

A new milestone for the study of variation in Montréal French: The Hochelaga-Maisonneuve sociolinguistic survey

Hélène Blondeau, Mireille Tremblay, Anne Bertrand and Elizabeth Michel



Electronic version

URL: <http://journals.openedition.org/corpus/6221>

DOI: [10.4000/corpus.6221](https://doi.org/10.4000/corpus.6221)

ISSN: 1765-3126

Publisher

Bases ; corpus et langage - UMR 6039

Electronic reference

Hélène Blondeau, Mireille Tremblay, Anne Bertrand and Elizabeth Michel, "A new milestone for the study of variation in Montréal French: The Hochelaga-Maisonneuve sociolinguistic survey", *Corpus* [Online], 22 | 2021, Online since 15 February 2021, connection on 16 February 2021. URL: <http://journals.openedition.org/corpus/6221> ; DOI: <https://doi.org/10.4000/corpus.6221>

This text was automatically generated on 16 February 2021.

© Tous droits réservés

A new milestone for the study of variation in Montréal French: The Hochelaga-Maisonneuve sociolinguistic survey

Hélène Blondeau, Mireille Tremblay, Anne Bertrand and Elizabeth Michel

AUTHOR'S NOTE

This research was supported by SSHRC research grants #435-2019-0640 *40 ans de changement linguistique à Montréal et à Welland: le rôle de l'individu et de la communauté* (M. Tremblay) and #412-2011-1005 *Le français à la mesure d'un continent* (F. Martineau). We would like to thank our colleagues, France Martineau, Yves Frenette, and Paul-André Linteau, as well as Raymond Mougeon for their contribution to this research endeavor, our research assistants (Guillaume Beauchamp, Geneviève Gauthier, Philippe Morneau), the participants at the conference *50 ans de linguistique sur corpus oraux* (U. of Orléans), NWAV48 2019 (U. Oregon) for their feedback, two anonymous reviewers and the editors of the volume for their useful comments. Any remaining errors are entirely our own.

1. Introduction

- 1 Sociolinguistic surveys conducted in Montréal since the early 1970s have contributed to the description of linguistic variation in spoken French from the perspective of apparent and real time. Longitudinal analyses of these corpora collected in three waves—1971 (Sankoff et al 1976), 1984 (Thibault & Vincent 1990) and 1995 (Vincent et al. 1995)—led to major advances in understanding linguistic change in the community and across the lifespan (Sankoff 2017, 2019).

- 2 In the context of the project *Le français à la mesure d'un continent*, a team of researchers put together the FRAN corpus which includes data representing French varieties spoken in North America (Martineau & Séguin 2016). This article introduces one of the components of the FRAN corpus: the Hochelaga-Maisonneuve variationist sub-corpus (HOMA) collected in Montréal in 2012 (Blondeau, Frenette, Martineau & Tremblay 2012).
- 3 In section 2, we provide background information on previous Montréal French corpora and explain the objectives prompting the collection of new data. In section 3, we discuss the methodological choices regarding the composition of the Hochelaga-Maisonneuve sub-corpus including the historical and demo-linguistic criteria that led to the selection of this neighborhood. In section 4, we introduce the socially stratified sample and criteria for participant selection, and detail the data collection protocols. The conditions of access to the corpus and their limits are also discussed. Section 5 is devoted to the treatment and analysis of the data with a focus on how technological advances have changed the way we process and access it. Finally, in section 6, we illustrate with two case studies the revival of research on linguistic variation made possible by the Hochelaga-Maisonneuve sub-corpus. We first focus on variation in the use of consequence markers over time. We then turn to a case of phonomorphological variation in the use of the preposition *dans*.

2. Previous sociolinguistic research on Montréal French (1971-1995)

- 4 This section summarizes previous sociolinguistic research on Montréal French with a focus on the three-pronged longitudinal study conducted in 1971, 1984 and 1995.

2.1. Milestone 1: The first sociolinguistic survey of Montréal French, 1971

- 5 In the 1970s, documenting local varieties of spoken French in Québec was a pressing issue (Blondeau 2020). The first large sociolinguistic survey undertaken focused on the variety of French spoken in Montréal and marked the development of the emerging field of sociolinguistics (Labov 2006). Drawing on the work of Labov in New York (1966), the Sankoff-Cedergren Corpus project (Sankoff et al. 1976) began in 1971. It involved the large-scale collection of spoken data conducted by a multidisciplinary team. Their main goal was to describe linguistic variation in Montréal French and discover the ordered heterogeneity (Weinreich et al. 1968) at play at the community level. In the researchers' own words, the contribution of the Sankoff-Cedergren corpus was as follows:

Nous voulions contribuer à une meilleure compréhension du français parlé au Québec en considérant ses aspects propres non comme des erreurs ou aberrations ou encore en terme 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 D. et al. 1976: 88).

- 6 This first corpus was based on a socially constructed sample of Montréal's Francophone population. It was comprised of 120 recorded interviews, followed by a reading task. All participants were native French speakers born and raised in Montréal and selected via

a random stratified sample.¹ The sampling grid, designed to ensure representativity, was comprised of six socioeconomic levels, balanced for gender and age (Sankoff et al. 1976; Sankoff 2017). This first data collection was characterized by an effort made to record spontaneous speech in conditions as close as possible to authentic, vernacular communication situations, while at the same time preserving sound quality.

