Charles (1990), « La prononciation de [t] après *quand* », *Linguisticae Investigationes*, XIV(1), 175-189.
- MÜLLER, Daniela et Marinus WIEDNER (2022), « Le schwa dans l'accent français méridional », 8^e *Congrès mondial de Linguistique française*, SHS Web Conf., 138, <https://doi.org/10.1051/shsconf/202213808006>.
- SAMPSON, Geoffrey (2001), « Liaison, nasal vowels and productivity », *Journal of French Language Studies*, 11, 241-258.
- TOUSIGNANT, Claude (1978), *La liaison consonantique à Montréal*, mémoire de maîtrise, Université de Montréal.

CHAPITRE 2

Convergence ou divergence en français laurentien

Quarante ans de changement
à Montréal et à Welland

*Hélène Blondeau, Claire Bourély, Raymond Mougeon
et Mireille Tremblay*

Cette étude examine le changement linguistique dans les français de Montréal et de Welland (Ontario), deux variétés laurentiennes évoluant dans des contextes démo-linguistiques différents. Nous évaluons dans quelle mesure ces deux variétés montrent une tendance à la convergence ou à la divergence. Notre étude s'appuie sur des corpus recueillis sur une période de 40 ans : une première fois dans les années 1970 (Montréal, 1971 et Welland, 1975) et une seconde fois, 40 ans plus tard (Montréal, 2012 et Welland, 2012).

Le corpus Sankoff-Cedergren recueilli en 1971 est le premier d'une série de corpus ayant servi de fer de lance au développement de la sociolinguistique au Québec et au Canada. Le recueil de ce premier corpus avait pour but de « contribuer à une meilleure compréhension du français parlé au Québec en considérant ses aspects propres, non comme des erreurs ou aberrations ou encore en matière de mélange non structuré d'anomalies grammaticales, mais en tant qu'éléments d'un système cohérent partagé par tous les membres de la communauté » (Sankoff *et al.*, 1976, 88-89). À peu près à la même période (1975), un autre corpus a été recueilli à Welland en Ontario, où réside une communauté francophone principalement issue de vagues migratoires provenant du Québec (Beniak *et al.*, 1985). Selon ces auteurs, la

description du français parlé à Welland devait nous permettre d'en savoir plus « sur le sort que connaît le français québécois une fois transplanté en milieu canadien anglais », et de « faire avancer la réflexion théorique » relativement aux problématiques « du contact des langues » et de « la restriction de l'usage d'une langue minoritaire » (Beniak *et al.*, 1985, 2). Ensemble, ces corpus constituent une ressource inestimable, ce sont des jalons permettant de mesurer non seulement l'évolution du français parlé à Montréal et Welland, mais aussi, d'une façon plus générale, celle des variétés de français au sein de la famille laurentienne, qui comprend le français québécois et les autres variétés qui en sont issues (Côté, 2020).

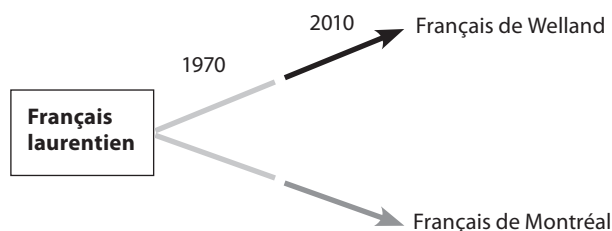
Parmi les études antérieures qui ont examiné la question de l'évolution des variétés de français laurentien hors Québec, on peut mentionner celles de Mougeon *et al.* (2009; 2016). La première portait sur l'alternance *j'vais/j'vas/m'as* et la seconde sur l'alternance *juste, seulement (que), rien que et (ne) ... que*. Dans ces deux études, le corpus de Welland (1975) est comparé à trois autres corpus recueillis dans les provinces de l'Ouest canadien durant les années 1970 ou au début des années 1980. La comparaison a révélé qu'à cette époque, ces quatre variétés de français laurentien partageaient le même inventaire de variantes et nombre des contraintes linguistiques et extralinguistiques qui pesaient sur leur emploi. Toutefois, ces deux études comparatives ont aussi mis au jour plusieurs aspects d'une tendance inverse : la divergence inter-variétés. Par exemple, Mougeon *et al.* (2009) ont trouvé qu'à Welland, l'âge n'avait pas d'effet sur l'alternance *j'vais/j'vas/m'as*, alors qu'à Saint-Boniface (Manitoba), on décelait des indices de changement en cours (montée de *je vais* et forte diminution de *m'as*). Enfin, les deux études ont révélé que la tendance à la convergence inter-variétés était plus forte dans le cas de l'alternance *juste, seulement, rien que et ne... que* que dans celui de l'alternance *j'vais/j'vas/m'as*. Enfin, dans une étude récente, Blondeau *et al.* (2022) ont réexaminé en diachronie courte la question de la convergence/divergence en français laurentien, ce que n'avaient pu faire les études de Mougeon *et al.* L'étude de Blondeau *et al.* (2022) est consacrée à la variation dans l'emploi des marqueurs de conséquence à Montréal et Welland et repose sur les corpus recueillis durant les années 1970 et 2010 dans ces deux villes. Nous reviendrons sur les résultats de Blondeau *et al.* dans ce chapitre. Contentons-nous ici de les résumer brièvement. L'analyse des données des années 1970 apporte un soutien relatif aux résultats des études de Mougeon *et al.* En effet, si Montréal et Welland ont en commun l'emploi des marqueurs consécutifs (*ça*) *fait (que), alors et donc*, à Welland, on utilise aussi la

conjonction anglaise *so*, une forme qui était alors presque uniquement employée par les moins de 35 ans. De plus, on constate que, si certains des facteurs extralinguistiques ont le même effet sur la variation dans les deux communautés, les autres facteurs ont un effet divergent. Par exemple, dans les deux communautés, *alors* est associé à la catégorie socioéconomique (CSE) haute, mais l'influence de l'âge sur ce marqueur est divergente. À Montréal, on observe un modèle de gradation selon l'âge tandis qu'à Welland, l'âge n'a pas d'influence significative sur l'emploi de cette forme. La comparaison des données des années 1970 avec celles des années 2010 a mis au jour une intensification de la divergence inter-variétés. Par exemple, dans les années 2010, à Montréal, la conjonction *so* est toujours absente, tandis qu'à Welland, cette forme rivalise désormais avec la variante traditionnelle (*ça*) *fait* (*que*). Par ailleurs, à Montréal, cette dernière est en passe de devenir le marqueur de conséquence par défaut. Somme toute, l'étude de Blondeau *et al.* (2022) a mis en avant l'intérêt d'élargir à la fois la profondeur temporelle de la recherche sur la convergence/divergence au sein de la famille des français laurentiens et l'éventail des cas de variation examinés.

Ce chapitre est pour nous l'occasion de poursuivre cet effort d'élargissement de la réflexion sur la convergence/divergence entre les français laurentiens en comparant l'évolution de l'usage des marqueurs de conséquence à celle des formes simples et complexes des pronoms forts au pluriel (*nous/vous/elles/eux* vs *nous-autres/vous-autres/eux-autres*). Ajoutons que, dans le cas des marqueurs de conséquence, on a affaire, dans les deux communautés, à un changement par le bas et, dans celui des pronoms forts, à un changement par le haut, ce qui confère un intérêt théorique supplémentaire à la comparaison.

Notre étude vise deux buts : 1) documenter les modèles de convergence/divergence intercommunautaires observables pour les deux cas de variation aux deux points dans le temps, et 2) avancer dans notre réflexion sur les facteurs systémiques et extrasystémiques susceptibles de révéler les points de convergence/divergence évolutifs. Dans cette réflexion, nous vérifions notamment si, dans le cas des pronoms forts, la divergence intercommunautaire était déjà visible dans les années 1970 et s'est intensifiée au fil du temps (voir le modèle illustré par la figure 2.1). De plus, dans notre réflexion sur l'impact des facteurs extrasystémiques sur la convergence/divergence, nous accordons une attention particulière à l'évolution du niveau d'éducation, des médias de langue française et de l'immigration francophone dans les deux communautés ainsi que du niveau de contact avec l'anglais à Welland.

FIGURE 2.1

Modèle évolutif des deux variétés de français laurentien

Nous avons déjà consacré deux études aux marqueurs de conséquence à Montréal et Welland (Blondeau *et al.*, 2019, 2022). Dans ce chapitre, nous accordons une attention particulière à l’alternance entre les formes simples et complexes des pronoms non clitiques pluriels.

La méthodologie***Les deux communautés***

Montréal est une grande métropole, située dans la province de Québec. Bien que Montréal soit culturellement hétérogène, le français est la langue de la majorité de la population et la transmission intergénérationnelle du français y est forte. De plus, à Montréal, au cours des dernières décennies, la population francophone s’est diversifiée : elle comprend notamment un nombre croissant de Québécois francophones provenant d’autres régions du Québec, ainsi que de nombreux nouveaux arrivants francophones nés à l’étranger (surtout en provenance de la France, d’Haïti, du Maghreb et de l’Afrique subsaharienne). En outre, la promulgation de la *Charte de la langue française* au Québec et l’instauration du français comme seule langue officielle de la province durant les années 1970 ont contribué à renforcer le statut du français à Montréal. Enfin, après la Révolution tranquille, le niveau d’instruction de la population francophone du Québec a considérablement augmenté.

Welland est une petite ville industrielle du sud de l’Ontario, située le long d’un canal reliant le lac Ontario et le lac Érié, à environ 700 km de Montréal. Elle comprend une communauté francophone qui représente environ 10 % de la population totale (50 000 en 2011). Cette communauté résulte de l’installation de francophones provenant surtout du Québec qui, des années 1910 au début des années 1980, sont venus travailler dans le secteur industriel, en forte expansion durant cette période. Pendant les années 1980,

Welland est entrée dans une phase de désindustrialisation, et l'arrivée des migrants francophones a pris fin. De plus, contrairement à Montréal, la communauté francophone de Welland n'a pas bénéficié de l'apport ultérieur de l'immigration francophone.

Depuis la fin des années 1960, les francophones de Welland ont droit à l'instruction en français aux paliers élémentaire et secondaire, droit octroyé par le gouvernement provincial de l'époque¹. En revanche, durant les décennies suivantes, l'accès à l'instruction postsecondaire en français n'a guère progressé dans la région de Welland. Selon les statistiques du recensement national, durant les dernières décennies, la connaissance de l'anglais parmi les francophones de Welland s'est accrue considérablement et, inversement, le maintien du français au foyer a fortement diminué. D'après le recensement de 2011, seulement 34 % de la population de langue maternelle française communiquait principalement en français à la maison. C'est donc l'école qui tend à se substituer au foyer dans la reproduction du français à Welland.

Les corpus

Le présent chapitre fait partie d'un projet de recherche subventionné par le Conseil de recherche en sciences humaines du Canada (CRSH) et intitulé 40 ans de changement linguistique à Montréal et à Welland : l'apport de la communauté et de l'individu (sous la direction de M. Tremblay). Ce projet combine l'étude des tendances communautaires, en temps apparent et en temps réel, ainsi qu'un suivi de cohortes dans les deux communautés. Ici, nous nous concentrons sur les tendances communautaires à Montréal et à Welland.

Pour mesurer l'évolution des tendances communautaires à Montréal, nous comparons les données du corpus Sankoff et Cedergren recueillies en 1971 (Sankoff *et al.*, 1976) avec celles du corpus FRAN-HOMA recueilli en 2012 (Martineau et Séguin, 2016) (Blondeau *et al.*, 2021). Pour Welland, nous analysons les données du corpus Mougeon et Beniak (Beniak *et al.*, 1985) recueilli en 1975, que nous comparons à celles du corpus FRAN-WELLAND recueilli en 2012 (Martineau et Séguin, 2016)². Les corpus des années 2010

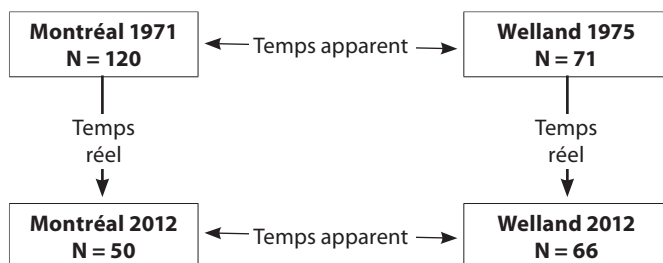
1. Durant les décennies précédentes, la population franco-ontarienne n'avait droit qu'à un enseignement bilingue anglais-français.

2. Sur 66 entrevues réalisées à Welland (corpus Martineau-Mougeon), 48 ont été transcrites dans le cadre du projet CRSH-GTRC *Le français à la mesure d'un continent: un patri-moine en partage*, et le reste des entrevues (corpus Mougeon-Tremblay), dans le cadre du

émanant du projet CRSH-GTRC Le français à la mesure d'un continent : un patrimoine en partage (dir. F. Martineau). Ensemble, les données de ces quatre corpus, totalisant 307 entrevues semi-dirigées recueillies selon des protocoles de recherche similaires, permettent de corroborer l'évolution différenciée du français dans les deux communautés sur une période d'environ 40 ans, comme le montre la figure 2.2.

FIGURE 2.2

Représentation de l'étude sur les tendances communautaires



Les deux cas de variation

Les marqueurs de conséquence

Inspirée de la recherche pionnière de Dessureault-Dober (1974), la première variable sociolinguistique à l'étude consiste en l'alternance entre les formes *alors, donc, (ça) fait (que)* (et ses variantes) – formes utilisées dans les deux communautés – et la conjonction *so*, empruntée de l'anglais et caractéristique du français de Welland. Les exemples en (1a) et (1b) illustrent l'utilisation des formes standard *alors* et *donc*; les exemples en (2a) et (2b) illustrent l'utilisation de la forme vernaculaire *(ça) fait (que)*, souvent réduite³ à [fak] ou [fɛk], et de l'emprunt *so*.

projet 40 ans de changement linguistique à Montréal et à Welland : le rôle de l'individu et de la communauté (dir. Mireille Tremblay).

3. Les variantes réduites mentionnées dans le texte ont été attestées à Montréal. Ce travail n'a pas encore été effectué pour Welland.

- 1) a. Le reste de notre famille habitait plutôt le centre de Montréal **alors** c'était bien commode d'avoir un téléphone pour s'inviter [...]. Montréal 1971, loc. 79, F60, CSE haute.
b. L'institutrice est pas venue me chercher **donc** j'ai été obligé de marcher. Welland 1975, loc. 14, CSE haute.
- 2) a. Pis tu sais ma petite sœur elle a dix-sept ans fait-qu'on s'entend elle a commencé Facebook jeune **fait-que** [fak] des fois elle m'écrit des choses pis je suis comme « Ariane, soit que tu utilises ton français d'école soit que je te parle plus ». FRAN-HOMA 013F27 CSE basse.
b. La directrice elle connaît beaucoup la technologie **so** elle est très bonne pour notre compagnie, FRAN-Welland 100F24, CSE intermédiaire.

Les pronoms forts (non clitiques) au pluriel

Le tableau 2.1 présente la deuxième variable sociolinguistique à l'étude, où les formes simples des pronoms forts au pluriel, qui correspondent au français de référence ainsi qu'au registre standard du français québécois, alternent avec les formes complexes, typiques du français laurentien et associées au vernaculaire (Auger, 1994; Blondeau, 2011; Morin, 1982).

TABLEAU 2.1

Paradigme des pronoms forts (non clitiques) au pluriel en français laurentien

	Simple (Standard)	Complexe (Vernaculaire)
1 ^{re} personne	nous	nous-autres
2 ^e personne	vous	vous-autres
3 ^e personne F	elles	eux-autres
3 ^e personne M	eux	

Les exemples ci-dessous montrent des cas de variation aux trois personnes du pluriel. Les exemples en (3a et 3c) illustrent des cas de variation inhérente, où les deux formes apparaissent dans un même énoncé.

- 3) a. On a une célébration qui va être un peu bizarre pour **nous** mais **nous-autres** on fête la journée de la Parade de Welland. FRAN-Welland 08F26, CSE haute.

- b. Mais quand c'est des jeunes comme **vous-autres**, tu achètes la maison, tu la renippes, tu la rénoves, tu restes là. FRAN-HOMA 021M51, CSE basse.
- c. **Eux-autres** ils étaient tout le temps chez L'Amère à Boire mais **eux** ils avaient/**eux** c'était pas pire parce qu'ils étaient un peu loin de l'action. FRAN-HOMA 024F36, CSE haute.

L'évolution des marqueurs de conséquence

Dans Blondeau *et al.* (2022), nous avons montré que si, dans les années 1970, les deux communautés partageaient un usage commun des deux principales variantes, la forme vernaculaire (*ça fait (que)*) et la forme standard *alors*, à Welland, on constatait un usage plus répandu qu'à Montréal d'une autre forme standard, *donc*, ainsi que l'émergence de la variante *so* empruntée à l'anglais. On observait donc le début d'une évolution divergente dès les années 1970, comme on peut le voir au tableau 2.2. Cette tendance à la divergence intercommunautaire s'est intensifiée dans les années 2010. Tout d'abord, à Welland, on observe une forte montée du connecteur *so* aux dépens de la variante vernaculaire traditionnelle (*ça fait (que)*) alors qu'à Montréal, on constate une augmentation marquée de (*ça fait (que)*). Enfin, l'analyse de l'évolution de la variante standard *alors* révèle une autre facette de la divergence. Alors qu'à Welland *alors* fait preuve de stabilité aux deux époques, à Montréal, cette forme subit un déclin marqué à la faveur de (*ça fait (que)*) et de *donc* (Blondeau *et al.*, 2022).

L'avancée de la variante lexicale (*ça fait (que)*) à Montréal s'explique par la montée de la variante sociophonétique *fek* [fɛk] illustrée en (4), en voie de

TABLEAU 2.2
Fréquence d'usage des marqueurs de conséquence à Montréal et Welland (années 2010) (Blondeau *et al.*, 2022)

	Montréal 1971 n = 3292	Montréal 2012 n = 4120	Welland 1975 n = 1166	Welland 2012 n = 3969
	%	%	%	%
Ça fait que	53	79	48	34
Alors	45	3	37	33
Donc	2	18	10	5
So	0	0	5	27

remplacer la variante *fak* [fak], un changement en cours mené par les femmes (Blondeau et Tremblay, 2022), comme l'illustre l'énoncé en (4) produit par une locutrice d'origine sociale modeste.

- 4) Ben c'est sûr que veut veut pas on s'entend-tu que il y a plein de condos **fait-que** [fɛk] le monde qui habitent dans les condos ils ont quand même un peu d'argent là quand même on s'entend-tu? HOMA 003F24, CSE basse.

Comme l'illustre le continuum⁴ en (5), tiré de Blondeau et Tremblay (2022), les années 1970 présentaient essentiellement une concurrence entre *alors* et *fak*. Par la suite, *alors* a subi un déclin à la faveur de *donc*. L'analyse en temps apparent des années 2010 montre un remplacement de *donc* par *fek* et un élargissement de la diffusion de *fek* à l'ensemble de la communauté. De plus, selon ces autrices, dans les années 2010, *fek* tend à remplacer *fak*: ainsi, elles notent 73 % d'usage de *fek* chez les jeunes (19 à 25 ans), comparativement à 44 % chez les 26-39 ans et 34 % chez les 40 à 60 ans.

- 5) Alors/fak → donc/fak → fek/fak → fek
 1970 à 2010

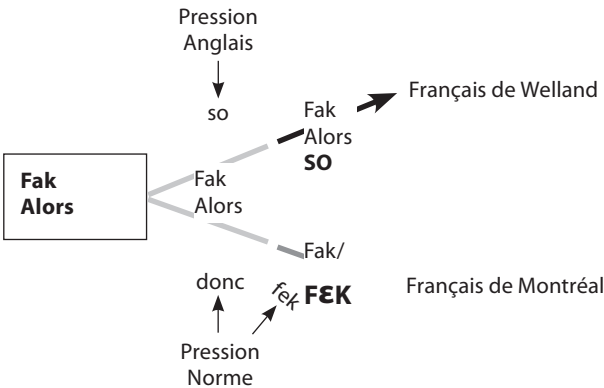
Dans l'analyse des données des corpus de Welland, c'est la trajectoire temporelle de *so* et celle des trois autres formes qui a retenu l'attention de Blondeau *et al.* (2022). Résumons ici ces évolutions. Durant les années 1970, la forme *so* était presque exclusivement employée par les jeunes locutrices bilingues équilibrées des CSE moyenne et basse (les agents probables de son entrée dans le parler local). En 2012, *so* s'est diffusé dans le parler des générations plus âgées de la CSE élevée et des bilingues franco-dominants et anglo-dominants, ces derniers et ces dernières étant désormais à la pointe de l'usage de *so*. Corollairement à la montée de *so* à Welland, on observe le déclin de (*ça*) *fait (que)*. En 2012, cette forme n'est plus employée que 6 % du temps par les jeunes de 15 à 34 ans. De plus, ce sont les bilingues franco-dominants qui utilisent surtout cette variante. Quant à la forme *alors*, on a vu que, si elle décline fortement à Montréal, à Welland sa fréquence n'évolue guère, stabilité qui en quelque sorte entrave la progression de *donc*, puisque cette dernière voit sa fréquence diminuer de moitié de 1975 à 2012.

4. Le schéma en (5) illustre l'évolution des trois marqueurs de conséquence dans la langue de tous les jours à Montréal. À noter qu'*alors* et *donc* demeurent les marqueurs par défaut du registre parlé soutenu, comme le révèlent les études de Bigot (2021) et Villeneuve (2018).

Somme toute, comme l'illustre la figure 2.3, la comparaison sur 40 ans des marqueurs de conséquence dans les deux communautés montre un renforcement considérable de la divergence intercommunautaire. D'une part, l'inventaire des variantes se diversifie différemment dans les deux communautés avec l'émergence de *so* à Welland et de *fek* à Montréal. D'autre part, les variantes communes connaissent des trajectoires bien différentes. Nous aborderons la question des causes externes possibles de cette évolution divergente et de celle que nous avons observée pour la variation pronominale, à la fin de ce chapitre.

FIGURE 2.3

Évolution des marqueurs de conséquence à Montréal et à Welland



Les pronoms forts

La présente section explore la variation pronominale entre les formes simples (*nous*, *vous*, *elles* et *eux*) et complexes (*nous-autres*, *vous-autres* et *eux-autres*) des pronoms forts (non clitiques) au pluriel. Après un bref examen des tendances générales, nous comparons la variation dans les deux communautés en temps apparent dans les années 1970. Par la suite, nous examinons le résultat de l'évolution de cette variation dans les années 2010, et ce, en temps apparent et en temps réel.

Tendances générales

Comme le montre le tableau 2.3, dans les années 1970, les formes complexes, traditionnellement associées au français laurentien vernaculaire, représentent 86 % des cas à Montréal, et 90 % des cas à Welland, alors que les formes simples, associées au français de référence, sont beaucoup moins fréquentes. Cette différence entre les deux communautés, déjà significative dans les années 1970 ($p=.001125$), s'accroît dans les années 2010. En effet, alors qu'on observe une importante montée des formes simples à Montréal qui passent de 14 % à 48 %, à Welland, la montée des formes simples est beaucoup plus lente (de 10 % à 22 %).

TABLEAU 2.3

Évolution des pronoms forts au pluriel à Montréal et à Welland

	Montréal 1971	Welland 1975	Montréal 2012	Welland 2012
	n = 3342	n = 1351	n = 814	n = 1385
	%	%	%	%
Formes simples	14	10	48	22
Formes complexes	86	90	52	78

Pour mieux saisir les différences et ressemblances entre les deux communautés au cours de la montée des formes simples, nous avons évalué l'impact des facteurs extralinguistiques sur le changement. Les résultats de cette évaluation sont présentés ci-dessous.

Années 1970

Le tableau 2.4 présente les résultats d'une analyse multivariée des données des années 1970, effectuée avec le logiciel GoldVarb Lion (Sankoff *et al.*, 2012). On voit que si, dans les deux communautés, l'utilisation des formes simples est favorisée par la CSE haute et par les femmes, on décèle des différences intercommunautaires relativement à l'effet du genre et de l'âge.

À Montréal, on observe un phénomène de gradation d'âge ; le groupe des 35 à 64 ans est le plus grand utilisateur des formes simples, ce qui correspond généralement à la période d'activité professionnelle. La situation est différente à Welland, où on remarque une augmentation de formes simples chez les plus jeunes, ce qui est un indice de changement en cours en faveur

TABEAU 2.4

**Influence des facteurs sociaux sur l'emploi des formes simples
(Montréal 1971 vs Welland, 1975)**

Facteurs		Montréal (n = 3250)			Welland (n = 1348)		
		% <i>simple</i>	N total	Poids relatif	% <i>simple</i>	N total	Poids relatif
Âge	15-34	11,4 %	171/1500	.44	15 %	75/491	.63
	35-64	18,8 %	262/1393	.64	8 %	48/639	.42
	65+	3,9 %	14/357	.24	7 %	15/218	.44
	Écart			39			21
CSE	Haute	37 %	237/641	.83	26 %	71/270	.79
	Intermédiaire	10,2 %	109/955	.48	7 %	37/527	.46
	Basse	6,5 %	101/1545	.35	5 %	30/551	.38
Genre	Écart			48			41
	Féminin	17 %	286/1680	.58	9 %	65/745	NS
	Masculin	10,3 %	161/1570	.41	12 %	73/603	
Bilinguisme	Écart			17			s. o.
	Franco-dominant				8 %	59/771	NS
	Équilibré				15 %	31/209	
	Anglo-dominant				13 %	48/368	
	Écart						s. o.
Total		13,8 %	447/3250		10 %	138/1348	
	Input 0,098 – Signification 0,000			Input 0,080 – Signification = 0,000			

des formes simples. De plus, alors qu'à Montréal les femmes favorisent les formes simples, à Welland, les hommes emploient les formes simples un peu plus souvent que les femmes, mais cette différence n'est pas significative.

Enfin, dans l'analyse multivariée des données de Welland, nous avons mesuré l'impact du bilinguisme. On voit que, si les personnes bilingues équilibrées et anglo-dominantes utilisent les pronoms simples un peu plus souvent que les personnes franco-dominantes, cette différence n'est pas statistiquement importante.

En résumé, malgré la claire prédominance des formes complexes dans les deux communautés, l'analyse en temps apparent révèle plusieurs différences quant à la configuration sociale de la variation. En particulier, le

facteur âge révèle une évolution divergente. À Montréal, alors que les personnes plus âgées utilisent très rarement les formes simples, on observe un phénomène de gradation d'âge, qui peut être attribué à l'effet de la pression normative durant la vie active. À Welland, on décèle plutôt les indices du début d'un changement en cours impulsé par les jeunes.

Années 2010

Le tableau 2.5 présente les résultats de l'analyse des facteurs sociaux dans les années 2010.

TABEAU 2.5

Influence des facteurs sociaux sur l'emploi des formes simples (Montréal 2012 vs Welland, 2012)

Facteurs		Montréal (n=814)				Welland (n=1390)		
		% simple	N total	Poids relatif		% simple	N total	Poids relatif
Âge	18-25	70 %	136/195	.74	15-34	34 %	65/193	NS
	26-39	49 %	112/229	.47	35-64	26 %	170/658	
	40-65	36 %	140/390	.38	65+	14 %	76/539	
	Écart			35	Écart			s. o.
CSE	Haute	63 %	204/324	.62	Haute	42 %	211/505	.76
	Intermédiaire	48 %	110/229	.57	Intermédiaire	14 %	74/515	.39
	Basse	28 %	74/261	.30	Basse	7 %	26/370	.29
Genre	Écart			33	Écart			47
	Féminin	60 %	262/438	.61	Féminin	27 %	193/726	.55
	Masculin	34 %	126/376	.37	Masculin	18 %	118/664	.44
	Écart			25	Écart			11
Bilinguisme	Franco-dominant	s. o.	s. o.		Franco-dominant	19 %	124/663	.49
	Équilibré	s. o.	s. o.		Équilibré	17 %	66/396	.40
	Anglo-dominant	s. o.	s. o.		Anglo-dominant	37 %	121/330	.65
	Écart			s. o.				25
Total			814			22 %	311/1390	
	Input 0.477 Signification = 0,000				Input 0.181 Signification = 0,006			

L'analyse factorielle multivariée des données de Montréal a retenu l'âge, la CSE et le genre. Pour ce qui est de l'âge, on constate que son effet est linéaire: la jeune génération (18 à 25 ans) emploie les formes simples plus souvent que la génération intermédiaire qui, elle-même, les emploie moins souvent que la génération plus âgée (40 à 65 ans). Selon le modèle du temps apparent, ce résultat suggère qu'on a affaire à un changement en cours impulsé par les jeunes. Cela contraste avec la situation des années 1970, où l'effet curvilinéaire de l'âge était symptomatique d'une gradation d'âge⁵. Les jeunes professionnels transmettent la nouvelle norme à leurs propres enfants (les plus jeunes en 2012), ce qui explique la rapidité du changement entre les deux périodes. S'agissant de la CSE, on voit que, comme dans les années 1970, les personnes de CSE élevée favorisent les formes simples et celles de CSE faible les défavorisent. Enfin, en ce qui concerne le genre, les femmes favorisent fortement la forme simple et, on peut le supposer, jouent un rôle de premier plan dans le changement. Somme toute, l'ensemble des résultats des analyses multivariées des données de Montréal indique que l'usage des formes simples est maintenant entré dans une phase dynamique et que l'on a affaire à un changement par le haut.

L'analyse des données de Welland révèle que l'effet de l'âge semble s'atténuer. Les différences entre les trois cohortes d'âges vont effectivement dans le même sens qu'en 1975, mais elles ne sont plus significatives. On constate aussi que la CSE et le genre exercent le même effet qu'à Montréal: les formes simples sont employées plus souvent par la CSE haute et les femmes, ce qui pourrait constituer deux indices de la poursuite du changement par le haut. Toutefois, l'absence de différence de fréquences significatives selon l'âge suggère qu'un tel changement serait en perte de vitesse. Pour ce qui est du facteur relatif au niveau de bilinguisme, alors que ce dernier n'était pas retenu dans l'analyse des données de 1975, il a une influence importante durant les années 2010 – les personnes anglo-dominantes se démarquant des autres par un taux d'emploi des formes simples deux fois plus élevé.

Pour mieux comprendre l'influence des facteurs CSE, genre et bilinguisme sur la montée des formes simples, nous avons croisé d'une part le genre et le bilinguisme, d'autre part, la CSE et le bilinguisme, à Welland. Ces croisements de facteurs dévoilent l'importance de l'impact du bilin-

5. Selon Labov (2010), l'opposition entre gradation d'âge et changement générationnel peut être trompeuse, car la gradation d'âge peut participer au mécanisme de certains types de changements, comme c'est le cas ici.

guisme anglo-dominant dans la montée des formes simples à Welland. En effet, quand on croise genre et bilinguisme (tableau 2.6), on voit que ce sont les femmes anglo-dominantes qui ont le taux le plus élevé d'emploi des pronoms simples (40 %). Quand on croise bilinguisme et classe sociale (tableau 2.7), on constate que ce sont les personnes anglo-dominantes de la CSE haute qui ont le taux le plus élevé (pas moins de 56 %). Somme toute, la prise en compte du facteur bilinguisme dans l'analyse des données de Welland fournit un éclairage complémentaire de la divergence intercommunautaire dans la dynamique sociale de la montée des pronoms simples.

En résumé, comme l'illustre la figure 2.4, l'étude de l'évolution de la variation pronominal indique que les grammaires des deux communautés étaient semblables dans les années 1970, malgré des divergences déjà observables à cette période.

TABLEAU 2.6

Usage des formes simples selon le genre et le bilinguisme (Welland, 2012)

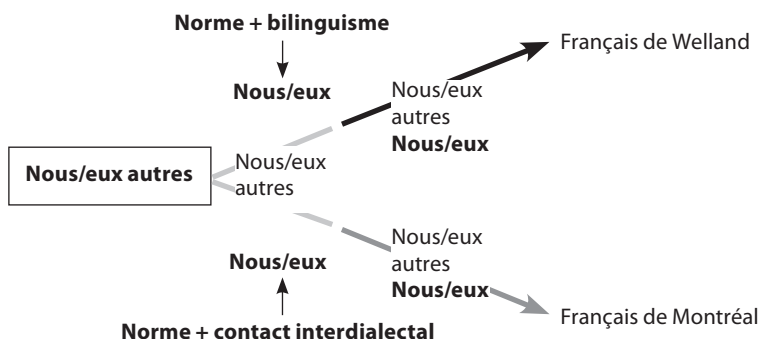
Genre	Bilinguisme	% simples	n simples	Total
Femmes	Franco-dominantes	26 %	75	288
	Bilingues équilibrées	13 %	29	223
	Anglo-dominantes	40 %	85	211
Hommes	Franco-dominants	13 %	48	371
	Bilingues équilibrés	21 %	37	173
	Anglo-dominants	26 %	31	117

TABLEAU 2.7

Usage des formes simples selon le genre et la classe sociale (Welland, 2012)

Genre	Bilinguisme	% simples	n simples	Total
Femmes	CSE haute	47 %	117	251
	CSE intermédiaire	18 %	62	348
	CSE basse	8 %	10	123
Hommes	CSE haute	36 %	90	251
	CSE intermédiaire	7 %	11	164
	CSE basse	6 %	15	246

FIGURE 2.4

Évolution des pronoms toniques au pluriel à Montréal et à Welland

Parmi les points de convergence entre les deux communautés figuraient la haute fréquence des pronoms complexes, l'effet de la CSE et l'absence d'effet du bilinguisme à Welland. Cependant, on constatait aussi des points de divergence, en particulier le contraste entre la gradation d'âge à Montréal et le changement en cours en faveur des formes simples à Welland.

L'analyse des données des années 2010 a fait apparaître deux points principaux de convergence. Dans les deux communautés: 1) les pronoms simples sont en phase d'ascension et 2) les femmes et la CSE haute sont à la pointe du changement. Cependant, notre analyse a aussi révélé que le changement est beaucoup plus rapide à Montréal, ce qui indique de la divergence. De surcroît, le rôle du bilinguisme anglo-dominant à Welland semble indiquer une dynamique particulière aux communautés en situation minoritaire.

* * *

L'étude de l'emploi des marqueurs de conséquence et des pronoms forts du pluriel dans les deux communautés et aux deux points dans le temps a révélé à la fois des points de convergence et des points de divergence. Nous les rappelons brièvement ici et passons en revue les principaux facteurs qui sont susceptibles de les expliquer.

Commençons par rappeler les principaux points de convergence. Durant les années 1970, on constate que, globalement, dans les deux communautés, les formes complexes des pronoms forts l'emportent très nettement sur les

formes simples, que *(ça) fait (que)* et *alors* sont les deux principaux marqueurs de conséquence et qu'ils sont employés à des niveaux de fréquence relativement semblables. De plus, dans les deux communautés, la CSE exerce la même influence sur les variantes. Les pronoms simples et *alors* sont associés à la CSE haute; *(ça) fait (que)* et les pronoms complexes le sont à la CSE basse. Il en est de même pour le connecteur *donc*: comme *alors*, il est associé à la CSE haute (Blondeau *et al.*, 2022). Selon nous, le principal facteur qui explique ces points de convergence est que, jusqu'à la fin des années 1960, la force des vagues migratoires en provenance du Québec à Welland a eu pour effet de préserver les principaux aspects de la grammaire variable des deux communautés relativement aux aspects de la morphosyntaxe du français ciblés par notre étude.

Cela dit, dès les années 1970, on décèle plusieurs points de divergence. À Montréal, on observe une gradation selon l'âge: dans le cas des pronoms forts et une association linéaire avec l'âge à Welland. Dans le cas des marqueurs de conséquence, on observe l'émergence à Welland de la variante vernaculaire *so* dans le parler des jeunes des CSE intermédiaire et basse. Pour ce qui est de la gradation d'âge à Montréal et son absence à Welland, elle pourrait refléter le fait que, durant les années 1970, les possibilités de travailler en français étaient fort réduites à Welland, mais pas à Montréal. À Montréal, les activités professionnelles étaient donc susceptibles d'avoir un effet standardisant, alors qu'à Welland, le monde du travail avait plutôt tendance à renforcer l'exposition à l'anglais. L'émergence de *so* reflète, pour sa part, le fait qu'avec l'arrêt du courant migratoire en provenance du Québec à la fin des années 1960, les plus jeunes générations sont majoritairement nées à Welland et sont nettement plus bilingues que les générations plus âgées. L'émergence de *so* dans le parler des plus jeunes pourrait donc être le signe avant-coureur de la montée d'une identité bilingue chez les jeunes en rupture avec l'identité plus franco-dominante des générations plus âgées.

L'analyse des données des années 2010 a révélé une tendance générale à l'intensification de la divergence intercommunautaire. Cette tendance est particulièrement évidente dans le cas des marqueurs de conséquence. En effet, nous avons vu qu'à Welland l'emploi de *so* a continué de progresser et qu'il s'est diffusé dans le parler des trois CSE et des trois cohortes d'âges. Dans le parler des adolescents, son emploi est devenu quasi systématique. Corollairement *(ça) fait (que)* est en chute libre dans le parler des moins de 35 ans. Quant aux variantes standard *alors* et *donc*, on a vu que la fréquence de la première est restée stable et que celle de la deuxième a diminué. Par

contraste, à Montréal, la variante vernaculaire (*ça*) *fait (que)* a poursuivi son ascension, allant même jusqu'à se scinder en deux formes sociophonétiques : *fak* se trouve confronté à un nouveau joueur *fek*, qui est venu le concurrencer et est en voie d'évincer les formes plus standard *alors* et *donc* (Blondeau et Tremblay, 2022). Cela dit, durant la dernière décennie, des observations qualitatives montrent que l'émergence de deux nouvelles variantes à Montréal, le connecteur *so* et la locution *du coup*, une innovation largement attestée en France (Abouda, 2022). Bien qu'elles soient absentes du corpus de 2010, il sera pertinent d'examiner à partir d'autres sources de données dans quelle mesure ces nouvelles variantes entreront en compétition avec les variantes locales.

L'étude de l'emploi des pronoms forts durant les années 2010 a aussi révélé une tendance à la divergence. Toutefois, celle-ci est moins évidente que dans le cas des marqueurs de conséquence. En effet, au cours de cette période, la divergence se manifeste essentiellement par une montée nettement plus rapide des pronoms simples à Montréal, par rapport à Welland.

Quant aux facteurs qui ont pu contribuer à l'intensification de la divergence relativement aux marqueurs de conséquence et aux pronoms forts, on peut les résumer en observant que, durant les quatre dernières décennies, la pression du français normé a été plus forte à Montréal qu'à Welland, et ce pour deux raisons principales : 1) l'accès accru à l'éducation postsecondaire en français à Montréal, mais pas à Welland ; et 2) la croissance des médias de langue française à Montréal, mais pas à Welland. Cette différence aurait été propice à la montée de *donc* (trait typique du français standard écrit) au détriment de *alors* à Montréal, à la stabilité de ce marqueur à Welland et à l'augmentation plus rapide des pronoms simples à Montréal qu'à Welland. Ajoutons que cette augmentation pourrait avoir été renforcée par l'exposition accrue aux variétés de français international à Montréal. Quant à la forte montée de *so* aux dépens de (*ça*) *fait (que)* à Welland et l'absence continue de *so* à Montréal, elles découleraient du fait que les bilingues équilibrés ou les anglo-dominants sont devenus majoritaires à Welland, mais pas à Montréal.

En conclusion, les analyses menées à partir de données comparables recueillies dans des conditions similaires à un intervalle de 40 ans ont permis de suivre en temps réel l'évolution de deux variables linguistiques. De surcroît, en s'appuyant sur le modèle du temps apparent, les analyses menées dans les années 1970 et 2010 ont mis en évidence le rôle des facteurs sociaux sur la divergence accrue entre ces deux communautés. En somme, cette étude

comparative de deux variétés apparentées de français laurentien a montré combien l'évolution différentielle de la configuration démolinguistique et la dynamique sociale propre aux deux communautés ont eu des retombées sur la variation sociolinguistique.

Références

- ABOUDA, L. (2022), L'émergence du marqueur méta-discursif *du coup*: de la conséquence à l'actualisation énonciative, *Langages*, 226(2), 99-116, <https://doi.org/10.3917/lang.226.0099>.
- AUGER, J. (1994), *Pronominal Clitics in Québec Colloquial French: A Morphological Analysis*, University of Pennsylvania.
- BENIAK, E., R. MOUGEON et D. VALOIS (1985), *Contact des langues et changement linguistique: étude sociolinguistique du français parlé à Welland (Ontario)*, Centre international de recherche sur le bilinguisme.
- BIGOT, D. (2021), *Le bon usage québécois: étude sociolinguistique sur la norme grammaticale du français parlé au Québec*, Presses de l'Université Laval.
- BLONDEAU, H. (2011), *Cet « autres » qui nous distingue: tendances communautaires et parcours individuels dans le système des pronoms en français québécois*, Presses de l'Université Laval.
- BLONDEAU, H., R. MOUGEON et M. TREMBLAY (2019), Analyse comparative de *ça fait que*, *alors*, *donc* et *so* à Montréal et à Welland: mutations sociales, convergences, divergences en français laurentien, *Journal of French Language Studies*, 29(1), 35-65, <https://doi.org/10.1017/S0959269518000169>.
- BLONDEAU, H., R. MOUGEON et M. TREMBLAY (2022), The Diverging Paths of Consequence Markers in Canadian French, dans E. PETERSON, T. HILTUNEN et J. KERN (dir.), *Discourse-Pragmatic Variation and Change* (1^{re} éd.), Cambridge University Press, 230-250, Cambridge University Press, <https://doi.org/10.1017/9781108864183.013>.
- BLONDEAU, H. et M. TREMBLAY (2022), The hidden dimensions of a change from below: Consequence markers in Montreal French, *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 67(12), 22-52, <https://doi.org/10.1017/cnj.2022.3>.
- BLONDEAU, H., M. TREMBLAY, A. BERTRAND et E. MICHEL (2021), A new milestone for the study of variation in Montréal French: The Hochelaga-Maisonneuve sociolinguistic survey, *Corpus*, 22, <https://doi.org/10.4000/corpus.6221>.
- CÔTÉ, M. H. (2020), De français canadien à québécois à laurentien: la résolution de trois siècles de débats terminologiques, *Études de lettres*, 312, 43-46, <https://doi.org/10.4000/edl.23-36>.
- DESSUREAULT-DOBER, D. (1974), *Étude sociolinguistique de /Ça fait que/: coordonnant logique et marqueur d'interaction*, mémoire de maîtrise, Université du Québec à Montréal.
- LABOV, W. (2010), *Principles of linguistic change, 1: Internal factors*, Wiley-Blackwell.

- MARTINEAU, F. et M.-C. SÉGUIN (2016), Le Corpus FRAN : réseaux et maillages en Amérique française, *Corpus*, 15, <https://doi.org/10.4000/corpus.2925>.
- MORIN, Y.-C. (1982), De quelques [l] non étymologiques dans le français du Québec : notes sur les clitiques et la liaison, *Revue québécoise de linguistique*, 11(2), 9-47, <https://doi.org/10.7202/602486ar>.
- MOUGEON, R., S. HALLION BRES, D. BIGOT et R. PAPEN (2016), Convergence et divergence sociolinguistique en français laurentien : l'alternance rien que/juste/seulement/seulement que/ne ... que, *Journal of French Language Studies*, 26, 115-154, <https://doi.org/10.1017/S095926951600003X>.
- MOUGEON, R., T. NADASDI et K. REHNER (2009), Évolution de l'alternance je vas/je vais/je m'en vas/je m'en vais/m'as dans le parler d'adolescents franco-ontariens (1978 vs 2005), dans L. BARONIAN et F. MARTINEAU (dir.), *Le français d'un continent à l'autre. Mélanges offerts à Charles Morin*, Presses de l'Université Laval, 327-373.
- SANKOFF, D., G. SANKOFF, S. LABERGE et M. TOPHAM (1976), Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale, *Cahier de linguistique*, 6, 85-125. <https://doi.org/10.7202/800043ar>.
- SANKOFF, D., S. A. TAGLIAMONTE et E. SMITH (2012), *Goldvarb Lion: A variable rule application for Macintosh*, Department of Linguistics, University of Toronto.
- VILLENEUVE, A.-J. (2018), Normes objectives et variation socio-stylistique : le français québécois parlé en contexte d'entrevues télévisées, *Arborescences*, 7, 49-66, <https://doi.org/10.7202/1050968ar>.

CHAPITRE 3

Corpus oraux et corpus de textos

La référence temporelle au futur en Belgique et au Québec

*Mireille Tremblay, Hélène Blondeau, Anne Dister
et Emmanuelle Labeau*

Ce chapitre porte sur la référence temporelle au futur (RTF) en Belgique et au Québec, et s'appuie sur des corpus de données provenant de productions orales et de textos. Depuis le travail fondateur d'Emirkanian et Sankoff (1985) sur la variation de l'usage des formes du futur dans le français parlé à Montréal, les cercles de la sociolinguistique variationniste entretiennent leur intérêt pour cette variable. Ce premier travail empirique sur le futur avait été mené à partir d'un sous-échantillon des données du corpus Sankoff-Cedergren de 1971. Nous nous inspirons ici de cette étude sociolinguistique pionnière tout en élargissant sa dimension comparative.

La plupart des travaux sur la variation sociolinguistique s'appuient sur la notion labovienne de communauté linguistique. Ils sont ancrés dans l'examen de la langue dans son contexte social, et, de ce fait, relativement peu d'études ont analysé de manière systématique, en faisant appel à des corpus comparables, les différences d'usage entre le français parlé en Europe et celui parlé en Amérique du Nord¹. En effet, c'est surtout depuis l'orée du

1. Cependant, la variation dialectale en français canadien est mieux décrite, par exemple, dans les travaux sur la convergence et la divergence entre les variétés de français acadienne et laurentienne et ou encore la variation à l'intérieur même des variétés laurentiennes, comme nous l'avons vu au chapitre 2 de ce volume portant sur une comparaison entre français parlé

21^e siècle que l'on assiste à des études plus larges, grâce au développement des approches en sociolinguistique comparative (Tagliamonte, 2013).

Le présent chapitre compare le français employé en Belgique et au Québec, deux communautés francophones qui se distinguent par leur distance géographique et symbolique de la France. Afin d'évaluer les différences entre les deux variétés, nous examinons l'usage du futur analytique (FA) et du futur synthétique (FS). L'objectif est d'établir les différences entre les deux variétés de français, compte tenu de l'effet du médium utilisé lors de l'interaction sur les usages observés. Pour ce faire, la comparaison fait appel à l'analyse de données orales provenant de deux corpus de français parlé, recueillis à Bruxelles et à Montréal, que nous mettons en relation avec les résultats d'une étude antérieure sur l'emploi de la référence au futur dans les textos belges et québécois (Tremblay *et al.*, 2020).

Après une présentation portant sur la variable soumise à l'examen et sur le cadre conceptuel de l'étude, une brève section établit l'état des connaissances sur la RTF afin de dégager les enjeux préalables à cette étude. S'ensuit une présentation des principaux résultats relatifs à la variation dans la référence au futur, dans les textos belges et québécois. La seconde partie dévoile les résultats de la comparaison de la variation à l'oral à partir de deux corpus de français parlé recueillis à Bruxelles et à Montréal. Enfin, la discussion synthétise les résultats de manière à faire le point sur la variation interdialectale en relation avec les différences interactionnelles entre les corpus oraux et les corpus de textos.

La variable et le cadre conceptuel de l'étude

Depuis les années 1980, l'étude de la RTF a suscité beaucoup d'intérêt en sociolinguistique variationniste (voir Comeau et Villeneuve, 2016). Les principales variantes pour exprimer la RTF en français sont : le futur synthétique, le futur analytique et le présent à valeur de futur. Les exemples 1) à 3), tirés du volet québécois du corpus de textos Textos4Science (Langlais *et al.*, 2012), illustrent ces trois variantes.

à Montréal et à Welland. Quant aux travaux comparatifs entre les variétés nord-américaines et les variétés européennes, on peut mentionner les travaux menés par Marie-Hélène Côté dans le cadre du projet PFC, en particulier le chapitre 1 du présent volume qui porte sur la liaison variable.

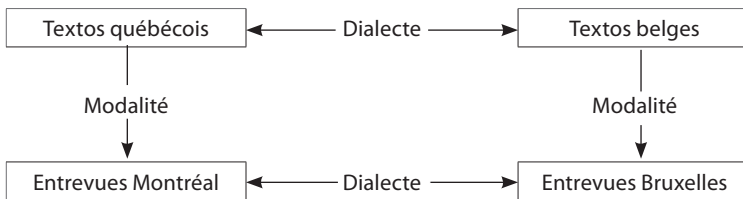
- 1) FS: Penses tu être debout ce soir quand je **reviendrai** du bureau (Texto4Science, 111 090)
- 2) FA: Je **vais le faire imprimer** demain matin a lecole. (Texto4Science, 112 314)
- 3) P: Ben j'avais p-e de koi mais je ten **reparle** tentot (Texto4Science, 112 314)

Si les études antérieures ont porté principalement sur l'alternance entre les variantes du FS et du FA, certains travaux ont également examiné le rôle de la variante du présent de l'indicatif (P) dans la variation (Grimm, 2015; Gudmestad *et al.*, 2018; Poplack et Turpin, 1999; Poplack et Dion, 2009; Tremblay *et al.*, 2020). Pour des raisons explicitées ci-dessous, l'analyse variationniste proposée dans ce chapitre se focalisera sur les deux variantes morphologiques du futur: le FS et le FA.

La figure 3.1 représente le cadre conceptuel du présent chapitre. Elle montre les différents axes de la comparaison. L'axe interdialectal vise à mesurer les différences entre le français de Montréal et celui de Bruxelles. L'axe intermodal vise à examiner dans quelle mesure le français oral se distingue du français des textos et si ces différences se répercutent sur l'axe dialectal.

FIGURE 3.1.

Cadre conceptuel



Pour ce faire, nous dégageons les principales différences interdialectales dans les corpus de textos à partir d'une étude antérieure (Tremblay *et al.*, 2020) que nous mettons ensuite en parallèle avec une analyse variationniste de l'alternance entre le FS et le FA à l'oral.

État des lieux sur la variable de la référence temporelle au futur en français

Types de corpus étudiés

La référence au futur en français a fait l'objet d'études portant sur plusieurs aires géographiques, en particulier en Amérique du Nord et chez des locuteurs de français langue première ou seconde (Comeau et Villeneuve, 2016; Tremblay *et al.*, 2020). Les études antérieures ont examiné plusieurs contextes d'usage. Les travaux variationnistes ont porté principalement sur des données tirées d'entretiens sociolinguistiques semi-dirigés de type labovien (Abouda et Skrovec, 2016; Blondeau, 2007; Deshaies et Laforge, 1981; Emirkanian et Sankoff, 1985; Grimm, 2010; Poplack et Turpin, 1999; Roberts, 2012; G. Sankoff, 2019; Wagner et Sankoff, 2011). Cependant, d'autres études sur l'oral ont examiné des conversations spontanées (Comeau, 2015; King et Nadasdi, 2003) ou de l'oral préparé (Blondeau et Labeau, 2016). D'autres travaux ont porté sur des productions écrites, soit dans une forme plus traditionnelle, comme les journaux (Lesage, 1991; Wales, 2002), ou encore dans le discours numérique, par exemple dans les textos (Labeau, 2014), un contexte de production qualifié d'hybride (Blondeau *et al.*, 2014; Blondeau et Tremblay, 2022; Tremblay, 2020).

Tendances générales et facteurs favorisant l'emploi des variantes à l'oral

Les études variationnistes antérieures ont porté non seulement sur les taux d'usage des différentes variantes, mais également sur l'effet de facteurs linguistiques et sociaux sur la variation. Un des enjeux importants pour cette variable consiste à déterminer si elle représente une variation stable conditionnée par des facteurs linguistiques et sociaux ou une variable en changement au sein des communautés étudiées. Si un changement est en cours, il y a lieu de se demander si les différences interdialectales reflètent les étapes d'un changement plus global à l'échelle du français contemporain.

Sur le plan des tendances générales, les travaux qui ont adopté une approche tripartite, en intégrant la variante du présent à l'alternance entre le FS et le FA, ont constaté que le P ne représentait qu'une faible portion du contexte variable, avec un taux en général inférieur à 10 % (Wagner et Sankoff, 2011). En outre, en français laurentien, l'emploi de cette variante serait assez stable au fil du temps, selon une analyse diachronique de Poplack et Dion (2009) portant sur plus de deux siècles. Les taux semblent fluctuer

davantage pour l'alternance entre les deux autres variantes (FS et FA). À ce titre, nous avons constaté des différences dans les taux de FS dans les variétés de français parlé, allant, en France continentale, de 34 % (Abouda et Skrovec, 2016 ; 2017) à 58 % (Jeanjean, 1988) ; ils atteignent 72 % en Martinique (Roberts, 2014). Pour les variétés de français parlé au Canada, on observe des différences entre les taux d'usage du FS en français acadien, allant de 40 % à 60 %, (Comeau, 2015 ; King et Nadasdi, 2003), ce qui les rapproche des taux européens, et les taux d'usage beaucoup moins élevés en français laurentien, souvent entre de 10 % et 20 % (Grimm, 2015 ; Poplack et Turpin, 1999 ; G. Sankoff et Wagner, 2020). Ces indications sur les taux dans différents dialectes pourraient non seulement conforter l'hypothèse que le FA évincerait graduellement le FS, mais également indiquer que ce changement serait beaucoup plus avancé en français laurentien.

L'analyse des contraintes linguistiques et sociales sur la variation a également permis de mieux cerner les dynamiques sociolinguistiques en cours. Parmi les facteurs les plus discutés, on retrouve la polarité négative, la contingence et la distance temporelle. D'autres facteurs ont été analysés, comme la présence d'adverbes non spécifiques (Poplack et Turpin, 1999) et la personne grammaticale (Poplack, 2001 ; Poplack et Turpin, 1999 ; Roberts, 2012). Toutefois, les différentes études ne s'intéressent pas toutes aux mêmes facteurs linguistiques, et il devient ardu de dégager les principales différences entre les variétés, surtout quand les comparaisons sont menées de manière indépendante. Quelques éléments retiennent cependant l'attention. D'une part, la polarité, un facteur qui entre en jeu dans plusieurs variétés, semble pertinente pour distinguer les variétés entre elles, les énoncés négatifs ayant tendance à favoriser le FS, mais à des degrés divers selon les communautés. Par exemple, la polarité joue un rôle plus important en français laurentien qu'en français hexagonal (Roberts, 2012). En français acadien, ce même facteur est non significatif (Comeau, 2015). Un autre facteur traditionnellement invoqué est celui de la distance temporelle. Il est cependant difficile à saisir, car il n'est pas toujours mesuré de la même manière selon les études. Cela dit, un consensus semble se dégager sur le rôle de la distance temporelle dans les variétés européennes et acadiennes, où le futur analytique est favorisé dans les contextes rapprochés (Comeau, 2015), et non pas (ou non plus) dans les variétés de français laurentien, où ce facteur n'est pas significatif (Poplack et Turpin, 1999). Les travaux antérieurs ayant intégré la variante P ont montré deux facteurs conditionnant son emploi, à savoir la spécification adverbiale (Poplack et Turpin, 1999) et la présence d'une indication spatio-temporelle

(Comeau, 2015; Edmonds *et al.*, 2017; Poplack et Dion, 2009). Selon Poplack et Turpin (1999), les contraintes linguistiques sur l'emploi du présent resteraient stables au fil du temps, ce qui justifierait d'exclure cette variante de l'analyse du changement en cours.

Quant aux facteurs sociaux, deux facteurs principaux ressortent des analyses, l'âge et l'origine sociale des locuteurs. Les études en temps apparent (Poplack et Turpin, 1999, entre autres) et en temps réel (G. Sankoff, 2019; Wagner et Sankoff, 2011b; Zimmer, 1994) ont montré l'existence d'un changement en cours mené par les jeunes en faveur de l'analytique en français laurentien, et une association entre le FS et la catégorie socioéconomique haute. Il semble, par ailleurs, que le facteur du genre ne joue pas de rôle significatif pour cette variable. Enfin, très peu de travaux ont abordé la question du style. Paradoxalement, on note souvent dans les études sur la question que la variable est peu marquée, sans toutefois que cette hypothèse ait été testée empiriquement. Enfin, encore moins de travaux ont examiné le rôle du médium utilisé lors de l'interaction en comparant les données de manière systématique, ce à quoi s'emploie le présent chapitre.

La variation dialectale dans les corpus de textos belges et québécois

Cette section présente les résultats d'une étude antérieure émanant d'un projet de comparaison de la variation observable dans des corpus de textos belges et québécois, recueillis respectivement en 2004 et en 2009-2010. L'analyse a montré que l'usage de la forme du présent ne distinguait pas les variétés entre elles. La variante P représentait 49 % de la variation au Québec et 46 % en Belgique, une différence non significative (Tremblay *et al.*, 2020). Par ailleurs, dans les deux variétés, les mêmes facteurs linguistiques contraignaient la variation, soit la spécification adverbiale, le type de verbe et la contingence.

En revanche, une différence dialectale indéniable ressortait des fréquences de l'alternance entre le FS et le FA : l'ordonnancement des deux variantes était pratiquement inversé d'une communauté à l'autre. Comme le montre le tableau 3.1, le FS compte pour 37,3 % au Québec, alors qu'il représente 71,6 % des formes en Belgique. La distinction entre les contextes négatifs et affirmatifs signale que c'est surtout en contexte affirmatif que s'observe la variation dialectale.

L'analyse multivariée d'un ensemble de facteurs linguistiques et sociaux a montré des différences significatives entre les deux variétés dans les textos.

TABLEAU 3.1

Distribution du FS dans les textos (adapté de Tremblay et al., 2020)

	Québec		Belgique	
	%	N	%	N
Polarité				
Négatif	85,0	47	81,0	82
Affirmatif	33,0	631	70,2	604
Total	37,3	678	71,6	686

On a en effet observé une configuration distincte selon la communauté. La polarité était fortement associée au FS au Québec, mais ne jouait pas de rôle significatif en Belgique. La présence d’adverbes non spécifiques était significative en Belgique, mais pas au Québec. Cependant, la distance temporelle conditionnait la variation dans les deux communautés, le FA étant associé à un futur rapproché. Enfin, les deux variétés étaient sensibles à la complexité morphologique du paradigme verbal, les verbes irréguliers comme *être*, *avoir* et les modaux favorisant le FS.

Questions de recherche

Comme nous venons de le voir, l’analyse des corpus de textos a révélé une variation dialectale importante dans le choix des variantes FA et FS. Il est possible que cette variation soit tributaire des particularités différentielles de ce médium dans les deux communautés – la Belgique alignant possible-ment davantage l’écrit texto sur l’écrit standard que le Québec – ou encore que l’écrit texto reflète une véritable distinction dialectale. Le présent chapitre, qui compare la même variable dans des corpus oraux belge et québécois, permettra de choisir entre ces deux hypothèses et, plus précisément, de répondre aux questions suivantes :

- 1) Les corpus oraux confirment-ils la variation dialectale observée dans les textos ?
- 2) Quel rôle les facteurs sociaux jouent-ils dans la variation à l’oral ?
- 3) Y a-t-il un changement en cours dans l’une ou l’autre communauté ?
- 4) Le médium a-t-il un effet sur la variable et, si oui, quelle est l’ampleur de cet effet ?

La variation dialectale dans les corpus oraux : méthodologie

Les corpus

Pour mesurer l'évolution des tendances communautaires à Montréal, nous comparons les données du corpus FRAN-HOMA recueilli en 2012 (Blondeau *et al.*, 2021; Martineau et Séguin, 2016) avec les données du Corpus de français parlé à Bruxelles (CFPB) (Dister et Labeau, 2017), recueilli à Bruxelles entre 2013 et 2018. Les données montréalaises sont prélevées d'un sous-échantillon de 38 entrevues semi-dirigées. L'échantillon est stratifié en termes d'âge, de genre et de catégorie socioéconomique (CSE). Pour les besoins du présent chapitre, nous avons catégorisé les 18 femmes et 20 hommes de l'échantillon, répartis en trois groupes d'âge (18-25; 26-39; 40-65) et trois CSE (haute, intermédiaire et basse). Pour Bruxelles, le corpus comprend 19 entrevues (8 femmes et 11 hommes). Cependant, il n'a pas été possible de distinguer trois groupes d'âge et trois CSE pour l'analyse statistique en raison du plus faible nombre de participants. Nous avons plutôt opté pour une catégorisation binaire : deux CSE (haute et basse) et deux groupes d'âge (18-39; 40 ans et plus).

Le contexte variable

L'analyse des textos met en évidence qu'il n'est pas pertinent de comparer le français parlé au Québec et en Belgique pour l'usage du présent à valeur de futur, ce qui confirme l'analyse de Poplack et Dion (2009). Cependant, il demeure judicieux d'opposer les deux variétés selon la répartition des deux variantes morphologiques du futur, à savoir le FS et le FA, et selon le conditionnement linguistique et social de leur usage.

Dans un premier temps, nous avons extrait toutes les formes synthétiques et analytiques, que nous avons classées ensuite selon leurs différentes interprétations, après exclusion des formes figées (ex. : *on verra bien*) et des verbes de mouvement (ex. : *il s'en va me porter, elle va courir dans le parc*). Les exemples 4 à 7 illustrent les 4 interprétations codifiées.

Référence à l'avenir

- 4) alors à ce moment-là il **va scanner** les documents ce **sera** plus net qu'une photocopie quoi (CFPB-1190-1, Charles, M81)²

Futur hypothétique

2. Dans le corpus CFPB, le code renvoie au code postal, suivi du numéro d'enregistrement, du pseudonyme, du sexe et de l'âge de la personne interviewée.

- 5) On **laissera** un message si c'est urgent (CFPB-1083-3b, Arlette, F51)
S'il y a un coup de vent euh l'échelle **va se retourner** (CFPB-1170-1, Robin, M21)

Futur habituel

- 6) Parce que même ceux qui sont d'origine maghrébiennne **vont parler** français entre eux autres souvent ils **parleront pas** arabe. (HOMA 027F60)³

Futur illustratif

- 7) Euh des des mots que tu **vas utiliser** avec tes amis que tu **utiliseras** pas avec ton boss. (HOMA 001M29)

Pour les besoins de la comparaison avec l'analyse des corpus de textos, nous avons exclu de l'analyse qui suit le futur habituel et le futur illustratif, et conservé deux interprétations: la référence à l'avenir et le futur hypothétique.

Les résultats

Les tendances générales dans les corpus oraux

Le tableau 3.2 présente la distribution des variantes FA et FS dans les corpus oraux. À Montréal, on constate un emploi beaucoup plus faible du FS (15,6 %) qu'à Bruxelles (40,4 %), ce qui rappelle la différence observée dans les textos (Québec: 37,3 % ; Belgique: 71,6 %).

Cette différence entre les deux communautés est significative ($p < .00001$).

TABLEAU 3.2

Distribution des variantes FA et FS dans les corpus oraux

	Montréal		Bruxelles	
	%	n	%	n
Futur analytique	84,4	1 145	59,6	180
Futur synthétique	15,6	212	40,4	122
TOTAL		1 357		302

3. Dans le corpus HOMA, le code renvoie au numéro d'identification de la personne interviewée, suivi du sexe et de l'âge.

L'analyse comparative des contraintes linguistiques à Montréal et à Bruxelles

Un examen approfondi des facteurs linguistiques éclaire la dynamique linguistique de la variation. Les cinq facteurs linguistiques sur lesquels nous nous penchons sont : l'interprétation (un facteur qui n'avait pas été testé dans l'analyse des textos), la polarité, le type de verbe, la contingence et la distance temporelle.

Le tableau 3.3 présente les résultats d'une analyse multivariée avec Goldvarb (D. Sankoff *et al.*, 2005) pour les quatre premiers facteurs. Nous rapportons séparément les résultats concernant la distance temporelle au tableau 3.4, ce facteur n'ayant pas été inclus dans l'analyse multivariée.

TABLEAU 3.3

Résultats de l'analyse multivariée pour les facteurs linguistiques en faveur du futur synthétique à Montréal et à Bruxelles

	Montréal Input = .077 n = 1 357			Bruxelles Input = .400 n = 302		
Facteur	Poids relatif	%	n	Poids relatif	%	n
Interprétation						
Référence à l'avenir	NS	15,2	1 216	.473	38,2	285
Futur hypothétique	NS	19,1	141	.860	76,5	17
Écart				39		
Polarité						
Négative	.985	84,0	194	NS	48,6	37
Positive	.333	4,2	1163	NS	39,2	265
Écart	65					
Type de verbe						
Être	.543	14,6	158	.858	79,4	34
Avoir	.592	24,5	110	.827	75,0	24
Modal	.808	32,4	37	.759	66,7	21
Autres irréguliers	.514	14,3	545	.348	26,0	123
Réguliers	.425	14,2	507	.387	31,0	100
Écart	38			51		
Contingence						
Quand	NS	13,2	38	s. o.	s. o.	s. o.
Si	NS	20,4	113			
Tant que	NS	57,1	7			
Non contingent	NS	15,0	1 199			

Interprétation

Le premier facteur porte sur l'interprétation et distingue le futur hypothétique de la référence à l'avenir. Ce groupe de facteurs ne joue un rôle significatif qu'à Bruxelles, le futur hypothétique favorisant fortement les formes synthétiques à un taux de 76,5 % (comparativement à 38,2 % pour la référence à venir), tel qu'illustré en (8) :

- 8) J'ai fait déjà une déclaration d'euthanasie et il est clair que si ça va pas euh je **passerai** l'arme à gauche (CFPB-1083-3a, Léon, M66)

Polarité

Pour le facteur Polarité, on note qu'à Montréal, comme l'illustre l'exemple (9), le contexte négatif est largement favorable à la variante synthétique avec un taux de 84 %. Cela contraste avec le contexte affirmatif, où l'on observe un taux de 4,2 % de FS. Cela pourrait suggérer que le changement vers l'analytique est pratiquement achevé en contexte affirmatif à Montréal. À Bruxelles, le contexte négatif favorise légèrement le FS à 48,6 %, mais la différence entre les contextes négatif et affirmatif n'est pas significative, peut-être en raison du faible nombre d'occurrences dans ce corpus.

- 9) Et là le professeur nous avait découragés en nous disant: « Ben vous **serez jamais** engagés. Les techniciennes **vont passer** avant vous. Vous **allez coûter** trop cher. » (FRAN-HOMA 009F55)

Type de verbe

En ce qui concerne les résultats du facteur Type de verbe, peu traité dans les études antérieures, on constate des effets importants dans les deux communautés. Même si on remarque une utilisation moins importante du FS à Montréal, ce facteur exerce un effet important similaire, avec des écarts de poids relatif de 38 à Montréal et de 51 à Bruxelles. À Montréal, comme illustré en (10), le FS est fortement favorisé pour les verbes modaux, contrairement aux autres types de verbes. On remarque que le FS est défavorisé par les verbes réguliers. À Bruxelles, le FS est fortement favorisé non seulement par les verbes modaux, mais aussi par les verbes *être* et *avoir*, comme on peut le constater dans l'exemple 11). Le FS est défavorisé par les verbes réguliers et les autres verbes irréguliers.

- 10) Fait-que c'est ça mais eux autres ils/ c'est ça ils **pourront** te parler/ elle elle **pourra** te parler à son tour. (FRAN-HOMA 012F53)

- 11) bientôt ce **sera** le nouveau les enseignants **seront** les seuls d’ailleurs à travailler jusqu’à l’âge de septante-cinq ans (CFPB-1200-2, Maminou, F70)

Contingence

Le facteur Contingence, analysé seulement dans le corpus montréalais, s’est avéré non significatif, comme dans les corpus de textos québécois et belges, et ne sera pas discuté ici.

Distance temporelle

Les résultats qui concernent la distance temporelle sont rapportés séparément au tableau 3.4, en raison des utilisations catégoriques observées.

TABEAU 3.4

Distribution du FS en fonction de la distance temporelle

Distance temporelle	Montréal		Bruxelles	
	%	n	%	n
Aujourd’hui	14,0	157	15,7	70
Demain	22,2	9	100,0	5
Après-demain et +	0,0	48	44,4	45
Non spécifié	16,4	1143	47,3	182

À Montréal, le contexte qui fait référence aux procès qui auront lieu le lendemain (Demain) présente le plus haut taux de futur synthétique, suivi des contextes Non spécifié et des événements qui auront lieu le jour même (Aujourd’hui). Les événements qui auront lieu dans un futur plus éloigné (par exemple, le surlendemain ou l’année suivante) ont suscité un emploi catégorique du futur analytique, comme en (12).

- 12) Cet été mon ami s’en va en voyage fait-que je **vas aller** nourrir là son chat. Fait-que je **vas y aller** en vélo. (FRAN-HOMA 016F44)

À Bruxelles, le contexte qui fait référence aux événements qui auront lieu le lendemain (Demain) présente le plus haut taux de futur synthétique. Dans ce corpus, cette utilisation est catégorique, mais porte sur peu d’occurrences ($n = 5$). Les contextes Non spécifié et Après-demain et + arrivent respectivement en deuxième et troisième positions. Les événements qui

auront lieu le jour même (Aujourd’hui) sont ceux qui présentent le plus faible taux de FS.

La grammaire normative qualifie le FA de futur proche. Les données du corpus belge viennent appuyer cette association. En revanche, l’utilisation catégorique du FA pour faire référence à des événements futurs plus éloignés contredit toute association entre le FA et le futur proche dans le corpus montréalais.

Analyse comparative des contraintes extralinguistiques à Montréal et à Bruxelles

Le tableau 3.5 présente les résultats de l’analyse multivariée des facteurs sociaux CSE, Genre et Âge. De ces trois facteurs, seul le genre avait pu être analysé dans les corpus de textos.

TABEAU 3.5
Résultats de l’analyse multivariée pour les facteurs extralinguistiques en faveur du futur synthétique à Montréal et à Bruxelles

	Montréal input=.150 n = 1 357			Bruxelles Input =.401 n = 302		
Facteur	Poids relatif	%	n	Poids relatif	%	n
Classe socioéconomique						
haute	.452	12,7	442	NS	40,1	152
intermédiaire	.411	11,0	429	NS	46,2	26
basse	.621	22,4	486	NS	39,5	124
écart	21					
Genre						
femme	NS	16,5	735	NS	45,2	104
homme	NS	14,6	622	NS	37,9	198
écart						
Âge						
jeunes	NS	14,9	281	.351	26,6	64
intermédiaire	NS	12,7	502			
âgés	NS	18,5	574	.541	44,1	238
écart				19		

Dans ce tableau, nous voyons qu'à Montréal, le seul facteur retenu par l'analyse multivariée est la CSE. On constate que ce sont les locuteurs de la catégorie socioéconomique basse qui favorisent le FS. En ce qui a trait à l'âge et au genre, bien que ces facteurs ne soient pas significatifs, on remarque une plus grande utilisation du FS chez les locuteurs plus âgés et chez les femmes.

