CAHIER DE RECHERCHE #1805E
Département de science économique
Faculté des sciences sociales
Université d'Ottawa

WORKING PAPER #1805E
Department of Economics
Faculty of Social Sciences
University of Ottawa

Linguistic Distance, Languages of Work and Wages of Immigrants in Montreal[*]

Ibrahim Bousmah[†], Gilles Grenier[‡], David Gray[§]

May 2018

[*] We thank José Galdo, Louis-Philippe Morin and Jean-François Tremblay for comments on an earlier version. We also thank participants at the colloquium Language Skills for Economic and Social Inclusion, Berlin, where the paper was presented in October 2017.
[†] Department of Economics, University of Ottawa, 120 University Private, Ottawa, Ontario, Canada, K1N 6N5; e-mail: ibous086@uottawa.ca
[‡] Department of Economics, University of Ottawa, 120 University Private, Ottawa, Ontario, Canada, K1N 6N5; e-mail: Gilles.Grenier@uottawa.ca
[§] Department of Economics, University of Ottawa, 120 University Private, Ottawa, Ontario, Canada, K1N 6N5; e-mail: David.Gray@uottawa.ca

Abstract

We use the Levenshtein linguistic distance measure to explore whether the distance between an immigrant's mother tongue and a Canadian official language (English or French) has an impact on his/her economic integration into the labour market. Using microdata from the master files of the 2001 and 2006 Canadian censuses and from the 2011 National Household Survey, we investigate the relationship between linguistic distance and the intensity of use of English and French at work in the Montreal metropolitan area. That region is characterized by the presence of sizeable French and English speaking communities, as well as of a large number of immigrants from a wide variety of linguistic backgrounds. Those elements of linguistic diversity interact in the context of English being the lingua franca. We find that linguistic distances between immigrants' mother tongues and English and French have an important impact on the relative intensities of use of the two Canadian official languages at work. We further investigate the role of the languages used at work on the earnings of immigrants by estimating earnings functions. We find that the use of both French and English are remunerated in the labour market, but that using English at work has a larger impact on earnings.

Key words: Linguistic distance, language of work, immigrants, Montreal, Canada, earnings

JEL Classification: C21, C25, F66, J01, J15, J31

Résumé

Nous utilisons la mesure de distance linguistique de Levenshtein pour évaluer si la distance entre la langue maternelle d'un(e) immigrant(e) et l'une des langues officielles du Canada (anglais ou français) a un impact sur son intégration au marché du travail. Avec les microdonnées des fichiers confidentiels des recensements de 2001 et de 2006, de même que de l'Enquête na- tionale auprès des ménages de 2011, nous étudions la relation entre la distance linguistique et l'intensité de l'utilisation de l'anglais et du français au travail dans la région métropolitaine de Montréal. Cette région se caractérise par la présence d'importantes communautés de langues françaises et anglaises, de même que de nombreux immigrants parlant une grande variété de langues. Ces éléments de diversité linguistique interagissent dans un contexte où l'anglais est la lingua franca. Nous trouvons que la distance linguistique entre la langue maternelle des immigrants et l'anglais et le français a une incidence sur l'intensité relative de l'utilisation des deux langues officielles canadiennes au travail. Nous étudions également le rôle des langues utilisées au travail sur les gains des immigrants avec des régressions salariales. Nous trouvons que le français et l'anglais sont utiles sur le marché du travail, mais que l'utilisation de l'anglais au travail a un impact plus important sur les gains que l'utilisation du français.

Mots clés: distance linguistique, langue de travail, immigrants, Montréal, Canada, gains

Classification JEL: C21, C25, F66, J01, J15, J31

1 Introduction

The Canadian labour market has performed well by international standards over much of the 21st century, exhibiting a relatively high employment-to-population ratio, a relatively high labour force participation rate, and a relatively low long-term unemployment rate. Despite a favourable economic profile relative to many OECD countries, however, Canada is also facing important future challenges, such as an aging population, a fairly low fertility rate, higher health care spending, labour market shortages for certain occupations, affordable housing shortages, and the erosion of competiveness in international trade. One of the remedies to reduce the impact of some of those challenges is to rely on immigration, which hopefully benefits the host country and contributes positively to the economy by augmenting the total supply of workers, increasing productivity, creating new businesses and jobs, improving innovation performance, and improving fiscal balances (OECD (2016)).

