PUM

(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

Les Presses de l'Université de Montréal

Table des matières

INTRODUCTION
Un anniversaire remarquable 7
Hélène Blondeau et Wim Remysen

Changement, diachronie et variation générationnelle

CHAPITRE 1
Cinquante ans de liaison à Montréal éclairée par le reste du monde 21
Marie-Hélène Côté

CHAPITRE 2
Convergence ou divergence en français laurentien 43
Quarante ans de changement à Montréal et à Welland
Hélène Blondeau, Claire Bourély, Raymond Mougeon et Mireille Tremblay

CHAPITRE 3
Corpus oraux et corpus de textos 63
La référence temporelle au futur en Belgique et au Québec
Mireille Tremblay, Hélène Blondeau, Anne Dister et Emmanuelle Labeau

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
Annette Gerstenberg et Marta Lupica Spagnolo

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
Claire Bourély

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
Mireille Elchacar, Angéline Martel et Nancy Gagné

CHAPITRE 7
La vie montréalaise racontée au sein de la Petite Italie et du Petit Portugal 145
Fabio Scetti

CHAPITRE 8
Approche de la mobilité discursive dans un corpus montréalais 163
Autour des expressions figées
Anne-Sophie Calinon et Katja Ploog

CHAPITRE 9
Devenir féministe et queer à Paris et à Montréal 187
Récits de vie dans le corpus CaFé
Julie Abbou et Heather Burnett

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
Wim Remysen et Hugo Saint-Amant Lamy

CHAPITRE 11
Moncton comme pôle urbain francophone 235
Quels corpus pour décrire le français en usage ici et maintenant ?
Laurence Arrighi, Basile Roussel et Isabelle Violette

CHAPITRE 12
Annotation et analyse de données sociophonologiques sur grands corpus 255
Présentation de la plateforme Phonometrica
Julien Eychenne et Léa Courdès-Murphy

CHAPITRE 13
Transcrire des corpus linguistiques à l'aide de la technologie intelligente 271
L'exemple du COLEM
Marimar Rufino Morales

Affiliations 291

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

INTRODUCTION

Un anniversaire remarquable

Hélène Blondeau et Wim Remysen

L'année 2021 marquait le 50e 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 diachronique[1]. 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 2022[2] 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 poursuivent[3]. 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 Mougeon 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'accentue 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 : Paméla 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 corpus[4].

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 Thibault (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 variabilty », 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 THIBAULT et Hélène BÉRUBÉ (1978), « Semantic field variability », dans David Sankoff (dir.), *Linguistic variation : models and methods*, New York, Academic Press, 23-43.

THIBAULT, Pierrette (2001), « Regard rétrospectif sur la sociolinguistique québécoise et canadienne », *Revue québécoise de linguistique*, *30*(1), 19-42.

THIBAULT, 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.

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

ISBN 978-2-7606-5262-0

Disponible en libre accès
www.pum.umontreal.ca