



HAL
open science

Réflexions épistémologiques sur de nouveaux apports méthodologiques et empiriques à l'étude géolinguistique des français d'Amérique

Mathieu Avanzi, André Thibault

► To cite this version:

Mathieu Avanzi, André Thibault. Réflexions épistémologiques sur de nouveaux apports méthodologiques et empiriques à l'étude géolinguistique des français d'Amérique. 6e Congrès Mondial de Linguistique Française, 2018, Mons, Belgique. <10.1051/shsconf/20184602001>. <hal-02531915>

HAL Id: hal-02531915

<https://hal.sorbonne-universite.fr/hal-02531915v1>

Submitted on 4 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Réflexions épistémologiques sur de nouveaux apports méthodologiques et empiriques à l'étude géolinguistique des français d'Amérique

Mathieu Avanzi¹ et André Thibault²

¹ FNRS & Université catholique de Louvain, Belgique – mathieu.avanzi@uclouvain.be

² Université de Paris-Sorbonne, France – andre.thibault@sorbonne-universite.fr

Résumé. Après un bref tour d'horizon historique de la géolinguistique du français au Canada, nous présenterons dans le détail une nouvelle technique d'enquêtes en ligne (inspirée du « crowdsourcing » ou « production participative ») permettant de porter un regard neuf sur la différenciation diatopique du français au Canada, envisagée du point de vue des auto-représentations. Nous enchaînerons avec une réflexion épistémologique portant sur les limites inhérentes à ce type d'approche (données élicitées vs. spontanées ; données de production et d'autoreprésentation ; paradoxe de l'observateur ; variation diasituationnelle ; problèmes de syntagmatique ; distinction entre signification et désignation), puis présenterons brièvement le traitement cartographique effectué, pour aboutir au cœur de l'exposé : une série de cartes et de graphes élaborés à partir de nos enquêtes et illustrant différentes distributions aréales prototypiques (aires complémentaires, types locaux, aires périphériques et effet d'âge).

Abstract. Epistemological considerations on a new methodological and empirical contribution to the geolinguistic study of North-American French. After a brief historical overview of Canadian French geolinguistic studies, we will offer a detailed presentation of a new on-line survey technique (crowdsourcing) that allows us to shed new light on dialectal differentiation in Canadian French, considered from the point of view of the speakers' self-representations. We will be following this up with an epistemological questioning focused on the limits of this approach (elicited vs. spontaneous data ; production vs. auto-representation data ; observer's paradox ; diasituational variation ; syntagmatic issues ; distinction between signification and designation), and then go on with a presentation of the cartographic treatment of the data. The core of the talk will be dedicated to a series of maps and graphs, elaborated on the basis of our on-line inquiries, that exemplify various prototypical areal distributions (complementary areas, local types, peripheral areas and age effect).

1 Introduction

Cette contribution a pour but de présenter un nouvel apport méthodologique à la géolinguistique francophone, basé sur des enquêtes en ligne destinées à sonder les locuteurs sur leurs autoreprésentations, qu'il s'agisse de variables phonologiques, morphosyntaxiques ou lexicales.

Après une application à la francophonie d'Europe (v. *Avanzi et al.* 2016), de nouvelles enquêtes ont été étendues à la francophonie nord-américaine, plus précisément aux provinces de l'est canadien. C'est de ce terrain qu'il sera question ici, raison pour laquelle nous allons d'abord faire un bref tour d'horizon de l'histoire de la géolinguistique et de l'atlantographie des français canadiens.

1.1 Champ disciplinaire : géolinguistique et atlantographie

Notre contribution s'inscrit dans le champ de la géolinguistique, branche de la dialectologie qui consiste en la représentation sous forme de cartes des aires d'extension de différentes variantes au sein de conglomérats dialectaux plus ou moins différenciés. Lorsque de nombreuses cartes sont réunies en un seul ouvrage, on parle d'atlas géolinguistiques (et le plus souvent aussi ethnographiques).

Dans le monde galloroman, les plus célèbres atlas géolinguistiques sont l'ALF, les atlas régionaux de deuxième génération (comme l'ALLY) et l'ALW (le plus interprétatif de tous). Il s'agit toutefois dans tous les cas d'ouvrages consacrés à la variation diatopique des dialectes primaires du galloroman et non à celle des régiolectes du français, dialectes « secondaires » dans la terminologie de Coseriu (1980). À ce jour, le seul atlas consacré à la variation géographique du français dit « régional » en Europe est celui d'Avanzi (2017).

1.2 Au Canada français

Au Canada français, qui ne connaît pas de dialectes primaires (les français canadiens sont par définition des dialectes secondaires, c'est-à-dire des variétés d'exportation du français), une longue tradition d'études géolinguistiques existe. Force est de constater toutefois que les données géolinguistiques n'y sont que très rarement présentées sous forme de cartes.

1.2.1 Recueils de données géolinguistiques

En ce qui concerne les français d'Acadie, l'ouvrage dialectologique fondamental reste *Les parlers français d'Acadie* de G. Massignon (1962), qui réunit des matériaux récoltés dans plus d'une quarantaine de points d'enquête et correspondant à près de deux mille concepts. L'ouvrage est d'une grande richesse, mais ne comporte aucune carte géolinguistique.

Le principal « atlas » des français canadiens est surtout connu sous le nom d'*ALEC* (1980), « Atlas linguistique de l'Est du Canada », mais il ne s'agit là que de son sous-titre. L'ouvrage s'appelle en fait *Le Parler populaire du Québec et de ses régions voisines* et ne comporte que très peu de cartes : celle des points d'enquête (vol. 1, p. 5, reprise à l'identique en p. 20) et celle de l'isoglosse du R (vol. 1, p. 31). Il s'agit d'un énorme ensemble de données (10 vol.) essentiellement lexicales (plus de 2300 concepts) recueillies dans 172 points d'enquête auprès de plusieurs centaines de témoins.

Lavoie *et al.* (1985) reprend la méthodologie de l'*ALEC* mais l'ouvrage est consacré à un domaine plus limité et présente un réseau d'enquêtes plus serré. Le mot « atlas » ne figure pas dans le titre, *Les parlers français de Charlevoix, du Saguenay, du Lac-Saint-Jean et de la Côte-Nord*. Les riches matériaux rassemblés dans les cinq volumes ne font l'objet d'aucune ébauche de cartographie.

1.2.2 Atlas

Il n'existe à ce jour qu'un seul ouvrage méritant pleinement le titre d'« atlas » en géolinguistique canadienne-française : il s'agit de Péronnet *et al.* (1998), *Atlas linguistique du vocabulaire maritime acadien*. L'ouvrage présente, sous forme de listes mais aussi de cartes, les réponses de 54 informateurs – répartis dans 18 localités – à 354 questions portant sur le lexique de la mer. Il propose en outre une analyse dialectométrique du vocabulaire.

Il s'agit d'une contribution importante, mais limitée aux côtes acadiennes et à un champ lexical très précis.

1.2.3 Monographies, articles

Il ne faudrait pas croire pour autant que ces recueils de données géolinguistiques sont restés inexploités. Un certain nombre d'auteurs se sont consacrés à l'analyse de ces matériaux, en particulier ceux tirés de l'ALEC. On relève des contributions sur la phonétique-phonologie ainsi que des articles portant sur le lexique.

Larochelle (1989) s'est livré à une « Analyse géolinguistique de quelques données phonétiques de l'*Atlas linguistique de l'est du Canada* ». C'est la première source qui propose des cartes interprétatives élaborées informatiquement à partir des données de l'ALEC. L'étude portait sur les voyelles fermées, l'affrication des dentales, l'aspiration des chuintantes, l'existence du /h/ aspiré et la chute de /t/, /r/ et /l/ comme second élément de groupe consonantique final. Friesner (2010) porte sur l'affrication des dentales et le relâchement des voyelles fermées en position entravée, à partir de données également tirées de l'ALEC.

Bergeron (1995) est une thèse consacrée à l'ALEC, rédigée par l'un de ses deux auteurs. Le travail propose une centaine de cartes présentant l'aréologie de certaines variables lexicales et identifie quatre aires fondamentales en laurentien. Verreault / Lavoie (1996), Laflamme (1996) et Verreault / Lavoie (2004) sont trois contributions basées sur les matériaux de l'ALEC ainsi que, dans une moindre mesure, sur ceux de l'enquête de la Société du parler français au Canada (début du 20^e siècle), de Massignon (1962) et de Lavoie *et al.* (1985).

1.2.4 Enquêtes PFC-Québec

On trouvera dans Côté (2014) une présentation d'ensemble du volet laurentien du projet PFC (*Phonologie du français contemporain*). Quelque 400 témoins répartis dans environ 35 localités ont effectué les tâches du protocole PFC (lecture de mots, lecture de texte, enquêtes semi-dirigées et libres). Une liste de mots complémentaires (v. Côté 2012 : 273-274), permettant de tester des particularités diatopiques et/ou diagénérationnelles laurentiennes, a également été soumise aux témoins.

Dans Côté / Saint-Amant Lamy / Milne (2016), les auteurs présentent quelques cartes consacrées à des shibboleths phonétiques (*baleine*, *connaissent*, *arrête* et *nage*), établies à partir de leurs études de production, et proposent un « clustering » des différents points d'enquête sur la base des données élicitées. Côté (2016a) et Côté (2016b) sont des articles consacrés respectivement à un locuteur de Montréal et de Trois-Rivières.