En revanche, à Bruxelles, le seul facteur clair est l'âge, les jeunes présentant un faible taux de FS à 26,6 %. Selon le modèle du temps apparent, on peut donc interpréter ces résultats comme un changement en cours en faveur de l'analytique. Les deux autres facteurs analysés, soit la CSE et la différence entre les hommes et les femmes, n'ont pas d'effet notable en Belgique. Toutefois, on remarque, comme à Montréal, une plus grande utilisation du FS chez les femmes. Pour la CSE, les effets sont inversés par rapport à Montréal, puisqu'à Bruxelles, ce sont les classes haute et basse qui présentent une plus grande utilisation du FS.

Discussion : analyse contrastive des corpus oraux

Contraintes linguistiques et variation dialectale

Les résultats de l'analyse des corpus oraux confirment la variation dialectale observée dans les corpus de textos, tant au niveau des fréquences que des contraintes linguistiques. En effet, on remarque en Belgique un emploi encore très productif du FS alors qu'à Montréal, cette variante ne représente qu'une faible partie de la variation et se cantonne aux contextes négatifs. À Bruxelles, on observe un emploi encore très productif tant en contexte affirmatif qu'en contexte négatif. À Bruxelles, c'est donc l'interprétation qui semble contraindre la distribution des variantes, le contexte hypothétique favorisant fortement le FS. À Montréal, cette contrainte ne semble pas opérationnelle : la distribution de la variante synthétique se répartit à peu près également dans les deux contextes interprétatifs.

On note toutefois une certaine similarité entre les deux communautés par rapport au rôle du type de verbe dans la variation. Dans les deux cas, on observe les mêmes tendances pour les verbes modaux (qui favorisent le FS) et les verbes réguliers (qui défavorisent le FS). En revanche, on constate des différences dans le cas des verbes *avoir* et *être*, qui favorisent fortement le FS à Bruxelles, mais dans une moindre mesure à Montréal.

Quant aux autres verbes irréguliers, ils se comportent comme les verbes réguliers à Bruxelles en défavorisant le FS, alors qu'à Montréal, ils le favorisent

légèrement, ce qui les regroupe avec les verbes *avoir* et *être*. Bien que le facteur Type de verbe soit significatif dans les deux communautés, cet effet pourrait être interprété davantage comme une contrainte sémantique à Montréal (modal vs non modal), mais morphologique à Bruxelles (réguliers vs irréguliers).

Quant à la distance temporelle, on observe dans les deux communautés une fréquence plus élevée du FS pour faire référence aux événements ayant lieu le lendemain, ce facteur étant même catégorique à Bruxelles. Cependant, les deux communautés s'opposent dans deux contextes. S'agissant de la référence aux événements ultérieurs (Après-demain et +), ce contexte est catégorique en faveur du FA à Montréal, contrairement à Bruxelles, où le FS est relativement plus fréquent. En ce qui concerne la référence au jour même (Aujourd'hui), ce contexte présente des fréquences relatives de FS beaucoup plus faibles à Bruxelles qu'à Montréal. Pour ce qui relève des contextes non spécifiés, ce facteur est par définition difficilement interprétable, mais semble jouer un rôle similaire à l'intérieur de chaque communauté, ne favorisant aucune des variantes.

En résumé, l'analyse des fréquences et des contextes linguistiques montre que, contrairement à Montréal, le FS est encore très productif à Bruxelles, ce qui n'exclut pas que cette variable soit présente dans un changement en cours en faveur du FA en français contemporain. L'analyse des facteurs sociaux nous permet d'évaluer cette possibilité et de mieux comprendre la dynamique propre à chaque communauté.

Contraintes sociales et changement en cours

L'analyse des facteurs sociaux nous permet d'aborder la question du changement en cours dans les communautés, une question à laquelle notre première étude sur les textos ne nous permettait pas de répondre.

À Montréal, comme mentionné à la section « Tendances générales et facteurs favorisant l'emploi à l'oral », les études antérieures ont montré un changement en cours avancé en faveur du FA. L'analyse des facteurs sociaux pour les données recueillies en 2012 indiquerait que ce changement est mené par les classes intermédiaire et haute et que les locuteurs de la CSE basse demeurent, avec les locuteurs les plus âgés, ceux qui y résistent le plus. À Montréal, le fait que les locuteurs de la CSE basse soient ceux qui emploient le plus la variante synthétique pourrait surprendre, étant donné que le FS, associé fortement à l'écrit, est considéré comme la variante de prestige en français contemporain. Toutefois, le changement est tellement avancé chez

les locuteurs des classes haute et intermédiaire que le FA est devenu la forme par défaut intégrée dans leur vernaculaire.

À Bruxelles, on observe un effet significatif de l'âge qui indique un changement en cours, quoique moins avancé, dans la communauté. Fait intéressant, même si ce facteur n'est pas significatif, ce sont les locuteurs de la CSE basse qui emploient le plus la variante analytique, ce qui correspond aux caractéristiques d'un changement par le bas. À Bruxelles, ce sont les locuteurs de la CSE haute et les femmes qui privilégient la variante synthétique, réputée plus traditionnelle. Ces résultats indiquent que, contrairement à Montréal, la variante synthétique continue à jouer un rôle de marqueur sociolinguistique.

En conclusion, l'analyse des facteurs sociaux vient confirmer la variation dialectale observée dans l'analyse des facteurs linguistiques.

Corpus oraux vs corpus de textos

La comparaison des résultats des analyses des corpus oraux et textos nous permet maintenant d'examiner l'incidence du médium utilisé et son impact sur la variation. Le tableau 3.6 présente l'analyse distributionnelle de la variante synthétique dans les quatre corpus, les contextes négatif et affirmatif étant distingués.

TABEAU 3.6

Distribution des variantes FS dans les corpus oraux et de textos au Québec et en Belgique

	Montréal/Québec		Bruxelles/Belgique	
	Entretiens	Textos	Entretiens	Textos
Négatif	84,4 %	85 %	48,6 %	81,0 %
Affirmatif	4,2 %	33 %	39,2 %	70,2 %
Total	15,6 % (n=1 357)	37,3 % (n=678)	40,4 % (n=302)	71,6 % (n=686)

Globalement, on peut constater que le médium a un effet sur la variable : dans les deux communautés, le contexte texto privilégie fortement le FS (+21,7 % au Québec ; +31,2 % en Belgique). Ces différences sont fortes dans les deux communautés ($p < ,00001$). Cependant, à Montréal, l'effet du médium

n'est observable qu'en contexte affirmatif (+28,8 %), alors qu'à Bruxelles, les deux contextes montrent un même effet du support (Aff. +32,4 %; Nég. +31 %).

Cette comparaison montre que les textos apparaissent plus traditionnels que l'oral observé dans les entretiens. On note tout de même que les taux de FS dans les textos sont moins élevés que ceux observés dans les études sur l'écrit normé (Lesage et Gagnon, 1992; Wales, 2002), où l'on rapporte des taux de 90 % et plus.

Nos résultats viennent appuyer l'hypothèse selon laquelle les textos seraient des formes hybrides à situer sur le continuum oral-écrit (Crystal, 2011, entre autres). D'une part, les taux de FS observés dans les corpus de textos sont bien au-dessus de ce qui est observé à l'oral. D'autre part, contrairement à l'écrit standard, hautement codifié et homogène, l'écrit familier des textos reflète une langue non codifiée beaucoup plus sensible à la variation diatopique et diastratique (Cougnon et Ledegen, 2008; Tremblay, 2020). En particulier, le fait que le FS soit beaucoup plus fréquent en contexte négatif qu'en contexte affirmatif dans les textos québécois démontre que ce médium peut être régi par les mêmes contraintes linguistiques que l'oral.

* * *

La dialectologie moderne bénéficie des apports méthodologiques et heuristiques de la sociolinguistique comparative. Cette approche permet d'intégrer des déterminants tant linguistiques que sociaux de la variation, afin de mieux cerner l'évolution différentielle des dialectes. Notre étude a examiné la variation de la RTF dans deux variétés de français (québécois et belge) selon deux supports (entretiens et textos). Les résultats de notre étude confirment la différence dialectale observée préalablement dans l'étude des textos. Il y a donc plus de FS en Belgique qu'au Québec, tous médiums confondus.

Pour les corpus oraux, nous avons vu que la similarité entre les deux communautés dépend de la direction générale du changement vers l'analytique, ainsi que du rôle du type de verbe sur la variation. En revanche, les facteurs linguistiques de la polarité et de la distance temporelle ne jouent pas le même rôle dans les deux communautés. Bien qu'elles participent, l'une et l'autre, au processus de changement en cours en faveur du FA, à Bruxelles, ce changement vers l'analytique est plus lent et non achevé.

Enfin, la comparaison entre les deux médiums a révélé que, dans les deux communautés, le FS est beaucoup plus fréquent dans les textos que

dans les corpus oraux, ce qui positionne les textos à mi-chemin sur le continuum oral-écrit.

Références

- ABOUDA, L. et M. SKROVEC (2016), Du rapport entre formes synthétique et analytique du futur : étude de la variable modale dans un corpus oral micro-diachronique, *Revue de sémantique et pragmatique*, 38, 35-57, <https://doi.org/10.4000/rsp.512>.
- ABOUDA, L. et M. SKROVEC (2017), Du rapport micro-diachronique futur simple/futur périphrastique en français moderne : étude des variables temporelles et aspectuelles, *Corela*, HS21, <https://doi.org/10.4000/corela.4804>.
- BLONDEAU, H. (2007), 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, <https://doi.org/10.7202/015840ar>.
- BLONDEAU, H. et E. LABEAU (2016), La référence temporelle au futur dans les bulletins météo en France et au Québec : regard variationniste sur l'oral préparé, *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 61(3), 240-258, <https://doi.org/10.1017/cnj.2016.26>
- BLONDEAU, H. et M. TREMBLAY (2022), Écrire son vernaculaire : variation et normes communautaires dans les messages textes en français québécois, *Journal of French Language Studies*, 32(2), 120-144, <https://doi.org/10.1017/S0959269522000096>.
- BLONDEAU, H., M. TREMBLAY, A. BERTRAND et E. MICHEL (2021), A new milestone for the study of variation in Montréal French : The Hochelaga-Maisonneuve socio-linguistic survey, *Corpus*, 22, <https://doi.org/10.4000/corpus.6221>.
- BLONDEAU, H., M. TREMBLAY et P. DROUIN (2014), Hybridité et variation dans les SMS : le corpus Texto4Science et l'oralité en français montréalais, *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 59(1), 137-165, <https://doi.org/10.1017/S0008413100000189>.
- COMEAU, P. (2015), Vestiges from the grammaticalization path : The expression of future temporal reference in Acadian French, *Journal of French Language Studies*, 25(3), 339-365, <https://doi.org/10.1017/S0959269514000301>.
- COMEAU, P. et A.-J. VILLENEUVE (2016), Future temporal reference in French : An introduction. *The Canadian Journal of Linguistics/La Revue Canadienne de Linguistique*, 61(3), 231-239, <https://doi.org/10.1353/cjl.2016.0022>.
- COUGNON, L.-A. et G. LEDEGEN (2008), « c'est écrire comme je parle » : une étude comparative de variétés de français dans l'écrit sms », *Actes du Congrès annuel de L'AFLS, Oxford, 3-5 septembre 2008*, Peter Lang, 39-57.
- CRYSTAL, D. (2011), *Internet Linguistics*, Routledge, <https://doi.org/10.4324/9780203830901>.
- DESHAIES, D. et E. LAFORGE (1981), Le futur simple et le futur proche dans le français parlé dans la ville de Québec, *Langues et linguistique*, 7, 21-37.
- EDMONDS, A., A. GUDMESTAD et B. DONALDSON (2017), A concept-oriented analysis of future-time reference in native and near-native Hexagonal French, *Journal of French Language Studies*, 27(3), 381-404, <https://doi.org/10.1017/S0959269516000259>.

- EMIRKANYAN, L. et D. SANKOFF (1985), Le futur « simple » et le futur « proche », dans M. LEMIEUX et H. CEDERGREN (dir.), *Les tendances dynamiques du français parlé à Montréal*, tome 1, Office québécois de la langue française, 189-204.
- GRIMM, R. D. (2010), A real-time study of future temporal reference in spoken Ontarian French, *University of Pennsylvania Working Papers in Linguistics*, 16(2), 83-92.
- GRIMM, R. D. (2015), *Grammatical variation and change in spoken Ontario French: The subjunctive mood and future temporal reference*, thèse de doctorat, Université York, Toronto.
- JEANJEAN, C. (1988), Le futur simple et le futur périphrastique en français parlé, dans C. BLANCHE-BENVENISTE, A. CHERVEL et M. GROSS (dir.), *Grammaire et histoire de la grammaire: hommage à la mémoire de Jean Stefanini*, Université de Provence, 235-257.
- KING, R. et T. NADASDI (2003), Back to the Future in Acadian French, *Journal of French Language Studies*, 13(3), 323-337, <https://doi.org/10.1017/S0959269503001157>.
- LABEAU, E. (2014), Quand l'analytique se fait synthétique: les formes verbales périphrastiques dans le texto, *Studii de lingvistica*, 4, 131-144.
- LANGLAIS, P., P. DROUIN, A. PAULU, A. ROMPRÉ BRODEUR et F. COTTIN (2012), Texto4Science: A Quebec French Database of Annotated Short Text Messages. *Proceedings, LREC*, 1047-1054.
- LESAGE, R. (1991), Notes sur l'emploi du présent à valeur de futur dans les quotidiens québécois, *Revue québécoise de linguistique théorique et appliquée*, 10, 117-131.
- LESAGE, R. et S. GAGNON (1992), Futur simple et futur périphrastique dans la presse québécoise, dans A. CROCHETIÈRE, J.-C. BOULANGER et C. OUELLON (dir.), *Les langues menacées: actes du XV^e Congrès international des linguistes*, Presses de l'Université Laval, 367-370.
- MARTINEAU, F. et M.-C. SÉGUIN (2016), Le Corpus FRAN: réseaux et maillages en Amérique française, *Corpus*, 15, <https://doi.org/10.4000/corpus.2925>.
- POPLACK, S. (2001), Variability, frequency and productivity in the irrealis domain of French, dans J. BYBEE et P. HOPPER (dir.), *Frequency and the Emergence of Linguistic Structure*, John Benjamins, 405-428.
- POPLACK, S. et N. DION (2009), Prescription vs praxis: The evolution of future temporal reference in French. *Language*, 85(3), 557-587, <https://doi.org/10.1353/lan.0.0149>.
- POPLACK, S. et D. TURPIN (1999), Does the *Futur* have a future in (Canadian) French? *Probus*, 11(1), <https://doi.org/10.1515/prbs.1999.11.1.133>.
- ROBERTS, N. S. (2012), Future temporal reference in hexagonal French, dans *University of Pennsylvania Working Papers in Linguistics* 18(2), 97-106.
- ROBERTS, N. S. (2014), *A sociolinguistic study of grammatical variation in Martinique French*, thèse de doctorat, Université de Newcastle.
- SANKOFF, D., S. A. TAGLIAMONTE et E. SMITH, (2005), *Goldvarb X: A variable rule application for Macintosh and Windows* [Logiciel].
- SANKOFF, G. (2019), Language change across the lifespan: Three trajectory types. *Language*, 95 (2), 197-229, <https://doi.org/10.1353/lan.2019.0029>.

- SANKOFF, G. et S. E. WAGNER (2020), The long tail of language change: A trend and panel study of Québécois French futures, *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 65(2), 246-275, <https://doi.org/10.1017/cnj.2020.7>.
- TAGLIAMONTE, S. A. (2013), Comparative Sociolinguistics, dans J. K. CHAMBERS et N. SCHILLING (dir.), *The Handbook of Language Variation and Change*, <https://doi.org/10.1002/9781118335598.ch6>.
- TREMBLAY, M. (2020), Le texto : une pratique langagière distincte?, dans Kristin REINKE (dir.), *Attribuer un sens. La diversité des pratiques langagières et les représentations sociales*, Presses de l'Université Laval, 79-100
- TREMBLAY, M., H. BLONDEAU et E. LABEAU (2020), Texting the future in Belgium and Québec: Present matters, *Journal of French Language Studies*, 30(1), 73-98, <https://doi.org/10.1017/S0959269519000188>.
- WAGNER, S. E. et G. SANKOFF (2011), Age grading in the Montréal French inflected future, *Language Variation and Change*, 23(3), 275-313, <https://doi.org/10.1017/S0954394511000111>.
- WALES, M. L. (2002), The relative frequency of the synthetic and composite futures in the newspaper *Ouest-France* and some observations on distribution, *Journal of French Language Studies*, 12(1), 73-93, <https://doi.org/10.1017/S0959269502000157>.
- ZIMMER, D. (1994), Le futur simple et le futur périphrastique dans le français parlé à Montréal, *Langues et Linguistique*, 20, 213-226.

CHAPITRE 4

LangAge à l’oral et à l’écrit, au cours de la vie

Madame Moreau et les péripéties
linguistiques de l’âge avancé

Annette Gerstenberg et Marta Lupica Spagnolo

Réalisé à Orléans, en France, le projet sociolinguistique LangAge corpora partage deux intérêts avec le projet pionnier de Gillian Sankoff, David Sankoff et Henrietta Cedergren (voir Sankoff, 2017 ; Sankoff et Blondeau, 2007). D’une part, le corpus est conçu, bien qu’à une échelle nettement plus modeste, dans une perspective longitudinale, et contient des entretiens de 2005 à 2023¹. D’autre part, dans l’analyse du corpus de Montréal, la prise en compte des références générationnelles des groupes d’âge ne se limite pas à une définition chronologique de *génération* ou à l’histoire de la politique linguistique. Clermont et Cedergren (1979) développent plutôt une notion sociohistorique de *génération*. En effet, ils émettent l’hypothèse que le groupe d’âge ayant connu la crise des années 1930 ainsi que la Seconde Guerre mondiale est à l’origine d’importants changements linguistiques. Les personnes représentées dans le corpus LangAge sont nées (à quelques exceptions près) avant la fin de la Seconde Guerre mondiale. En sociologie, cette date constitue un point de rupture entre les générations d’avant et d’après la guerre en France (Drouin, 1995).

1. Le choix d’Orléans pour collecter les données pour le corpus LangAge a été motivé par la disponibilité des données de l’Étude sociolinguistique sur Orléans, ESLO1 (1968-1971 ; Bergounioux *et al.*, 1992), qui a été relayée, depuis 2008, par ESLO2 (Baude et Dugua, 2011).

Ces deux projets de corpus diffèrent cependant de manière importante. LangAge se concentre sur un seul groupe d'âge, c'est-à-dire des locuteurs et locutrices qui ont en majorité plus de 70 ans. La plupart des personnes de la première série de 2005 ont été successivement réenregistrées, et la même personne a mené tous les entretiens. Dans un cas particulier, une reprise du premier entretien a été réalisée quelques mois après le premier enregistrement. En effet, une participante, présentée ici sous le pseudonyme de Mme Moreau², avait manifesté le désir de réitérer l'entretien. L'intérêt marqué qu'elle portait au projet s'est exprimé dès la première rencontre dans un atelier d'écriture spontanée dont elle était, à 83 ans, l'une des animatrices. À l'intérieur d'une période de 9 ans, de nombreux échanges ont suivi cette première reprise de 2005, à savoir entre ses 83 et 92 ans, âge de son décès. Des échanges de lettres, de cartes postales et de vœux ont également accompagné ces entretiens. Seules les lettres et les cartes postales rédigées par Mme Moreau restent disponibles.

À l'exemple des traits linguistiques indicatifs pour situer un échantillon entre les pôles communicatifs de l'immédiat et de la distance en fonction des déterminants situationnels et contextuels (Koch et Oesterreicher, 2001, p. 586), nous suivons l'évolution de cet échange sur la base de la transcription de ces témoignages oraux et écrits. L'analyse de la variation « intrinsèque » (Gadet, 1996)³ qui en résulte nous aide à comprendre la manière dont Mme Moreau conçoit l'oralité de l'entretien et dont elle gère la scripturalité de l'échange épistolaire.

Le corpus et Mme Moreau

Le projet LangAge et son axe longitudinal

L'idée qui préside au corpus LangAge est d'étudier un seul groupe d'âge, souvent situé aux marges des échantillons sociolinguistiques, soit la tranche d'âge de 70 ans et plus, et ce, dans une même zone géographique. Les personnes, résidant à Orléans ou dans ses environs, ont été sollicitées en tant

2. Les pseudonymes utilisés dans LangAge sont choisis parmi les noms les plus fréquents listés par l'INSEE (2018) pour la décade et le département de naissance des individus.

3. Avec ce terme, Gadet souligne la place fondamentale de la variabilité : « Quel que soit le locuteur, son usage de la langue manifeste de l'hétérogène, de l'inattendu, de l'instable, de l'imprévisible, de l'éphémère, plutôt que du prévisible, de l'homogène et du monotone ; ceci se traduit en une grande souplesse et une adaptabilité à la diversité des situations » (Gadet, 1996, 32).

que témoins pour un projet d'histoire orale⁴. L'objectif était de faire ressortir la structuration interne des pratiques linguistiques de la population âgée, qui n'est que partiellement visible si l'analyse est centrée sur la comparaison intergénérationnelle. Or, cette perspective intragénérationnelle trouve sa continuation dans la série longitudinale. La majorité des 56 personnes interrogées en 2005 ont à nouveau été invitées à raconter leurs expériences à la même chercheuse en 2012 (34), en 2015 (23+25 nouvelles recrues) et en 2023 (7+3). La discussion des résultats a conduit à distinguer des traits linguistiques générationnels et des marques du vieillissement linguistique, autant en termes de grammaire (Hekkel, 2021) que de prosodie (Gerstenberg *et al.*, 2018; Fuchs *et al.*, 2021) et d'organisation des récits narratifs (Gerstenberg et Hamilton, 2023).

Les études longitudinales qui comparent des données de corpus écrits et oraux sont, à notre connaissance, peu nombreuses⁵. Dans le contexte des études sur le vieillissement langagier (Gerstenberg, 2015a), on observe souvent une approche marginalisante ou pathologisante des personnes âgées (décrite par Pichler *et al.*, 2018). En résulte le risque d'une image biaisée, compte tenu du fait que les identités «vieilles» découlent du contexte interactionnel, comme l'ont montré les études pionnières de Coupland et Coupland (1990). Par contre, les entretiens menés pour bâtir le corpus LangAge mettent les personnes participantes dans une position forte avec leurs expériences et leur savoir (Gerstenberg, 2015b). Elles expliquent à l'intervieweuse – plus jeune, soit de l'âge de leurs petits-enfants environ, et francophone non native –, les contextes historiques comme elles les ont vécus. Des informations biographiques et sociales denses ressortent, dont celles concernant les activités à la retraite. Le cadre d'histoire orale et l'écoute active de l'intervieweuse sollicitaient un discours largement monologique, contribuant à la construction de structures narratives et linguistiques complexes, tout en gardant le caractère spontané et souvent expressif de l'échange personnel.

4. Les variables extralinguistiques comprennent le sexe (sans cas de non-binarité déclaré), l'éducation sur trois échelons et le statut socioéconomique sur quatre échelons. Voir Eman El Sherbiny *et al.* (2022) pour une description plus détaillée du corpus LangAge; Gerstenberg *et al.* (2018) pour la transcription. Le corpus LangAge est géré avec l'outil LaBB-CAT (Fromont et Hay, 2012) et publié en libre accès en ligne.

5. Par exemple, Vandeweerd *et al.* (2023) au sujet des apprenants du français ou Mitzner et Kemper (2003) au sujet des manifestations linguistiques du vieillissement atypique, tel qu'observé chez les patients anglophones atteints de la maladie d'Alzheimer. Pour l'étude de lettres en perspective longitudinale, voir Voeste (2015).

Présentation de Mme Moreau

Mme Moreau (1922-2015) est issue d'une famille nombreuse qui exploitait une grande ferme près de la ville de Vendôme (région Centre-Val de Loire). Son père était actif dans la vie culturelle et politique de la ville. De plus, il transmettait son instruction, notamment musicale, à ses enfants. Bonne élève, Mme Moreau n'a pas eu l'occasion de continuer l'école jusqu'au Brevet élémentaire, ce qui aurait pu lui ouvrir des possibilités professionnelles : vers l'âge de 14 ans environ, elle fut obligée, à son grand regret, de retourner à la ferme pour s'occuper du plus jeune de ses frères.

Pendant l'occupation allemande, elle fut recrutée comme aide-soignante par un hôpital militaire allemand ; la période de la Seconde Guerre mondiale et les nuits de bombardement prennent une grande place dans ses souvenirs. Mariée à un ancien combattant des Forces françaises de l'intérieur (FFI) après la guerre, elle eut deux enfants, une fille et un garçon. En raison de sa scolarité écourtée, elle n'avait pas d'activité professionnelle propre, mais aidait son mari dans sa petite entreprise. De plus, elle participa à une chorale pendant plusieurs dizaines d'années. Elle était aussi membre d'une organisation qui animait des ateliers d'écriture spontanée. Elle ne faisait pas qu'y participer, elle animait également certains de ces ateliers.

Après la mort de son mari, Mme Moreau éménagea dans un appartement près de chez sa fille ; elle gardait de bonnes relations avec cette dernière et son mari, sa petite-fille et son arrière-petit-fils. La relation avec son fils ne fut pas aussi facile, surtout lorsqu'il commença à vivre et à travailler à Montréal. En 2012, à la suite de problèmes de santé, elle dut emménager en maison de retraite. Elle n'eut ensuite d'autre choix qu'interrompre ses activités de chant et de participation aux ateliers d'écriture spontanée. Selon les informations de sa fille, un diagnostic de démence type Alzheimer (DTA) fut établi en 2013.

Le témoignage de Mme Moreau

Mme Moreau porte un intérêt marqué au contexte sociohistorique de ce qu'elle a vécu. Ainsi, elle évoque souvent des sujets historiques, comme le rôle de la femme dans la société, les conditions de vie en France rurale et l'arrivée de la première radio, tout en racontant certaines de ses expériences personnelles (Gerstenberg et Hamilton, 2023).

Les sujets des lettres et des cartes postales sont plutôt de nature personnelle. Il y est question des nouvelles, du temps, des visites en famille ou des

visites non réalisées et du regret qu'elle en éprouve. Elle fournit dans ces correspondances, de façon régulière, des informations précises et chronologiques. Ainsi, elle spécifie le numéro du département des lieux dont elle parle et elle mentionne son âge lorsqu'elle aborde des sujets la concernant, tout comme elle le fait pour celui des enfants de la famille : l'arrière-petit-fils est toujours introduit par la formule figée [*Prénom*], *X ans*.

L'âge joue non seulement un rôle dans la chronologie de ses écrits, mais aussi dans les effets du vieillissement. Le plus souvent, Mme Moreau commente les inconvénients liés à l'âge avec sur un ton laconique (1).

1. Lettre du 20.12.2012

Je sais très bien, qu'à mon âge, je ne les reverrai pas!.. [...] enfin, avec la vieillesse, il faut « savourer » les bons moments, et accepter le reste!...

Plusieurs fois, Mme Moreau s'excuse de son écriture, de son *bavardage* ou de ses répétitions, qu'elle attribue à l'âge. Les lettres témoignent de la conscience de ne plus répondre à ses propres attentes. Cette évolution est compensée par l'effort de rester active et de maintenir un certain humour face aux défis du vieillissement (2).

2. Lettre non datée, début 2012

excusez-moi, si je vous raconte tout cela, [...], et j'ai tellement de temps à passer seule, que je m'occupe comme je peux!... Surtout, n'allez pas croire que je suis malheureuse

Analyse

Notre analyse du corpus, qui s'échelonne de la période de 2005 à 2014, se base sur un premier sous-échantillon de 9 entretiens et sur un second sous-échantillon composé de 11 cartes postales et de 6 lettres manuscrites (voir l'annexe pour une vue d'ensemble du corpus). Par ailleurs, six lettres adressées à la famille de la fille (fille, gendre et petite-fille) servent de matériel de comparaison.

Contact et situation communicative

Dans l'analyse des traits linguistiques qui établissent et gèrent le contact, nous tenons compte d'une possible évolution de la situation communicative et de la relation entre l'intervieweuse et l'interviewée, qui sont des aspects

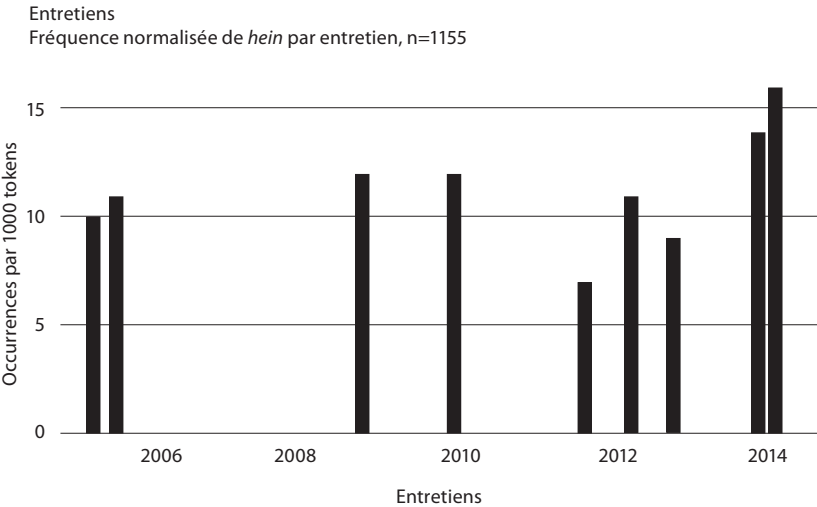
importants dans la discussion d'études longitudinales (Sankoff, 2017 ; Wagner et Tagliamonte, 2017 ; Rickford, 2021).

Pour les deux sous-échantillons, la forme d'adresse est le vouvoiement, réciproque et stable pendant toute la période. L'asymétrie de l'âge est visible lorsque Mme Moreau utilise le prénom pour appeler l'intervieweuse, alors que cette dernière emploie *Madame* [*Nom de famille*]⁶.

Dans les entretiens, le plus fréquent des marqueurs discursifs analysés par Beeching (2007) pour leur aspect pragmatico-sémantique est *hein*. D'après son analyse (ibid., p. 82), *hein* exprime une « demande d'assentiment » ou de l'« intersubjectif ». Il représente donc un moyen d'établir un contact direct entre la personne qui parle et l'interlocutrice, autrement dit un marqueur « phatique » (Kerbrat-Orecchioni, 2009). Mme Moreau fait un usage intensif de ce marqueur (figure 4.1), ce qui exprime l'idée d'interaction, même si, dans le contexte monologique de l'entretien, il n'y a que des réponses minimales de la part de l'intervieweuse (ex. mouvement de tête, interjections). Dans une perspective longitudinale, nous n'observons pas de tendance claire de sa fréquence, mais les derniers entretiens montrent une fréquence normalisée nettement élevée de *hein*.

FIGURE 4.1

Fréquence normalisée de *hein* entre 2005 et 2014



6. Dans les rencontres personnelles, Mme Moreau adoptait des comportements de type familial, comme la bise et des salutations affectives, documentés dans les notes de champ.

À l'écrit, les formules d'adresse et de salutation sont partiellement figées et hautement ritualisées (Lüger, 2007, p. 454), elles remplissent une fonction constitutive pour le genre des lettres et la structuration du texte (Biber et Conrad, 2009). Cependant, des variations sont possibles en fonction de la nature du contact à établir.

Dans les premières lettres et cartes postales, la salutation initiale se limite au prénom de l'intervieweuse (e20052, e20062, e20061, e20074, e20073, e20072, e20071). En revanche, à partir de décembre 2008, l'attribut (*ma*) *chère* apparaît avant le prénom (e20081, e20092, e20121, e20123, e20131), soulignant ici le caractère affectif de la relation. Cette salutation est la même que celle que Mme Moreau utilise le plus souvent dans ses lettres écrites à sa fille et la famille de celle-ci.

La salutation finale se compose d'éléments différents, qui peuvent être combinés librement, tels que des évocations de contact, comme la bise, les vœux et les variantes de signature. À noter que la signature est toujours formée du nom de famille – tant dans les lettres écrites à l'intervieweuse que celles écrites à sa fille.

La première lettre (e20052) exprime un haut niveau de familiarité et d'amitié : *avec des grosses bises d'une mamy de bientôt 84 printemps*. Le rapport inverse de l'âge est réinterprété comme une forme de rôle familial, de grand-mère (*mamy* est le même terme que celui qu'utilise Mme Moreau pour se désigner elle-même dans les lettres adressées à la famille de sa fille). Dans les années qui suivent (2006-2008), nous trouvons surtout des variantes de la formule *avec toutes mes amitiés* ou *amitiés sincères*, nettement plus neutres que celles citées dans la première lettre. Les exceptions sont la lettre du 2 avril 2007 et la carte postale du 24 décembre 2007, où la déclaration d'amitié est complétée par une autre formule semi-fixe, mais plus informelle *je vous embrasse* (la forme d'adresse reste toujours «vous»). Si, dans la lettre, la variation peut être attribuée à une plus grande disponibilité d'espace, ce n'est pas le cas pour la carte postale.

À partir de 2009, on observe une plus grande variabilité dans les formules de vœux, qui deviennent plus longues, affectueuses et innovantes. L'effet des formules phatiques et expressives est particulièrement renforcé dans les cas où elles s'accumulent et sont complétées par des vœux ou des expressions d'une forte affection⁷. Cela témoigne d'une évolution de la relation entre les participants vers une confiance accrue.

7. Par exemple : *Soyez heureuse là, où vous vivez, [...] Je vous embrasse très fort* (e20123).

Variables grammaticales

Nous avons choisi deux variables sociolinguistiques bien étudiées en français parlé, à savoir la non-réalisation vs réalisation de la particule de négation *ne* et la préférence pour le pronom *on* – au lieu du pronom *nous* – pour l’expression de la première personne du pluriel. Les deux variantes du *ne* réalisé et du *nous* comme première personne du pluriel représentent les formes par défaut de la distance communicative (Koch et Oesterreicher, 2011, p. 172-173 ; p. 183). De plus, leur usage nous donne une idée de ce que Mme Moreau estime approprié à l’usage dans l’entretien vs la correspondance. Ces phénomènes ont été largement étudiés en français hexagonal, notamment par Ashby (2001), Armstrong et Smith (2002) et Massot et Rowlett (2013). Ils ont également fait l’objet d’une étude entre le français hexagonal et le français du Québec dans une perspective comparative par Martineau et Mougeon (2003).

Nous avons comparé les variantes comme elles se présentent dans le premier entretien de 2005 (a013) avec celles de 2009 (fo13), de 2012 (jo13) et de 2014 (qo13). Sur le plan thématique, les entretiens sont comparables par la dominance des sujets biographiques et les récits sur le passé, complétés par des descriptions concernant sa situation actuelle, les nouvelles familiales et la santé⁸.

Ainsi, à l’oral, nous observons dès le début chez Mme Moreau un niveau bas, mais remarquable, du maintien de *ne*. Il y a 19 occurrences (occ.), soit un taux de 18,4 %, dans l’entretien de 2005. Ici, Mme Moreau utilise *ne* même dans deux expressions figées, habituellement dépourvue de négation préverbale : voir « *alors après c’est qu’il n’y avait plus de cartes d’alimentation* ». Le taux de réalisation de *ne* diminue en 2009 (11 occ., 6,6 %) et en 2012 (5 occ., 3,0 %) pour augmenter à nouveau en 2014 (9 occ., 11,8 %). Par ailleurs, en excluant les expressions figées, la tendance reste la même : 22,4 % de *ne* maintenus en 2005 ; 10,1 % en 2009 ; 3,8 % en 2012 et 18,4 % en 2014⁹.

8. L’annotation manuelle a été réalisée dans PRAAT (2023) avec le soutien de Friederike Schulz et Christian Löser. Pour les statistiques et les graphiques, nous avons utilisé le programme R sur la base de R Studio (RCoreTeam2022), avec les paquets de dplyr (Wickham et al., 2021) et ggplot (Wickham, 2016).

9. En chiffres : 2005 : *ne* maintenu n=17 (n=12 avec *pas*) plus n=2 avec une expression figée (exF) vs *ne* omis n=57 (n=43 avec *pas*) plus n=27 exF ; 2009 : *ne* maintenu n=11 (n=6 avec *pas*) vs *ne* omis n=98 (n=69 avec *pas*) plus n=57 exF ; 2012 : *ne* maintenu n=5 (n=4 avec *pas*) vs *ne* omis n=126 (n=98 avec *pas*) plus n=34 exF ; 2014 : *ne* maintenu n=9 (n=9 avec *pas*) vs *ne* omis n=40 (n=34 avec *pas*) plus n=27 exF. Nous considérons comme exF les suivantes : *c’est/c’était pas*, *je sais pas*, *il y a/avait pas*, *(il) faut/fallait pas*.

Quant à la réalisation des pronoms *on* vs *nous*, nous observons, dans la position du sujet, une fréquence assez réduite de la variante *nous*. Le pourcentage du taux de réalisation du *nous* dans la position du sujet à la première personne du pluriel diminue au fil des ans, c'est-à-dire à un taux de 6,2 % (10 occ.) en 2005; de 0,8 % (2 occ.) en 2009; de 0,5 % (2 occ.) en 2012 et de 0,5 % (1 occ.) en 2014, et ce, en comptabilisant les cas de redoublement (*nous...*, *on*) comme occurrences de *on*¹⁰.

À l'écrit, la particule de négation *ne* est maintenue dans l'ensemble des 42 occurrences, avec l'adverbe de négation *pas* comme seconde particule la plus fréquente (n=37), qui se positionne avant des occurrences très rares de *plus* (n=2), *rien*, *guère* et *que* (n=1).

Le pronom *nous* apparaît 27 fois dans la position du sujet, le plus souvent sous sa forme abrégée *ns* (21 occ.). Cette abréviation témoigne de la présence de traits graphiques renvoyant à un style de communication plus immédiat et moins planifié¹¹, présent dès le début, à la fois dans les cartes postales et dans les lettres. Par contre, Mme Moreau n'utilise pas la variante *on* à l'écrit, à l'exception d'une seule occurrence, qui se trouve dans un contexte d'un petit récit de voyage avec une exclamation soulignée, qui pourrait mettre en évidence l'aspect ludique du contenu (3).

3. Lettre du 2 juin 2009

C'est le pays natal de mon gendre, qui a une fille installée dans cette région, ns avons fêter [sic] les anniversaires de mon gendre 75 ans et ses 2 filles 45 et 40 ans!.. on a fait la fête!...

On pourrait y reconnaître une perception de sa valeur « orale » : la présence du pronom *on* en tant que première personne du pluriel, exceptionnelle dans les lettres, fait partie de l'exclamation et complète l'expression « jeune » *faire la fête*.

À côté des variantes de référence *ne* et *nous*, le sous-échantillon de l'écrit se caractérise par le choix systématique de *car* au lieu de *parce que* (7 vs 0 occ.),

10. En chiffres: 2005: *nous* n=14 (incluant 1 occurrence de *nous* en phrase clivée et 3 occurrences de *nous* répétées) vs *on* n=234, *nous on* n=14; 2009: *nous* n=2 vs *on* n=234, *nous on* n=14; 2012: *nous* n=2 vs *on* n=373, *nous on* n=32; 2014: *nous* n=1 vs *on* n=195, *nous on* n=13. Il n'y a pas de cas de *on...*, *nous* dans les entretiens.

11. Cette utilisation évoque l'ancrage historique de certains traits qui marquent la communication écrite récente, du langage texto (SMS, Blondeau et Tremblay, 2022) à la Communication Médiatisée par Ordinateur (CMO, Cougnon et Ledegen, 2010).

qui, en revanche, est la seule forme employée à l'oral. Par conséquent, la variance au niveau longitudinal dans le corpus écrit de Mme Moreau est presque inexistante, du moins pour les variables considérées. Cependant, elle préfère les variantes contemporaines, à savoir la non-réalisation de *ne* et de *on*, en communication orale, en outre plus sensible au passage du temps.

Variables pragmatiques

Au niveau pragmatique, les marques d'expressivité montrent la manière dont Mme Moreau réalise la différence entre les deux modalités, orale et écrite. Ces marques aident à situer les sous-échantillons entre l'immédiat et la distance, d'après le paramètre de l'émotionalité de Koch et Oesterreicher (2001).

Dans les données orales, nous avons mesuré la fréquence de deux adverbes d'intensité, en l'occurrence les adverbes *très* et *toujours*, des indicateurs de l'expression de l'émotion selon Biber et Conrad (2009, p. 196). L'adverbe *très* souligne souvent la valeur affective de ce qui est dit ; de plus, Mme Moreau l'utilise souvent de façon répétée (4).

4. Entretien du 11.04.2005

on est très très famille [pause] très famille

L'adverbe *toujours* comporte d'abord son sens lexical d'illimité. Dans les entretiens biographiques, son emploi souligne souvent l'importance d'un souvenir ou de l'activité même d'y penser ou d'en parler. C'est le cas des expressions *je me rappelle toujours* (9 occ.), *je me souviens toujours* (13 occurrences), *il/elle disait toujours* (11 occ.) ou *mon frère/elle/j'ai dit toujours* (3 occ.). Par ailleurs, *toujours* est utilisé avec différentes connotations, comme le montrent les occurrences dans l'entretien mené en avril 2005 (5). Le sens d'affirmation renforce la signification positive d'une description ; par contre, l'euphémisme de la forme négative (*pas toujours*) utilisée avec un adjectif positif exprime la difficulté de l'expérience vécue.

5. Entretien du 11 avril 2005

et ça je ça est toujours resté là hein

[...]

c'était pas toujours drôle hein

[...]

on avait toujours des beaux jouets

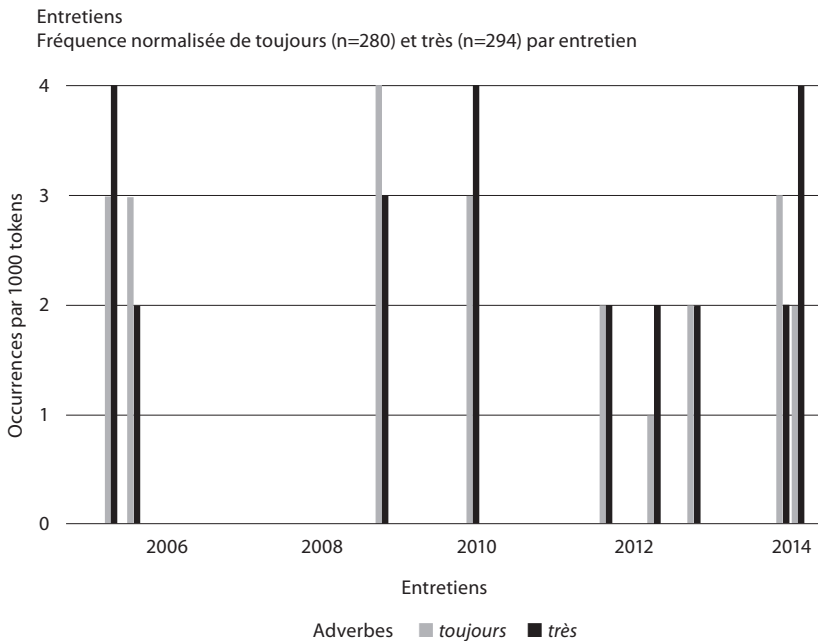
[...]

j'ai eu une vie pas toujours facile
 [...] mais on on on est toujours bien

Pour mieux estimer la valeur de ces adverbes, nous en avons comparé la fréquence normalisée (figure 4.2). Statistiquement, les deux séries de fréquences ne sont pas significativement différentes¹². Cette observation nous conduit à supposer que, au-delà des différences sémantiques, les deux adverbes peuvent être traités comme des indices interconnectés du degré de l'expressivité que Mme *Moreau* montre dans ses entretiens.

FIGURE 4.2

Adverbes d'intensité entre 2005 et 2014



Cependant, nous constatons également une évolution non linéaire de la fréquence normalisée de ces deux adverbes, qui est plus faible dans les blocs d'entretiens de 2011 et de 2012 (i013, j013, l013, o013).

12. Test de rang signé de Wilcoxon avec correction de continuité, apparié, non directionnel, médianes de la fréquence normalisée: $V = 12$, $p = 0,82$.

De plus, une analyse qualitative a montré que le dernier entretien (q013, 2014) est marqué par une utilisation partiellement différente de *très* et de *toujours* par rapport aux autres entretiens. En effet, à l'intérieur de cet entretien, *très* est fréquemment répété, comme dans l'exemple (4) ci-dessus. De plus, l'adverbe *toujours* est utilisé, dans ce même entretien, avec une référence au présent, qui semble se confondre avec le passé : au moment de l'entretien, la ferme n'était plus une propriété familiale (6).

6. Entretien du 19 octobre 2014

on avait des caves dans le rocher où on allait faire le vin c'était aménagé
pour faire du vin mh qu'on a toujours d'ailleurs mh

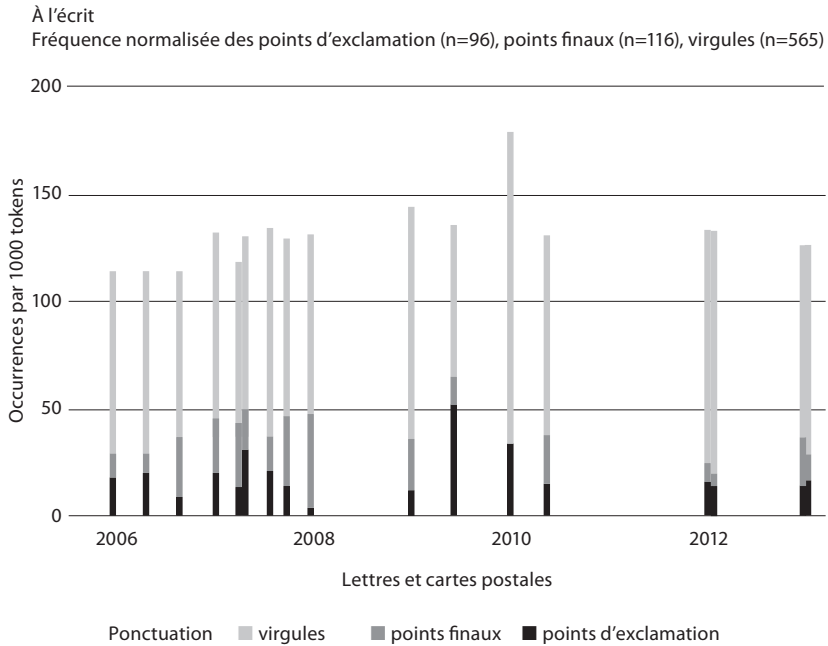
À l'écrit, ces adverbes ne sont pas utilisés aussi fréquemment que dans les entretiens (8 occ. de *toujours* et 30 occ. de *très*) et ne présentent pas les mêmes usages idiosyncratiques.

Pour analyser les traits linguistiques expressifs des textes écrits, nous avons codé la ponctuation. En tant que marqueurs d'énoncés exclamatifs, les points d'exclamation ont été considérés comme des indicateurs d'expressivité au niveau sémantico-syntaxique, comparable aux intensificateurs au niveau lexical, tels que *très* et *toujours* (Anscombe et Tamba, 2013; Bikialo et Rault, 2022). Nous en avons calculé la fréquence par rapport à la fréquence des autres signes courants, à savoir les points finaux et les virgules. À noter que Mme Moreau préfère une variante idiosyncratique du point d'exclamation composée de <!> et de <...> ou de <..>.

L'observation d'une évolution non linéaire se confirme dans les données écrites (figure 4.3), si bien que l'on ne peut pas parler d'une augmentation ou d'une diminution de la fréquence relative des signes de ponctuation examinés, mais d'une tendance curviligne, non monotone, du moins pour la période considérée.

On notera une fréquence élevée des points d'exclamation dans la carte postale déjà mentionnée, soit celle du mois de juin 2009 (e20092), dans laquelle Mme Moreau raconte un voyage en Bretagne. L'état d'esprit exprimé à l'exemple (3) pourrait motiver l'usage intensif des points d'exclamation (fréquence normalisée par 1000 *tokens*: 52). De plus, Mme Moreau fait un usage assez fréquent du point d'exclamation dans une autre carte postale de décembre de la même année (e20091), où elle évoque l'entretien (f013) et son bon déroulement (fréquence normalisée: 34). Enfin, on remarque une fréquence assez notable du point d'exclamation dans une carte postale datant

FIGURE 4.3

Signes de ponctuation entre 2005 et 2013


de 2007 (e20074; fréquence normalisée: 31), toujours centrée sur le thème des voyages.

La fréquence variable des virgules mérite aussi que l'on s'y intéresse: nous observons deux pics (fréquence normalisée >100), l'un pour les années 2008-2009 et l'autre pour les années 2012-2013 (figure 4.3).

Comme le montre l'extrait (7) de la carte postale de décembre 2009 (e20091), un emploi intensif des virgules peut correspondre à une présentation expressive et immédiate du récit.

7. Lettre du 27.12.2009

J'ai passé Noël chez ma fille, en famille, et j'attends, [sic] ma petite fille [prénom f] avec le petit [prénom m], pour le nouvel an!..

Cette expressivité est encore plus marquée dans les dernières lettres de 2012 et de 2013, caractérisées par une augmentation de l'utilisation des virgules et une diminution de l'utilisation des autres signes de ponctuation (figure 4.3). En plus, les lettres de ces années sont marquées par de nombreuses ratures (notées ici avec <c> </c>), des segments ajoutés au-dessus ou en dessous des lignes (<sup> ou <sub>), ou des erreurs d'accent (8).

8. Lettre non datée (2013)

Je t<c>e</c>nais à vous prévenir, que <c>depuis</c> 3 mois, je suis en maison de retraite médicalisée, suite à une chute que j'ai eu chez moi, le soir, <c>où</c> [accent illisible, MLS] j'ai été incapable de me relever, c'est mon gendre qui m'a retrouvée le lendemain matin. j'ai dormi par terre dans ma cuisine, toute la nuit, bien sûr, suite à cela, ^{j'ai} vu le <c>Dr</c> et mes enfants, ils n'ont pas voulu que je reste chez moi, j'ai été 1 mois à l'Hôpital (j'avais une fracture au bras, et j'ai des difficultés à marcher, maintenant Je suis donc en maison de Retraite médicalisée <c>à</c> [nom de lieu] (région _{Orléans})

Les énoncés sont enchaînés, sans délimitation forte, comme l'est un point final; cette structuration limitée du texte s'accompagne d'une cohérence moindre au niveau du contenu.

Niveaux et directions de la dynamique communicative en coupe longitudinale

Bien que les entretiens, et surtout les lettres, abordent des sujets privés, intimes, caractérisés par une forte émotivité, la relation entre l'interviewée et l'intervieweuse se caractérise par la permanence de traits de distance communicative, comme le montre l'utilisation du *vous* et d'autres formes d'adresse. À côté de ce trait partagé, la comparaison des sous-échantillons oral et écrit révèle que Mme Moreau les distingue de manière conséquente. Elle les gère avec une certaine flexibilité, ce qui se répercute dans l'analyse longitudinale. Cela nous amène à modifier l'observation de Gregersen *et al.* (2018) sur l'écart stable entre différents registres (voir aussi Kemper *et al.*, 1989) : dans le sous-échantillon écrit, dès le début, nous trouvons des signes variables d'échange amical et même familial.

Les derniers documents – écrits comme oraux – sont marqués de traces très particulières, qui sont l'expression d'une gestion plus complexe des

ressources cognitives dans la planification et la réalisation du discours. Cela concerne les derniers documents écrits, mais seulement sous l'angle de l'organisation du discours, plus vif et spontané (voir exemple 6). Quelques ratures témoignent de l'effort consenti. Dans les entretiens suivants, Mme Moreau ne répond pas avec la même facilité qu'avant et trouve différents mécanismes pour maintenir le dialogue ; l'emploi intensif du marqueur de rétroaction *hein* pourrait être considéré comme un outil de compensation (Freund et Baltes, 2002 ; Lemaire, 2015).

La combinaison de données longitudinales et de données sur la variation des registres écrit et oral peut soulever de nombreuses questions pour approcher une image plus variée et plus complexe du développement longitudinal des répertoires linguistiques des individus (voir aussi Bowie, 2019 ; Rickford, 2021) – et des aspects qui ne sont pas si dynamiques. Même si, dans le cadre de l'étude de cas, les caractéristiques ne pouvaient être traitées exhaustivement, il en résulte tout de même que le maintien du français standard à l'écrit fait partie des aspects stables. À l'oral, l'adhésion au rôle de témoin assure un trait de stabilité ; la relation amicale qui se manifeste à l'extérieur du contexte de l'entretien enregistré ne semble pas affecter l'interaction.

* * *

Nous devons, avec toute notre estime et notre gratitude, à la participation active de Mme Moreau et à son intérêt le fait que nous pouvons disposer de cette série de témoignages écrits et oraux. Sortie de l'école à un jeune âge contre sa volonté, elle a développé tout au long de sa vie un répertoire linguistique propre, par l'usage intense de l'écriture dans la correspondance privée et dans le cadre d'ateliers d'écriture.

Mme Moreau nous a permis de découvrir son usage nuancé des genres communicatifs et sa gestion adaptative des ressources linguistiques, consciente de ses conditions physiques de plus en plus fragiles, comme le montrent les remarques dans ses lettres. Autant pour la communication parlée que pour la communication écrite, ses pratiques individuelles d'écriture pourraient être considérées comme un complément stable et possiblement stabilisant de ses usages. Cela contrebalance l'effort cognitif élevé manifesté dans les témoignages des dernières années, une hypothèse qui mérite d'être explorée.

Références

- ARMSTRONG, Nigel et Alan SMITH (2002), The Influence of Linguistic and Social Factors on the Recent Decline of French “ne”, *Journal of French Language Studies*, 12, 23-41, <http://www.doi.org/10.1017/S0959269502000121>.
- ASHBY, William J. (2001), Un nouveau regard sur la chute du *ne* en français parlé tourangeau: s'agit-il d'un changement en cours?, *Journal of French Language Studies*, 11(1), 1-22, <http://www.doi.org/10.1017/S0959269501000114>.
- ANSCOMBRE, Jean-Claude et Irène TAMBA (2013), Autour du concept d'intensification, *Langue française*, 177(1), 3-8, <http://www.doi.org/10.3917/lf.177.0003>.
- BAUDE, Olivier et Céline DUGUA (2011), (Re)faire le corpus d'Orléans quarante ans après: «quoi de neuf, linguiste», *Corpus*, 10, 99-118, <http://www.doi.org/10.4000/corpus.2036>.
- BEECHING, Kate (2007), La co-variation des marqueurs discursifs «bon», «c'est-à-dire», «enfin», «hein», «quand même», «quoi» et «si vous voulez»: une question d'identité?, *Langue française*, 154(2), 78-93.
- BERGOUNIOUX, Gabriel, Jean BARADUC et Céline DUMONT (1992), L'étude sociolinguistique sur Orléans (1966-1991): 25 ans d'histoire d'un corpus, *Langue française*, 93, 74-93, <http://www.doi.org/10.3406/lfr.1992.5812>.
- BIBER, Douglas et Susan CONRAD (2009), *Register, Genre, and Style*, Cambridge, Cambridge University Press.
- BIKIALO, Stéphane et Julien RAULT (2022), Expressivité, exclamation et ponctuation, *Langages*, 228(4), 57-72.
- BLONDEAU, Hélène et Mireille TREMBLAY (2022), Écrire son vernaculaire: variation et normes communautaires dans les messages textes en français québécois, *Journal of French Language Studies*, 32(2), 120-144, <http://www.doi.org/10.1017/S0959269522000096>.
- BOWIE, David (2019), Individual variation in the development of the western vowel system of Utah, *Linguistics Vanguard*, 5(2), <http://www.doi.org/10.1515/lingvan-2018-0020>.
- CLERMONT, Jean et Henrietta J. CEDERGREN (1979), Les «r» de ma mère sont perdus dans l'air, dans Pierrette THIBAUT (dir.), *Le français parlé. Études sociolinguistiques*, Amsterdam, Linguistic Research, 13-28.
- COUGNON, Louise-Amélie et Gudrun LEDEGEN (2010), «c'est écrire comme je parle». Une étude comparatiste de variétés de français dans l'écrit sms, dans Michaël ABECASSIS et Gudrun LEDEGEN (dir.), *Les voix des Français*, Oxford/New York, Lang, 39-70.
- COUPLAND, Nikolas et Justine COUPLAND (1990), Language and Later Life dans Howard Giles et Peter Robinson (dir.), *Handbook of Language and Social Psychology*, Chichester, Wiley, 451-468.
- DROUIN, Vincent (1995), *Enquêtes sur les générations et la politique 1958-1995*, Paris, L'Harmattan.
- EMAN EL SHERBINY, Ismail, Annette GERSTENBERG, Marta LUPICA SPAGNOLO, Friederike SCHULZ et Anne VANDENBROUCKE (2022), L'âge avancé en perspective longitudinale et ses outils: LangAge, un corpus au pluriel, *SHS Web Conf.* (SHS

- Web of Conferences*), 8^e Congrès mondial de Linguistique française, 138(10003), 1-14. <http://www.doi.org/10.1051/shsconf/202213810003>.
- FREUND, A. M. et P. B. BALTES (2002), Life-Management Strategies of Selection, Optimization, and Compensation: Measurement by Self-Report and Construct Validity, *Journal of Personality and Social Psychology*, 82(4), 642-662.
- FROMONT, Robert et Jennifer HAY (2012), LaBB-CAT: an Annotation Store, dans Paul COOK et Scott NOWSON (dir.), *Proceedings of Australasian Language Technology Association Workshop 2012*, Dunedin (N.-Z.), Otago University, 113-117, <https://aclanthology.org/U12-1015.pdf>.
- FUCHS, Susanne, Laura L. KOENIG et Annette GERSTENBERG (2021), A Longitudinal Study of Speech Acoustics in Older French Females: Analysis of the Filler Particle *eah* across Utterance Positions, *Languages*, 6(4), 211, <http://www.doi.org/10.3390/languages6040211>.
- GADET, Françoise (1996), Niveaux de langue et variation intrinsèque, *Palimpsestes*, 10, 17-40.
- GERSTENBERG, Annette (2015a), Langue et générations: enjeux linguistiques du vieillissement, dans Claudia POLZIN-HAUMANN et Wolfgang SCHWEICKARD (dir.), *Manuel de linguistique française*, Berlin, de Gruyter, 314-333, <http://www.doi.org/10.1515/9783110302219-016>.
- GERSTENBERG, Annette, (2015b), A Sociolinguistic Perspective on Vocabulary Richness in a Seven-Year Comparison of Older Adults, dans Annette GERSTENBERG et Anja VOESTE (dir.), *Language Development: the Lifespan Perspective*, Amsterdam, John Benjamins, 109-127, <http://www.doi.org/10.1075/impact.37.06ger>.
- GERSTENBERG, Annette, Valerie HEKKEL et Julie KAIRET (2018), *Corpus LangAge: Transcription Guide*, University of Potsdam, Department of Romance Studies, <http://www.doi.org/10.5281/zenodo.6444538>.
- GERSTENBERG, Annette et Heidi E. HAMILTON (2023), « Older adults » conversations and the emergence of « narrative crystals »: a new approach to frequently told stories », *Narrative Inquiry*, 33(1), 27-60, <http://www.doi.org/10.1075/ni.21075.ger>.
- GREGENSEN, Frans, J. Normann JØRGENSEN, Janus SPINDLER MØLLER, Nicolai PHARAO et Gert FOGET Hansen (2018), Sideways: five methodological studies of sociolinguistic interviews, *Acta Linguistica Hafniensia*, 50(1), 1-35, <http://www.doi.org/10.1080/03740463.2017.1335111>.
- HEKKEL, Valerie (2021), *Eine soziolinguistische Betrachtung von parce que-Strukturen in Synchronie und Diachronie*, Universität Potsdam, publishUP, <http://www.doi.org/10.25932/publishup-51396>.
- INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES (2018), *Fichier des noms* (paru le 22/05/2018), Paris, INSEE, <https://www.insee.fr/fr/statistiques/3536630>.
- KEMPER, Susan, Donna KYNETTE, Shannon RASH et Kevni O'BRIEN (1989), Life-span changes to adults' language. Effects of memory and genre, *Applied Psycholinguistics*, 10, 49-66.
- KERBRAT-ORECCHIONI, Catherine (1990), *Les interactions verbales*, tome 1, Paris, A. Colin.

- KOCH, Peter et Wulf OESTERREICHER (2001), Langage parlé et langage écrit, dans Günter HOLTUS et Michael METZELTIN, Christian SCHMITT (dir.), *Lexikon der romanistischen Linguistik*: Tübingen, Niemeyer, 584-627.
- KOCH, Peter et Wulf OESTERREICHER (2011), *Gesprochene Sprache in der Romania*, Berlin, De Gruyter.
- LangAge corpora, Université de Potsdam. www.langage-corpora.org.
- LEMAIRE, Patrick (2015), *Vieillessement cognitif et adaptations stratégiques*, Louvain-la-Neuve, De Boeck supérieur.
- LÜGER, Heinz-Helmut (2007), Pragmatische Phraseme: Routineformeln, dans Harald BURGER, Dmitrij DOBROVOL'SKIJ, Peter KÜHL et Neal R. NORRICK (dir.), *Phraséologie/Phraseology*, tome 1, Berlin, De Gruyter, 444-459.
- MITZNER, Tracy L. et Susan KEMPER (2003), Oral and written language in late adulthood: findings from the Nun Study, *Experimental aging research*, 29(4), 457-474, <http://www.doi.org/10.1080/03610730303698>.
- MARTINEAU, France et Raymond MOUGEON (2003), A Sociolinguistic Study of the Origins of *ne* Deletion in European and Quebec French, *Language* 79(1), 118-152, <http://www.jstor.org/stable/4489387>.
- MASSOT, Benjamin et Paul ROWLETT (2013), Le débat sur la diglossie en France: aspects scientifiques et politiques, *Journal of French Language Studies*, 23, 1-16, <http://www.doi.org/10.1017/S0959269512000336>.
- PICHLER, Heike, Suzanne E. WAGNER et Ashley HESSON, (2018), Old-age language variation and change: Confronting variationist ageism, *Language and Linguistics Compass*, 12(6), 1-21, <http://www.doi.org/10.1111/lnc3.12281>.
- POSIT SOFTWARE, *Posit (formerly RStudio) 22.12.0*, Boston (MA), PBC, 2022. <https://posit.co/>.
- PRAAT = BOERSMA, Paul et David WEENINK. *Praat: Doing Phonetics by Computer*, 6.3.09, 2023, <http://www.praat.org>.
- R PROJECT = R DEVELOPMENT CORE TEAM, *R: A Language and Environment for Statistical Computing (R 4.2.2)*, 2022. <http://www.r-project.org>.
- RICKFORD, John R. (2021), Stylistic Variation in Panel Studies of Language Change, dans Karen V. BEAMAN et Isabelle BUCHSTALLER (dir.), *Language variation and language change across the lifespan. Theoretical and empirical perspectives from panel studies*, New York, Routledge, 77-97.
- SANKOFF, Gillian (2017), Before There Were Corpora, dans Suzanne E. Wagner et Isabelle Buchstaller (dir.), *Panel Studies of Variation and Change*, New York, Routledge, 21-51.
- SANKOFF, Gillian et Hélène BLONDEAU (2007), Language change across the lifespan: /r/ in Montreal French, *Language*, 83(3), 560-588, <http://www.doi.org/10.1353/lan.2007.0106>.
- VANDEWEERD, Nathan, Alex HOUSEN et Magali PAQUOT (2023), Comparing the longitudinal development of phraseological complexity across oral and written tasks, *Studies in Second Language Acquisition*, 45(4), 787-811, <http://www.doi.org/10.1017/S0272263122000389>.

- VOESTE, Anja (2015), Tired mind or tired hand Linguistic changes in the private letters of a Baltic German nobleman, dans Annette Gerstenberg et Anja Voeste (dir.), *Language Development: the Lifespan Perspective*, Amsterdam, Benjamins, 167-188.
- WAGNER, Suzanne E. et SALI A. Tagliamonte (2017), What Makes a Panel Study Work Researcher and Participant in Real Time, dans Suzanne E. WAGNER et Isabelle BUCHSTALLER (dir.), *Panel Studies of Variation and Change*, New York, Routledge, 2-32.
- WICKHAM, Hadley, *ggplot2: Elegant graphics for data analysis (Use R!)*, New York (NY), Springer, 2016, <http://www.doi.org/10.1007/978-3-319-24277-4>.
- WICKHAM, Hadley, François ROMAIN, Lionell HENRY et Kirill MÜLLER (2021), *dplyr: A Grammar of Data Manipulation*, 2021, <https://CRAN.R-project.org/package=dplyr>.

Annexe

TABLEAU 4.1

**Corpus témoignages de Mme Moreau. Type : écrit (E), oral (O),
code interne (code), date, âge, mots graphiques (tokens)**

type	code	date	âge	token
O	a013	2005-04-11	83	10 469
O	b013	2005-07-26	83	10 992
E	e20052	2005-12-20	83	282
E	e20062	2006-04-24	84	203
E	e20061	2006-08-24	84	216
E	e20077	2007-01-05	84	196
E	e20075	2007-04-02	85	302
E	e20074	2007-04-23	85	160
E	e20073	2007-07-24	85	193
E	e20072	2007-09-24	85	276
E	e20071	2007-12-24	85	249
O	e013	2008-09-29	86	12 042
E	e20081	2008-12-29	86	421
E	e20092	2009-06-02	87	154
O	f013	2009-12-06	87	12 491
E	e20091	2009-12-27	87	89
E	e20101	2010-05-14	88	474
O	i013	2011-00-00	88	13 929
O	j013	2012-08-16	89	15 686
E	e20122	2012-01-05	89	345
E	e20121	2012-12-20	90	995
E	e20123	2012 (non daté)	90	901
O	l013	2012-03-30	90	10 084
O	o013	2013-10-07	90	12 115
E	e20131	2013 (non daté)	91	421
O	q013	2014-10-19	92	6 835
Total E				5 877
Total O				94 174
Total E+O				100 051

Récits de vie, communauté et appartenance sociale

CHAPITRE 5

Le pouvoir des mots

La toponymie et la gentrification dans le quartier Hochelaga-Maisonneuve entre 2010 et 2012

Claire Bourély

Ce chapitre porte sur deux corpus, l'un est constitué d'entrevues sociolinguistiques recueillies entre 2012 et 2014 à Montréal dans le quartier Hochelaga-Maisonneuve (HM) et l'autre, d'articles de médias traditionnels écrits, montréalais, évoquant HM, de 2010 à 2012.

Enclavé entre le fleuve Saint-Laurent au sud, l'arrondissement Ville-Marie à l'ouest, l'arrondissement Rosemont-La-Petite-Patrie au nord et le quartier Mercier à l'est, HM est une portion de l'arrondissement comprenant également Mercier. Au début du siècle, le paysage urbain de HM a connu de grands changements : « Pas moins de 6 280 logements s'y sont construits entre 2002 et 2012, selon la Société canadienne d'hypothèques et de logement. Seuls le centre-ville et Saint-Laurent ont vu davantage de maisons et de condos pousser durant cette période. » (*Mercier-Hochelaga-Maisonneuve*, 2013)

Nous nous sommes donné pour objectif de comprendre, pendant cette période de mutation urbaine, comment les habitants de HM et les médias ont utilisé les toponymes pour désigner le quartier. Nous explorons les préférences toponymiques des locuteurs du corpus, ainsi que les tendances des médias, pour examiner ensuite quels noms sont en progression, qui les utilise et pourquoi.

Notre étude se concentre sur l'emploi des toponymes officiels (Hochelaga, la partie ouest du quartier, Hochelaga-Maisonneuve, le double quartier à

l'étude, Mercier-Hochelaga, l'association des parties est et ouest, et Mercier-Hochelaga-Maisonneuve, le nom de l'arrondissement), et des toponymes non officiels (HoMa, et Hochelag) dans les entrevues (citations 1) et dans les pages des trois journaux montréalais (*Le Devoir* [citations 2], *La Presse* [citations 3] et *Le Journal de Montréal* [citations 4]), avec une approche combinant les méthodes quantitative et qualitative.

Cadre contextuel

L'action de nommer un quartier est porteuse de sens. Comme le souligne Dugas :

l'attribution d'un nom à un lieu marque son appartenance à un milieu humain particulier; il fait désormais partie, il appartient au sens littéral du mot, à un univers culturel, à un groupe de personnes dont il reflète la spécificité, et cela, même si la dénomination fut attribuée par un nommant, celui-ci véhiculant le bagage culturel de la communauté à laquelle il se rattache. (1987, p. 441)

Certaines appellations ne sont pas innocentes et, parfois, « au “dénotatif” s'adjoint le “connotatif” : une plus-value toponymique » (Christian, 1999, cité dans Boyer, 2008, p. 5). Dans le cas de HM, le nom Hochelaga est très symbolique, puisqu'il représente le premier nom connu pour désigner le territoire où est située la Ville de Montréal. Lorsqu'il est associé à Maisonneuve, pour évoquer le quartier qui s'étend plus à l'est, Hochelaga est en contact avec son premier bâtisseur, Paul Chomedey, sieur de Maisonneuve. L'origine de la ville vient largement connoter le nom officiel du quartier. Boyer (2008) parle du fait de nommer comme d'un acte « glot-topolitique », visant à « revendiquer l'inscription du toponyme dans une communauté linguistique et/ou culturelle ». Mais qu'en est-il du nom HoMa, créé de toutes pièces par des promoteurs sur le modèle SoHo (Germain et Rose, 2010) ou de l'appellation apocopée Hochelag?

D'un point de vue historique, HM est un ancien quartier industriel peuplé d'anciens ruraux de la Grande Plaine de Montréal (Linteau, 1981). Au début du xx^e siècle, c'est un quartier prospère qui offre de nombreux emplois aux ouvriers, mais dans la deuxième partie du xx^e siècle, comme un peu partout en Amérique du Nord, HM connaît une importante phase de désindustrialisation. S'ensuit alors une période de déclin économique qui a pour conséquence l'augmentation du chômage et de la pauvreté. À la fin du xx^e siècle, on assiste à une gentrification inhérente à l'arrivée de nouveaux résidents issus de la classe moyenne. Cette cohabitation de populations issues

de milieux sociaux différents permet d'examiner les effets de la gentrification et de la mixité sociale (Blondeau et Tremblay, 2012).

La gentrification est un processus qui transforme des quartiers délabrés, habités par des populations à faible revenu, en zones rénovées attirant une population aisée, ce qui cause souvent une hausse des prix immobiliers et l'exclusion économique et sociale des résidents d'origine (Lees et Ley, 2008). Ce processus entraîne l'augmentation du coût de la vie et l'éviction des habitants historiques.

Mais quel rapport entretient la toponymie avec la gentrification ? D'une part, la gentrification, qui modifie le paysage urbain, et l'appropriation de l'espace, qui reflète le rapport des résidents avec leur environnement, sont interdépendantes (Bélanger, 2010; Nastase, 2021; Paré et Mounier, 2021; Rose, 2006). D'autre part, le lieu de vie et l'appartenance au quartier sont importants pour la construction de l'identité individuelle. Les résidents d'origine se sentent attachés à leur quartier en raison de leur histoire familiale et de leur communauté, ainsi que des caractéristiques physiques du quartier (Ilbury, 2022). En entraînant des changements structurels, la gentrification affecte la construction de l'identité des résidents d'origine.