- 7 This study design paved the way for quantitative studies at different levels of linguistic structure, particularly in phonetics, phonology and morphosyntax (for a review of early work, see Thibault & Vincent 1990; Daveluy 1994). These studies relied on the Apparent-Time model (Labov 1972), which postulates linguistic change based on synchronous data. According to this approach, change is accounted for by an indirect measure of the time factor, which in the case of the 1971 study was participant age. The first quantitative analysis of the change in the pronunciation of the /r/ phoneme in Montréal French (Clermont & Cedergren 1979) illustrated the relevance of this model, and the key role played by women and youth in the decline of the apical variant.
- 8 This first period was characterized by innovations in data analysis and methodology. A first multivariate analysis program known as Varbrul was designed to identify trends governing sociolinguistic variation (Cedergren & D. Sankoff 1974; Rousseau & Sankoff 1978). The social dimensions of the analysis were based on the speakers' educational and occupational features (Sankoff 2019), but very early on the conceptualization of the social dimension was improved with the development of an index of participation in the language market² (Sankoff & Laberge 1978).

2.2. Longitudinal studies of Montréal French

Milestone 2: The Montréal-1984 corpus

- 9 Thirteen years later, a longitudinal project was launched by Thibault and Vincent, two former doctoral students of G. Sankoff, in collaboration with D. Sankoff. It led to the collection of the Montréal-1984 corpus (Thibault & Vincent 1990). The goals were to verify how changes postulated according to the apparent-time model were validated in real-time and to answer questions related to interactional sociolinguistics and discourse analysis.
- 10 The team succeeded in tracing and re-interviewing 50% (n=60) of the initial participants to the 1971 study, despite the fact that a second wave of data collection had not been part of the initial 1971 study-design (Daveluy & Laforest 1994). To ensure balanced representation across age groups, 12 interviews were also conducted with new participants between the ages of 15 and 25. In total, the corpus was comprised of 72 interviews. The data collection protocol was similar to that of 1971 with the exception of the reading task, which was discarded in 1984.
- 11 This kind of follow-up with a community after an initial sociolinguistic survey was an innovative direction in sociolinguistics (see Trudgill 1988 for English and Cedergren 1988 for Spanish). Although Thibault & Vincent (1990) recognized that the reduction in size of the 1984 sample resulted in some loss of representativity—particularly with regard to the lower socio-economic status (SES)—the corpus was innovative in that it offered two avenues for monitoring language change over time: community trend follow-up and panel study (Sankoff 2017). For the sixty participants who were re-interviewed, it was possible to follow changes in individuals across a span of 13 years. It

was also possible to compare a sample of the community at Time 1 (120 speakers in 1971) with Time 2 (72 speakers in 1984) in order to identify community trends.

- 12 This re-examination of community speech patterns provided a better understanding of the relationship between community change and age-grading phenomena; a relationship that apparent-time cannot capture (Labov 1994). This distinction was well illustrated by a real-time analysis of the use of discourse markers over time (1971-1984) using both the trend and the panel components (Thibault & Daveluy 1989). In particular, the use of the consequence marker *alors* 'then'—initially considered to be part of a change in progress at the community level due to its association with older speakers (Dessurault-Dober 1973)—was argued to be associated instead with an age-grading phenomenon.
- 13 In sum, research based on this second wave of Montréal data had a significant impact on real-time sociolinguistic research (Sankoff & Blondeau 2007; Sankoff 2019). The methodological distinction between trend study and panel study provided a better understanding of the dynamics of language change both at the community level and at the individual level. This distinction was impossible to discern based on a strict analysis of the effect of age on synchronic variation and has contributed to refinements in the theory of language variation and change (Sankoff 2017).

Milestone 3: The Montréal-1995 corpus

- 14 In 1995, Vincent and her collaborators returned to the Montréal community to compile a third corpus. Smaller in size, the Montreal 1995 corpus was comprised of sociolinguistic interviews with twelve of the participants interviewed in both 1971 and 1984. In addition, the sample included two of the twelve young speakers who were added to the Montréal 1984 corpus (Vincent et al. 1995). These fourteen 1995 sociolinguistic interviews were similar in form and content to those conducted in 1971 and 1984. Additionally, the 1995 protocol included recordings of spontaneous speech of four participants in their respective homes, giving access to a variety of speech events involving the participant and other family members or friends without the presence of an interviewer. Despite the reduced sample size, the longitudinal nature of this corpus allowed for the examination of intra-individual variation over time, from 1971 to 1995. Moreover, the situational range of the data collected broadened the potential of the analyses from an interactional perspective.
- 15 The research based on the Montréal corpora has played a major role in the development of the research stream on language change over the lifespan (Buchstaller & Wagner 2017). Until the 1990s, the phenomena related to the sociolinguistic development of individuals over the lifespan was under-explored (Sankoff 2019), and thus the 1995 Montréal French data contributed to a better understanding of this dimension. Although less suitable for the identification of community trends, the Montréal 1995 corpus did provide an opportunity to examine intra-individual variation over time and its social significance. Indeed, it extended the time scope for the study of the sociolinguistic positioning of the individual over time for various linguistic variables (Sankoff et al. 2001; Sankoff 2017; Blondeau 2001; Blondeau et al. 2018). Studies of intra-individual variation over time remain rare for the French language. Apart from the Montréal project, two corpora on French varieties include a similar longitudinal dimension. In Welland, Ontario, Raymond Mougeon has been documenting community and lifespan change (Beniak et al. 1985; Martineau & Mougeon 2013).