2 Présentation de la méthodologie d'enquête

2.1 La méthode des enquêtes en ligne

La méthode du *crowdsourcing*ⁱ consiste à faire appel au grand public en vue de réaliser certaines tâches. Si la méthode ne date pas d'aujourd'hui, c'est seulement au début des années 2000 qu'elle a connu un essor important, avec la démocratisation du web. Dans le domaine des sciences du langage, la méthode a été exploitée pour documenter l'aire d'extension et la vitalité de particularismes linguistiques locaux en anglais (Vaux / Golder 2003), en allemand (*Atlas zur deutschen Alltagssprache*), en italien (Tosques / Castellarin) ou en français européen (Avanzi *et al.* 2016, Avanzi 2017). Devant les potentialités que présente cette méthode (elle permet de récolter de nombreux témoignages individuels à des échelles géographiques variables sur une période de temps restreinte et sans moyens techniques particuliers), nous avons souhaité la mettre à profit pour l'étude des régionalismes du français parlé en Amérique du Nord.

2.2 L'enquête nord-américaine

En pratique, nous avons mis en place deux enquêtes (nov. 2016 – août 2017 et sept. 2017 – nov. 2017). Chacune d'entre elles proposait une quarantaine de questions. Les participants étaient invités à prendre part à une enquête en ligne, mise en place sur la plateforme Qualtricsⁱⁱ. On les informait qu'ils allaient participer à une enquête sur les usages actuels du français, dont le but était d'évaluer la vitalité de certaines tournures régionales, ainsi que leur extension géographique. On leur demandait de répondre de façon anonyme aux questions, de la façon la plus naturelle possible. Ils étaient avertis que leur tâche n'était pas de dire quelle(s) variante(s) est/sont correcte(s) du point de vue du français standard ou des dictionnaires. Après avoir fourni quelques informations relatives à leur profil sociologique (pays et localité où ils ont passé la plus grande partie de leur jeunesse, déménagements, âge, sexe, métier et plus haut diplôme obtenu), les participants voyaient apparaître sur leur écran une image accompagnée d'une définition, d'un contexte d'emploi ou d'une phrase à trous. Ils devaient ensuite choisir, dans une liste fermée, la réponse correspondant à la forme qu'ils utiliseraient de préférence pour prononcer un mot, désigner un objet ou se référer à une situation. Ils pouvaient cocher une ou plusieurs réponses, et/ou cocher une case « autre » pour citer un terme qui ne figurait pas dans la liste.

2.3 Participants

Le lien de l'enquête a été diffusé auprès de collègues, d'amis et de parents, via les réseaux sociaux. Nous avons pris le parti de clôturer les enquêtes une fois que le nombre d'au moins 4 000 participants avait été atteint. Comme on peut voir dans le tableau ci-dessous, le nombre de participants par province est similaire d'une enquête à l'autre. La province de Québec est celle qui cumule le plus grand nombre de participants, suivie par l'Ontario, le Nouveau-Brunswick et le Manitoba. Pour la Nouvelle-Écosse, les chiffres sont beaucoup plus bas :

Province	Enquête n°1		Enquête n°2	
	Nb. part.	% du total	Nb. part.	% du total
Québec	3707	86,5	3690	84,7
Ontario	330	7,7	216	5,0
Nouveau-Brunswick	148	3,5	256	5,9
Manitoba	89	2,1	191	4,4
Nouvelle-Écosse	10	0,2	5	0,1
TOTAL	4284	100	4358	100

Table. 1. Nombre de participants par province et par enquête.

Il est intéressant de constater que, malgré le format du sondage (en ligne, diffusé sur les réseaux sociaux), le panel de participants est relativement varié du point de vue de l'âge, comme on peut le voir sur les figures 1 et 2 :

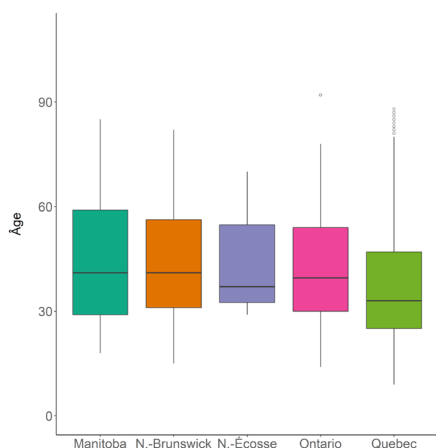


Fig. 1. Âge des participants en fonction de leur province de jeunesse [enquête n°1]

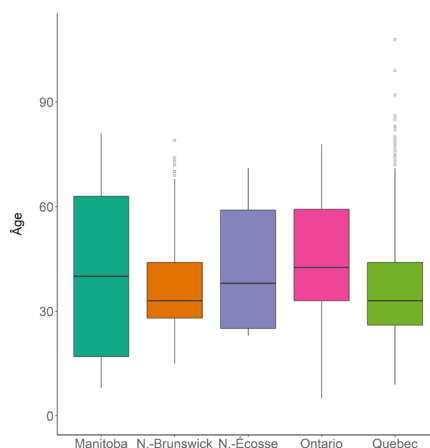


Fig. 2. Âge des participants en fonction de leur province de jeunesse [enquête n°2]

Quant au sexe des participants, il ressort sur les figures 3 et 4 que les femmes ont été à peu près deux fois plus nombreuses à participer que les hommes, puisqu'en moyenne 70,1% des participants étaient de sexe féminin (il s'agit d'une constante, le même phénomène ayant été observé dans les enquêtes sur le français d'Europe) :

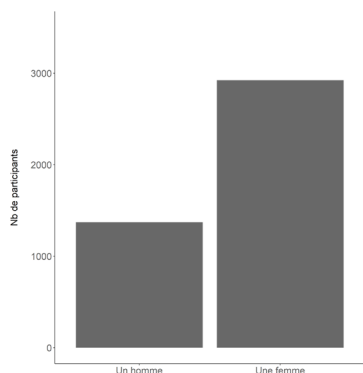


Fig. 3. Âge des participants en fonction de leur sexe [enquête n°1]

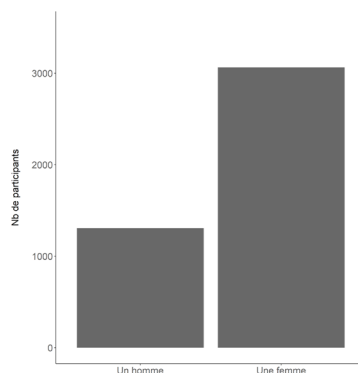


Fig. 4. Âge des participants en fonction de leur sexe [enquête n°2]

3 Questions épistémologiques

La nature même de cette nouvelle démarche doit être soumise à questionnement car, comme tout protocole expérimental, elle porte en elle ses points forts mais également ses limites et ses lacunes. Il convient de la comparer à d'autres façons de recueillir des données, de la caractériser par rapport à ces dernières et de préciser quelles en sont les limites.

3.1 Données spontanées vs. données élicitées

Les données élicitées n'ont pas toujours bonne presse. On soupçonne à bon droit les témoins de ne pas s'exprimer de la même façon que s'ils produisaient des données spontanément, dans un échange non contraint et surtout non enregistré. Cela dit, la spontanéité absolue dans le cadre d'enquêtes sociolinguistiques est un objectif difficile à atteindre. Les enquêtes variationnistes classiques mettent les enquêtés dans une situation jusqu'à un certain point artificielle (et aux allures d'interrogatoire), et même les enquêtes dites « écologiques » ne peuvent faire l'économie d'un micro, voire parfois d'une caméra, sans compter tout le protocole de signature de formulaires de consentement qui confère à toute la démarche une connotation juridico-policière du plus mauvais effet.

Les différentes démarches visant à recueillir des données linguistiques de première main doivent en fait tout simplement s'adapter à l'objet qu'elles entendent décrire. Lorsqu'il s'agit d'étudier des variables dont la fréquence relative dans le discours est extrêmement élevée (simplification de groupes consonantiques en fin de mot, réalisation du *ne* de négation, alternance entre futur analytique et synthétique), un corpus recueilli avec un minimum de contraintes et d'exigences envers les témoins fait parfaitement l'affaire. Toutefois, s'il s'agit de savoir comment 5 000 francophones canadiens prononcent la voyelle tonique du mot « baleine », un corpus d'entretiens libres est tout à fait inopérant. Le mot n'est tout simplement pas susceptible d'apparaître par hasard, systématiquement et à répétition, dans des milliers d'entretiens menés partout à travers le pays (à supposer que l'on ait les moyens de faire transcrire des milliers d'heures d'entrevues, ce qui ne s'est jamais vu).

Les exigences de comparabilité qui caractérisent le travail des géolinguistes déterminent le protocole des enquêtes dialectologiques qui, depuis le 19^e siècle déjà, proposent des listes de questions uniformisées à de nombreux témoins répartis de façon plus ou moins uniforme sur un espace dialectal donné. C'est cette démarche qui a donné naissance à tous les atlas géolinguistiques du monde roman. Malgré ses défauts (questions mal comprises, influence de la langue de l'enquêteur sur celle de l'enquêté, nivellement de la variation diasituationnelle, mauvaise représentation de la synonymie, etc.), l'élicitation directe de données à

grande échelle est la condition *sine qua non* d'une représentation cartographique de la variation dialectale.