Les effets de la gentrification sur les perceptions identitaires ont été étudiés dans un quartier londonien, où les résidents d'origine ont exprimé leur sentiment de perte d'identité et de marginalisation culturelle en raison des changements socioéconomiques du quartier (Ilbury, 2022). Il en est ressorti que le quartier changeait d'image selon que les locuteurs étaient « gentrifiés » ou « gentrificateurs ».

Plusieurs études ont déjà fait état de l'avancée des connaissances sur la gentrification à HM (Ghaffari, 2020; Nastase, 2021; Paré et Mounier, 2021). Nastase montre que la gentrification, qui met en présence plusieurs catégories de population, remet en question l'identité de quartier des habitants. Paré et Mounier examinent l'impact de l'immigration et de la gentrification sur la transformation du paysage économique et culturel de HM, et l'image que les médias renvoient de ce quartier. Les autrices ont relevé les noms « Hochelaga-Maisonneuve » et « HoMa » dans un corpus de 6 quotidiens montréalais pendant 8 ans, entre 2010 et 2018. Au total, elles ont réalisé une étude portant sur 1 420 articles jugés pertinents. Dans leur étude, Paré et Mounier montrent qu'il ne faut pas sous-estimer l'impact de la gentrification sur l'identité des résidents du quartier; en effet, elles constatent que les habitants historiques perçoivent le statut aisé des nouveaux résidents comme une source de tensions plus forte que ne l'est leur origine ethnoculturelle :

Dans notre quartier cible, il semblerait que le statut socioéconomique plus élevé des nouveaux habitants soit plus perturbateur que leur origine ethnoculturelle, car il est associé à un changement dans la façon dont les gens vivent, font leurs courses et interagissent dans l'espace¹. (Paré et Mounier, 2021) (Traduction libre)

Le déplacement croissant des populations à faible revenu est causé par le manque de logements abordables dans le quartier. Les changements sociaux dans la façon dont les gens vivent, font leurs courses et interagissent dans l'espace public sont aussi un aspect de la gentrification. Enfin, les autrices constatent que les perceptions de HM incluent des changements sociaux, de nouvelles habitudes sociales, et génèrent souvent des conflits socioterritoriaux. Paré et Mounier concluent en soulignant l'importance de surveiller les impacts des processus de gentrification et d'immigration sur les populations locales, notamment pour l'accès aux logements abordables et les changements sociaux dans l'espace public. Elles estiment qu'il est important de mieux comprendre les perceptions et les conflits susceptibles de découler de ces changements, pour trouver des solutions permettant une coexistence harmonieuse entre les différentes populations de HM qui vivent des tensions liées à l'appropriation de l'espace.

Dans leur méthodologie, les autrices ont sélectionné les articles en utilisant les toponymes Hochelaga-Maisonneuve et HoMa, mais en ignorant les autres noms donnés au quartier ainsi que la représentativité de la population à travers l'identité véhiculée par chaque média. De ce fait, il reste encore des questions ouvertes sur la participation des médias à la difficulté d'appropriation de l'espace et sur la représentation des perceptions des habitants du quartier dans les médias telles que véhiculées par les discours dans les entrevues. Comment se manifeste cette difficile appropriation de l'espace ? Est-ce que les médias participent à cette difficulté ? Est-ce qu'ils la modèrent ? L'étude des médias rend-elle compte des perceptions des habitants de HM ?

Cadre méthodologique

Notre approche se situe au carrefour des méthodes quantitative et qualitative. Nous avons procédé de manière itérative (Glaser et Strauss, 1999) en partant des entretiens sociolinguistiques du sous-corpus FRAN-HOMA 2012, et en

1. *In our target neighbourhood, it would appear that a higher socio-economic status of newcomers is more disruptive than their ethno-cultural background because it is associated with a change in the way people live, shop and interact in public space.*

nous concentrant sur les discours sur HM. Nous avons ensuite travaillé sur trois médias montréalais afin de comparer le discours médiatique à ceux des locuteurs, ce qui a fait émerger de nouvelles catégories.

Le corpus FRAN-HOMA 2012 : description et méthode d'analyse

Le sous-corpus du corpus FRAN-HOMA 2012 (Blondeau *et al.*, 2021 ; Martineau et Séguin, 2016) regroupe 38 des 50 locuteurs interviewés entre 2012 et 2014. Ce corpus stratifié (18 femmes et 20 hommes) présente un nombre équivalent de locuteurs regroupés en trois catégories d'âge et six catégories socioprofessionnelles, tout en comportant une égalité presque parfaite entre les hommes et les femmes.

Recrutés majoritairement de bouche à oreille, les participants devaient parler français, être âgés de plus de 18 ans et résider dans HM depuis au moins 5 ans. Les entretiens ont duré en moyenne une heure et comportaient tous un volet sur le quartier, son évolution et la question de la gentrification.

Après avoir importé les entretiens dans QDA Miner, un logiciel d'analyse qualitative (Miner, 4), nous avons isolé les moments au cours desquels les locuteurs parlaient de HM. À la lumière de ce qui est ressorti des articles médiatiques, nous avons codé les mots clés Hochelag, Hochelaga, Hochelaga-Maisonneuve et HoMa, en utilisant une grille d'analyse qui catégorise les termes selon leur connotation, leur fréquence et leur contexte d'utilisation. Cette approche nous a permis de déceler des tendances dans le discours et d'identifier les attitudes des locuteurs face au quartier.

Les médias : description du corpus et méthode d'analyse

Les trois grands médias à l'étude, *Le Devoir*, *La Presse* et *Le Journal de Montréal* entre 2010 et 2012 s'adressent à une population francophone diversifiée. *Le Devoir* est considéré comme un quotidien intellectuel tandis que *La Presse* est davantage accessible. Enfin, *Le Journal de Montréal* est un quotidien populaire. Cela permet une bonne représentation des différentes couches de population, à l'instar du corpus.

Pour constituer le corpus, nous avons interrogé Eureka pour *Le Devoir*, *La Presse* et *Le Journal de Montréal* entre le 1^{er} janvier 2010 et le 31 décembre 2012 avec les requêtes suivantes :

« Hochelag* », puis « Hochelaga-Maisonneuve », « Hochelaga » sans « Maisonneuve », et enfin, « HoMa ». Parmi ces trois sources, 1 383 articles sont sortis et ont été triés selon plusieurs critères. Nous avons conservé tous les articles, sans nombre de mots minimum². Tous les articles qui ne concernaient pas HoMa ont d’abord été exclus (la ville Homa en Syrie, ainsi que les noms de personnes). Nous avons ensuite éliminé les doublons et pratiqué des exclusions plus fines. Le groupe de musique Hochelaga et toutes les mentions y faisant référence ont été exclus, la rue Hochelaga, l’école Hochelaga, le couvent Hochelaga, le village Hochelaga, ainsi que toute référence historique aux origines de Montréal. Enfin, nous avons exclu les enseignes contenant le nom Hochelaga ou HoMa (Phoenix Homa, Zone Homa, la boulangerie Arhoma, Salon Hochelaga mon Amour, Taxi Hochelaga, Restaurant Jardin Hochelaga, Brasserie Hochelaga, Princesses d’Hochelaga et Sportifs [d’] Hochelaga, ainsi que le titre du film *Hochelaga*). L’analyse a porté sur 1 153 articles. Le tableau 5.1 présente les trois phases d’extraction des articles des trois quotidiens.

TABLEAU 5.1
Médias faisant référence au quartier entre 2010 et 2012

Média	Articles extraits	Articles uniques et pertinents géographiquement	Articles analysés
<i>La Presse</i>	558	528	462 (206 927 mots)
<i>Le Journal de Montréal</i>	424	407	350 (277 763 mots)
<i>Le Devoir</i>	401	375	337 (165 771 mots)
Total	1 383	1 310	1 153 (484 614 mots)

- Après nettoyage des données, nous avons procédé en quatre étapes :
- Une lecture rapide des articles pour identifier les thèmes pertinents (Deslauriers et Kérisit, 1997; Lincoln *et al.*, 2018; Olivier de Sardan, 2008; Poupart *et al.*, 1997).

2. Paré et Mounier (2021), dans leur collecte de données assez similaires, ont choisi d’exclure les articles de moins de 200 mots ainsi que les articles traitant de sujets généraux. Cette étude les conserve, car ils mentionnent le quartier et témoignent d’une toponymie factuelle et dépourvue d’attitude, ce qui permet d’étudier des usages plus spécifiques.

- Une analyse de mots clés par QDA pour les termes Hochelag, Hochelaga, Hochelaga-Maisonneuve, Mercier-Hochelaga, Mercier-Hochelaga-Maisonneuve et HoMa. Cette phase a montré la variabilité des toponymes pour décrire le quartier.
- Une analyse des différences significatives entre les quotidiens en ce qui concerne les toponymes et l'utilisation des toponymes entre 2010 et 2012.
- Une analyse qualitative de certains des articles les plus pertinents pour les comparer aux constats dégagés par l'analyse quantitative.

Les résultats chez les locuteurs du sous-corpus FRAN-HOMA 2012

Description quantitative

Dans notre analyse sociolinguistique du sous-corpus, nous avons observé que, si tous les locuteurs ont mentionné au moins une fois l'un des noms du quartier, plusieurs d'entre eux ont utilisé différents toponymes pour le désigner. Pour explorer ce phénomène, nous avons procédé à une analyse quantifiée des transcriptions des entretiens. À cette fin, nous avons écarté seulement trois occurrences où les locuteurs faisaient référence à la rue Hochelaga plutôt qu'au quartier.

Le tableau 5.2 présente la distribution des toponymes parmi les locuteurs et leur fréquence dans les discours.

La majorité des locuteurs utilisent Hochelaga (N=137) et Hochelaga-Maisonneuve (N=118), des termes plutôt neutres. Nous montrerons, en

TABEAU 5.2

Nombre d'occurrences des toponymes dans le corpus FRAN-HOMA 2012

Toponyme	Nombre d'occurrences	Locuteurs utilisant le toponyme	% des occurrences par locuteur
Hochelaga	137	31	81,60 %
Hochelaga-Maisonneuve	118	27	71,10 %
Hochelag	16	10	26,30 %
HoMa	9	5	13,20 %
Mercier-Hochelaga	1	1	2,60 %
Mercier-Hochelaga-Maisonneuve	1	1	2,60 %
	N=282	N-38	

analysant les discours des locuteurs, qu'Hochelag et HoMa présentent des variations identitaires significatives. Ces différences linguistiques sont porteuses de sens et scindent l'espace urbain en deux.

Analyse qualitative

L'analyse du discours des locuteurs montre qu'il y a deux espaces associés à des valeurs différentes : l'espace nommé Hochelaga/Hochelag vs l'espace Hochelaga-Maisonnette/HoMa. Cette double appellation peut entraîner des problèmes d'identification au lieu et d'appartenance individuelle (Dugas, 1987). Afin d'illustrer les enjeux de l'appartenance associés à la question du choix des noms pour désigner HoMa, voici une analyse du discours de trois locuteurs représentatifs.

Gabrielle : contre HoMa, pour Hochelag

Gabrielle, âgée de 18 ans et résidente de HM depuis sa naissance, issue d'une catégorie socioéconomique basse, met en lumière les enjeux de l'identité et de l'appartenance associés à l'utilisation des différents toponymes pour désigner le quartier. Selon Gabrielle, au moment de la grande grève étudiante de 2012, le terme Hochelag s'est démocratisé en étant largement utilisé, mais certains considèrent l'utilisation de ce nom comme une perte de prestige social. D'un autre côté, HoMa s'est mis à émerger. En tant que résidente, Gabrielle craint que l'utilisation de ce toponyme soit associée à la gentrification et à l'éviction de la population défavorisée. Elle souhaite que le quartier conserve son identité actuelle sans devoir déménager. Gabrielle précise dans son discours que sa mère, qui n'habite pas le quartier, préfère le nom HoMa, qui représente pour elle une promesse de prestige social et de développement économique :

- 1.a. Hochelag il y a des gens qui disent qu'on peut pas dire ça qu'on habite ici. [...] Hochelag... Mais moi c'est comme ça que je l'appelle parce que... Avec la grève tout le monde appelait un peu le quartier comme ça. [...] Mais l'autre appellation qui existe c'est HoMa. Pis HoMa, on en rit un peu parce que c'est vraiment trop pompeux. Ça fait un peu huppé. [...] Le mot qui désigne tout ce qui est en train de changer dans le quartier là. On veut pas nécessairement... Ma mère, elle trouve ça super. Elle comprend pas pourquoi moi je trouve ça pas correct elle est comme... « Gabrielle le quartier se développe on va peut-être être le prochain Plateau-Mont-Royal. » Mais non (rire) c'est pas ça que je veux. Je veux que le quartier reste pareil. Et que les gens soient pas obligés de déménager. *Gabrielle, FRAN-HOMA 2012.*

L'utilisation d'HoMa reflète l'image gentrifiée du quartier souhaitée par les promoteurs immobiliers pour attirer une clientèle plus aisée. Cette image est rejetée par certains habitants, comme Gabrielle, qui considèrent HoMa comme le symbole de gentrification et de menace d'expulsion de la population locale.

Paul : contre HoMa, pour Hochelaga-Maisonneuve

De même, Paul, un animateur de 49 ans, né dans l'est de Montréal, mais qui vit dans HM depuis l'âge de 3 ans, fait aussi le lien entre HoMa et la gentrification du quartier.

- 1.b. Tu sais dans l'histoire, le quartier... et c'est là qu'il est devenu le nom que tous les gens d'Hochelaga-Maisonneuve rebutent le plus. C'est HoMa. Qui est les deux premières lettres de Hochelaga et Maisonneuve. Mais ça vient SoHo. Ben c'est parce que c'est une tradition américaine. C'est que SoHo c'est un quartier comme Hochelaga-Maisonneuve qui après la gentrification est devenu un quartier où les gens mieux nantis plus cultivés allaient s'installer. Mais nous on s'est toujours appelés Hochelaga-Maisonneuve, tu sais. À la rigueur j'aimerais mieux un mot aussi laid que Hochelag que HoMa. Paul, FRAN-HOMA 2012.

Dans son intervention, Paul qualifie HoMa du « nom que tous les gens rebutent le plus », c'est-à-dire le nom dont les résidents de HM se dissocient à l'unanimité. Il souligne que le symbole d'HoMa représente la situation financière (« mieux nantis ») et socio-intellectuelle (« plus cultivés ») des nouveaux résidents. Le pronom « nous » lui permet de séparer HoMa d'Hochelaga-Maisonneuve, le quartier réel dans lequel il vit. Il ajoute préférer le nom Hochelag, qu'il dit détester, si cela peut le mettre à une distance suffisante du HoMa créé de toutes pièces par les promoteurs.

Robert : d'Homa à Hochelag

C'est de cet espace imaginé que parle Robert, un homme sans-emploi de 51 ans, né dans l'ouest de Montréal, mais qui vit dans HM depuis ses 22 ans. Dans son entretien, il parle de la prostitution et de la drogue dans HM et montre combien son quartier n'est pas l'illusion que souhaitent vendre les promoteurs :

- 1.c. Tu t'en vas au Parc Morgan là le soir là quand tous les jeunes tu-sais que les parents surveillent pas. Pis même il y en a certains de leurs parents qui sont là pis sont sur le speed ou whatever. Il y a une pauvreté extrême tu-sais. Ces jeunes-là, ils côtoient les filles qui font la rue les clients tu-sais qui les ramassent tout dans un tas là. Ça fait ça fait pitié. Ça fait dur. Là tu t'aperçois que t'es encore dans Hochelag là tu-sais. T'es pas dans HoMa. Robert, FRAN-HOMA 2012.

Robert utilise Hochelag sous sa forme apocopée pour illustrer la pauvreté du quartier et les problèmes qui en découlent. En comparant «t'es encore dans Hochelag» et «t'es pas dans HoMa», Robert souligne la déception liée à la promesse d'un ailleurs fictif soutenue par la gentrification et nommée HoMa.

Si les locuteurs de HM associent des valeurs différentes à Hochelaga-Maisonneuve et à HoMa pour désigner leur quartier, qu'en est-il des médias? *Le Devoir*, *La Presse* et *Le Journal de Montréal* ont-ils la même utilisation des différents toponymes pour désigner HM? Ou utilisent-ils plus les dénominations neutres?

Les résultats dans les trois médias à l'étude

Description quantitative

Le tableau 5.3 présente l'utilisation des toponymes dans les trois quotidiens et le total des utilisations des toponymes tous médias confondus. Les deux toponymes les plus utilisés, tous médias confondus, sont Hochelaga-Maisonneuve (58,9 %), puis Hochelaga (22,44 %) et Mercier-Hochelaga-Maisonneuve (15,6 %). Les trois médias présentent la même répartition de ces trois toponymes.

TABLEAU 5.3
Nombre d'occurrences des toponymes par média entre 2010 et 2012

Toponyme	<i>La Presse</i>	%	<i>Le Devoir</i>	%	<i>Le Journal de Montréal</i>	%	TOTAL par toponyme	%
Hochelaga-Maisonneuve	383	62 %	278	60,7 %	228	51,6 %	889	58,6 %
Hochelaga	111	18 %	86	18,8 %	143	32,4 %	340	22,4 %
Mercier-Hochelaga-Maisonneuve	110	17,8 %	68	14,9 %	57	12,9 %	235	15,5 %
HoMa	13	2,1 %	23	5 %	12	2,7 %	48	3,2 %
Hochelag	0	0,0 %	2	0,4 %	0	0,0 %	2	0,1 %
Mercier Hochelaga	0	0,0 %	1	0,2 %	2	0,4 %	3	0,2 %
TOTAL par média	617		458		442		1 517	

La fréquence d'utilisation des toponymes est différente dans les trois médias ($\chi^2=15,5$; $p=0,003$).

Le Devoir utilise près de deux fois plus HoMa que les autres journaux (23 occurrences vs 13 dans *La Presse* et 12 dans *Le Journal de Montréal*). En outre, *Le Devoir* se distingue également par son recours au toponyme Hochelag qu'il est le seul à utiliser entre 2010 et 2012 (2 occurrences).

En comparant la fréquence des différents toponymes dans les articles des médias et dans les discours des locuteurs du sous-corpus FRAN-HOMA 2012, nous constatons qu'Hochelaga et Hochelaga-Maisonneuve sont les plus utilisés. Cependant, nous remarquons une inversion entre les deux corpus : les habitants de HM emploient plus Hochelaga (86,6 % vs 22,4 % dans les médias) qu'Hochelaga-Maisonneuve (71,1 % vs 58,6 % dans les médias). HoMa a aussi un poids plus important chez les locuteurs (13,2 % vs 3,16 % dans les médias) : même sans y adhérer, les résidents montrent que ce toponyme occupe l'espace discursif. En outre, l'utilisation d'Hochelag est plus largement répandue chez les locuteurs que dans les médias (26,3 % vs 0,2 %).

Le toponyme HoMa étant fortement connoté à la faveur du changement de paysage urbain dans le quartier, un nom associé à la gentrification, nous avons réalisé une régression logistique pour répondre à deux questions : 1) l'année a-t-elle un impact sur l'utilisation du toponyme HoMa ? et 2) existe-t-il une différence dans l'utilisation d'HoMa parmi les différents journaux ? Le tableau 5.4 présente notre modèle dans lequel l'année 2010 et le journal *La Presse* sont les références.

Tableau 5.4

Paramètres du modèle de régression logistique sur le toponyme HoMa

	Rapport de cotes	Valeur de p
Année 2011	1,06	0,9
Année 2012	2,81	0,006
<i>Le Devoir</i>	2,35	0,016
<i>Le Journal de Montréal</i>	1,15	0,7

Pour la question 1, on remarque une augmentation non significative de la probabilité d'utiliser le toponyme HoMa en 2011 par rapport à 2010 ; tandis qu'en 2012, la probabilité d'utiliser le toponyme HoMa est 3 fois plus importante qu'en 2010 ($p=0,006$).

Pour la question 2, la probabilité d'utiliser le toponyme HoMa dans *Le Devoir* est 2 fois plus élevée par rapport à *La Presse*, à année égale ($p=0,016$). Dans *Le Journal de Montréal*, la probabilité d'utiliser le toponyme HoMa diminue de 32 % par rapport à *La Presse*, mais cette diminution n'est pas significative ($p>0,05$). Pourquoi observe-t-on une augmentation de l'utilisation du nom HoMa en 2012 ? Pourquoi *Le Devoir* est-il plus enclin à utiliser le toponyme HoMa que les autres journaux ? Et pourquoi *Le Devoir* est-il le seul journal à utiliser le nom Hochelag au cours des trois années ?

L'analyse qualitative nous permettra de répondre à ces questions et de fournir une compréhension plus approfondie des facteurs influençant l'utilisation des noms de lieux dans les journaux étudiés.

Analyse qualitative

Dans un article publié dans *Le Devoir* en juin 2012, Amélie Gaudreau annonce les Spectacles de la Fête nationale du Québec à Montréal dans le Parc de Maisonneuve et souligne la division profonde au sein du quartier Hochelaga-Maisonneuve :

- 2.a. Avant d'être le semi-branché « HoMa » et le défavorisé « Hochelag », ce coin de Montréal fut d'abord la ville de Maisonneuve. *Le Devoir*, juin 2012, Amélie Gaudreau.

Ces deux noms représentent deux espaces distincts, avec des connotations différentes. Hochelag est perçu comme un quartier défavorisé associé à la criminalité, la prostitution, la drogue et la précarité, en raison de sa représentation dans la littérature et les films (*Je voudrais qu'on m'efface* de Barbeau-Lavalette [2010]). De son côté, HoMa est considéré comme un quartier en expansion, avec l'ambition de devenir le nouveau Plateau ou le SoHo québécois. La présence de guillemets autour des deux toponymes indique leur nature récente et leur fabrication respectivement artificielle et populaire. La journaliste met ainsi de l'avant une division imaginaire de l'espace. Cette dichotomie linguistique, basée sur une simple toponymie, renvoie à des perceptions divergentes d'un même espace, comme en témoignent les deux espaces imaginaires créés dans le quartier est de Londres (Ilbury, 2022). Elle reflète une perception exogène, telle que la désignation de l'espace en tant que « semi-branché », et une perception endogène, telle que l'identification de l'espace comme étant « défavorisé ». Cette fracture médiatique trouve un écho direct dans les discours des locuteurs : Gabrielle (1.a) et Paul (1.b) opposent explicitement HoMa, perçu comme un artifice de promotion immobilière, à

Hochelag, symbole d'authenticité populaire. Cette convergence confirme que les toponymes cristallisent des rapports de pouvoir sociospatiaux, bien au-delà de leur fonction de simple désignation géographique. Cette fracture médiatique corrobore les perceptions des locuteurs.

De plus, dans nombre d'occurrences, *Le Devoir* utilise HoMa pour l'agenda culturel. Trois exemples peuvent être cités à titre d'illustration :

- 2.b. Plus loin dans HoMa, le plus récent opus de Sébastien Harrisson, *Musique pour Rainer Maria Rilke*, est toujours à l'affiche jusqu'à la fin de semaine. *Le Devoir*, 2012, Michel Bélair.
- 2.c. En poussant encore plus loin dans les deux directions de la ligne verte du métro, les Carrefour Angrignon (à LaSalle) et StarCité (dans HoMa) sont également sous la gouvernance de Cineplex, qui règne, là aussi, sans partage. *Le Devoir*, 2012, Martin Bilodeau.
- 2.d. À tout cela, il faut rajouter quelques nouvelles idées séduisantes, comme l'événement Lire ensemble, un défi lecture lancé aux habitants du quartier HoMa durant toute la durée du festival et une séance de lectures sur le sport organisée dans une taverne du coin à l'intention des papas. *Le Devoir*, 2011, Michel Bélair.

Musique, cinéma et défi lecture viennent alimenter la description d'HoMa. S'il ne semble pas de prime abord s'agir d'un quartier semi-branché, l'évocation de la culture dans un quartier historiquement ouvrier et industriel revient à le classer du côté des quartiers semi-branchés. Les toponymes Hochelag, Hochelaga et Hochelaga-Maisonneuve serviraient-ils dans *Le Devoir* à évoquer la dure réalité de HM, et HoMa, son versant culturel ?

Du côté du *Journal de Montréal*, on voit un tout autre visage d'HoMa. Deux exemples montrent que, quel que soit son nom, HM conserve son taux de dangerosité :

- 3.a. Notre reportage dans le « Red-light » d'HoMa nous a permis de constater que la prostitution comporte son lot de dangers, à la fois pour les prostituées, les clients et les policiers. *Le Journal de Montréal*, 2012, Daniel Renaud.
- 3.b. Tentative de meurtre dans HoMa. *Le Journal de Montréal*, 2012, Élisabeth Laplante.

Dans l'extrait 3.a., le quartier HoMa équivaut au quartier Hochelaga ou Hochelaga-Maisonneuve et recouvre la même réalité que le nom ne s'entête pas à dissimuler : celle de la drogue et de la prostitution. Élisabeth Laplante va même jusqu'à titrer, en 3.b., « Tentative de meurtre dans HoMa », ce qui contraste avec le « semi-branché HoMa » dont parlait *Le Devoir*.

Afin de modaliser notre propos, Christian Lavigne met en avant la distinction entre HoMa et Hochelaga-Maisonneuve dans cet extrait :

- 3.c. Je me rendais au lancement d'un intéressant projet de condos dans HoMa, le nom « gentrifié » du quartier Hochelaga-Maisonneuve. *Le Journal de Montréal*, 2012, Christian Lavigne.

Sous la plume du journaliste, les deux espaces, HoMa et Hochelaga-Maisonneuve, font référence de manière explicite au même quartier, puisqu'il précise qu'il s'agit du « nom gentrifié », mais on voit le début d'une dichotomie franche, semblable à celle opérée par Amélie Gaudreau, dans *Le Devoir*. Le projet de condos s'attache à décrire la réalité gentrifiée, celle de HoMa qui, à l'instar de SoHo, s'intéresse au côté huppé.

La comparaison entre les deux types de quartier est prégnante en 2012, mais il sera bon de s'interroger sur sa pérennité pour les années à venir. Le pont entre HoMa et Hochelaga-Maisonneuve est très présent dans *La Presse*, comme en témoignent les deux exemples suivants :

- 4.a. HoMa, donc, se transforme doucement. D'anciennes manufactures sont converties en appartements et en espaces de bureaux par des architectes allumés, des jeunes créatifs y déménagent, ne pouvant trouver ailleurs en ville des loyers aussi abordables. *La Presse*, 2012, Marie-Claude Lortie.
- 4.b. Hochelaga-Maisonneuve, ou HoMa si vous préférez, se fait dire depuis longtemps qu'il est le nouveau Plateau, le nouveau quartier encore abordable en devenir où il faut investir... *La Presse*, 2012, Marie-Claude Lortie.

Dans ces deux citations, Marie-Claude Lortie établit un parallèle entre les deux dénominations de ce même lieu et montre que l'une, HoMa, représente le côté fortuné, tandis que l'autre, Hochelaga-Maisonneuve, se contente de se rapporter à la réalité géographique, en ce qu'elle peut revêtir de péjoratif. C'est Patrick Lagacé qui, dans *La Presse*, propose une critique acerbe de ces deux noms, ayant pour but de créer deux espaces là où il n'y en a qu'un, et un seul, dans toute sa dure réalité :

- 4.c. Pis, un jour, quand tu côtoies la pauvreté, c'est parce que tu viens d'acheter un triplex dans Hochelaga-Maisonneuve. Oups, HoMa. Là, tu les vois, les pauvres. Pas sûr que tu trippes, non plus... *La Presse*, 2012, Patrick Lagacé.

L'interjection « oups » montre le sarcasme du chroniqueur qui écrit de manière ostentatoire que le seul objectif d'HoMa est de faire vendre un produit qui ne sera jamais que le même produit, dans sa réalité brute, celle d'Hochelaga-Maisonneuve avec ses « pauvres » qui ne s'inscrivent pas dans le *rêve américain* associé à HoMa.

D'après une évaluation des articles des trois quotidiens de 2010 à 2012, HoMa est utilisé plus fréquemment par *Le Devoir* que par les deux autres journaux. Ce choix terminologique s'accompagne également de l'utilisation du toponyme Hochelag que *Le Devoir* est le seul à employer. Nous nous sommes alors demandé si le toponyme HoMa pouvait évoquer le côté culturel du quartier. Pour en avoir le cœur net, nous avons placé tous les paragraphes contenant le toponyme HoMa dans des fichiers textes classés par quotidien et avons soumis ce sous-corpus à Hyperbase en ligne (*Hyperbase*, s. d.) afin de regarder quels étaient les termes du cotexte de droite et de gauche. Dans *La Presse* et *Le Journal de Montréal*, le toponyme HoMa apparaît avec des termes fonctionnels tels que « quartier », « condos », « adresse » ou « restaurant ». Parallèlement, dans *Le Devoir*, ce sont plutôt les termes liés au contexte artistique qui sont en vedette, avec « culture », « lecture », « artistique » ou « événement ». Si la présente analyse s'est concentrée sur HoMa, une étude qualitative montre que Hochelag est fréquemment associé à la précarité dans les médias, ce qui renforce la dichotomie avec HoMa. Cette asymétrie médiatique (HoMa culturel vs Hochelag rare ou associé à la précarité) reflète la fracture observée chez les locuteurs, où Hochelag incarne une résistance identitaire face à l'appropriation symbolique par HoMa.

Au terme de cette partie, nous constatons que l'augmentation de l'utilisation du nom HoMa en 2012 est due à une recrudescence du nombre de constructions et de ventes de condos dans HM. Les deux autres questions posées n'obtiennent cependant pas de réponse à la suite de notre analyse qualitative. Nous présenterons des éléments de réponse dans la discussion.

Les trois grands médias ne se comportent donc pas de la même manière. *Le Devoir* sépare HoMa des autres toponymes et l'utilise plutôt pour évoquer le caractère culturel du quartier. *Le Journal de Montréal* ne met pas de l'avant l'aspect bourgeois d'HoMa et se cantonne à sa réalité géographique interchangeable avec Hochelaga ou Hochelaga-Maisonneuve. Chez les journalistes du *Journal de Montréal*, le nom HoMa semble plutôt être un synonyme capable d'éviter une répétition. En revanche, le point de vue de Patrick Lagacé dans *La Presse* montre qu'une certaine partie de la population montréalaise a conscience d'une forme de manipulation : HoMa n'est qu'un maquillage sur le visage brut d'Hochelaga-Maisonneuve.

Toutefois, les trois médias s'entendent sur un point : HM est un quartier pauvre qui ne suscite pas un engouement de résidence de la part des habitants des autres quartiers montréalais, à part peut-être lorsqu'il s'agit d'investir dans un condo dans HoMa.

Discussion

Les études qui cherchent à décrire et comprendre le phénomène de la gentrification ne prennent pas en considération l'aspect toponymique du quartier. Qu'il s'agisse d'HoMa à Montréal, de SoHo à New-York, de SoMa à San Francisco (Phillips, 2012), ou de LoDo à Denver (Prentiss, 2021), le fait de renommer un quartier en contractant son nom pour le rendre plus attrayant n'est pas un geste innocent. Il serait intéressant d'étudier ce phénomène sous un angle plus sociopsychologique afin de comprendre les enjeux identitaires que ces transformations sous-tendent.

À la suite des grands changements du paysage urbain survenus au début des années 2000, les médias, comme les locuteurs résidant dans le quartier, témoignent de cette modification sociospatiale. Dans ce contexte, l'utilisation croissante du toponyme HoMa, en écho à la diminution du toponyme Hochelaga-Maisonneuve, entre 2010 et 2012, semble refléter cette transformation sociale et urbaine.

Après avoir montré que le toponyme HoMa est davantage utilisé par *Le Devoir* que par les deux autres journaux, et qu'il se trouve plus souvent dans un contexte culturel qu'Hochelag, on pourrait se demander si le faible nombre de mots par article pourrait favoriser l'utilisation du toponyme HoMa en raison d'un contenu annonçant des événements à venir ou résumant des événements passés.

De son côté, le toponyme Hochelag, dans sa forme apocopée, est utilisé pour évoquer le quartier vu sous son angle populaire, souvent associé à la « culture trash³ ». Cette utilisation différenciée des termes pour évoquer un même quartier montre l'importance de la toponymie dans la construction de l'identité et de l'image des lieux. Le fait que *Le Devoir* mène ce mouvement renvoie probablement à la ligne éditoriale du quotidien ainsi qu'à son lectorat, constitué majoritairement d'intellectuels issus de catégories sociales aisées. Dans sa « Politique d'information », *Le Devoir* affirme « s'intéresser à tous les débats qui ont un impact significatif sur le développement de la société québécoise » (2005). Le quotidien se doit donc de broser un portrait complet du quartier en utilisant toute une palette identitaire.

Ces différences d'utilisation des termes influencent nécessairement la perception des quartiers par les différents acteurs sociaux. Elles reflètent les

3. Voir pour une définition de la « trash culture » le volume 28, n° 3 de la revue *French cultural studies*, 2017 : les habitudes inhérentes aux milieux défavorisés sont mises de l'avant et leur hégémonie vise à remplacer celle des classes dominantes.

enjeux identitaires, sociaux et culturels qui y sont associés. Cette étude apporte un nouvel éclairage sur le rapport entre la toponymie et la gentrification, dans un corpus qui, jusqu'alors, avait toujours été étudié pour sa forme linguistique et non pour son contenu. Pourtant, les locuteurs racontent leurs trajectoires à Montréal, leur vie dans le quartier, et partagent leur perception de leur identité de quartier.

Références

- BARBEAU-LAVALLETTE, A. (2010), *Je voudrais qu'on m'efface*, Hurtubise.
- BLONDEAU, H. et M. TREMBLAY (2012), Social mixing in HOMA: Young urban francophones and language variation, communication présentée au colloque *Socio-linguistics Symposium*, Freie Universität Berlin, 21-24 août.
- BLONDEAU, H., M. TREMBLAY, A. BERTRAND et E. MICHEL (2021), A new milestone for the study of variation in Montréal French: The Hochelaga-Maisonneuve socio-linguistic survey, *Corpus*, 22.
- BOYER, H. (2008), Fonctionnements sociolinguistiques de la dénomination toponymique, *Mots. Les langages du politique*, 86(1), 9-21.
- CHRISTIAN, G. (1999), Les processus redénotatifs dans les noms de communes françaises depuis 1943. Étude socio-toponymique de la variation dans la nomenclature administrative, *Noms et re-noms. La dénomination des personnes, des populations, des langues et des territoires*, dans S. AKIN (dir.), *Noms et re-noms. La dénomination des personnes, des populations, des langues et des territoires*, Rouen, Publications de l'Université de Rouen, 209-227.
- DESLAURIERS, J.-P. et M. KÉRISIT (1997), Le devis de recherche qualitative, dans J. POUPART, L. GROULX, A. LAPERRIÈRE, R. MAYER et A. PIRÈS (dir.), *La recherche qualitative: enjeux épistémologiques et méthodologiques*, Gaëtan Morin, 85-111.
- DUGAS, J.-Y. (1987), La toponymie québécoise comme mode d'appropriation symbolique du territoire, *Proceedings of the Meeting of the French Colonial Historical Society*, 11, 237-250.
- GERMAIN, A. et D. ROSE (2010), La mixité sociale programmée en milieu résidentiel à l'épreuve des discours critiques internationaux: le cas de Hochelaga à Montréal, *Lien social et Politiques*, 63, 15-26.
- GHAFFARI, L. (2020), *Pour une gentrification socialement acceptable, le cas d'Hochelaga-Maisonneuve à Montréal et Madeleine-Champ-de-Mars à Nantes*, thèse de doctorat, Université du Québec à Montréal.
- GLASER, B. G. et A. L. STRAUSS (1999), *The discovery of grounded theory: Strategies for qualitative research*, New Brunswick, Aldine publishing company.
- Hyperbase. (s. d.), Analyse de données textuelles en ligne, <http://hyperbase.unice.fr/>.
- ILBURY, C. (2022), A tale of two cities: The discursive construction of «place» in gentrifying East London, *Language in Society*, 51(3), 511-534. <https://doi.org/10.1017/S0047404521000130>.

- La Presse+* (18 octobre 2013), *Mercier-Hochelaga-Maisonnette*, <https://plus.lapresse.ca/screens/4c80-e6e8-52605393-9e40-151cac1c606d|3~DoFZA.A9YN.html>.
- Le Devoir (25 mai 2005), Politique d'information du Devoir (1^e partie), *Le Devoir*, <https://www.ledevoir.com/culture/medias/82552/politique-d-information-du-devoir-1re-partie>.
- LEES, L., et D. LEY (2008), Introduction to special issue on gentrification and public policy, *Urban Studies*, 45(12), 2379-2384.
- LINCOLN, Y. S., E. G. Guba et S. A. Lynham (2018), Paradigmatic controversies, contradictions, and emerging confluences, revisited, dans N. K. Denzin et Y. S. Lincoln (dir.), *The Sage handbook of qualitative research* (5^e éd.), Sage, 108-150.
- LINTEAU, P.-A. (1981), *Maisonnette ou comment des promoteurs fabriquent une ville : 1883-1918*, Boréal express.
- MARTINEAU, F. M.-C. et M.C. SÉGUIN (2016), Le Corpus FRAN : réseaux et maillages en Amérique française, *Corpus*, 15.
- MINER, Q. D. A. (4), Lite (2020), *QDA Miner Lite free qualitative data analysis software*.
- NASTASE, I. A. (2021), *La gentrification du quartier Hochelaga-Maisonnette : le discours sur la mixité à l'épreuve de ses réalités socio-spatiales*, Université de Montréal.
- OLIVIER DE SARDAN, J.-P. (2008), *La rigueur du qualitatif : les contraintes empiriques de l'interprétation socio-anthropologique*, Academia.
- PARÉ, S. et S. MOUNIER (2021), Perceptions of Hochelaga-Maisonnette Neighborhood in Montreal: A Textual Analysis of Written Medias, *Culture and Local Governance*, 7(12), 17-39, <https://doi.org/10.18192/clg-cgl.v7i1-2.4848>.
- PHILLIPS, L. K. (2012), *Revitalized Streets of San Francisco : A Study of Redevelopment and Gentrification in SoMa and the Mission*, Clairmont College.
- POUPART, J., J.-P. DESLAURIERS, L. GROULX, A. LAPERRIÈRE, R. MAYER et A. PIRÈS (1997), Échantillonnage et recherche qualitative : essai théorique et méthodologique, dans *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Gaetan Morin, 113-169.
- PRENTISS, M. A. (2021), *New-Build Gentrification in the American Urban Core : The Case of Denver's Lower Downtown Neighborhood*, mémoire de maîtrise, University of Colorado at Denver.

CHAPITRE 6

Le corpus « Traits d’union »

Perceptions de la langue française
dans le travail d’intervenants
auprès de citoyens issus de l’immigration

Mireille Elchacar, Angéline Martel et Nancy Gagné

Selon le dernier bilan démographique du Québec, le nombre d’immigrantes et d’immigrants accueillis durant les années précédant la pandémie de COVID-19 a atteint des niveaux record. Au premier semestre de 2022, « les gains migratoires reviennent aux niveaux record observés avant la pandémie, et les dépassent même légèrement » (Institut de la statistique du Québec, 2022). La majorité de ces immigrants s’établit à Montréal et dans les environs. Selon les dernières données du ministère de l’Immigration, de la Francisation et de l’Intégration, sur les 509 270 personnes immigrantes admises au Québec entre 2010 et 2019, 69,6 % résident dans la région métropolitaine ; Laval et Longueuil sont également des secteurs d’accueil importants (ministère de l’Immigration, de la Francisation et de l’Intégration, 2021).

C’est sur la toile de fond de la ville de Montréal en tant que société d’accueil de la majorité des immigrants de la province que l’Institut Jacques-Couture [IJC] a été mis sur pied. Affilié à l’Université TÉLUQ, l’IJC a pour principale mission de soutenir des innovations en matière de relation accueillants-accueillis, et d’offrir des activités d’enseignement, de recherche et de services à la collectivité en rapport avec les thématiques de l’accueil, des échanges et des sociétés en contact¹. Ce texte rend compte d’une

1. L’IJC offre par exemple toute une série de formations courtes destinées autant aux personnes immigrantes qu’aux personnes et collectivités qui travaillent auprès d’elles (« Le

recherche découlant du projet *Traits d'union* : compétences interculturelles en action, une initiative conjointe de l'IJC et du Groupe d'expertise pour le développement des cités interculturelles au Québec, un organisme à but non lucratif qui a pour mission de renforcer les compétences interculturelles. Ce projet de recherche-action vise à répertorier les besoins en compétences et communication interculturelles des intervenants de divers services publics de Montréal et des environs, tels qu'ils les ont répertoriés lors des entretiens. Dans la foulée du dépôt du rapport final du projet, en 2020, nous avons souhaité étudier plus en détail les perceptions qu'ont de la langue quelques intervenants du secteur public qui travaillent avec des nouveaux arrivants (Gagné *et al.*, 2020).

L'accueil d'immigrants amène des besoins en matière d'intégration, de communication et de francisation. Les intervenants du secteur public qui œuvrent auprès des populations immigrantes doivent être formés en matière de compétences interculturelles (Potvin *et al.*, 2018; Pouliot, Gagnon et Pelchat, 2015; UNESCO, 2015). De faibles compétences interculturelles peuvent mener à des interventions moins réussies, voire à des malentendus, entre autres liés à des incompréhensions de nature linguistique (Bleton, Martel et Gagné, 2022).

Le projet *Traits d'union* n'avait pas au départ une visée strictement linguistique. Néanmoins, durant les entretiens, les participants ont clairement exprimé « ne pas disposer d'outils assignés à la résolution de problèmes liés à la langue de communication et doivent souvent improviser pour les résoudre (Gagné *et al.*, 2020, p. 35). » Nous avons voulu creuser les enjeux de nature linguistique auxquels sont confrontés les intervenants qui travaillent avec des citoyens issus de l'immigration.

Recension des écrits : la perception du français chez les nouveaux arrivants

De nombreuses études se sont intéressées à la perception et aux représentations du français des nouveaux arrivants, et aux difficultés de nature linguistique auxquelles ils sont confrontés.

L'étude d'Amireault et Lussier (2008) a fait ressortir la représentation du français en tant que langue noble et romantique. Les répondants de cette

Canada en cinq dates», « Montréal, ville de carrefours et de rencontres », « Entreprendre au Québec », « Communication et compétences interculturelles », etc.).

étude disent comprendre que le français soit une langue protégée au Québec, en regard de son histoire et de sa situation sociolinguistique. Le français est perçu comme étant la langue de l'avenir et un outil essentiel d'intégration, puisque c'est la langue du marché du travail ou de l'école (que ce soit pour un projet de retour aux études ou pour la scolarisation des enfants) (Amireault et Lussier, 2008; Ataepour, 2016).

On souligne la barrière de la langue : il est difficile d'acquérir une aisance suffisante du français avec les cours de francisation, de se sentir à l'aise dans des conversations quotidiennes (Beaulieu *et al.*, 2021; Amireault et Lussier, 2008). Calinon (2015) note que les situations où immigrants en apprentissage du français et francophones se rencontrent sont rares, ce qui nuit à l'acquisition de la langue. Dans plusieurs études, on constate une différence entre le français appris dans les cours de francisation et le français dit « de la rue » (Calinon, 2015, Harvey, 2016; Arbour, 2012; Ataepour, 2016). Un lien peut être fait ici avec les études portant sur la prise en compte de la variation dans l'enseignement du français langue seconde (FLS) (Eggington, 2004; Gadet, 2004; Roy, 2015; Thibault *et al.*, 2015). Les manuels utilisés pour enseigner le FLS ne prennent pas tous comme référence le français de la société d'accueil, soit, ici, le français québécois, et les éléments de variation, géographique ou diaphasique, ne sont pas toujours bien contextualisés; la prise en compte de ces éléments dépend alors de la connaissance et de l'expérience de chaque enseignant de FLS (Elchacar, Steffens et Baiwir; 2021). Par ailleurs, malgré le virage amorcé depuis le dernier Cadre européen commun de référence (CECR) pour les langues, en 2001, qui prône l'utilisation en classe de documents authentiques (plutôt que des classiques de la littérature, par exemple), le choix et la qualité des documents effectivement fournis aux apprenants sont inégaux (Giroud et Surcouf, 2016). En outre, advenant un bon choix de textes authentiques, l'écart entre l'oral familier et l'écrit soigné demeure, et l'apprenant peut sentir un décalage entre le français auquel il est exposé en classe et les conversations courantes (Amireault et Lussier, 2008).

Quelques études soulignent les difficultés que représente le français québécois en raison de particularismes au niveau de la prononciation ou du lexique (Harvey, 2016; Benzakour, 2004). Benzakour explique que le français québécois n'est pas la variété géographique de français à laquelle les immigrants avaient été exposés avant leur arrivée au Québec, particulièrement chez les personnes originaires du Maghreb. Par ailleurs, les nouveaux arrivants identifient parfois des traits linguistiques familiers ou populaires comme étant propres au français québécois, même s'ils sont présents dans

d'autres variétés géographiques de français, comme la chute du /e/ caduc [schwa] ou l'assimilation consonantique (Harvey, 2016). Ceci, couplé au fait que le français « de la rue » diffère du français enseigné en classe de FLS, rappelle une certaine vision du français québécois qui le résume au registre familier ou à une langue populaire (Villeneuve, 2017; Guertin, 2017). Un amalgame se fait entre le français familier et le français québécois, sans égard à la déclinaison de ce dernier en divers registres.

Certaines recherches abordent la représentation du français québécois par rapport au français de France (ou européen), et concluent que les immigrants perçoivent ce dernier de manière plus positive que le français québécois (Kircher, 2012; Maurais, 2008; Benzakour, 2004, Remysen et Rheault, 2023 pour une synthèse).

Enfin, les nouveaux arrivants disent que connaître le français est important, mais insuffisant pour s'intégrer, particulièrement sur le marché du travail (Benzakour, 2004). C'est également la conclusion à laquelle parviennent Godin et Renaud (2005, 167) : « La clé du succès d'une intégration professionnelle résiderait davantage dans l'apprentissage des rouages de la culture québécoise (us et coutumes, modes de vie de la population québécoise, façons de faire pour s'intégrer au marché du travail, etc.) ».

Certaines études relèvent les perceptions de l'anglais qu'ont les nouveaux arrivants du Québec (Kircher, 2014). Ces derniers notent une dichotomie entre, d'une part, l'importance accordée au français par rapport à l'anglais dans le discours public et, d'autre part, la réelle présence constatée de l'anglais sur le terrain (Amireault et Lussier, 2008). Les nouveaux arrivants perçoivent l'anglais comme étant plus « facile »; ceci pourrait s'expliquer par le fait qu'ils ont été plus exposés à l'anglais en tant que langue véhiculaire mondiale avant leur arrivée au Québec. En attendant de pouvoir suivre les cours de francisation, ce qui peut prendre plusieurs mois², puis d'acquérir une certaine aisance en français, il peut être tentant de se trouver un emploi en anglais.

En revanche, à notre connaissance, aucune étude ne s'est penchée sur la perception du français qu'ont les intervenants œuvrant auprès des nouveaux arrivants. Est-ce qu'ils ont conscience des difficultés linguistiques vécues par les nouveaux arrivants? Est-ce que ces éléments expliquent cer-

2. Ces délais peuvent s'expliquer par la pénurie d'enseignants et par la pandémie de COVID-19 (<https://www.ledevoir.com/societe/597864/une-trentaine-de-cours-de-francisation-reportes>; <https://www.journaldequebec.com/2022/11/21/longue-attente-pour-la-francisation>).

tains défis que vivent les intervenants? Ce sont les questions auxquelles nous tenterons de répondre en interrogeant le corpus Traits d'union.

Le corpus Traits d'union

La méthode choisie pour déterminer les besoins en matière de formation interculturelle des intervenants est le groupe de discussion. L'équipe de recherche en a créé 11, qui rassemblent 85 personnes (39 femmes et 46 hommes) travaillant en proximité avec les collectivités ethnoculturelles de Montréal, Laval et Longueuil. L'âge moyen des participants est de 43 ans, et ils ont en moyenne 12,5 années d'expérience dans leur poste actuel.

Cinq secteurs d'intervention ont été retenus : les bibliothèques municipales (BM), les services d'incendie (SI), de police (SP) et d'ambulance (Urgence Santé, US), ainsi que Service Québec (SQ), auquel sont intégrés les centres locaux d'emploi. Il s'agit de secteurs où des intervenants du secteur public sont en contacts directs et fréquents avec les populations immigrantes, dans des régions où ces populations sont en plus forte proportion.

Les rencontres semi-dirigées, d'une durée d'environ quatre heures chacune, se sont déroulées à partir d'un canevas de questions touchant cinq thématiques : interprétation du concept de culture, description du contexte d'intervention, savoirs en regard des cultures particulières, compétences interculturelles et besoins de formation. Les discussions n'étaient pas dirigées directement sur la langue. Or, la thématique de la langue a émergé spontanément comme étant fondamentale dans chaque groupe de discussion. Nous avons donc repris le corpus Traits d'union afin d'en faire une utilisation secondaire dans une perspective linguistique.

Nous avons étiqueté la transcription des entretiens dans NVivo afin de procéder à une analyse qualitative. La transcription a été faite de manière large et ne conviendrait pas nécessairement si le matériau de l'étude était la langue parlée par les intervenants. Notre intérêt est plutôt ce que les intervenants disent à propos de la langue, informations que le corpus Traits d'union permet de faire émerger³.

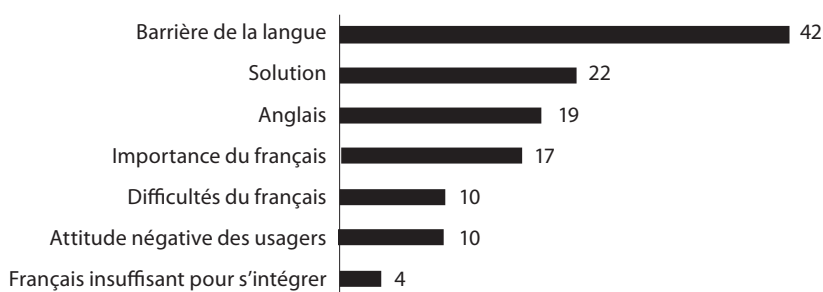
3. Le projet a obtenu un certificat d'éthique pour la recherche auprès de sujets humains et les données, anonymisées, ont été utilisées dans l'analyse des propos recueillis.

Analyses : perceptions du français chez les intervenants

Nous avons relevé 124 segments concernant le français et l'anglais dans le corpus. Un segment est un passage des entretiens où il est question de langue. La longueur d'un segment varie d'une à plusieurs phrases. Nous avons classé les segments sous sept thématiques générales, présentées à la figure 6.1. Un même segment peut être codé sous plus d'une thématique.

FIGURE 6.1

Les thématiques linguistiques dans le corpus Traits d'union

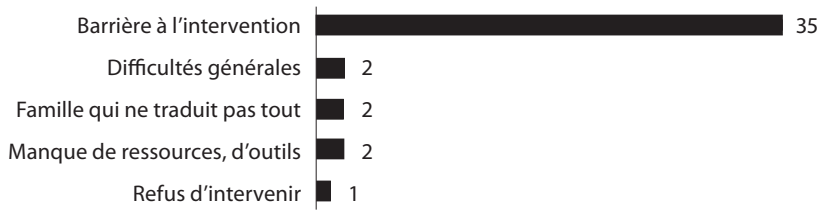


La catégorie « Barrière de la langue » fait référence aux difficultés rencontrées par les intervenants en raison de l'absence d'une langue commune entre eux et les immigrants. La catégorie « Solution » concerne les outils mis en place par les intervenants, ou ceux dont ils souhaitent l'implantation, pour surmonter la barrière de la langue. Nous avons regroupé sous une catégorie tous les segments concernant l'anglais. La catégorie « Importance du français » concerne autant la place du français dans l'espace public québécois que son rôle dans l'intégration des immigrants. La catégorie « Difficultés du français » concerne plutôt les difficultés que les intervenants perçoivent chez les usagers.

La barrière de la langue

La thématique qui revient de loin le plus souvent, soit dans plus du tiers des segments, et qu'on retrouve dans les cinq secteurs d'intervention, est la barrière de la langue. Il s'agit de problèmes d'ordre linguistique qui représentent une difficulté pour mener une intervention. La figure 6.2 présente les différentes sous-catégories de ces segments.

FIGURE 6.2

La barrière de la langue, sous-catégories

La barrière de la langue est la première difficulté nommée par les intervenants des centres locaux d'emploi (1), d'Urgences-Santé (2) et des services d'incendie (3). Ces intervenants utilisent souvent eux-mêmes le mot *barrière*:

- 1) « C'est ça, s'il y a une difficulté au niveau de la langue... souvent il y a une barrière. [...] C'est difficile. Si on ne se comprend pas, si au niveau de la langue il y a une difficulté, le reste de l'intervention ou quoi que ce soit va être difficile, parce qu'il y a cette barrière-là au départ. Moi je pense que c'est ça la plus grosse difficulté rencontrée ». (SQ2⁴)
- 2) « Dans votre travail, à quoi vous êtes confronté? – Problème de langue, le plus difficile, s'ils parlent pas français, le minimal. En tout cas pour moi c'est difficile. Je maîtrise pas d'autres langues [...] ». (US1)
- 3) « [L]e problème de communication, le problème de langue. La barrière linguistique est énorme. » (SI2)

Chez les policiers, on indique que la barrière de la langue est un enjeu qui empêche parfois de mener une intervention adéquate auprès des populations immigrantes, mais également que ces dernières ne font pas toujours appel à eux pour cette même raison :

- 4) « On a quand même une difficulté d'aller rejoindre certaines populations qui sont plus vulnérables. Ouais, c'est ça, ils appellent pas. Il y a des barrières de langage des fois qui sont vraiment importantes. Tu te ramasses en intervention, la personne ne parle ni français, ni anglais,

4. À des fins de respect de la confidentialité, nous gardons les informations liées aux intervenants au niveau minimal : les initiales renvoient au secteur d'intervention et le chiffre, à la numérotation qui résulte du processus d'anonymisation.

puis là, tu te ramasses à avoir besoin d'un interprète bengali pour intervenir. Il y a tout cet enjeu-là.» (SP2)

Un seul segment évoque un refus d'intervenir auprès de citoyens issus de l'immigration, et il fait référence à une époque passée. Ce segment vient d'un employé des centres locaux d'emploi :

- 5) « Moi je me souviens, à mes débuts, il y a des gens qui ont toujours travaillé dans l'Est, qui ne voulaient rien savoir de l'Ouest comme Côte-des-Neiges, Saint-Laurent. À l'époque c'était à ville Saint-Laurent, parce qu'il y a l'anglais, d'autres qui ne veulent pas parler à des immigrants, carrément. » (SQ3)

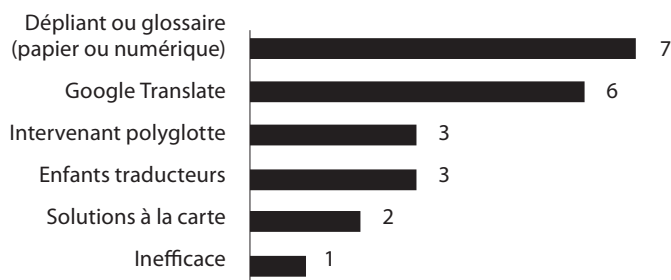
On perçoit néanmoins que le refus d'intervenir est également relié à d'autres facteurs, dont le fait qu'on ne veuille pas parler anglais ou travailler auprès d'une population immigrante.

Les solutions

La deuxième thématique abordée regroupe les solutions : celles que les intervenants utilisent déjà ou celles qu'ils voudraient voir mises en place pour les soutenir dans leur travail. Dans tous les secteurs, on déplore qu'aucun outil ne soit fourni d'emblée pour les cas où la barrière de la langue empêcherait les employés de mener leur intervention adéquatement.

FIGURE 6.3

Les solutions, sous-catégories



La solution la plus évoquée, et que les intervenants jugent aisée à implanter, est le recours à un simple glossaire ou dépliant avec quelques

mots clés ou phrases liés au secteur d'intervention, traduits en quelques langues :

- 6) « ... des mots clés : détresse respiratoire, douleurs de type cardiaque, as-tu des engourdissements ? As-tu des maux de tête ? Les mots clés ! » (US₂)

Dans l'extrait suivant, l'intervenant entrevoit un tel outil dans la double perspective de se faire comprendre dans l'immédiat, mais aussi de créer des ponts, pour établir une certaine confiance à plus long terme entre intervenants et usagers :

- 7) « Quand j'arrive admettons [...] dans le quartier chinois... eux autres, ils veulent pas évacuer. Bien là, tu prends le micro dans le bâtiment puis tu dis : "c'est les pompiers, évacuez". Évacuez en chinois. Ça évacue un peu plus, pas complètement... Ou tu dis : "tout est sous contrôle". [...] une dizaine de mots de même. Écoute, c'est drôle parce que, même mes pompiers restent surpris souvent. Écoute, je l'ai refait encore cette semaine. Je vous le dis, ça crée des ponts parce que déjà en partant "OK, il me dit *bonjour*". On pourrait faire un petit pense-bête. Faut pas l'imposer, mais on le laisse traîner. Tu le laisses traîner puis admettons tu fais comme des mots clés "je veux prendre votre pouls", "on est pompier, on est ici pour vous aider". Puis qu'on le met admettons dans la langue ou qu'on le met phonétique. Une petite *flip chart*. Moi je l'ai dans mon téléphone que je me suis fait personnel. [...] Ça fonctionne. » (SI₂)

Quelques intervenants nomment des solutions à plus grand déploiement :

- 8) « Si on avait un centre désigné, un centre désigné multiculturel, mais quand on a des gens vulnérables qu'on les transporte, qu'on les amène voir un consultant, médecin, que j'ai du monde qui peuvent parler sa langue, quand ils le reçoivent. Quelqu'un qui peut les accueillir, quelqu'un qui peut, comme on dit... la culture. Ça, ça serait important. » (US₂)

La solution la plus souvent employée à l'heure actuelle est l'outil Google Translate à partir du téléphone cellulaire :

- 9) « Aujourd'hui, on a quand même beaucoup plus d'outils quand même. Tu sais, Google translate... Une madame arabe qui parlait zéro, bien heureusement Google translate nous a comme un peu sauvé la mise. [...] C'est les outils qui nous manquent... » (Approbations) (US2)

Parfois, ce sont les enfants immigrants qui accompagnent leurs parents à des rendez-vous et assurent la traduction. Ceci peut mener à des situations délicates, par exemple lorsqu'il est question d'argent :

- 10) « [D]es fois c'est les petits enfants, des enfants de quatre ans qui font la traduction au comptoir, [...] c'est très gênant. Et la dernière fois c'était une petite fille de six, sept ans, puis son papa ; il parlait au papa, et le papa est mal à l'aise, la petite fille elle nous regarde avec ses grands yeux puis nous on est débouchés parce qu'on parle d'argent, de revenu [...] c'est très délicat. » (SQ1)

Les intervenants doivent parfois bricoler des solutions à la carte, selon leur débrouillardise, leurs connaissances et leurs contacts :

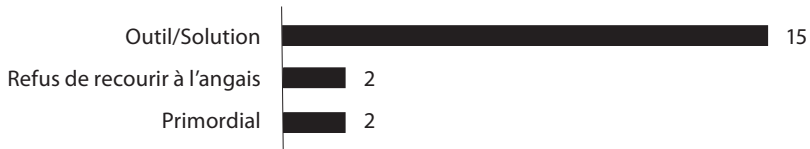
- 11) « [C]'est arrivé à l'occasion où des gens doivent répondre à des appels, des paramédics puis les gens arrivent là puis ils parlent seulement mandarin. C'est arrivé à deux occasions où j'ai eu l'idée de faire une conférence à trois avec des paramédics de l'hôpital chinois à Montréal. Pis les paramédics peuvent faire le lien avec la personne [...] parce qu'ils ont toujours des gens à l'accueil, à l'hôpital chinois. [...] C'est une idée que j'ai eue parce que dans le fond, on n'a pas d'outils » (US2)

L'anglais

La catégorie « anglais » arrive en troisième position des thématiques linguistiques qui émergent. De manière générale, les intervenants présentent l'anglais comme un outil indispensable, particulièrement en situation d'urgence. On voit dans l'exemple 11 que l'anglais est employé dès le début d'une intervention policière :

- 12) « La première fois, tu es obligé de prendre ton cellulaire pour faire *translate* parce que français/anglais, ça ne marche pas » (SP2)

FIGURE 6.4

L'anglais, sous-catégories

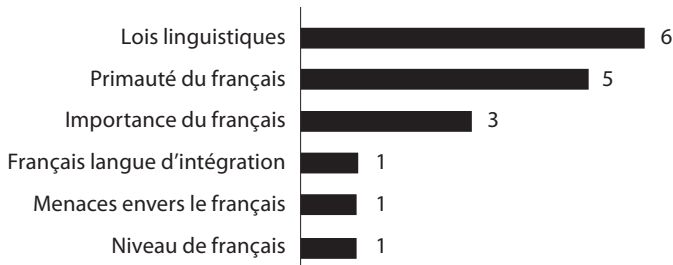
Chez les pompiers, on affirme que connaître l'anglais est une nécessité :

- 13) « [O]n a encore des pompiers qui parlent pas anglais, moi je trouve ça aberrant. » (SIM2)

L'importance du français

La catégorie « importance du français » concerne autant la place du français dans l'espace public québécois que son rôle dans l'intégration des immigrants.

FIGURE 6.5

L'importance du français, sous-catégories

Lorsque les intervenants mentionnent les lois linguistiques, ce n'est jamais pour décrier le fait qu'elles ne sont pas respectées ou connues des nouveaux arrivants. Par exemple, un pompier souligne que les immigrants connaissent les lois linguistiques :

- 14) Les gens savent aussi. Faut pas prendre les minorités ethniques pour des idiots. Ils savent qu'ils sont au Québec, qu'on parle français au Québec en général. (SIM2)

Seules deux interventions rappellent que le français est la langue publique au Québec. Elles viennent d'intervenants des Centres locaux d'emploi :

- 15) « [C]e que tu fais dans ta cuisine, quelle langue que tu parles, ça t'appartient. Je change pas ça. Ce que moi je parle dans ma cuisine versus les immigrants, ça m'appartient. Et j'ai le droit, je suis dans ma cuisine. Mais quand je quitte la maison, on se retrouve sur le terrain commun. » (SQ2)

Chez les policiers, on fait référence au fait que les lois linguistiques du passé ont plutôt favorisé l'école anglaise pour les immigrants et leurs enfants⁵ :

- 16) « [Q]uand la personne peut pas te parler en français là, bien, il y a une raison. Même si ça fait 50 ans, c'est parce qu'elle a pas eu le droit d'aller à l'école francophone ! Parce qu'elle avait pas la bonne religion ! Bien ça [...] fait partie de l'histoire. Il faut que le monde le sache, faut que les gens soient au courant que c'était comme ça que ça marchait avant. Il faut que tu le dises ! Il faut pas que tu dises : "Bah, il a pas su s'adapter !" [...] C'est nous autres qui a fait qu'il a pas su s'adapter ! Parce que c'est nous autres qui l'a empêché d'aller à l'école. » (SP2)

Alors que dans les secteurs d'intervention qui travaillent dans l'urgence l'anglais est perçu comme un outil essentiel pour aider les citoyens, cette perception diffère dans les centres locaux d'emploi, où les intervenants ne veulent pas avoir recours à l'anglais. Ils s'appuient parfois sur le cadre législatif québécois :

- 17) « [...] [M]oi je parle pas anglais, je sais que le gouvernement a pas d'obligation de servir les gens en anglais. [...] Ma chef d'équipe parle pas anglais, la moitié du bureau parle pas anglais [...] c'est pas une obligation. [...] On veut pas, on veut pas l'offrir en anglais. Non parce qu'après ça.... – Sauf que la réalité est là – Non. On est au Québec. » (SQ)

5. « Il faut [...] rappeler que, avant 1964, les écoles françaises étaient catholiques, l'Église contrôlait traditionnellement l'éducation. Cette situation a des répercussions importantes sur l'intégration des immigrants, les immigrants non catholiques s'orientaient vers le secteur protestant et anglophone, plus ouvert à la diversité religieuse. » (Deshaies *et al.*, 1998, p. 112)

Cette attitude peut s'expliquer par le fait que les intervenants des centres locaux d'emploi travaillent moins dans l'urgence et à plus long terme, pour l'intégration sur le marché du travail.

Dans les centres locaux d'emploi apparaît également la thématique que nous avons nommée « la primauté du français », c'est-à-dire l'importance de faire passer le français avant d'autres langues, notamment l'anglais, mais pas uniquement :

- 18) « [L]a langue du service est le français, à la limite l'anglais. C'est pas l'espagnol, l'arabe, j'sais pas moi, le portugais, machin... [...] [O]n vous aide, mais c'est ça, il faut respecter les règles. » (SQ2)

Alors que les services d'urgence affirment utiliser autant l'anglais que le français lors de leurs interventions, chez Service Québec, on privilégie le français :

- 19) « On doit essayer toujours de parler le français, c'est comme [...] le plus important, alors que, par exemple, moi je parle espagnol, quelqu'un me regarde et dit "oh tu parles espagnol!", et il commence à me parler espagnol, et comme ça je dis: "Est-ce que tu parles un petit peu français? On va essayer de parler en français et si vous comprenez pas, je vais vous traduire en espagnol" » (SQ2)

Nous avons distingué le thème de la « primauté du français » de celui de « l'importance du français », où les intervenants expliquent l'importance d'apprendre la langue pour s'intégrer ou travailler. Par exemple, cet employé de Service Québec explique devoir parfois rappeler aux nouveaux arrivants qu'une certaine maîtrise du français est nécessaire avant d'intégrer le marché du travail :

- 20) « Je commence par le référer dans un organisme qui est spécialisé, qui a l'expertise pour travailler auprès des immigrants, et eux autres ils feront un bilan de compétences. Et suite à ce bilan de compétences, parce que eux autres ils vont vraiment lui expliquer et lui ramener les pieds sur la terre. Exemple: secrétaire de direction en... je sais pas moi, en Colombie, qui arrive ici, pis qui veut travailler comme adjointe administrative. "Combien d'années vous avez travaillé en français? Combien de rapports vous avez rédigés en français?" Ah! Donc là qu'est-ce qu'on fait avec? Qu'est-ce qu'on fait avec elle? » (SQ2)

Le seul secteur où on trouve un élément de « menace au français » est dans les bibliothèques :

- 21) « [O]n a vécu quelque chose de similaire avec notre programme de lecture d'été, [...] il y avait quelque chose d'écrit en deux-trois langues. Une signalisation pour les enfants, quelque chose de très ludique... Et c'est un Québécois pure laine qui est venu nous dire "ici c'est une bibliothèque francophone". » (BM2)

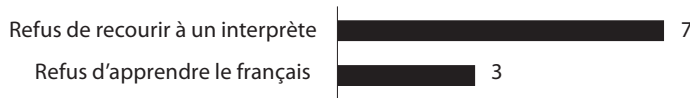
Ici, la peur de voir le français menacé provient de la population et non pas des intervenants.

L'attitude négative des usagers

La thématique « attitude négative des usagers » n'est présente dans notre corpus que pour le secteur des centres locaux d'emploi. Les intervenants soulignent la réticence qu'ont certains immigrants à venir accompagnés d'un interprète, ce qui complexifie l'intervention.

FIGURE 6.6

L'attitude négative des usagers, sous-catégories



- 22) « Moi on venait me chercher tout le temps, mais le problème qui se produit, c'est quand la personne elle sait que quelqu'un parle sa langue au bureau, même si on l'aide la première fois et qu'on lui dit "la prochaine fois vous venez avec un interprète parce que ça va pas fonctionner comme ça", mais ils s'en viennent et c'est "je veux madame là-bas". » (SQ2)
- 23) Ça crée un précédent [...]. – Pis tu lui expliques: "ça se peut que tu arrives une deuxième fois, puis le monsieur n'est pas là, la madame n'est pas là, comment tu vas t'en sortir? Faut que t'essayes de faire des efforts". » (SQ2)

Quelques intervenants déplorent le fait que des immigrants n'essaient pas d'apprendre le français :

- 24) « Pis tu vois il fait un an, deux ans, trois ans... il a même pas appris un traître mot en français. C'est comme s'il veut pas... je sais pas ce qui se passe, mais tout ce qu'il veut c'est que ce soit toi qui le sers, parce que t'es là, tu comprends sa langue, [...] il est comme "OK moi là, c'est lui". [...] – Mais ils cherchent la solution facile pour eux. – Mais on parle des besoins essentiels, la personne est très vulnérable, quelques fois [...] c'est de l'aide sociale, n'a pas l'argent, ou des personnes qui arrivent qui sont demandeurs d'asile... qui ont vécu des situations très difficiles... Alors, il y a des situations que la personne est vraiment dans une crise de... de tout. Parce qu'il vient d'arriver, il connaît rien, et toujours les démarches sont pas faciles, c'est pas juste avec nous la démarche, il faut... le numéro d'assurance sociale, la carte d'assurance maladie... c'est plein de démarches à faire... l'école pour les enfants... » (SQ2)

Un autre intervenant de Service Québec tempère ces propos en évoquant les situations parfois difficiles que vivent les immigrants :

- 25) « Il y en a qui sont ouverts à ça, mais il y en a... ça les bouleverse. Ils ne veulent pas du tout faire l'effort. [...] justement dernièrement il y en a un qui me dit "mais je veux pas apprendre le français, je veux pas. [...] je veux pas trouver un travail, je vais trouver dans la communauté puis je vais rester dans la communauté". Pis je dis "tu veux vivre ici ou non?" il me dit "non, je veux vivre dans la communauté". Complètement apeuré. » (SQ3)

Les difficultés du français

La catégorie « Difficultés du français » concerne les difficultés que les intervenants perçoivent chez les usagers en ce qui concerne la langue. Alors que les difficultés liées à l'apprentissage de la langue française et à l'écart entre les cours de francisation et le français « de la rue » étaient très présentes dans les études réalisées auprès des nouveaux arrivants, seul un intervenant du corpus Traits d'union les évoque, une seule fois :

FIGURE 6.7

Les difficultés du français, sous-catégories

- 26) « Comment tu peux demander à quelqu'un de s'intégrer pis chialer parce qu'il s'est pas intégré, alors qu'à la base, la langue française qui est super difficile à apprendre ? » (SP2)

Toutefois, on fait parfois référence à la langue technique qui rend les démarches difficiles à accomplir et les formulaires laborieux à remplir, et ce, même pour des francophones :

- 27) « [E]n principe, normalement pour une demande de service ils doivent remplir des formulaires. Des formulaires qui sont pas faciles à remplir. [...] Quand t'es Québécois sont pas faciles. » (SQ2)

Les intervenants ne s'expriment pas spécifiquement sur le français québécois⁶. Seulement trois commentaires mentionnent au passage le français québécois, et le seul qui fait état d'une difficulté provient d'un intervenant qui est lui-même issu de l'immigration :

- 28) « [D]onc moi je suis arrivé ici adolescent. Le choc culturel ça a été le franglais. [...] Ils viennent te dire "Watch ton bus" [...] Oui c'est surveille ton autobus. » (SQ3)

Le français, insuffisant pour s'intégrer

Tout comme l'avaient mentionné les nouveaux arrivants, les intervenants estiment que, même si la francisation est une étape importante vers l'intégration, elle n'est pas suffisante. On évoque le fait qu'une intégration réussie relève également de facteurs culturels autres que linguistiques.