Similarly, a project in Orléans, France (1968-2012), tracks the evolution of hexagonal French (Abouda & Baude 2007; Baude & Dugua 2016). In both cases a cohort of 12 speakers were reinterviewed after almost 40 years.

3. Montréal 2012

3.1. The need for a new corpus

- 16 The 20th century corpora were successful in their attempts to model change both in progress at the community level (apparent- and real-time) and across the lifespan (real-time). As demonstrated above, the three-pronged longitudinal study provided a unique opportunity to describe Montréal French over time. Research that developed out of these three milestone corpora contributed to the theory of language change, in particular towards the establishment of a better understanding of the relationship between age-grading and generational change at the community level (Blondeau 2011; Buchstaller & Wagner 2017; Sankoff 2019; Sankoff & Blondeau 2007; Thibault & Daveluy 1989; Wagner & Sankoff 2011).
- 17 However, because these three corpora were not initially designed to model community change in real-time, the 1984 and 1995 corpora did not allow for a perfect comparison with the 1971 corpus. The trend comparisons between 1971 and 1984 offered a relatively short window of 13 years for the possible identification of linguistic change. Although a longer period of 24 years elapsed between 1971 and 1995 corpora, the small sample size of the 1995 corpus (only 12 were interviewed in both 1971 and 1995) made any comparison with the much larger 1971 corpus difficult. The 1995 corpus was also limited in its potential for the identification of community trends as it was composed of data from the same speakers surveyed in previous corpora. There was also an over-representation of the middle and upper classes in the 1984 and 1995 corpora.
- 18 It was therefore clear that a new stratified corpus was needed to observe community changes in real time over a wider time span. Such a corpus would allow to further verify the hypothesis that linguistic and social synchronic variation was the basis for diachronic variation (Labov 1994). By comparing community linguistic dynamics from 1971 and 2012, we can further test the apparent-time model to see if its predictions can be confirmed in real time.

3.2. Choice of the fieldwork site

- 19 While the greater Montréal area still bears the evidence of the traditional split between Francophone and Anglophone communities along its eastern and western parts, most neighborhoods are increasingly culturally and socially diverse. This poses a challenge for any comparison with previous corpora assembled at a time when Montréal's Francophone neighborhoods were more homogeneous. Over the last forty years, the social configuration of Montréal has evolved due to the effects of globalization and the implementation of the French Language Charter (Bill 101) in 1976. The latter measure has reduced the presence of English, increased the use of French in the workplace and its visibility in the linguistic landscape, and required new immigrants to Québec to send their children to Francophone schools. At the same time, an influx of more affluent residents into previously working class neighborhoods has produced an effect of

gentrification. In the 2010s, neighborhoods are less socially and culturally homogenous than they once were. The 1971 sample selection procedure cannot be used because neighborhoods can no longer be selected on the assumption that they reflect a particular social class. In the 21st century, the sample selection procedure must take into account both the cultural and social diversification of neighborhoods.

- 20 In the context of the project *Le français à la mesure d'un continent: un patrimoine en partage* (Martineau et al. 2011-2018) (continent.uottawa.ca), a multidisciplinary team explored various Francophone communities across North America (Martineau & Seguin 2016). New fieldwork was undertaken in order to explore the situation of 21st century Montréal and to take into account how these social changes affected language practices in the Francophone community. A neighborhood-based approach was adopted, and sub-corpora were collected in two Francophone neighborhoods of Montréal: Hochelaga-Maisonneuve and St-Michel-Montréal-Nord.³ In this article, we focus on the Hochelaga-Maisonneuve sub-corpus (Blondeau et al. 2012).

3.3. Focus on Hochelaga-Maisonneuve

- 21 Hochelaga-Maisonneuve is one of the neighborhoods that best lends itself to comparison with the previous corpora; it is a predominantly French-speaking, socially diverse urban neighborhood. This ensures a solid comparison with previous corpora across social and cultural criteria. Situated in the eastern part of the city where Francophones constitute a majority,⁴ Hochelaga-Maisonneuve is a former industrial area whose inhabitants were former rural residents from the Great Plain of Montréal (Linteau 1981). At the turn of the 20th century, this prosperous neighborhood provided many jobs for workers. This era of prosperity was followed by a period of economic decline brought about by de-industrialization. This decline led to an increase in poverty and unemployment, and to the development of many community networks. More recently, with the arrival of new residents from the middle class, the neighborhood has become gentrified.⁵ This cohabitation of populations from different social backgrounds has led to social diversification, making it possible to examine the effects of gentrification and social mixing (Blondeau & Tremblay 2012). Moreover, this social diversification facilitates comparisons with the socially stratified 1971 Montréal French corpus.