Having mentioned the benefits of immigration, one must also address the potential costs. Indeed, it is important to take into account the fact that many of those newcomers are ethnically, linguistically and culturally different from the host population, and that without adequate social and economic support from the receiving country, the process of integration can be costly for both the immigrants and the host country. It is therefore important to investigate different aspects of the economic integration of immigrants into the Canadian labour market and the associated adjustment costs. This study will focus on the linguistic aspect of immigrants' integration and its effect on their economic well-being. Immigrants' capacity to communicate in either one of Canada's official languages is essential in order to succeed socially and economically, as linguistic barriers constitute one of immigrants' principal barriers to integration. To the extent that they are not addressed, immigrants are expected to experience lower earnings and productivity profiles. It is therefore relevant to conduct research on linguistic characteristics of immigrants in the Canadian labour market.

The metropolitan area of Montreal is an interesting case for research purposes, as it is

characterized by the presence of important French-and English-speaking communities, as well as by a large number of immigrants from a wide variety of linguistic backgrounds. Unlike most other destination areas, where only one major language is used, immigrants in Montreal are exposed to two languages that are commonly used in the labour market, and they usually have to make choices between the two. Using microdata from the master files of the 2001 and 2006 Canadian censuses and from the 2011 National Household Survey, we investigate the impact of the linguistic distances between an immigrant's mother tongue and both English and French on the intensity of use of those languages at work in the Montreal metropolitan area. We also explore the role of the languages used at work in affecting the earnings of immigrants. As a proxy for linguistic distance, we employ an index developed by the Max Planck Institute for Evolutionary Anthropology, which is derived from an algorithm based on the Levenshtein distance.[1] We find that linguistic distances between immigrants' mother tongues and English and French have an important impact on the relative intensities of use of the two Canadian official languages at work. We also find that immigrants in the greater Montreal gain a higher payoff from using English at work than from using French.

This paper is structured as follows. The second section consists of a review of selected studies in the existing literature regarding the effect of language skills on the labour market in Canada. The third section lays out the conceptual framework. The fourth section presents some general background information and a description of the dataset. It also explains the sample restrictions that are imposed and presents some descriptive statistics. The fifth section presents the econometric model and the various specifications. The sixth section contains the discussion of the results of the empirical work, and the last section contains the conclusion.

[1] We describe how the index is calculated in section 4.

2 Previous Literature

Many studies have focused on the effect of language skills and characteristics on labour market outcomes. The traditional line of research focused on trying to estimate the wage premium received from knowing or using an official language. A more recent line of research upon which we draw involves the introduction of the concept of linguistic distance that is accompanied by a quantitative measure of the distance between languages.

Vaillancourt (1980), which is one of the earliest influential studies, investigates the role of the languages known by an individual in explaining earnings in the labour market of the province of Quebec. Drawing on the public use microdata file of the 1971 Census of Canada, Vaillancourt excluded from his estimating sample women, individuals who are neither Francophones nor Anglophones, non-whites, and those without positive earnings in 1970. The estimation method employed is a semi-logarithmic earnings equation that includes a set of four dichotomous language variables as the key independent variables: Unilingual Anglophone, Bilingual Anglophone, Unilingual Francophone, and Bilingual Francophone. That scheme for categorizing the language groups was adopted in many subsequent studies. The main empirical findings are that knowing English brings higher earnings to males in Quebec City and in Montreal, but that bilingualism brings no monetary returns to Anglophones. In contrast, bilingualism does afford wage premiums of ten per cent to Francophones in Quebec City and six per cent in Montreal. That paper is one of the first to show evidence of the importance of possessing English skills in the predominantly French-speaking province of Quebec.

Carliner (1981) analyzes wage differences amongst language groups in the French-speaking and English-speaking parts of Canada, specifically examining the supply and demand for language skills and the factors that can influence them. The data source is the 1971 Canadian Census. One interesting result relates to the link between education and language skills; unilingual Anglophones have 2.7 more years of education on average than the unilingual

Franophones, and have 0.2 more years than the bilingual Francophones. In addition, one of the findings regarding Montreal indicates that the group whose native tongue is not an official language receives substantial economic rewards for learning French or English. Furthermore, unilingual Francophones in this region earn less than the unilingual English workers, suggesting that the possession of English skills for a native French speaker in Montreal is more beneficial than is the case for English native speakers possessing French skills. In contrast, the results for the English-speaking part of Canada show that the bilingual English workers are the group earning the highest income, followed closely by the unilingual English speakers. Carliner (1981) thus concludes that in English Canada, the economic rewards for learning French are lower than they are in the province of Quebec, where learning English as a second official language leads to significantly higher wages.