3.2 Données de production vs. données d'autoreprésentation

Parmi les données élicitées, on peut distinguer des données de « production » et des données que nous appellerions « d'autoreprésentation ». Dans le domaine du lexique, cela revient à peu près au même : on demande au témoin quel mot il utilise pour désigner un certain référent et sa réponse est consignée sur une fiche papier par l'enquêteur (à l'ancienne) ou dans un formulaire électronique en ligne (de nos jours). La réponse « produite » correspond en même temps à la « représentation » que le locuteur se fait (ou veut bien donner) de son propre usage.

En ce qui concerne la phonétique, il convient de distinguer deux cas de figure. L'enquêteur peut demander au témoin de produire du matériau phonique, pour ensuite le transcrire. On dira alors qu'il cartographie des données de *production* (cette même production pouvant varier selon qu'il s'agisse de lire une liste de mots, un texte, ou de répondre à des questions ciblées). Il est toutefois aussi possible de demander directement au témoin s'il prononce un mot de telle ou telle façon (v. section 2 ci-dessus). Une cartographie basée sur des matériaux élicités de cette manière repose sur des données de *représentation* (il ne s'agit pas de l'usage constaté, mais bien de l'usage *déclaré*). Dans une autre terminologie, on parlera d'usage *perçu*, ce qui nous amène du côté de la dialectologie perceptive, à cette différence près qu'ici on ne demande pas aux témoins ce qu'ils pensent de la langue d'autrui mais bien seulement de la leur. On peut raisonnablement considérer que les locuteurs connaissent mieux leur propre usage que celui de locuteurs vivant à l'autre bout du pays ; c'est pourquoi nous tenons à préciser qu'il s'agit de données d'*autoreprésentation*.

3.3 Le paradoxe de l'observateur

Toute démarche visant à recueillir des matériaux linguistiques de première main pose la question du paradoxe de l'observateur, que Labov (1970) avait résumée de façon concise en disant qu'il s'agit d'observer les locuteurs pour savoir comment ils parlent... lorsqu'ils ne sont pas observés. La situation d'enquête elle-même, qu'elle soit semi-dirigée, libre, voire « écologique », peut toujours bien sûr entraîner un biais dans la nature des données recueillies. De nombreuses stratégies ont tenté d'éliminer ou à tout le moins d'adoucir les effets de cette situation (choix d'enquêteurs issus du milieu, conversations libres entre témoins auxquelles l'enquêteur ne participe pas, etc.).

Dans le cas qui nous occupe, il s'agit surtout de contraster l'enquête en ligne, totalement anonyme et sans confrontation « face à face » entre enquêtés et enquêteurs, avec les enquêtes variationnistes traditionnelles où linguistes et témoins sont physiquement co-présents. Un témoin amené à prendre la parole devant un universitaire qui lui a explicitement signalé qu'il l'enregistrait pour étudier sa façon de parler peut vouloir donner une bonne image de lui-même en adoptant des variantes socialement plus valorisées – ou, bien au contraire, faire le bravache et choisir de s'exprimer systématiquement dans les registres les plus stigmatisés. Il nous semble toutefois qu'un internaute derrière son écran n'a pas grand-chose à gagner à donner une meilleure image de lui-même à l'enquêteur, puisque le processus est entièrement anonyme et dématérialisé. Bien sûr, cela n'empêche pas qu'un répondant puisse se mentir à lui-même ou avoir une perception erronée de son propre usage, mais l'anonymat absolu de la démarche favorise certainement un certain détachement face à l'enquêteur. En outre, on ne comprend vraiment pas pourquoi un enquêté devrait donner une réponse en désaccord avec son propre usage lorsqu'aucune des variantes proposées ne fait l'objet d'une stigmatisation sociale (nous reviendrons à ce point ci-dessous sous 3.4.1).

3.4 Données d'autoreprésentation : leurs limites

Toutes les variables ne se prêtent pas aussi bien à ce type d'enquête en ligne, qui ne sont évidemment pas une panacée. Comme nous allons le voir, un certain nombre de problèmes se posent, dont le concepteur d'une telle enquête doit tenir compte pour, en amont, bien élaborer son questionnaire et, en aval, jauger la portée de ses résultats et les croiser avec d'autres types de sources afin de mieux évaluer la valeur documentaire.

3.4.1 La variation diasituationnelle

Parmi les axes de la variation linguistique, certains facteurs dépendent du locuteur (qui a un certain âge, vit dans une certaine localité et appartient à un certain milieu) mais d'autres dépendent de la situation, de l'intention pragmatique, de la stylistique, etc. En d'autres mots, la même personne ne s'exprime pas toujours de la même manière. Cela exerce évidemment une influence sur le choix des questions à poser.

Par exemple, s'il semble pertinent de demander à des Québécois comment ils prononcent « arrête » (car la longueur de la voyelle tonique dans ce mot ne dépend en rien de facteurs situationnels), cela n'aurait en revanche aucun sens de leur demander s'ils prononcent *moi* [mwa] ou [mwe], car l'on sait pertinemment qu'il est possible de dire les deux mais pas dans les mêmes circonstances, ni surtout avec la même intention pragmatique. Il faut donc avoir un minimum d'intuition sur sa langue pour poser des questions appropriées : toutes les variables ne se prêtent pas aussi bien à ce genre d'enquête. De façon générale, toute variante faisant l'objet d'une stigmatisation sociale n'a pas vocation à figurer dans un questionnaire de ce type, les répondants étant susceptibles de ne pas admettre (et de ne pas s'avouer à eux-mêmes) qu'ils l'emploient couramment.

Parmi les commentaires récoltés, on relève aussi des témoignages de locuteurs disant opter pour une certaine variante lorsqu'ils retournent dans la région de leur enfance alors qu'ils l'avaient abandonnée entre temps au profit de la variante de leur nouvel environnement. C'est ce qui fait que certains choix lexicaux affichent des aires très larges et qu'il y a beaucoup de chevauchement entre les aires. Nous revendiquons cette situation. Les enquêtes dialectologiques traditionnelles mettaient un point d'honneur à ne sélectionner que des témoins « purs », âgés, ruraux, ayant été le moins possible exposés au système scolaire et totalement dépourvus de mobilité. La réalité langagière est bien évidemment très éloignée du portrait que l'on obtient à partir de tels sujets – et ce, en particulier s'il s'agit de cartographier le français régional et non les dialectes primaires. Comme les répondants avaient la possibilité de cocher plus d'un choix, le portrait d'ensemble est souvent complexe, mais cette complexité est le reflet de la réalité langagière.

3.4.2 L'axe syntagmatique

Dans nos enquêtes, les mots proposés aux enquêtés leur sont présentés dans une contextualisation extralinguistique (ex. : « Vous avez laissé traîner le pain trop longtemps et il est devenu impropre à la consommation car il est : a) canif ; b) moisi ; c) autre [précisez] ») mais pas nécessairement dans un énoncé concret (de style « phrases à trou »). Or, certaines variantes sont sensibles à l'environnement syntaxique immédiat. Par exemple, on peut très bien imaginer un témoin qui dirait *jus de canneberges* mais *sauce* ou *gelée d'atocas* (comme nous l'a suggéré une répondante dans la case consacrée aux commentaires libres – on n'insistera jamais assez sur l'importance du discours métalinguistique ainsi recueilli, l'un des points forts de ce type d'enquête). Si on le sait d'avance, il est possible de raffiner la question ou de la dédoubler (quitte à la poser de nouveau sous une formulation révisée dans une autre enquête), mais dans le cas contraire le protocole risque de laisser cette importante précision dans l'ombre. Quoi qu'il en soit, une étude distributionnelle dans de

vastes banques de données textuelles (de type « Europresse » ou « Google Recherche Avancée de Livres ») est nécessaire pour comparer l'usage déclaré des répondants et la situation dans la langue écrite (en l'occurrence, dans les corpus en ligne, *canneberges* et *atocas* se combinent tous les deux autant avec *jus* qu'avec *sauce* ou *gelée*, mais pas nécessairement dans les mêmes proportions).

3.4.3 Signification vs. désignation (référence)

Le protocole d'enquête demande aux répondants d'établir un lien entre un signe linguistique et un stimulus visuel (et/ou une contextualisation extralinguistique). Cela ne renseigne pas sur le contenu sémantique du signe ; les valeurs connotatives, entre autres, restent inaccessibles au chercheur. Les données récoltées relèvent de la désignation et non de la signification.

Le discours métalinguistique recueilli peut aider à corriger ce défaut. Par exemple, *shoe-claque* a été présenté par quelques enquêtés (dans la case « commentaires ») comme désignant préférentiellement une vieille paire de chaussures de sport, usées. La question portait sur des chaussures de sport en général, et l'image censée illustrer le concept représentait une paire de chaussures de sport vues comme prototypiques par les concepteurs de l'enquête, mais des « Converse » et des « Nike » sont-elles nécessairement désignées par le même mot ? Et qu'en est-il de la relation entre hypéronymes et hyponymes ? Si le nombre de commentaires critiques exprimés par des internautes insatisfaits des choix qu'on leur propose s'avère relativement élevé, l'enquêteur doit réviser le tir et reformuler la présentation de la question, voire la laisser de côté.