4. The Hochelaga-Maisonneuve corpus (Blondeau, Frenette, Martineau & Tremblay variationist corpus)

4.1. The stratified sample

- 22 Initially, the team planned for a sample of 48 speakers selected on the basis of age, socio-economic status and gender. Roughly an equal number of male and female participants of diverse age and socio-economic status (SES) were intended in order to construct a speech corpus suitable for the sociolinguistic analysis of language variation. A stratified sample scheme was initially devised: 48 participants (24 M and 24 F), from 4 age groups (emerging adults 18-25, young adults 26-39, older adults 40-60, and mature adults 61+), and six socioeconomic backgrounds. The categorization of speakers according to SES was based on the occupational history of the speakers or their

families⁶: SES 1 (Liberal profession/Business Person), SES 2 (Bachelor Degree/Intellectual), SES 3 (Technician /Supervisor), SES 4 (White-collar/office worker), SES 5 (Blue-collar/manual worker), and SES 6 (history of no stable employment). Participants were required to have grown up in the Great Plain area of Montréal, to have been educated in French and to have been living in the neighborhood for at least five years.

- 23 In the end, data was collected from 50 participants of which 23 were female and 27 were male, and ranging from ages 18 to 89 distributed as follows: 13 participants (18-25), 12 participants (26-39), 12 participants (40-60), and 13 participants (61+). The construction of a socially-balanced speaker sample was somewhat hampered by the fact that it was difficult to recruit older participants situated at the higher end of the social hierarchy. For this reason, speaker sample suffered from a shortage of higher SES speakers in the oldest age-group, as shown in table 1.

Table 1. Sample

	SES 1 Liberal profession/ Business Person		SES 2 Bachelor Degree / Intellectual		SES 3 Technician/ Supervisor		SES 4 White- collar/ Office worker		SES 5 Blue- collar/ manual worker		SES 6 Unemployed	
	F	M	F	M	F	M	F	M	F	M	F	M
Emerging Adults 18-25	1	1	1	1	1	1	1	1	1	2	1	1
Younger Adults 26-39	1	1	1	1	1	1	1	1	1	1	1	1
Older Adults 40-60	1	1	2	1	1	1	1	1	0	2	1	1
Mature Adults 61+	0	0	0	1	0	4	1	1	2	2	2	0

4.2. Data collection protocol

- 24 To preserve comparability with previous sociolinguistic corpora of spoken Montréal French, the data was collected via face-to-face sociolinguistic interviews conducted by different interviewers, all native speakers of Québec French. Thirteen of the older speakers were interviewed by project co-investigators (two women and a man in their fifties), while the remaining interviews were conducted by students (one man and two women ranging in age from 24 to 31). The interviews were intended to mimic the flow of natural conversation and were loosely based on a script which included (i) questions designed to elicit spontaneous responses on topics such as childhood memories, and important historical and biographical events (ii) questions on life in the neighborhood and changes it may have undergone, and (iii) a series of questions on the use and status of the French language in Montréal and Québec. While similar in content and style to

the 1971 and 1984 interviews, some of the topics covered in 2012-2015 reflected changes in the cultural and socioeconomic context of Montréal as well as the local context of the neighborhood. The interviews lasted between one and two hours and were usually conducted in the home of the participant.

5. Data management

5.1. Transcription

- 25 All 50 interviews were transcribed and audio-aligned: 38 interviews were transcribed in ELAN (Wittenburg et al. 2006) at the Université de Montréal under the supervision of M. Tremblay and 12 interviews were transcribed in Praat at the University of Ottawa under the supervision of F. Martineau. The transcriptions followed the protocol for the FRAN corpus (Martineau & Séguin 2016), designed to facilitate the retrieval of comparable morphosyntactic variables (Martineau et al. 2014). It aimed to produce accurate transcriptions while maintaining uniformity across sub-corpora. Morpho-phonological phenomena were not transcribed systematically. However, because the transcriptions were audio-aligned, it remains possible to retrieve information on specific phonological variables from their context within the transcription. The transcriptions also recorded conversational events of a more pragmatic nature such as interruptions, back channeling, hesitation, false starts, etc. The transcriptions were made using standard French orthography when doing so did not obfuscate morphosyntactic characteristics of the varieties. The FRAN corpus is located at the Université de Sherbrooke, and is available online via catfran.flsh.usherbrooke.ca/txm. While the texts of the FRAN Corpus are not accessible in their entirety, it is possible to query the available interviews and obtain results in the form of a concordance.

5.2. Annotation

- 26 The audio-alignment of interviews allows for easy tracking and annotation of relevant occurrences. Annotation directly under the segment means that it is possible to listen to the interview without having to search for the matching segment in a separate audio file. This alignment also allows for on-going modification. In particular, Elan can be used to codify various types of linguistic information, which can then be extracted and analyzed. The information can be codified according to a predetermined scheme on individual tiers, simplifying the subsequent linguistic analyses as the information is wellorganized in a unique environment.

6. Illustration

- 27 This section illustrates the advantages of the corpus with its sound-text aligned transcription and dynamic annotation in the renewal of the questioning of linguistic variation. We first discuss variation in the use of the consequence markers *alors*, *donc*, and *(ça) fait que*. This variable is involved in a vigorous change in progress at the community level. We then turn to preposition-determiner contraction, a phenomenon that is losing momentum in Montréal French.