6. Rappelons que les études sur les perceptions des populations immigrantes ont relevé que le français québécois est parfois mal perçu ou mal compris, par exemple parce qu'il ne s'agit pas de la variété de français que les nouveaux arrivants connaissaient avant leur arrivée au Québec, ou parce qu'il ne s'agit pas de la variété mise de l'avant dans les cours de francisation.

- 29) « Je dirais que parfois les gens se disent “je suis francophone, ça va tout de suite marcher, parce que j’ai déjà la langue”, mais c’est pas une question de langue. [...] [O]n idéalise beaucoup l’immigration. » (SQ3)

Ici se pose toute la question de la reconnaissance des acquis et des diplômes obtenus à l’étranger. Cette intervenante de Service Québec rappelle qu’une personne immigrante doit parfois franchir plusieurs étapes avant de pouvoir exercer le métier ou la profession qu’elle exerçait dans son pays d’origine (équivalence de la formation antérieure, francisation, nouvelle formation, etc.):

- 30) « [D]ans leur pays ils étaient ingénieurs, ils étaient médecins, ils étaient... “Voici ici, au Québec, [ce que] ça donne.” Une fois que ça c’est fait, on peut [les] aider aussi en formation. Mais si, dans leur pays d’origine, ils avaient pas de formation, on peut dire “[...] on va commencer par faire de l’orientation.” On a un service d’orientation qu’on peut leur offrir. Ils passent par l’orientation. [...] j’en ai eu cette semaine, qui parlent pas le français, qui parlent pas l’anglais, ils parlent juste le hongrois ou le russe; on commence par la francisation. Il y a des étapes à faire, avant même de pouvoir penser intégrer ces gens-là sur le marché du travail québécois. » (SQ1)

Toutes ces étapes impliquent un processus s’échelonnant sur plusieurs années:

- 31) « Là c’est un parcours qui dure une, deux, trois, quatre, cinq ans, avant de réussir à intégrer quelqu’un sur le marché du travail. Quelques fois je prends des gens qui sont au niveau pré-secondaire, qui ont jamais fréquenté l’école, [...] j’en ai eu cette semaine, là, une femme elle a pas son primaire. Je pars de là. Elle parle pas le français, pas l’anglais puis elle a pas le primaire. Je pars d’où, là? Puis je veux l’intégrer, [...] écoute c’est des étapes là, puis c’est pas rien qu’un an, c’est deux, trois, quatre ans avant même d’espérer pouvoir. Puis elle est pas fonctionnelle, elle est pas capable d’aller à l’épicerie. » (SQ1)

La durée de l’intégration s’allonge dans le cas des personnes allophones qui n’étaient pas scolarisées dans leur langue première.

Les convergences et les divergences

Certaines thématiques et perceptions linguistiques nommées par les nouveaux arrivants dans les études que nous avons recensées au début de ce chapitre se retrouvent chez les intervenants du corpus Traits d'union. L'apprentissage du français est perçu comme une étape importante de l'intégration, autant chez les immigrants que chez les intervenants, même si les deux groupes admettent que la connaissance de la langue, à elle seule, est insuffisante. La barrière de la langue est désignée comme un enjeu important dans les deux groupes. Cependant, alors que les immigrants regrettent l'insuffisance des cours de francisation et la difficulté de pratiquer la langue avec des francophones, cet aspect n'est pas directement évoqué chez les intervenants.

Parmi les divergences relevées en ce qui concerne les thèmes abordés, ceux de la primauté et de l'importance du français sont très présents chez les intervenants de Service Québec. Or, ce frein n'apparaît pas dans les études auprès des nouveaux arrivants que nous avons présentées dans la recension des écrits : ces dernières concluaient que les immigrants comprenaient les raisons derrière la protection du français au Québec (Amireault et Lussier, 2008). Lorsqu'on sort du secteur des centres locaux d'emploi, les intervenants n'invoquent pas la primauté du français. Les différents contextes d'intervention pourraient expliquer ce constat.

Les intervenants vivent d'importantes difficultés d'ordre linguistique ; ceci ajouté au fait que les parcours d'intégration s'échelonnent sur plusieurs années semble soutenir les critiques envers certaines nouvelles dispositions de la *Charte de la langue française*, qui réduisent à six mois la période pendant laquelle les intervenants de l'État peuvent s'adresser aux immigrants dans une autre langue que le français⁷ ; six mois de francisation ne seraient pas toujours suffisants pour atteindre des niveaux fonctionnels pour tous les nouveaux arrivants⁸. C'est le cas par exemple pour les arrivants faiblement scolarisés dans leur langue d'origine ou ceux dont la langue est très éloignée du français.

7. La nouvelle mouture de la *Charte de la langue française* stipule à son article 22.4 : « Un organisme de l'Administration doit mettre en œuvre des mesures qui assureront, à la fin d'une période de six mois, des communications exclusivement en français avec les personnes immigrantes, lorsque, afin de fournir des services pour l'accueil de ces personnes au sein de la société québécoise, il utilise une autre langue que le français en vertu de l'article 22.3 ».

8. Rappelons que nos données, tirées d'un corpus visant les relations interculturelles et non spécifiquement les enjeux de francisation, n'abordent pas directement les niveaux de compétence atteints dans l'apprentissage du français.

Les intervenants n'évoquent presque pas les particularités du français québécois : sont-ils sensibilisés au fait que certains immigrants ont probablement été exposés à d'autres variétés du français avant leur arrivée au Québec, et que les cours de francisation n'abordent pas tous la variation de la même manière ?

Implications en matière de formation et d'enseignement

Le projet de recherche-action Traits d'union avait pour but de répertorier les besoins de formation des intervenants auprès des populations immigrantes. En plus des formations de base déjà existantes en communication interculturelle, l'IJC a développé une formation courte en rédaction interculturelle⁹, qui explore les façons selon lesquelles les différences culturelles peuvent se manifester à travers les écrits, et les manières d'en tenir compte pour parvenir à une communication la plus efficace possible, en réduisant les irritants. D'autres formations pourraient éventuellement être développées.

Mises à part les formations ponctuelles qui peuvent pallier certains manques, une réflexion plus globale sur la francisation pourrait contribuer à une meilleure intégration des nouveaux arrivants. Compte tenu de l'écart entre les cours de francisation et la langue courante, on comprend la nécessité de veiller à ce que les cours de francisation soient suffisants et assez proches de la langue de tous les jours, si l'on veut qu'ils mènent à une intégration réussie dans la société. Est-ce que les parcours de francisation permettent aux nouveaux arrivants de développer toutes les compétences nécessaires à une intégration réussie ? Une réflexion en ce sens semble s'imposer.

Références

- AMIREAULT, Valérie et Denise LUSSIER (2008), *Représentations culturelles, expériences d'apprentissage du français et motivations des immigrants adultes en lien avec leur intégration à la société québécoise : étude exploratoire*, Québec, Office québécois de la langue française.
- ATAEPOUR, Niloufar (2016), *La perception d'apprenants immigrants iraniens adultes de leurs cours de français langue seconde à Montréal, de leur investissement social et de leur intégration linguistique*, mémoire de maîtrise, Montréal, Université du Québec à Montréal.

9. <https://www.telugu.ca/institut-jacques-couture/formations/formations-courtes/redaction-interculturelle/>

- BEAULIEU, Suzie *et al.* (2021), Cours de français langue seconde pour personnes immigrantes à Québec: portrait des habiletés orales en fin de parcours, *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, 24(3), 1-29.
- BENZAKOUR, Fouzia (2004), Les stéréotypes associés aux constructions sur la langue dans le contexte d'immigration récente au Québec, dans Denise DESHAIES et Diane VINCENT (dir.), *Discours et constructions identitaires*, Québec, Presses de l'Université Laval, 155-163.
- BLETON, Paul, Angéline MARTEL et Nancy GAGNÉ (2022), Malentendus interculturels: analyses de représentations d'intervenants et intervenantes de secteurs publics québécois en vue de la formation aux compétences interculturelles, *Enjeux et société*, 9(1), 33-65.
- BOUDREAU, Annette et Marie-Ève PERROT (2005), Quel français enseigner en milieu minoritaire? Minorités et contact des langues: le cas de l'Acadie, *Glottopol*, 6, 7-21.
- CALINON, Anne Sophie (2015), Légitimité interne des politiques linguistiques au Québec: le regard des immigrants récents, *Minorités linguistiques et sociétés*, 5, 122-142.
- CONSEIL DE L'EUROPE (2001), *Un cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*, Paris, Didier.
- DESHAIES, Denise, Conrad OUELLON et Claude ROCHELEAU (1998), L'aménagement linguistique au Québec: historique et perspectives d'avenir, dans Alain BÉLANGER, Nubia HANCIAUS et Sylvie DION (dir.), *L'Amérique française: introduction à la culture québécoise*, Editora da Furg, Rio Grande, 89-132.
- GADET, Françoise (2004), Quelle place pour la variation dans l'enseignement du français langue étrangère ou seconde?, dans Jørgen U. Sand (dir.), *Pré-textes franco-danois*, Roskilde, Université de Roskilde, 17-28.
- GAGNÉ, Nancy *et al.* (2020), *Projet Traits d'union: compétences interculturelles en action. Rapport de recherche portant sur cinq secteurs d'intervention à Montréal, Longueuil et Laval (Québec)*, Montréal, Institut Jacques-Couture, <https://r-libre.telug.ca/2020/>.
- GIROUD, Anick et Christian SURCOUF (2016), De « Pierre, combien de membres avez-vous? » à « Nous nous appelons Marc et Christian »: réflexions autour de l'authenticité dans les documents oraux des manuels de FLE pour débutants, *SHS Web of Conferences* 27/07017, 1-18.
- HARVEY, Marie-Hélène (2016), *Enseignement du français québécois et exposition à ses différents usages: représentations d'apprenants immigrants adultes à Montréal*, mémoire de maîtrise, Université du Québec à Montréal.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2022), *Le bilan démographique du Québec. Édition 2022*, Québec, statistique.quebec.ca/fr/fichier/bilan-demographique-quebec-edition-2022.pdf.
- KIRCHER, Ruth (2012), How Pluricentric is the French language? An Investigation of Attitudes towards Quebec French compared to European French, *Journal of French Language Studies*, 22(3), 345-370.
- KIRCHER, Ruth (2014), Thirty years after Bill 101, A Contemporary Perspective on Attitudes towards English and French in Montreal, *Revue Canadienne de linguistique appliquée*, 17(1), 55-80.

- MAURAI, Jacques (2008), *Les Québécois et la norme*, Québec, Office québécois de la langue française.
- MINISTÈRE DE L'IMMIGRATION, DE LA FRANCISATION ET DE L'INTÉGRATION (2022), *Présence et portraits régionaux des personnes immigrantes admises au Québec de 2010 à 2019*, Gouvernement du Québec.
- POTVIN, Maryse *et al.* (2018), La formation des enseignants sur la diversité et les rapports ethniques, *Regard comparatif France, Québec, Belgique et Suisse, Éducation et francophonie*, 46(2), 30-50.
- POULIOT, Sophie, Susie GAGNON et Yolande PELCHAT (2015), *La formation interculturelle dans le réseau québécois de la santé et des services sociaux. Constats et pistes d'action* [rapport], Québec, Institut national de santé publique du Québec.
- REMYSEN, Wim et Amélie-Hélène RHEAULT (2023), La linguistique populaire dans la Romania : le français au Québec, dans Lidia BECKER, Sandra HERLING et Holger WOCHLE (dir.), *Manuel de linguistique populaire*, Berlin/Boston, De Gruyter, 423-447.
- ROY, Sylvie (2015), Discours et idéologies en immersion française, *Revue canadienne de linguistique appliquée*, 18(2), 125-143.
- THIBAUT, Laurence *et al.* (2015), Pour que les étudiants de FLS comprennent et participent à la francophonie canadienne, *Cahier de l'ILOB*, 7, 125-140.
- UNESCO (2015), *Éducation à la citoyenneté mondiale. Préparer les apprenants aux défis du XXI^e siècle*, Paris, UNESCO.
- VILLENEUVE, Anne-Josée (2017), Normes objectives et variation socio-stylistique, le français québécois parlé en contexte d'entrevues télévisées, *Arborescences*, 7, 49-66.

CHAPITRE 7

La vie montréalaise racontée au sein de la Petite Italie et du Petit Portugal

Fabio Scetti

Les communautés italienne et portugaise sont de véritables pièces du puzzle urbain de la ville de Montréal. Les quartiers de la Petite Italie et du Petit Portugal ont été fondés respectivement dans les années 1910 et 1950 et, depuis, ils se sont étendus sur le territoire urbain au-delà de leur « enclave ethnique ». La Petite Italie, *Piccola Italia*, qui chevauche le marché Jean-Talon au nord du Mile End, et le Petit Portugal ou *Comunidade* (communauté), installé sur Le Plateau-Mont-Royal et longeant le boulevard Saint-Laurent, entre le centre-ville et le Mile End, demeurent des lieux de référence du passage de ces deux flux migratoires dans la métropole québécoise. La fondation des deux communautés a marqué l'histoire de la ville et de la province. Les deux communautés ont joué un rôle important pendant la Révolution tranquille en ce qui concerne l'intégration linguistique, notamment à la suite des politiques linguistiques, dont notamment la loi 22 et la loi 101¹, et leurs répercussions sur la scolarisation des enfants immigrés dans le système provincial.

Nos recherches, qui concernent l'observation des pratiques langagières au sein de ces deux groupes, ouvertes en 2011, se sont prolongées jusqu'en

1. La loi 22 (*Loi sur la langue officielle* [1974]) fit du français la langue officielle dans certains secteurs (législation, administration, entrepreneuriat, enseignement, affaires et commerce). La loi 101 (*Charte de la langue française* [1977]) est une loi fondamentale qui fait partie du statut de la province francophone comme d'autres lois, telles la *Charte des droits et libertés de la personne du Québec* (1975), qui garantit les droits culturels des citoyens, et la *Déclaration sur les relations interethniques et interraciales* (1985).

2024. Durant les différents séjours sur le terrain, nous avons recueilli deux corpus sociolinguistiques, composés d'observations ethnographiques de la vie des quartiers et d'entretiens. Ils nous ont permis de raconter l'histoire de Montréal et la vie de ses citoyens y ayant vécu pendant plus d'un siècle, en considérant à la fois les récits des plus jeunes membres interviewés et les histoires des plus âgés et de leurs ancêtres. Grâce aux récits recueillis, nous pouvons redessiner l'histoire politique, sociale et linguistique de la ville et de la province. De plus, les narrations individuelles peuvent se regrouper dans une seule et unique histoire collective, celle des discours circulant au sein de ces communautés, des pratiques langagières diverses selon les répertoires des locuteurs et des locutrices, ainsi que des représentations identitaires.

Tous ces témoignages nous montrent le rôle des langues et des dialectes au sein des familles et de la communauté, cette dernière étant devenue un véritable lieu d'ancrage de l'identité du groupe. Après une description historique et sociopolitique des deux contextes étudiés, cette contribution fournit des exemples de pratiques dans les langues parlées, tels qu'ils émanent des récits de vie, ainsi que des observations de la vie quotidienne. Ces récits sont précieux pour brosser le portrait du Montréal d'aujourd'hui, et apportent des éléments complémentaires dans le processus de construction de l'identité montréalaise au *xxi*^e siècle. Ces communautés et leurs quartiers respectifs deviennent alors de véritables îlots de la mémoire communautaire, mais aussi de la mémoire urbaine montréalaise.

Ce chapitre mettra en évidence l'importance de ce travail de recueil des corpus multilingues dans le contexte urbain montréalais, en tant qu'outil pour analyser la manière dont les communautés linguistiques entretiennent un rapport subjectif et social avec les langues de leur répertoire et avec les autres communautés. L'exploitation de ces corpus permet ainsi de mieux comprendre comment ces langues participent à la construction, à l'affirmation ou à la transformation de leur identité individuelle et collective.

Montréal, porte d'entrée du Canada

Tout au long de son histoire, Montréal a été l'une des portes d'entrée du Canada et constitue aujourd'hui une ville cosmopolite qui attire des personnes de partout dans le monde. Au cours des *xix*^e et *xx*^e siècles, la ville a d'abord accueilli de nombreux migrants arrivant d'Europe, puis d'Asie, d'Afrique et d'Amérique centrale et méridionale. À la fin du *xix*^e siècle, la ville était le point de rencontre, tout particulièrement durant l'hiver, des

travailleurs saisonniers qui revenaient des forêts ou des plaines, en quête d'un nouvel emploi pour la saison suivante. Ces travailleurs saisonniers se sont installés progressivement dans la ville, alors que Montréal s'affirmait comme le centre industriel et commercial du pays (Linteau, 2000).

Aujourd'hui devenue une métropole, elle rassemble dans et autour d'elle environ 4 millions d'habitants : un puzzle multicolore et « multisonore », où de nombreuses langues ont imprimé leurs marques sur le tissu urbain. Définis comme des « enclaves ethniques » (Zucchi, 2007), il s'agit d'espaces originellement créés selon l'axe central qui coupe la ville entre l'est et l'ouest, le boulevard Saint-Laurent, le long duquel s'égrenent le quartier chinois de Ville-Marie, le Petit Portugal de Saint-Louis sur Le Plateau-Mont-Royal, la communauté juive hassidique du Mile End, la Petite Italie et le quartier hellénique de Parc-Extension. D'autres quartiers dits « ethniques », plus tardifs, se sont développés dans d'autres arrondissements de la ville, comme le quartier libanais de Saint-Laurent, également nommé « Saint-Liban », le quartier haïtien de Saint-Michel, le Petit Maghreb de Villeray et les communautés sud-américaines de La Petite-Patrie.

Afin de raconter l'histoire de Montréal au travers des yeux des Montréalais « d'ailleurs », cette étude présente deux groupes que nous avons étudiés dans le cadre de deux corpus : un premier lors d'une étude doctorale, la communauté portugaise (2011-2016) et un second pendant un projet postdoctoral, la communauté italienne (2019-2024). Bien que différentes par leur histoire et leur composition, ces deux communautés présentent de nombreuses similitudes dans leurs pratiques langagières, et dans les représentations qu'elles construisent d'elles-mêmes et qu'elles projettent auprès des autres dans le contexte urbain montréalais. Le premier corpus, tiré du projet MontPORT, se compose de 52 entretiens réalisés entre 2011 et 2014, principalement en portugais (deux en français, un en anglais). Le corpus inclut aussi des observations qui ont été enregistrées au magnétophone, conformément à la pratique ethnographique (Copland et Creese, 2015). Du point de vue de la langue, on a pu retrouver différentes variétés de portugais, notamment celle des îles (Açores et Madère) et celle du Portugal continental. Clos en 2016, ces travaux ont été publiés en 2019, dans un livre tiré de la thèse (Scetti, 2019). Pour le deuxième corpus, la collecte de données a été plus laborieuse en raison des restrictions dues à la pandémie de la COVID-19. Le corpus du projet MontreITal se compose de 100 entretiens réalisés en deux temps, entre 2018 et 2021 et entre 2022 et 2023 (un premier terrain exploratoire avait débuté en 2011 avec quelques observations, mais ce projet a

commencé officiellement en 2019 pour se terminer en 2024. Entre la fin de 2019 et la fin de 2021, la recherche a été réalisée en ligne). La majorité des entretiens se sont déroulés en italien (un seul en français, 14 en anglais et un dans les trois langues), et ont été complétés par des observations de la vie quotidienne où nous pouvons répertorier l'usage de plusieurs autres langues et dialectes d'Italie (Avolio, 2009). Au cours de l'analyse, nous avons étudié la variation de l'italien selon la provenance des locuteurs et des locutrices, du nord au sud, et nous avons pu retrouver la présence d'autres langues et dialectes d'Italie exerçant une forte influence sur cet italien.

La collecte de ces deux corpus nous a menés à certaines interrogations concernant leurs propriétés, mais aussi leur pérennité. S'agissant de deux communautés cristallisées dans le temps, nous nous sommes ainsi questionnés sur notre rôle d'enquêteur, notamment au moment de la sélection des informateurs et des informatrices durant l'enquête de terrain.

Un contexte montréalais multicolore et « multisonore »

Montréal est perçue comme étant une ville séparée et unie par les langues, entre autres du fait des enjeux relatifs au contact des langues et à la dualité linguistique anglais-français, qui fait, très concrètement, partie de son histoire et de son quotidien. Cette situation de « double majorité » (Ancil, 1984) peut aussi être interprétée à la lumière des idéologies politiques, Montréal se trouvant au carrefour de l'interculturalisme francophone issu de la province du Québec et du multiculturalisme fédéral canadien (Bouchard, 2012 ; McAll, 1992). Ces deux langues partagent pourtant un même espace avec d'autres langues, ce qui donne à cette ville québécoise une allure de kaléidoscope.

Différentes communautés vivent dans cette réalité multiculturelle où le débat linguistique est souvent à l'ordre du jour. Les communautés italienne et portugaise ont su construire leurs propres espaces autour de pratiques culturelles, traditionnelles et langagières. Il faut tout particulièrement souligner que ces deux communautés, même si elles sont plus nombreuses à Toronto qu'à Montréal depuis les années 1980, sont plus enclines à maintenir leurs langues à Montréal que dans le contexte uniquement anglophone de Toronto (Da Rosa et Poulin, 1986), ce qui contribue à la spécificité de la métropole québécoise dans son identité de « ville plurielle ».

Le Petit Portugal, cœur de la communauté portugaise de Montréal

La date officielle de l'arrivée des premiers Portugais au Canada est le 13 mai 1953 (Moura et Soares, 2003), date importante pour le groupe puisqu'elle symbolise la première apparition de la Vierge Marie à Fatima, dont c'était le 70^e anniversaire en 2023. De nombreux Portugais sont arrivés au Canada à la suite de ces pionniers, et beaucoup se sont installés à Montréal, où ils ont fondé la première église et la première association portugaise du pays. Plusieurs membres de la communauté se sont installés dans le quartier Saint-Louis, le long du boulevard Saint-Laurent, entre la rue Sherbrooke et l'avenue Mont-Royal, endroit aujourd'hui surnommé le Petit Portugal, le « quadrilatère portugais » ou, tout simplement, *a Comunidade*. Les premiers arrivés étaient originaires des Açores et de l'île de Madère. Plus tard sont venus ceux des régions du nord et du centre du Portugal, à savoir les « continentaux ». Depuis les années 1950, certains ont déménagé et se sont installés dans les banlieues du nord, du nord-est et du sud du fleuve Saint-Laurent (Teixeira et Da Rosa, 2009). Néanmoins, le Petit Portugal est resté le centre de la *Comunidade* et conserve une concentration d'institutions, d'associations ainsi que des lieux de divertissement et de restauration : des boutiques, des bars, des cafés, des pâtisseries et des restaurants. On y trouve, par exemple, l'Associação Portuguesa do Canadá (APC-1956), située entre la rue Saint-Urbain et la rue Rachel. À côté de l'APC se dresse l'église de la Missão Santa Cruz (1984), centre de la vie catholique. Son parvis est en partie occupé par l'école communautaire née de la fusion entre l'Escola Santa Cruz (primaire) et l'école Lusitana (secondaire). Dans cette école, on enseigne encore aujourd'hui la langue portugaise aux enfants de 6 à 18 ans. Les cours ont lieu le samedi matin, bien que les inscriptions aient connu une baisse à partir des années 2000 (Eusébio, 2001). En remontant le boulevard Saint-Laurent, on trouve la banque de la communauté, la Caixa portuguesa Desjardins, le parc du Portugal, décoré par une murale représentant la célèbre fadiste Amália Rodrigues, et le Clube Portugal de Montréal (1965).

La communauté portugaise est la deuxième plus grande du Canada après celle de Toronto, avec 51 585 personnes dans la région du Grand Montréal (Statistique Canada, 2023). D'autres communautés sont implantées à Vancouver, à Hamilton, dans la région d'Ottawa et de Gatineau, à London dans l'Ontario et à Winnipeg. La communauté portugaise de Montréal diffère beaucoup des autres communautés portugaises du Canada, étant la

seule où les Portugais du continent sont presque aussi nombreux que les «iliens». La question de la langue se pose alors, la norme du portugais continental s'imposant comme langue du groupe et unique drapeau de la communauté. Toutefois, dernièrement, la communauté brésilienne s'est rapprochée de la *Comunidade* et, avec elle, on observe la présence d'une autre variété de portugais, le portugais brésilien.

La communauté italienne à Montréal et la naissance de la Petite Italie

On commence à parler de la migration italienne au Canada dès le début du xx^e siècle. Lors du recensement de 1901, on comptait déjà 10 834 Italiens au Canada, et ce chiffre a vite augmenté, en passant à 66 769 en 1921, et à environ 100 000 dix ans plus tard (Grassi *et al.*, 2014).

Montréal a été la première ville où les Italiens se sont installés au Canada. Les premières décennies du xx^e siècle ont été celles de la transition d'une «colonie» italienne vers une véritable «enclave ethnique», appelée pour la première fois «*Little Italy*», en anglais (Ramirez et Del Balso, 1980), suivant l'exemple d'autres enclaves italiennes du même genre qui s'étaient développées au Canada et aux États-Unis (Harney, 1979)². Pourtant, ce n'est qu'après la Deuxième Guerre mondiale que les chiffres ont vraiment changé, car, selon le recensement de 1971, on comptait 730 830 Italiens résidant au Canada (Grassi *et al.*, 2014), alors qu'en 2006, selon Statistique Canada, il y avait 299 965 individus qui se déclaraient de nationalité italienne et 1 445 330, d'origine italienne dans tout le pays. L'italien était alors la langue parlée par 467 905 citoyens canadiens, représentant la troisième langue la plus parlée du pays après l'anglais et le français (Statistique Canada, 2009). Ce chiffre doit toutefois être discuté par rapport à la réalité au sein de ces communautés, où nous avons observé que les dialectes sont le plus souvent parlés au foyer. Lors du recensement suivant, ce classement a changé, l'italien ayant été dépassé par les langues chinoises (Statistique Canada, 2016). Avec un total de 267 240 personnes d'ascendance italienne dans sa région urbaine (Statistique Canada, 2023), Montréal accueille tout de même la deuxième plus grande communauté italienne au Canada, après celle de Toronto. En

2. Il y avait les «*Little Italies*» de East Harlem et de Lower East Side à New York, le North End de Boston, celle de Federal Hill à Providence (Rhode Island), de Fishermen's Wharf à San Francisco, celle d'Oswego (New York), de Baltimore, de Saint-Louis (Missouri), de Philadelphie, de Newark (New Jersey), de Tampa Bay (Floride), et enfin celles de Thunder Bay et de Toronto, en Ontario (voir Gabaccia, 2006; Harney et Scarpaci, 1981; Mormino, 1986; Immerso, 1999).

matière d'origines, cette communauté se caractérise par une prédominance de personnes venant du Frioul et de la Vénétie (nord-est), du Latium, des Abruzzes et du Molise (centre-sud) et du Sud, notamment la Campanie, les Pouilles, la Calabre et la Sicile.

Contrairement à la communauté portugaise, les Italiens se sont installés dans différents quartiers de Montréal et ont construit de nombreuses « *Little Italies* » : au tout début, dans les quartiers Sainte-Agathe et Saint-Timothée, où, en 1905, ils ont fondé l'église Madonna del Carmelo (Notre-Dame-du-Mont-Carmel) ; dans l'arrondissement de Ville-Émard, autour des années 1910, attirés par le développement des usines dans les environs du canal Lachine ; et autour du marché Jean-Talon, un peu plus tard, où ils ont fondé la paroisse de Madonna della Difesa (Notre-Dame de la Défense) en 1911. C'est dans cette partie de la ville qu'entre 1920 et 1930, des familles italiennes se sont installées et ont ouvert des épiceries et des restaurants. C'est en 1936 qu'est ouverte la Casa d'Italia, centre névralgique important de la vie des Italiens. Ce quartier a aussi changé de nom, passant de l'anglais « *Little Italy* » au français « la Petite Italie » (Ramirez, 2007).

Après les vagues migratoires qui suivirent la Seconde Guerre mondiale, de nombreuses familles se sont installées dans d'autres parties plus abordables de la ville, à l'Est et au Nord-Est, entre autres à Saint-Michel, à Montréal-Nord, à Saint-Léonard, à Rivière-des-Prairies, à Anjou et à Laval. Saint-Léonard est devenu le quartier le plus italien de Montréal, l'église Madonna del Carmelo y ayant été relocalisée à cause des travaux engendrés par l'Expo 67. On y a également fondé l'hôpital Santa Cabrini en 1960 et le centre culturel Leonardo da Vinci en 2002. Parmi tous ces quartiers, la Petite Italie du Mile End est considérée comme le centre de la communauté. Ce quartier est le symbole du groupe et un lieu de la mémoire. Il est devenu le lieu où l'on achète des produits italiens et où l'on trouve des restaurants et des cafés italiens. C'est à cet endroit que l'on peut vivre l'« italianité », étudiée comme processus d'auto-identification ethnique.

À propos de la langue, il faut noter que la communauté italienne n'a pas de véritable école, comme la *Comunidade*, mais plutôt un système d'écoles³, qui proposent des cours d'italien gérés par deux institutions : le Patronato Italo Canadese per l'Assistenza agli Immigrati (PICAI)⁴ et le Centro Scuola

3. Une vingtaine d'écoles (primaires et secondaires) dans la région métropolitaine de Montréal.

4. Patronat italo-canadien pour l'assistance aux immigrés.

Dante Alighieri (CESDA)⁵. Ici, la question des langues se pose non seulement en matière de variété d'italien par rapport aux origines différentes de ses locuteurs et locutrices, mais aussi en termes d'autres langues et dialectes qui se parlent encore au sein du groupe (sicilien, napolitain, sarde, frioulan, etc.). On ne trouve pas leur trace dans les statistiques canadiennes (sauf probablement dans la partie *Langues italiques (romanes)*, *n.i.a.* – non incluses ailleurs), et c'est grâce à des études qualitatives comme celle qui nous occupe ici qu'il est possible de prendre en compte la variété dialectologique au sein des communautés italiennes, au Canada, mais ailleurs également.

Exploitation des corpus du point de vue linguistique et discursif

Dans cette contribution, nous exploitons nos données dans le but de reconstruire la relation entre langue et identité au sein des deux communautés. Les extraits que nous employons ici sont présentés avec le prénom ou le pseudonyme du locuteur ou de la locutrice, la génération⁶, l'année de naissance et l'origine géographique. En note, on trouve chaque énoncé en langue originale sans aucune correction ou modification.

Dans notre travail de comparaison des deux corpus étudiés, nous observons la relation entre pratiques langagières et représentations identitaires, pour comprendre le processus de construction de l'identité d'héritage au sein des deux communautés (Adserà et Pytliková, 2016; Canagarajah et Silberstein, 2012; Norton et Toohey, 2011; Kerswill, 2006; Esser, 2006; Pavlenko et Blackledge, 2003; Errington, 2001; Irvine et Gal, 2000). Pour, aussi, comprendre la relation qui existe entre ces communautés et le contexte d'accueil montréalais. L'«identité d'héritage», fondée sur une identité collective, tient ici un rôle important, car elle est envisagée en tant que processus de construction identitaire envers une même origine, ou langue dite d'«héritage», associée au groupe ou à la communauté.

5. Centre scolaire Dante Alighieri.

6. Les informateurs et informatrices ont été classés en trois groupes : première génération (1G), qui comprend des personnes arrivées d'Italie ou du Portugal et qui se sont installées pour la première fois à Montréal; deuxième génération (2G), qui comprend des personnes arrivées d'Italie ou du Portugal en bas âge ou nées au Canada d'au moins un parent italien; troisième et quatrième générations (3G et 4G), qui comprennent des personnes d'origine italienne ou portugaise, dont au moins l'un des grands-parents ou des arrière-grands-parents est arrivé au Canada en provenance de l'Italie ou du Portugal.

Langues parlées et répertoires langagiers multiples

Durant nos enquêtes de terrain, l'observation des pratiques langagières nous a permis de dévoiler la variété des répertoires langagiers de ces Montréalais. Le choix de parler une langue, ou plusieurs, peut être interprété comme un « acte d'identité » (Le Page et Tabouret-Keller, 1985). Dans notre travail d'analyse, nous avons cherché à comprendre comment l'italien est devenu le symbole identitaire de cette communauté, sachant que plusieurs de ses membres ne connaissent qu'à peine cette langue. Nous avons observé que l'italien est parlé dans le contexte familial et associatif au sein de la communauté, mais que parfois il se mélange non seulement avec le français et, plus fréquemment, l'anglais, mais aussi avec d'autres langues régionales de l'Italie. Il faut savoir que, durant nos entretiens, nous avons rencontré des langues et des dialectes d'Italie connus, comme le sicilien et le napolitain, mais aussi des langues plus rares, comme l'arbëresh (dialecte albanais d'Italie) et le faetar (franco-provençal des Pouilles), ce qui rend la communauté très diversifiée du point de vue de la variation linguistique. Nous nous sommes intéressés aussi aux attitudes linguistiques et au rapport aux langues qui changent d'une génération à l'autre. Comme le dit Angela (1G – 1944, Sicile, IT) à propos de ses enfants : « Ils (ses enfants) n'ont pas voulu apprendre l'italien. Je leur parle italien et ils comprennent. Ils me répondent en français⁷ ». La transmission de la langue d'une génération à l'autre devient alors un élément majeur de la construction identitaire. Domenico (2G – 1963, Pouilles, IT), par exemple, souligne l'importance de parler italien pour ses enfants. Il met de l'avant le rôle des « nonni » (grands-parents) qui s'occupaient de ses enfants quand ils étaient petits, mais il note aussi l'importance des cours d'italien du PICAI : « Mes enfants parlent italien, tout a commencé avec les grands-parents qui allaient les chercher à l'école, ils sont allés tous les deux à l'école d'italien (PICAI) et ils n'aimaient pas ça quand ils étaient petits. Après, ils ont compris que c'est fantastique le fait d'être capables de parler une autre langue⁸ ». Même si c'était difficile de comprendre, le sacrifice d'aller à l'école le samedi matin, aujourd'hui, ses enfants se considèrent chanceux de pouvoir parler une « autre langue ». Cette notion de « sacrifice »

7. « *non hanno voluto imparare l'italiano, ci parlo italiano e mi capiscono, rispondono in francese* ».

8. « *i miei figli parlano italiano, è iniziato con i nonni che li andavano a prendere a scuola, sono andati tutti e due a scuola di italiano e non erano contenti da piccoli. Dopo hanno arrivato a capire che è fantastico il fatto che sono capaci di parlare un'altra lingua* ».

par rapport aux cours de langues du samedi matin, on la retrouve aussi dans les entretiens avec les Montréalais d'origine portugaise, comme le raconte Esmeralda (2G – 1966, Minho, PT) : « Maintenant nous obligeons nos enfants aussi, parce que nous voyons le résultat, avec l'âge, quand nous sommes de leur âge (les enfants), nous ne comprenons pas la raison parce que, c'est embêtant, parfois parce qu'il (son fils) peut être en train de faire autre chose le samedi⁹ ». Dans cette communauté aussi, le lien intergénérationnel est important et la langue en est la base. Jeff (3G – 1991, Minho, PT) remarque l'importance de parler le portugais quand ses grands-parents sont présents : « Mais c'est seulement quand nous sommes, toute la famille, à Noël ou Pâques, quand toute la famille est réunie et avec les grands-parents, nous ne voulons pas parler l'anglais parce qu'ils ne comprennent pas et alors, nous parlons tous en portugais, pour que tout le monde comprenne¹⁰ ».

Nous avons aussi traité la question de l'évaluation et de l'autoévaluation concernant l'usage que l'on fait d'une langue. Souvent, selon le parcours scolaire et la profession, nous pouvons rencontrer différents profils de locuteur et locutrice ayant une conscience plus ou moins marquée de leur rapport aux pratiques langagières. En général, nous avons noté plus d'aisance au sein de la communauté portugaise que dans la communauté italienne en ce qui concerne respectivement la pratique du portugais et de l'italien. L'italien standard est perçu comme « véritable » par rapport à celui qui est parlé par les membres, jugé « tordu ». En analysant l'italien parlé à Montréal, Filippa (2G – 1966, Campanie, IT) dit que « les Italiens (de Montréal) le parlent tordu, pas le vrai italien¹¹ », et Rosamaria (2G – 1969, Calabre, IT) anticipe aussi, avant notre entretien : « Je ne parle pas très bien ! J'essaie de parler le vrai italien¹² ». Cette idéologie du « vrai » ou « véritable » italien est intéressante si l'on considère qu'il a été très rare d'entendre des conversations uniquement dans cette langue lors de nos observations et même durant les entretiens. Certes, l'italien se parle, mais on le mélange beaucoup et Vittorio (1G – 1953, Calabre, IT) l'explique d'ailleurs en ces termes : « Je suis analphabète, une

9. « agora a gente obriga os nossos filhos também, porque a gente vê o resultado com a idade quando somos da idade dele não percebe a razão porque, é chato, as vezes porque ele pode estar a fazer outras coisas ao sábado ».

10. « mas é só quando somos a família toda com, Natal, a Páscoa, quando toda a família está conjunta e com os avós, nós não queremos falar em inglês por é, sabes, eles não compreender então em tudo nós falamos sempre em português pa' toda a gente compreender ».

11. « gli italiani lo parlano storto, non il vero italiano ».

12. « io non parlo molto bene! provo di parlare italiano vero ».

fois en dialecte, italien, une fois français, anglais¹³». Les pratiques sont ainsi le plus souvent le résultat du mélange entre dialectes, italien, français et anglais. Pour éviter de mélanger les langues, de véritables politiques linguistiques familiales sont élaborées, sauf qu'il est parfois difficile de s'y tenir, comme l'explique Elizabeth (2G – 1980, Açores, PT) : « Avec mon mari, nous essayons de parler toujours portugais avec notre fille, qui a seulement deux ans, mais ce n'est pas facile de parler (portugais) entre nous¹⁴ ».

Ces exemples illustrent la situation linguistique dans les deux communautés, ainsi que leur rapport aux langues. Parmi les Italiens et les descendants de cette communauté, nous observons une variation dialectale plus ample, qui s'avère un élément important quant à la décision de transmettre l'italien ou de l'apprendre à l'âge adulte. Dans les deux cas, l'influence des langues dites « dominantes », à savoir le français et l'anglais, est perceptible, bien que des discours sur la « pureté de la langue » se fassent entendre au sein des deux communautés, lorsqu'il est question des phénomènes de contact entre les langues, comme l'alternance codique, les interférences et les mélanges (Scetti, 2020).

Convergence identitaire : du quartier « ethnique » à la ville « multiethnique »

On peut définir les espaces géographiques des pratiques langagières par des limites qui nous aident à redessiner les frontières des quartiers « ethniques ». Définir l'identité d'un groupe reste alors une manière de se démarquer de ceux qui n'appartiennent pas à ce groupe (Tabouret-Keller, 1997).

Ces limites existent non seulement entre deux communautés distinctes, mais également au sein d'une même communauté, les pratiques langagières étant parfois suffisantes. Laura (3G – 1990, Latium, IT) remarque une différence entre le « parler du quartier » des zones à haute concentration de familles d'origine italienne et l'italien standard : « À Saint-Léonard et à Rivière-des-Prairies, tout le monde parle de la même manière [...] ils ont le même accent en anglais aussi¹⁵ ». Elle remarque toutefois la même chose en anglais – un anglais qui aurait une construction tonique « à l'italienne », notamment chez les hommes, comme les études de Charles Boberg le

13. « sono analfabeto, una volta in dialetto, italiano, una volta francese, inglese ».

14. « com o meu marido tentamos falar sempre português com a nossa filha, que só tem dois anos, mas não é fácil falar entre nós ».

15. « a Saint-Léonard e a Rivière-des-Prairies tutti parlano alla stessa maniera [...] hanno lo stesso accento anche in inglese ».

montrent (2004 ; 2014). La notion d'accent, ici, pour nous, ne s'observe pas seulement du point de vue strictement linguistique, mais aussi du point de vue sociolinguistique, en tant que « marqueur » ou « indicateur » d'identité (Gasquet-Cyrus, 2010, p. 181). Ces questionnements à propos de l'accent font écho à la notion de l'intégration, qui est aussi un sujet fréquemment abordé lors de nos entretiens. La perspective change ici, puisque « l'accent » se manifeste en fonction de la langue d'intégration dans le processus de construction de l'identité « nationale », comme dans le cas de Dora (2G – 1979, Centre, PT), qui, à la question sur la nationalité et son ressenti, répond d'un air très ferme : « Je me sens canadienne, je suis canadienne de même que mon mari, je parle et je n'ai pas d'accent (en français)¹⁶ ». Filomeno (1G – 1948, Campanie, IT), au contraire, fait un constat plus général et décrit la société montréalaise d'aujourd'hui en tant que « bilingue », même si des extrêmes existent encore : « Ici (à Montréal), ils sont presque tous bilingues, presque toute la population, il y a quelques extrémistes anglais ou français, ils ne veulent pas se soumettre et dire un mot (dans l'autre langue)¹⁷ ».

À propos d'intégration, une autre question souvent discutée est celle de la scolarisation, tout particulièrement à la suite des politiques linguistiques des années 1970 et 1980, notamment celles concernant la loi 101, qui a créé un décalage entre les immigrants arrivés au Québec avant et après 1977 (Mc Andrew, 2002). La communauté italienne a été au centre du débat au cours de cette période ; on se rappelle la « crise de Saint-Léonard » (Linteau, 1989, p. 179), qui fut un moment important de cette « bataille », parce que plusieurs parents italiens étaient désireux d'offrir à leurs enfants un enseignement en anglais plutôt qu'en français et que certains discours suggéraient que les francophones ne voulaient pas des Italiens dans leurs écoles. Teresa (1G – 1943, Molise, IT), qui a travaillé dans une commission scolaire, a un point de vue de l'intérieur et dit que personne n'était vraiment préparé aux changements sociaux liés aux flux migratoires de la seconde moitié du xx^e siècle : « Je l'ai vu de dedans, les Français (francophones) n'étaient pas préparés, ce n'était pas qu'ils (francophones) ni les (Italiens) voulaient pas¹⁸ ». Cependant, l'histoire de Ida (2G – 1949, Molise, IT), qui avait déjà terminé le secondaire en Italie, est aussi un bon exemple pour mieux comprendre la situation réelle

16. « *eu sento-me canadiana, eu sou canadiana tal e qual como o meu esposo, eu falo e não tenho sotaque nenhum* ».

17. « *qua sono quasi tutti bilingui, quasi tutta la popolazione, ci sono qualche estremisti inglesi o francesi, non vogliono essere sottomessi a dire una parola* ».

18. « *io l'ho visto da dentro, i francesi non erano preparati, non è che non li volevano* ».

dans les classes et le fait que cet « accueil » n'avait pas été très bien préparé. Elle raconte : « Ça a été un désastre, je voulais aller à l'école française, mais j'ai dû aller à l'école anglaise. Six mois d'enfer, je me sentais mal avec les enfants de six, sept ans, je pleurais tous les jours, je voulais rentrer en Italie. Mon père avait demandé si j'aurais pu passer la 4^e ou 5^e (années d'école) et ils ont dit non. Alors, j'ai commencé à travailler dans le textile¹⁹ ».

Nous avons ensuite analysé les discours liés à l'identité des communautés et à la relation que les communautés cultivent avec la ville de Montréal. Plus précisément, nous nous sommes concentrés sur l'apport de ces communautés à la ville et le lien que ces quartiers et leur « fonctionnement » entretiennent avec l'identité multiethnique urbaine. Nous observons une convergence des discours qui vont d'une identité « de quartier » et « communautaire » assez bien marquée à une identité montréalaise où tout le monde serait inclus. On parle alors d'identité « plurielle » où les langues deviennent des marqueurs identitaires. Virgilio (1G – 1954, Centre, PT) nous parle de cette distinction comme étant « normale » : « Je suis toujours Canadien et Portugais, OK ? je suis les deux mais c'est normal²⁰ ». Cependant, dans cette pluralité, on observe aussi une division, et Vittorio (1G – 1953, Calabre, IT) souligne l'importance de la vie de quartier et de son indépendance par rapport au reste de la ville : « Nous avons (les Italiens) deux quartiers qui vont très bien. Le premier est Saint-Léonard, puis nous avons Rivière-des-Prairies (RDP). Là-bas aussi, c'est la deuxième communauté. Ils ont (à RDP) des centres sportifs, (un) centre où on y danse tous les jours²¹ ». Cette existence en tant que communauté distincte coïncide, dans une certaine mesure, avec l'identité multiethnique de Montréal.

Bien que différentes, ces communautés convergent vers d'autres réalités, du quartier à la ville, et de la province à l'État. Mário (2G – 1967, Centre, PT) remarque qu'il ne faut pas oublier qu'on fait partie du Canada : « Même si nous nous considérons Portugais, nous faisons partie du Canada²² ». On se rappelle aussi que ces communautés ont contribué à la diversité et à la

19. « *è stato un disastro, volevo andare alla scuola francese ma sono dovuta andare alla scuola inglese. Sei mesi d'inferno, mi sentivo male con i bambini di sei, sette anni, piangevo tutti i giorni, volevo tornare in Italia. Mio padre chiese se potevo passare il 4° o 5° e dissero no. Allora, cominciai a lavorare nel tessile* ».

20. « *eu sempre sou canadiano e português, ok ? Eu sou os dois mas é normal* ».

21. « *c'abbiamo due quartieri che camminano molto bene. Il primo è San Leonardo, poi c'abbiamo Rivière-des-Prairies. Pure là è la seconda comunità. C'hanno centri sportivi, centro che si balla tutti i giorni* ».

22. « *mesmo que a gente se considera português, a gente faz parte do Canadá* ».

multiethnicité de ce pays, ce que confirme Luciana (1G – 1952, Latium, IT) : « Ici, ils (les Italiens) ont retrouvé des conditions favorables. Le Canada a encouragé la multiethnicité²³ ». Angela (1G – 1944, Sicile, IT) souligne, enfin, qu'il s'agit d'une reconnaissance et qu'il ne faut pas oublier ce que ces communautés ont réalisé. Elle souhaite que personne n'oublie le fait que les Italiens ont été très importants dans la construction de la ville montréalaise et de la province québécoise, non seulement du point de vue matériel (industrie du bâtiment), mais aussi du point de vue social : « Les Italiens ont fait le Québec, Montréal²⁴ ! ».

Tous ces exemples soulignent l'apport de ces corpus dans l'analyse des récits sur l'histoire de Montréal, mais aussi du Québec et du Canada, sous un regard différent. Nous insistons ici sur une diversité de discours à propos des moments historiques qui ont marqué la vie de ces communautés à Montréal. La construction identitaire y est perçue comme une juxtaposition de faits et de marqueurs qui démontrent le caractère malléable et instable de ce processus. On parle d'un « bricolage » des identités qui se négocie selon la situation et les participants ; les locuteurs et locutrices utilisent les langues et les formes culturelles tout en essayant de les afficher, de les cacher, de les mélanger, et ce, pour affirmer leur identité ou résister à celle qui leur a été attribuée (Moore et Brohy, 2013).

* * *

Ce chapitre nous a permis de parcourir l'histoire des deux communautés dans la ville de Montréal depuis leur arrivée et la fondation de leurs quartiers ethniques respectifs, et de mieux comprendre comment se construit l'identité plurielle de chaque groupe.

Nos corpus constituent un apport majeur, qui met en relief la diversité des langues utilisées à Montréal. Il reste cependant à trouver un endroit où archiver les données, et une manière de mieux les exploiter. S'agissant des pratiques langagières, on parle d'un ensemble de variétés de langues et de dialectes qui, aujourd'hui, sont parfois à peine parlés dans leurs localités d'origine. Du point de vue des représentations, nous avons pu observer des discours intéressants à propos de l'identité de chacune des communautés et

23. « *qui hanno trovato condizioni favorevoli. Il Canada ha incoraggiato la multiethnicità* ».

24. « *gli Italiani hanno fatto il Quebec, Montreal* ! »

de leur rapport à leurs pratiques langagières et culturelles. Nous pouvons ainsi nous questionner sur la réalité de ces deux enclaves ethnolinguistiques. De plus, en ce qui concerne la période des enregistrements, nous pouvons envisager une analyse des changements survenus au sein de ces communautés après la pandémie de COVID-19, qui a entraîné la fermeture de nombreuses activités communautaires que les jeunes ont du mal à relancer.

Enfin, par rapport à la situation actuelle dans les deux communautés, notre étude nous permet, grâce à une analyse des discours institutionnels existants dans les corpus, de prendre conscience du rôle des institutions et des associations dans le soutien à l'usage des langues d'héritage. À propos de l'usage de l'italien au sein de son institution, par exemple, la directrice de la Casa d'Italia soulignait d'ailleurs, en 2019, que « tout le monde parle italien à la Casa [d'Italia]²⁵ ». Notre étude éclaire aussi la vitalité des activités proposées par ces institutions et associations dans les deux groupes. D'ailleurs, l'implication des jeunes générations dans la perpétuation de l'identité de ces communautés est un facteur crucial, non seulement pour préserver leur nature d'enclaves « exotiques » du panorama urbain montréalais, mais également en tant que pierres angulaires de la ville de Montréal telle qu'on la conçoit aujourd'hui.

Références

- ADSERÀ, Alícia et Mariola PYTLIKOVÁ (2016), Language and migration, dans *The Palgrave Handbook of Economics and Language*, Londres, Palgrave Macmillan, 342-372.
- ANCTIL, Pierre (1984), Double majorité et multiplicité interculturelle à Montréal, *Recherches sociographiques*, 3(25), 441-456.
- AVOLIO, Francesco (2009), *Lingue e dialetti d'Italia*, Rome, Carocci.
- BOBERG, Charles (2004), Ethnic patterns in the phonetics of Montreal English, *Journal of Sociolinguistics*, 8(4) 538-568.
- BOBERG, Charles (2014), Ethnic divergence in Montreal English, *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 59(1), 55-82.
- BOUCHARD, Gérard (2012), *L'interculturalisme. Un point de vue québécois*, Montréal, Boréal.
- CANAGARAJAH, Suresh et Sandra SILBERSTEIN (2012), Diaspora identities and language, *Journal of Language, Identity, et Education*, 2(11), 81-84.
- COPLAND, Fiona et Angela CREESE (2015), *Linguistic Ethnography: Collecting, Analysing and Presenting Data*, Londres, Sage Publications.

25. « le persone parlano tutte italiano nella Casa ».

- DA ROSA, Victor Manuel PEREIRA et Richard POULIN (1986), Espaces ethniques et question linguistique au Québec : à propos des communautés italienne et portugaise, *Canadian Ethnic Studies/Études ethniques au Canada*, 18(2), 143-150.
- ERRINGTON, Joseph (2001), Ideology, dans Alessandro DURANTI (dir.), *Key Terms in Language and Culture*, Malden, Blackwell, 110-112.
- ESSER, Hartmut (2006), *Migration, Language and Integration*, Berlin, WZB.
- EUSÉBIO, Joaquim (2001), *Falando Português em Montréal*, Montréal, Quebec World.
- GABACCIA, Donna Rae (2006), Global Geography of “Little Italy”. Italian Neighborhoods in Comparative Perspective, *Modern Italy*, 11, 9-24.
- GASQUET-CYRUS, Médéric (2010), L’accent : concept (socio)linguistique ou catégorie de sens commun ?, dans Henri Boyer (dir.), *Pour une épistémologie de la sociolinguistique. Actes du colloque international de Montpellier 10-12 décembre 2009*, Limoges, Éditions Lambert-Lucas, 179-189.
- GRASSI, Tiziana, Enzo CAFFARELLI et Mina CAPPUSI (2014), *Dizionario enciclopedico delle migrazioni italiane nel mondo*, Rome, SER.
- HARNEY, Robert F. (1979), Men Without Women, Italian Migrants in Canada, 1885-1930, *Canadian Ethnic Studies/Études ethniques au Canada*, 11(1), 29-47.
- HARNEY, Robert F. et J. Vincenza SCARPACI (1981), *Little Italies in North America*, Toronto, The Multicultural History Society of Ontario.
- IMMERSO, Michael (1999), *Newark’s Little Italy: The Vanished First Ward*, New Brunswick NJ, Rutgers University Press.
- IRVINE, Judith T. et Susan GAL (2000), Language ideology and linguistic differentiation, dans Paul V. KROSKRITY (dir.), *Regimes of Language: Ideologies, Politics, and Identities*, Santa Fe NM, School of American Research Press, 35-84.
- KERSWILL, Paul (2006), Migration and Language, dans Klaus MATTHEIER, Ulrich AMMON et Peter TRUDGILL (dir.), *Sociolinguistics/Soziolinguistik: An International Handbook of the Science of Language and Society* (2^e éd.), Berlin, De Gruyter, tome 3, 2271-2285.
- LE PAGE, Robert B. et Andrée TABOURET-KELLER (1985), *Acts of Identity: Creole-Based Approaches to Language and Ethnicity*, Cambridge, Cambridge University Press.
- LINTEAU, Paul-André (2000), *Histoire de Montréal depuis la Confédération* (2^e éd.), Montréal, Boréal.
- LINTEAU, Paul-André (1989), The Italians of Quebec: Key Participants in Contemporary Linguistic and Political Debates, dans Roberto PERIN et Franc STURINO (dir.), *Arrangiarsi. The Italian Immigration Experience in Canada*, Montréal, Guernica, 179-207.
- MCALL, Christopher (1992), *Class, ethnicity, and social inequality*, Montréal, McGill-Queen’s University Press.
- MC ANDREW, Marie (2002), La loi 101 en milieu scolaire, *Revue d’aménagement linguistique*, n^o hors-série, 69-82.
- MOORE, Danièle et Claudine BROHY (2013), Identités plurilingues et pluriculturelles, dans Jacky SIMONIN et Sylvie WHARTON (dir.), *Sociolinguistique du contact. Dictionnaire des termes et concepts*, Lyon, ENS Éditions, 289-315.

- MORMINO, Gary R. (1986), *Immigrants on the Hill, Italian-Americans in St. Louis, 1882-1982*, Urbana/Chicago, University of Illinois Press.
- MOURA, Manuel de Almeida et Imitério SOARES (2003), *Pionniers: l'avant-garde de l'immigration portugaise, Canada 1953*, Montréal, Direcção Geral dos Assuntos Consulares e Comunidades Portuguesas.
- NORTON, Bonny et Kelleen TOOHEY (2011), Identity, language learning, and social change, *Language Teaching*, 44(4), 412-446.
- PAVLENKO, Aneta et Adrian BLACKLEDGE (2003), *Negotiation of Identities in Multilingual Contexts*, Clevedon, Multilingual Matters.
- RAMIREZ, Bruno et Michael DEL BALSO (1980), *The Italians of Montreal: From Sojourning to Settlement, 1900-1921*, Montréal, Éditions du Courant.
- RAMIREZ, Bruno (2007), Quartiers italiens et Petites Italies dans les métropoles canadiennes, dans Marie-Claude BLANC-CHALÉARD, Antonio BECHELLONI, Bénédicte DESCHAMPS et al. (dir.), *Les Petites Italies dans le monde*, Rennes, Presses universitaires de Rennes, 73-87.
- SCETTI, Fabio (2019), *La communauté portugaise de Montréal: langue et identité*, Québec, Presses de l'Université Laval.
- SCETTI, Fabio (2020), Le portugais est la 7^e langue la plus utilisée au monde! Promotion de la langue portugaise dans deux communautés portugaises en Amérique du Nord, *Langage et Société*, 170(2), 87-108.
- STATISTIQUE CANADA (2009), *Census of Canada (2006), Citizenship, Immigration, Birthplace, Generation Status, Ethnic Origin, Visible Minorities, and Aboriginal Peoples*, Ottawa, Industrie Canada.
- STATISTIQUE CANADA (2016), *Census of Canada (2011), Immigration and Ethnocultural Diversity*, Ottawa, Industrie Canada.
- STATISTIQUE CANADA (2023), *Census of Canada (2021): Mobility and Migration*, Ottawa, Industrie Canada.
- TABOURET-KELLER, Andrée (1997), *Le nom des langues: Les enjeux de la nomination des langues*, Louvain-la-Neuve, Peeters, tome 1.
- TEIXEIRA, Carlos et Victor Manuel PEREIRA DA ROSA (2009), *The Portuguese in Canada: Diasporic Challenges and Adjustment* (2^e éd.), Toronto, University of Toronto Press.
- ZUCCHI, John (2007), *A history of Ethnic Enclaves in Canada*, Ottawa, Canadian Historical Association.

CHAPITRE 8

Approche de la mobilité discursive dans un corpus montréalais

Autour des expressions figées

Anne-Sophie Calinon et Katja Ploog

Alors que le souci de tout locutaire est de produire un discours le plus en phase possible avec l'espace discursif dans lequel il se déploie et de ne se démarquer que positivement, « de nombreuses phrases grammaticales produites par les apprenants sonnent de manière peu naturelle et étrangère » (Ellis *et al.*, 2008, 378, *notre traduction*¹), même pour ceux dont l'habileté est fort développée. La démarcation positive peut se faire par l'élaboration d'éléments dotés d'une socio-indexicalité forte, emblématique, comme les expressions, locutions ou proverbes, élaboration qui reste structurellement très contrainte, par un « figement » relatif. Leur plus-value socio-indexicale comporte ainsi un risque pour la face positive si le locutaire ne les élabore pas de manière adéquate. Nous observerons cette tension en termes de *mobilité discursive* du *dire* (les caractéristiques de l'élaboration séquentielle) et du *dit* (les caractéristiques des constructions représentées; Ploog *et al.*, 2020), dans une situation qui exacerbe les deux vecteurs: l'argumentation d'un projet de mobilité académique.

Notre étude se fonde sur un corpus d'entretiens de recherche recueillis entre 2013 et 2017 auprès de personnes maghrébines étudiantes, engagées

1. « many grammatical sentences generated by language learners sound unnatural and foreign ».

dans un processus de mobilité dans l'espace francophone, actuellement installées au Québec², et s'investissant, de ce fait, dans un processus d'appropriation d'un nouvel espace discursif au Québec, ce qui entraîne la reconfiguration de l'ensemble de leurs réseaux (repères) sociaux. Le corpus d'étude comporte 256 000 mots au total (paroles des enquêtrices comprises)³ pour 10 entretiens. Les locutaires témoins rencontrés ont entre 20 et 35 ans, s'expriment toutes et tous couramment en français et possèdent des compétences variables dans une ou plusieurs autres langues, dont, en particulier, l'arabe littéral (fusha), le berbère, le kabyle, la darija, l'anglais. Les entretiens ont été menés en français, par des chercheuses françaises.

La maîtrise de la langue française est une nécessité pour mener à bien le projet migratoire de ces locutaires; elle est aussi garante de la légitimité même de celui-ci. Naturellement, la dynamique interactionnelle fait émerger des commentaires explicites à propos de cet enjeu⁴:

Extrait n° 1

ENQ je suis au Québec alors je tutoie tout le monde mais

AMR ben non ben c'est ça euh + justement là le tutoiement + c'est une particularité québécoise

ENQ oui {rires} c'est pour ça je m'adapte quand même à la situation du Canada

AMR ben oui

ENQ mais

AMR ben moi je m'adapte là parce que euh vous êtes + vous venez de la France j'essaie de de modifier un peu ma façon de parler (E42)

Ces commentaires indiquent que les locutaires sont conscients des enjeux socio-discursifs et rendent manifeste leur désir de s'adapter, par

2. Les données sont issues du projet CEM Dynamiques spatiales, langagières, identitaires de la circulation migratoire étudiante Maghreb–France–Canada, mené en collaboration avec N. Thamin (Université Marie et Louis Pasteur, ex-Université de Franche-Comté, CRIT), A. Mahieddine et M. Z. Ali-Bencherif (Université de Tlemcen, Dylandimed, Algérie)

3. L'étude est longitudinale; la plupart des personnes ont été rencontrées au moins deux fois.

4. Tous les exemples sont extraits du sous-corpus québécois du corpus CEM. La transcription des paroles suit les règles orthographiques (hors ponctuation). Conventions supplémentaires: pauses (+, ++); éléments paralinguistiques entre {...}; segments avec accent d'emphase en majuscules; segments peu audibles et doutes de transcription entre parenthèses; # pour les segments inaudibles; [pour un chevauchement de parole]; pour un allongement syllabique. La mise en exergue des observables est rendue par des caractères italiques (balises, voir section « Cartographies de l'espace discursif: autour du *Dit* ») et gras (expressions, voir section « Ressources, constructions et expressions figées »).

exemple, comme ici, à l'interlocutaire. Nous ferons l'hypothèse que l'emploi d'expressions figées (EF), le cas échéant, contribue à la présentation de soi comme locutaire compétent en ce qu'elles représentent un gain immédiat de mobilité symbolique. L'objectif de notre contribution est d'appréhender l'élaboration en discours, en temps réel, des EF par les locutaires dans leur contexte de mobilité migratoire.

Espaces discursifs

Mobilité

Dans une expérience de mobilité, l'investissement d'une terre « d'accueil » s'accompagne de l'élargissement de l'espace langagier, construit par des ressources symboliques (linguistiques) qui indexicalisent l'espace socio-géographique de référence. Cet investissement d'un espace langagier nouveau s'effectue sur la base de l'espace initial en le restructurant et en s'intriquant à lui.

Les caractéristiques de la *convocation* en discours de l'ensemble des ressources hétérogènes cartographient la mobilité du locutaire dans l'espace socio-discursif. Autrement dit, chaque élaboration en instance témoigne de la manière singulière dont il se représente la complexité de l'espace social et témoigne, par là même, de son positionnement dans cet espace. En tant que sujet énonciateur et acteur social, la mobilité du locutaire se définit par sa capacité à coordonner, dans son discours, les différents ordres relatifs à l'espace social.

Nous concevons en effet la *mobilité discursive* comme une disposition individuelle générale qui n'est pas directement accessible, mais donnée à voir en discours. Elle peut être explicitée en termes de *Dire* et de *Dit* : le *Dire* désigne l'activité d'élaboration par tous les mécanismes de construction dans le processus de formulation séquentiel, telles les marques épilinguistiques ; le *Dit* désigne les structures construites, en tant que résultat produit et socialement représenté, telles les expressions figées. L'approche se distingue donc de celle de la Polyphonie (Ducrot, 1984), où les concepts du Dire et du Dit désignent respectivement les parties métalinguistique et propositionnelle du sens d'un énoncé, en retenant l'intérêt de problématiser la réflexivité du locutaire, qui place l'analyse de son discours dans un rapport intersubjectif.

Ressources, constructions et expressions figées

Les ressources langagières sont des constructions, des assemblages de formes et de fonctions de complexité variable. La notion de « langue », problématique par ailleurs en sociolinguistique, se justifie par la relative stabilité d'un ensemble conséquent de constructions générales, sous-déterminées dans la mesure où leur élaboration lexicale en discours est relativement imprévisible.

Parmi les constructions, on compte un certain nombre d'« expressions », constructions qui se caractérisent par une relative stabilité de la forme concrète (lexicale) et une lecture synthétique. Le domaine d'étude des EF, foisonnant et multiforme, inclut un large éventail de constructions, allant des collocations (cooccurrences préférentielles) aux expressions figées au sens strict. Cette variabilité repose sur plusieurs éléments, dont une restriction de la combinatoire, la non-compositionnalité sémantique (un emploi au moins partiellement non littéral), la possible réduction du paradigme de flexion (ex. temps verbaux, détermination) et une intégration articulatoire majeure (fluence), du moins pour les collocations fréquentes (Legallois et Tuttin, 2013; Towell *et al.*, 1996; Chambers, 1998; Wray et Perkins, 2000; Wood, 2004). Ces vecteurs descriptifs montrent en effet que les EF couvrent une gamme étendue de constructions différentes, certaines étant métaphoriques et ne permettant quasiment plus de lecture littérale (*coup de pouce*), d'autres le permettant un peu plus (*prendre/tirer profit*) :

Extrait n° 2

AMR mais il y a quand même *comme on dit* le **coup de POUce** (E15; DetC 783; coup de pouce=1)⁵

AMR euh c'est c'est comme ça par contre il y a des il y a d'autres non qui qui v- qui qui savent *euh comment euh euh* **prendre profit** des des des expériences des d'autrui (E15; RetC 670; tirer/prendre profit=2/2)

Typiquement, une expression polylexicale représente une seule unité conceptuelle, globale, semblable à celle d'un « mot »; ex. *coup de pouce* serait équivalent à *piston*, *soutien* et *prendre profit* à *profiter*, les deux variant dans le degré de littéralité.

5. Le référencement des extraits mentionne : 1) le numéro d'entretien dans la collection CEM; 2) la ou les référence(s) d'ouvrage qui recensent l'expression et la page; le nombre d'occurrences dans le corpus au format token/nombre de locutaires distincts). La méthodologie est présentée plus en détail dans la section intitulée « Le relevé des expressions ».

La mobilité discursive des locutaires non-natifs

Les EF chez les locutaires en situation de mobilité

Le locutaire, dans le cadre d'une expérience de mobilité, s'approprie un nouvel espace discursif, structuré par des normes différentielles, dont il doit progressivement établir les « coordonnées ». Ce processus comprend l'expérimentation de nouvelles constructions, basée sur la représentation de leurs contraintes structurelles et leur socio-indexicalité, ainsi que la ré-indexicalisation des ressources disponibles. Il ne s'agit donc ni de substituer des ressources à d'autres, ni de simplement ajouter des ressources à l'inventaire, mais de restructurer l'ensemble des relations d'affinité entre ressources : ex., la construction *faire ses études* sera mise en relation avec *être aux études*, construction relevant plus spécifiquement de l'espace québécois. Or, les EF sont des constructions particulières en ce qu'elles présentent des contraintes à la fois socio-indexicales et structurelles particulières.

Nous posons que 1) **l'indexicalité de l'EF donne à voir l'inscription du locutaire dans l'espace socio-discursif**, le cas échéant, travaillée par lui comme véritable ressource symbolique et 2) que **les contraintes des EF sont congruentes avec leur état de figement**, plus ou moins accentué, qui oblige le locutaire à mémoriser l'EF comme figure invariable.

- 1) La relative invariabilité des EF leur confère un caractère de *gestalt*, c'est-à-dire une indexicalité particulière qui permet « à un membre de la communauté langagière de la reconnaître et de la réutiliser » (Schmale, 2013, p. 41). Ainsi, l'emploi des EF contribue à la présentation de soi comme membre compétent de la communauté discursive, à double titre, en attestant l'expérience de l'espace discursif et en indexicalisant des pratiques comme valeurs partagées. L'émblématique des EF va de la simple intégration comme collocation non-marquée à la reconnaissance normative de certaines expressions comme véritable patrimoine culturel de la langue (Burger, 1998), qui, à l'instar des proverbes, véhiculent des valeurs morales ou sociales. Les EF faiblement indexicales sont les plus fréquentes : *laisser tomber* est une EF relativement courante – attestée dix fois dans notre corpus d'étude chez cinq locutaires distincts – en comparaison à *comme un poisson dans l'eau*, qui n'a qu'une seule occurrence :

Extrait n° 3

ADL c'est-à-dire je peux me fondre là-dedans et ce n'est pas je peux me sentir
comme un poisson dans l'eau (E12; AetM 337; poisson=1)

Dans une perspective d'observation des traces de la mobilité discursive, conçue comme cartographie *sociale* des ressources langagières, notre étude se limitera aux EF à valeur emblématique forte, déterminées par un effet de reconnaissance de l'EF comme figure au sein de la communauté (voir section intitulée « Le relevé des expressions »). En cela, elles présentent une plus-value pour la face du locutaire.

- 2) La projection en discours d'une EF est soumise à des contraintes fortes, plus fortes que celles des constructions abstraites au sens de la Grammaire des Constructions⁶. Si la mémorisation va de pair avec l'expérience en discours, l'acquisition des EF est entravée par leur faible fréquence d'emploi. Ces contraintes s'appliquent à tous les locuteurs, mais la contrainte est exacerbée pour les locuteurs non natifs, dont l'input est réduit. En ce sens, il est plausible que les EF témoignent fréquemment, chez les non natifs, d'un déficit d'idiomaticité (Ellis *et al.*, 2008; Granger, 1998; Howarth, 1998; Pawley et Syder, 1983).

Nous proposons d'explicitier ce déficit d'idiomaticité, lorsqu'il existe, comme créativité langagière *stricto sensu*: en cours d'élaboration discursive, la recherche de l'élément – lexical, grammatical, syntagmatique – juste constitue un bricolage au sens de Lévi-Strauss, qui mobilise l'ensemble de la mémoire discursive. Or l'hétérogénéité de la mémoire discursive des locutaires est accentuée ici, puisque les ressources de l'espace d'origine ne coïncident que très partiellement avec l'expérience de l'espace discursif d'accueil, et que le processus de restructuration est intense. Il est raisonnable d'admettre que cette créativité constructrice exploite (aussi) les ressources de la L1 (Nesselhauf, 2003), dont le vecteur principal dans notre corpus est l'arabe. Dans une approche inspirée de la Grammaire de Construction, nous dirons que la projection en discours d'une EF mobilise un nuage de constructions, dont potentiellement d'autres EF, et potentiellement relatives à d'autres langues. Autrement dit, la représentation des constructions dans la mémoire

6. Déterminées par des contraintes catégorielles/grammaticales plus que lexicales dont l'élaboration laisse davantage de marge de manœuvre.

du locutaire, sous forme de réseau, transcende les frontières des langues : face à la distinction externe en ce qui concerne les « langues », la mémoire discursive est certes indexicalisée, mais dans une structuration unique.

Sous l'angle de la mobilité discursive, nous étudierons leur élaboration précise dans leur contexte, en faisant l'hypothèse que **cette créativité donne à voir les tensions pesant sur la restructuration en cours de la mémoire discursive**. La résonance d'une autre construction, figée ou non, dans la projection de l'EF observée⁷ ne pourra être abordée qu'en guise de perspective pour des études futures, car des prérequis méthodologiques de cette question restent non résolus à ce jour.

Par les enjeux socio-discursifs qu'elles représentent, les EF constituent une cible pour l'activité réflexive dans le processus de construction du discours.

L'activité épilinguistique

Dans le foisonnement des conceptualisations et dénominations (par exemple, Culioli, 1991 ; Gombert, 1990 ; Canut, 1998), nous retiendrons l'activité épilinguistique comme indice de la réflexivité du locutaire en train d'élaborer son discours. Prise comme commentaire de l'investissement de l'espace discursif, cette activité met en exergue l'intrication de sa pluralité. Le processus de cartographie en discours fait émerger des marques spécifiques qui donnent à voir ce travail de circulation entre plusieurs espaces géographiques et reconfigure la mémoire discursive du locutaire en situation de mobilité.