3.4.4 Oppositions binaires, multilatérales et graduelles

Les variables qui se prêtent le mieux à ce genre d'enquêtes sont celles qui correspondent à une opposition binaire. Les oppositions phonologiques sont particulièrement bien appropriées, car il s'agit par définition d'oppositions fondamentales d'un point de vue émique pour le locuteur, lesquelles sont souvent très difficiles à identifier de façon expérimentale, et à plus forte raison à l'oreille. À preuve ces passages de Walter (1977) où l'enquêtrice exprime la difficulté qu'elle éprouve à différencier des réalisations phoniques qui, pour les enquêtés, ont une valeur clairement phonématique : « bien que l'informateur ait une conscience très nette de toujours distinguer entre /a/ et /ɑ/, il est souvent difficile d'identifier les deux phonèmes à l'audition » (*id.*, 1977 : 81) ; « Quoiqu'elle nous ait déclaré qu'elle distinguait parfaitement et très naturellement *brun* de *brin*, nous avons eu du mal à identifier la voyelle /œ/ au début, car l'informatrice la réalise le plus souvent sans arrondissement des lèvres, soit avec une certaine longueur [œ:], soit avec une mise en valeur accentuelle [œ:]. » (*id.*, 1977 : 86). Sans le témoignage explicite des témoins, il y a fort à parier que l'auteure ne se serait jamais risquée à poser l'existence d'une opposition phonologique sur la seule base d'études de formants ; elle n'aurait probablement même pas entendu de différence significative entre les réalisations. D'où l'importance fondamentale, en phonologie, d'avoir recours aux autoreprésentations du sujet parlant.

Lorsque le choix proposé aux enquêtés est multilatéral (ternaire ou davantage), le problème qui se pose est celui du caractère fermé ou ouvert de la liste ; nous l'avons déjà évoqué au point précédent en référence aux différentes dénominations des chaussures de sport, qui posent la question des équivalences sémantiques et référentielles.

Enfin, les différences graduelles ne se prêtent pas du tout à ce genre d'enquêtes. Nous pensons ici, par exemple, au fait que les diphtongues en français québécois sont *plus ou moins* fortement réalisées. Il est impossible de demander aux témoins si leur prononciation correspond à une diphtongaison plus ou moins forte (*a fortiori* d'ailleurs en raison de la stigmatisation qui affecte les diphtongues les plus audibles, mais cela est une autre histoire).

De même, les voyelles fermées /i/, /y/ et /u/ en position entravée sont *plus ou moins* relâchées en français québécois : ici encore, il serait peu avisé de tenter d’interroger directement les témoins sur leur prononciation. Il s’agit de cas où on ne peut faire l’économie d’une étude de production menée avec des outils expérimentaux appropriés.

4 Traitement cartographique

Pour générer les cartes que nous présentons dans les sections suivantes, nous nous sommes basés sur le code postal de la localité dans lesquels les participants ont indiqué avoir passé la plus grande partie de leur jeunesse. Dans un premier temps, nous avons géocodé semi-automatiquement chacun de ces codes postaux avec l’outil ezGeocodeⁱⁱⁱ. Cette procédure nous a permis d’obtenir les coordonnées (longitudes et latitudes), la division de recensement (Municipalité Régionale de Comté, comté et district) et la province de chacune des entrées. Nous avons ensuite regroupé les participants selon leur division de recensement, ce qui nous a permis de disposer d’un réseau de 111 points, comprenant chacun entre 4 (MRC Les Jardins-de-Napierville ou MRC Mékinac) et 644 informateurs (Communauté-Urbaine-de-Montréal) :

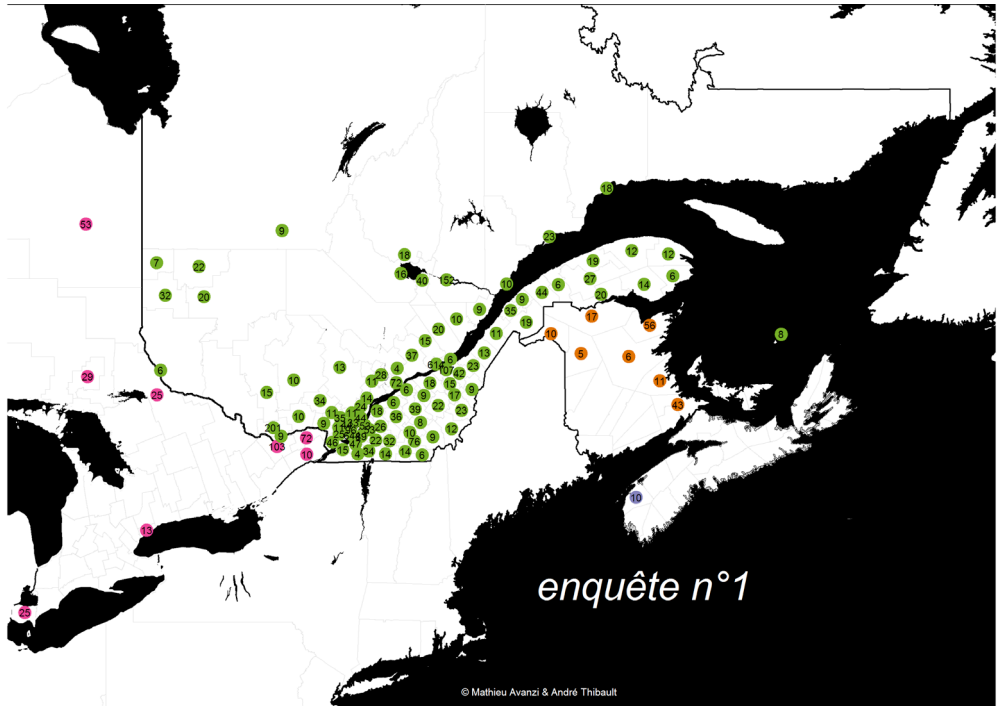


Fig. 5. Nombre de participants par MRC/district/comté de jeunesse [enquête n°1]

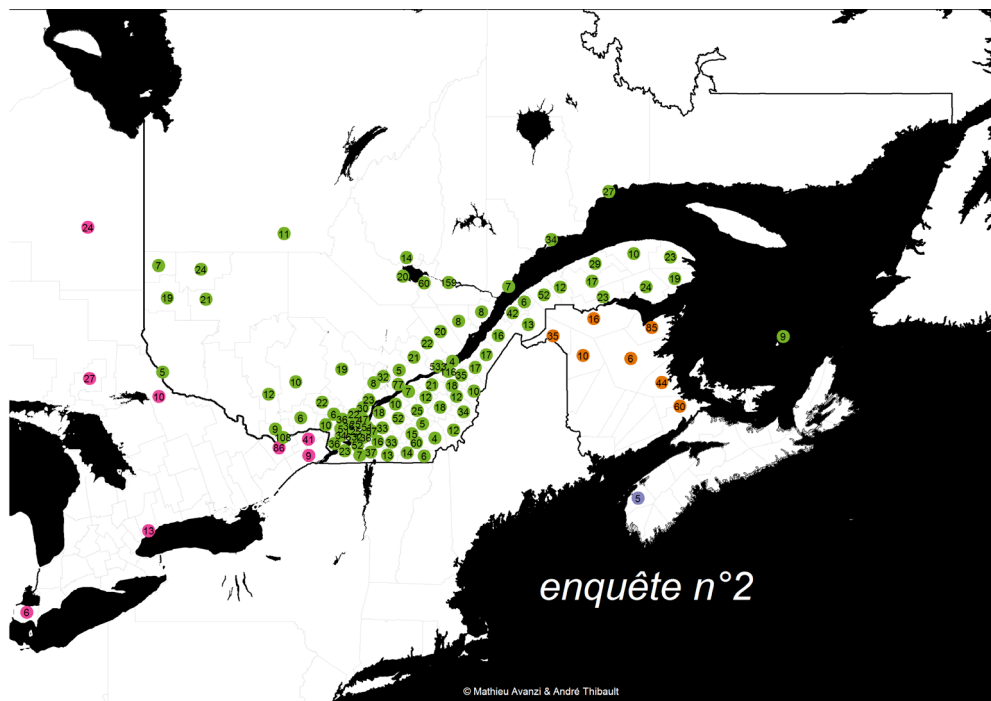


Fig. 6. Nombre de participants par MRC/district/comté de jeunesse [enquête n°2]

Pour chaque point du réseau d'enquête ainsi défini, on a alors calculé le rapport du nombre de réponses en fonction du nombre de participants, ce qui nous a permis d'obtenir un pourcentage allant potentiellement de 0% (= pas de participants) à 100% (= tous les participants). À partir de ces valeurs, nous avons utilisé la méthode du *krigeage* (Gratton 2002) pour obtenir une surface lisse et continue^{iv}. Les figures 7 et 8, obtenues à partir des réponses à la question portant sur la prononciation du mot *baleine*, permettent de rendre compte de la visualisation des résultats avant et après interpolation :