Shapiro and Stelcner (1997) examine linguistic earnings disparities over twenty years among Francophones, Anglophones, and Allophones situated in Quebec. They draw on data from the 1991 Census and compare their results with those obtained from previous studies from 1971 and 1981 census data. Restricting their samples to full-time workers aged 18 to 65 years, they estimate earnings regressions including controls for human capital, personal characteristics, labour supply patterns, and indicators for type of occupation and industry. The main findings suggest that since the implementation of Bill 101 in 1977, the earnings gap between bilingual and unilingual Anglophones and bilingual Francophones has decreased, but that the relative earnings of Allophones and unilingual Francophones have fallen. With unilingual Anglophone as a reference group in all their regressions, they found that the earnings disadvantage for unilingual Francophone men increased to approximately 8 to 9 percent in 1990, compared to a range from 5 to 7 percent in 1980. The results for Allophones show that men who speak only French had a wage penalty of about 18 percent in 1990 compared to approximately 15 percent in 1980. Allophone men who speak only English had a penalty of 13 to 15 percent in 1990, up from 8 to 10 percent in 1980, and the earnings disadvantage experienced by bilingual Allophones was 6 to 8 percent in 1990 compared to

4-6 percent a decade earlier.

Albouy (2008) focuses on the wage gap between Francophones and Anglophones in Canada from 1970 to 2000 based on a sample of 20 to 59 year-old males born in Canada who speak French or English fluently. The data sources are the Canadian censuses of 1971, 1981, 1986, 1991, 1996 and 2001. The results suggest that from 1970 to 2000, the wage premium for the Quebecois francophones for learning English fell from 11 percent to 8 percent. In the case of Anglophones in Quebec, the returns to learning French increased from 0 to 5 percent. The author suggests that this evolution is the result of the large number of laws and regulations implemented in Quebec. When examining the English-speaking part of Canada, Albouy (2008) did not find significant returns for Anglophones to acquiring French skills.

A common feature of the literature cited above is the focus of the attribute of knowledge of a language as opposed to a focus on the premium of using French or English in the workplace. This emphasis was mainly due to the fact that data on the language used at work were not collected before the 2001 Census. This relatively new variable is used by Li and Dong (2007), Christofides and Swidinsky (2010) and Grenier and Nadeau (2016). Li and Dong (2007) use the language-of-work variable, but in a different context than the preceding studies. They employ it as a proxy for participation in an enclave labour market, defined as a geographic area with a high ethnic concentration. The main purpose of their paper is to test if Chinese immigrants in the Canadian labour market who participate in the enclave economy earn a lower return than their counterparts who are working in the mainstream economy. The authors examine this wage gap and find (as expected) that earnings are substantially lower for immigrants in the enclave economy. They explain this key finding by the difference in the types of jobs concentrated in the enclave and by the fact that Chinese immigrants in the enclave economy are more likely not to speak the official languages.

Christofides and Swidinsky (2010) estimate the earnings advantage acquired by learning a second official language in Canada and using it at work, focussing on these effects for Francophones in Quebec and Anglophones in the rest of Canada. Their data source is the

individual file of the 2001 Census Public Use Microdata File. They omit immigrants from their sample and include only Canadian-born individuals of 15-64 years of age who worked full-time, full-year and had at least a high school certificate. Their results reveal that in Canada outside Quebec, the wages of men who are English Anglophones are 3.8 percent lower than those of bilingual men who work only in English, 5.4 percent lower than those of bilingual men who frequently use French in the workplace, and 8.4 percent higher than the small number of bilingual men who work either equally, mostly, or exclusively in French. In the case of Anglophone women, the outcomes are quite similar, but with a few differences. Compared to unilingual, English women, bilingual women who use French frequently at work earn 9.3 percent more, and the women who are fluent in French but use only English at work earn a premium of 6 percent.

In comparison, Christofides and Swidinsky (2010) findings for the province of Quebec show that bilingual men who use only French at work earn a premium of 7 percent relative to the unilingual Francophones. Moreover, bilingual Quebec residents who regularly use English at work earn 20.9 percent more than their unilingual counterparts. These results again confirm that knowledge and use of the English language is an important factor for success in the Canadian labour market irrespective of the official language of the province.