Les indices de cette réflexivité sont des marques de différents types. Outre les balises méta- ou épilinguistiques, marques lexicales ou expressions linguistiques plus complexes, on observe de nombreuses traces formelles d'un « travail d'ajustement » (Authier-Revuz, 1995, p. 175), comme les reformulations, reprises ou marques d'hésitation. Les reformulations, particulièrement concernées par notre propos, reposent sur un schéma de succession séquentielle d'un segment source (qu'on appellera *reformulé*) par un segment paraphrastique (*reformulant*), les deux pouvant être formellement associés par un marqueur de reformulation, médian ou final (Gülich et Kotschi, 1987).

Les contraintes accentuées des EF augmentent en effet le coût de traitement et conditionnent des délais, sous la forme de phénomènes de disfluente.

7. Cette perspective est traditionnellement décrite comme une « contamination » (suivant Paul, 1870 et Bergström, 1906).

L'ensemble de ces marques rend compte de «la dimension foncièrement dialogique du discours, qui doit se frayer un chemin, négocier à travers un espace saturé par les mots et les énoncés autres» (Maingueneau, 2009, p. 87) en tant que manifestations des mouvements de modalisation, de distanciation ou de ratification du travail langagier en cours.

En manifestant la réflexion sur le travail effectué à propos du *Dit* (la formule/expression) ou du *Dire* (le processus de formulation), **les marques épilinguistiques congruentes aux EF témoignent du travail de cartographie de l'espace discursif** (francophone) et assurent l'indexicalisation des ressources en conformité avec les espaces du locutaire, tout en rendant explicite l'attention qu'il y accorde. En effet, les EF sont des constructions qui «ne vont pas de soi»: un manquement aux contraintes rend(ra)it indue l'argumentation déployée pour légitimer le projet de mobilité.

Le relevé des expressions

Différentes approches sont utilisées pour la mise à jour des EF dans les corpus:

- 1) L'approche textométrique (*corpus driven*, déterminée par le corpus) consiste à mettre à jour des fréquences de cooccurrence, c'est-à-dire des collocations candidates EF; une telle approche ne met en exergue que les cooccurrences fréquentes; or, les EF, telles que nous les appréhendons, ont une fréquence très faible. Lorsqu'elles ne se démarquent pas formellement (par exemple, par une réduction de paradigme grammatical), ces expressions sont difficiles à distinguer parmi les constructions abstraites.
- 2) Le recours aux dictionnaires d'EF par requêtage automatique permet d'établir la présence dans le corpus des expressions répertoriées. Cette méthode reste encore difficile à mettre en œuvre pour les corpus oraux, dont les caractéristiques (de la transcription) diminuent les possibilités des traitements automatiques. Par ailleurs, dans un contexte de plurilinguisme, le requêtage devrait idéalement s'appuyer sur autant de dictionnaires qu'il y a de langues en jeu pour le locutaire.
- 3) La découverte «manuelle» des EF, coûteuse en temps, revêt un caractère principalement éthique. Dans cette démarche, l'effet de reconnaissance

expérimenté par l'auditeur est une reformulation de la fréquence de token à l'échelle d'un grand corpus. La démarche du relevé manuel a pour avantage enfin de repérer de nombreuses expressions qui auraient échappé à tout requêtage automatique en raison de leurs caractéristiques inattendues (puisqu'elles ne respectent que partiellement le figement) et qui en constituent des « variantes ». Mais s'il est vrai que les caractéristiques des EF permettent aux locutaires de les reconnaître, la reconnaissance est ici conditionnée par la subjectivité de l'observateur.

Nous avons objectivé cette approche par le recours aux dictionnaires spécialisés des expressions : nous avons vérifié systématiquement l'existence des EF relevées dans le corpus par leur mention dans des ouvrages de référence relatifs aux différents espaces discursifs. Ne seront prises en considération que les expressions polylexicales répertoriées dans l'un des ouvrages de référence. Le fait même d'être recensé dans un tel ouvrage constitue un indicateur de la valeur emblématique de l'EF. À titre informatif, le nombre d'occurrences dans le corpus de l'élément pivot de l'expression figure entre parenthèses après la référence textuelle de l'extrait présenté, précédé, le cas échéant, du « pointage » de la référence. Nous avons choisi trois ouvrages grand public consacrés au recensement des expressions figées du français : Ashraf et Miannay, 1995 (référéncé AetM), Rey et Chantreau, 2015 (RetC) et Duneton et Claval, 1990 (DetC). Le cadrage définitionnel est hétérogène, mais recense des expressions polylexicales, à l'exclusion des lexies métaphoriques isolées, des collocations usuelles et des proverbes⁸. Conformément à leur projet éditorial, les dictionnaires recensent en effet le plus souvent les expressions à forte teneur métaphorique, alors que les cooccurrences avec une socio-indexicalité moindre relèvent, le cas échéant, des dictionnaires généraux.

Ces trois ouvrages ne sont pas explicitement centrés sur un espace francophone spécifique. Ce caractère géographiquement non marqué reste cependant implicitement centré sur l'usage franco-français. Ponctuellement, nous avons consulté en complément le *Trésor de la Langue française informatisé* (TLFi) accessible en ligne depuis le site du Centre national de ressources textuelles et lexicales (CNRTL). Pour retracer l'indexicalisation de l'espace québécois, nous avons eu recours au dictionnaire USITO ; au dictionnaire anglais de Cambridge (DC) ; au Grand dictionnaire terminologique

8. À l'exception de Duneton et Claval, qui les recensent (mais que nous excluons de l'étude).

(GDT) et à la Banque de dépannage linguistique (BDP) de l'Office québécois de la langue française.

La prise en compte des espaces discursifs d'origine n'a été que très sommaire⁹. Pour confronter les expressions relevées dans le corpus à l'étude à celles potentiellement représentées dans la mémoire discursive arabe, nous avons dépouillé le *Dictionnaire des expressions idiomatiques arabes* de Moussaoui (2015) (M), puis, en inversant la perspective, cherché d'éventuelles correspondances dans le corpus à l'étude en découvrant des expressions arabes équivalentes.

L'élaboration des EF en discours

Cartographies de l'espace discursif: autour du Dit

Les balises épilinguistiques (BEL) au sens habituel, de nature lexicale ou syntagmatique, exercent de nombreuses fonctions pragmatiques et métadiscursives. En guise de commentaire d'une EF, elles témoignent du travail de cartographie de l'espace discursif.

Parmi les fonctions les plus fréquentes des BEL, on trouve la **fonction emphatique**, pour souligner, renforcer ou insister sur un argument discursif, représenté par l'EF (*justement, il y a quand même une différence, et même, comme ça, quand même, sacré*).

Extrait n° 4

AMR donc euh c'est c'est *justement* euh euh **la pierre angulaire** de de de toute intégration (E15; RetC 621; pierre angulaire=1)

La fonction emphatique s'ajoute aux différents procédés d'augmentation du nombre ou de l'intensité, notamment pour amplifier une élaboration en liste, qu'il s'agisse de reformulations ou d'exemples corroborant un même argument (*et cetera*).

Une autre fonction fréquente en est la **modalisation**, lorsque la balise se rapporte à la négociation interactionnelle (*bien sûr, il faut dire que, voilà*):

9. Notre compétence et les outils à disposition ne nous ont pas permis de prendre en compte le kabyle.

Extrait n° 5

AMR TOUT système il y a un ANTisystème qui va parallèlement donc il y a **le noir et le blanc** *bien sûr* que euh il y a il y a tout le temps un réseau euh il y a il (E15; AetM 290, RetC 77, DetC 386, 464, 629; noir/blanc=2/2)¹⁰

Extrait n° 6

SAN on est colonisé **jusqu'à l'os** {rire} (E26; AetM 302; jusqu'à l'os=1)

Le dernier exemple montre comment un rire peut modaliser l'énonciation : en accentuant la représentation, la qualité métaphorique de la figure de l'EF accentue un point de vue qui pourrait être perçu comme quelque peu catégorique par l'interlocutrice, accentuation que le rire permet de modaliser en l'atténuant.

Lorsque l'EF exprime une catégorie sociale, elle semble s'adresser directement au locataire en tant qu'acteur social, et le balisage semble contribuer à la **construction d'une image de soi** valorisante :

Extrait n° 7

SAN honnêtement ça me pose des Vrais vrais problèmes d'éthique un peu aussi parce que je me dis est-ce que c'est une forme d'abandon euh que de ne pas **mener le combat** *entre guillemets* (E26; RetC 199; combat=1)

Lorsque l'EF consiste à convoquer un référent socioculturel, comme dans l'exemple suivant, par la citation d'une référence littéraire, le locataire élabore l'image de soi en donnant à voir sa compétence dans une culture autre que celle de son origine et argumente implicitement la légitimité de son projet de mobilité :

Extrait n° 8

AMR la conception qu'on qu'on a euh de l'autre euh en France + *c'est tout le temps* **l'autre c'est l'enfer** *comme disait Sartre*
ENQ {rire} (E15; RetC 317; enfer=2/2)

Bien des fois, la portée de la balise est ambivalente, en ce qu'il est difficile d'établir si elle se rapporte au contenu de ce qui est dit ou au choix de la formulation. C'est notamment le cas avec les balises, fréquentes, comme *un peu* ou *entre guillemets* : parfois les locutaires peuvent baliser l'expression

10. Dans les extraits de corpus, les EF sont mises en exergue par des caractères gras et les marques épilinguistiques par des caractères italiques.

convoquée comme telle et faire explicitement référence à un espace discursif particulier. Dans l'exemple suivant, si l'élaboration d'une EF qui indexicalise l'espace québécois, *tricoté serré*, fait apparaître de facto le locutaire comme membre compétent de la communauté discursive correspondante, le caractère marqué de cet espace est potentiellement excluant pour l'interlocutrice française, ce qui peut favoriser un balisage :

Extrait n° 9

TRK mais je sais qu'ils ont plus de rejet par rapport à l'étranger euh parce qu'ils sont différents puis ils sont contre le multiculturalisme ils sont contre donc c'est ça donc ils vivent *un peu dans la comment dirais-je dans la comment euh en tout cas dans le tricoté serré* ou euh on ne peut pas nécessairement euh c'est chez nous tu sais (E12 ; BDL ; tricoté serré=1)

Dans le prolongement de la fonction modalisatrice, la balise fait acte de réparation (anticipée) lorsque l'énonciation porte potentiellement atteinte à la face positive de l'interlocutrice, qui est associée à l'espace (universitaire) français :

Extrait n° 10

AMR en tant que # en tant que moDÈLE euh je pense que que les études supérieures en France *a un peu un peu euh perdu de son de de comment de son ÉCLAT* (E12 ; TLFi ; éclat=1)

L'EF synthétise en effet différentes fonctions, qui peuvent apparaître séparément en discours : la fonction sémantico-référentielle, rendue dans nos exemples par un rapport métaphorique ; la fonction identitaire, présente dans l'indexicalité patrimoniale de l'EF ; la fonction discursive, actualisée par la qualité des formulations. Nous pensons que cette complexité socio-indexicale est caractéristique de la construction discursive de soi dans un espace discursif où les exigences sont multiples, en référant à des temporalités diverses (l'argument, la temporalité de l'entretien, le projet de mobilité, l'identité...).

Le Dire entre fluence et indexicalité des espaces

On note que de nombreuses balises ont pour fonction de commenter la forme même de la formulation en cours, en explicitant l'idée que l'expression **ne va pas de soi** :

Extrait n° 11

MLK ah oui ce jour-là + j'étais **au bord du gouffre** euh enfin + *une façon de parler* mais + tu sais c'est – ça fait **remonter des souvenirs à la surface** et là tu te dis + ah oui mais finalement euh ma vie elle a pas toujours été un : *un fleuve un long fleuve tranquille* [et il y a eu quand même des euh

ENQ [oui

MLK il y a eu *quand même des hauts et des bas* il y a eu *des des- des trucs assez euh :*

ENQ ben oui [quand même quel parcours

MLK [donc euh : donc oui oui non mais c'est bien aussi + c'est- c'est + faire cet exercice ça permet aussi de + de remettre un peu certaines choses dans : dans leur contexte + puis euh : justement ça évite de + de **s'endormir sur ses acquis** (E36; CNRTL)

On peut noter que cet extrait relativement court fait état d'un nombre remarquable d'EF. Bien que nous n'ayons pas pu faire de relevé systématique pour des questions méthodologiques dans lesquelles s'insère la réflexion ci-présente, notre expérience du corpus nous fait penser que la condition des locutaires de ce corpus, hautement compétents en français et arguant leur mobilité dans l'espace francophone, favorise chez la plupart d'entre eux l'élaboration d'EF emblématiques. Reste que nous avons choisi cet extrait pour sa force illustrative, non seulement de la quantité, mais aussi de la diversité des EF : la métaphore aquatique richement élaborée, la métaphore spatiale ordinaire et la dernière hautement emblématique.

Dans le processus d'élaboration du discours en temps réel, la projection locale d'une EF induit donc une **vigilance** supplémentaire pour en contrôler le plus possible l'élaboration, du fait de la socio-indexicalité des EF. La manière de les élaborer, en instance, rend manifeste la mobilité du *Dire* singulier du locutaire. Or la vigilance ne se limite pas au cas où elle est attestée par une balise verbale (*une façon de parler*) : face à la profusion de marques d'activité épilinguistique constatée, et aux particularités observées dans l'élaboration même des EF, force est de constater que leur intégration fluide à la construction, superordonnée, est plutôt rare.

La gestion de l'hétérogénéité de l'espace discursif et des inscriptions différentielles (en l'occurrence, celle du locutaire témoin et celle de l'enquêtrice) abordée plus haut est souvent source de reformulation :

Extrait n° 12

AMR ça aide énormément ça ça nous aide pas à oublier mais ça nous aide à à gérer ouais ouais ouais + ça nous donne **du punch** euh *de la force* et donc euh du coup voilà (E42; CNRTL: avoir du punch; donner du punch=1)

La reformulation de *du punch* par *de la force* fait apparaître l'EF comme non satisfaisante, soit qu'elle ne décrive pas précisément le contenu visé, soit qu'elle soit indexicale d'un espace moins accessible (aux yeux du locataire) pour l'enquêtrice française. L'acte de reformulation ici serait orienté sur l'interlocutrice, comme acte de politesse, mais par lequel le locataire revendiquerait une mobilité plus étendue que celle de son interlocutrice. L'intention de thématiser l'hétérogénéité de l'espace discursif ressort de façon plus nette encore dans les cas où la construction (EF) relevant de l'espace québécois (dans l'exemple suivant, *deux de pique*) succède à la construction plutôt hexagonale (*vrais bosseurs*):

Extrait n° 13

SAN c'est les enfants sont des vrais bosseurs là *c'est pas des* + **c'est pas des deux de pique** puis euh + donc c'est ça donc euh il y a quand même beaucoup euh mais en même temps parallèlement (E26; USITO; pique=2)

Enfin, certains cas témoignent de l'enjeu même à « placer » des EF, lorsque reformulant et reformulé sont des EF, métaphoriques:

Extrait n° 14

TRK ah c'était pas nécessairement ça mais euh un mauvais calcul j'ai fini ma maîtrise en 2005 je croyais oh **les jobs vont pleuvoir** ouais pas mal **avec un tapis rouge** en même temps j'ai connu quelqu'un (E12; AetM 332; pleuvoir des X=1; AetM 387; tapis rouge=1)

Malgré l'élaboration inexacte (*dérrouler le tapis rouge* vs *avec un tapis rouge*), la force figurative est telle que, non seulement le sens reste intact, mais le caractère réduit de l'EF reformulante ne se remarque qu'à peine à l'écoute.

D'autres reformulations semblent témoigner plutôt du travail de formulation au sens strict. Dans l'exemple suivant, après une première BEL modalisatrice (*comme si*), l'EF est amorcée avec une réouverture du paradigme de détermination (*le bâton/des bâtons*), poursuivie d'un piétinement (répétition de *dans*) avant d'arriver à complétion puis d'être reformulé avec une construction ouverte, plus littérale:

Extrait n° 15

AMR ce qui ce qui ajoutait à la à la problématique donc c'était POUR moi *comme si* on **mettait le bâton dans dans les roues** donc euh *il y avait plus moyen de de de bouger* (E15; RetC 61; mettre (des) bâtons dans les roues=1)

La recherche de la formulation juste conduit donc au retardement de la complétion des projections, qui peut aussi prendre la forme de disfluences – souvent mineures, jusqu'à n'être rendues visibles que par la transcription – mais qui témoignent des limites d'intégration de ces constructions dans le flux du discours. Si la césure prosodique entre deux constructions n'est pas remarquable en soi, ces ralentissements au milieu de l'EF provoquent une suspension temporaire de la projection de l'EF, projection réputée particulièrement forte et habituellement articulée en un seul geste :

Extrait n° 16

AMR en tant que x- en tant que moDÈLE euh je pense que que les études supérieures en France *a un peu un peu euh* **perdu de son de de comment de son ÉCLAT** (E12; TLFi=éclat; éclat=1)

Les répétitions d'EF à courte distance peuvent, à l'inverse, renforcer leur caractère marqué et la plus-value escomptée :

Extrait n° 17

SAN il peut y avoir + *euh tu sais cette espèce de je sais pas comment on peut l'appeler là mais cette espèce de genre regarde on* **jette l'éponge** *là tu vois* et euh + et en fait c'est pour ça que en ce moment j'ai très très peur que il y ait beaucoup beaucoup de personnes qui **jettent l'éponge** et qui ne ##suite (E26; RetC 326; jeter l'éponge=2)

Le travail de formulation est manifesté ici par différentes traces qui indiquent le caractère marqué, littéralement, par les indices de la gestion interlocutive en amont et en aval (*tu sais, tu vois*), une phase d'approximation complexe (*espèce de*), un commentaire métadiscursif (*je sais pas comment on peut l'appeler*). L'élaboration de l'EF elle-même est fluide : son intégration dans un discours animé (*regarde on jette l'éponge*) en soutient l'idiomaticité et la répétition est élaborée dans une autre unité de construction du même tour (*beaucoup de personnes qui jettent l'éponge*). La répétition apparaît davantage comme ressource d'emphasis que comme ralentissement de la progression thématique. Au second plan, l'élaboration réitérée d'une EF en contexte « rhétorique » signe la mobilité discursive singulière du locuteur.

Résonances

Il n'aura pas échappé aux observateurs que certains extraits produits ci-dessus sonnent étranges/étrangers. Dès lors, comment saisir cette étrangeté en termes de mobilité discursive ?

D'abord, certains **décrochages énonciatifs** au fil de l'interaction rendent explicite la dimension intersubjective de la réflexivité à propos des EF :

Extrait n° 18

TRK bah ça te permet de t'entraîner euh donc ton énergie n'est pas vraiment mais en même temps mh *faut pas* (se, ce) justement **il faut pas dormir** euh
 ENQ (*sur ses deux oreilles*) *sur ses lauriers* {rire}
 TRK **sur ses lauriers** et de
 ENQ *sur ses deux oreilles c'est pas du tout la même chose* {rire}
 TRK ah oui OK je savais pas ça (E40 ; AetM 244 ; (s'en)dormir sur ses lauriers= 2/2)

L'enquêtrice se montre investie dans la coconstruction en proposant deux complétions possibles pour la projection *il faut pas dormir* suspendue après *euh*. Le locuteur reprend la seconde tout en reconnaissant son expertise limitée des constructions proposées. La césure induite par la suspension initiale de la projection conduit dans la complétion par l'enquêtrice à une fusion qui marque formellement un défigement : *dormir* / *sur ses lauriers* puise dans *dormir sur ses deux oreilles* et *s'endormir sur ses lauriers*. Les caractéristiques communes des deux constructions en amont et en aval de la césure favorisent l'intégration malgré les différences. Ce n'est pas tant la singularité de l'expression émergente qu'il convient de souligner, mais le contexte de césure, qui résulte d'une lecture analytique de l'expression. Celle-ci caractérise la mobilité du locuteur, dont le projet d'élaborer une construction métaphorique avec le noyau *dormir* peine à aboutir, puisque les EF proposées par la suite ne sont pas disponibles en temps « réel ».

Cet exemple articule trois aspects de l'élaboration des espaces discursifs : **1) la compétence différentielle des EF et l'appropriation de leur indexicalité ; 2) la résonance d'une pluralité de constructions dans celle qui se trouve effectivement élaborée (fusion) ; 3) les conséquences pour le processus d'élaboration.**

1) De nombreux exemples comportent une « réouverture » de position syntaxique qui devrait être bloquée par la lecture globale :

Extrait n° 19

AMR euh un supérieur peut **mettre sa main dans la pâte** avec euh un subalterne {rire} *si euh j'emploie l'expression {rire} justement* (E15; RetC 495; main/pâte=1)

Cette élaboration, qui **déroge aux contraintes habituelles** de l'EF (*mettre *sa/la main *dans/à la pâte*) comprend une activité épilinguistique qui témoigne de la sensibilité au risque du travail en cours, et signale le doute du locuteur sur sa capacité à utiliser l'expression de façon appropriée. Dans l'exemple suivant, la modalisation (*un peu*), l'amorce (*ils ne sont pas*) sont suivies d'une balise épilinguistique double portant sur l'énonciation avec un premier segment en anglais. L'expression est néanmoins fluide et on a l'impression que le locuteur cherche à compenser le risque pris à l'emploi de l'EF par un balisage qui accentue l'hétérogénéité de l'espace discursif et qui le positionne ostensiblement comme acteur agile :

Extrait n° 20

TRK je n'ai pas envie de ça donc l'image que j'ai des Québécois *ils ne sont pas* ils ont *un peu* **jeté le bébé avec l'eau de bain** ou *whatever ça se dit* (E12; AetM 34; jeter le bébé avec l'eau (du) bain=2/2)

Dans l'exemple suivant, le locuteur élabore une EF en reprise de l'enquêtrice :

Extrait n° 21

ENQ par euh justement ce qu'on entend **de bouche à oreille** c'est vous pensiez à quoi quand on parle de propagande ou sur le site du gouvernement ou
 AMR non c'est c'est c'est c'est le discours officiel
 ENQ discours officiel qu'on va retrouver euh
 AMR non non *euh mais* **la bouche à oreille** *euh je je je* je n'y prête pas tellement attention mais c'est plutôt le discours officiel
 ENQ plutôt discours officiels mh d'accord
 AMR euh discours officiel qui euh qui m'intéresse est-ce que la voix que j'entends euh euh vers les canaux qui sont reconnus euh au Canada concorde euh réellement avec euh
 ENQ la réalité quand on y est
 AMR la réalité mh quand on y est
 ENQ mh mh d'accord
 AMR c'est ça euh qui euh qui m'interpellent le plus mais **(le) bouche à oreille** euh il y a des histoires euh donc au pluriel (E15; AetM 47; bouche à oreille=3/2)

La réutilisation, similaire à la **répétition** décrite plus haut, met l'accent sur l'EF dans l'espace du discours et souligne son indexicalité. La première réutilisation fait l'objet d'un travail de formulation visible/audible : privilégiant le genre grammatical de la lexie *bouche*¹¹ au lieu de celui de l'EF (masculin), l'élaboration est analytique et entourée, qui plus est, de disfluences. L'intégration discursive laisse penser que le locuteur reconnaît l'expression comme telle et qu'elle est sémantiquement représentée. Ainsi, le processus illustre la façon dont le locuteur cartographie les ressources langagières en temps réel, en identifiant l'EF dans le flux pour aussitôt amorcer un travail d'appropriation.

La présence d'une **reformulation impliquant l'EF** en reformulé ou en reformulant est l'un des révélateurs du fait que son emploi ne va pas de soi. La reformulation d'une EF représente donc un doute, soit sur son élaboration, soit sur la capacité à se faire comprendre par l'interlocutrice française :

Extrait n° 22

AMR il y a quand même une catégorie **il faut pas se voiler la face** hein donc euh euh *tiens je vais lâcher le mot* la discrimination ça existe aussi ici hein c'est *il faut pas se* **il faut pas se mentir** donc euh euh bien sûr qu'il y a quand même une euh il y a quand même euh une une différence hein euh (E15; DetC 499, 766, AetM 410; se voiler la face=2/1)

On note ici la manière dont la mobilité discursive est mise au service de l'argumentation. Le locuteur fait montre d'une capacité à asseoir son propos – la critique envers le Québec – en l'atténuant précautionneusement par plusieurs BEL (*quand même, bien sûr, tiens je vais lâcher le mot*) qui inscrivent en outre son discours dans plusieurs espaces : le reformulé *il faut pas se voiler la face* et le reformulant *il faut pas se mentir* possèdent une socio-indexicalité différente. Tandis que le second relève de l'espace discursif oral courant du langage jeune, le premier relève du patrimoine commun du « bon français », catégorie citée par ce locuteur dans un autre contexte :

Extrait n° 23

AMR est-ce que je dois euh je dois enseigner le français français + ce qu'on appelle standard hein + avec l'accent euh français euh ++ entre guillemets le BON français comme vous dites

11. Le genre est féminin dans la première reformulation, alors que la réalisation est floue pour la seconde.

2) Les choix de navigation entre les ressources disponibles font émerger des créations singulières :

Extrait n° 24

AMR j'ai fait ma demande en tant que résident permanent
 ENQ donc dans un deuxième temps
 AMR sachant euh
 ENQ ouais
 AMR dans un deuxième temps et parce que je savais que pour faire des études
 ++ en tant qu'étudiant ÉTRANger ça coûtait le blanc des yeux
 ENQ {rire}
 AMR {rire} donc euh
 ENQ nous on dit **les yeux de la tête** + *c'est excellent le blanc des yeux* {rire}
 AMR ouais ouais ouais
 ENQ ah c'est chouette ouais
 AMR le blanc des yeux ouais c'est une formule personnelle {rire}
 ENQ {rire} d'accord (E12, DetC 714; les yeux de la tête=1)

Le locuteur persiste et assume le *blanc des yeux* comme « formule personnelle¹² », ratifiée seulement en conclusion par l'enquêtrice. À l'occasion d'un échange qui émerge dans une relation plutôt verticale (au vu des statuts, origines, et rôles interactionnels respectifs), le locuteur revendique sa mobilité singulière dans l'espace discursif par une créativité différentielle assumée...

Au-delà des québécismes lexicaux simples, qui ne font pas l'objet de cette étude, on constate la présence en somme assez discrète de l'**espace québécois** dans les EF. S'il est impossible de raisonner l'absence d'un phénomène dans le corpus, celle-ci peut en effet épouser deux logiques contraires : la première serait liée au contexte interactionnel où le locataire se trouve face à une enquêtrice catégorisée comme appartenant à l'espace français, qui plus est, dans une relation verticale, qui induit des accommodations convergentes. Le cas échéant, la capacité à la neutralisation des emblèmes différentiels au profit de constructions non marquées signifie une mobilité certaine. La seconde serait, au contraire, liée à une expérience encore trop limitée de l'espace d'accueil pour amplifier l'indexicalisation. Dans ce cas, l'absence d'EF ne serait pas un comportement accommodant, mais serait révélatrice d'une appropriation ténue du patrimoine linguistique, et à interpréter plutôt comme mobilité limitée.

12. L'expression métaphorique *le blanc des yeux* n'est pas référencée dans le dictionnaire de Moussaoui.

Les discours comportent toutefois des traces de l'espace discursif à travers la présence d'EF, dont un certain nombre de contextes d'alternance interphrastique avec l'anglais¹³:

Extrait n° 25

SAN puis de toute façon il avait un financement malien et cetera donc Paul c'était un un gars très euh +

ENQ ah oui

SAN **by the book** *tu vois* il faisait des affaires (E38; DC; by the book=1)

L'exemple suivant contient l'EF *ouvrir les horizons*:

Extrait n° 26

AMR et par la suite je me retrouve en train d'étudier euh à l'Université de Montréal euh pour justement *euh ou – ouvrir les horizons euh donc pour euh plus de chance* (E15; RetC 437; USITO; ouvrir les horizons=1)

Outre son existence dans l'espace francophone mondial, elle peut être rapprochée de

واسع الأفق

vaste / DEF-horizon-OBL

« qui a les horizons vastes »

érudit (arabe fusha; M 316)

houl moukhek

« ouvrir notre cerveau »

se cultiver (darija)

Certaines expressions existent de façon plus ou moins symétrique dans plusieurs langues potentiellement concernées. Ces cas de syncrétisme limitent l'indexicalité au caractère métaphorique, valorisable en tant que tel. L'indexicalité spatiale est alors neutralisée.

3) Pour mémoire, les expressions proches sont représentées en contiguïté, toutes « langues » relevant de l'expérience du locutaire confondues. L'activation en discours de plusieurs constructions similaires constitue la base de phénomènes comme le lapsus ou les élaborations en liste. Cependant, les constructions (EF) similaires ne sont pas seulement convoquées en dis-

13. De par la convocation d'une alternance codique, cet exemple est à distinguer de celui incluant *donner du punch* dans l'exemple 12, expression figée ordinaire dans l'espace francophone dont l'emprunt de *punch* à l'anglais n'est plus identifié.

cours alternativement, elles se trouvent parfois, aussi, dans une intégration séquentielle :

Extrait n° 27

AMR un allophone qui apprend présentement c'est pas un allophone qui apprend présentement mais lorsqu'il apprend cette langue + lorsqu'il est au Québec + dans la rue il entend tout le temps présentement {accent différentiel}

ENQ ben oui +

AMR et ben moi **je suis dans un dilemme entre le l'enclume et le marteau** +

ENQ {rire}

AMR alors euh : et et moi je pense ici euh il y a tout un discours à faire (E15 ; RetC 509 ; entre l'enclume et le marteau=1)

Si un *dilemme* est un choix impossible entre deux options et *être coincé entre l'enclume et le marteau* décrit une situation inconfortable qui prive de la capacité d'agir, le sémantisme de l'immobilité contrainte est présent dans les deux éléments. Élaborée avec une forte intégration prosodique et syntaxique, la séquence *être dans un dilemme entre l'enclume et le marteau*, présente un dilemme, pour ainsi dire, surdéterminé par l'EF. Ici aussi, l'EF correspondante existe en arabe (fusha) :

بَيْنَ الْمَطْرَقَةِ وَالسُّنْدَانِ

entre / DET marteau-OBJ / CJ DET enclume

«entre le marteau et l'enclume»

se trouver entre deux situations difficiles (*sic*, M 67)

On appréciera la traduction libre proposée par l'auteur. Les cas de fusion, celui-ci tout comme celui cité en début de section, montrent que la cartographie crée des résonances entre ressources langagières aux indexicalités différentielles. L'intrication dans la mobilité du locutaire génère des tensions, localement notables. Marqués dans leur hétérogénéité, ces exemples donnent à voir la restructuration de la mémoire discursive, qui est une dynamique particulièrement forte chez les locutaires néo-arrivants, mais en réalité permanente chez toutes ces personnes.

* * *

Dès la prise en main du corpus CEM, la présence récurrente d'EF nous est apparue comme remarquable. Si la catégorisation étique des figures par l'observateur est méthodologiquement délicate, le type de corpus ne permet pas de connaître les catégories émiques correspondantes. Dans ce travail,

nous avons proposé de saisir l'intersubjectivité par le recours aux travaux qui recensent les expressions figées en leur donnant une existence symbolique, et par l'analyse du balisage discursif des expressions par le locutaire. Les balises montrent en effet le travail réflexif du locutaire : il nous importe moins de témoigner d'une fragilité des compétences (c'est-à-dire visant des réalisations « non standard »), que de retenir l'effort consenti par le locutaire pour élaborer en temps réel son discours avec des figures emblématiques. Le locutaire en contexte de mobilité conçoit l'utilisation d'EF comme une plus-value pour son discours. Dès lors, les études à venir devront détailler les ressources sur lesquelles repose la créativité et les modalités de l'identification intersubjective des EF en instance.

Références

- ASHRAF, Mhatab et Denis MIANNAY (1995), *Dictionnaire des expressions idiomatiques françaises*, Paris, Librairie Générale française.
- AUTHIER-REVUZ, Jacqueline (1995), *Ces mots qui ne vont pas de soi : boucles réflexives et non-coïncidence du dire*, Paris, Larousse.
- BERGSTRÖM, Gustaf Adolf (1906), *On blendings of Synonymous or Cognate Expressions in English*, Lund, Hakan Ohlsson.
- BURGER, Harald (1998), *Phraseologie. Eine Einführung am Beispiel des Deutschen*, Berlin, Erich Schmidt.
- CANUT, Cécile (1998), Pour une analyse des productions épilinguistiques, *Cahiers de praxématique*, 31, 69-90.
- CHAMBERS, J. K. (1998), Social Embedding of Changes in Progress, *Journal of English Linguistics*, 26(1), 5-6.
- DUNETON, Claude et Sophie CLAVAL (1990), *Le bouquet des expressions imagées : encyclopédie thématique des locutions figurées de la langue française*, Paris, Seuil.
- DUCROT, Oswald (1984), *Le dire et le dit*, Paris, Éditions de Minuit.
- ELLIS, Rod et al. (2008), The Effects of Focused and Unfocused Written Corrective Feedback in an English as a Foreign Language Context, *System*, 36(3), 353-371.
- GRANGER, Sylviane (1998), *Prefabricated patterns in advanced EFL writing; collocations and formulae*, dans A. P. COWIE (dir.), *Phraseology: Theory, Analysis and Applications*, Oxford, Clarendon Press, 145-160.
- GÜLICH, Elisabeth et Thomas KOTSCHI (1987), Les actes de reformulation dans la consultation « La dame de Caluire », dans P. BANGE (dir.), *L'analyse des interactions verbales. La dame de Caluire, une consultation*, Berne, Peter Lang, 15-81.
- HOWARTH, Peter (1998), Phraseology and Second Language Proficiency, *Applied Linguistics*, 19(1), 24-44.
- LEGALLOIS, Dominique et Agnès TUTTIN (2013), Présentation : Vers une extension du domaine de la phraséologie, *Langages*, 189, 3-25.

- MAINGUENEAU, Dominique (2009), *Les termes clés de l'analyse du discours*, Paris, Seuil.
- NESSELHAUF, Nadja (2003), The Use of Collocations by Advanced Learners of English and Some Implications for Teaching Get Access Arrow, *Applied Linguistics*, 24(2), 223-242.
- PAUL, Hermann (1870), *Principien der Sprachgeschichte*, Niemeyer, Halle.
- PAWLEY, Andrew et F. H. SYDER (1983), Two puzzles for linguistic theory: Nativelike selection and nativelike fluency, *Language and Communication*, 191-227.
- PLOOG, Katja, Anne-Sophie CALINON, et Nathalie THAMIN (2020), *Mobilités. Histoire d'un concept en sociolinguistique*, Paris, L'Harmattan.
- REY, Alain et Sophie CHANTREAU (2015), *Dictionnaire d'expressions et locutions*, Paris, Le Robert.
- SCHMALE, Günter (2013), Qu'est-ce qui est préfabriqué dans la langue? Réflexions au sujet d'une définition élargie de la préformation langagière, *Langages*, 189(1), 27-45.
- TOWELL, Richard *et al.* (1996), The Development of Fluency in Advanced Learners of French, *Applied Linguistics*, 17(1), 84-119.
- WRAY, Alison et Michael R. PERKINS (2000), The functions of formulaic language: An integrated model, *Language et Communication*, 20, 1-28.
- WOOD E. J. (2004), Problem-Based Learning: Exploiting Knowledge of how People Learn to Promote Effective Learning, *Bioscience Education*, 3(1), 1-12.

CHAPITRE 9

Devenir féministe et queer à Paris et à Montréal

Récits de vie dans le corpus CaFé

Julie Abbou et Heather Burnett

Ce chapitre a deux buts principaux. Tout d'abord, présenter le corpus Cartographies linguistiques du féminisme (CaFé), composé d'entretiens sociolinguistiques réalisés à Paris, à Montréal et à Marseille avec des militant.e.s féministes et queers. Par ailleurs, mettre en lumière la façon dont les entretiens de ce corpus peuvent nous renseigner sur les situations politiques, linguistiques et idéologiques des militant.e.s féministes et/ou queers dans ces trois villes. Comme les responsables de ce livre l'ont observé dans leur introduction, les chercheurs et chercheuses variationnistes se sont surtout intéressé.e.s aux schèmes linguistiques dans leurs corpus, laissant de côté l'analyse du contenu des entretiens et ont, ce faisant, laissé de côté de précieuses informations politiques et sociologiques que des analyses discursives pourraient produire. Nous analyserons une étude d'une partie des récits de vie présents dans le corpus CaFé : celle de l'entretien où les locuteurices racontent comment elles, ils et iels, sont devenu.e.s féministes et/ou queers. En nous concentrant sur une comparaison détaillée des entretiens à Paris et à Montréal, nous montrerons que les trajectoires personnelles et militantes des participant.e.s des deux villes sont en partie très semblables. Des différences importantes seront observées cependant dans l'articulation entre le féminisme et le queer. En effet, dans notre corpus, la signification du mot *queer* varie légèrement : dans la métropole française, le queer est plutôt vu comme une orientation politique au sein du féminisme, tandis que, dans la

métropole québécoise, cette dénomination est plus souvent comprise comme associée à une orientation sexuelle ou même une catégorie de genre. Ces différences dans la façon dont les participant.e.s de l'étude comprennent le queer sont apparues dès la constitution du corpus lui-même, notamment en ce qui concerne les genres des locuteurices et donc la représentativité. Dans la double lignée des réflexions méthodologiques élaborées par les études féministes (voir Passerini, 2008, ainsi que Charron et Auclair, 2016 pour une synthèse) qui invitent à toujours réinterroger nos pratiques d'investigation, et des approches par les récits de vie en sciences sociales (Veith, 2004, Costa et dos Santos, 2021, Bernard *et al.*, 2021), nous souhaitons montrer l'importance de l'étude des récits de vie pour l'interprétation des propriétés démographiques d'un corpus variationniste, dans la mesure où ces récits proposent des catégories émiques propres.

Le corpus CaFé

Le corpus Cartographie linguistique du féminisme (CaFé) est un corpus francophone d'entrevues semi-dirigées consacrées au thème de l'engagement féministe et queer. Il a été réalisé dans le contexte du projet européen Formal Models of Social Meaning and Identity Construction through Language (SMIC, 2020-2025, ERC StG no 850539, dirigé par Burnett). Il comprend 102 entrevues avec des personnes se reconnaissant comme engagé.es dans le féminisme, la cause des femmes, le queer ou les luttes pour les sexualités, à Paris (42 entrevues), à Montréal (40 entrevues) et à Marseille (20 entrevues), au cours desquelles sont recueillies les positions des locuteurices sur des questions de genre et de sexualités et les liens qu'ils, elles et iels voient avec le langage. Le corpus est constitué d'un total de 168 heures de parole et de 2 millions de mots, consignés en transcription orthographique enrichie. La durée moyenne d'une entrevue est d'environ 1 h 30. La collecte des données s'est étalée sur presque deux ans (entre janvier 2021 et novembre 2022), la pandémie de COVID-19 ayant ralenti les possibilités d'entrevue, toujours menées en personne.

Le corpus a pour objectif la réalisation de cartographies qualitatives et quantitatives des positionnements idéologiques féministes et queers francophones, dans une dimension comparative. Au-delà des apports sociolinguistiques de ce corpus, dégager de telles cartographies idéologiques – et donc discursives – dans lesquelles se situent les locuteurices permet de donner une image, non exhaustive, mais qualitativement et quantitativement significative, du féminisme contemporain dans les villes concernées, afin de répondre à

quelques interrogations. Quel est l'espace idéologique du féminisme francophone contemporain ? Quelles reconfigurations ont-elles cours ? Comment les féministes et les queers de différents endroits de la francophonie énoncent-ils et elles ces reconfigurations ? Comment ces configurations idéologiques sont-elles articulées par rapport à des usages linguistiques ?

L'originalité sociolinguistique de ce corpus est ainsi de mettre l'accent sur les orientations idéologiques plutôt que sur les caractéristiques sociodémographiques habituellement employées pour étudier la variation linguistique. Ce qui soulève un certain nombre de questions, mais produit aussi, croyons-nous, un certain nombre d'apports. Du point de vue sociolinguistique, il permettra d'étudier les communautés de pratiques linguistiques au sein du féminisme et du queer, et de rechercher des variantes, pour voir si celles-ci sont des marqueurs de positionnement idéologiques ou de rapports sociaux exprimés, ceci dans une démarche similaire à celle que Levon (2010) avait adoptée pour son analyse des positionnements idéologiques des communautés queers israéliennes. Cette approche permet l'étude variationniste de variantes lexicales et grammaticales, difficiles à étudier avec des corpus classiques comme Sankoff-Cedegren (1972), Montréal 84 (Thibault et Vincent, 1990) ou HOMA (Blondeau *et al.*, 2021), qui s'appuient sur la référence non genrée (1), les termes qui dénotent les femmes (2) ou les violences (3).

- 1) a. pis on dit « ah les jeunes s'impliquent pas » mais je pense que les jeunes s'impliquent comme **ils** peuvent et que souvent **ils et elles** ne sont pas invitées à participer non plus
(LocMon10, 26 ans)
- b. j'ai trouvé un peu intéressant ce que **elles et ils** faisaient dans l'asso
(LocPa11, 38 ans)
- c. je pense que les trentenaires ont plus de chance de se dire féministes si **iels** sont allé.e.s à l'université
(LocMon34, 30 ans)
- 2) a. puis ma mère en plus c'est **une nana** qui a une grande gueule donc elle c'est pas **la femme** qui ferme sa gueule
- b. j'ai relancé **une fille** que j'avais rencontrée l'année dernière qui lançait sa boîte
- c. il y avait pas mal de **nanas** et surtout des justement des **meufs** leaders
(LocPa01, 29 ans)

- 3) a. j'étais alors pas tout seule mais quand même en première ligne pour ce qui était la lutte contre les **violences sexistes et sexuelles** dans cette université

(LocPa21, 25 ans)

- b. c'est la première fois que des avocats montent une association de lutte contre les **violences faites aux femmes**

(LocPa15, 40 ans)

- c. c'est aussi un peu une espèce de courte sociohistoire de l'évolution des politiques publiques de lutte contre les **violences de genre violences faites aux femmes**

(LocMono9, 31 ans)

Il fallait donc commencer par définir le périmètre du féminisme. Nous suivons Moser et Tillous (2020), qui défendent l'expression de féminisme/queer, afin de saisir ensemble les luttes autour du genre et autour des sexualités, de les articuler sans les confondre. Nous pensons que cette articulation est cruciale d'un point de vue théorique comme politique. Mais nous devons être également en mesure de saisir les différents positionnements à ce sujet. Dès l'élaboration du périmètre, il s'agissait donc d'être en mesure de saisir le féminisme et le queer dans leurs différentes dénominations, ainsi qu'en fonction de leur frontière. Nous avons cherché à recruter des personnes « engagé.es dans le féminisme, la cause des femmes, le queer ou les luttes pour les sexualités », en introduisant notre étude selon une formulation précise lors de la prise de contact avec de potentielles recrues.

Il aurait également été méthodologiquement problématique de structurer notre recrutement selon des types idéologiques, théoriques ou politiques, de féminismes que nous présupposions, mais qui ne sont pas nécessairement ceux qui structurent effectivement le champ féministe (du point de vue des locuteurices, comme au-delà), et qui ne sont pas forcément nommés de manière univoque : certaines dénominations s'appliquent à différentes positions idéologiques (par exemple *radical*), certaines dénominations s'emploient majoritairement en hétéroqualification (par exemple *TERF*), ou ne désignent pas les mêmes positions selon qu'elles sont revendiquées ou attribuées à d'autres (par exemple *intersectionnel*) (Bendifallah *et al.*, 2023). Partir de telles catégories nous aurait fait courir le risque soit de trouver ce que nous cherchons – sans garantie que ces catégories soient représentatives d'autre chose que de notre propre lecture et de nos propres pré-construits – soit d'invalider en permanence les catégories choisies par les actrices.

En l'absence de consensus clair sur le découpage idéologique du champ féministe, et afin de ne pas présupposer les positions idéologiques que nous voulions observer (en d'autres termes pour ne pas risquer que notre classification soit auto-réalisatrice), nous avons donc préféré ne pas pré-concevoir ce découpage, pour plutôt saisir la complexité du champ féministe en termes de pratiques. En effet, les différentes pratiques du féminisme sont diverses et observables. Nous avons, en ce sens, choisi de recruter les personnes enregistrées selon leur type d'engagement, traduit en modalités d'activités. De quelles façons sont-elles en activité sur les questions de genre et de sexualités ? Les personnes sont-elles engagées professionnellement dans le féminisme ou bien de manière militante bénévole ? S'emploient-elles à la création, à la diffusion de discours féministes, et dans quels cadres ? Ont-elles une pratique individuelle ou collective du féminisme ?

Nous avons identifié 5 modalités :

- **Universitaire :** L'engagement professionnel pour la formation, la création et la production de connaissances féministes. Il s'agit ici principalement d'universitaires travaillant en études de genre, mais également de personnes s'employant à diffuser une lecture féministe de la science (universitaires, médecins) ou à promouvoir les femmes dans les sciences.
- **Professionnelle :** L'activité professionnelle de fabrique, de négociation ou d'application des lois, règlements, procédures en lien avec le féminisme, la cause des femmes ou les droits sexuels. Il s'agit ici typiquement de professionnel.les de l'égalité, des questions de genre et/ou de sexualité (chargées de mission ou référent.es égalité femmes/hommes, travailleuses de la diversité du secteur public ou privé, travailleuses communautaires, avocates, formatrices).
- **Associative :** Les activités associatives ou communautaires bénévoles au sein d'associations ou d'organismes. Ce travail peut être du lobbyisme sur un secteur spécifique ou, plus largement, un travail d'élaboration de cadre d'action sociale. Ce sont ici des militant.es, porte-parole, etc., d'associations, de collectifs, ou de syndicats.
- **Média :** Le travail, salarié ou bénévole, de diffusion médiatique et numérique des idées féministes. Cette modalité concerne les médias, le monde du livre et de la presse (papier comme numérique) et rassemble des autrices, éditrices, bibliothécaires, rédactrices de revues, traductrices, personnalités publiques, journalistes, etc.

- **Collectif:** les activités réticulaires de contestation ou de solidarité, activités bénévoles qui prennent place dans des espaces militants non-organisés (militance en ligne, collectifs non-organisés, regroupement autour de lieux, réseaux informels, etc.).

Nous faisons l'hypothèse que la façon de s'investir dans des « causes » de manière scientifique, salariée, organisée ou numérique offre un curseur sur différents axes : degré d'institutionnalisation, axe majoritaire/radical, axe collectif/individuel, etc., sans investir idéologiquement ces termes. Cela nous permet de rassembler dans les entrevues des personnes situées très différemment sur l'échiquier politique (sur la dichotomie gauche/droite, universaliste/intersectionnaliste, légaliste/anti-autoritaire, etc.) avec pour point commun une appropriation des questions féministes ou de sexualités. L'analyse permettra d'établir si des corrélations apparaissent entre type de pratique du féminisme (engagement) et positionnement idéologique au sein du champ féministe.

Au-delà du critère de l'engagement féministe/queer, nous avons également appliqué un critère géographique. Il fallait qu'une partie de la trajectoire, personnelle ou professionnelle, se soit déroulée dans une des trois villes du corpus, à savoir Montréal, Paris ou Marseille. Cette approche géographique à l'échelle de la ville est pertinente pour plusieurs raisons. Elle permet tout d'abord de garder un contexte comparable entre les personnes d'un même sous-corpus évoluant dans un espace social commun (la ville), ce qui n'aurait pas été possible à l'échelle nationale (disparité entre urbain et rural, tailles diverses des agglomérations, disparité des histoires féministes dans chaque ville, disparité des mobilités géographiques et sociales selon les villes, etc.), tout en offrant la possibilité de comparer les sous-corpus (Paris, Marseille, Montréal), à l'échelle nationale comme internationale.

Cela permet également de repérer d'éventuels phénomènes linguistiques propres à des variétés du français, sans avoir à les généraliser à l'échelle de variétés nationales (par exemple France/Québec), mais aussi sans essentialiser les locuteurices, en évitant de les ramener à une qualité de « natifs » qui fait bien peu sens dans les contextes de mobilité géographique et sociale contemporains (qui parmi la population parisienne pourraient prétendre à être natif.ve, et en quoi seraient-ils ou elles plus représentatifs de la population parisienne réelle?). D'un point de vue analytique, cette approche permet de sortir d'une lecture dissociant centre et périphérie (ce qui aurait été le cas dans une comparaison Marseille/Paris) ou d'une lecture prenant en compte

les variétés nationales coexistantes (ce qui aurait été le cas dans une comparaison Montréal/Paris), pour basculer dans une lecture pluricentrique, où les rapports de centralité s'entremêlent. Idéalement un quatrième terrain aurait permis de travailler également la tension centre/périphérie dans l'espace nord-américain francophone, mais cela n'a pas été possible pour des raisons de temps. Enfin, d'un point de vue socio-idéologique, Montréal, Paris et Marseille correspondent actuellement à un triangle de circulation majeur dans certains espaces féministes queers de la francophonie occidentale, ce qui confirme l'intérêt du choix de ces trois villes en particulier¹.

Le fait de constituer notre corpus sur la base de l'engagement permet ainsi d'éviter des écueils méthodologiques et de capturer des dynamiques sociolinguistiques au-delà des macrostructures sociales traditionnelles que sont l'âge, la classe, le genre, la race, le degré d'étude, le pays, etc. Cependant, se pose tout de même la question de la représentativité. Si notre méthode de constitution nous préserve de l'effet de bulle idéologique, comment garantir une certaine homogénéité des profils et à quelle représentativité pouvons-nous prétendre? La taille du corpus (102 entrevues) ne se prête pas à faire un échantillon statistiquement représentatif de la population féministe/queer, ni à représenter l'ensemble des expériences et des coordonnées sociales du féminisme dans une ville donnée. Mais il nous permet d'aller au-delà d'une approche ethnographique. Cette échelle n'est donc pas problématique pour l'interprétation des résultats de nos analyses, tant que l'on ne cherche pas à généraliser excessivement. Elle permet de recenser une partie des mouvements féministes, de repérer des tendances et de montrer des récurrences sans les affirmer pour l'ensemble du féminisme ou d'une ville.

Comme cela a été déjà remarqué (Bertaux, 1997), cette question de représentativité est aussi une question théorique pour les sciences sociales. Appliquée à la sociolinguistique, elle permet d'éclairer la tension entre approche sociolinguistique classique et approche sociolinguistique prenant en compte le discours. En effet une démarche cheminant selon les déterminismes sociaux uniquement est trop limitée pour saisir des variations sociolinguistiques plus fines, car les personnes ne sont jamais seulement féministes ou queer. Elles sont prises dans des rapports sociaux, notamment de genre, de classe et de race, qui s'articulent avec leur positionnement idéologique. Il est donc nécessaire

1. Cette circulation est peut-être encore trop récente pour être documentée, mais elle apparaît dans la trajectoire et le discours d'un certain nombre de participant.es du corpus, qui témoignent également de la vivacité du tissu social queer et féministe dans ces trois villes francophones.

d'adosser la constitution de notre corpus à une vision complexe du social, qui prenne en compte l'activité critique des acteurs et actrices sociales à la façon de la sociologie pragmatique (Boltanski, 2009) ou de la théorie queer (Butler, 2005), sans faire disparaître pour autant les rapports sociaux, tels que les a décrits, par exemple, le féminisme matérialiste (Guillaumin, 1992, Delphy, 1997 ou plus récemment Bilge, 2010). En d'autres termes, notre cadre théorique est, lui aussi, féministe/queer.

Nous avons donc pris garde d'équilibrer les types d'engagement, les orientations politiques et les niveaux d'éducation, et nous avons veillé à ne pas constituer un corpus uniquement blanc ou cis. Au sein de l'université, nous avons également veillé à diversifier les statuts et les disciplines.

Nous présentons ici quelques cas limites qui éclairent cette tension entre ancrages idéologiques et structures sociales, et ses conséquences sur la représentativité du corpus.

La majorité des personnes enregistrées sont blanches. Nous avons bien sûr cherché à intégrer des personnes non-blanches, mais que signifie cette inclusion ? Elle permet certainement une visibilisation des voix non-blanches du féminisme et du queer, et de ne pas reproduire le blanchissement du féminisme. À ce titre, elle est cruciale. Mais cette inclusion peut-elle prétendre pour autant à une quelconque représentativité ? Il faudrait connaître la structuration du féminisme dans chaque ville. Par exemple à Paris, des personnes ayant des expériences minoritaires de la racisation ont été enregistrées, mais est-ce que cela est représentatif des présences minoritaires dans les milieux féministes de cette ville ? Que signifie cette expérience minoritaire du point de vue de l'analyse que l'on peut en faire, sinon une expérience du racisme ? À Montréal, uniquement des personnes blanches et des personnes noires ont participé aux entrevues. Ce n'est certainement pas représentatif du féminisme montréalais francophone. Mais sur quoi pouvons-nous nous baser pour faire pencher la balance vers une répartition plus représentative ? Enfin, pour Paris au moins, les personnes blanches sont assez majoritaires dans le féminisme et le queer. Souhaite-t-on représenter cet effet-là, qu'il est important de souligner ? Il serait en effet problématique qu'une représentativité méthodologique masque le fait que certains espaces sont majoritairement blancs.

D'une manière différente, la question de la représentativité concerne également l'âge. Le corpus comprend des locuteurices âgées de 19 à 83 ans, avec une surreprésentation des 25-30 ans, et une sous-représentation des 35-59 ans, dans une proportion relativement similaire pour les trois volets du corpus (répartition de l'ensemble du corpus : 18-34 ans : 45 % ; 35-59 ans :

23 % ; 60-83 ans : 30 %. Montréal 18-34 ans : 59 % ; 35-59 ans : 27 % ; 60-83 ans : 15 % ; Paris : 18-34 ans : 41 % ; 35-59 ans : 32 % ; 60-83 ans : 27 %. Marseille : 18-34 ans : 55 % ; 35-59 ans : 25 % ; 60-83 ans : 20 %). Une distribution paritaire des âges à tout prix ne serait pas significative, pour deux raisons. D'une part, les milieux militants sont généralement plus jeunes et les postes de responsabilité, institutionnels sont souvent occupés par des personnes plus âgées. D'autre part, cela reflète probablement que – pour la France au moins – les femmes nées dans les années 1970 sont, selon Bereni (communication personnelle), « moins nombreuses à s'être engagées en féminisme que leurs aînées (politisées dans les années 1970) ou leurs cadettes nées dans les années 1980 et surtout 1990 (qui se sont politisées à partir de la fin des années 2000, au moment de la résurgence du féminisme en France) ». Malgré quelques grands événements féministes pour cette génération, les années 1990 sont surtout la période de l'institutionnalisation des questions de genre. Il est donc nécessaire de chercher la variabilité autant que possible, mais sans la forcer afin de ne pas défigurer la représentation de la structuration sociale du féminisme observé.

Il faut enfin préciser que nous avons souhaité ne pas nous limiter à des francophones « natif.ves », et que le corpus contient des locuteurices bilingues (anglais/français, mais aussi espagnol/français, arabe/français, etc.) ou issues de familles mixtes (par exemple père anglophone, mère francophone).

Toutes ces questions nous ont conduites à développer l'idée de *représentativité qualitative*, qui permet de se décaler du dilemme de la saturation théorique des catégories (Glaser et Strauss, 1967) dans les approches qualitatives. Le corpus doit représenter un assez grand échantillonnage pour recueillir une variété de subjectivations politiques. Cela permet de ne pas reproduire les biais de représentations majoritaires qui écartent les personnes minoritaires de l'intelligibilité, et les réduisent au silence. Cela permet également d'articuler de manière fine positionnements idéologiques et positionnements sociaux sans réduire les uns aux autres. Cela implique enfin de ne pas tirer de conclusions trop générales sur la structuration des milieux féministes et queers.

Cette question de l'entremêlement des positions idéologiques et des rapports sociaux, et de ses implications linguistiques, a donc des conséquences méthodologiques très concrètes sur la constitution du corpus et donc sur les données. C'est là, croyons-nous, l'un des apports originaux de ce corpus.

Concernant le contenu, il s'agit d'entrevues semi-dirigées structurées autour d'un questionnaire en trois parties : une présentation de la trajectoire

biographique des locuteurices, une discussion sur leur positionnement politique et idéologique, et enfin une discussion sur leur rapport au langage. Chaque partie contient des questions telles que :

- **Trajectoire biographique :** lieux (sociaux et géographiques) de l'enfance, formation, trajectoire professionnelle, rencontre avec le féminisme, le queer, les sexualités ou la cause des femmes.
- **Positionnements idéologiques :** rapport au militantisme, vision du politique, rapport au féminisme, alliés et ennemis, rapport à l'action publique et l'État, prison, lois, violence, intersectionnalité, légitimité et expertise, rapport aux questions de classe, de racisme, de générations, géographique intellectuelle ;
- **Questions de langage :** contrôle du langage, auto-dénomination, écriture inclusive et pronom, termes employés, rapport au tabou, place du langage dans la lutte. Nous demandons ici explicitement si les personnes repèrent des éléments de leurs pratiques langagières qui encodent des sens sociaux. Nous présentons également un certain nombre de termes relevant du domaine du genre et des sexualités en demandant aux locuteurices d'en donner les connotations (explication du sens social).

Cette structure du questionnaire permet de capter les processus de subjectivation féministe et queer. Les sujets abordés correspondent à de grandes lignes de partition actuelles sur les questions de genre et de sexualités (place de l'antiracisme, articulation genre et sexualités, rapport à l'État, etc.), sans se focaliser sur les sujets les plus polémiques (travail du sexe/prostitution, laïcité/islamophobie, identités trans) pour éviter aux personnes de devoir polariser leur discours.

Devenir féministe et/queer à Paris et à Montréal

Le questionnaire du corpus CaFé comporte un grand nombre de questions sollicitant des récits de vie de nos participant.e.s, telles que [B1] *Quelle est votre trajectoire biographique?* [B2] *Comment avez-vous rencontré le féminisme?* [P7] *Est-ce que votre féminisme a changé/Avez-vous toujours eu les mêmes positions?* Dans cette section, nous présentons une étude détaillée des réponses à la question [B2] dans les sous-corpus parisiens et montréalais (nous laissons Marseille de côté afin que la taille des deux sous-corpus soit équivalente et comparable). Ces récits éclairent plusieurs aspects de la fré-

quentation du féminisme dans les deux villes, ainsi que différents sens attribués à la dénomination « queer », qui se révèlent structurants pour notre corpus et permettent d'expliquer certains enjeux de représentativité qualitative.

La question du devenir féministe est très complexe et peut être décomposée en au moins trois sous-questions :

- **La prise de conscience :** *Comment avez-vous commencé à être sensibilisé.e aux inégalités de genre ?*
- **L'engagement :** *Comment avez-vous commencé votre militantisme féministe/queer ?*
- **La dénomination :** *Comment avez-vous commencé à employer le mot « féministe » (ou « queer ») pour vous décrire ?*

Cela dit, dans le corpus CaFé, les locuteurices entremêlent souvent la prise de conscience, l'engagement et l'adoption du terme dans leurs réponses. Nous les traiterons donc ensemble. Nous repérons cinq types de prise de conscience : l'observation et/ou l'expérience d'inégalités ou d'injustices ; les expériences internationales ; les lectures et réseaux amicaux ; le militantisme sur d'autres questions ; enfin l'attrance queer.

La prise de conscience féministe et queer

Le récit de prise de conscience féministe le plus représenté dans le corpus CaFé est, de loin, celui qui s'articule autour de l'observation et/ou de l'expérience d'inégalités ou d'injustices liées au genre. On retrouve ce discours aussi bien à Paris (15 participant.e.s) qu'à Montréal (14 participant.e.s). Comme le montrent les exemples en (4 et 5), plusieurs participant.e.s ont été sensibles aux inégalités de genre dès l'enfance ou l'adolescence, même si, parfois, l'utilisation du terme *féministe* vient seulement après.

- 4) j'ai **toujours l'impression d'avoir plutôt** eu une **sensibilité assez forte** aux inégalités notamment liées au genre mais que je percevais pas de cette manière-là avant mais grosso modo en fait ça m'énervait de pas pouvoir faire la même chose que les garçons et du coup j'ai un jour juste décidé que j'étais féministe
(LocPa17, 24 ans)

- 5) j'ai autour de 12, 13, 14, 15 ans, ces années-là [...] **je suis très très très très rapidement consciente des inégalités entre les femmes et les hommes**, des choses que en principe je ne peux pas faire mais je fais quand même

(LocMono2, 57 ans)

Les expériences internationales sont aussi citées par nos participant.e.s comme contribuant à une prise de conscience féministe et/ou queer. C'est surtout le cas dans le sous-corpus parisien, où les séjours à l'étranger, et notamment au Québec, ont permis aux locuteurices de suivre des cours universitaires en études de genre, ce qui n'était pas possible jusqu'à très récemment en France (6).

- 6) a. ce que je préférerais c'était les théories queer j'en ai même profité je suis + deux ans après je suis allé bah **je suis parti faire un semestre au Canada aussi j'en ai profité pour suivre un cours sur les théories queer** passionnant [ah bah ouais là (rire)] ah bah c'était autre chose que ce qu'on avait en France aussi hein

(LocPao4, 28 ans)

b. ouais alors pour le féminisme en fait c'est lié à mon parcours [...] je pense qu'en fait ça a été un un double mouvement c'est-à-dire que comprendre le patriarcat comprendre les questions de consentement enfin commencer à travailler là-dessus [...] et **sur le queer je dirais plutôt bah c'est Montréal** en fait

(LocPa41, 33 ans)

À Montréal, les formations universitaires sont effectivement un des lieux principaux de prise de conscience féministe et queer; neuf locuteurs ou locutrices mentionnent des cours en études de genre dans ce sous-corpus (7) contre seulement deux locutrices du sous-corpus parisien.

- 7) Mon féminisme a évolué je pense à travers le temps **surtout dans mes études universitaires** je pense [...] les lectures sur l'intersectionnalité m'ont donné envie de creuser à la fois mes réflexions féministes mais aussi de les mettre encore plus de l'avant.

(LocMon30, 30 ans)

Pour les Parisien.ne.s, l'éducation féministe et queer se fait (ou se faisait) souvent de façon informelle, par les lectures (8a), les réseaux amicaux ou amoureux (8b) ou les réseaux sociaux et publications en ligne (9).

- 8) a. eh bah c'est Beauvoir nan mais franchement hein [...] c'est-à-dire que c'est vraiment théoriquement au début hein **c'est vraiment l'enseignement quoi par les livres** seule effectivement à la maison chez mes parents c'est avec Simone de Beauvoir que je me suis dit « tiens là il y a un truc »
(LocPa40, 62 ans)
- b. j'ai eu ma première relation avec une nana à la même période qui m'a foutu en main **les textes de Delphy** euh des années soixante-dix en disant que je déconnais complètement à penser qu'il y avait que des questions économiques
(LocPa11, 38 ans)
- 9) j'ai découvert *Mademoizelle.com* je sais plus comment mais **j'ai découvert Mademoizelle et ça m'a ouvert la porte vers tout un tas de concepts** assez cools
(LocPa27, 30 ans)

Dans les deux sous-corpus, certains discours sur la prise de conscience mentionnent des événements dans l'actualité (comme l'affaire Dominique Strauss Kahn en 2011 (LocPa18, 36 ans) ou l'élection du président progressiste François Mitterrand en 1981 (LocPa28, 68 ans). Quatorze participant.e.s parisien.ne.s et douze montréalais.e.s décrivent une prise de conscience ou une décision d'engagement survenues lors de leurs réflexions sur d'autres questions sociales, comme le racisme et l'islamophobie, les droits LGBT+ ou l'écologie. Par exemple, la locutrice LocPa24 raconte que sa prise de conscience féministe et contre l'islamophobie a été déclenchée par la découverte de l'hostilité envers des militantes du groupe Femen (10).

- 10) ma première image que j'associais au mot féministe quand j'étais plus petite c'était en primaire et c'était souvent des manifestations de Femen et je me disais « mh il y a de l'idée moi je le ferais pas mais j'aime bien le concept » [...] enfin **j'aimais pas le portrait qu'on faisait d'elles** et souvent à ce moment-là j'aimais pas non plus le portrait qu'on faisait de l'islam à la télé donc **mon cerveau a fait le rapprochement** des deux de (rire) **pourquoi on parle mal des femmes et de l'islam tout le temps**

(rire) du coup je me suis dit ben c'est « je suis les deux donc ça ne me plaît pas et voilà »

(LocPa24, 19 ans)

Il faut noter que tous.tes les locuteurices susceptibles de subir de l'islamo-phobie n'ont pas eu la même réaction que LocPa24 face aux manifestations des Femen. Au contraire, d'autres locutrices, comme LocPa42 ou LocPao8, identifient une tension entre ces représentations du féminisme très visibles dans les médias au début des années 2000 et le fait d'être musulmane ou racisée (11). En conséquence, certaines participantes racisées dans les deux sous-corpus ne se revendiquent pas comme féministes à cause de cette tension (12).

- 11) **la vision qu'on avait du féminisme** c'était une vision ben c'était les Femen c'était Caroline Fourest c'était bah **des gens qui nous ressemblaient pas** et qui en plus nous mettaient toujours à une place un peu un peu problématique donc même si on avait conscience effectivement de l'effet de notre genre sur ben notre vie et sur notre éducation et que c'est des choses desquelles dont on voulait s'émanciper [Julie: ouais le terme féminisme était pas disponible] ah non [...] et encore aujourd'hui il y a des femmes proches de mon entourage qui refusent cette étiquette féministe

(LocPa42, 40 ans)

- 12) a. c'est que juste en étant née comme une femme noire je suis une femme noire alors que je le veuille ou non je vais faire face à certains systèmes d'oppression que j'ai pas le choix de lutter contre alors **c'est pas que je suis féministe parce que je veux être féministe** ou je vais être « pro-black » parce que je veux être « pro-black » je suis on dirait dans ma tête c'est juste je le suis [...] **je ne dirais pas que je suis féministe** mais le travail que je fais je sais que ça tient de théories et de courants féministes

(LocMon12, 26 ans)

b. **je dis pas de moi que je suis féministe** c'est les autres qui le disent mais moi je le dis pas [...] et ça me dérange pas de le dire au contraire tu vois mais je le revendique pas [...] c'est difficile de dire « je suis ça je suis ça ou je suis ça » parce que je suis plein de choses en fait donc si je commence à dire je suis féministe ça veut dire que je suis pas autre chose tu vois si on dit que je suis si je commence à dire je suis française ça veut dire que je suis pas algérienne si on dit que je suis algérienne ça

veut dire que je suis plus française donc du coup ça c'est [...] c'est compliqué en fait de se revendiquer

(LocPao8, 44 ans)

Enfin, plusieurs participant.e.s dans les deux sous-corpus arrivent au féminisme par leur implication dans des mouvements de gauche. On observe ici une légère différence entre Paris et Montréal : pour les Parisien.ne.s, il s'agit souvent d'une implication dans des syndicats ou organisations politiques d'extrême gauche (13a) ou de gauche libérale (13b), alors que pour les Montréalais.e.s, il s'agit plus souvent d'une implication dans le mouvement étudiant et les grèves de 2005 ou 2012 (14).

- 13) a. ma première formation politique s'est faite par **le trotskisme en l'occurrence à Lutte Ouvrière** quand j'étais au lycée en prépa où clairement la question féministe n'était pas beaucoup prise en charge c'est plus à Lyon que peu à peu dans le cadre de **la section Sud Éducation** qu'on avait montée avec des camarades là-bas il y a eu peu à peu des formations des lectures fémin- des ateliers féministes qui ont été organisés même si c'est plus par la suite pendant les quelques années que j'ai passées à Dijon que j'ai commencé à militer plus spécifiquement sur les questions féministes et vraiment à m'approprier le terme

(Paris, Loc25, 33 ans)

- b. je me suis engagée en 2012 donc j'allais avoir quinze ans **au mouvement des Jeunes Socialistes** donc c'est là que j'ai reçu une première on va dire formation en termes formels aux questions féministes mais bon du coup après comment dire ma position a beaucoup évolué depuis

(LocPa17, 24 ans)

- 14) j'ai milité dans le mouvement étudiant donc nous au cégep là donc **à travers le mouvement militant étudiant** j'ai rencontré le féminisme

(LocMon28, 42 ans)

Être queer à Paris et à Montréal

Un dernier type de prise de conscience présent dans le corpus est celui de l'attraction sexuelle (15). Il concerne plus particulièrement la rencontre avec le queer, et on le trouve uniquement dans le sous-corpus montréalais.

- 15) toutes mes ami.es là se définissent féministe queer du bac en sexo tsé y a personne c'est ça qui est cis hétéro là tsé pis là «ah euh je sais pas trop» pis à moment donné je suis comme «ben non mais tsé je suis quand même flexible» [...] pis genre pour moi **c'est une femme là dans un film** québécois peu importe pis genre tsé comme y a plein de trucs je suis comme «**ben oui t'as eu un "crush"** sur telle personne t'as eu un "crush" sur, tsé comme "of course"»
(LocMon29, 35 ans)

Cette présence spécifique au corpus montréalais s'explique par le fait que le mot *queer* y est plus souvent compris comme la dénomination d'une orientation sexuelle et/ou d'une catégorie de genre, en opposition avec la catégorie *cis hétéro* (16a), et se rapprochant de *non-binaire*, sans être la même chose (16b).

- 16) a. ça a quand même été un un débat longtemps de comme «est-ce que queer signifie comme **toutes les personnes de genre la communauté qui sont pas comme cis et hétéro**» tsé ou c'est juste comme spécifique à tsé quelque chose qui est un peu est contre comme pas dans les normes
(LocMon23, 25 ans)
- b. **queer comme quelqu'un qui justement qui sort qui veut pas avoir avoir une étiquette** ou avoir une qu'on lui appose une définition précise liée à son identité de genre **alors que non-binaire on est plus dans comme si y avait juste les deux pôles** pour moi tsé l'échelle avec les deux pôles homme femme
(LocMon26, 37 ans)
- c. cette cette acceptation-là du du du féminisme ou cette revendication-là est venue avant mes revendications queer je pense que j'avais besoin de me donner le droit en fait d'exister sans subir ou tsé au moins j'étais comme consciente du patriarcat pis des systèmes d'oppression fait-que là après ça je pouvais réfléchir plus sur mon orientation sexuelle je m'identifie comme bisexuelle [...] puis j'ai fait mon bout de chemin avec ça c'est bien le fun on aime ça puis là **je suis rendue plus à me questionner sur euh mon identité de genre** tsé je suis comme plus euh tsé euh je je «relate» pas au complet à l'étiquette de femme fait-que là je suis plus comme dans le demi-femme «gender queer» c'est là que je suis en ce moment
(LocMon34, 26 ans)

Pour leur part, quand on leur demande de décrire leur découverte du queer, les locuteurices du corpus parisien décrivent plutôt le queer comme une orientation politique au sein du féminisme (au même titre que le matérialisme), représenté par les travaux de Judith Butler, entre autres, comme le montrent les exemples en (17).