6.1. Variation in the use of consequence markers

- 28 Since Dessurault-Dober (1974), sociolinguistic studies have examined the variation at play in the usage of consequence markers *alors*, *donc* and *(ça) fait que* mainly from the perspective of lexical variation by examining its discursive and social constraints (Blondeau et al 2019). The example in (1) illustrates the use of the consequence marker *(ça) fait que*, but both *alors* and *donc* can be used in the same context without any change in meaning.

(1) *Ben ma mère est pas super scolarisée mais elle s'intéressait à plein de choses fait-que c'est possible que elle apprenait beaucoup de mots mais elle les utilisait pas comme il faut tu sais.* (HOMA_008F27).

'Well, my mother's not very well educated, but she was interested in a lot of things, so it's possible that she learned a lot of words, but she didn't use them properly, you know.'

- 29 The new data collected in 2012 provides a new perspective on this variable undergoing change. The analysis of the lexical variation in 2012 and the comparison with the 1971 data confirm the spread of the *(ça) fait que* variant within the community, and show that the use of *alors* declined sharply between 1971 and 2012 (Martineau & Séguin 2016; Blondeau et al. 2019). The standard variant *donc* appears to be a good candidate for the replacement of *alors* as the prestigious variant because, while it was almost absent in 1971, its use has increased over time.
- 30 The apparent-time analysis of the 2012 data set demonstrates that the use of *donc* was short-lived as it was mainly favored by speakers ages 40 to 60 years with a high socio-economic status, most of whom were women. A subsequent examination of its usage by speakers in the youngest age group indicates that this variant is neglected by women and the youngest group of male speakers. Initially associated with a lower socioeconomic status, *(ça) fait que* is now being used by speakers of the upper end of the social scale to the detriment of the other variants. While real-time analyses confirmed the rapid rise of *(ça) fait que*, an analysis focusing only on lexical variation left the popularity of this vernacular form unexplained.
- 31 The sound-text aligned transcription and dynamic annotation facilitate the analysis of the sociophonetic realization of the variant *(ça) fait que*, an aspect that was never quantitatively analyzed, due in part to the lack of tools available for previous corpora of Montréal French. A careful analysis of the various phonological realizations of the variants sheds light on the role of sociophonetic factors in the change at play.
- 32 We examined the social distribution of two phonological variants of *(ça) fait que* that are distinguished by vowel quality: /fak/ and /fɛk/, illustrated in (2).
- (2) a. *Au besoin il y a les taxis pis tout est proche à Montréal fait-que [fak] si jamais on a besoin d'un taxi on s'en sort à dix piastres.* (HOMA_001M29)
'If we need, there are taxis and everything is close by in Montreal, so if we ever need a taxi, we get by for ten bucks.'
- b. *Pis sa fille a accepté fait-que [fɛk] on est parti ensemble.* (HOMA_002F52)
'And his daughter agreed so we left together.'
- 33 An apparent-time variationist analysis demonstrates that the choice of variant is socially conditioned; /fɛk/ is favoured by women and youth, while /fak/ is associated with speakers with a low SES. We attribute the success of *(ça) fait que* in this lexical competition to its ability to phonologically encode social variation. The analyses indicate that /fak/ as a vernacular variant has been replaced by the variant /fɛk/.

Instead of a tripartite lexical variable as is generally proposed in the literature, the 2012 apparent-time study demonstrates that there are four variants in competition over time (alors/fak → donc/fak → fɛk/fak). This model provides a better explanation of the puzzling results provided by previous lexical studies.

6.2. Contraction of the preposition *dans*

- 34 While standard French allows P+Det fusion only in the context of the prepositions *de* “of” and *à* “to”, Québec French extends this phenomenon to the prepositions *dans* “in” and *sur* “on”, as shown in (3) and (4).

(3) a. *Oui je me rappelle avoir patiné dans la [dã:] rue.* (HOMA_004M25)

‘I remember skating in the street.’

b. *J’étais bon dans les [dɛ:] dictées.* (HOMA_010M31)

‘I was good at dictation.’

(4) a. *Sur la [sa:] rue Bourbonnière là c’est vraiment des vieux logements.*

(HOMA_016F44)

‘On Bourbonnière Street, it’s really old housing’

b. *On disait que les gens en campagne ils étaient un peu en retard sur les [se:] autres.*

(HOMA_002F52)

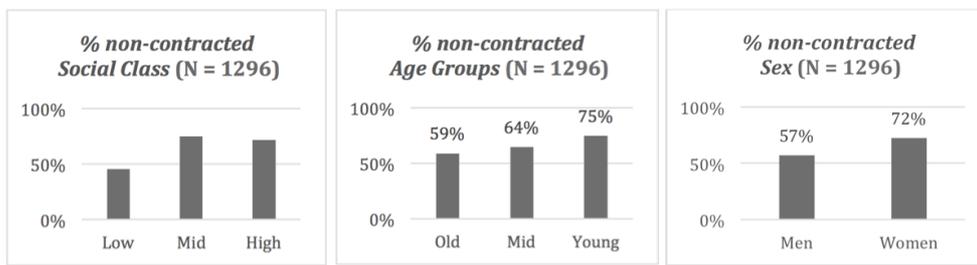
‘We used to say that people in the countryside they were a little behind the others.’