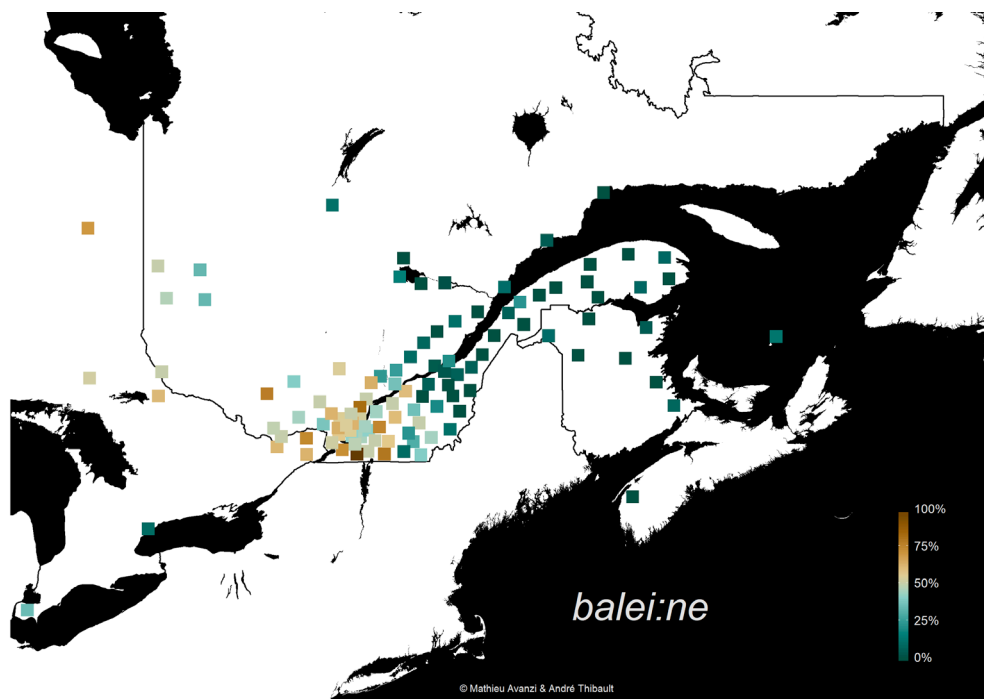


Fig. 7. Prononciation du mot *baleine* avec [ʒ] avant interpolation [enquête n° 1]

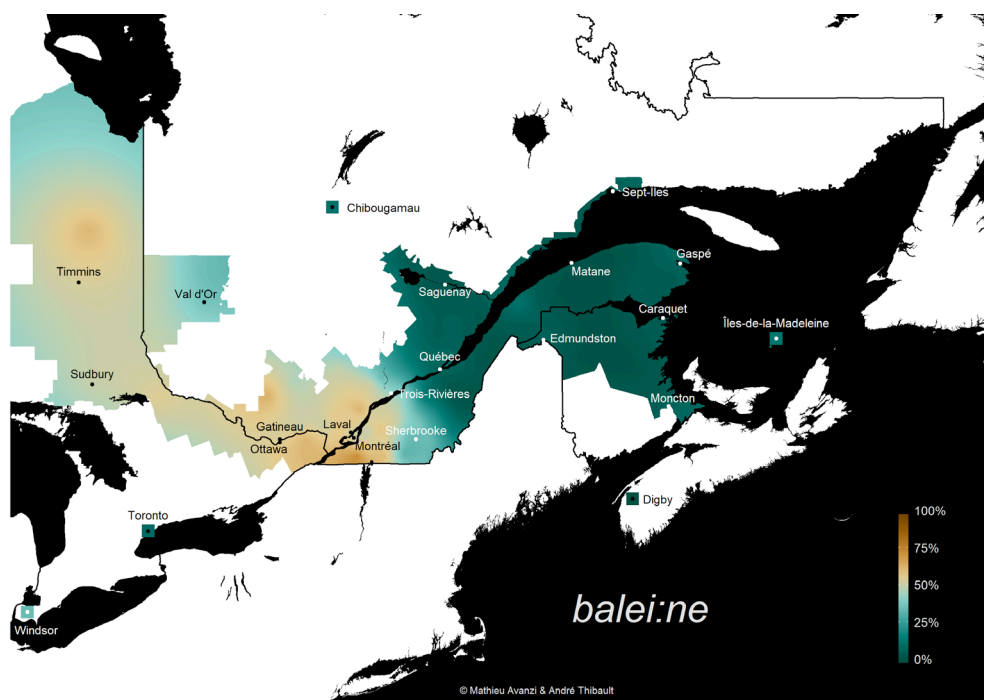


Fig. 8. Prononciation du mot *baleine* avec [ʒ] après interpolation [enquête n° 1].

5 Résultats

Nous allons présenter ci-dessous quelques résultats préliminaires tirés de nos enquêtes. Ils se répartissent en quatre catégories : a) aires complémentaires ; b) types locaux ; c) aires périphériques ; d) effet d'âge.

5.1 Aires complémentaires

Nous avons réuni dans cette catégorie des situations aréologiques où plusieurs types se partagent le territoire, sans qu'aucun d'entre eux ne domine vraiment les autres.

Il s'agit du concept de « sécher les cours, faire l'école buissonnière » et de celui de « chaussures de sport ».

5.1.1 Concept « sécher les cours, faire l'école buissonnière »

Quatre variantes se répartissent le territoire : *foxer*, *jigger*, *loafer* et *skipper*. On aura remarqué que ces termes d'argot scolaire sont tous des emprunts à l'anglais nord-américain.

Nous avons consacré une carte à part entière à chacun de ces types, afin que l'on puisse mieux visualiser leurs pourcentages d'utilisation respectifs sur l'ensemble du territoire. La variante *foxer* (figure 9) est typiquement montréalaise (elle est également attestée – dans une moindre proportion – en Gaspésie) ; *jigger* (figure 10) a été relevé au Madawaska (Nouveau-Brunswick) ; *loafer* (figure 11) couvre pratiquement tout l'est du Québec ; quant à *skipper* (figure 12), on le retrouve dans des zones latérales : l'Ontario, les Cantons de l'Est et la plus grande partie du Nouveau-Brunswick.

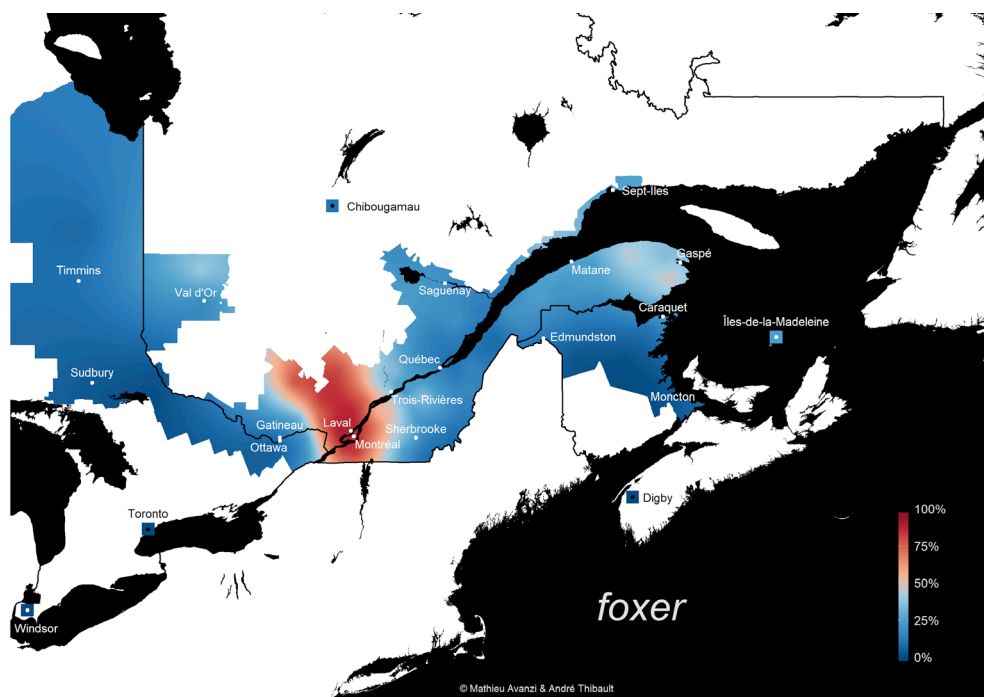


Fig. 9. Aire de *foxer* [enquête n°1]

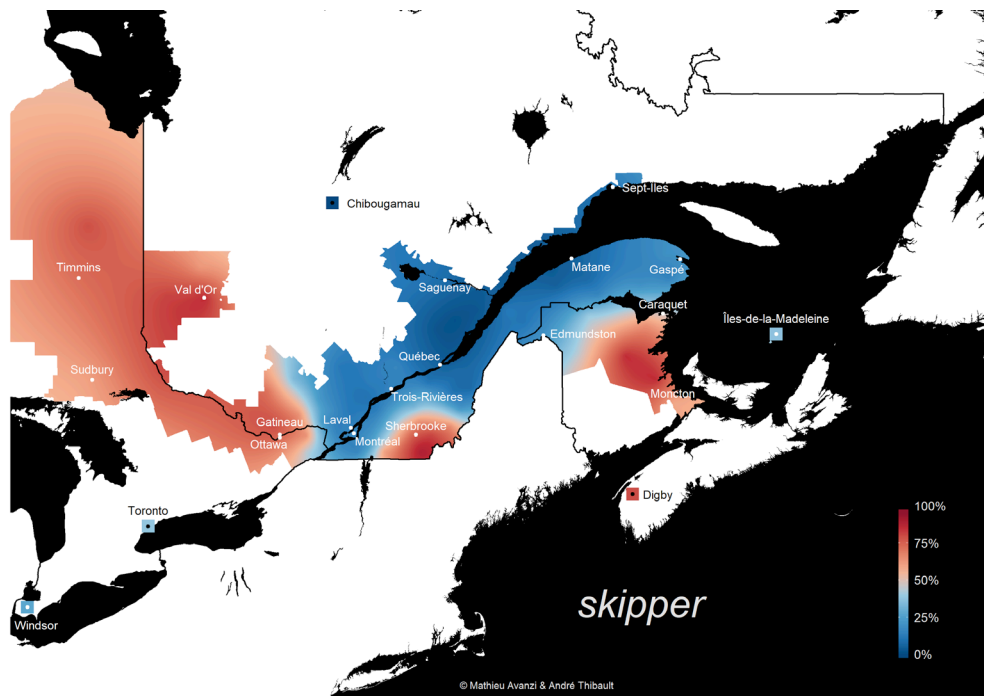


Fig. 12. Aire de *skipper* [enquête n°1].