- 17) a. mais sinon moi je me suis toujours dit euh féministe je crois [Julie: et queer du coup] **queer non c'est venu beaucoup plus tard [...] c'était plutôt à la fac** quand j'ai rencontré à Lyon des gens qui s'intéressaient au rapport entre matérialisme et queer et du coup on faisait souvent des débats [...] j'ai vraiment rencontré après **les lectures scientifiques elles sont venues plus tard** parce que j'avais un peu du mal avec Butler et compagnie

(LocPa36, 27 ans)

- b. queer euh je sais pas moi je pense pas que je suis très queer parce que je pense que genre aussi le féminisme il a pas forcément [...] bien sûr que **idéalement on vivrait dans une société sans genre et du coup sans oppression basée sur le genre** et ça serait fantastique mais le truc c'est que pour moi **le queer ça a tendance justement à flouter** en fait enfin ça a vocation à flouter prématurément les questions de genre notamment axées autour de la binarité si ce n'est réelle du moins perçue par la société

(LocPa16, 25 ans)

La dénomination *queer* est aussi utilisée, parfois, en tant qu'« orientation politique » à Montréal, mais, dans ce cas, elle peut être vue comme en conflit avec la dénomination *féministe* (18).

- 18) j'en suis venue presque dans ma conception de où **je me situe à opposer féminisme et queer** je me disais « moi je ne veux plus rien savoir d'une librairie féministe je veux une librairie queer » parce que [...] de me faire bombarder par des féministes blanches universitaires qui veulent lire du Mona Chollet j'ai rien contre Mona Chollet mais comme (rire) ça ça me rend un peu fatiguée

(LocMono8, 28 ans)

Les différents sens de *queer* à Paris et à Montréal sont aussi visibles dans le commentaire en (19), par un.e Néo-Montréalais.e immigré.e de France: iel

est surpris.e (et énervé.e) de voir que les personnes dans les espaces jeunesse à Montréal traitent *queer* et *non-binaire* de façon équivalente.

- 19) dans certains espaces communautaires jeunesse j'ai pu voir et ça **ça m'énerve beaucoup une espèce d'équivalence entre queer et non-binaire** alors que je trouve que ça a rien à voir
(LocMono3, 29 ans)

Catégories sociales et positionnements idéologiques : un enjeu sociolinguistique

Cette différence de sens du terme *queer* à Montréal et à Paris se traduit par le fait que plus de personnes d'identité de genre masculine correspondent à nos critères de sélection (« engagé.e dans le féminisme, la cause des femmes, le queer ou les luttes pour les sexualités ») à Montréal qu'à Paris. En conséquence, si, dans les trois sous-corpus, la majorité des participant.e.s ont une identité de genre féminine, ce que reflète la composition habituelle des milieux féministes en Europe et en Amérique du Nord, le sous-corpus montréalais est beaucoup plus masculin que les sous-corpus français. La composition démographique du corpus est donc étroitement liée aux significations en circulation dans les différents contextes : les catégories sociales susceptibles de faire varier les usages linguistiques dépendent aussi de ces divers sens dans des contextes sociolinguistiques donnés.

Le corpus ayant été planifié en France, nous ne nous attendions pas à cette différence de composition du corpus ni de sens du terme *queer*. Cependant, il n'aurait pas été pertinent de chercher à équilibrer démographiquement le corpus parisien en matière de genre, car la raison de la présence masculine dans le corpus montréalais est due à un sens de queer qui ne circule pas ou peu à Paris. La représentativité qualitative, telle que nous l'avons décrite plus haut, consiste ainsi à tenir un équilibre entre comparabilité et spécificité des sous-corpus.

Il ressort de l'analyse que la source des différences dans la composition démographique du sous-corpus québécois et du sous-corpus français relève d'une différence dans les articulations entre le féminisme et le queer, entre Montréal et Paris et, plus particulièrement, de différences dans la signification du terme *queer* selon que l'on se trouve en France ou au Québec. Cette interrelation du sens et des catégories sociodémographiques a pu être saisie grâce à l'analyse des récits biographiques. Nous espérons ainsi avoir montré

comment l'analyse de discours peut apporter une contribution à la sociolinguistique de corpus.

Références

- BENDIFALLAH, Lina *et al.* (2023), Conceptual Spaces for Conceptual Engineering? Feminism as a Case Study, *Review of Philosophy and Psychology*, 16, 199-229.
- BERNARD, Marie-Claude, Hervé BRETON, et Emmanuelle JOUET (2021), Récits de vie et savoirs: enjeux des enquêtes narratives, *Recherches qualitatives* 40(2), 1-11.
- BERTAUX, Daniel (1997), *Les récits de vie. Perspective ethnosociologique*, Nathan.
- BILGE, Sirma (2010), De l'analogie à l'articulation: théoriser la différenciation sociale et l'inégalité complexe, *L'Homme et la Société*, 176-177, 43-64.
- BLONDEAU, Hélène, Mireille TREMBLAY, Anne BERTRAND et Elizabeth MICHEL (2021). A new milestone for the study of variation in Montréal French: The Hochelaga-Maisonneuve sociolinguistic survey, *Corpus*, 22, <https://doi.org/10.4000/corpus.6221>.
- BOLTANSKI, Luc (2009), *De la critique*, Paris, Gallimard.
- BUTLER, Judith (2005), *Trouble dans le genre*, La Découverte.
- COSTA, Luciano R. et Yumi G. DOS SANTOS (2021), Revisiter le « récit de vie », une approche en sciences sociales. Entretien avec Daniel Bertaux, *Recherches qualitatives*, 40(2), 146-175.
- DELPHY, Christine (1997), *L'ennemi principal*, tome 1, Syllepse.
- GLASER, Barney G. et Anselm L. STRAUSS (1967), *The discovery of grounded theory, Strategies for qualitative research*, Aldine Transaction.
- GUILLAUMIN, Colette (1992), *Sexe, Race et Pratique du pouvoir. L'idée de Nature*, Côté-Femmes.
- LEVON, Erez (2010), *Language and the Politics of Sexuality. Lesbians and Gays in Israel*. Springer.
- MOSER, Cornelia et Marion TILLOUS (2020), *Avec, sans ou contre. Critiques queer/féministes de l'État*, Éditions Ixe.
- PASSERINI, Luisa (2008), La scrittura storica come forma di intersoggettività, *Contemporanea*, 11(3), 534-538.
- SANKOFF, David et Henrietta CEDERGREN (1972), Sociolinguistic research on French in Montréal, *Language in Society*, 1(1), 173-174.
- THIBAUT, Pierrette et Diane VINCENT (1990), Un corpus de français parlé. Montréal 84: Historique, méthodes et perspectives de recherche, *CIRAL*, 1.
- VEITH, Blandine (2004), De la portée des récits de vie dans l'analyse des processus globaux, *Bulletin of Sociological Methodology/Bulletin de Méthodologie sociologique*, 84(1), 49-61.

Instrumentation des corpus

CHAPITRE 10

Les corpus oraux montréalais à l'ère numérique

Des bobines magnétiques à la diffusion
sur la plateforme du Fonds de données
linguistiques du Québec

Wim Remysen et Hugo Saint-Amant Lamy

[...] en ce moment, je suis très angoissée pour l'avenir de nos bandes sonores. Nous n'avons pas encore trouvé le médium parfait pour en assurer une bonne conservation. Les données originales, ce n'est pas la transcription, c'est l'enregistrement. (Pierrette Thibault, citée dans Daveluy et Laforest, 1994, 19-20)

Maintenant que les corpus existent...¹

Cinquante ans après leur genèse, force est de constater que les corpus socio-linguistiques montréalais de 1971, 1984 et 1995 demeurent de puissants catalyseurs pour l'analyse et la description du français parlé au Québec, notamment dans la perspective variationniste. Il n'est pas exagéré d'affirmer que les sociolinguistes qui ont pris l'initiative de constituer le premier de ces corpus – Gillian Sankoff, David Sankoff et Henrietta Cedergren – font figure

1. Ce titre est un clin d'œil au chapitre que Gillian Sankoff a consacré, en 2018, à la recherche sur le changement linguistique, intitulé « Before there were corpora » (avant que les corpus existent).

de véritables pionnières². Non seulement leur entreprise a inspiré quantité d'autres chercheuses par la suite, mais de nombreux travaux réalisés à partir des données recueillies ont aussi grandement contribué à une meilleure connaissance de plusieurs phénomènes caractéristiques du français québécois³.

Dans un contexte encore fortement marqué par des préjugés négatifs envers cette variété de français, de tels travaux avaient, en plus de leur importance scientifique, une pertinence sociale certaine⁴. L'élaboration de corpus subséquents en 1984 et 1995, où certains locuteurs de 1971 sont interviewés à nouveau, a permis d'ajouter une dimension diachronique précieuse au riche portrait sociolinguistique du projet original (Sankoff, 2018; Blondeau, 2020). Cette richesse a suscité l'envie et l'intérêt de sociolinguistes partout dans le monde. Aujourd'hui, Montréal 1971 et ses suites n'ont aucunement perdu de leur intérêt, bien au contraire. Bien plus qu'une simple pièce d'anthologie, la désormais célèbre trilogie ouvre la possibilité d'étudier le changement qui affecte le français parlé à Montréal, notamment par des comparaisons avec des corpus constitués à une date plus récente, et permet de prendre le pouls des représentations de la langue qui existaient à l'époque. La granularité sociale et la profondeur longitudinale inégalées des trois corpus en font en outre un étalon de choix à l'aune duquel il est possible d'interpréter globalement la variation en français québécois.

Cela dit, au fil du temps, la préservation des données recueillies et leur accessibilité sont devenues des enjeux majeurs. Ainsi, la numérisation des corpus, que souhaitaient déjà les auteures de la trilogie montréalaise⁵, est relativement récente et leur mutualisation ainsi que leur dépôt n'avaient jamais été réalisés jusqu'à la mise sur pied récente du Fonds de données linguistiques du Québec (FDLQ), en 2020. C'est de cette initiative qu'il sera question dans ce chapitre, que nous dédions aux auteures qui nous ont pré-

2. Nous utilisons régulièrement dans ce chapitre le féminin générique pour souligner l'apport des femmes à la constitution des premiers corpus québécois ainsi que des corpus subséquents.

3. En fait foi l'impressionnante bibliographie des recherches menées à partir de leur corpus (ainsi que des deux autres qui ont suivi). Cette bibliographie, réalisée par Hélène Blondeau et Marty Laforest en 2015, est disponible sur le site du Fonds de données linguistiques du Québec (<https://fdlq.recherche.usherbrooke.ca/bibliographie-travaux-grands-corpus-sociolinguistiques-mtl.html>).

4. Il peut être instructif de relire la partie programmatique de l'article de Sankoff *et al.* (1976) pour s'en convaincre.

5. Voir à ce sujet les propos recueillis par Michelle Daveluy et Marty Laforest auprès des initiatrices des trois corpus (et publiés dans Daveluy et Laforest, 1994).

cédés. Soucieux d'établir des ponts entre le passé et l'avenir, et désireux de donner la parole à nos prédécesseurs, nous ponctuerons notre texte de citations de leur cru, ce qui nous permettra d'illustrer comment ce nouveau Fonds répond à des préoccupations exprimées depuis l'émergence de la linguistique de corpus au Québec dans les années 1960 relativement à la pérennisation des données.

Le Fonds de données linguistiques du Québec et la place accordée aux corpus montréalais

Je me souviens, quand Pierrette [Thibault] était étudiante, avoir eu des conversations avec elle et Diane [Vincent] sur un projet de création d'une banque de données ou d'un centre de recherches, quelque chose qui donnerait une continuité à ce travail, une institutionnalisation. (Gillian Sankoff, citée dans Daveluy et Laforest, 1994, 14)

Le Fonds de données linguistiques du Québec est une plateforme numérique documentaire destinée à mutualiser des corpus oraux et écrits constitués depuis les années 1960 en vue de décrire et d'étudier le français du Québec. Il contient des collections diverses de documents jugés représentatifs du français tel qu'il est utilisé au Québec, en tenant compte de la variabilité et de l'évolution de ses usages. Les différents corpus réunis sur la plateforme témoignent ainsi de la langue ayant cours dans diverses régions du Québec, dans des contextes d'utilisation et auprès de milieux variés ainsi qu'à différentes époques de son histoire. Certains corpus témoignent par ailleurs des rapports normatifs et identitaires particuliers que les Québécois ont entretenus ou entretiennent avec leur langue. Ces corpus métalinguistiques se prêtent bien à l'analyse des discours qui ont circulé au Québec concernant la qualité de la langue et l'acceptabilité des particularismes du français québécois.

S'inscrivant dans le domaine des humanités numériques, le Fonds est développé par une équipe de linguistes-chercheurs associés au Centre de recherche interuniversitaire sur le français en usage au Québec (CRIFUQ), en étroite collaboration avec des analystes-informaticiens du Service des technologies de l'information de l'Université de Sherbrooke (STI), et il est financé par le ministère de la Langue française du Québec. Il poursuit essentiellement trois grands objectifs: 1) assurer la pérennité des corpus linguistiques

constitués au Québec (dont plusieurs ont été réalisés avant l'ère numérique), 2) permettre la consultation et l'interrogation de leurs contenus (souvent peu accessibles aux chercheurs), 3) valoriser les pratiques linguistiques que les corpus donnent à entendre. D'outils utilisés ponctuellement pour des recherches spécifiques, les corpus intégrés au Fonds sont ainsi appelés à devenir des ressources d'usage courant, facilement accessibles et partagées par la communauté des chercheurs. Le projet contribue à la patrimonialisation des corpus concernés : outre qu'il permet la découverte de ces collectes de données, il assure une reconnaissance de leur valeur scientifique et culturelle. Et la langue préservée dans ces corpus accède à ce même rang : la plateforme préserve des traces tangibles et les inscrit dans la conscience linguistique collective. De ce point de vue, le Fonds partage certaines caractéristiques avec d'autres plateformes spécialisées permettant d'héberger des corpus, comme Cocoon et Ortolang (deux initiatives soutenues par l'infrastructure Huma-Num) en France. En même temps, il s'en démarque à plus d'un titre, notamment pour ce qui est de la consultation et de la visualisation des données. La plateforme FDLQ permet notamment d'interroger, grâce à un moteur de recherche unique, l'ensemble des corpus qu'elle diffuse, et de peaufiner les résultats d'une requête par l'activation de filtres (<https://fdlq.crifuq.usherbrooke.ca/>). Des recherches transversales de ce type ne sont pas disponibles sur les autres plateformes, où la consultation se fait corpus par corpus. Le moteur de recherche permet de trouver des mots simples, des suites de mots, des collocations ainsi que des lemmes⁶, et de formuler des requêtes incluant des métacaractères (jokers) ou des expressions régulières (regex)⁷. La plateforme FDLQ se démarque également par les différentes possibilités de navigation qu'elle offre à l'utilisateur, qui peut, par exemple, parcourir la liste des corpus et des documents contenus dans le Fonds, et mieux comprendre ainsi la composition exacte des différents contenus qu'il consulte. Il peut aussi naviguer dans le lexique du Fonds (c'est-à-dire dans la liste de l'ensemble des mots-formes⁸ attestés, triés par ordre alphabétique ou par fréquence) ou encore dans les lexiques d'un corpus ou d'un document précis. Les entrées de tous

6. Les corpus versés au Fonds sont lemmatisés avec SpaCy (lemmatisation automatique).

7. Les métacaractères correspondent à un ou plusieurs éléments non spécifiés dans une requête (ex. « pi?e » renvoie les résultats *pire*, *pige*, *pile*, etc. alors que « pi*e » renvoie aussi *piastre* et *pierre*). Les expressions régulières, de leur côté, permettent de chercher des éléments répondant à différentes conditions logiques. Par exemple, l'expression régulière « \w+z » renvoie tous les mots (\w+) se terminant par la lettre Z.

8. La lemmatisation des corpus, réalisée seulement récemment, permettra la constitution d'autres lexiques, comprenant cette fois-ci les mots attestés dans les corpus sous forme de lemmes.

ces lexiques se présentent sous forme d'hyperliens grâce auxquels on repérera par la suite les mots en contexte. Le site de présentation du Fonds apporte des informations complémentaires sur le Fonds, son équipe et ses partenaires, et contient des fiches de présentation pour l'ensemble des corpus qu'il héberge. Ce site comprend aussi un guide de consultation détaillé qui explique les principales fonctionnalités de la plateforme (<https://fdlq.recherche.usher-brooke.ca/guide/Content/>).

Les corpus montréalais occupent une place de choix parmi l'ensemble des données réunies sur la plateforme FDLQ. À l'heure actuelle, le Fonds inclut 10 corpus montréalais, dont certains ont déjà été versés intégralement (ou partiellement) alors que d'autres sont en cours de traitement. Il s'agit plus particulièrement de Montréal 1971 (Sankoff *et al.*, 1976), Montréal 1984 (Thibault et Vincent, 1990), Montréal 1995 (Vincent, Laforest et Martel, 1995), PFC-Montréal (un sous-ensemble de PFC-Québec; Côté, 2014; Côté et Saint-Amant Lamy, 2023), le corpus variationniste Hochelaga-Maisonneuve 2012 et le corpus variationniste Saint-Michel-Montréal-Nord 2013 (à propos de ces deux corpus, constitués dans le cadre du projet *Le français à la mesure d'un continent* (sous la direction de France Martineau) et versés au corpus FRAN, voir Gadet et Martineau, 2017; Martineau et Séguin, 2016), le corpus Ahuntsic-Cartierville (constitué par Davy Bigot) et Mont(REA)L 2016 (Rea, 2020). Ces corpus sont présentés de manière plus détaillée dans le tableau 10.1 ci-dessous, où ils apparaissent dans l'ordre chronologique de leur constitution. À cette liste devraient s'ajouter quelques autres corpus qui permettraient de compléter le volet montréalais du Fonds, dont le corpus Bibeau-Dugas (constitué au début des années 1960; Dugas, 1986), le corpus Centre-Sud de Claire Lefebvre (avec des entrevues menées de 1976 à 1978; Doran, Drapeau et Lefebvre, 1982), le corpus Friesner du français montréalais et du contact français-anglais-espagnol (réalisé en 2007-2008; Friesner, 2009) et le corpus réalisé auprès de Montréalais d'origine haïtienne (par Mireille Tremblay), auxquels s'ajoute la série d'entrevues menées en français auprès d'Anglo-Montréalais en 1993-1994 (par Gillian Sankoff et Pierrette Thibault, voir Thibault et Sankoff, 1993 et Sankoff *et al.*, 1997).

L'état physique sous lequel se présentent les enregistrements varie considérablement d'un corpus à l'autre. Les plus anciennes entrevues ont été sauvegardées sur des bandes magnétiques, un support d'enregistrement couramment utilisé dans les années 1960 et 1970, mais dont la consultation nécessite des machines de moins en moins disponibles dans les centres de soutien audiovisuel, ou encore sur des cassettes audio pouvant être lues sur

TABLEAU 10.1

Les corpus montréalais versés dans le Fonds de données linguistiques du Québec

Corpus	Auteurs et auteurs	Année de constitution	Nombre de participantes et de participants	Nombre et durée des entrevues	Support et format d'enregistrement
Montréal 1971	Gillian Sankoff David Sankoff Henrietta Cedergren	1971	120 personnes (60 femmes, 60 hommes)	120 entrevues (± 135 heures)	analogique (rubans magnétiques)
Montréal 1984	Pierrette Thibault Diane Vincent	1984	72 personnes (39 femmes, 33 hommes)	72 entrevues (± 100 heures)	analogique (cassettes)
Montréal 1995	Diane Vincent Marty Laforest Guylaine Martel	1995	14 personnes (5 femmes, 9 hommes)	14 entrevues (± 26 heures) auxquelles s'ajoutent des enregistrements écologiques (± 45 heures)	numérique (disque magnéto-optique, fichiers ATRAC)
PCF-Montréal (sous-corpus de PFC-Québec)	Marie-Hélène Côté	2010-2017	43 personnes (22 femmes, 21 hommes)	31 conversations libres (± 18 heures) et 43 entretiens guidés (± 14 heures) auxquels s'ajoute la lecture de deux listes de mots et d'un court texte	numérique (mémoire flash, fichiers WAV)
HoMa variationniste (Hochelaga-Maisonneuve)	Hélène Blondeau France Martineau Mireille Tremblay Yves Frenette	2012	48 personnes (24 femmes, 24 hommes)	46 entrevues (± 82 heures)	numérique (fichiers WAV)
MoMu variationniste (Saint-Michel-Montréal-Nord)	Hélène Blondeau France Martineau Mireille Tremblay	2013	20 personnes (15 femmes, 5 hommes)	20 entrevues (± 26 heures)	numérique (fichiers WAV)
MoMu écologique (Saint-Michel-Montréal-Nord)	Hélène Blondeau France Martineau Mireille Tremblay	2013	2 personnes (2 femmes)	1 entrevue (± 1 heure)	numérique (fichiers WAV)
HoMa écologique (Hochelaga-Maisonneuve)	France Martineau	2014	6 personnes (1 femme, 5 hommes)	1 entrevue (± 1 heure)	numérique (fichiers WAV)
Ahuntsic-Cartierville	Davy Bigot	2015	63 personnes (39 femmes, 24 hommes)	63 entrevues (± 42 heures)	numérique (fichiers MP3)
Mont(REA)L 2016	Béatrice Rea	2016	53 personnes (27 femmes, 26 hommes)	36 entrevues (± 60 heures)	Numérique (fichiers WAV)

des lecteurs de cassettes portatifs. Certaines bandes avaient déjà été recopiées sur des cassettes (c'est le cas notamment de Montréal 1971). Le corpus de Montréal 1995 marque une étape charnière, en ce sens que les entrevues ont été pour la première fois enregistrées sur un support numérique, notamment sur des disques magnéto-optiques, une technologie apparue vers le milieu des années 1980, mais somme toute assez rapidement disparue. Les enregistrements plus récents, réalisés à partir des années 2000, ont été faits directement sur un support numérique et non plus analogique. Nous verrons plus loin que le traitement des corpus en vue de leur diffusion sur la plateforme FDLQ tient nécessairement compte de l'hétérogénéité qui caractérise les formats d'enregistrement.

L'ensemble des corpus montréalais réunis au FDLQ représente aujourd'hui déjà près de 550 heures d'enregistrements. Cette impressionnante somme de données témoigne bien des efforts collectifs considérables déployés par les sociolinguistes pour documenter le français parlé dans la métropole (et ailleurs au Québec). Au travers de leurs choix relatifs aux informatrices et informateurs avec lesquels elles se sont entretenues, on comprend que la plupart des auteures des corpus versés au FDLQ conceptualisent le français montréalais comme étant la variété de français parlée par les francophones natifs originaires de Montréal (ou, à tout le moins, arrivés dès le jeune âge, généralement avant l'entrée à l'école), et vivant dans la ville (ou dans un quartier en particulier) au moment de l'entrevue. Cela dit, les critères de sélection précis ont pu varier d'un corpus à l'autre. En 1971, par exemple, l'équipe de Sankoff-Cedergren a recruté dans les quartiers majoritairement francophones de la ville des « individus de 15 ans et plus qui étaient à la fois francophones [de langue maternelle française] et Montréalais d'origine », donc tout individu « né à Montréal ou [...] arrivé à Montréal avant le début de l'école primaire, c'est-à-dire avant l'âge de six ans » (Sankoff *et al.*, 1976, 91-92). En 1984 et 1995, lorsque plusieurs informateurs de 1971 ont été retrouvés et réinterviewés, certains d'entre eux avaient quitté Montréal, si bien que les corpus subséquents incluent des entrevues menées auprès de personnes vivant en périphérie de Montréal (Rive-Nord ou Rive-Sud), voire (dans quelques cas plus rares) ailleurs dans la province⁹. Cette approche a permis d'évaluer l'impact de la mobilité géographique sur les usages linguistiques d'un individu (Daveluy, 2005). L'élargissement progressif de la conception même du

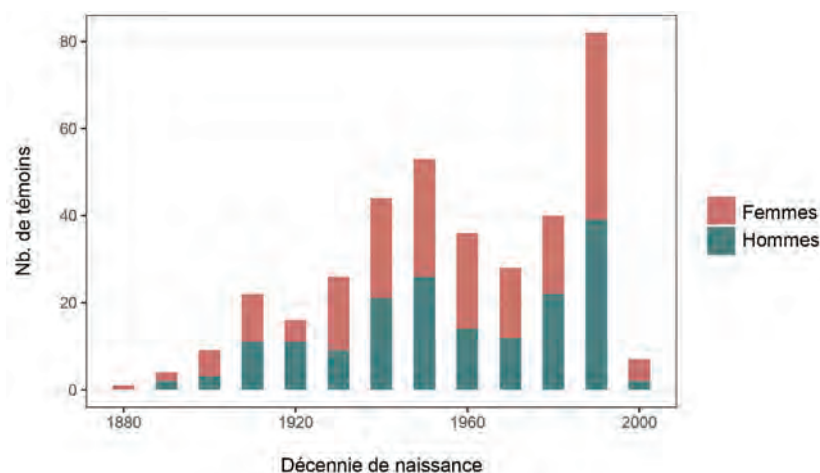
9. Selon Thibault et Vincent (1990, 22), 35 % des personnes interviewées à nouveau en 1984 vivaient à l'extérieur de l'île de Montréal au moment de leur deuxième participation.

français montréalais explique aussi que certaines collectes plus récentes incluent des entrevues menées auprès de Montréalaises et de Montréalais dont le français n'est pas la première langue apprise. On comprend mieux ainsi les effets des contacts linguistiques et de la diversité socioculturelle et linguistique croissante de la population montréalaise depuis les années 1970 sur le français parlé dans la métropole (Blondeau, 2014).

Si l'on tient compte des années de naissance des locutrices et locuteurs interrogés depuis les années 1970, le Fonds de données linguistiques du Québec permet de couvrir des pratiques linguistiques de personnes nées sur une période de plus d'un siècle (voir figure 10.1). En effet, environ 120 ans séparent la plus vieille locutrice, née en 1885 dans l'ancienne municipalité de Saint-Jean-Baptiste (maintenant intégrée à l'arrondissement du Plateau-Mont-Royal), des personnes les plus récemment rencontrées, nées dans les années 2000. Une telle richesse permettra de compléter considérablement, dans les prochaines années, notre compréhension de l'évolution du français montréalais, et de vérifier certaines hypothèses relatives à la théorie du changement linguistique. L'étude de cas présentée dans la dernière section de ce chapitre illustre concrètement l'apport du Fonds à cet égard.

FIGURE 10.1

Nombre de locuteurs par décennie de naissance dans les corpus montréalais



La mutualisation et la diffusion des corpus oraux sur la plateforme FDLQ soulèvent d'importantes questions relatives aux droits d'auteur, à la propriété intellectuelle et à la confidentialité des données. Deux principes fondamentaux encadrent l'action du FDLQ à ce propos : la reconnaissance du travail des concepteurs des corpus, et la protection de la vie privée des personnes interrogées, toute l'entreprise de la linguistique de corpus reposant sur ces dernières. Chaque corpus intégré au Fonds fait l'objet d'une licence d'utilisation signée par les auteurs (ou par la ou les personnes qui les représentent) et autorisant l'intégration du corpus concerné au Fonds. Cette entente balise essentiellement la diffusion qui peut être faite des données, garantit l'anonymisation des entrevues, protège l'intégralité du corpus et assure le respect de la propriété intellectuelle et des droits moraux¹⁰. Plus particulièrement, l'entente-type prévoit que les corpus ne soient jamais rendus accessibles dans leur totalité sur la plateforme et que les résultats de recherche se limitent à un contexte de 60 mots avant et après le mot-pivot ou les mots-pivots (une limite de 12 secondes s'applique également aux extraits audio), à moins qu'une telle diffusion soit autorisée explicitement par l'auteur. Chaque contenu extrait d'un corpus est systématiquement accompagné d'une référence bibliographique qui signale l'auteur et le nom du corpus. De toute évidence, le contenu des ententes varie d'un corpus à l'autre, suivant le consentement obtenu auprès des personnes participantes et suivant le souhait exprimé par les auteurs. Dans certains cas, en effet, il arrive que les entrevues puissent être diffusées de manière intégrale (c'est le cas notamment du corpus PFC-Québec), alors que, dans d'autres, l'accès est plus restrictif.

Le traitement des corpus : des enregistrements à la préparation des données

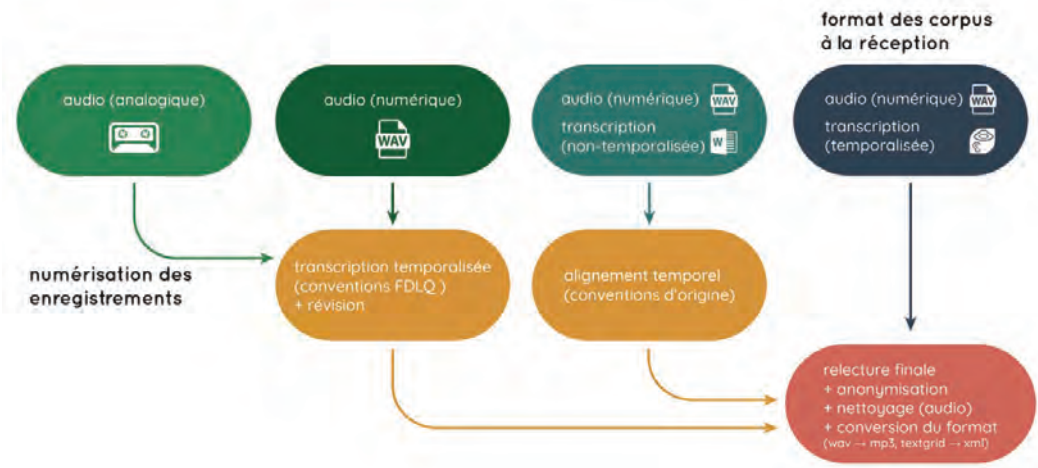
[N]ous avons passé des heures et des heures à discuter sur les conventions de transcription en se disant que c'était un outil qu'on faisait une fois pour toutes, mais ce n'est plus vrai. La transcription n'est aujourd'hui qu'un point de départ, un canevas de base.
(Pierrette Thibault, citée dans Daveluy et Laforest, 1994, 20)

10. Il va de soi que chaque auteur de corpus garde l'entière responsabilité de ses droits sur son œuvre. Au Canada, il est d'ailleurs impossible de céder les droits moraux, catégorie de droits qui « permet à l'auteur d'une œuvre d'en revendiquer la création » (source : <https://educaloi.qc.ca/capsules/le-droit-dauteur-pour-la-protection-de-la-creation/>).

Les problèmes et les enjeux, nombreux pour ce qui est du traitement des corpus, concernent toutes les étapes menant à la mise à disposition des données, incluant leur anonymisation. De façon générale, le déroulement des opérations est le même pour tous les corpus : numérisation des fichiers sonores, transcription temporalisée des contenus et préparation à la diffusion (voir la figure 10.2 pour une représentation schématique de la chaîne de traitement). Dans certains cas, par exemple celui du corpus Mont(REA)L 2016, seuls les fichiers sonores sont disponibles. L'équipe du FDLQ s'occupe alors de l'entièreté du processus, à commencer par la numérisation, si les données sont stockées sur un médium analogique. Dans d'autres, une partie du traitement a déjà été effectuée. Par exemple, dans le cas des corpus Montréal 1971, Montréal 1984 et Montréal 1995, il existe déjà une transcription, mais celle-ci n'est pas temporalisée, c'est-à-dire il n'y a pas d'association structurée entre le son et le texte. L'alignement temporel des transcriptions existantes est alors assuré par l'équipe du FDLQ. Enfin, certains corpus comme PFC-Montréal sont transmis sous une forme presque complète. Il ne reste alors qu'à faire quelques ajustements en vue d'une intégration harmonieuse à la base de données.

FIGURE 10.2

Le traitement des corpus oraux en vue de leur diffusion sur la plateforme



Afin d'assurer la cohérence du Fonds, un certain nombre de principes de transcription sont respectés. Ceux-ci varient légèrement selon que les corpus sont déjà transcrits, ou non, au moment de leur réception. S'ils sont transcrits, la préoccupation principale est le respect des orientations éditoriales établies par les auteurs, ce qui comprend la préservation des conventions de transcription et le maintien des marquages extralinguistiques existants (codages, marquages, etc.). L'intervention du FDLQ dans la transcription se limite alors à la correction de coquilles et d'autres erreurs évidentes. Lorsque les transcriptions n'existent pas déjà, elles sont réalisées par l'équipe du FDLQ à l'aide du logiciel Praat (Boersma et Weenink, 2023), au format TextGrid. Les principes de base guidant cette opération sont la lisibilité des données et leur relative transparence pour un public composé en grande partie de non-spécialistes. Les conventions employées sont fondées sur celles du projet Phonologie du français contemporain (PFC; Durand, Laks et Lyche, 2002), et plus particulièrement celles développées pour son volet québécois (Côté, 2014; Côté et Saint-Amant Lamy, 2023).

Afin d'assurer un équilibre entre la représentation fidèle du contenu du signal sonore et la lisibilité par l'usager, les conventions PFC préconisent une transcription en orthographe standard de l'ensemble des éléments lexicaux effectivement produits par les locuteurs (et seulement de ceux-là). Par exemple, la séquence [ʒāvwajpɔsɔa:bɔs] serait transcrite « j'envoie pas ça à la boss ». Le *ne* de négation n'est pas transcrit puisqu'il est complètement absent du signal, alors que l'article *la* est inclus car une trace de sa présence persiste dans l'allongement du [a]. Suivant les conventions PFC-Québec, la variation de nature phonétique/phonologique n'est pas représentée, mais la variation morphologique l'est. Ainsi, le mot *pas* est transcrit de la même façon, qu'il soit prononcé [pa], [pɔ] ou [pɔ̃]. Toutefois, la distinction entre « envoie » et « envoie » est faite afin de marquer l'opposition entre deux radicaux existants du verbe *envoyer*. Pour faciliter davantage encore la consultation par le public, les transcriptions faites par l'équipe du FDLQ ne comprennent pas de codages ou de marquages idiosyncrasiques¹¹. De plus, dans un souci de contrôle de la qualité, ces transcriptions font toutes l'objet

11. Par exemple, le recours aux codes et marquages dans la transcription (fictive) suivante permet de relever des phénomènes absents de la transcription orthographique standard, comme la liaison, l'intonation et l'alternance codique: « je suis T à veille de dire comme ↑ [angl] that's enough drama for me ». Bien qu'ils communiquent des renseignements intéressants pour les chercheurs, ces éléments graphiques additionnels alourdissent la transcription et rendent son accès par le grand public plus difficile.

d'une révision complète. Les conventions adoptées pour les différents corpus, qu'ils soient déjà transcrits ou non, sont constamment mises à jour dans des guides de transcription détaillés.

Un des principaux intérêts de la plateforme du FDLQ pour l'exploitation des corpus oraux est la possibilité d'écouter, dans son contexte de production, un mot ou une expression tapée dans le moteur de recherche. Cette fonctionnalité nécessite un balisage temporel serré, obtenu par une segmentation des transcriptions en intervalles courts, n'excédant généralement pas 7 ou 8 secondes. Les séquences de silence dépassant 150 millisecondes sont également identifiées et isolées. Enfin, les passages où les voix de plusieurs locuteurs se chevauchent sont clairement délimités afin de bien représenter la temporalité des interactions et de mettre en évidence les séquences où l'environnement acoustique est plus chargé. La segmentation du signal sonore et sa transcription sont réalisées dans des fichiers TextGrid, sous Praat. Lorsque l'entièreté du traitement est effectuée par l'équipe du FDLQ, la transcription et la segmentation sont simultanées. Les corpus comprenant des transcriptions non temporalisées, de leur côté, font l'objet d'une procédure d'alignement particulière en trois étapes : le traitement préalable des transcriptions originales, le transfert du contenu textuel et l'ajustement des rendus finaux. Lors de la phase de prétraitement, le contenu des fichiers originaux est transféré dans un fichier de format TXT où chaque ligne correspond à un tour de parole. Les longs tours de parole dépassant souvent la durée maximale visée pour les intervalles, ils sont donc divisés en séquences plus courtes par l'insertion de ruptures après les points et les virgules. Dans Praat, un membre de l'équipe d'alignement insère ensuite des balises temporelles au long du fichier sonore, balises correspondant à chaque saut de ligne présent dans le fichier TXT. Le contenu textuel peut ensuite être transféré automatiquement du TXT au TextGrid, en associant chaque ligne à un intervalle. Une fois la transcription importée dans Praat, elle est révisée de façon à obtenir le même format que les transcriptions effectuées en totalité par l'équipe du FDLQ (ajustement des frontières d'intervalles, segmentation des intervalles trop longs, isolement des silences, ajouts d'éléments manquants, etc.).

Tous les fichiers, peu importe les particularités associées à leur traitement, font l'objet d'une procédure de finalisation en vue de leur publication. Cette procédure comporte une relecture complète de la transcription, l'anonymisation des informations sensibles, l'archivage des différentes versions des fichiers et le nettoyage des fichiers sonores. La tâche de relecture a pour

but d'assurer que les fichiers définitifs soient de la plus haute qualité. Elle comprend la correction des coquilles restantes, les ajustements nécessaires à l'alignement temporel ainsi que la résolution de problèmes divers soulevés dans les phases précédentes du traitement¹². La personne responsable de la finalisation de chaque fichier doit s'assurer de la protection des renseignements personnels des locuteurs qui y interviennent. La procédure d'anonymisation prévoit l'élimination de tout élément permettant une identification directe (nom, prénom) ou indirecte (adresse, date de naissance, employeur, noms des amis/parents/enfants, etc.). Dans le fichier sonore, cela se fait en remplaçant le passage sensible par un bruit périodique de la même durée et suivant la courbe intonative¹³. Dans le texte, l'élément retiré est remplacé par un terme générique précédé d'un N majuscule. Comme l'illustre bien l'exemple ci-dessous (tableau 10.2), la méthode peut alourdir le texte et rendre plus difficile son interprétation. Pour pallier ces inconvénients et pour conserver la dimension humaine de l'interaction, il pourrait être pertinent de réfléchir à une forme de pseudonymisation qui permettrait minimalement de remplacer les anthroponymes par des équivalents appropriés et vraisemblables, leur choix étant basé sur les données disponibles relativement à la fréquence des prénoms selon l'année de naissance du locuteur¹⁴.

TABLEAU 10.2

Anonymisation des entrevues (exemple fictif)

Version non anonymisée	Ma fille Claudine, elle est gérante au Steinberg puis elle a engagé Mario.
Version anonymisée	Ma fille Nprénom, elle est gérante au Nentreprise puis elle a engagé Nprénom.
Version pseudonymisée	Ma fille {Christiane}, elle est gérante au Nentreprise puis elle a engagé {Alain}.

12. Généralement, la personne qui s'occupe de cette tâche est la troisième à travailler sur le fichier. Elle suit celle ayant effectué la première transcription et celle ayant procédé à la révision (dans le cas des fichiers entièrement traités par le FDLQ) ou à l'alignement (dans le cas des transcriptions initialement non temporalisées). Les différentes personnes intervenant sur un fichier peuvent partager à même le fichier TextGrid leurs questions et leurs incertitudes par rapport à la transcription.

13. L'opération est effectuée avec un script développé par Daniel Hirst (2013).

14. Les données pour la période 1890-1980 sont tirées de Duchesne (1997) et celles pour la période 1980-2020 de la banque de prénoms de Retraite Québec (<https://www.retraite-quebec.gouv.qc.ca/fr/services-en-ligne-outils/banque-de-prenoms>).

Des considérations liées à la gestion de l'espace et aux fonctionnalités de lecture mènent à convertir les fichiers audio (en général au format WAV) au format MP3. Les fichiers sont également nettoyés pour assurer leur facilité d'écoute et leur uniformité. Le nettoyage est effectué dans le logiciel de restauration audio iZotope RX Advanced et comprend, entre autres, une correction du ronflement de secteur, une diminution de la réverbération et du bruit ainsi qu'une normalisation du volume sonore. Les fichiers de transcription sont convertis au format XML. Après ces dernières interventions, les fichiers sont prêts pour une diffusion sur la plateforme FDLQ, étape qui est assurée par l'outil de gestion décrit dans la section suivante.

La gestion des corpus : des métadonnées à la publication des entrevues

La constitution de mégacorpus comprend un ensemble d'activités de gestion de données qui demandent un investissement et des efforts considérables pour assurer la sauvegarde des renseignements généraux et les transcriptions. Il est certain qu'une politique d'accessibilité des données pour d'autres chercheurs contribue à assurer un rendement satisfaisant, compte tenu du travail fourni. (Thibault et Vincent, 1990, 29)

La gestion des contenus diffusés sur le FDLQ constitue un autre défi important dans la réalisation du projet, ne serait-ce qu'en raison de l'hétérogénéité des corpus, du nombre élevé de documents concernés ou encore de la confidentialité de certaines données et des droits d'auteur à respecter. Cette gestion est assurée par une interface spécialement conçue pour les besoins du projet, baptisée Comet. Il s'agit d'un outil collaboratif qui se présente sous la forme d'une application Web accessible à l'aide d'un navigateur. Comet permet plus particulièrement de consulter, d'éditer et de publier les différents corpus du Fonds (et les documents qui les constituent), mais aussi de gérer l'ensemble de leurs métadonnées. Les interfaces Comet et FDLQ communiquent entre elles par un intergiciel à messages. Lorsqu'un document est publié par un administrateur, Comet envoie un message à la plateforme FDLQ qui constitue son canal de diffusion. L'outil est également doté d'un moteur de recherche qui permet de faire des requêtes textuelles dans les corpus.

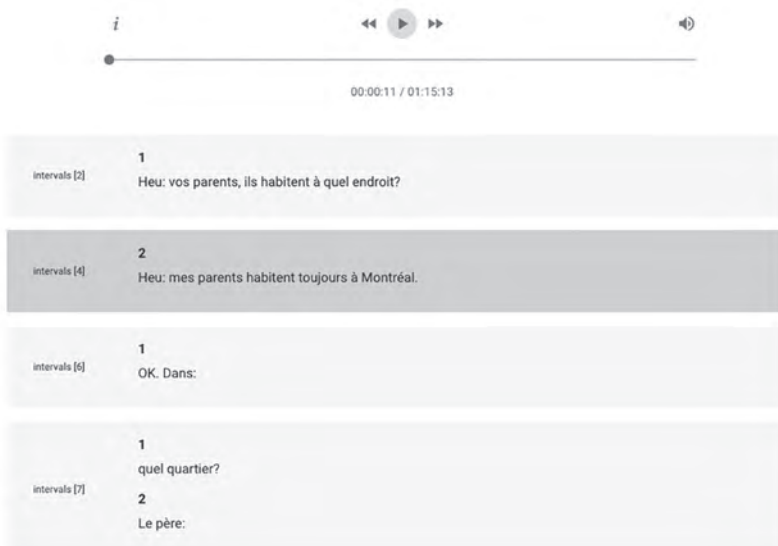
L'interface Comet permet de gérer conjointement les documents sources (fichiers audio) et les transcriptions (fichiers textuels) de toutes les entrevues

faisant partie des corpus oraux. Lorsqu'une entrevue est jugée prête en vertu du protocole de traitement présenté dans la section précédente, les différents fichiers ainsi que les métadonnées qui s'y rattachent sont importés. Cette importation peut se faire entrevue par entrevue ou encore en lot, lorsqu'on veut importer un corpus complet (ou un sous-corpus composé de plusieurs entrevues). Tous les fichiers sont stockés sur un serveur sécurisé, hébergé à l'Université de Sherbrooke. La transcription peut être visualisée de différentes manières – sous forme d'intervalle (voir figure 10.3a), ou sous forme de dialogue, avec des tours de parole qui alternent (voir figure 10.3b) – grâce à une feuille de style CSS qui est adaptée aux particularités de chaque corpus.

FIGURE 10.3

Visualisation sous forme d'intervalle (a) ou sous forme de dialogue (b) de la transcription de l'entrevue menée auprès de Bertrand P. (Corpus Montréal 1984)

a)



b)



L'alignement temporel permet par ailleurs d'écouter l'entrevue en suivant la transcription en temps réel. Cette fonctionnalité est très utile dans la mesure où elle permet de retourner facilement dans l'enregistrement original lorsqu'on veut écouter la prononciation d'un mot ou encore lorsqu'on veut valider un élément de la transcription. Lorsqu'un collaborateur détecte une erreur ou une coquille, il est facile de la corriger à l'aide de l'éditeur XML intégré à l'interface Comet et de publier la version corrigée. Il est également possible d'ajouter des pièces jointes à un corpus ou à une entrevue. Cette fonctionnalité peut être utile pour conserver des documents qui témoignent des différentes étapes dans la constitution du corpus, comme le protocole d'édition, ou pour archiver d'autres versions du fichier audio.

La gestion des métadonnées est au cœur des fonctionnalités de Comet. Dans la mesure où le projet adhère aux principes définis par l'initiative FAIR (<go-fair.org>) – qui encourage à rendre les données scientifiques faciles à trouver, accessibles, interopérables et réutilisables –, les contenus diffusés sur la plateforme doivent en effet être dotés de données de référence riches, structurées et pérennes. L'outil permet d'associer des métadonnées autant à des corpus qu'aux documents qui les constituent, qu'il s'agisse de métadonnées descriptives (l'auteur, la date et le titre d'un document, par exemple, ou encore les métadonnées de nature sociologique et démographique relatives

aux locuteurs enregistrés¹⁵), structurelles (comme l'appartenance du document à un corpus précis) ou administratives (informations concernant les droits d'accès ou les droits d'auteur, par exemple, ou concernant l'état d'un document dans le processus d'édition).

Comet est accessible seulement aux personnes autorisées par le responsable du projet. Contrairement à la plateforme FDLQ, qui ne requiert pas d'authentification (pour la simple raison que l'affichage des données se fait suivant les limites autorisées par les ententes signées), l'interface Comet est dotée de plusieurs niveaux d'autorisation qui permettent de contrôler minutieusement l'accès aux données. Les principaux niveaux d'accès concernent la lecture (consultation) et l'exploration (requêtes textuelles) des documents (accès minimal), l'édition (ajout d'un corpus ou d'un document, ajout de métadonnées, intervention au besoin dans la transcription, etc.), l'approbation et la diffusion. Seuls les administrateurs du projet ont la possibilité de publier un document ou un corpus sur la plateforme FDLQ et de gérer les accès accordés.

L'exploitation des données : l'exemple des mots en /ar#/

Le bien durable a été obtenu en raison d'une problématique justifiant la captation d'un ensemble de données qui pourront être observées sous différents angles, au-delà de ce qui avait été prévu par les concepteurs du projet. Si l'entreprise est coûteuse, elle se rentabilise par la quantité d'observables qu'elle offre aux chercheurs et par l'étalement dans le temps et dans l'espace de son utilisation. (Vincent, 2009 : 86)

Pour les linguistes, les bases documentaires de grande taille, comme le FDLQ, présentent l'avantage indéniable de permettre l'étude de phénomènes moins fréquents, difficiles à cerner de façon satisfaisante dans des ensembles plus limités. Afin d'illustrer les nouvelles possibilités de recherche ainsi créées, nous présentons ici l'analyse d'un phénomène de ce type : la neutralisation des voyelles basses antérieure et postérieure devant un /r/ final. Nous limitons cette analyse aux corpus montréalais du Fonds.

15. Les données sociodémographiques (âge, scolarité, profession, etc.) peuvent déjà être consultées dans l'interface Comet par les chercheurs qui ont les autorisations requises, mais ne sont pas encore disponibles sur la plateforme du Fonds. L'ajout est prévu dans une version ultérieure qui sera diffusée depuis le site Web du projet.

Si l'opposition /a/ - /ɒ/ se maintient bien en français québécois (ex. *patte-pâte, mal-mâle*), on observe une tendance à favoriser la postérieure devant une rhotique finale (ex. *égare* [egaʁ] → [egɒʁ], *égard* [egɒʁ]; voir Côté, 2012, 2016). Côté évoque comme facteur contributif à ce phénomène l'important déséquilibre statistique qui existe entre l'ensemble, très limité, des mots où l'on trouve traditionnellement la voyelle antérieure (catégorie /aʁ#/; ex. *gare, guitare, prépare*) et celui, beaucoup plus important, où l'on trouve la voyelle postérieure (catégorie /ɒʁ#/; ex. *phare, départ, part*). Or, ce déséquilibre représente un obstacle de taille pour l'exploration du phénomène. Par exemple, dans une étude des voyelles prérhotiques produites en parole spontanée, Saint-Amant Lamy (2022) ne recense que 42 occurrences de mots appartenant à la catégorie /aʁ#, contre 623 occurrences pour /ɒʁ# (sur un total de plus de 13 000 voyelles prérhotiques). Dans ces conditions, les possibilités d'analyses sont évidemment très limitées.

Une solution possible au problème de la rareté consiste à amener délibérément les locuteurs à produire les variables étudiées, par exemple dans le cadre d'une tâche de lecture. C'est l'approche adoptée par Côté (2012, 2016), qui met en lumière une distinction générationnelle claire dans les données de PFC-Trois-Rivières: alors que les locuteurs plus âgés (nés entre 1931 et 1958) favorisent nettement [a] dans les mots de la catégorie /aʁ#, les plus jeunes (nés entre 1986 et 1989) adoptent à divers degrés la variante [ɒ]¹⁶. Toujours à partir d'un exercice de lecture, Michaud et Remysen (2025), observent aussi une utilisation nettement plus fréquente de [ɒ] pour ces mots à Sherbrooke et à Victoriaville chez les jeunes de moins de 30 ans, nés depuis 1990. La tâche de lecture a toutefois le désavantage de favoriser l'hypercorrection, associée à la variante antérieure. Ainsi, deux locutrices de l'échantillon de Côté (2012, 2016) ont dû être exclues en raison de leur propension à antérioriser toutes les voyelles basses, y compris dans les mots de la catégorie /ɒʁ# (ex. *homard* [omaʁ]).

En agrégeant de grandes quantités de données orales issues d'entrevues semi-dirigées ou de conversations spontanées, le FDLQ permet de régler en bonne partie le problème de la rareté et d'éviter celui de l'hypercorrection en contexte de lecture. Nous avons ainsi effectué des recherches dans le Fonds pour chacun des mots de la classe /aʁ#/ identifiés dans Dumas (1986, 1987)

16. Quatre jeunes sur cinq favorisent [ɒ] dans *guitare* et *gare*, deux d'entre eux optent aussi pour cette voyelle dans *bulgare*. Le jeune locuteur JG la produit également dans *prépare*, ce qui suggère que la neutralisation est totale chez lui.

et Côté (2012, 2016), soit *bagarre*, *bulgare*, *cigare*, *cithare*, *déclare*, *démarre*, *gare*, *guitare*, *tare* et tous les verbes se terminant par le morphème *pare* (ex. *prépare*, *sépare*, *compare*, *répare*, *accapare*). Ces requêtes ont permis d'identifier 237 occurrences dans les corpus Montréal 1971, Montréal 1984, Montréal 1995, PFC-Montréal et Mont(REA)L 2016, qui totalisent ensemble 4,24 millions de mots. Le nombre de cas reste modeste, mais il s'avère suffisant pour une analyse générale des facteurs sociaux et linguistiques qui conditionnent la variation.

Les 237 voyelles ont été classées de façon impressionniste par accord inter-juge¹⁷. Seules deux catégories ont été distinguées : [a] et [ɒ]. L'exercice s'est avéré plutôt aisé, puisque les deux voyelles s'opposent non seulement par le timbre, mais souvent aussi par la durée et la diphtongaison¹⁸. Trois facteurs temporels sont disponibles pour l'analyse : l'année de naissance des locuteurs, leur âge au moment de l'entrevue et l'année de l'enregistrement. Deux autres facteurs externes ont été étudiés, soit le sexe et la catégorie professionnelle. Le classement de la catégorie professionnelle est tiré de Thibault et Vincent (1990), mais rapporté sur une échelle de 1 à 3, soit : 1-cols bleus, 2-cols blancs, 3-cadres et professionnels. Les facteurs internes retenus pour l'analyse sont la catégorie grammaticale (nom, verbe) et la fréquence du lexème dans l'ensemble des corpus. La catégorie grammaticale est retenue car elle est associée à deux réalités morphologiques favorisant potentiellement le maintien du [a] dans la catégorie verbale : la domination du morphème *-pare*, qui pourrait assurer une certaine robustesse paradigmatique, et l'existence de nombreuses formes dérivées où la neutralisation vers [ɒ] est moins productive (ex. *séparer*, *réparais*, *comparez*, *préparerai*). La prise en compte de la fréquence permet de vérifier si un phénomène de diffusion lexicale est à l'œuvre.

Avant de procéder à la modélisation du phénomène de neutralisation, une question doit être posée quant à la nature de celui-ci : s'agit-il d'un cas de gradation selon l'âge ou d'un changement en cours ? Les études antérieures, basées sur des données synchroniques, ne permettent pas d'apporter une réponse définitive. La présence dans le FDLQ de corpus montréalais élaborés à différentes époques rend toutefois la chose possible. Une série de régressions logistiques a été effectuée afin de déterminer la proportion de la

17. Dans les cas de désaccord entre les deux auteurs, une locutrice de français québécois ayant une formation en phonologie (Marie-Pier Picard) a été sollicitée pour trancher.

18. La voyelle [a] est la seule voyelle ne pouvant être diphtonguée en syllabe finale fermée par /ʁ/.

variance [aʁ#] - [ɔʁ#] expliquée par chacun des facteurs temporels (voir tableau 10.3). On constate que l'année de naissance est de loin le facteur le plus explicatif, ce qui indique qu'un changement est en cours¹⁹. L'âge au moment de l'entrevue et l'année de l'enregistrement ont donc été écartés de l'analyse qui suit.

TABEAU 10.3

Portion de la variance expliquée par chacun des facteurs temporels

Facteur	R ² de Nagelkerke
Année de naissance	0,239
Âge lors de l'entrevue	0,126
Année de l'enregistrement	0,066

Les statistiques descriptives concernant les facteurs catégoriques sont présentées à la figure 10.4 et celles pour les facteurs continus, à la figure 10.5. Afin de dégager les tendances générales dans les données, nous avons d'abord testé la relation statistique existant entre chaque facteur et le taux d'emploi de la voyelle postérieure, notamment grâce à des tests t et des mesures de corrélation. La corrélation entre l'année de naissance et le taux de réalisations [ɔʁ#] a été mesurée à partir des moyennes de chaque locuteur. Dans le cas de la fréquence lexicale, la corrélation a été mesurée à partir des moyennes pour chaque lexème. Les tests suggèrent que les variables jouant un rôle significatif sont la catégorie professionnelle, la classe grammaticale et l'année de naissance. Des comparaisons par paire pour la catégorie professionnelle montrent qu'il n'y a pas de distinction significative entre les cols bleus et les cols blancs, mais que ces deux groupes se distinguent de la catégorie des cadres et professionnels ($p < 0,001$). Aucune tendance claire n'émerge pour le sexe et la fréquence lexicale. En ce qui a trait à la classe grammaticale, il semble que les noms soient nettement plus propices à la neutralisation que les verbes.

Nous avons par la suite élaboré un modèle de régression logistique binaire (fonction *glm* dans R) afin de mesurer de façon intégrée le rôle des différents facteurs de variation. La catégorie vocalique (0 = [aʁ#], 1 = [ɔʁ#]) a été retenue comme variable dépendante, et l'année de naissance du locuteur, son sexe

19. Dans un cas de gradation selon l'âge, le facteur le plus explicatif serait l'âge au moment de l'entrevue.

FIGURE 10.4

Taux de réalisation du type [nɛ#] en fonction du sexe, de la catégorie professionnelle et de la classe grammaticale

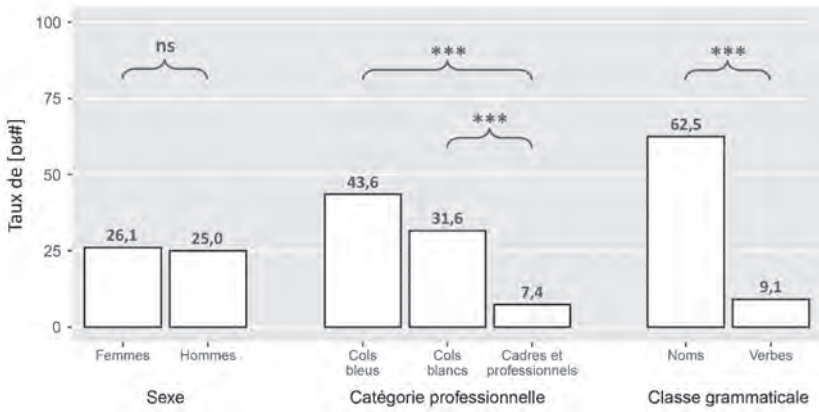
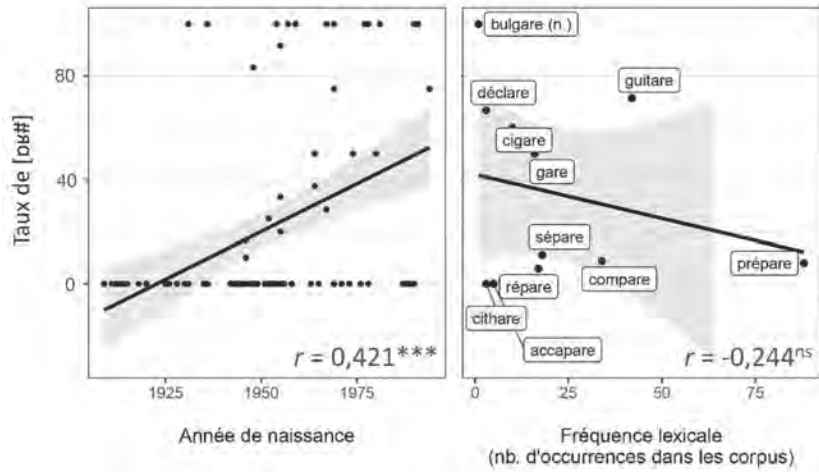


FIGURE 10.5

Taux de réalisation du type [nɛ#] en fonction de l'année de naissance et de la fréquence lexicale



(F = 0, H = 1), sa catégorie professionnelle, la fréquence lexicale du mot et la catégorie grammaticale de celui-ci (nom = 0, verbe = 1), comme variables indépendantes. Le modèle de régression logistique s'avère significatif²⁰ et permet d'expliquer 55,9 % de la variance (R^2 de Nagelkerke) et de classer correctement 87,3 % des observations. Lorsque les facteurs sont pris en compte ensemble, on relève un effet significatif de l'année de naissance, de la catégorie professionnelle et de la classe grammaticale (voir tableau 10.4). Pour l'année de naissance, les chances que la variante postérieure soit produite augmentent d'environ 5 % par année. Elles diminuent toutefois d'environ les deux tiers pour chaque échelon dans la catégorisation professionnelle. Enfin, les chances que la variante [ɒʁ#] soit réalisée sont environ 20 fois plus élevées dans les noms que dans les verbes. Le sexe et la fréquence lexicale n'ont pas d'effet significatif sur le choix de la voyelle.

TABLEAU 10.4

Tableau récapitulatif de la régression logistique binaire

Variable	B	Erreur type	p	Rapport des cotes (RdC)	IC 95 % pour RdC	
					Borne inf.	Borne sup.
Année	0,049	0,012	<0,001	1,051	1,027	1,077
Catégorie professionnelle	-1,094	0,328	<0,001	0,335	0,170	0,624
Classe grammaticale (verbe)	-3,055	0,517	<0,001	0,047	0,016	0,123
Fréquence lexicale	-0,002	0,008	0,793	0,998	0,982	1,014
Sexe (homme)	-0,904	0,483	0,061	0,405	0,151	1,024

Ces résultats permettent de reconstituer sommairement l'évolution de la neutralisation de /a/ et /ɒ/ devant un /ʁ/ final (voir figure 10.6). Il semble que le processus débute avec quelques locuteurs nés dans le deuxième quart du 20^e siècle et appartenant aux catégories des cols bleus et des cols blancs²¹. Les cadres et les professionnels suivent avec un décalage d'une génération. Ce schéma pourrait suggérer que le changement se situait initialement sous

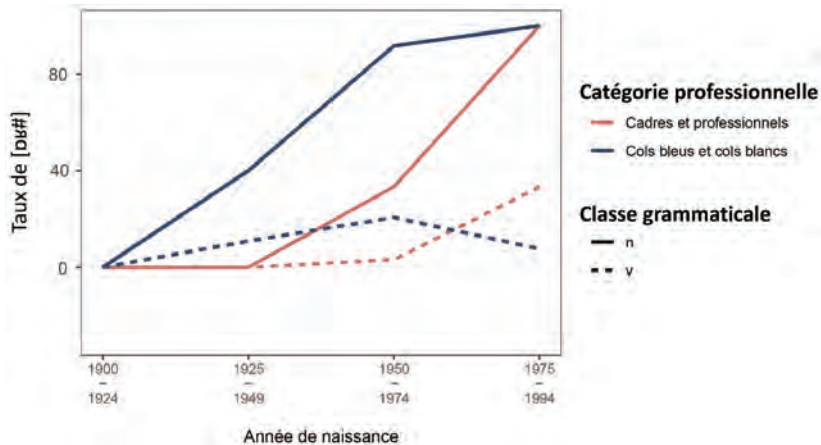
20. $\chi^2(5, N = 237) = 22,337, p < 0,001$

21. Les deux catégories ont été combinées à la figure 10.6 puisqu'aucune différence significative dans l'usage des variantes [aʁ#] et [ɒʁ#] n'a pu être identifiée avec les tests t.

le seuil de la conscience. Le changement s'observe essentiellement dans les noms (ex. *gare, guitare, cigare*), la réalisation [aʁ#] demeurant largement dominante dans les verbes (ex. *prépare, sépare*). Il est possible que l'alternance du radical en /aʁ#/ avec des formes dérivées, où le [a] se maintient de façon plus robuste, ait contribué à freiner la neutralisation dans cette dernière catégorie. Toutefois, chez les jeunes cadres et professionnels, la variante [ɔʁ#] semble gagner du terrain au sein de la classe verbale, avec un tiers des occurrences étudiées. Cette tendance est à confirmer (elle est ici basée sur un petit nombre d'observations), mais elle pourrait indiquer qu'un changement dans le statut de la variante postérieure est en cours. La neutralisation étant complétée (et le conditionnement social éliminé) pour la catégorie nominale, il est probable que la pression systémique en faveur de la forme [ɔʁ#] soit devenue suffisamment forte pour contrebalancer les facteurs ayant jusque-là permis le maintien de [aʁ#] dans le paradigme *prépare-sépare-répare*.

FIGURE 10.6

Taux de réalisation du type [ɔʁ#] en fonction de la classe grammaticale et de la catégorie professionnelle



* * *

Alors, nous avons voulu faire quelque chose comme ça, sur le français de Montréal, une grosse étude sociolinguistique. C'était des folies de jeunesse! (Henrietta Cedergren, citée dans Daveluy et Laforest, 1994, 12)

Outre qu'elle pérennise des collectes de données dont la valeur patrimoniale n'est plus à démontrer, la plateforme FDLQ permet de mutualiser des données qui, autrement, auraient pu rester fragmentées et dispersées. Grâce aux « folies de jeunesse » des sociolinguistes québécoises de la première heure et aux progrès récents des humanités numériques, nous sommes en mesure de rassembler des données en nombre considérable et de reconstituer certaines étapes dans l'évolution récente du français au Québec. Le FDLQ ouvre ainsi la porte à de nouvelles avancées en matière d'analyse, par la taille des corpus rassemblés ainsi que par la profondeur chronologique couverte par les données réunies. La plateforme favorise aussi les analyses comparatives, puisqu'elle permet de rapprocher et de croiser plus facilement des données recueillies dans différentes régions du Québec.

Nous avons illustré les nombreuses possibilités d'analyse offertes par le FDLQ à partir de l'exemple d'un phénomène de prononciation en particulier, la neutralisation des voyelles basses en position prérhotique. S'il reste beaucoup à découvrir sur ce phénomène, en particulier sur ses conditions d'apparition, il apparaît nettement qu'une meilleure compréhension repose, d'abord et avant tout, sur une meilleure description du phénomène. La rareté des occurrences des mots appartenant à la classe /aʁ#/ dans la parole spontanée avait jusqu'ici rendu cette tâche difficile. Un outil comme le FDLQ peut alors se révéler une aide précieuse pour la description et l'analyse de ce genre de phénomènes peu fréquents.

Références

- BLONDEAU, Hélène (2014), La nature métropolitaine du Montréal « d'aujourd'hui » et le français « d'ici », dans Wim REMYSEN (dir.), *Les français d'ici: du discours d'autorité à la description des normes et des usages*, Québec, Presses de l'Université Laval, 205-239.
- BLONDEAU, Hélène (2020), Du cheminement sociolinguistique des individus: l'apport des études longitudinales à l'étude de la variation et du changement en français parlé, dans Wim REMYSEN et Sandrine TAILLEUR (dir.), *L'individu et sa langue: hommages à France Martineau*, Québec, Presses de l'Université Laval, 13-46.

- BLONDEAU, Hélène et Marty LAFOREST (2015), *Bibliographie des travaux fondés sur les données des grands corpus sociolinguistiques de Montréal*, <https://fdlq.recherche.usherbrooke.ca/bibliographie-travaux-grands-corpus-sociolinguistiques-mtl.html>.
- BOERSMA, Paul et David WEENINK (2023), *Praat, doing phonetics by computer*, version 6.3.10., <http://www.praat.org/>.
- CÔTÉ, Marie-Hélène (2012), Laurentian French (Quebec): extra vowels, missing schwas and surprising liaison consonants, dans Randall GESS, Chantal LYCHE et Trudel MEISENBURG (dir.), *Phonological variation in French: illustrations from three continents*, Amsterdam, John Benjamins, 235-274.
- CÔTÉ, Marie-Hélène (2014), Le projet PFC et la géophonologie du français laurentien, dans Jacques DURAND, Gjert KRISTOFFERSEN et Bernard LAKS (dir.), *La phonologie du français: normes, périphéries, modélisation*, Nanterre, Presses universitaires de Paris Ouest, 175-198.
- CÔTÉ, Marie-Hélène (2016), Variation in Canada: Trois-Rivières in Quebec, dans Sylvain DETEY, Jacques DURAND, Bernard LAKS et Chantal LYCHE (dir.), *Varieties of spoken French*, Oxford, Oxford University Press, 449-462.
- CÔTÉ, Marie-Hélène et Hugo SAINT-Amant LAMY (2023), The Phonologie du français contemporain project in Quebec: methodological and dialectometric considerations, dans Elissa PUSTKA, Carmen QUIJADA VAN DEN BERGHE et Verena WEILAND (dir.), *Corpus dialectology: from methods to theory (French, Italian, Spanish)*, Amsterdam/Philadelphie, John Benjamins, 60-83.
- DAVELUY, Michelle (2005), *Les langues étendards: allégeances langagières en français parlé à Montréal*, Québec, Éditions Nota Bene.
- DAVELUY, Michelle et Marty LAFOREST (1994), Entretiens: *Tout ça a commencé en 1970...*, *Culture*, 14(2), 11-26.
- DORAN, Michèle, Lyne DRAPEAU et Claire LEFEBVRE (1982), Le projet Centre-Sud: corpus et méthodologie, dans Claire LEFEBVRE (dir.), *La syntaxe comparée du français standard et populaire: approches formelles et fonctionnelles*, tome 2, Québec, Office québécois de la langue française, 401-420.
- DUCHESNE, Louis (1997), *Les prénoms, des plus rares aux plus courants au Québec*, Saint-Laurent, Trécaré.
- DUGAS, André (1986), *Le corpus Bibeau-Dugas*, document interne, Université du Québec à Montréal.
- DUMAS, Denis (1986), Le statut des « deux a » en français québécois, *Revue québécoise de linguistique*, 5(2), 167-196.
- DUMAS, Denis (1987), *Nos façons de parler: les prononciations en français québécois*, Québec, Presses de l'Université Laval.
- DURAND, Jacques, Bernard LAKS et Chantal LYCHE (2002), La phonologie du français contemporain: usages, variétés et structure, dans Claus D. PUSCH et Wolfgang RAIBLE (dir.), *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache*, Tübingen, Gunter Narr, 93-106.
- FRIESNER, Michael (2009), *The Social and Linguistic Predictors of the Outcomes of Borrowing in the Speech Community of Montréal*, thèse de doctorat, University of Pennsylvania.

- GADET, Françoise et France MARTINEAU (2017), Le maillage du français en Amérique du Nord, dans un cadre de francophonie, dans Beatrice BAGOLA et Ingrid NEUMANN-HOLZSCHUCH (dir.), *L'Amérique francophone: carrefour culturel et linguistique*, Bern, Peter Lang, 11-40.
- HIRST, Daniel (2013), Anonymising long sounds for prosodic research, dans Brigitte BIGI et Daniel HIRST (dir.), *Tools and Resources for the Analysis of Speech Prosody*, Aix-en-Provence, Laboratoire Parole et Langage, 36-37.
- MARTINEAU, France et Marie-Claude SÉGUIN (2016), Le Corpus FRAN: réseaux et maillages en Amérique française, *Corpus*, 15, 55-87.
- MICHAUD, Josiane et Wim REMYSEN (2025), L'analyse géophonologique du français québécois et la diffusion de quelques traits de prononciation dans le Centre du Québec, dans Luc BARONIAN et Sandrine TAILLEUR (dir.), *Les français d'ici en mouvement: hier et maintenant*, Québec, Presses de l'Université Laval, 323-346.
- REA, Beatrice (2020), « Je m'ai fait mal quand j'ai tombé »: a real- and apparent-time study of auxiliary alternation in intransitive and pronominal verbs in spoken Montréal French (1971-2016), thèse de doctorat, University of Oxford.
- SAINT-AMANT, Lamy (2022), *Initiation des changements phoniques: le cas de l'écosystème rhotique du français laurentien*, thèse de doctorat, Université de Lausanne.
- SANKOFF, David, Gillian SANKOFF, Suzanne LABERGE et Marjorie TOPHAM (1976), Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale, *Cahier de linguistique*, 6, 85-125.
- SANKOFF, Gillian (2018), Before there were corpora: the evolution of the Montreal French Project as a longitudinal study, dans Suzanne EVANS WAGNER et Isabelle BUCHSTALLER (dir.), *Panel studies of variation and change*, New York, Routledge, 21-51.
- SANKOFF, Gillian, Pierrette THIBAUT, Naomi NAGY, Hélène BLONDEAU, Marie-Odile FONOLLOSA et Lucie GAGNON (1997), Variation in the use of discourse markers in a language contact situation, *Language Variation and Change*, 9(2), 191-218.
- THIBAUT, Pierrette et Gillian SANKOFF (1993), Diverses facettes de l'insécurité linguistique: vers une analyse comparative des attitudes et du français parlé par des Franco- et des Anglo-Montréalais, *Cahiers de l'Institut de linguistique de Louvain*, 19(34), 209-2019.
- THIBAUT, Pierrette et Diane VINCENT (1990), *Un corpus de français parlé. Montréal 84: historique, méthodes et perspectives de recherche*, Québec, Département de langues et linguistique, Université Laval.
- VINCENT, Diane (2009), Corpus, banques de données, collections d'exemples: réflexions et expériences, *Cahiers de linguistique*, 33(2), 81-96.
- VINCENT, Diane, Marty LAFOREST et Guylaine MARTEL (1995), Le corpus de Montréal 1995: adaptation de la méthode d'enquête sociolinguistique pour l'analyse conversationnelle, *Dialangua*, 6, 29-45.