- 35 While many have focused on describing this phenomenon (Baronian 2006, Dumas 1974, Walker 1984), no quantitative study had yet attempted to measure its full extent. Alignment of the audio files with transcriptions allowed us to fill this gap.
- 36 Our data come from a sub-corpus of 38 semi-directed interviews of the corpus. We present the results for the preposition *dans* combined with the determiners *la*, *les*, *un* and *une*. Each P+DET combination was classified as non-contracted or contracted (vernacular) according to its phonological form. In addition, each occurrence was coded according to linguistic (definiteness and gender of the determiners) and extralinguistic (gender, age, socio-economic status) factors.
- 37 Our multivariate analysis was based on 1296 occurrences. Table 2 provides the distribution of the non-contracted variant for the determiners *la*, *les*, *un* and *une*⁷. The overall results show a much higher rate of vernacular forms with the definite determiners *la* and *les* (grammatical gender was not significant).

Table 2. General tendencies (38 speakers)

Definite				Indefinite				
<i>Dans la</i> [dãla]~ [dã:]		<i>Dans les</i> [dãle(z)]~ [dɛ:(z)]		<i>Dans un</i> [dã(z)œ:(n)] ~ [dœ:(n)]		<i>Dans une</i> [dã(z)yn]~ [dyn]		TOTAL
N	%	N	%	N	%	N	%	N
468	48.1	382	53.4	271	86.3	175	94,9	1 296

- 38 Turning now to the social dimension of the variation, the three factor groups considered were significant ($p < .05$). Figure 1 shows the effect of extralinguistic factors on the choice of the non-contracted variant. SES plays the most important role in the variation, followed by age and gender. The non-contracted variant is strongly disfavored by older speakers and those with a low SES; the non-contracted variant is favored by younger speakers and women.

Figure 1. Influence of extralinguistic factors on the use of non-contracted vs. contracted *dans*

- 39 In light of the apparent-time model, the P+DET fusion of the preposition *dans*—although still relatively frequent—would be a regressive phenomenon in Montréal French. The variable is characterized by a regular age distribution and the analysis suggests a shift towards more standard forms; older Montrealers favor vernacular forms, while younger Montrealers disfavor these variants in a shift that appears to be led by women.

Conclusion

- 40 This article introduced the methodological choices that guided the design of the Hochelaga-Maisonneuve corpus, a data set that complements previous sociolinguistic corpora and represents a new milestone for the study of sociolinguistic variation in Montréal French. This corpus collected in 2012 connects in many ways with previous corpora collected in 1971, 1984, and 1995. It is therefore a valuable resource for the microdiachronic study of spoken French, particularly with respect to community trends spanning 1971 to 2012.
- 41 After providing information on previously available sociolinguistic corpora, we discussed the historical and demo-linguistic information that led to the choice of the Hochelaga-Maisonneuve neighborhood. The presentation of the sample showed how social stratification across age, gender and SES ensured comparability with previous Montréal French corpora. Additionally, we provided information on data collection protocols and type of sociolinguistic interview, which are another guarantee of comparability. One notable difference with previous Montréal French corpora is the methodological advances in transcription and annotation. As demonstrated, transcription softwares allowing dynamic annotations and audio-aligned transcriptions provide greater flexibility and precision in the analysis of the dynamics of sociolinguistic variation.
- 42 A study of two sociolinguistic variables involved in community changes in progress highlighted the potential for new analyses. We summarized the real-time analysis on lexical variation in the use of consequence markers between 1971 and 2012 and compared it with a 2012 apparent-time analysis of the sociophonetic variation. In addition, we examined in apparent time the phonomorphological variation in the use of preposition *dans* ‘in’ when followed by a determiner and demonstrated a case of de-vernacularisation.
- 43 In sum, the characteristics of the Hochelaga-Maisonneuve corpus ensure comparability with previous studies and allow for a better understanding of the dynamics of change in real- and apparent-time. Likewise, this corpus allows us to revisit questions surrounding well-studied sociolinguistic variables, and to investigate variables that have not yet been the subject of quantitative studies.