5.1.2 Concept « chaussures de sport »

Le mot le plus répandu au Canada francophone pour désigner des chaussures de sport est *espadrille*. Il existe toutefois de nombreuses appellations propres à certaines régions : *runnings* (de l'angl. *running shoes*) est surtout montréalais (figure 13), alors que *shoe-claques* (< angl. *shoe* + fr. *claques*) domine dans la sphère d'influence de Québec (figure 14) ; quant à *sneakers* (ou sa variante apocopée, *sneaks*), ce sont les Cantons de l'Est et surtout le Nouveau-Brunswick qui le préfèrent (figure 15).

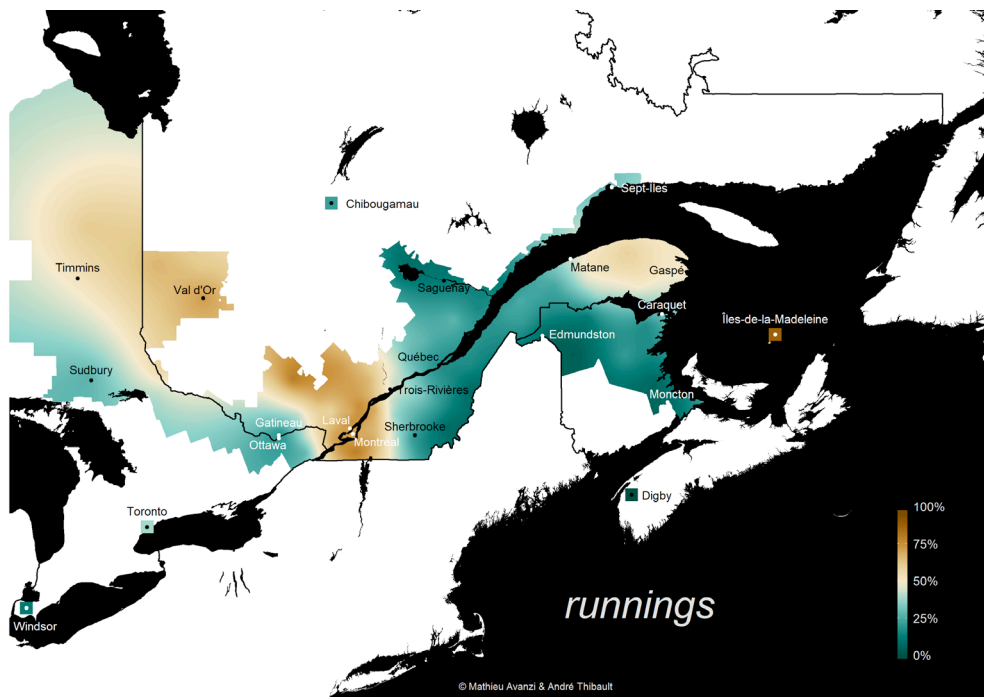


Fig. 13. Aire de *runnings* [enquête n°1].

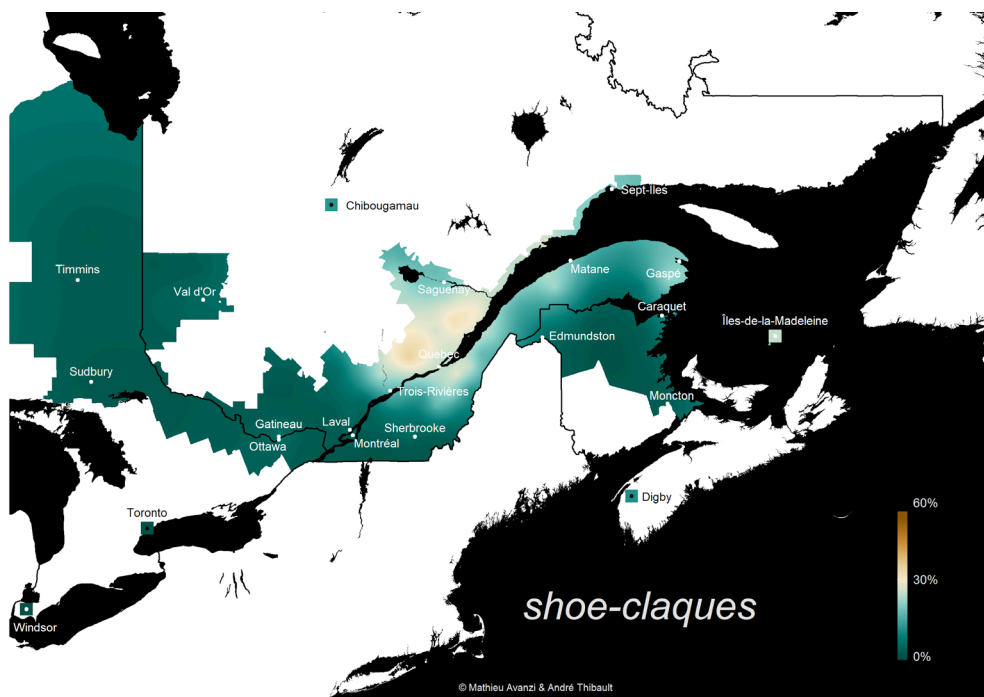


Fig. 14. Aire de *shoe-claques* [enquête n°1].

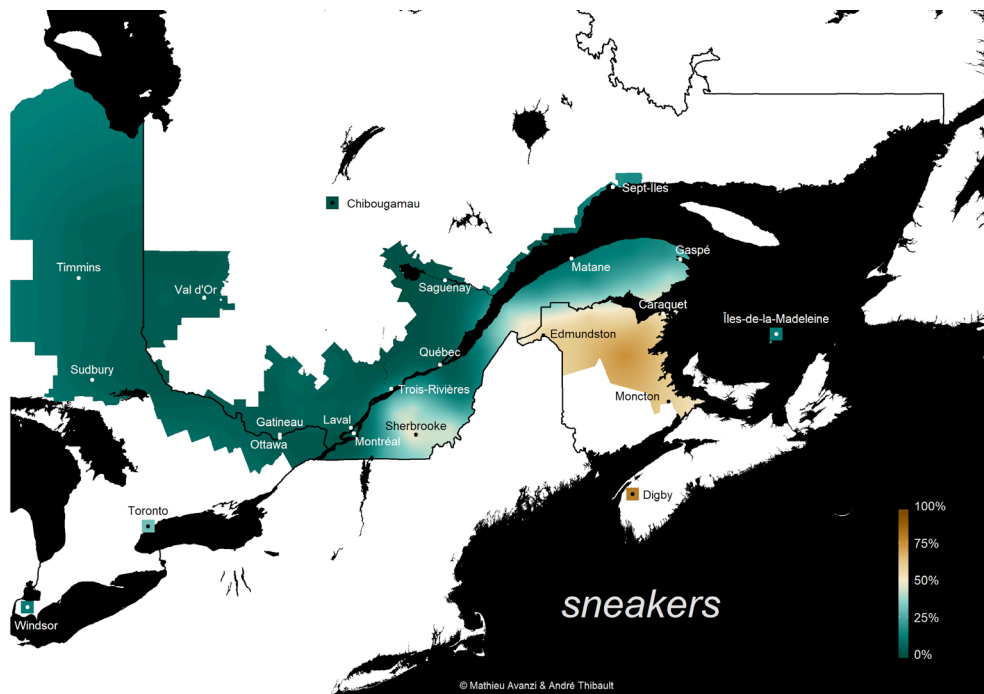


Fig. 15. Aire de *sneakers* [enquête n°1].

5.2 Types locaux

Certains types lexicaux ont une extension très locale. Nous illustrerons cette situation ci-dessous avec la locution verbale *faire simple* « avoir l'air simplet », typique de Charlevoix et du Saguenay–Lac-Saint-Jean mais inusitée dans le reste du pays (figure 16), ainsi qu'avec la locution nominale *pommes de pré* (figure 17) qui désigne ce qu'on appelle partout ailleurs au Canada *canneberges* ou *atocas* (en anglais et en français de France : *cranberries*).

5.3 Aires périphériques

Certains termes, plutôt que de s'être diffusés autour de l'influence d'un grand centre démographique, connaissent une aréologie qui dénote en fait une survivance archaïque en zones périphériques. Nous avons retenu deux exemples, *bête puante* « mouffette » (figure 18) et *siffleux* « marmotte » (figure 19), des zoonymes renvoyant à des animaux bien connus à la campagne mais dont les citadins n'ont une connaissance que livresque.