CHAPITRE 11

Moncton comme pôle urbain francophone

Quels corpus pour décrire le français
en usage ici et maintenant ?

Laurence Arrighi, Basile Roussel et Isabelle Violette

Si Montréal est indéniablement la « métropole et capitale culturelle du Canada français, ainsi [qu’une] ville internationale et cosmopolite attrayante » (Breton-Carbonneau et Heller, 2021, 84), d’autres centres urbains au Canada constituent eux aussi des pôles de francophonie dynamiques, attractifs et singulièrement intéressants pour la sociolinguistique. Moncton, principal centre urbain du Nouveau-Brunswick, présente une situation remarquable pour l’Acadie, et, au-delà, pour la francophonie, dans la mesure où, depuis les années 1970, la ville n’a cessé d’accroître son pouvoir d’attraction et son rayonnement comme pôle francophone au niveau local, régional, national et international. L’une des manifestations de cette importance croissante de Moncton comme symbole contemporain de l’Acadie se manifeste dans le fait que le vernaculaire local, communément nommé le chiac (Arrighi, 2020 ; Boudreau, 2012), est parfois présenté comme la langue de l’ensemble de la population acadienne, tandis que les pratiques linguistiques en Acadie, y compris au sein même de la ville de Moncton, sont largement, et sans doute de plus en plus, hétérogènes.

Dans ce contexte, Moncton s’avère un objet d’étude de premier plan pour une meilleure connaissance de la francophonie canadienne au 21^e siècle. Toutefois, si cette ville a connu, grâce à l’existence du CRLA (anciennement Centre de recherche en linguistique appliquée et aujourd’hui Centre de

recherche sur la langue en Acadie), un travail de documentation des pratiques linguistiques par la confection de nombre de corpus, la plupart datent désormais d'au moins 20 ans. De plus, presque tous ont été bâtis à partir de présupposés méthodologiques et de questions de recherche qui ne permettent pas toujours de répertorier les pratiques linguistiques dans toute leur diversité. Cet état de fait atteste de la nécessité d'impulser de nouveaux chantiers de confection de corpus. Quelles peuvent être alors les nouvelles avenues pour en faire une récolte à la fois plus écologique dans ses méthodes de collecte, et plus représentative de la diversité de la francophonie monctonienne actuelle ?

Moncton métropole

La recevabilité de Moncton comme objet urbain et, qui plus est, métropolitain, pose question. Moncton¹ est une agglomération de 171 608 habitants², située au sud-est de la petite province excentrée du Nouveau-Brunswick, province qui affiche le taux de ruralité le plus élevé au pays (37 % selon Landry, 2021, p. 17). À première vue, Moncton ne satisfait pas aux critères habituels retenus pour la définition d'une métropole (ou grande ville) : concentration du pouvoir, densité démographique, diversité économique, hétérogénéité de la population, aire d'influence. Or, Moncton présente des spécificités, notamment sociolinguistiques, et a connu, à partir des années 1990, une croissance démographique et un développement économique qui lui ont progressivement valu de devenir un centre urbain d'importance, ainsi qu'un objet de travaux scientifiques. En quelques décennies, elle est passée de ville industrielle ferroviaire au bord de la faillite à chef de file de la nouvelle économie, et « plaque tournante des Maritimes » (Entreprise Grand Moncton, cité dans LeBlanc, 2014, p. 157). La transformation a été telle que le *New York Times* a qualifié le phénomène de « miracle monctonien » (Allain et Chiasson, 2016, p. 16). La devise de la ville *Resurgo* (« Je me relève »), adoptée à la suite d'un premier revirement économique concernant

1. Le Grand Moncton est une agglomération urbaine (région métropolitaine de recensement) regroupant, par ordre d'importance, trois municipalités : Moncton, Dieppe et Riverview. Le centre originel de la ville étant Moncton, et Dieppe et Riverview constituant ses banlieues (bien que ce qualificatif soit de plus en plus contesté, notamment par Dieppe), la dénomination Moncton tend parfois à désigner le centre urbain dans son ensemble.

2. Selon les chiffres de Statistique Canada cités dans « Moncton et Halifax sont les deux villes qui grandissent le plus vite au pays », Radio-Canada, Ici Nouveau-Brunswick, 11 janvier 2023.

tant les chantiers navals que les chantiers ferroviaires, à la fin du 19^e siècle, en fait foi. Cette renaissance contemporaine fut attribuée en grande partie au dynamisme de la présence acadienne dans la ville : les travaux du sociologue Greg Allain (2006 et avec Chiasson, 2016) ont montré l'impact de l'entrepreneuriat acadien et du développement culturel acadien sur l'économie régionale. Cette renaissance est également due à la présence d'une main-d'œuvre bilingue, en forte majorité acadienne, dans le secteur des télécommunications (les centres d'appels), qui est devenu un moteur économique clé pour attirer des nouvelles entreprises et a valu à Moncton le sobriquet de *phone city* (Dubois, LeBlanc et Beaudin, 2006).

Bien que Moncton ne puisse pas figurer comme métropole dans le sens classique du terme, à partir des années 2000, tout un ensemble de travaux de recherche se consacre à sa réalité métropolitaine, vécue à plus petite échelle. Dès lors certains auteurs n'hésitent pas à en parler comme d'une « mini-métropole » (Desjardins, 2002), tandis que d'autres s'attardent au processus de « métropolisation » (Allain, 2006), et que d'autres encore qualifient le lieu de « microcosmopolite » (Bruce, 2005). On recense à l'heure actuelle un important corps de travaux sur Moncton consacrés à divers phénomènes propres à l'urbanité (migration, gouvernance, mobilité, économie, identité)³.

On peut aussi retenir, à l'instar de Blondeau (2014), une définition sociolinguistique du terme « métropole », à savoir un lieu de contact entre groupes linguistiques, exerçant une force d'attraction migratoire. C'est le cas de Moncton qui constitue un terrain de contacts entre les anglophones et les francophones au sein de la province, et ce, depuis la fondation de la ville, dans la deuxième moitié du 19^e siècle. C'est par ailleurs le centre urbain qui attire le plus grand nombre de migrants francophones. Ainsi, le caractère urbain de Moncton revêt une importance encore plus grande lorsque l'on tient plus précisément compte de la population francophone.

3. Pensons notamment aux numéros thématiques de revues suivants, qui regroupent plusieurs contributions au sujet de Moncton : « Urbanité et durabilité des communautés francophones du Canada » dans *Francophonies d'Amérique* (2006, n° 22) et « Le français en milieu minoritaire, défis et enjeux. La situation du français au Nouveau-Brunswick » dans *Minorités linguistiques et société* (2014, n° 4).

Moncton francophone

Que Moncton mérite une attention comme objet urbain ne va pas forcément de soi, mais que Moncton serait, en outre, une ville francophone relève sans doute d'une proposition que d'aucuns jugeront incongrue. Quand il s'agit de francophonie, Moncton a longtemps été pensée, et continue à l'être, à travers le prisme de la minorisation du français face à l'anglais, langue historiquement, démographiquement, socialement et symboliquement dominante. La ville est alors ressentie comme un « lieu de perte linguistique » (Boudreau, 2021, p. 101) pour les francophones qui s'y retrouvent en contexte minoritaire, contrairement au milieu rural qui abrite des communautés francophones homogènes⁴. Les représentations négatives de l'urbanité ne manquent pas dans l'imaginaire collectif acadien : le vécu francophone y serait marqué par l'exiguïté, la fragmentation et l'aliénation, ce qui fait écho d'ailleurs à l'histoire de la ville marquée par une hostilité affichée face au français. Selon ce point de vue, le français est soit en perte (par le phénomène du transfert vers l'anglais, dû en grande partie à l'exogamie que favorise la ville), soit en dégradation en raison des interférences avec l'anglais (dont la manifestation suprême serait le *chiac*, souvent réduit à sa part d'hybridité ; voir Arrighi et Violette, 2013 et Arrighi et Urbain, 2016-2017). Les discours sur la faible qualité du français parlé à Moncton sont légion et historiquement attestés⁵ depuis la fin du 19^e siècle (Boudreau, 2021).

Toutefois, depuis les années 1990, cet espace francophone urbain prend de l'expansion, le français y est plus visible. La vie en français présente une certaine complétude institutionnelle. La ville dispose de plusieurs institutions francophones dans lesquelles des analystes des situations linguistiques minoritaires canadiennes voient la condition du maintien et de l'épanouissement des communautés (Forgues *et al.*, 2020). À Moncton, la population francophone jouit d'un réseau d'écoles en pleine expansion, d'une université unilingue française qui attire une clientèle internationale (30 % des étudiantes et étudiants n'y sont pas originaires du Canada). La région offre aussi des possibilités intéressantes d'emplois aux francophones qui ont des compé-

4. Les recherches de Beaudin et Landry (2003) et de Landry, Allard et Deveau (2005) font ressortir les effets de l'urbanisation sur la francophonie minoritaire. Selon leur analyse, les contacts accrus avec la majorité menacent la vitalité du fait français.

5. Du fait que les premières études sur le français acadien ont surtout mis l'accent sur sa nature conservatrice dite archaïque, des régions urbaines telles que Moncton ont longtemps été négligées en tant que terrains d'enquête (voir Gesner, 1986).

tences en anglais (ex. au sein de la fonction publique fédérale ou dans le domaine de la santé) ou qui n'en ont pas (dans le domaine de l'éducation). C'est encore à Moncton que se trouve l'antenne francophone des provinces atlantiques du radio et télédiffuseur public national, Radio-Canada. Enfin, Moncton propose une vie culturelle en français assez prolifique : un Festival international du cinéma francophone en Acadie (FICFA) ainsi qu'un festival littéraire bilingue (Festival Frye). S'ajoutent aussi à ces événements récurrents des institutions culturelles pérennes, telles que le Centre culturel Aberdeen, le théâtre l'Escaouette, la Galerie d'art de l'Université de Moncton (GAUM), la Librairie acadienne, les maisons d'édition Perce-Neige et Bouton d'or d'Acadie, institutions qui ont toutes leur siège à Moncton. La scène culturelle acadienne du sud-est du Nouveau-Brunswick est aussi nourrie par la présence de nombreux artistes affiliés, ou non, aux institutions mentionnées précédemment. Si certains sont originaires de l'endroit, d'autres viennent d'ailleurs, mais tendent à considérer Moncton comme leur « chez eux » et le chiac comme leur outil de création ; (citons le cas récent de l'artiste multidisciplinaire Céleste Godin qui s'est vu commander à titre de dramaturge une pièce de théâtre en chiac⁶). Enfin, Moncton a alimenté, et continue d'inspirer le travail de poètes (tels Gérard LeBlanc⁷, Paul Bossé, Gabriel Robichaud, Dyane Léger, Hélène Harbec et Sarah Marylou Brideau), de romanciers (Gérald LeBlanc, France Daigle, Jean Babineau ou Ulysse Landry), mais aussi d'artistes visuels (le peintre Yvon Gallant a de nombreuses fois peint Moncton, voir Doyon-Gosselin, 2022).

Une autre lecture de l'urbanité et de la francophonie acadienne s'impose alors, celle de la vitalité et de l'affirmation identitaire. La ville devient un lieu de pouvoir, où les francophones exercent une influence sur la gouvernance et le développement (Allain et Chiasson, 2016). On assiste également à un contre-discours sur le chiac, vernaculaire acadien fortement associé à l'urbanité, qui devient symbole de la résistance, de la créativité de la francophonie de Moncton, voire de l'Acadie dans le regard extérieur (Leclerc, 2006 ; Cormier, 2010 ; Berger, 2020). Témoin précoce de ce changement, l'historien Maurice

6. La pièce *Overlap* est en effet une commande de la compagnie Satellite théâtre qui a demandé à l'artiste d'écrire un « chœur tragique en chiac » (voir Mousseau, 2019).

7. Gérard LeBlanc va travailler à ériger sa ville d'adoption au rang de « capitale culturelle » (voir Boudreau, 2007). Si certains estiment que LeBlanc construit un mythe plutôt que le compte rendu d'une réalité, Boudreau nous rappelle que l'on ne peut négliger la dimension instituante du mythe. En 2012, Brun del Re consacre son mémoire de maîtrise à Moncton (et Ottawa) comme « capitales littéraires ».

Basque (2005) a merveilleusement bien résumé l'interdépendance croissante entre Moncton et l'Acadie en suggérant, dans un titre de communication, que l'on assiste à la « monctonisation de l'Acadie et à l'acadianisation de Moncton ». Pour toutes ces raisons, Moncton constitue un terrain d'enquête privilégié pour l'étude des dynamiques linguistiques en Acadie.

Les données disponibles

Le recensement des pratiques linguistiques en usage à Moncton possède déjà une certaine profondeur historique. Le milieu des années 1970 voit la constitution des premiers corpus, récoltés entièrement ou partiellement dans le but de répertorier des pratiques vernaculaires. Il s'agit d'abord du corpus de Marie-Marthe Roy, alors étudiante à la maîtrise à l'Université du Québec à Montréal, et du corpus de Gérard Étienne, professeur d'information-communication à l'Université de Moncton. Ces deux corpus présentent l'avantage d'être entièrement à la disposition des linguistes puisqu'ils ont été déposés au CRLA⁸. Dans la foulée, ce centre va accompagner toute une génération de chercheuses et chercheurs qui, durant les années 1980 et 1990, vont se lancer dans la constitution de corpus. Afin d'appréhender ces corpus⁹ en un coup d'œil, nous en proposons une présentation sous forme de tableau synoptique.

Ce tableau fait d'emblée apparaître un clivage net entre les corpus des années 1970 et 1980 (incluant celui de Perrot au tout début de la décennie 1990) et ceux des années 1990 et 2000. Ceux de la première période ont voulu servir d'assise à la description du vernaculaire, et ceux qui viendront ensuite dans les années 1990 ont plutôt adopté une approche fondée sur l'étude des représentations linguistiques. Plus largement, l'analyse de discours individuels, institutionnels, médiatiques venus de différents groupes sociaux (jeunes, artistes, militants, chroniqueurs, immigrants) devient alors l'une des marques de fabrique de la sociolinguistique monctonienne. Sous l'égide de la sociolinguiste Annette Boudreau, qui a influencé toute une génération de jeunes chercheuses et chercheurs, les travaux produits dans le giron du CRLA s'attacheront plutôt à la prise en compte des mécanismes sociostructurels et symboliques de la domination linguistique dans les rapports, entre différents groupes de francophones ainsi qu'entre francophones et anglophones. On connaît bien les atti-

8. Ces corpus sont consultables sur deux supports : cassettes audio et transcriptions papier.

9. À noter : d'autres corpus illustrent les pratiques du français en Acadie. Voir notamment les travaux de Wiesmath (2006), Comeau (2011), King (2013) et Roussel (2020).

TABLEAU 11.1

État des lieux des corpus du Centre de recherche sur la langue en Acadie

Nom du corpus	Année de création	Créateur(s) du corpus	Région ciblée	Heures d'enregistrements ou de mots transcrits	Nbre de personnes	Âge	Objectifs
Le corpus Roy	1975	Marie-Marthe Roy	Moncton	223 937 mots	22	15-89 ans	Description du vernaculaire
Le corpus Étienne	1975	Gérard Étienne	Nord-est du N.-B.	17 056 mots	41	9-89 ans	Contes et récits de vie
Le corpus Péronnet	1985, 1988, 1989	Louise Péronnet	Sud-est du N.-B.	24 heures	39	18-55 ans	Description du vernaculaire
Le corpus Perrot-CRLA	1991	Marie-Ève Perrot	Moncton	≈ 11 heures	44	16-19 ans	Description du vernaculaire
Le corpus Libotte	1991	non mentionné	Dieppe et Shediac	non mentionné	25	16-19 ans	Tester la variation stylistique (passage d'un registre familial à un registre soutenu)
Le corpus Boudreau-Dubois	1989-1994	Annette Boudreau et Lise Dubois	Plusieurs régions du N.-B.	≈ 30 heures	≈ 60	17-18 ans	Évaluer l'insécurité linguistique
Le corpus Boudreau	1994-1996	Annette Boudreau	Plusieurs régions du N.-B.	≈ 50 heures	100	25-60 ans	Évaluer l'insécurité linguistique
Le corpus Sévin	1995	Muriel Sévin	Moncton	23 heures	26	31-60 ans	Mesurer les attitudes linguistiques des professionnels francophones
Le corpus Parkton	1998	Annette Boudreau, Lise Dubois et Guylaine Poissant	Moncton	≈ 20 heures	22	21-83 ans	Examiner la situation sociolinguistique d'un quartier populaire
Le corpus Poissant	2004	Guylaine Poissant	Différentes régions du N.-B.	266 124 mots	31	20-60 ans	Savoir comment les femmes perçoivent les questions de santé
Le corpus CREFO-CRLA	1997-2000	Membres du Centre de recherche en éducation franco-ontarienne (CREFO) et du CRLA	Plusieurs régions de l'Acadie des Maritimes	296 heures	148	18 ans et plus	Récits de vie, discours sur la langue et l'identité

tudes et les perceptions de la population à l'égard des pratiques du français à Moncton, ainsi que l'évolution des discours à l'égard du chiac. Cependant, on ne dispose que de peu de données empiriques sur les caractéristiques linguistiques du français parlé *hic et nunc* et sur lequel, faute justement de données, tous les discours de dénigrement sont permis. En effet, périodiquement, des discussions publiques sur la qualité du français à Moncton reviennent. En 2012, quelques membres du corps professoral de l'Université de Moncton ainsi que certains journalistes avaient mené une véritable croisade face à ce qu'ils considéraient comme un français de piètre qualité circulant à Moncton, en particulier chez les jeunes (voir Arrighi et Violette, 2013). À peu près dix ans plus tard, le Conseil de la langue française de cette institution remet cette question de l'avant, et propose des journées de réflexion sur la qualité de la langue française à l'Université¹⁰. Faute de données disponibles, il est toutefois fort difficile pour les linguistes conviés d'esquisser un portrait représentatif de l'hétérogénéité des pratiques linguistiques à l'échelle de Moncton (Violette, 2022; Vernet, 2022). Les linguistes concernés peinent à se prononcer sur la question dans la sphère publique et médiatique.

Par ailleurs, ce tableau fait apparaître une seconde faille : l'absence de corpus conservés institutionnellement et recueillis après l'an 2000¹¹. De fait, les corpus disponibles correspondent à des collectes antérieures aux grandes transformations subies par Moncton, notamment en termes de mobilités (géographique et sociale) de la population francophone.

Une francophonie qui a changé¹²

À Moncton, la population francophone a d'abord et avant tout augmenté du fait d'un certain exode rural des régions du nord du Nouveau-Brunswick. Cette migration interne s'explique par le départ des jeunes de ces régions

10. Ces deux épisodes ne sont que deux exemples récents de débats sur la qualité du français parlé à Moncton, qui perdurent depuis le début du 20^e siècle, voir Boudreau (2021 et 2014).

11. Cela n'empêche pas l'existence de corpus « privés », recueillis par des linguistes de divers horizons (par exemple celui de Wiesmath, 2006).

12. On remarquera que, voulant illustrer un changement qui se poursuit actuellement (mais qui a débuté il y a plus de 20 ans), nous mobilisons des chiffres qui parfois datent un peu. En l'absence de données issues d'un recensement long et récent, qui permettraient d'isoler exclusivement la population francophone de Moncton, nous nous fondons sur des données prises dans les médias ou générées par des initiatives privées (comme la Stratégie d'immigration du Grand Moncton 2020-2024). S'il n'est pas toujours aisé de chiffrer le changement (la présence accrue des francophones du nord et de gens venus d'ailleurs dans la francophonie internationale), ce changement existe et l'on ne peut l'ignorer.

vers les grands centres, qui offrent formations postsecondaires et emplois qualifiés. En 2006, Beaudin et Forgues soulignent que 71 % des francophones nouvellement arrivés dans le Grand Moncton proviennent du nord-est de la province. Dix ans plus tard, cette tendance semble toujours d'actualité (voir Allain et Chiasson, 2016, p. 83). Dans le même ordre d'idées, Boudreau souligne que :

[l]a présence [à Moncton] de francophones d'ailleurs, une relative nouveauté depuis les années 1990, exerce sans aucun doute une influence importante sur les pratiques langagières et contribue à redéfinir les modalités d'expression rattachées à l'acadianité. [...] Avec ces apports nouveaux, la ville de Moncton joue un rôle dans la redéfinition de la/des norme(s) du français pour les Acadiennes et les Acadiens, et devient l'un des lieux importants de la production et de la (re)production de l'identité acadienne contemporaine. (Boudreau, 2014, p. 176)

Il faut noter que la francophonie du nord-est du Nouveau-Brunswick est, à plusieurs égards, différente de celle du sud-est. Elle est notamment réputée plus homogène linguistiquement (97 % de la population y est francophone) et le français qui y a cours est souvent présenté comme peu, voire pas influencé par l'anglais. Aucune étude ne s'est cependant encore intéressée à cette population néo-monctonienne de francophones originaires du Nouveau-Brunswick. Ainsi, on ne sait pas dans quelle mesure elle aurait adopté des traits linguistiques locaux, ni jusqu'à quel point elle aurait influencé le français en usage à Moncton.

À Moncton, la population francophone a aussi augmenté grâce aux entrées internationales. Selon les données du recensement de 2021, le Nouveau-Brunswick a accueilli 28 509 nouveaux résidents entre 2016 et 2021, parmi lesquels 13 347 se sont établis à Moncton, soit 46 % (Radio-Canada, « Qu'est-ce qui attire les gens vers le Nouveau-Brunswick? », 11 février 2022). En matière d'immigration, on constate que Moncton devient la porte d'entrée canadienne pour davantage de nouveaux arrivants (bien que les difficultés de rétention demeurent, les départs vers le Québec et l'Ontario étant fréquents). Le centre urbain s'est doté, en 2014, d'une stratégie en matière d'immigration. Depuis, Moncton est devenu le pôle d'attraction des flux migratoires et accueille désormais la majorité des immigrants francophones qui choisissent le Nouveau-Brunswick. Soixante-sept pour cent d'entre eux se sont établis dans le centre urbain en 2021 (Radio-Canada, « Moncton est le premier choix des immigrants francophones au Nouveau-Brunswick », Océane Doucet, 13 juin 2022). Parmi les nouveaux arrivants à

Moncton entre 2011 et 2016, 28 % ont le français comme première langue officielle, ce qui représente le taux le plus élevé sur l'ensemble des centres urbains, hors Québec (Stratégie d'immigration du Grand Moncton 2020-2024, p. 12). Selon ce dernier document, il y aurait 8 000 immigrants dans le Grand Moncton, auxquels s'ajoutent 1 500 étudiants internationaux et 700 travailleurs étrangers temporaires (*Ibid.*, p. 12). Cette population présente en outre un bon niveau de formation¹³ : 52 % de la population immigrante détient un diplôme universitaire, ce qui est nettement plus élevé que la population née au Canada (*Ibid.*, p. 13). De plus, 36 % des immigrants dans le Grand Moncton sont bilingues dans les deux langues officielles : « Parmi les immigrants dont la langue maternelle n'est pas une langue officielle, un sur quatre s'exprime à la fois en français et en anglais dans le Grand Moncton, comparativement à 14 % dans l'ensemble du Nouveau-Brunswick et à 11 % dans tout le Canada » (*Ibid.*, p. 13). Le taux de rétention des immigrants est de 49,3 % selon un chiffre couvrant une période de 5 ans en 2015, ce qui mène à constater une amélioration liée aux efforts fournis en la matière. Ce taux est supérieur à ceux des autres centres urbains de l'Atlantique (*Ibid.*, p. 25). Il faut noter que, selon l'approche ethnisante – qui retient comme légitimes représentants du français les seuls locuteurs et locutrices ayant des ancêtres acadiens – et qui a toujours dominé au sein de la linguistique (tout type d'approches confondues), cette population n'a jamais été prise en compte.

À cet égard, soulignons que les deux travaux fondateurs qui ont décrit le fonctionnement linguistique du français parlé à Moncton, Roy (1979) et Perrot (1995), démontrent une absence de considération envers l'immigration, ce qui était commun à l'époque. En effet, dans l'ensemble des sciences humaines et sociales, avant les années 2010 et ce que l'on a qualifié de tournant migratoire (Faist, 2013), les populations avaient tendance à être construites comme historiquement ancrées dans un territoire fixe, le mouvement et le mélange y étaient vus comme des éléments tantôt marginaux, tantôt perturbateurs, de l'intégrité communautaire. Ainsi, la population acadienne était de facto conçue comme homogène et définie par son lieu de naissance, sans que cela soit source de questionnements. Le français à décrire était également considéré comme une variété sociomaternelle, transmise dans l'enceinte familiale par la prime

13. C'est aussi le cas des « migrants » venus d'ailleurs dans la province. Ainsi, selon Forgues *et al.*, 2009, les francophones qui viennent du nord sont plus instruits que la moyenne de la population.

socialisation. Roy et Perrot adoptent par ailleurs des critères natifs stricts pour sélectionner les locuteurs et locutrices susceptibles de participer à leur étude : être de langue maternelle française et être originaire de Moncton (ou d'une ville ou d'un village voisin). Dans le cas de Perrot, les critères étaient encore plus rigides à cet égard, dans la mesure où les jeunes témoins devaient toujours avoir habité Moncton, de même que leurs parents.

La transformation de la population francophone à Moncton est marquée aussi par la mobilité sociale de cette dernière, ce que traduit une augmentation du niveau de son éducation et de ses revenus. Selon le *Portrait des minorités de langue officielle au Canada, les francophones du Nouveau-Brunswick* (Statistique Canada, 2011), on assiste à une augmentation constante du niveau de scolarisation des francophones depuis 1971¹⁴. C'est un peu avant cette date (1966) que les francophones de la province ont pu, pour la première fois, suivre des cursus d'enseignement entièrement en français, des premières classes jusqu'aux éventuelles études collégiales ou universitaires. On peut envisager que cela ait une incidence importante sur les pratiques linguistiques. Notons aussi que, pour la confection des corpus du CRLA énumérés ci-haut, plusieurs chercheurs et chercheuses ont volontairement exclu les personnes dotées d'un certain capital financier et/ou éducatif (par exemple dans les corpus Parkton et Sévin, ou encore celui Perrot-CRLA). En ce qui concerne le revenu, on note une croissance du revenu moyen des francophones depuis 1971 et une diminution de l'écart avec les anglophones¹⁵.

Pistes pour de nouveaux corpus

Au sujet des travaux de constitution des grands corpus montréalais, Blondeau (2014) rappelle que les grandes enquêtes variationnistes, nourries de la tradition dialectologique, sont marquées par l'idéologie du locuteur natif et

14. Selon des données datant de 2006, pour la population francophone âgée de 25 ans et plus, les chiffres sont les suivants : 36,9 % aucun diplôme, 18,9 % études secondaires, 12,3 % école de métiers, 16,5 % collège, 3,7 % certificat ou diplôme inférieur au baccalauréat et 12,7 % baccalauréat ou plus (Lepage, Bouchard-Coulombe et Chavez, 2011, p. 57). Or, on note une différence de pourcentage notable entre les groupes d'âge. Par exemple, chez les 25-34 ans, 22 % sont titulaires d'un diplôme universitaire, contrairement aux 5,6 % des 65 ans et plus (*Ibid.*, p. 61). On note aussi une différence du niveau de scolarisation entre régions : dans le sud-est, 15 % des francophones ont un diplôme universitaire.

15. Toutefois un écart demeure présent, voir Lepage, Bouchard-Coulombe et Chavez, 2011, p. 71. L'étude de Béland, Forgues et Beaudin (2008) souligne par ailleurs que les unilingues francophones demeurent le groupe le plus désavantagé au niveau du revenu et des possibilités d'insertion socioprofessionnelle.

d'un idéal de sédentarité. Il en est résulté une vision ethnoculturelle de la communauté linguistique qui ne permet pas de fournir un portrait global et intégré des pratiques linguistiques. Dans le contexte de Montréal, Blondeau se demande comment la relation plus étroite entre le local et le global nécessite d'élargir les avenues de recherche de la sociolinguistique montréalaise. Elle interroge aussi la prise en compte de la dimension métropolitaine de Montréal dans les descriptions et corpus du français montréalais (Blondeau, 2014, p. 206-207, voir également Blondeau *et al.*, 2016). La question se pose également pour Moncton. Somme toute, on peut affirmer que le facteur migratoire n'a jamais été intégré à la description du français monctonien. À notre connaissance, aucune étude sur le français de Moncton n'a pris en compte les néo-Monctoniens et néo-Monctoniennes francophones.

Pourtant, le tissu social urbain s'est densifié et complexifié à Moncton et on ne sait pas quels effets linguistiques cela a pu avoir. *A contrario*, les représentations du français parlé à Moncton continuent d'être plutôt homogénéisantes et dominées par une vision anglicisée ou du moins hybride, ce qui, en retour, peut affecter les pratiques. Ainsi, dans certains milieux, il devient « chic » de parler chiac (pour reprendre le mot de Leclerc, 2005¹⁶). Blondeau ajoute que la dimension multiculturelle et plurilingue de Montréal a surtout été prise en compte par des études plus anthropologiques axées sur le choix de langues et sur la transmission des langues patrimoniales. Au Québec, c'est la concurrence entre le français et l'anglais qui a occupé l'agenda scientifique, au détriment des problématiques de contacts avec d'autres langues. Par ailleurs, la prise en compte du facteur migratoire a été incorporée à la problématique du choix de langues plutôt qu'à celle de la variation linguistique. On pourrait affirmer la même chose pour Moncton¹⁷ et, comme nous y invite Blondeau, mieux réfléchir à la notion de « communauté linguistique ». Cette dernière fournit deux pistes intéressantes pour des recherches à Moncton. Tout d'abord, il serait bon d'établir des portraits des répertoires linguistiques des francophones, notamment pour mettre en relief les pratiques plurilingues ; ensuite, d'adopter la distinction de Hymes

16. Une tendance vers l'emploi d'un français hybride comme norme de communication semble exister chez une part de l'élite culturelle francophone, dont les membres ne proviennent pas tous de Moncton. À cet égard, il serait intéressant d'analyser les pratiques linguistiques entourant la promotion des événements du Festival Acadie Rock à Moncton. Voir aussi le cas de Céleste Godin évoqué plus haut.

17. Voir, par exemple, Guignard (2007), pour les migrants intraprovinciaux ; et Violette (2018-2019), pour les migrants internationaux.

(1974) entre appartenance et participation, dans le découpage de la communauté linguistique (Daveluy, 2006; voir aussi Blondeau, 2015, p. 218). Ce faisant, on pourrait inclure des francophones ayant déménagé, des allophones parlant français ainsi que des anglophones bilingues. Il serait également profitable d'adopter une perspective de sociolinguistique urbaine. Dans son ouvrage pionnier consacré à la question, Calvet (1994) appelle à saisir la ville, comme objet, à travers les langues et, pour ce faire, à développer trois thématiques: la ville comme facteur d'unification linguistique (émergence de langues véhiculaires); la ville comme lieu de conflit de langues et la ville comme lieu de coexistence et de métissage linguistique (création de formes spécifiques des parlers urbains). Calvet établit une distinction entre la prise en compte du facteur urbain (isoler l'effet de la ville sur les pratiques linguistiques, à l'opposé de ce que l'on observe en milieu rural) et l'analyse d'un corpus urbain (données linguistiques recueillies dans un cadre urbain) (Calvet, 1994, p. 16). Cette distinction est intéressante pour la dimension évaluative de notre contribution, puisqu'il semble que l'analyse des pratiques linguistiques à Moncton ait surtout été menée selon la seconde perspective (corpus Roy, Perrot, Parkton)¹⁸, qui n'est pas celle que prône Calvet au titre d'un programme de « sociolinguistique urbaine ». Pour Calvet, l'intérêt de la ville réside dans le phénomène d'attraction des migrants et, plus précisément, dans l'étude de la contrepartie linguistique à l'affluence de ces migrants. Le sociolinguiste affirme que « l'urbanisation est un phénomène qui va croissant, alors la réflexion sur l'avenir des situations linguistiques ne peut faire l'économie d'une réflexion sur le facteur urbain, notamment l'analyse des productions langagières des migrants et, surtout, des enfants de migrants » (*Ibid.*, p. 14).

Si on suit cette logique, Moncton présenterait une valeur sociolinguistique en raison des flux migratoires importants qui y convergent, ceux des francophones d'ailleurs¹⁹ au Nouveau-Brunswick, d'ailleurs au Canada et d'ailleurs dans le monde, ce qui donne trois niveaux de mobilité: intraprovinciale,

18. Cela rejoint la critique que fait Hélène Blondeau (2014) au sujet de la sociolinguistique montréalaise: exclusion des immigrants des corpus, seule prise en compte de locuteurs nés à Montréal et vivant dans des quartiers spécifiques de la métropole. Cette façon de faire repose sur une conception ethnoculturelle de la communauté linguistique. Blondeau introduit alors la distinction entre français montréalais et français parlé à Montréal, que nous pourrions reprendre à notre compte et distinguer français monctonien et français parlé à Moncton.

19. On peut noter également les populations allophones (ex. de la Syrie) qui font le choix de scolariser leurs enfants en français, soit dans le système scolaire francophone, soit au sein des programmes d'immersion française du système anglophone.

interprovincial et international. En ce sens, il y aurait lieu de constituer des corpus avec des migrants pour étudier les effets linguistiques de la ville sur leurs pratiques, c'est-à-dire étudier l'appropriation de variantes identifiées comme acadiennes par des francophones d'ailleurs (provenant d'une autre région que celle du sud-est et ceux issus de l'immigration). Ajoutons que la considération de facteurs tels que celui du réseau social (voir Beaulieu et Cichocki, 2002) permettrait de mieux saisir l'impact des affiliations sociales des individus sur leurs propres pratiques linguistiques. Au vu des fonctions identitaires et d'intégration associées au chiac (ce que des travaux ont déjà démontré, voir notamment Violette, 2009 et 2021), on tendrait à observer un phénomène de vernacularisation, mais l'on connaît très mal quels traits sont adoptés, dans quels contextes et à quelle génération de migrants, la première ou la deuxième.

L'effet linguistique inverse mérite aussi d'être étudié, c'est-à-dire la diffusion de traits caractéristiques d'autres variétés de français, et en premier lieu le français laurentien²⁰, ou d'autres langues, dans la population francophone de Moncton. Cette influence est sans doute plus difficile à isoler comme facteur, puisque le contact se fait également via les médias, la culture et les voyages. De plus et avant tout, le français standard est la variété cible à l'école comme à l'université (Vernet, 2016). Étant donné l'augmentation du niveau de scolarisation des francophones de Moncton, on peut s'attendre à une standardisation accrue du français de la ville. Demeurent aussi les questions de style. Les corpus du CRLA ont tous recherché un style familier pour « faire sortir le vernaculaire », mais les francophones à Moncton ont assurément d'autres styles à leur disposition. Enfin, tous les corpus disponibles ont été recueillis par la méthode de l'entretien semi-dirigé dont on comprend bien les divers avantages. Il faudrait pourtant préconiser des formes plus écologiques de recueil des pratiques linguistiques afin de documenter un plus grand éventail de registres stylistiques. La compilation de données issues des médias sociaux (émissions en baladodiffusion produites localement) pourrait être l'une des voies prometteuses (voir Arrighi et Berger, 2021 pour une exploration d'un balado produit localement dans un but de description des pratiques linguistiques), mais non la seule.

20. Voir, par exemple, Cichocki et Perreault (2019) qui montrent l'existence d'un changement linguistique observé à la fois en français laurentien et dans trois régions du nord du Nouveau-Brunswick (Shippagan, Paquetville et Néguaac) en ce qui concerne l'utilisation du R dorsal au détriment du R apical.

Comme l'ont fait remarquer Allain et Chiasson (2016, p. 6), les études sur l'urbanité ont tendance à se concentrer sur les métropoles au détriment des villes de taille moyenne qui, pourtant, présentent tout de même des réalités urbaines. Ces villes sont généralement moins bien étudiées sous divers aspects, notamment sociodémographiques ou sociolinguistiques²¹, mais aussi en matière de gouvernance urbaine. Le travail de description du français parlé à Moncton pourrait alors y participer.

* * *

Dans ce chapitre, nous avons souligné dès l'introduction le caractère métropolitain de Moncton et sa valeur comme terrain de choix pour une approche en sociolinguistique urbaine. De manière croissante depuis les années 2000, Moncton agit comme lieu de convergences pour des francophones provenant d'horizons divers et présentant des répertoires linguistiques plurilingues et pluridialectaux. Or, un état des lieux des corpus disponibles sur le français parlé à Moncton pointe une absence de données représentatives de cette réalité démolinguistique. Un chantier plus systématique dans l'étude du français parlé à Moncton reste toujours à entreprendre.

Références

- ALLAIN, Greg (2006), « Resurgo ! » : la renaissance et la métropolisation de Moncton, la ville-pivot des provinces maritimes et nouvelle capitale acadienne, *Francophonies d'Amérique*, 22, 95-119, <https://doi.org/10.7202/1005381ar>.
- ALLAIN, Greg et Guy CHIASSON (2016), *Minorités francophones et gouvernance urbaine: Moncton, Sudbury, Edmonton et Ottawa*, Presses de l'Université Laval.
- ARRIGHI, Laurence et Tommy BERGER (2021), « Cosser tu parles », le chiac : un français non standard ou une langue à part ? Description de pratiques linguistiques au sud-est du Nouveau-Brunswick (Canada), dans Iwona PIECHNIK et Marta WICHEREK (dir.), *Langues romanes non-standard*, Presses de l'Université Jagellonne, 11-41, <https://ruj.uj.edu.pl/xmlui/handle/item/285-998>.
- ARRIGHI, Laurence (2020), De quoi le chiac est-il le nom ? Une étude du parcours définitoire du chiac et de ses enjeux dans la littérature savante et de vulgarisation scientifique, *Francophonies d'Amérique*, 50, 95-118, <https://doi.org/10.7202/1073711ar>.

21. Il existe pourtant une certaine tradition sociolinguistique de constitution de corpus dans des villes de taille moyenne, ainsi les corpus d'Orléans en France (une première vague dans les années 1970, une deuxième vague dans les années 2000).

- ARRIGHI, Laurence et Émilie URBAIN (2016), « Wake up Québec »: du recours aux communautés francophones minoritaires dans le discours visant l'émancipation nationale du Québec, *Francophonies d'Amérique*, 42-43, 105-124, <https://doi.org/10.7202/1054037ar>.
- ARRIGHI, Laurence et Isabelle VIOLETTE (2013), De la préservation linguistique et nationale: la qualité de la langue de la jeunesse acadienne, un débat linguistique idéologique, *Revue de l'Université de Moncton*, 44(2), 67-101, <https://doi.org/10.7202/1031001ar>.
- BASQUE, Maurice (2005), De l'acadianisation de Moncton à la monctonisation de l'Acadie, communication présentée au colloque *Cultures minoritaires et urbanité, explorations, théories et méthodes*, Université de Moncton, 22-24 septembre.
- BEAUDIN, Maurice et Rodrigue LANDRY (2003), L'attrait urbain; un défi pour les minorités francophones au Canada, *Thèmes canadiens/Canadian Issues*, février, 19-22.
- BEAULIEU, Louise et Wladyslaw CICHOCKI (2002), Le concept de réseau social dans une communauté acadienne rurale, *Revue canadienne de linguistique*, 47, 123-150.
- BÉLAND, Nicolas, Éric FORGUES et Maurice BEAUDIN (2008), Évolution du salaire moyen des hommes de langue maternelle française ou anglaise au Québec et au Nouveau-Brunswick, 1970-2000, Office québécois de la langue française, http://www.oqlf.gouv.qc.ca/etudes/etude_13.pdf.
- BERGER, Tommy (2020), *Le chiac: entre langue des jeunes et langue des ancêtres*, mémoire de maîtrise, Université de Montréal.
- BLONDEAU, Hélène (2014), La nature métropolitaine du Montréal « d'aujourd'hui » et le français « d'ici », dans Wim REMYSEN (dir.), *Les français d'ici, du discours d'autorité à la description des normes et des usages*, Presses de l'Université Laval, 205-239.
- 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, 159-182, <https://doi.org/10.3917/cisl.1602.0159>.
- BOUDREAU, Annette (2021), *Dire le silence: l'insécurité linguistique en Acadie: 1867-1970*, Prise de parole.
- BOUDREAU, Annette (2012), Discours, nomination des langues et idéologies linguistiques, dans Davy BIGOT, Michael FRIESNER et Mireille TREMBLAY (dir.), *Les Français d'ici: description, représentation et théorisation*, Presses de l'Université Laval, 89-109.
- BOUDREAU, Annette (2014), Des voix qui se répondent: analyse discursive et historique des idéologies linguistiques en Acadie, l'exemple de Moncton, *Minorités linguistiques et société/Linguistic Minorities and Society*, 4, 175-199, <https://doi.org/10.7202/1024697ar>.
- BOUDREAU, Raoul (2007), La création de Moncton comme « capitale culturelle » dans l'œuvre de Gérald Leblanc, *Revue de l'Université de Moncton*, 38(1), 33-56, <https://doi.org/10.7202/018403ar>.
- BRETON-CARBONNEAU, Gabrielle et Monica HELLER (2021), Le rapport centre-périphérie et les mobilités structurées: les jeunes Franco-Manitobains et Montréal, *Francophonies d'Amérique*, 52, 85-103, <https://doi.org/10.7202/1082863ar>.

- BRUCE, Clint (2005), Gérald LeBlanc et l'univers micro-cosmopolite de Moncton, *Études canadiennes*, 58, 205-220.
- BRUN DEL RE, Ariane (2012), *Portrait de villes littéraires: Moncton et Ottawa*, mémoire de maîtrise, Université McGill.
- CALVET, Louis-Jean (1994), *Les voix de la ville: introduction à la sociolinguistique urbaine*, Payot.
- CICHOCKI, Wladyslaw et Yves PERREAULT (2019), Contribution à l'étude phonétique et géolinguistique du R en français parlé au Nouveau-Brunswick, dans Sandrine HALLION et Nicole ROSEN (dir.), *Les français d'ici, des discours et des usages*, Presses de l'Université Laval, 177-206.
- COMEAU, Philip (2011), *A Window on the Past, a Move toward the Future: Sociolinguistic and Formal Perspectives on Variation in Acadian French*, thèse de doctorat, York University.
- CORMIER, Julie (2010), *Représentations, dynamiques langagières et Internet: le cas du chiac en Acadie*, mémoire de maîtrise, Université de Moncton.
- DAVELUY, Michelle (2006), *Les langues étendards: allégeances langagières en français parlé*, Québec, Nuit Blanche.
- DESJARDINS, Pierre-Marcel (2002), *La périphérie n'est pas homogène: trois régions du Nouveau-Brunswick, Madawaska, Gloucester et Kent-Westmorland*, Institut canadien de recherche sur le développement régional, INRS Urbanisation, culture et société.
- DOYON-GOSSELIN, Benoit (2022), *Moncton mentor: géocritique d'une ville*, Perce-Neige.
- DUBOIS, Lise, Mélanie LEBLANC et Maurice BEAUDIN (2006), La langue comme ressource productive et les rapports de pouvoir entre communautés linguistiques, *Langage et société*, 4(118), 17-41, <https://doi.org/10.3917/l.s.118.0017>.
- FAIST, Thomas (2013), The mobility turn: a new paradigm for the social sciences, *Ethnic and Racial Studies*, 36(11), 1637-1646, <https://doi.org/10.1080/01419870.2013.812229>.
- FORGUES, Éric, Maurice BEAUDIN, Josée GUIGNARD-NOËL et Jonathan BOUDREAU (2009), *Analyse de la migration des francophones au Nouveau-Brunswick entre 2001 et 2006*, Institut canadien de recherche sur les minorités linguistiques.
- FORGUES, Éric, Anne ROBINEAU, Dominique PÉPIN-FILION et Marc-André BOUCHARD (2020), La construction d'espaces francophones comme projet de société en milieu minoritaire, *Minorités linguistiques et société/Linguistic Minorities and Society*, 13, 29-48. <https://doi.org/10.7202/1070389ar>.
- GESNER, Edward B. (1986), *Bibliographie annotée de linguistique acadienne*, Centre international de recherche sur le bilinguisme.
- GUIGNARD, Josée (2007), *Les migrants francophones du nord du Nouveau-Brunswick dans le territoire urbain de Moncton-Dieppe: réseaux sociaux et vitalité ethnolinguistique*, mémoire de maîtrise, Université de Moncton.
- HYMES, Dell (1974), *Foundations in Sociolinguistics: An Ethnographic approach*, University of Pennsylvania Press.
- KING, Ruth (2013), *Acadian French in Time and Space*, Durham, N.C., Duke University Press.

- LANDRY, Rodrigue, Réal ALLARD et Kenneth DEVEAU (2006), Revitalisation ethno-linguistique: un modèle macroscopique, dans André MAGORD (dir.), *Innovation et adaptation, expériences acadiennes contemporaines*, Éditions Peter Lang, 105-124.
- LANDRY, Michelle (2021), Acadie rurale ou Acadie urbaine?, dans Michelle LANDRY, Dominique PÉPIN-FILION et Julien MASSICOTE (dir.), *L'état de l'Acadie. Un grand tour d'horizon de l'Acadie contemporaine*, Del Busso, 17-20.
- LEBLANC, Matthieu (2014), Les atouts et les avantages du bilinguisme à Moncton, entre discours et réalité, *Minorités linguistiques et société*, 4, 154-174, <https://doi.org/10.7202/1024696ar>.
- LECLERC, Catherine (2006), Ville hybride ou ville divisée: à propos du chiac et d'une ambivalence productive, *Francophonies d'Amérique*, 22, 153-165, <https://doi.org/10.7202/1005384ar>.
- LEPAGE, Jean-François, Camille BOUCHARD-COULOMBE et Brigitte CHAVEZ (2011), *Portrait des minorités de langue officielle au Canada: les francophones du Nouveau-Brunswick*, Statistique Canada, Gouvernement du Canada.
- MASSIGNON, Geneviève (1962), *Les parlers français d'Acadie: enquête linguistique*, Klincksieck, 2 tomes.
- MUSSEAU, Sylvie (2019), Satellite Théâtre convie le public à « un chœur tragique à cinq voix », *L'Acadie nouvelle*.
- PÉRONNET, Louise (1989), *Le parler acadien du sud-est du Nouveau-Brunswick: éléments grammaticaux et lexicaux*, Peter Lang.
- PERROT, Marie-Ève (1995), *Aspects fondamentaux du métissage français/anglais dans le chiac de Moncton (Nouveau-Brunswick, Canada)*, thèse de doctorat, Université de Paris III.
- RADIO-CANADA (2023), Moncton et Halifax sont les deux villes qui grandissent le plus vite au pays, 11 janvier.
- RADIO-CANADA (2022), Qu'est-ce qui attire les gens vers le Nouveau-Brunswick?, 11 février.
- RADIO-CANADA (OCÉANE DOUCET) (2022), Moncton est le premier choix des immigrants francophones au Nouveau-Brunswick, 13 juin.
- ROUSSEL, Basile (2020), *À la recherche du temps (et des modes) perdu(s), une étude variationniste en temps réel du français acadien parlé dans le nord-est du Nouveau-Brunswick*, thèse de doctorat, Université d'Ottawa.
- ROY, Marie-Marthe (1979), *Les conjonctions but et so dans le français de Moncton, étude sociolinguistique de changements linguistiques provoqués par une situation de contact*, mémoire de maîtrise, Université du Québec à Montréal.
- S. A. (2020), *Stratégie d'immigration du Grand Moncton 2020-2024*, Ville de Moncton, Ville de Dieppe, Ville de Riverview.
- VERNET, Samuel (2016), *Discours, idéologies linguistiques et enseignement du français à l'Université de Moncton*, thèse de doctorat, Université de Moncton/Université de Grenoble Alpes.
- VERNET, Samuel (2022), Les idéologies linguistiques en milieu universitaire, *Journée de réflexion sur la langue française*, Université de Moncton, campus de Shippagan.

- VIOLETTE, Isabelle (2010), Discours, représentations et nominations: le rapport au « chiac » chez les immigrants francophones à Moncton (Acadie), dans Carmen LEBLANC, France MARTINEAU et Yves FRENETTE (dir.) *Vues sur les français du Canada*, Presses de l'Université Laval, 267-284.
- VIOLETTE, Isabelle (2018-2019), De quelques interprétations de la minoration: postures d'immigrants francophones face aux contacts de langues à Moncton, Nouveau-Brunswick, *Francophonies d'Amérique*, 46-47, 51-71, <https://doi.org/10.7202/1064887ar>.
- VIOLETTE, Isabelle (2021), « Parler chiac, c'est même pas *fake* »: immigration, authenticité et français vernaculaire en Acadie du Nouveau-Brunswick (Canada), dans Émilie URBAIN et Laurence ARRIGHI (dir.), *Retour en Acadie: penser les langues et la sociolinguistique à partir des marges. Ouvrage en hommage à Annette Boudreau*, Presses de l'Université Laval, 135-156.
- VIOLETTE, Isabelle (2022), Qualité de la langue et insécurité linguistique: quels rapports?, *Journée de réflexion sur la langue française*, Université de Moncton, campus de Moncton.
- WIESMATH, Raphaele (2006), *Le français acadien: analyse syntaxique d'un corpus oral recueilli au Nouveau-Brunswick, Canada*, L'Harmattan.

CHAPITRE 12

Annotation et analyse de données sociophonologiques sur grands corpus

Présentation de la plateforme Phonometrica

Julien Eychenne et Léa Courdès-Murphy

Le tournant du xxi^{e} siècle a révélé un intérêt croissant pour l'analyse des corpus oraux, comme l'atteste, par exemple, le développement de la sociophonologie et de la phonologie de corpus (Durand *et al.*, 2014, Detey *et al.*, 2016, Honey, 2017, Przewozny *et al.*, 2020). Ce phénomène a considérablement élargi la communauté des spécialistes de la parole au-delà de son noyau d'origine, ce qui a naturellement suscité de nouveaux besoins et de nouvelles attentes en termes de gestion et d'analyse de données orales. Nous présentons ici la plateforme logicielle Phonometrica, un outil destiné à l'annotation et l'analyse de corpus oraux alignés sur le signal sonore, et conçu pour faciliter un certain nombre de tâches courantes en (socio)phonologie et (socio)phonétique, pour le codage et l'analyse de variables discrètes aussi bien que l'extraction de mesures acoustiques continues.

Présentation du logiciel

La « ruée vers l'or des données » (Winter et Wedel, 2016, p. 80) à laquelle on assiste à l'heure actuelle a donné naissance à un riche écosystème de logiciels de traitement et d'analyse de données linguistiques, et en particulier de corpus oraux, chacun ciblant une communauté particulière de personnes utilisatrices. Sans en faire un examen exhaustif, on mentionnera à titre

d'exemple la plateforme ELAN (Sloetjes et Wittenburg, 2008) pour l'annotation de données audio et vidéo, Phon (Hedlund et Rose, 2020) dans le domaine de l'acquisition phonologique, et bien sûr Praat (Boersma et Weenink, 2023), qui s'est imposé comme un standard de facto pour l'annotation et l'analyse acoustique du signal de parole. Le logiciel Phonometrica représente quant à lui l'aboutissement d'un développement itératif d'outils dans le domaine de la (socio)phonologie de corpus, d'abord centrés sur les besoins du projet Phonologie du français contemporain (PFC) (Durand *et al.*, 2002), mais qui a évolué au fil du temps pour permettre de traiter tout type de corpus (socio)phonologique et/ou (socio)phonétique¹.

Phonometrica est un logiciel libre, disponible² sous la licence publique générale GNU, qui fonctionne sur les trois principaux systèmes d'exploitation (Windows, macOS et GNU/Linux). Les principes qui ont guidé la conception de ce logiciel sont la facilité d'utilisation, la portabilité, la performance et le respect des standards. Il s'appuie sur un certain nombre de bibliothèques logicielles existantes et éprouvées pour les opérations primitives, telles que la gestion du son, les expressions régulières ou le rendu de l'interface graphique, et il emprunte plusieurs routines de traitement du signal au Speech Signal Processing ToolKit (SPTK Working group 2008).

Ce logiciel a été conçu pour traiter de grandes collections de données linguistiques : il n'opère donc pas sur des fichiers individuels, mais sur un projet, qui est un ensemble de cinq types de données : 1) des *données de corpus*, c'est-à-dire des données audio primaires ainsi que les annotations correspondantes éventuelles ; 2) des *requêtes*, qui peuvent être sauvegardées et réutilisées ; 3) des *tableaux de données*, qui sont des données provenant généralement d'une requête et se présentent sous forme tabulée ; 4) des *scripts*, qui permettent d'automatiser certaines tâches à l'aide du moteur de script (voir section « Moteur de script ») ; 5) des *signets*, grâce auxquels on peut marquer certains items dans le projet, par exemple un instant spécifique dans une annotation.

L'interface graphique se décompose en quatre composants principaux (voir figure 12.1). Le *gestionnaire de fichiers* (à gauche) est utilisé pour afficher et organiser les fichiers du corpus, selon les cinq types énumérés plus haut³.

1. Le nom du logiciel est un hommage au travail pionnier de Zwirner et Zwirner (1936/1970), qui avaient développé une conception unifiée de la phonétique et de la phonologie : la phonométrie.

2. <http://www.phonometrica-ling.org>

3. Il est utile de préciser que l'organisation hiérarchique des fichiers dans un projet ne suit pas nécessairement leur organisation physique sur l'ordinateur de la personne utilisatrice.

Le champ de recherche rapide, situé en bas du navigateur, permet de trouver un ou plusieurs fichiers à partir d'une chaîne de caractères contenue dans le nom du fichier ou dans ses métadonnées. Le deuxième composant est le *panneau d'information* (à droite), qui est utilisé pour afficher et éditer les métadonnées des fichiers sélectionnés. En bas de l'interface se trouve la console qui permet d'exécuter directement des commandes ou de visualiser le résultat de certaines opérations, telles que l'exécution d'un script ou une mesure de formants. Enfin, la partie centrale de l'interface est occupée par la *visionneuse*, qui permet de visualiser chaque fichier ou requête sous forme d'onglets, de la même manière que les navigateurs Web affichent les pages Internet. Hormis la visionneuse, chaque composant peut être affiché ou masqué en fonction des besoins, de manière à optimiser l'utilisation de l'espace disponible sur l'écran.

FIGURE 12.1

Interface graphique de Phonometrica

8179 matches												Active target:
	File	Layer	Start time	End time	Left context	Target	Right context	Age	Gender	Speaker	Survey	Task
1	91oaligg_TextGrid	2	0.0000	4.8888	E: Mois euh <LA; C'est pas facile	0413	-> tu as beaucoup d'écart d'avec ton frè	27	Femme	SIAM1	91A	Guided conversation
2	91oaligg_TextGrid	2	4.8888	8.3200	tu saur euh. LA: Oui j'ai, euh mon avec	0412	ma saur0412 j'ai trois ans et avec0412	27	Femme	SIAM1	91A	Guided conversation
3	91oaligg_TextGrid	2	4.8888	8.3200	LA: Oui j'ai, euh non avec0412 ma saur	0412	j'ai trois ans et avec0412 mon frère041	27	Femme	SIAM1	91A	Guided conversation
4	91oaligg_TextGrid	2	4.8888	8.3200	0412 ma saur0412 j'ai trois ans et avec	0412	mon frère0412 j'en ai dix0413. E: Ah ou	27	Femme	SIAM1	91A	Guided conversation
5	91oaligg_TextGrid	2	4.8888	8.3200	412 j'ai trois ans et avec0412 mon frère	0412	j'en ai dix0413. E: Ah ouais quand même	27	Femme	SIAM1	91A	Guided conversation
6	91oaligg_TextGrid	2	4.8888	8.3200	ms et avec0412 mon frère0412 j'en ai dix	0413	E: Ah ouais quand même <LA: Et ou...> a	27	Femme	SIAM1	91A	Guided conversation
7	91oaligg_TextGrid	2	13.3721	15.2522	fort quoi de beau euh. LA: Mo saur elle	0411	est diététicienne0413. LA: Elle0412 tra	27	Femme	SIAM1	91A	Guided conversation
8	91oaligg_TextGrid	2	13.3721	15.2522	LA: Mo saur elle0411 est diététicienne	0413	LA: Elle0412 travaille0411 à l'hôpital	27	Femme	SIAM1	91A	Guided conversation
9	91oaligg_TextGrid	2	15.2522	18.5997	elle0411 est diététicienne0413. LA: Elle	0412	travaille0411 à l'hôpital0412 privé du	27	Femme	SIAM1	91A	Guided conversation
10	91oaligg_TextGrid	2	15.2522	18.5997	diététicienne0413. LA: Elle0412 travaille	0411	à l'hôpital0412 privé du Vol0412 d'Yerr	27	Femme	SIAM1	91A	Guided conversation
11	91oaligg_TextGrid	2	15.2522	18.5997	LA: Elle0412 travaille0411 à l'hôpital	0412	privé du Vol0412 d'Yerr0412 donc0412 j	27	Femme	SIAM1	91A	Guided conversation
12	91oaligg_TextGrid	2	15.2522	18.5997	availle0411 à l'hôpital0412 privé du Vol	0412	d'Yerr0412 donc0412 juste0412 à côté.	27	Femme	SIAM1	91A	Guided conversation
13	91oaligg_TextGrid	2	15.2522	18.5997	à l'hôpital0412 privé du Vol0412 d'Yerr	0412	donc0412 juste0412 à côté. LA: Elle0411	27	Femme	SIAM1	91A	Guided conversation
14	91oaligg_TextGrid	2	15.2522	18.5997	0412 privé du Vol0412 d'Yerr0412 donc	0412	juste0412 à côté. LA: Elle0411 a eu son	27	Femme	SIAM1	91A	Guided conversation
15	91oaligg_TextGrid	2	15.2522	18.5997	vé du Vol0412 d'Yerr0412 donc0412 juste	0421	à côté. LA: Elle0411 a eu son B.T.S. bo	27	Femme	SIAM1	91A	Guided conversation
16	91oaligg_TextGrid	2	18.5997	22.4516	0412 donc0412 juste0421 à côté. LA: Elle	0411	a eu son B.T.S. bah l'0411 y a un an mo	27	Femme	SIAM1	91A	Guided conversation
17	91oaligg_TextGrid	2	18.5997	22.4516	0412. LA: Elle0411 a eu son B.T.S. bah il	0411	y a un an mainte312nant tout juste0423	27	Femme	SIAM1	91A	Guided conversation
18	91oaligg_TextGrid	2	18.5997	22.4516	u son B.T.S. bah l'0411 y a un an mainte	3122	nant tout juste0423. LA: Donc0411 elle04	27	Femme	SIAM1	91A	Guided conversation
19	91oaligg_TextGrid	2	18.5997	22.4516	0411 y a un an mainte312nant tout juste	0423	LA: Donc0411 elle0412 travail0412 de	27	Femme	SIAM1	91A	Guided conversation
20	91oaligg_TextGrid	2	22.4516	23.8272	mainte312nant tout juste0423. LA: Donc	0411	elle0412 travail0412 de3122puis un an	27	Femme	SIAM1	91A	Guided conversation
21	91oaligg_TextGrid	2	22.4516	23.8272	3122nant tout juste0423. LA: Donc0411 elle	0412	travail0412 de3122puis un an. LA: Ell	27	Femme	SIAM1	91A	Guided conversation
22	91oaligg_TextGrid	2	22.4516	23.8272	ste0423. LA: Donc0411 elle0412 travaille	0412	de3122puis un an. LA: Elle0411 est resp	27	Femme	SIAM1	91A	Guided conversation
23	91oaligg_TextGrid	2	22.4516	23.8272	LA: Donc0411 elle0412 travail0412 de	1222	puis un an. LA: Elle0411 est responsable	27	Femme	SIAM1	91A	Guided conversation
24	91oaligg_TextGrid	2	23.8272	27.0830	travail0412 de3122puis un an. LA: Elle	0411	est responsable0421 au service0412 de31	27	Femme	SIAM1	91A	Guided conversation
25	91oaligg_TextGrid	2	23.8272	27.0830	puis un an. LA: Elle0411 est responsable	0421	au service0412 de3122 diététicien0413. l	27	Femme	SIAM1	91A	Guided conversation
26	91oaligg_TextGrid	2	23.8272	27.0830	Elle0411 est responsable0421 au service	0412	de3122 diététicien0413. LA: Elle0411 a é	27	Femme	SIAM1	91A	Guided conversation
27	91oaligg_TextGrid	2	23.8272	27.0830	11 est responsable0421 au service0412 de	1122	diététicien0413. LA: Elle0411 a été pron	27	Femme	SIAM1	91A	Guided conversation

Traitement de données de corpus

Visualisation et annotation du signal de parole

Phonometrica est compatible avec la plupart des formats de fichier son, notamment WAV, AIFF et FLAC. Les annotations peuvent être importées depuis, ou exportées vers, le logiciel Praat, mais aussi être stockées dans le format natif du logiciel, fondé sur le standard XML. Les annotations utilisent

le formalisme des graphes d'annotation développé par Bird et Liberman (2001) : une annotation contient un ou plusieurs *niveaux d'annotation*, sur lesquels sont indiqués des *événements* (des *instants*, qui sont ponctuels, ou des *intervalles*, qui sont bornés) ; l'ensemble des bornes temporelles et des événements forment un graphe. Ce modèle de données permet de représenter explicitement des relations entre les différents niveaux d'annotation (par exemple entre phonèmes et syllabes, ou entre mots et phrases) et de partager des bornes temporelles entre niveaux, celles-ci restant synchronisées sur tous les niveaux lorsqu'elles sont modifiées.

Phonometrica offre un module d'annotation et de visualisation du signal convivial et facile d'utilisation (voir figure 12.1). De par sa conception, il tente de remédier à un problème identifié par Meunier et Nguyen (2013, p. 97-98) : le besoin constant qu'ont les annotateurs et annotatrices de naviguer entre une représentation large et une représentation étroite du signal de parole. Pour ce faire, le logiciel offre une visualisation de l'ensemble du fichier sous la forme d'une onde (la *barre d'onde*) en bas de la fenêtre de visualisation, ainsi qu'un zoom sur la portion du signal sur laquelle on souhaite travailler. Cette approche permet de se repérer et de naviguer aisément dans le signal : on peut ainsi sélectionner une portion du signal directement dans la barre d'onde, puis ajuster le niveau de zoom à l'aide de boutons dans la barre d'outils ou de la molette de la souris.

Phonometrica offre les principaux types de visualisation attendus des logiciels d'analyse phonétique, à savoir l'onde sonore, le spectrogramme ainsi que l'estimation des formants, de l'intensité et de la fréquence fondamentale⁴. Ces trois derniers paramètres acoustiques peuvent être mesurés manuellement (voir figure 12.1, qui illustre la mesure de formants) ou de manière automatique à l'aide de requêtes acoustiques (voir section « Requêtes acoustiques »). Outre la sélection des visualisations à afficher, il est possible de sélectionner les niveaux d'annotation et/ou les canaux, ainsi que de moyenner tous les canaux pour obtenir une représentation « mono » du signal. Il est donc possible de n'afficher que les informations essentielles à la tâche en cours, sans surcharger l'interface graphique, ce qui est un problème fréquent lorsque l'on travaille sur des fichiers contenant des annotations multiniveaux riches et/ou plusieurs canaux audio.

4. Les formants sont estimés selon la méthode *Linear Predictive Coding* (LPC) à l'aide de l'algorithme de Burg (Andersen, 1974). L'estimation de la fréquence fondamentale s'appuie, à l'heure actuelle, sur l'algorithme SWIPE de Camacho (2007).

Métadonnées

Les métadonnées jouent un rôle essentiel. Elles permettent non seulement d'organiser et d'identifier les fichiers, mais aussi de filtrer les résultats d'une requête (voir section « Requêtes textuelles »). Le mécanisme d'annotation qu'offre Phonometrica, qui se veut à la fois simple et flexible, repose pour l'essentiel sur un ensemble de paires clé/valeur appelées *propriétés*. On peut étiqueter chaque fichier à l'aide d'un nombre arbitraire de propriétés. La clé d'une propriété représente une catégorie, qui est toujours une chaîne de caractères, alors que la valeur peut être une chaîne de caractères, une valeur booléenne (vrai ou faux) ou une valeur numérique, selon la nature de la métadonnée. À titre d'exemple, on pourrait créer, en fonction des besoins du projet, une catégorie « Participant », avec pour valeurs possibles les identifiants des sujets, une catégorie « Âge » associée à la valeur numérique correspondante, ou encore une catégorie « Tâche » en fonction de la tâche effectuée dans un enregistrement (ex. : « lecture de texte », « répétition de mots », « entretien semi-dirigé », etc.). Les propriétés sont gérées, au choix : a) manuellement, à l'aide de l'éditeur de propriétés, qui est accessible à partir du panneau d'information ou du menu contextuel du gestionnaire de fichiers ; b) de manière semi-automatique en les important d'un fichier tabulé au format CSV ; c) automatiquement, en utilisant le moteur de scripts.

Les fichiers peuvent par ailleurs être enrichis d'une description, qui est un champ pouvant contenir n'importe quelle séquence textuelle Unicode valide, et servir à trouver des fichiers à l'aide de la fonction de recherche rapide dans le navigateur de fichiers, ou encore à filtrer les résultats d'une requête. Enfin, Phonometrica permet de lier explicitement une annotation à un fichier audio : cette liaison se fait de manière automatique si le fichier audio et l'annotation ont le même nom de fichier (à l'extension près), mais elle peut aussi être effectuée manuellement ou automatiquement à l'aide d'un script.

Les métadonnées des annotations enregistrées dans le format natif de Phonometrica sont stockées directement dans le fichier d'annotation. Les métadonnées des annotations au format TextGrid de Praat sont quant à elles stockées dans une base de données embarquée dans le logiciel, puisque Praat ne permet pas l'ajout de métadonnées. Ce mécanisme est néanmoins entièrement transparent pour la personne utilisatrice, si bien qu'il est possible de combiner des annotations natives et au format TextGrid dans un même projet. Notons enfin que toutes les métadonnées d'un projet peuvent être exportées au format CSV afin de maximiser l'interopérabilité du logiciel.

Requêtes

Outre ses fonctionnalités d'analyse et d'annotation du signal, Phonometrica possède un moteur de recherche grâce auquel on peut exécuter des requêtes d'extraction d'informations textuelles ou acoustiques. À noter: ce logiciel distingue explicitement la *requête* proprement dite (ex.: «trouver tous les schwas réalisés en syllabe initiale de mot produits par des femmes»), du résultat qu'elle produit, à savoir une *concordance*, qui est une liste ordonnée d'occurrences représentée sous forme de tableau de données. Requête et concordance peuvent être sauvegardées et modifiées indépendamment l'une de l'autre: cela permet notamment de relancer aisément une requête complexe si le corpus a été mis à jour, par exemple si des fichiers ont été ajoutés, supprimés ou modifiés. Ces fonctionnalités de recherche sont disponibles à partir du menu Analysis.

Requêtes textuelles

Phonometrica permet d'extraire toutes les occurrences d'une chaîne de caractères correspondant à un événement ou contenues dans un événement. Il peut s'agir de chaînes de texte brutes, sensibles ou non à la casse selon les besoins, mais aussi de chaînes formulées à l'aide d'expressions régulières, un formalisme qui permet d'extraire des motifs tels que «tout mot commençant par *re-*» ou «toute séquence de 4 chiffres commençant par 0 ou 1» (voir Friedl, 2006 pour une introduction aux expressions régulières). Par défaut, la recherche est effectuée dans tous les niveaux de toutes les annotations du corpus. Il est toutefois possible d'ajouter des contraintes pour filtrer les résultats pertinents en spécifiant un niveau d'annotation particulier, un ensemble de fichiers et/ou un ensemble de métadonnées. Par exemple, si le corpus possède des métadonnées pour la tâche et le sexe, on pourrait limiter les résultats d'une requête aux personnes participantes de sexe féminin en tâche de lecture.

Les résultats d'une requête sont affichés sous forme de concordances dans un nouvel onglet de la visionneuse: la figure 12.2 illustre l'extraction de codages pour le schwa dans le cadre du projet PFC (voir section «Protocoles de requête» pour une discussion plus détaillée). Le mode de recherche par défaut est le «mot clé en contexte» (*keyword in context* ou KWIC), ce qui signifie que Phonometrica extrait toute occurrence de la chaîne de caractères ou du motif ainsi qu'une séquence de *n* caractères à gauche et à droite (40 par défaut). Il est possible d'extraire à la place le texte des événements à

gauche et à droite de celui dans lequel l'occurrence a été trouvée, ou encore de n'extraire aucun contexte. Outre le texte de l'occurrence et son contexte éventuel, Phonometrica extrait des informations sur le fichier et l'événement dans lequel l'occurrence a été trouvée, ainsi que les métadonnées associées au fichier. Chacun de ces éléments est représenté par une colonne dans la concordance, les occurrences apparaissant sous forme de ligne ; les colonnes à afficher peuvent être ajustées pour chaque concordance.

Pour chaque occurrence, plusieurs opérations sont disponibles à partir du menu contextuel (clic gauche sur l'occurrence) ou de la barre d'outils de l'onglet de concordance. On peut tout d'abord écouter l'occurrence, ce qui est notamment utile pour vérifier que le texte correspond bien au son, et donc valider l'occurrence si besoin. On peut également modifier le texte de l'événement de l'annotation, directement depuis la concordance, ou encore supprimer l'occurrence s'il s'agit d'un résultat erroné. Enfin, il est possible d'ouvrir directement le fichier son et l'annotation dans un nouvel onglet à l'endroit de la concordance, afin de vérifier le signal de parole. Si les fichiers d'annotation proviennent du logiciel Praat, Phonometrica offre par ailleurs la possibilité d'ouvrir les fichiers son et annotation dans Praat directement.