BIBLIOGRAPHY

- Abouda L. & Baude O. (2007). "Constituer et exploiter un grand corpus oral, choix et enjeux théoriques: le cas des ESLO", in *Actes du colloque Corpus en lettres et sciences sociales, Des documents numériques à l'interprétation. Colloque d'Albi, Langages et signification, juin 2006*. Toulouse: Presses universitaires de Toulouse, 161-168.
- Baude O. & Dugua C. (2016). "Les ESLO, du portrait sonore au paysage digital", *Corpus* 15.
- Baronian L. V. (2006). "Preposition Contractions in Quebec French", in P. Saint-Dizier (ed.), *Syntax and Semantics of Prepositions* 29: 27-42.
- Beniak É., Mougeon R. & Valois D. (1985). *Contact des langues et changement linguistique: étude sociolinguistique du français parlé à Welland*. Québec: Centre International de Recherche sur le Bilinguisme.
- Blondeau H. (2001). "Real-time changes in the paradigm of personal pronouns in Montreal French", *Journal of Sociolinguistics* 5(4): 453-474.
- Blondeau H. (2011). *Cet "autres" qui nous distingue Tendances communautaires et parcours individuels dans le système des pronoms en français québécois*. Québec: Presses de l'Université Laval.
- Blondeau H. (2020). "La sociolinguistique au Québec des années 1960 à 2000", in G. Dostie (ed.), *Combats pour la linguistique au Québec. Courants, théories, domaines (1960-2000)*. Bruxelles: Peter Lang, 251-278.
- Blondeau H., Mougeon R. & Tremblay M. (2018). "Variation and change in real time in two French-Canadian communities", Paper presented at Discourse-pragmatic variation and change (DiPVaC) 4, University of Helsinki, June 2018, Helsinki.
- Blondeau H., Mougeon R. & Tremblay M. (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: 35-65.
- Blondeau H. & Tremblay M. (2012). "Social mixing in HOMA: Young urban Francophones and language variation", Paper presented at the Sociolinguistics Symposium 19, Freie Universität Berlin, June 22-24, Berlin.
- Blondeau H. & Tremblay M. (2016). "Le traditionnel et l'émergent. L'apport de jeunes montréalais issus de l'immigration au français vernaculaire", in H. Blondeau & W. Remysen (eds.), Special Issue: Du local au global: pratiques et idéologies linguistiques en contexte montréalais, *Cahiers Internationaux de Sociolinguistique* 10: 19-45.
- Blondeau H., Frenette Y., Martineau F. & Tremblay M. (dir.) (2012). *Sous-corpus variationniste de Hochelaga-Maisonneuve du Corpus FRAN*.
- Bourdieu P. & Boltanski L. (1975). "Le fétichisme de la langue", *Actes de la recherche en sciences sociales* 4: 2-32.
- Buchstaller I. & Wagner S. E. (2018). "Introduction. Using panel data in the sociolinguistic study of variation and change", in S. E. Wagner & I. Buchstaller (eds.), *Panel Studies of variation and change*. New York: Routledge, 1-18.
- Cedergren H. (1988). "The spread of language change: Verifying inferences of linguistic diffusion", in P. H. Lowenberg (ed.), *Language spread and language policy: Issues, implications, and case studies*, GURT 1987. Washington, DC: Georgetown University Press, 45-60.

- Cedergren H. & Sankoff D. (1974). "Variable rules: performance as a statistical reflection of competence", *Language* 50: 333-355.
- Clermont J. & Cedergren H. (1979). "Les R de ma mère sont perdus dans l'air", in P. Thibault (ed.), *Le français parlé. Études sociolinguistiques*. Edmonton, Alberta: Linguistic Research, 13-28.
- Daveluy M. (éd.) (1994). "Thematic section for the tenth anniversary of the Montreal 1984 corpus: Spoken French, a retrospective", *Culture* 14 (2); 15 (1) Montréal: Société canadienne d'anthropologie.
- Dessurault-Dober D. (1974). Étude sociolinguistique de (ça) fait que: "coordonnant logique" et "marqueur d'interaction". M.A. thesis, Montréal: Université du Québec à Montréal.
- Dumas D. (1974). "La fusion vocalique en français québécois", *Montréal Working Papers in Linguistics* 2, Montreal: McGill University, Université de Montréal, Université du Québec à Montréal.
- Labov W. (2006 [1966]). *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Labov W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
- Labov W. (1994). *Principles of Linguistic Change. Volume I: Internal Factors*, Oxford and Malden: Blackwell.
- Linteau P.-A. (1981). *Maisonneuve ou comment des promoteurs fabriquent une ville 1883-1918*. Montréal: Boréal.
- Martineau F. & Mougeon R. (2013). Sous-corpus variationniste de Welland du Corpus FRAN.
- Martineau F. & Séguin M.-C. (2016). "Le corpus FRAN: Réseaux et maillages en Amérique française", *Corpus* 15.
- Rousseau P. & Sankoff D. (1978). "Advances in variable rule methodology", in D. Sankoff (ed.), *Linguistic Variation, Models and methods*, New York: Academic Press, 57-69.
- Sankoff D. & Laberge S. (1978). "The linguistic market and the statistical explanation of variability", in D. Sankoff (ed.), *Linguistic Variation, Models and methods*. New York: Academic Press, 239-250.
- Sankoff D., Sankoff G., Laberge S. & Topham M. (1976). "Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale", *Cahiers de linguistique de l'Université du Québec* 6: 85-125.
- Sankoff G. (2017). "Before there were corpora: The evolution of the Montréal French project as a longitudinal study", in S. Wagner & I. Buchstaller (dir.), *Panel studies of language variation and Change*, New York, NY: Routledge, 21-52.
- Sankoff G. (2019). "Language change across the lifespan: Three trajectory types" *Language* 62(2): 197-229.
- Sankoff G. & Blondeau H. (2007). "Language change across the lifespan: /r/ in Montréal French". *Language* 83(3): 560-588.
- Sankoff G. Blondeau H. & Charity A. (2001). "Individual roles in a real-time change: Montreal R- r 1947-1995", in H. Van de Velde & R. van Hout (eds.), *r-atics: Sociolinguistic, phonetic and phonological characteristics of /r/*, Brussels, Belgique: ILVP, 141-157.