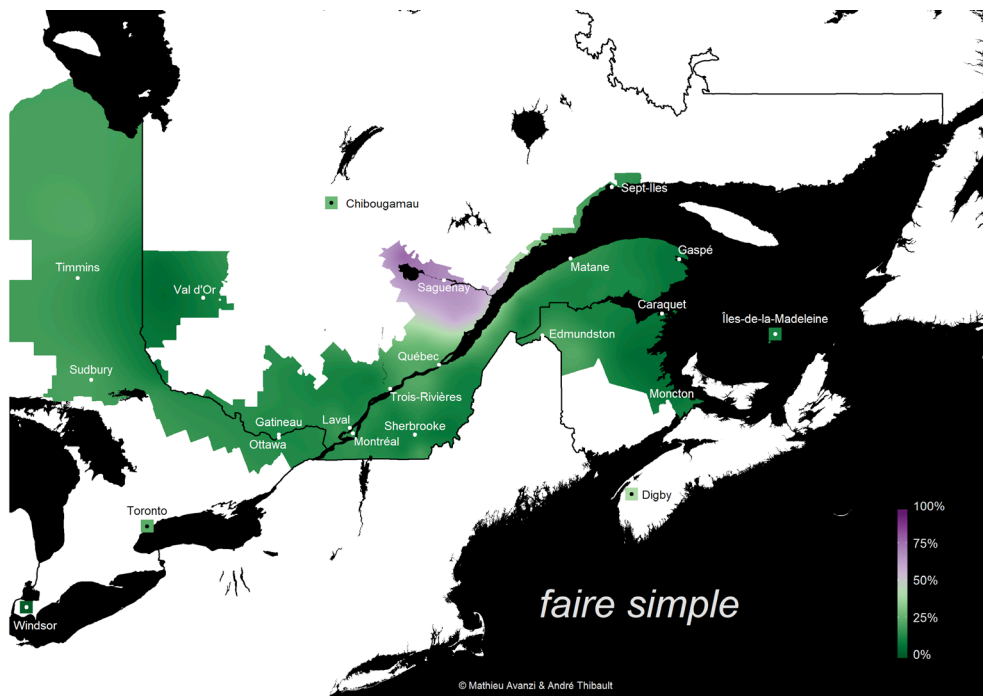


Fig. 16. Aire de faire simple [enquête n°2].

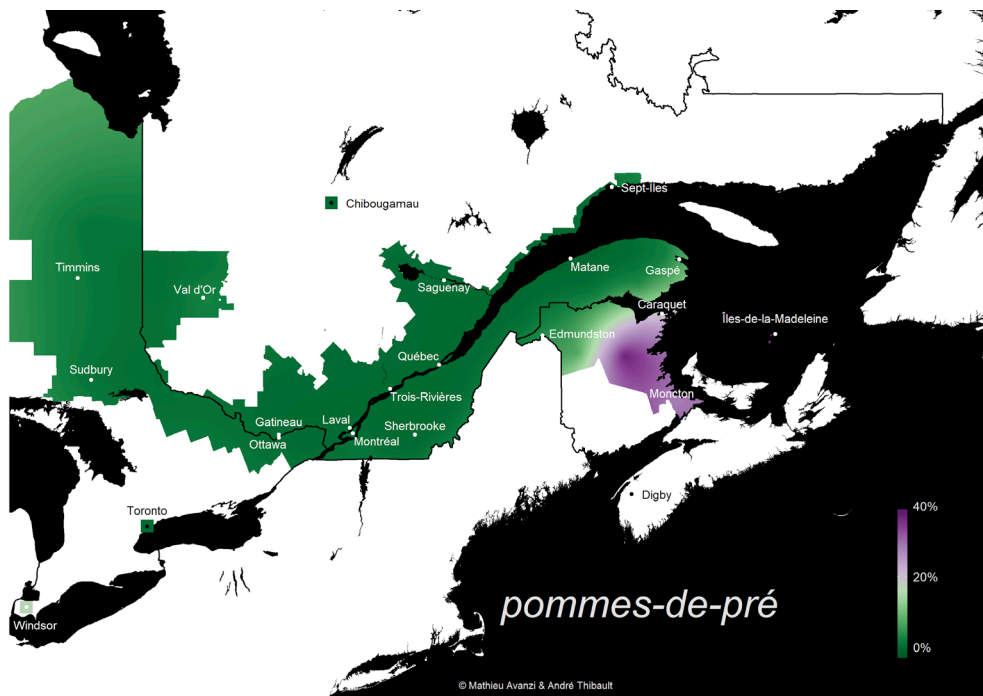


Fig. 17. Aire de pommes de pré [enquête n°2].

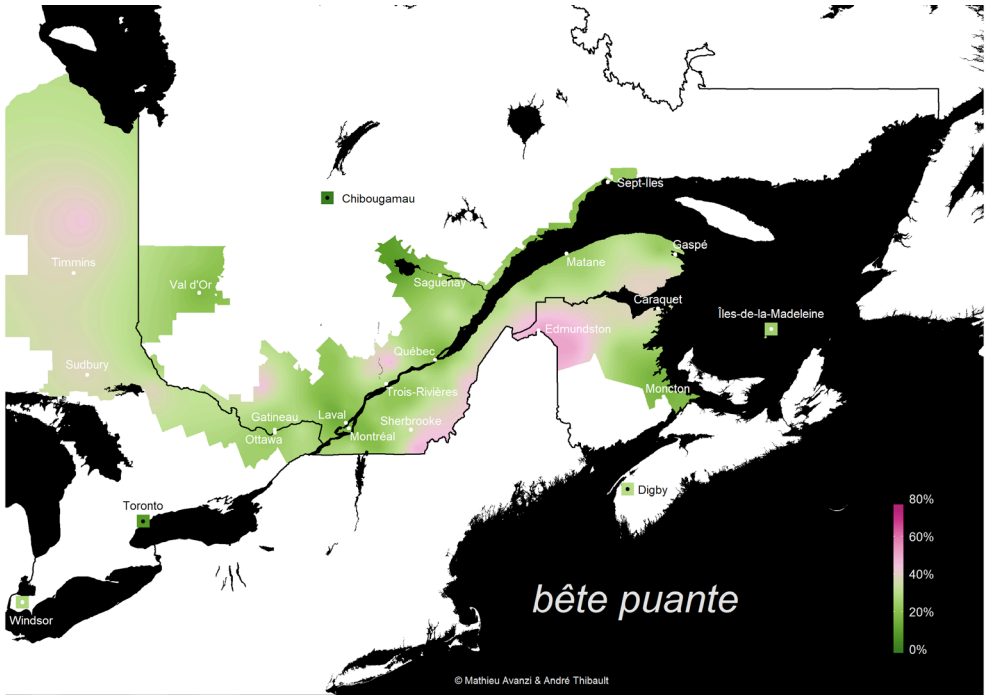


Fig. 18. Aire de *bête puante* [enquête n°1].

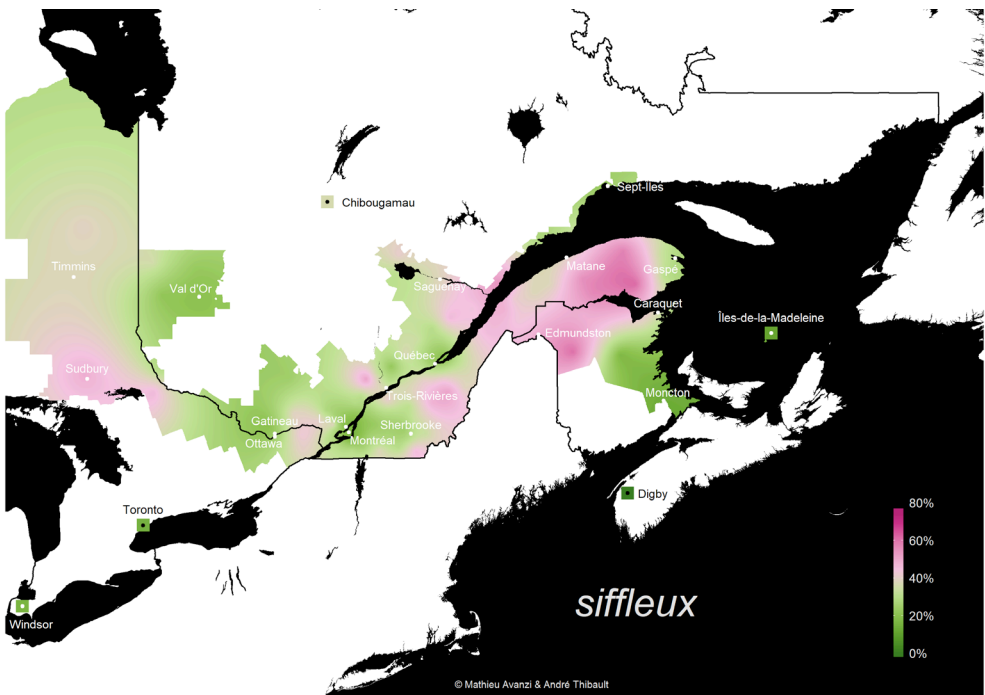


Fig. 19. Aire de *siffleux* [enquête n°2].

5.4 Effet d'âge

Certaines variables sont plus sensibles au facteur diagénérationnel qu'au facteur diatopique. C'est le cas du concept de « grand-mère », variable dont les trois variantes (*mémère*, *grand-maman* et *mamie*) ne se répartissent pas également selon l'âge des participants, comme on peut le voir sur la figure 20 ci-dessous. La variante *mémère* est la plus archaïque, suivie de *grand-maman*, qui occupe le milieu de l'échelle ; *mamie* clôt la marche, étant majoritairement utilisé par les moins de trente ans.