Chaque concordance peut être sauvegardée dans le format XML natif de l'application ; on peut donc la conserver et la modifier au besoin, lors des sessions d'utilisation. Le logiciel offre ici une approche intéressante pour la validation manuelle de grosses concordances contenant plusieurs milliers d'occurrences, puisqu'il permet un va-et-vient aisé entre la concordance et

FIGURE 12.2

Exemple de concordance pour le codage du schwa dans le projet PFC

2202 matches										Active dialog			
ID	Layer	Start time	End time	Left context	Target	Right context	Age	Gender	Speaker	Survey	Tools		
1	Phonlog.TextGrid	2	4.8888	8.3289	tu saur tuh. LA: Qui j'ai, tuh non avec 0412	no exp0412 j'ai trois ans et avec0412	27	Female	91A01	91A	Guided conversation		
2	Phonlog.TextGrid	2	4.8888	8.3289	LA: Qui j'ai, tuh non avec0412 tu saur 0412	j'ai trois ans et avec0412 non frère0412	27	Female	91A01	91A	Guided conversation		
3	Phonlog.TextGrid	2	4.8888	8.3289	0412 tu saur0412 j'ai trois ans et avec 0412	non frère0412 j'ai trois ans et avec 0412	27	Female	91A01	91A	Guided conversation		
4	Phonlog.TextGrid	2	4.8888	8.3289	012 j'ai trois ans et avec0412 non frère 0412	j'ai trois ans et avec0412 non frère 0412	27	Female	91A01	91A	Guided conversation		
5	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	travail0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
6	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
7	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
8	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
9	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
10	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
11	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
12	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
13	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
14	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
15	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
16	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
17	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
18	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
19	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			
20	Phonlog.TextGrid	2	15.2522	18.5997	elle0412 est distastieuse0413. LA: Elle 0412	privé du 0412 à l'hôpital0412 privé du 27	Female	91A01	91A	Guided conversation			

le signal de parole et qu'il n'est pas nécessaire d'effectuer la validation en une seule fois. Une fois la concordance validée, il est possible de l'exporter au format tabulé CSV, ce qui permet d'importer les résultats dans un tableur (ex. : Excel^{MD}) ou encore dans un logiciel d'analyse qualitative (ex. : Nvivo^{MD}) ou quantitative (ex. : SPSS^{MD} ou R).

Requêtes acoustiques

L'extraction de mesures acoustiques afin de les mettre en relation avec diverses variables sociales, notamment l'âge, le genre et la catégorie socio-économique est une tâche courante en sociophonétique. L'effectuer dans un logiciel tel que Praat demande certaines compétences en programmation, puisqu'il est nécessaire d'écrire (ou d'adapter) un script pour extraire ces informations dans un ensemble de fichiers. Une fois les résultats obtenus, le travail de validation, qui est une étape essentielle en analyse acoustique, se montre particulièrement fastidieux, du fait de l'absence de lien direct entre les mesures extraites et le signal de parole. Cette approche est donc complexe, chronophage et peu flexible.

Phonometrica simplifie considérablement ce travail grâce à son moteur de recherche et son système de métadonnées. Pour ce faire, il offre une interface permettant d'effectuer des requêtes acoustiques pour les formants, l'intensité ou la fréquence fondamentale. Le principe de fonctionnement est similaire aux requêtes textuelles présentées à la section « Requêtes textuelles », mais laisse la possibilité de spécifier les paramètres acoustiques pertinents pour le type de mesure (ex. : la bande de fréquence et le nombre de pôles pour l'extraction de formants) ainsi que le nombre de mesures à effectuer et leur position dans l'événement. Les mesures acoustiques apparaissent dans la concordance comme des colonnes supplémentaires. Les mêmes opérations que pour les requêtes textuelles sont alors possibles, ainsi que les allers-retours rapides et aisés entre les mesures extraites et le signal de parole, et la correction ou la suppression des mesures erronées.

Opérations logiques sur les concordances

Le système de métadonnées, couplé à l'utilisation d'expressions régulières, constitue un système puissant et flexible pour l'extraction et le filtrage de données dans un corpus. Phonometrica offre un mécanisme supplémentaire, qui permet de créer une nouvelle concordance à partir d'opérations logiques

sur deux concordances existantes. Les opérations possibles sont l'*union* (la fusion des deux concordances, avec élimination des doublons éventuels), l'*intersection* (l'ensemble des occurrences qui apparaissent dans les deux concordances) et le *complément* (l'ensemble des occurrences de la seconde concordance qui n'apparaissent pas dans la première).

Pour illustrer l'utilité de cette fonctionnalité, considérons le problème de l'extraction de formants. La personne utilisatrice souhaite généralement obtenir un tableau de données unique contenant tous les résultats du corpus. Néanmoins, pour des raisons physiologiques, il est habituellement recommandé d'analyser séparément les voix des hommes et celles des femmes, car elles nécessitent des paramétrages différents. On pourra donc, par exemple, effectuer une requête pour les hommes avec une bande de fréquence de 5 000 Hz, une requête pour les femmes avec une bande de fréquence de 5 500 Hz, puis fusionner les deux concordances ainsi obtenues.

Programmabilité et extensibilité

Bien que le logiciel ait été conçu avant tout pour être convivial et facile d'utilisation, il propose également des fonctionnalités avancées permettant de le personnaliser ou de l'étendre pour des besoins particuliers.

Moteur de script

Phonometrica possède un moteur de script spécifique, ainsi qu'une interface de programmation d'application qui permet d'accéder à certaines fonctionnalités et de contrôler certains aspects du logiciel⁵. Le langage de script est proche des langages de script modernes, tels que Python ou Ruby : il a été conçu pour être simple d'utilisation, lisible (sa syntaxe est proche du pseudo-code) tout en étant puissant et adapté au traitement du signal. Il dispose de fonctionnalités avancées de traitement de chaînes de caractères (Unicode, expressions régulières), de structures de données flexibles (listes, tableaux associatifs) ainsi que de tableaux numériques efficaces (vecteurs ou matrices) pour les opérations mathématiques. Un aspect original, et qui s'est avéré en pratique un choix judicieux, est que le langage de script est un surensemble du format d'échange de données JSON (Javascript Object Notation). En d'autres termes, tout fichier

5. Nous renvoyons à la documentation du logiciel (disponible en ligne et depuis le menu d'aide) pour une présentation détaillée du langage de script et des possibilités qu'il offre.

JSON est également un script Phonometrica valide. JSON est un format très utilisé, souvent qualifié de «XML léger», du fait qu'il excelle dans la représentation de données structurées, sans avoir la verbosité du XML. Ce format est donc abondamment utilisé par Phonometrica et joue un rôle pivot dans son système d'extension (voir section «Protocoles de requête»).

Les scripts peuvent être créés, édités ou exécutés de façon simple et conviviale grâce à l'éditeur de script intégré au logiciel. Parmi ses fonctionnalités, on mentionnera la coloration syntaxique, l'indentation automatique du code, la complétion de commande, des infobulles d'aide pour les fonctions intégrées ou encore l'exécution partielle du code et la localisation des erreurs d'exécution. Ces fonctionnalités simplifient et accélèrent considérablement le développement de scripts.

Protocoles de requête

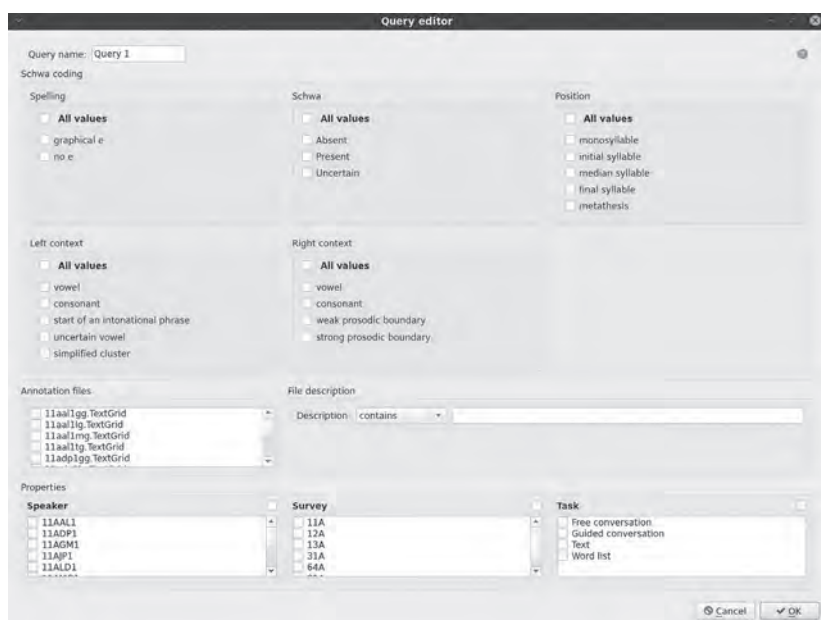
La recherche en sociolinguistique, et notamment en sociophonologie, est très souvent fondée sur le codage de variables linguistiques. L'interface standard de Phonometrica permet déjà d'extraire facilement ce type de codages, soit en cherchant chaque occurrence d'un codage donné, soit en utilisant des expressions régulières. Néanmoins, cette approche «brute» ignore complètement la sémantique du codage, puisque celui-ci est généralement composé de différents champs, chacun ayant une signification précise (la réalisation de la variable, son contexte phonologique, morphologique ou syntaxique, la partie du discours ou toute autre information utile pour l'analyse). Pour pallier ce problème, Phonometrica permet de définir, pour un système de codage donné, un *protocole de requête*, qui spécifie explicitement sa structure. Le protocole est implémenté dans un fichier JSON qui décrit, pour chaque champ, les valeurs possibles et leur signification. Ce fichier est ensuite interprété par le logiciel et transformé en un éditeur de requête sur mesure grâce auquel la personne utilisatrice peut sélectionner les champs à extraire, à l'aide de la souris, sans devoir manipuler le codage directement.

Nous illustrerons cette approche à l'aide du système développé pour le codage du schwa (ou *e* caduc, *e* instable, *e* muet) dans le cadre du projet PFC (Racine *et al.*, 2016). Cette voyelle peut être réalisée (ex.: *cette semaine* [setsəmə]) ou non (ex.: *la s(e)maine* [lasmə]) en fonction de l'interaction complexe de divers facteurs linguistiques et extralinguistiques. Le système de codage, qui est un système numérique à quatre chiffres, est présenté à l'annexe A. Le fichier JSON qui implémente le protocole de requête corres-

pondant est, quant à lui, présenté à l'annexe B. Lorsque le logiciel charge ce fichier, il produit un éditeur de requête spécifique pour ce système de codage (voir figure 12.3). La personne utilisatrice peut alors simplement cliquer sur les valeurs qu'elle souhaite obtenir (ou ne cliquer sur rien si elle souhaite extraire tous les codages, sans avoir à nécessairement connaître les détails du système de codage, ou à s'en souvenir. Par ailleurs, le logiciel connaît la structure du codage et peut (optionnellement) séparer les champs dans la concordance qu'il produit, ce qui facilitera l'analyse si la concordance est exportée vers un logiciel d'analyse qualitative ou quantitative.

FIGURE 12.3

Éditeur de requête pour le codage du schwa dans PFC



Greffons

Bien que le développement de scripts et de protocoles de requête requière certaines compétences en informatique, Phonometrica en facilite très largement l'utilisation grâce à un système de *greffons* (*plugins*). Un greffon est une extension qui permet de configurer ou d'ajouter des fonctionnalités à un logiciel. Phonometrica permet de regrouper des scripts et des protocoles

dans une archive au format ZIP, qui peut ensuite être redistribuée et installée sur la machine d'utilisateurs tiers. Une fois le greffon installé, les fonctionnalités qu'il offre sont disponibles dans le menu Tools de l'application.

Cette approche est particulièrement utile pour mettre à la disposition d'une communauté d'utilisateurs et d'utilisatrices un ensemble de fonctionnalités pour le traitement et l'analyse d'un type de corpus particulier. En guise d'exemple, le greffon réalisé pour le projet PFC offre un script permettant d'extraire automatiquement certaines métadonnées à partir du nom des fichiers, ainsi que des protocoles de requête pour l'analyse du schwa et de la liaison.

* * *

Ce tour d'horizon a mis en exergue les principales fonctionnalités de Phonometrica, un logiciel en constante évolution, qui s'améliore au fil des versions et en fonction des retours de sa communauté d'utilisateurs et d'utilisatrices. Notre chantier principal, à l'heure actuelle, porte sur la mise en œuvre de requêtes complexes d'identification des relations de dominance (ex. : une syllabe domine un phonème), de précédence (ex. : un déterminant précède un nom) ou d'alignement (ex. : un mot est aligné avec une partie du discours). Plusieurs prototypes sont en cours d'évaluation afin de déterminer la manière la plus conviviale de construire des requêtes et d'en afficher les résultats dans une concordance.

Nous travaillons par ailleurs à l'implémentation d'une méthode proposée par Weenink (2015) pour améliorer l'analyse de formants. La méthode d'estimation des formants utilisée par la plupart des logiciels de traitement du signal, à savoir la méthode LPC, est une approche paramétrique : les résultats qu'elle produit peuvent donc varier en fonction de la valeur des paramètres (la bande de fréquence choisie et le nombre de pics spectraux attendus dans cette bande). La méthode de Weenink permet de déterminer automatiquement les paramètres optimaux pour chaque segment analysé, en sélectionnant les valeurs qui produisent les pistes de formants les plus lisses, et donc celles qui contiennent *a priori* le moins d'erreurs d'estimation.

À plus long terme, nous envisageons d'ajouter des fonctions statistiques pour pouvoir effectuer des analyses quantitatives directement dans Phonometrica, sans qu'il soit nécessaire d'exporter les résultats vers un logiciel tiers. Nous sommes par ailleurs ouverts à toute suggestion de la communauté pour répondre à des besoins que nous n'aurions pas anticipés.

Références

- ANDERSEN, N. (1974), On the calculation of filter coefficients for maximum entropy spectral analysis, *Geophysics*, 39(1), 69-72.
- BIRD, Steven et Mark LIBERMAN (2001), A Formal Framework for Linguistic Annotation, *Speech Communication*, 33(12), 23-60.
- BOERSMA, Paul et David WEENINK (2023), Praat: doing phonetics by computer [Programme informatique], version 6.1.38, <http://www.praat.org/>.
- CAMACHO, Arturo (2007), A sawtooth waveform inspired pitch estimator for speech and music, thèse de doctorat, University of Florida.
- DETEY, Sylvain, Jacques DURAND, Bernard LAKS et Chantal LYCHE (dir.) (2016), *Varieties of Spoken French*, Oxford University Press.
- DURAND, Jacques, Ulrike GUT et Gjert KRISTOFFERSEN (dir.) (2014), *The Oxford Handbook of Corpus Phonology*, Oxford University Press.
- DURAND, Jacques, Bernard LAKS et Chantal LYCHE (2002), La phonologie du français contemporain: usages, variétés et structure, dans Claus D. PUSCH et Wolfgang RAIBLE (dir.), *Romanistische Korpuslinguistik – Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language*, Gunter Narr Verlag, 93-106.
- FRIEDL, Jeffrey E. F. (2006), *Mastering Regular Expressions* (3^e éd.), O'Reilly Media.
- HEDLUND, Gregory et Yvan ROSE (2020), Phon [Programme informatique], version 3.1, <https://phon.ca>.
- HONEY, John (2017), Sociophonology, dans Florian COULMAS (dir.), *The Handbook of Sociolinguistics*, Wiley-Blackwell, 92-106.
- MEUNIER, Christine et Noël NGUYEN (2013), Traitement et analyse du signal de parole, dans Noël NGUYEN et Martine ADDA-DECKER (dir.), *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, Hermès Science Publications, 85-119.
- PRZEWOZNY ANNE, Cécile VIOLLAIN et Sylvain NAVARRO (dir.) (2020), *The Corpus Phonology of English: Multifocal Analyses of Variation*, Edinburgh University Press.
- RACINE, Isabelle, Jacques DURAND et Helene N. ANDREASSEN (2016), PFC, codages et représentations: la question du schwa, *Corpus*, 15, 1-16.
- SLOETJES, Han et Peter WITTENBURG (2008), Annotation by category: ELAN and ISO DCR, dans *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- SPTK WORKING GROUP (2018), Speech Signal Processing Toolkit (SPTK) [Programme informatique], version 3.11, <http://sp-tk.sourceforge.net>.
- WEENINK, David (2015), Improved formant frequency measurements of short segments, dans *Proceedings of ICPhS 2015*.
- WINTER, Bodo et Andy WEDEL (2016), Commentary. Desiccation and tone within linguistic theory and language contact research, *Journal of Language Evolution*, 1, 80-82.
- ZWIRNER, Eberhard et Kurt ZWIRNER (1936), *Grundfragen der Phonometrie*, Metten.
- ZWIRNER, Eberhard et Kurt ZWIRNER (1936), *Principles of Phonometrics*, University Press Of Alabama, 1970. Traduction de Zwirner et Zwirner.

Annexes

A. Système de codage du schwa dans le projet PFC (d'après Racine et al., 2016)

- Champ 1: Réalisation
 - o 0: absent
 - o 1: présent
 - o 2: incertain
- Champ 2: Position dans le mot
 - o 1: monosyllabe
 - o 2: syllabe initiale de polysyllabe
 - o 3: syllabe interne de polysyllabe
 - o 4: syllabe finale
 - o 5: métathèse
- Champ 3: Contexte gauche
 - o 1: VC
 - o 2: CC
 - o 3: début de groupe intonatif
 - o 4: schwa incertain à gauche
 - o 5: groupe consonantique simplifié
 - o 6: voyelle immédiatement à gauche
- Champ 4: Contexte droit
 - o 1: voyelle
 - o 2: consonne
 - o 3: frontière intonative forte
 - o 4: frontière intonative faible

B. Protocole de requête pour le codage du schwa dans PFC

```
{
  "name": "Schwa coding",
  "version": "0.2",
  "type": "coding_protocol",

  "separator": "",
  "layer_index": 2,
  "fields_per_row": 3,
```

```

"fields": [
  {
    "name": "Spelling", "match_all": ".",
    "values": [
      {
        "match": "e", "text": "graphical e",
        "match": "[^e]", "text": "no e"
      }
    ]
  },

  {
    "name": "Schwa", "match_all": "[0-2]",
    "values": [
      {
        "match": "0", "text": "Absent",
        "match": "1", "text": "Present",
        "match": "2", "text": "Uncertain"
      }
    ]
  },

  {
    "name": "Position", "match_all": "[1-5]",
    "values": [
      {
        "match": "1", "text": "monosyllable",
        "match": "2", "text": "initial syllable",
        "match": "3", "text": "median syllable",
        "match": "4", "text": "final syllable",
        "match": "5", "text": "metathesis"
      }
    ]
  },

  {
    "name": "Left context", "match_all": "[1-6]",
    "values": [
      {
        "match": "1", "text": "vowel",
        "match": "2", "text": "consonant",
        "match": "3", "text": "start of an intonational phrase",
        "match": "4", "text": "uncertain schwa",
        "match": "5", "text": "simplified cluster",
        "match": "6", "text": "vowel directly preceding schwa"
      }
    ]
  },

  {
    "name": "Right context", "match_all": "[1-4]",
    "values": [
      {
        "match": "1", "text": "vowel",
        "match": "2", "text": "consonant",
        "match": "3", "text": "weak prosodic boundary",
        "match": "4", "text": "strong prosodic boundary"
      }
    ]
  }
]
}

```


CHAPITRE 13

Transcrire des corpus linguistiques à l'aide de la technologie intelligente

L'exemple du COLEM

Marimar Rufino Morales

Contextualisation de la redite

La révolution instrumentale des dernières décennies a métamorphosé la méthodologie de préparation des corpus linguistiques. Elle a conduit à l'automatisation de plusieurs tâches en allégeant, entre autres, les étapes de collecte et de stockage de matériel, d'alignement et d'annotation. En revanche, elle semble n'avoir que peu transformé la mécanique ou les outils employés pour la transcription des enregistrements destinés à l'étude de la langue parlée; celle-ci est généralement effectuée à l'aide d'un lecteur et d'un clavier¹. C'est le constat que nous avons fait lorsque nous nous sommes intéressée aux recherches en dialectologie du Dr Pato. Il s'agirait à bâtir le Corpus oral de la langue espagnole à Montréal (COLEM), un ensemble de données qui reflètent l'espagnol de la région métropolitaine, que l'on peut voir comme une pépinière de création langagière sans égal: d'une part, on y trouve la plus grande concentration d'Hispano-Américains au Canada, et donc les conditions propices aux contacts entre de nombreuses variantes de l'espagnol; d'autre part, leur interaction avec le français et l'anglais donne

1. Le clavier actuel est un périphérique de l'ordinateur qui sert à lancer une commande ou à saisir une information. Il s'agit d'une version à peine améliorée du clavier initial des machines à écrire, lequel associe chaque touche à un caractère (Gubern, 2010).

lieu à d'autres phénomènes intéressants de variation linguistique (Pato, 2022). Le chercheur constituait le COLEM à partir d'un échantillonnage non probabiliste, en sélectionnant les informateurs selon la technique de la boule de neige, rendue possible grâce aux réseaux des personnes étudiant à la Section d'études hispaniques de l'Université de Montréal, les mêmes qui menaient les entrevues et les transcrivaient à l'aide d'un lecteur et d'un clavier.

Ce constat a été pour le moins surprenant car, travaillant à la télévision nous² avions depuis longtemps délaissé le clavier traditionnel – et même la version syllabique, plus rapide, le clavier vélotype – au profit de la technologie intelligente et de la saisie de texte au moyen, entre autres, de notre voix. La mutation s'inscrivait dans la dynamique des changements technologiques touchant l'ensemble de l'industrie audiovisuelle à l'échelle internationale, régie par un cadre législatif fondé sur le principe, reconnu par les Nations Unies, de chances égales d'accès à l'information. Au travers du prisme des télédiffuseurs, qui s'étaient engagés à sous-titrer progressivement la totalité de leurs émissions, les méthodes de transcription rapide encore en place au tournant du 21^e siècle s'avéraient financièrement contraignantes. En effet, on peinait à trouver du personnel qualifié dans un secteur considéré comme moribond, alors que les avancées dans les technologies de traitement automatique de la parole laissaient présager l'automatisation du sous-titrage. Même si, dans les faits, malgré l'utilisation de logiciels conçus sur mesure et entraînés spécifiquement pour la tâche à accomplir, la qualité des sous-titres automatiques était décevante. Outre les performances limitées des logiciels face à la variabilité de la langue spontanée et à la diversité des profils linguistiques, la compréhension des textes affichés à l'écran était amoindrie par l'absence, dans les sous-titres automatiques, d'un ensemble de codes normalisés généralement insérés par les professionnels et professionnelles (tels que la ponctuation, l'identification des voix intervenant dans l'interaction ou encore les descriptions des bruits environnants). L'industrie audiovisuelle a néanmoins choisi de traiter un obstacle à la fois, en misant, dans un premier temps, sur les avantages de l'automatisation. Une fois les coûts de production réduits grâce au recours aux sous-titres automatiques, il ne restait qu'à trouver un moyen de transformer ces transcriptions brutes en textes acceptables à l'écran. Les efforts des préposés et préposées au sous-

2. Au début des années 2000, je sous-titrais en temps réel des émissions télévisées pour les chaînes du Groupe TVA.

titrage ont permis d'atteindre les résultats espérés, grâce à l'introduction de pratiques d'interactions vocales avec les logiciels de transcription automatique, converties ensuite en méthodes transférables. Dans le meilleur des cas, il leur était même possible de retoucher la transcription automatique avant qu'elle ne soit transformée par un autre logiciel en sous-titres affichés et diffusés dans un très court délai (Baaring, 2006 ; Brousseau *et al.*, 2003 ; Lambourne, 2006). Cette manière de produire des sous-titres en temps réel s'est fait connaître sous le nom de *respeaking* (Romero-Fresco, 2011).

Le parallélisme entre l'oralité présente dans la plupart des scénarios des médias audiovisuels et celle qui caractérise les données collectées pour créer un corpus de la langue parlée nous a suggéré l'hypothèse que l'adaptation des stratégies du *respeaking* à la transcription du matériel oral destiné à l'analyse linguistique pouvait améliorer les méthodes en usage en recherche appliquée.

La technique

En *respeaking*, c'est un être humain qui – installé dans un environnement exempt de bruits et muni d'un casque, d'un micro et d'un ordinateur avec un clavier, d'une souris et, quelquefois, d'une manette de jeu vidéo – fait office de médiateur entre la situation communicative à transcrire et un logiciel de reconnaissance automatique de la parole (RAP). Il s'affaire à : a) écouter et parler en même temps ; b) anticiper les possibles erreurs de son outil en reformulant sans délai forme et contenu et c) ajouter les signes de ponctuation et toute autre information paralinguistique et prosodique requise. En réduisant certaines variabilités propres à l'oralité (voix d'homme, de femme ou d'enfant aux accents différents, chevauchement des voix, bruits ambiants captés par des micros de plus en plus performants, mais aussi des pauses, des hésitations, des reprises ou des autocorrections), le nouvel intrant améliore les résultats du logiciel.

En français, le terme pour désigner cette technique, qui consiste à écouter le son original d'une émission et à le « redire » à un logiciel de RAP en utilisant une série de stratégies qui l'aideront à produire de meilleurs résultats, n'a jamais été normalisé. Certains font du *respeaking* (Swiss TXT), tandis que d'autres font le perroquet (TF1, France 2), de la reconnaissance vocale³ (CRTS, TVA), du sous-titrage vocal (TERMIUM Plus) ou encore

3. Emploi métonymique critiqué.

du sous-titrage en direct (SOVO). Nous avons adopté l'identificateur unique « redite », qui fait référence à la technique de production des transcriptions vocales, indépendamment du mode employé, en direct (*on-line*) ou en différé (*off-line*).

État des lieux : transcrire des enregistrements pour la recherche linguistique

Transcrire est une pratique habituelle dans plusieurs disciplines qui exploitent des enregistrements comme matériel d'analyse. Elle stimule les réflexions méthodologiques et théoriques, et apporte rigueur, précision et maniabilité : le matériel écrit est plus facilement manipulable et partagé que l'oral (Ives, 1995). En recherche appliquée, une transcription peut servir à la fois à formuler des questions et à répondre à des questions prédéfinies. Dans un cas comme dans l'autre, en linguistique, en sociologie ou en sociolinguistique, on souhaitera recueillir une représentation adéquate de l'expression orale spontanée d'une communauté, et les procédés d'échantillonnage conduiront à effectuer un grand nombre d'entretiens ; la transcription des enregistrements constitue une bonne façon de rendre accessibles les données recueillies (Sankoff *et al.*, 1976). Auparavant, la prise de notes sur le terrain était couramment utilisée pour garder une trace écrite des communications orales, mais les notes de terrain ne fournissaient pas le niveau de détail nécessaire au travail de recherche. Ainsi, dès que la technologie a rendu possible l'enregistrement des conversations avec un magnétophone, la transcription a gagné du terrain.

La transcription de l'oralité cherche à rendre compte d'un acte de communication complexe, qui se déroule dans une situation d'interaction à la fois personnelle et contextuelle. Cet acte s'exprime par des échanges de signes et de gestes verbaux et non verbaux, conscients et inconscients, dans le flux irrégulier des prises de parole. Il mobilise trois canaux complémentaires – verbal, paralinguistique et kinésique – correspondant à ce que nous disons, à la manière dont nous le disons, et à la façon dont nous le gesticulons ou le somatisons. Ces trois canaux partagent des fonctions lexicales et grammaticales. Ils présentent à la fois des caractéristiques segmentaires (mots, pauses, gestes, postures) et non segmentaires (intonation, modulations de la voix, mouvements, positions). Ils peuvent se substituer les uns aux autres et sont tous les trois spécifiques à une langue et à une culture, tout en étant sujets à de nombreuses variations. En communication orale, chaque élément

compte – y compris ce qui ne se manifeste pas à travers des mots – ; la transcription, elle, s'efforce d'en restituer la dynamique (Poyatos, 2002).

L'enregistrement de l'entrevue, sa transcription et son codage, les notes de terrain ainsi que d'autres notes décrivant la situation de communication et le contexte dans lequel elle s'est produite sont autant de matériaux qui forment le corpus de recherche (Lapadat et Lindsay, 1999).

Du matériel à transcrire au produit transcrit

La transcription d'une conversation, loin d'être une tâche purement mécanique, nécessite de nombreuses prises de décisions, fondées sur les théories que l'on tente de soutenir, et qui visent à représenter authentiquement par écrit à la fois les données orales et toutes les informations contextuelles de l'enregistrement. C'est une quête d'équilibre entre représentativité et exhaustivité dans la saisie d'événements de communication évanescents, afin de les rendre accessibles pour une analyse ultérieure. Toutefois, l'action en elle-même a comme résultat de limiter le champ des interprétations possibles des données, puisque ce sont ces décisions prises pendant l'acte de la transcription qui définissent le résultat et qu'il n'existe pas de décision neutre. Chaque décision influence et contraint les lectures et utilisations futures du produit. Il s'agit de savoir, avant toute chose, si l'enregistrement mérite d'être transcrit. Il convient ensuite de déterminer l'étendue du matériel enregistré à inclure dans la transcription. Si la quantité ne définit pas à elle seule la qualité du résultat, cette dernière ne peut néanmoins être envisagée sans une certaine mesure quantitative. Par ailleurs, il n'existe ni règles absolues ni critères universels permettant de mesurer la qualité d'une transcription. Celle-ci demeure, avant tout, un outil interprétatif. Néanmoins, plusieurs spécialistes de la linguistique réclament que le matériel (oral et écrit) soit aussi complet que possible, dont Rojo (2021). En prenant le Brown Corpus⁴ comme exemple, il soutient que c'est, historiquement, en raison de la nécessité de limiter la taille des échantillons, que des fragments d'enregistrements ont été employés. Maintenant que la technologie a repoussé ces frontières, la meilleure façon d'étudier les phénomènes linguistiques est de travailler avec des textes complets. Il faut aussi déterminer quels codes linguistiques

4. Le *Standard Sample of Present-Day American English* était constitué d'environ un million de mots extraits de 500 textes écrits et imprimés aux États-Unis en 1961, et sélectionnés selon des critères spécifiques afin d'assurer la représentativité des locuteurs natifs de l'anglais américain (Francis et Kucera, 1979).

et autres codes non verbaux accompagnant la parole doivent être reflétés à l'écrit, lesquels ne devront pas l'être et pourquoi. Enfin, quelles conventions de transcription devront être utilisées. On peut notamment représenter les mots au moyen de l'orthographe conventionnelle, sachant malgré tout que ces conventions ne rendent pas compte de toute l'information perçue dans la parole enregistrée. Tantôt ce qui a été dit n'est pas catalogué, tantôt il n'y a pas de consensus quant à sa représentation graphique dans les ouvrages destinés à préserver l'unité de la langue et qui recueillent les éléments en usage général dans une communauté. C'est pourquoi on observe que, lorsqu'il s'agit de rendre compte de la variabilité de la parole, et bien que cela implique des efforts supplémentaires, certains chercheurs et chercheuses incorporent une représentation phonétique ou phonologique aux échantillons écrits. Il faut également préciser le niveau de détail, tout comme le niveau de littéralité. Ces informations incluent notamment la nécessité de se prononcer sur la représentation de chacune des dimensions de la voix, celle des phénomènes physiques qui y sont associés, ainsi que certains facteurs contextuels et environnementaux. Par exemple, certaines qualités de la voix identifient le sexe, l'âge, la région ou le dialecte d'origine de la personne qui parle; certains gestes renseignent sur l'état affectif, attitudinal et émotionnel; l'aboïement d'un chien, le son d'une cloche ou l'apparition d'un chat peuvent également modifier le contenu de la communication.

Plusieurs de ces décisions, qui ont à voir avec le niveau de détail – ou ce que Mondada (2008, 151) appelle « la granularité de la transcription » – ne concernent pas uniquement des éléments verbaux (ce qui a été dit), mais aussi des éléments prosodiques (la façon dont cela a été dit) et des éléments non verbaux (extralinguistiques et paralinguistiques). Habituellement, les informations prosodiques, extralinguistiques et paralinguistiques sont ajoutées lors d'une phase ultérieure à la transcription, sous forme d'étiquettes descriptives ou grâce à des annotations. Les étiquettes et les annotations peuvent apparaître, ou non, dans le corps du texte. Les personnes qui les placent dans le texte assurent qu'elles facilitent les recherches et aident à quantifier les données. Celles qui les rejettent estiment qu'elles entravent la lisibilité et la fluidité du texte, ce qui empêche la détection de certains phénomènes. Il est toutefois admis, de part et d'autre, que ces marques s'avèrent indispensables à la recherche linguistique et doivent être intégrées dans le processus de préparation du matériel.

Finalement, le matériel écrit nécessite des révisions, des corrections, des codifications avant de pouvoir être exploité. Autant de phases distinctes de

l'action de transcrire, même si leur relation d'interdépendance explique en partie qu'elles soient souvent assimilées à une seule étape de transcription (Adolphs, 2013), ce qui entraîne qu'elles soient fréquemment négligées lors de la planification d'un corpus.

Dans tous les cas, il n'existe pas de correspondance biunivoque entre l'enregistrement d'une conversation et sa transcription, comme il n'existe pas de système de transcription capable de coder tout ce que la parole exprime – ni tout ce qui doit être mis à la disposition d'un domaine d'étude – en une seule opération (Tilley et Powick, 2002). Il en est ainsi parce que, malgré leur continuum, la langue employée dans la communication orale et celle en usage dans la communication écrite ne sont pas les deux faces d'un même système ; ce sont deux systèmes parallèles et dissociés (Du Bois, 1991), du fait, entre autres, de leurs paramètres et de leurs caractéristiques : a) à l'oral, la communication est multidimensionnelle, par opposition à la dimension uniquement spatiale de la langue écrite ; b) la structure de la communication orale est flexible, sa forme discursive étant le dialogue et les locuteurs et locutrices s'adaptant au langage de l'autre, alors que la communication écrite est linéaire, et c) la langue employée dans la communication orale est spontanée, subjective et transitoire, avec un fort caractère implicite, du fait que tous ses participants se trouvent dans le même contexte, dont elle dépend (Willems, 1989).

Les outils et les méthodes de transcription

Toutes les questions pertinentes (longueur, format, représentativité, diversification, marquage, étiquetage, provenance, taille de l'échantillon, classification) sont soumises « aux objectifs de chaque chercheur et [aux] possibilités technologiques disponibles » (Parodi, 2007, p. 21). Dès lors, la constitution d'un corpus de recherche linguistique impose, pour garantir la rigueur de la transcription, le recours aux outils les plus appropriés et aux techniques les plus performantes.

William Soskin et Vera John ont réalisé, en 1953, le premier enregistrement acoustique d'une conversation naturelle d'un couple de jeunes mariés lors d'un camp d'été pour professeurs-es et étudiants-es diplômés-es de l'Université de Chicago. Les techniques de collecte de données basées sur l'enregistrement de la parole dans un contexte naturel, qu'ils ont préconisées, ont modifié à jamais la sociolinguistique et l'ethnographie (Jones, 2011). Bien que la transcription, destinée à recenser les modèles phonologiques et grammaticaux des

interactions sociales, n'ait été intégrée à la recherche que plus tard (il a fallu attendre l'émergence d'outils permettant d'enregistrer les interactions orales en temps réel de manière pratique et précise), l'introduction de la technologie d'enregistrement pour conserver les interactions sociales invitait à obtenir davantage de données.

Indépendamment des outils et des méthodes employés, toute tentative de capture de l'acte communicatif transitoire et éphémère de la transcription doit représenter une reproduction fidèle et littérale de l'enregistrement (Poland, 1995, 2002), sur un support fixe (papier ou tout autre dispositif visuel), dont les propriétés garantissent une certaine pérennité, tout en facilitant l'usage.

Transcription manuelle

Le débit moyen de production de la parole (ou vitesse d'élocution normale) à la radio française se situe entre 120 et 400 mots par minute (Pimsleur *et al.*, 1977). La parole spontanée des professionnels et professionnelles de la télévision en Espagne atteint régulièrement des pics allant jusqu'à 500 mots par minute (Rodero Antón et Campos Parra, 2005). Que ce soit avant ou après l'arrivée des ordinateurs, le processus de transcription dite manuelle commence toujours par lancer l'enregistrement, qui doit être arrêté chaque fois que l'on souhaite écrire ; il faut ensuite relancer le matériel audiovisuel en reculant légèrement jusqu'à ce qu'il coïncide avec le moment transcrit, puis reprendre l'opération autant de fois que nécessaire, jusqu'à ce que la transcription du matériel audiovisuel soit terminée. Bien que certaines personnes privilégient une pédale analogique ou numérique pour ces actions, il demeure que la transcription orthographique d'un enregistrement à l'aide d'un clavier impose de surmonter le décalage entre la vitesse de production orale en temps réel et la capacité à produire un texte écrit. En effet, personne n'est en mesure de taper à la vitesse de la parole, pas même Barbara Blackburn, qui détient le record du monde Guinness de vitesse de frappe pour avoir tapé 212 mots par minute pendant une minute et 150 mots par minute pendant cinquante minutes.

La transcription manuelle d'une heure d'enregistrement peut ainsi prendre de 4 à 60 heures (Markle *et al.*, 2011). Face à l'ampleur de la tâche, on est généralement amené à engager du personnel professionnel de la transcription. Les sténographes professionnels captent environ 200 mots par minute en temps réel grâce à une méthode d'écriture rapide apprise dans

des formations de deux ans (Del Valle López, 2004). Or, le financement adéquat⁵ fait souvent défaut, si bien que le travail est fréquemment confié à des auxiliaires ou à des étudiants.

Transcription automatique

Les logiciels de transcription automatique sont dans la mire de plusieurs personnes qui font de la recherche linguistique et qui espèrent accélérer le travail de transcription.

C'est ce qui a conduit Nagy et Sharma, en 2014, à examiner les « logiciels de transcription », Transcriber, Praat, CLAN, ELAN, TypeCraft et Fieldworks Language Explorer (FLEEx). Ces logiciels et d'autres, largement utilisés en recherche, tels que NVivo, HyperResearch, MaxQDA, Transana et Audacity, sont surtout des outils d'analyse de données qualitatives assistée par ordinateur. Ils permettent l'alignement (ou la synchronisation) du texte et du matériel audiovisuel, dans le but de faciliter l'annotation d'un corpus. Par exemple, plusieurs automatisent l'encodage des caractéristiques suprasegmentales de la prosodie, telles que l'intonation pour la transcription phonétique (Elvira-García *et al.*, 2015). Toutefois, la réalisation de ces manipulations requiert des transcriptions orthographiques fiables. Bien que certains de ces outils intègrent la technologie de RAP, au moment d'entreprendre notre recherche, les transcriptions générées automatiquement contenaient encore trop d'erreurs pour être considérées comme une solution de rechange fiable à la transcription manuelle (Llisterri, 2009; Tatham et Morton, 2005), en particulier lorsqu'il s'agit de langues disposant de ressources ou de données limitées.

Transcription vocale

Au cours des années 2000, de nombreuses personnes engagées dans la recherche qualitative sont conscientes que la transcription manuelle constitue une tâche chronophage et pénible (particulièrement pour celles qui rencontrent des difficultés physiques), et nourrissent encore l'espoir que la technologie simplifie le processus de transcription (Matheson, 2007). Elles observent que les logiciels de RAP reposent sur deux principes incompatibles

5. À titre d'exemple, on peut consulter les tarifs réglementés par le gouvernement provincial (www.legisquebec.gouv.qc.ca) et fédéral (www.cas-satj.gc.ca) pour la transcription judiciaire.

avec la réalité de la parole spontanée : d'une part, ils sont conçus pour reconnaître une seule voix ; d'autre part, cette voix n'est pas censée interagir avec d'autres, ni se superposer à elles. Pour contourner ces limites, elles proposent une stratégie qui consisterait à écouter l'enregistrement contenant plusieurs voix segment par segment, à l'arrêter pour répéter les propos entendus dans un microphone connecté au logiciel de RAP, puis à relire et à corriger chaque passage avant de continuer (Fogg et Wightman, 2000 ; MacLean *et al.*, 2004 ; Park et Echo Zeanah, 2005 ; Roberts Powers, 2005). De l'aveu même de ces personnes, la transcription vocale s'avère moins exigeante que le travail au clavier ; toutefois, le temps gagné lors de la dictée est en partie perdu dans la correction des erreurs du logiciel et des reformulations humaines.

Une autre initiative de saisie de texte avec la voix et un outil de RAP qui mérite que nous nous y intéressions ici concerne le domaine de la traduction. Diverses études ont montré que l'utilisation d'un logiciel de RAP pour dicter des traductions accroît la productivité (Dragsted *et al.*, 2009), ce qui permet de traduire presque quatre fois plus vite sans perdre de précision (Vidal *et al.*, 2006). Cette approche s'inspire de la traduction à vue, qui était largement répandue avant l'avènement de l'ordinateur personnel (Delisle et Otis, 2016). Elle consistait à lire mentalement phrase par phrase le texte source et à produire oralement un discours correspondant à la traduction du texte, lequel était ensuite enregistré sur un dictaphone avant d'être transcrit (Jiménez Ivars et Hurtado Albir, 2003).

Nous avons analysé les pratiques de transcription vocale assistée de la RAP et utilisées en recherche qualitative et en traduction. Nous les avons comparées avec le *respeaking* employé pour le sous-titrage en direct. Nous avons constaté que ce sont les stratégies fondées sur la communication multimodale humain-machine, telles qu'employées à la télévision, qui permettent d'optimiser la technologie. C'est pourquoi, en adaptant du *respeaking* à la transcription vocale en différée destinée à la recherche linguistique (redite *off-line*), nous avons inscrit parmi nos objectifs l'élaboration d'une marche à suivre et entrepris de recenser et d'analyser chaque geste posé par l'humain pour réduire les erreurs du logiciel.

Démarche expérimentale

Instruments

Les 36 enregistrements destinés à la collecte des données ont été extraits du COLEM: un échantillonnage⁶ constitué de 40,07 heures non transcrites d'interviews réalisées auprès d'immigrants originaires de 20 pays hispanophones. Y sont représentés tous les pays hispanophones énumérés dans le rapport de l'Instituto Cervantes (Fernández Vítóres, 2021) à l'exception de la Guinée équatoriale. Suivant la répartition censée coïncider avec la population, nous avons choisi un nombre égal d'hommes et de femmes. Il aurait été possible d'ajouter d'autres critères à cette sélection (par exemple, en tenant compte du fait que chaque locuteur n'est pas l'unique représentant de la langue de son pays et que sa manière de parler est influencée par son éducation, son niveau d'études ou son âge), mais la large représentation des pays hispanophones, équilibrée entre hommes et femmes, devait suffire à fournir un cadre représentatif de l'espagnol spontané.

Nous avons utilisé un ordinateur dont la puissance du processeur graphique dépassait les exigences minimales fixées par les fournisseurs des logiciels, le traitement de la parole avec des réseaux neuronaux profonds nécessitant de nombreux calculs simultanés.

Les logiciels de RAP ont été sélectionnés en fonction de leurs spécificités⁷: leur compatibilité avec l'espagnol, leur capacité à effectuer une reconnaissance à la fois indépendante et continue, leur robustesse en termes de performances et leur accessibilité. Nous avons retenu un logiciel payant qui permettait l'amélioration du répertoire lexicque et l'ajout de raccourcis, ainsi qu'un logiciel gratuit qui n'acceptait que l'ajout de raccourcis. Une fois le premier logiciel acquis, installé et entraîné, les mêmes raccourcis ont été mémorisés dans les deux logiciels sélectionnés.

6. Pour être représentatif, un échantillon de langue naturelle doit tenir compte de toutes ses variétés (Leech, 1991); en ce qui nous concerne, la compilation doit inclure le maximum de possibilités phonétiques, sémantiques et grammaticales des différentes variétés d'espagnol (Bouzouita *et al.*, 2018).

7. Les propriétés de chaque outil varient selon la configuration définie dans sa conception. Le logiciel est soit indépendant, soit dépendant du locuteur (dans ce cas, un entraînement est nécessaire); il peut être conçu pour un seul utilisateur ou pour plusieurs, ce qui le rend moins précis, mais accessible à un plus grand nombre; la reconnaissance peut porter sur des mots isolés ou sur la parole en continu; la robustesse ou la tolérance au bruit peut varier en fonction des conditions sonores de paramétrage; enfin, la taille du vocabulaire influe sur la vitesse de traitement: plus le répertoire lexicque est étendu, plus le temps nécessaire à l'estimation des probabilités augmente, de même que le risque de correspondances approximatives et d'erreurs.

Par ailleurs, pour la redite, nous avons fait usage d'un casque monaural équipé d'un microphone. Enfin, pour la lecture des enregistrements, nous avons employé le lecteur VLC (www.videolan.org), qui permet de contrôler la vitesse de lecture, et un casque BQuietComfort45.

Séquençage du travail

L'objectif était de comparer toutes les transcriptions obtenues pour chaque entrevue (manuelle, vocale, automatique), afin de déterminer laquelle pouvait être considérée comme la plus fidèle à l'enregistrement et la plus conforme au protocole de transcription du COLEM. Le séquençage des tests a débuté par la redite des 36 entrevues⁸. De l'unique enregistrement oral redit par entrevue, nous avons obtenu une transcription avec chacun des logiciels (donc, deux transcriptions par entrevue). Puis, nous avons soumis les 36 enregistrements originels des entrevues uniquement au logiciel acheté, et avons obtenu 36 transcriptions automatiques.

Pour chaque transcription, nous avons réalisé les révisions des transcriptions conformément aux exigences internationales relatives à la qualité des services linguistiques (ISO 17100:2015). Puis nous avons effectué une double révision comparée (en écoutant l'enregistrement original) et une double révision non comparée, tous les temps de travail étant comptabilisés.

L'évaluation de la qualité

Pour évaluer la qualité des transcriptions, nous avons mis en relation leur exactitude et leur niveau de détail et avons mesuré l'effort requis pour les produire.

Les performances de la reconnaissance automatique de la parole se mesurent selon le taux d'erreur par mot, à partir d'une formule connue sous l'acronyme WER, qui correspond à l'expression *Word Error Rate* ou *Word Error Ratio*. Le WER permet d'évaluer l'exactitude (*accuracy*) des transcriptions ou la proximité de la mesure par rapport à la valeur réelle à partir du nombre de mots insérés, substitués et supprimés, en mettant en parallèle le texte généré automatiquement et le modèle soumis comme valeur réelle. Or, puisque chacune de nos transcriptions devait être conforme à des normes

8. La redite de chaque entrevue se fait en temps réel, mais, pour maintenir un bon niveau de concentration, il faut compter une pause de 5 à 10 minutes toutes les 30 minutes.

et conventions de transcription bien précises, limiter l'évaluation à l'unité du mot posait problème. Avec l'aide du Centre de recherche informatique de Montréal (CRIM), nous avons modifié la formule pour prendre en compte non seulement les mots, mais tous les signes qui font qu'une transcription est conforme au protocole du COLEM. Le traitement automatique des données a été effectué à l'aide de la boîte à outils Kaldi.

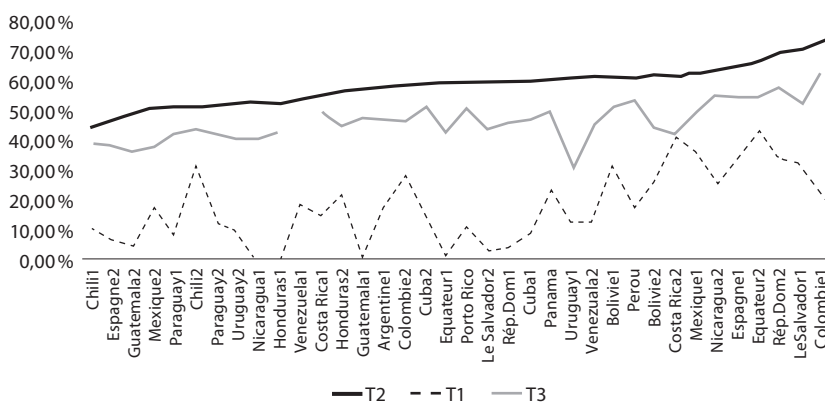
Résultats et discussion

En cohérence avec les observations déjà formulées dans le domaine du sous-titrage, la supériorité de la transcription vocale par rapport à la transcription manuelle dans le cadre du COLEM a été corroborée lors d'un atelier tenu à l'Université de Montréal, auquel participaient des étudiants et étudiantes aux cycles supérieurs en études hispaniques, ayant l'habitude de travailler sur des corpus. Nous y avons mesuré une vitesse moyenne de frappe de 63,33 mots par minute pour la saisie de texte à l'écran, proche de celle d'une personne expérimentée qui, pour l'espagnol, se situe autour de 80-90 mots par minute. Leur taux d'exactitude moyen était de 95,66 % ; il correspond à la moyenne établie pour la transcription humaine manuelle. Lorsque ces mêmes personnes ont été confrontées à l'écoute, durant 30 minutes, d'un enregistrement du COLEM, leur vitesse de transcription manuelle a chuté à 12,6 mots par minute ; le taux d'exactitude moyen a baissé à 65,53 %, et personne n'a transcrit au-delà de 3 minutes d'enregistrement (2,29 minutes en moyenne) (Rufino Morales, 2020).

Pour la transcription automatique et vocale, nous proposons d'observer ici la comparaison des transcriptions automatiques générées par le logiciel acheté (T1), celles obtenues à partir de la redite avec le même logiciel (T2) et, enfin, celles produites avec un logiciel gratuit en utilisant la même redite (T3). La figure 13.1 est tirée des statistiques descriptives et illustre bien les résultats d'exactitude. La ligne inférieure représente les taux d'exactitude des transcriptions T1. La ligne médiane indique le taux d'exactitude des transcriptions T3 (la transcription T3 de Venezuelaz a été écartée de cette analyse en raison d'un bogue logiciel ayant compromis sa conformité aux critères établis). La ligne supérieure illustre les taux d'exactitude des transcriptions T2.

Notre démarche expérimentale démontre que ce sont les transcriptions redites en temps réel qui produisent les résultats les plus stables ; et que celles effectuées avec un logiciel que l'on peut préconfigurer sont supérieures aux

FIGURE 13.1

Comparaison des performances (%) de haut en bas : T2, T3, T1

autres. Le nombre de signes codifiés corrects de T1 est significativement inférieur à celui de T3 et de T2. Le nombre de signes codifiés corrects de T3 l'est aussi par rapport à T2. Le taux d'exactitude de la meilleure transcription automatique (T1) a été celui de l'entrevue Ecuador2, correspondant à un taux d'exactitude de 44,76 %, plus bas que ceux de T2 et T3, pour la même interview redite, qui se situent respectivement à 67,15 % et 56,26 %.

D'autre part, la comparaison entre les transcriptions automatiques (T1) et les transcriptions issues de la redite (T2), obtenues à partir du même logiciel, met en évidence les éléments essentiels suivants :

- Les signes de ponctuation sont généralement absents dans les transcriptions T1; en l'absence de ces signes de détachement, le cotexte transcrit est parsemé d'erreurs de suppression et de substitution additionnelles. Les signes de ponctuation dictés ont été correctement représentés dans les transcriptions T2.
- Le changement d'interlocuteur ou d'interlocutrice n'est pas identifié dans les transcriptions T1; l'absence de signes d'identification déclenche plusieurs erreurs de suppression et de substitution additionnelles. Les étiquettes utilisées pour identifier les prises de parole dictées sont correctement représentées dans les transcriptions T2.
- Les transcriptions T1 ne sont pas anonymisées. Puisque les renseignements personnels sont volontairement omis en redite et remplacés par

des étiquettes d'anonymisation, les transcriptions T2 sont plus respectueuses des normes et conventions de transcription.

- Les événements non verbaux et paralinguistiques sont à l'origine de plusieurs erreurs d'insertion, de substitution et de suppression dans les transcriptions T1. Ils ont été omis de façon intentionnelle en redite et remplacés par des étiquettes qui permettent d'identifier chacun des phénomènes, ce qui rend les transcriptions T2 plus respectueuses des normes et conventions de transcription.
- Les prononciations inattendues ou atypiques sont à l'origine de plusieurs erreurs (insertions, substitutions, suppressions) dans les transcriptions T1. L'autocorrection intentionnelle en redite a fait baisser le nombre d'erreurs dans les transcriptions T2.
- Chaque mot absent du répertoire lexicale du logiciel de transcription est à l'origine de plusieurs erreurs (insertions, substitutions, suppressions) dans les transcriptions T1. L'omission intentionnelle en redite des termes non répertoriés a fait baisser le nombre d'erreurs de reconnaissance automatique dans les transcriptions T2.
- Les mots tronqués, les hésitations, les répétitions volontaires ou involontaires sont à l'origine de plusieurs erreurs (insertions, substitutions, suppressions) dans les transcriptions T1. Leur omission intentionnelle en redite et l'ajout d'étiquettes d'identification des phénomènes a fait baisser le nombre d'erreurs dans les transcriptions T2.
- L'alternance codique et les emprunts (p. ex. « todas tienen su *charme*, lo que les encantaba era el *charme* » (Costa Rica²)) sont à l'origine de plusieurs erreurs (insertions, substitutions, suppressions) dans les transcriptions T1. Des stratégies de redite incluant l'omission intentionnelle et l'ajout d'étiquettes d'identification des phénomènes ont fait baisser le nombre d'erreurs dans les transcriptions T2.

En général, les calques, formés à partir de mots propres à l'espagnol auxquels on attribue un sens qui appartient à la langue d'emprunt (ex. « estufas a gas » (Bolivia²); « allí pues *habité*, perdón, viví », « donde *habito* » (Colombia¹), ne font pas partie des erreurs observées dans les transcriptions, pour autant que les éléments lexiques mis à contribution soient intégrés dans le répertoire du logiciel.

Les transcriptions obtenues à partir des T2 sont non seulement les plus riches, pleinement conformes aux conventions du COLEM, mais ont également été réalisées dans un délai réduit. On a utilisé en moyenne 8,3 minutes

par minute d'enregistrement pour obtenir les transcriptions finales à partir des T2, la redite en temps réel, les révisions et les corrections étant comprises dans ce laps de temps. En raison des fluctuations des performances de la RAP, aucune transcription T1 ne remplit de façon satisfaisante les conditions requises pour gagner du temps ou réduire les efforts de révision.

* * *

À partir des archives orales du Corpus oral de la langue espagnole à Montréal, nous avons testé plusieurs méthodes de préparation des données destinées à la recherche linguistique : la transcription manuelle, la transcription automatique et la transcription vocale redite assistée d'un logiciel de reconnaissance automatique de la parole (RAP)⁹. Nous avons évalué, pour chacune de ces méthodes, la qualité des transcriptions initiales et le temps nécessaire à produire des transcriptions finales. Pour ce faire, l'exactitude et le niveau de détail ont été mis en relation avec l'effort : plus le document initial est exact, moins le temps nécessaire à la production d'un document final est important. Les transcriptions générées grâce à des technologies intelligentes et des stratégies vocales de redite sont plus exactes ; et celles réalisées avec un logiciel installé sur l'ordinateur, qui peut être entraîné, sont supérieures à celles obtenues avec un logiciel en ligne. Ainsi, la redite vocale assistée s'est avérée être la meilleure option pour convertir, en moins de temps que d'autres méthodes, des enregistrements contenant des données précieuses pour l'étude de la variation en matériel écrit.

9. Nous souhaitons saluer la célébration des 50 ans de corpus montréalais. Elle est arrivée à point nommé pour briser l'isolement pandémique et a été riche en rencontres de grande valeur. Plusieurs prises de parole ont alimenté toutes sortes de réflexions. L'une d'entre elles a résonné, et nous avons décidé de devenir les porte-étendard de la voie de la redite pour le travail de transcription des corpus linguistiques. En effet, Anne-José Villeneuve a été une pionnière dans le domaine de l'optimisation de la transcription vocale assistée par la technologie intelligente. Tout a commencé alors qu'elle avait obtenu un contrat de traduction pour une compagnie biomédicale et qu'on lui avait suggéré de faire de la traduction à vue avec un logiciel de RAP. Transcrire avec sa voix, c'était une façon de faire qu'elle trouvait plus rapide – ce sont ses mots – que la transcription manuelle de corpus avec la pédale qu'elle avait utilisée en tant qu'auxiliaire de recherche. Elle a adopté une démarche intuitive d'essais-erreurs, et n'a pas cessé de peaufiner de nombreuses pratiques. Elle connaît très bien, par exemple, l'importance de « nettoyer » son dictionnaire ou de dicter sa ponctuation, nous a-t-elle avoué. De plus, même si elle ralentit l'écoute de ses enregistrements de 50 % et s'impose de faire chaque redite au complet et d'une seule traite, il ne fait aucun doute, pour elle, que la redite est plus rapide, moins exigeante et plus fiable que la transcription manuelle.

L'optimisation de la transcription redite repose sur la collaboration étroite entre l'humain et la machine, fondée principalement sur le travail préparatoire réalisé en amont de la RAP, afin de produire des textes conformes au protocole établi. Ce processus exige l'identification et le traitement de la variabilité du signal sonore (qu'il s'agisse de changements de voix, de sexe ou d'origine géographique des locuteurs, ou encore de bruits environnants), ainsi que la variabilité propre à l'évolution dialectologique de l'espagnol (notamment les harmonies vocaliques, les assimilations, les épenthèses consonantiques et divers procédés d'abrègement). Il apparaît dès lors légitime de s'interroger sur la manière dont la linguistique entend contribuer à l'amélioration des résultats de la technologie intelligente (Alcántara Plá, 2008) et de suggérer, entre autres pistes, de concentrer les efforts sur la normalisation et la codification des phénomènes liés à l'oralité et à la variation.

L'approche méthodologique retenue pour rendre compte de faits linguistiques propres à l'espagnol en usage, ainsi que des transformations qu'il subit de facto, risque de devenir rapidement obsolète. Toutefois, l'analyse de l'oralité demeurera un enjeu central, dont l'optimisation repose tant sur la réduction de l'écart entre les corpus oraux et écrits que sur le recours à des outils garantissant une meilleure disponibilité des matériaux et, surtout, une interopérabilité renforcée des vastes ensembles de données orales qui, bien qu'existants, demeurent largement sous-exploités. Par ailleurs, la performance de ces technologies serait accrue si les matériaux d'entraînement faisaient l'objet d'un encodage plus rigoureux, conçu en étroite collaboration avec les chercheuses et chercheurs en linguistique.

Bien que l'évolution des technologies suscite l'espoir d'une automatisation totale du traitement des matériaux oraux pour leur analyse, cette perspective demeure inatteignable. Dans cette phase transitoire, nous soutenons une collaboration interactive et multimodale entre l'humain et la machine, une approche qui permettrait d'enrichir la dimension théorique et descriptive des données, tout en répondant à leurs exigences analytiques spécifiques. Manipulée par une personne experte de la langue, une interface plus souple et réactive, intégrant notamment le traitement de la variation linguistique, de la segmentation du discours, des pauses, des aspects prosodiques et des signaux non verbaux, faciliterait l'optimisation tant des processus que des résultats issus de cette collaboration humain-machine.

Références

- ADOLPHS, Svenja (2013), *Spoken corpus linguistics: from monomodal to multimodal*, Routledge.
- ALCÁNTARA PLÁ, Manuel (2008), El análisis lingüístico en la transcripción automática de la lengua hablada, el Proyecto COAST, dans Antonio MORENO SANDOVAL (dir.) *Actas del VIII Congreso de Lingüística General*, Universidad Autónoma de Madrid, 184-197, elvira.llf.uam.es/clg8/actas/.
- BAARING, Inge (2006), Respeaking-based online subtitling in Denmark, *Intralinea*, 8, intralinea.org.
- BOUZOUTA, Miriam *et al.* (2018), Dialectos del español. Una nueva aplicación para conocer la variación actual y el cambio en las variedades del español, *Dialectologia*, 20, 61-83.
- BROUSSEAU, Julie *et al.* (2003), Automated closed-captioning of live TV broadcast news in French, 8th *European Conference on Speech Communication and Technology (Eurospeech 2003)*, Genève, 1245-1248, doi: 10.21437/Eurospeech.2003-398.
- DEL VALLE LÓPEZ, Ángela (2004), La taquigrafía en España: origen y evolución, *Revista de Educación*, 333, 441-461.
- DELISLE, Jean et Alain Otis (2016), *Les douaniers des langues: grandeur et misère de la traduction à Ottawa, 1687-1967*, Presses de l'Université Laval.
- DÉSILETS, Alain *et al.* (2008), Evaluating productivity gains of hybrid ASR-MT systems for translation dictation, *Proceedings of the 5th International Workshop on Spoken Language Translation (IWSLT)*: Waikiki, 158-165, aclanthology.org.
- DRAGSTED, Barbara *et al.* (2009), «Experts exposed», dans Inger Margrethe MEES *et al.* (dir.), *Methodology, Technology and Innovation in Translation Process Research*, Samfundslitteratur Press, 38, 293-317.
- DU BOIS, John W. (1991), Transcription design principles for spoken discourse research, *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 1(1), 71-106.
- ELVIRA-GARCÍA, Wendy *et al.* (2015), Una herramienta para la transcripción prosódica automática con etiquetas Sp_ToBI en Praat, dans Adrián CABEDO NEBOT (dir.), *Perspectivas actuales en el análisis fónico del habla: tradición y avances en la fonética experimental*, Universitat de València, 455-464.
- FERNÁNDEZ VÍTORES, David (2021), *El español: una lengua viva*, Dirección Académica del Instituto Cervantes, cvc.cervantes.es.
- FOGG, Terry et Colin W. WIGHTMAN (2000), Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology, *Creating Knowledge in the 21st Century: Insights From Multiple Perspectives (American Educational Research Association Annual Meeting)*, New Orleans.
- FRANCIS, W Nelson et Henry KUCERA (1979 [1964]), *Brown corpus manual*, korpus.uib.no/icame/brown/bcm.html.
- GUBERN, Román (2010), *Metamorfosis de la lectura*, Anagrama.
- HAMMERSLEY, Martyn (2010), Reproducing or constructing? Some questions about transcription in social research, *Qualitative Research*, 10(5), 553-569.

- IVES, Edward D. (1995), *The tape-recorded interview: A manual for fieldworkers in folklore and oral history* (2^e éd.), University of Tennessee Press.
- JIMÉNEZ IVARS, Amparo et Amparo HURTADO ALBIR (2003), Variedades de traducción a la vista. Definición y clasificación, *TRANS. Revista de Traductología*, 7, 47-57.
- JONES, Rodney H. (2011), Data collection and transcription in discourse analysis, dans K. HYLAND et B. PALTRIDGE (dir.), *Continuum companion to discourse analysis*, Continuum International Publishing Group, 9-21.
- KNIGHT, Dawn et Svenja ADOLPHS (2010), Building a spoken corpus. What are the basics?, dans A. O'KEEFE et M. MCCARTHY (dir.), *The Routledge Handbook of Corpus Linguistics*, Routledge, 38-52.
- LAMBOURNE, Andrew D. (2006), Subtitle respeaking. A new skill for a new age, *Intralinea*, 8, intralinea.org.
- LAPADAT, Judith C., et Anne C. LINDSAY (1999), Transcription in research and practice: From standardization of technique to interpretive positionings, *Qualitative inquiry*, 5(1), p. 64-86.
- LEECH, Geoffrey (1991), The state of the art in corpus linguistics, dans K. AIJMER et B. ALTENBERG (dir.), *English corpus linguistics: studies in honour of Jan Svartvik*, Longman Group Limited, 20-41.
- LLISTERRI, Joaquim (2009), Las tecnologías del habla en las lenguas románicas ibéricas, *Studies in Hispanic and Lusophone Linguistics*, 2(1), 133-180.
- MACLEAN, Lynne M. et al. (2004), Improving accuracy of transcripts in qualitative research, *Qualitative health research*, 14(1), 113-123, doi.org/10.1177/104973230325980.
- MARKLE, Thomas D. et al. (2011), Beyond transcription: technology, change, and refinement of method», *Forum Qualitative Sozialforschung*, 12(3), doi.org/10.17169/fqs-12.3.15-64.
- MATHESON, Jennifer L. (2007), The voice transcription technique: Use of voice recognition software to transcribe digital interview data in qualitative research, *The Qualitative Report*, 12(4), 547-560.
- MONDADA, Lorenza (2008), Contributions de la linguistique interactionnelle, *CML 2008: Congrès mondial de Linguistique française*, Paris, doi.org/10.1051/cmlfo8348.
- NAGY, Naomi et Devyani SHARMA (2014), Transcription, dans R. J. PODESVA et D. SHARMA (dir.), *Research Methods in Linguistics*, Cambridge University Press, 235-256.
- PARK, Julie et Angelica ECHO ZEANAH (2005), An evaluation of voice recognition software for use in interview-based research: A research note, *Qualitative Research*, 5(2), 245-251.
- PARODI, Giovanni (dir.) (2007), *Lingüística de corpus y discursos especializados: puntos de mira*, Ediciones Universitarias de Valparaíso.
- PATO, Enrique (2022), El español como lengua comunitaria y lengua de herencia en Quebec, *Lengua y migración*, 14(1), 7-26.
- PATO, Enrique (dir.) (2006), El español en Montreal: COLEM, esp-montreal.jimdo.com.

- PIMSLEUR, Paul *et al.* (1977), Vitesse d'élocution dans les émissions radiophoniques. Étude comparative des présentateurs français et américains, *Études de linguistique appliquée*, 27, 12-18.
- POLAND, Blake D. (2002), Transcription quality», dans J. GUBRIUM et J. HOLSTEIN (dir.), *Handbook of interview research: context and method*, Sage Publications, 629-649.
- POLAND, Blake D. (1995), Transcription quality as an aspect of rigor in qualitative research, *Qualitative inquiry*, 1(3), 290-310.
- POYATOS, Fernando (2002), *Nonverbal Communication Across Disciplines*, vol. 2: *Paralanguage, Kinesics, Silence, Personal and Environmental Interaction*, John Benjamins Publishing.
- ROBERTS POWERS, Willow (2005), *Transcription techniques for the spoken word*, Rowman Altamira Press.
- RODERO ANTÓN, Emma et Gloria CAMPOS PARRA (2005), Las voces de los presentadores de informativos en televisión, *Comunicar*, 25(2), revistacomunicar.com.
- ROJO, Guillermo (2021), *Introducción a la lingüística de corpus en español*, Routledge.
- ROMERO-FRESCO, Pablo (2011), *Subtitling through speech recognition: respeaking*, St. Jerome Publishing.
- RUFINO MORALES, Marimar (2020), Estudio comparativo de métodos de transcripción para corpus orales: el caso del español, *Revista Nebrija de Lingüística aplicada a la enseñanza de lenguas*, 14(29), 126-146, doi.org/10.26378/rnlael1429406.
- SANKOFF, David *et al.* (1976), Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale, *Cahier de linguistique*, 6, 85-125, erudit.org/.
- TATHAM, Mark et Katherine MORTON (2005), *Developments in speech synthesis*, John Wiley et Sons.
- TILLEY, A. Susan et Kelly D. POWICK (2002), Distanced data: Transcribing other people's research tapes, *Canadian Journal of Education/Revue canadienne de l'éducation*, 27(2/3), 291-310.
- VIDAL, Enrique *et al.* (2006), Computer-assisted translation using speech recognition, *IEEE Transactions on Audio, Speech, and Language Processing*, 14(3), 941-951, doi.org/10.1109/TSA.2005.857-788.
- WILLEMS, Dominique (1989), Lenguaje escrito y lenguaje oral, *Historia y fuente oral*, 1, 97-105, jstor.org/stable/27753231.

Affiliations

Julie Abbou, Università di Torino
Laurence Arrighi, Université de Moncton
Hélène Blondeau, University of Florida
Claire Bourély, Université de Montréal
Heather Burnett, Université Paris Cité
Anne-Sophie Calinon, Université de Franche-Comté
Marie-Hélène Côté, Université de Lausanne
Léa Courdès-Murphy, Université de Toulouse 2
Anne Dister, UCLouvain – Saint-Louis Bruxelles
Mireille Elchacar, Université TÉLUQ
Julien Eychenne, Université de Sherbrooke
Nancy Gagné, Université TÉLUQ
Annette Gerstenberg, Universität Potsdam
Emmanuelle Labeau, Universiteit Antwerpen
Marta Lupica Spagnolo, Universität Potsdam
Angéline Martel, Université TÉLUQ
Raymond Mougeon, Glendon College – York University
Katja Ploog, Université d'Orléans
Wim Remysen, Université de Sherbrooke
Basile Roussel, Université de Moncton
Marimar Rufino Morales, Université de Montréal
Hugo Saint-Amant Lamy, Université du Québec à Rimouski
Fabio Scetti, Université du Québec à Trois-Rivières, CRIEM de McGill
Mireille Tremblay, Université de Montréal
Isabelle Violette, Université de Moncton

Table des matières

INTRODUCTION

Un anniversaire remarquable	7
<i>Hélène Blondeau et Wim Remysen</i>	

Changement, diachronie et variation générationnelle

CHAPITRE 1

Cinquante ans de liaison à Montréal éclairée par le reste du monde	21
<i>Marie-Hélène Côté</i>	

CHAPITRE 2

Convergence ou divergence en français laurentien	43
Quarante ans de changement à Montréal et à Welland	
<i>Hélène Blondeau, Claire Bourély, Raymond Mougeon et Mireille Tremblay</i>	

CHAPITRE 3

Corpus oraux et corpus de textos	63
La référence temporelle au futur en Belgique et au Québec	
<i>Mireille Tremblay, Hélène Blondeau, Anne Dister et Emmanuelle Labeau</i>	

CHAPITRE 4

LangAge à l'oral et à l'écrit, au cours de la vie	83
Madame Moreau et les péripéties linguistiques de l'âge avancé	
<i>Annette Gerstenberg et Marta Lupica Spagnolo</i>	

Récits de vie, communauté et appartenance sociale

CHAPITRE 5

Le pouvoir des mots	105
La toponymie et la gentrification dans le quartier Hochelaga-Maisonneuve entre 2010 et 2012	
<i>Claire Bourély</i>	

CHAPITRE 6	
Le corpus « Traits d'union »	123
Perceptions de la langue française dans le travail d'intervenants auprès de citoyens issus de l'immigration	
<i>Mireille Elchacar, Angéline Martel et Nancy Gagné</i>	
CHAPITRE 7	
La vie montréalaise racontée au sein de la Petite Italie et du Petit Portugal	145
<i>Fabio Scetti</i>	
CHAPITRE 8	
Approche de la mobilité discursive dans un corpus montréalais	163
Autour des expressions figées	
<i>Anne-Sophie Calinon et Katja Ploog</i>	
CHAPITRE 9	
Devenir féministe et queer à Paris et à Montréal	187
Récits de vie dans le corpus CaFé	
<i>Julie Abbou et Heather Burnett</i>	
Instrumentation des corpus	
CHAPITRE 10	
Les corpus oraux montréalais à l'ère numérique	209
Des bobines magnétiques à la diffusion sur la plateforme du Fonds de données linguistiques du Québec	
<i>Wim Remysen et Hugo Saint-Amant Lamy</i>	
CHAPITRE 11	
Moncton comme pôle urbain francophone	235
Quels corpus pour décrire le français en usage ici et maintenant ?	
<i>Laurence Arrighi, Basile Roussel et Isabelle Violette</i>	
CHAPITRE 12	
Annotation et analyse de données sociophonologiques sur grands corpus	255
Présentation de la plateforme Phonometrica	
<i>Julien Eychenne et Léa Courdès-Murphy</i>	
CHAPITRE 13	
Transcrire des corpus linguistiques à l'aide de la technologie intelligente	271
L'exemple du COLEM	
<i>Marimar Rufino Morales</i>	
Affiliations	291