- Tagliamonte S. (2015). *Making waves: The story of variationist sociolinguistics*. Malden, MA: John Wiley & Sons.
- Thibault P. & Daveluy M. (1989). "Quelques traces du passage du temps dans le parler des Montréalais, 1971-1984", *Language Variation and Change*, 1(1): 19-45.
- Thibault P. & Vincent D. (1990). *Un corpus de français parlé: recherches sociolinguistiques*. Québec: Université Laval.
- Trudgill P. (1988). "Norwich revisited: recent changes in an English urban dialect", *English World Wide* 9: 33-49.
- Vincent D. Laforest M. & Martel G. (1995). "Le corpus de Montréal 1995: adaptation de la méthode d'enquête sociolinguistique pour l'analyse conversationnelle", *Dialangue* 6: 29-46.
- Wagner S. E. & Sankoff G. (2011). "Age grading in the Montréal French inflected future", *Language Variation and Change* 23: 275-313.
- Walker D. C. (1984). *The Pronunciation of Canadian French*. Ottawa: University of Ottawa Press.
- Weinreich U., Labov W. & Herzog M. (1968). "Empirical foundations for a theory of language change", in W. Lehmann & Y. Malkiel (eds.), *Directions for historical linguistics: A symposium*. Austin: University of Texas Press.
- Wittenburg P. Brugman H. Russel A. Klassmann A. & VI Sloetjies H. (2006). *ELAN: a Professional Framework for Multimodality Research*. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.

NOTES

1. The sample universe was based on the 188 Census tracts with a minimum of 64% of Francophones.
2. Inspired by Bourdieu's theory of sociosymbolic market dynamics (Bourdieu & Boltanski 1975), this numerical index was developed to serve as an independent variable to establish the relationship between linguistic and social variation. This index, constructed a posteriori, proved to be more revealing than the simple use of social origin (Sankoff D. 1980).
3. Information on the variationist corpus collected in St-Michel-Montréal-Nord (Blondeau et al. 2012) is available in Blondeau & Tremblay (2016).
4. In the borough of Mercier-Hochelaga-Maisonneuve, among the population that declared a single mother tongue, individuals with French as their mother tongue are in the vast majority (79%) (Montréal en statistiques 2011). Moreover, there are more individuals who communicate in French at home than individuals with French as their mother tongue. Therefore, French is gaining relative ground over other languages as the language of communication in the home in this neighborhood. In comparison, 54.3% of the entire Montreal population reported speaking mostly French at home, 25.3% mostly English and the remaining 20.4% mostly other languages. It is worth noting that in the greater Montréal, Francophones represent 69.9% and in the province of Québec 85.2% of the population. Data on knowledge of official languages (specifically the ability to hold a conversation in the official languages) reveal that in Mercier-Hochelaga-Maisonneuve area, of the individuals whose mother tongue is French, 63% can only converse in French, and 37% can converse in both French and English.
5. In the 1971 data collection, most of the speakers interviewed in this area of Montréal originated from Hochelaga (the western part of the neighborhood) and were classified at the lower end of the social ladder.

6. For comparability purposes, these SES categories were based on the Montréal-1984 categorization scheme (Thibault & Vincent 1990).

7. In this analysis, the category “non-contracted” for indefinites includes both [dãz] (with liaison) and [dã] (without liaison). In this category, the variant with liaison was used 87% of the time.

ABSTRACTS

This article introduces the 2012 Montréal FRAN-HOMA corpus, collected in the Francophone neighborhood of Hochelaga-Maisonneuve, and how it relates to the heritage corpora of Montréal French collected since the 1970s. We discuss the methodological choices made regarding the composition of this corpus including the historical and demo-linguistic information that led to the selection of the Hochelaga-Maisonneuve neighborhood for fieldwork. A presentation of the socially stratified sample and criteria for participant selection is followed by a discussion on data collection and the sociolinguistic interview. After a brief review of the transcription protocol and the advantages of audio-aligned transcription, we provide two illustrations of the revival of research on linguistic variation made possible by this new corpus.

Cet article présente le corpus FRAN-HOMA de Montréal 2012 recueilli dans le quartier francophone d'Hochelaga-Maisonneuve, et son rapport avec les corpus patrimoniaux de français montréalais recueillis depuis les années 1970. Nous discutons des choix méthodologiques effectués concernant la composition de ce corpus, y compris les informations historiques et démo-linguistiques qui ont conduit à la sélection d'Hochelaga-Maisonneuve comme site pour le travail de terrain. Une présentation de l'échantillon socialement stratifié et des critères de sélection des participants est suivie d'une discussion sur la collecte des données et l'entretien sociolinguistique. Après un bref exposé sur le protocole de transcription et les avantages de la transcription audio-alignée, nous fournissons deux illustrations de la relance de la recherche sur la variation linguistique rendue possible par ce nouveau corpus.

INDEX

Mots-clés: Variation, Changement, Corpus, Temps apparent, Temps réel, français québécois

Keywords: Variation, Change, Corpus, Apparent Time, Real Time, Québec French

AUTHORS

HÉLÈNE BLONDEAU

University of Florida

MIREILLE TREMBLAY

Université de Montréal

ANNE BERTRAND

University of British Columbia

ELIZABETH MICHEL

Université de Montréal