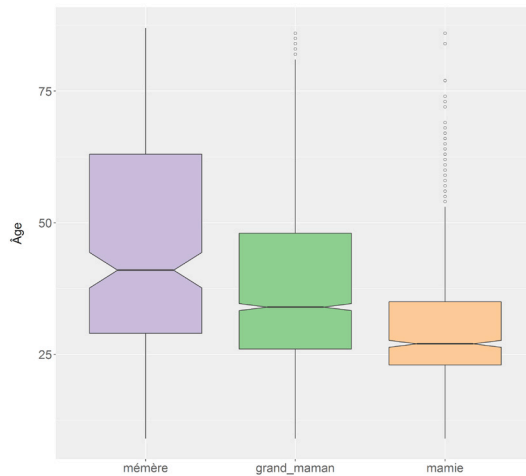


Fig. 20. Âge des participants en fonction de leur préférence lexicale pour la variable « grand-mère » [enquête n°1].

6 Bilan et conclusions

Après un bref tour d'horizon historique de la cartographie de la variation diatopique du français au Canada (qui n'en est encore qu'au stade embryonnaire, malgré l'abondance des données dialectales disponibles), nous avons présenté une nouvelle technique de recueil de données dialectologiques utilisant les ressources de l'Internet et la mobilisation des répondants sur les réseaux sociaux, technique qui nous a permis d'obtenir le témoignage de plusieurs milliers d'internautes répartis dans tout le Canada français. Les précieuses données ainsi recueillies ont fait l'objet d'un traitement cartographique entièrement informatisé. Cette nouvelle approche a été soumise à un questionnement épistémologique ayant pour but de mettre en relief ses points forts mais également d'attirer l'attention sur ses limites et ses lacunes : il s'agit bien de cartographier les *autoreprésentations* des locuteurs. Les résultats concrets présentés dans la dernière section montrent que la démarche peut s'avérer féconde : nous avons ainsi pu illustrer plusieurs cas de figure (aires complémentaires, types locaux, aires périphériques et effet d'âge) qui pourraient tous donner lieu à des recherches plus approfondies.

Les variables présentées dans cette contribution sont toutes lexicales. Faute de place, nous n'avons pu montrer le fonctionnement des variables phonétiques (v. toutefois les fig. 7 et 8), mais une analyse préliminaire des matériaux nous a déjà permis d'entrevoir tout leur potentiel. Ce sera l'objet d'une autre contribution.

7 Références

- ALEC = Dulong, G. / Bergeron, G. (1980). *Le Parler populaire du Québec et de ses régions voisines. Atlas linguistique de l'Est du Canada*. Québec : Gouvernement du Québec, 10 vol.
- ALF = Gilliéron, J. / Edmont, E. (1903-1910). *Atlas linguistique de la France*. Paris : Champion (35 fascicules en 17 volumes).
- ALLY = Gardette, P. (1950-1976). *Atlas linguistique et ethnographique du Lyonnais*. Lyon / Paris : CNRS (5 vol.).
- ALW = *Atlas linguistique de la Wallonie* (en cours de parution depuis 1953). Tableau géographique des parlers de la Belgique romane d'après l'enquête de Jean HAUST et des enquêtes complémentaires. Liège, 10 vol. parus.
- Atlas zur deutschen Alltagssprache : <http://www.atlas-alltagssprache.de/>
- Avanzi, M. *et al.* (2016). Présentation d'une enquête pour l'étude des régionalismes du français. In Bergounioux, G. / Côté, M.-H. / Fournier, J.-M. / Hriba, L. / Prévost, S. (dir.), *Actes du 5^e congrès mondial de linguistique française*. Tours : Institut de Linguistique Française, 1-15.
- Avanzi, M. (2017). *Atlas du français de nos régions*. Paris : Dunod.
- Bergeron, G. (1995). *L'Atlas Linguistique de l'Est du Canada et les aires de variations linguistiques au Québec*, thèse non publiée, Université de Sherbrooke (Canada).
- Chauveau, J.-P. / Lavoie, Th. (1993). À propos des origines dialectales du lexique québécois. *Revue de linguistique romane* 57, 373-420.
- Coseriu, E. (1980). « "Historische Sprache" und "Dialekt" ». In Göschel, J. / Ivić, P. / Kehr, K. (éds), *Dialekt und Dialektologie*. Wiesbaden : Steiner, 106-122.
- Côté, M.-H. (2012). Laurentian French (Québec). Extra vowels, missing schwas and surprising liaison consonants. In Gess, R. / Lyche, Ch. / Meisenburg, Tr. (dir.), *Phonological variation in French. Illustrations from three continents*. Amsterdam : John Benjamins, 235-274.
- Coté, M.-H. (2014). Le projet PFC et la géophonologie du français laurentien. In Durand, J. / Kristoffersen, G. / Laks, B., avec la coll. de J. Peuvergne (éds), *La phonologie du français : normes, périphéries, modélisation*. Nanterre : Presses Universitaires de Paris Ouest, 175-198.
- Coté, M.-H. (2016a). French in Quebec : A speaker from Montreal. In Detey, S. / Durand, J. / Laks, B. / Lyche, Ch. (eds), *Varieties of spoken French*. Oxford : OUP, 268-278.
- Coté, M.-H. (2016b). Variation in Canada: Trois-Rivières in Quebec. In Detey, S. / Durand, J. / Laks, B. / Lyche, Ch. (éds), *Varieties of spoken French*. Oxford : Oxford University Press, 449-462.
- Coté, M.-H. / Saint-Amant Lamy, H. (2012). D'un [r] à l'aut[ɣ]e : Contribution à l'étude de la chute du R apical au Québec. In Neveu, Fr. / Muni Toke, V. / Blumenthal, P. / Klingler, Th. / Ligas, P. / Prévost, S. / Teston-Bonnard, S. (éds), *3^e Congrès Mondial de Linguistique Française*. Paris : Institut de Linguistique Française, 2012, 1441-1453.
- Côté, M.-H. / Saint-Amant Lamy, H. / Milne, P. (2016). PFC-Québec : État des lieux et perspectives. Communication présentée aux *Journées PFC-Floral, 17-19 novembre 2016*.
- Français de nos régions : <https://francaisdenosregions.com/>
- Friesner, M. (2010). Une prononciation 'tsipéquement' québécoise ? : La diffusion de deux aspects stéréotypés du français canadien. *Canadian Journal of Linguistics / Revue canadienne de linguistique* 55(1), 27-53.
- Gratton, Y. (2002). Le krigeage. La méthode optimale d'interpolation spatiale. *Les articles de l'Institut d'Analyse Géographique*, juin 2002, www.iag.asso.fr.
- Katz, J. (2016). *Speaking American: How Y'all, Youse, and You Guys Talk: A Visual Guide*. Boston : Houghton Mifflin Harcourt.

- Labov, W. (1970). The study of language in its social context. *Studium generale* 23, 30-87.
- Laflamme, C. (2004). Distribution de quelques variantes géolinguistiques dans les parlers populaires de l'Est du Canada : essai de comparaison diachronique. In Mercier, L. (éd.), *Français du Canada – Français de France VI. Actes du sixième Colloque international d'Orford, Québec, du 26 au 29 septembre 2000*. Tübingen : Niemeyer, 123-149.
- Larochelle, P. (1898). *Analyse géolinguistique de quelques données phonétiques de l'Atlas linguistique de l'est du Canada*. Mémoire de maîtrise, Sherbrooke : Université de Sherbrooke (Canada).
- Lavoie, Th. (1991). La formation de régions linguistiques québécoises. *Dialangue* 2, 37-42.
- Lavoie, Th. et al. (1985). *Les parlers français de Charlevoix, du Saguenay, du Lac-Saint-Jean et de la Côte-Nord*. Québec : Gouvernement du Québec (5 vol.).
- Massignon, G. (1962). *Les parlers français d'Acadie*, Enquête linguistique. Paris : Klincksieck.
- Péronnet, L. et al. (1998). *Atlas linguistique du vocabulaire maritime acadien*. Sainte-Foy (Québec) : PUL.
- PFC = *Phonologie du français contemporain*, <https://www.projet-pfc.net/>
- R development core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna (Austria).
- Vaux, B. / Golder, S. (2003). *The Harvard Dialect Survey*. Cambridge : Harvard University Linguistics Department. <http://dialect.redlog.net/>
- Verreault, Cl. / Lavoie, Th. (1996). Genèse et formation du français au Canada : l'éclairage de la géographie linguistique. *Revue de linguistique romane* 60, 413-462.
- Verreault, Cl. / Lavoie, Th. (2004). Les parlers de l'Est et de l'Ouest québécois : essai de caractérisation linguistique. In Mercier, L. (éd.), *Français du Canada – Français de France VI. Actes du sixième Colloque international d'Orford, Québec, du 26 au 29 septembre 2000*. Tübingen : Niemeyer, 71-121.
- Walter, H. (1977). *La phonologie du français*. Paris : PUF.

ⁱ On dit aussi *production participative* ou *externalisation ouverte*.

ⁱⁱ © 2017 Qualtrics, Provo, UT, USA, <http://www.qualtrics.com>

ⁱⁱⁱ <https://ez34.net/ezapps/ezgeocode>

^{iv} Les cartes ont été générées avec différents 'packages' du logiciel R (R development core Team, 2017).