Grenier and Nadeau (2016) focus on the effects of the languages used at work on wages in the metropolitan area of Montreal while treating the mother tongue as the principal variable of interest, which is divided into three categories: French, English, and others. Their specification allows for the estimation of three distinct effects on wages of working in a second language: i) the effect of using French at work for anglophones, ii) the effect of using English at work for francophones, and iii) the effect of using either official language at work for those with other mother tongues. Their data source is the 2006 Census master file. They find that Anglophones do not benefit from the use of French at work. By comparison, however, when considering francophones and the group with other mother tongues, they discern an important positive effect for the usage of English in their workplace. A further interesting

finding is that French-speaking immigrants increase their use of English at work as they spend more time in Canada, whereas immigrants in the other-mother-tongues group seem to use English at first and then to move gradually to using French. Overall, their findings indicate that knowledge and command of English is an important element for success in the Montreal labour market, and that financial incentives are an important determinant of the decision to learn English for Francophones and for members of the other-mother-tongues group.

More recently, an emerging literature has developed a new approach that introduces the concept of linguistic distance into the process that involves estimating the effects of linguistic differences on the labour market outcomes of immigrants. This variable is accompanied by a quantitative measure of the distance between languages. We will briefly describe the three most popular techniques employed to measure these linguistic distances. For a more detailed explanation of those techniques, one can consult Ginsburgh and Weber (2016).

Several measures of linguistic distances have been proposed. One popular metric is based on the Levenshtein distance, which is an algorithm that measures the distance between words in two different languages that have the same meaning by determining the minimum number of transformations required to convert the word from one language into the corresponding word in the other language.

Isphording and Otten (2013) apply the Automated Similarity Judgment Program (ASJP), a measure based on the Levenshtein distance and developed by the Max Planck Institute for Evolutionary Anthropology, to analyze the economic success of immigrants in the labour market of some host countries. In order to carry out an international comparison, they use data from the 2000 U.S Census, the German Socio-Economic Panel, and the National Immigrant Survey of Spain. They estimate a probit regression whose dependent variable is equal to one if the individual reported having "good" or " very good" language abilities in the host country language. The key independent variable is the linguistic distance between the immigrant's native language and the language of the host country, which is entered in

the specifications as a percentile measure in order to be able to compare the results with the test-scores approach (described below). The other control variables are those typically included in the immigration literature, such as age at migration, years since migration, years of education, marital status, and number of children. In some specifications they also include the geographical distance between national capitals in kilometers as a proxy for migration costs. For all specifications their results show a significant negative effect of immigrants' linguistic distance on the probability of reporting either good or very good ability in the host country language. In the U.S., for example, an individual situated in the first percentile of the Levenshtein distance distribution has an increased probability of 20 percentage points of reporting good or very good linguistic skills compared to another individual coming from the highest percentile. In Germany and Spain, these increases in probability reach 40 and 20 percentage points, respectively.

Gunduz (2017) employs the Levenshtein distance measure in order to investigate the degree of immigrant-native substitutability in production for Canada by applying methodologies proposed by Borjas (2003) and Ottaviano and Peri (2012). Immigrants' skill types are allowed to vary by language skills, for which linguistic difference serves as the proxy. She divides her sample into three principal groups (low, medium and high) based on the value of their linguistic distance from English or French. Her results show that when the analysis is conducted separately by language groups, the estimates for low-language-skill immigrants suggest an imperfect degree of immigrant-native substitutability, and the estimates for medium and high-language-skill immigrants suggest a perfect degree of substitutability.

Another measure of linguistic distances is the method of learning scores, which consists of following over time a group of people who learn a language, and then measuring their progress at the end of some learning period. This measure is used by Chiswick and Miller (2005), who developed a quantitative measure of the distance between English and other languages based on the difficulty that a typical American would have in learning those languages within a fixed period of time. It was based on a standardized proficiency test

developed by the U.S. State Department School of Language Studies that teaches foreign languages to English-speaking Americans in preparation for placement as diplomats. A test was designed and conducted on which a higher or lower score indicates a lower or higher linguistic distance from English. A limitation of this index is that it only considers the distance between English and other languages, which poses a problem when applying it to analyze the Canadian labour market, a task for which one also needs a measure of the distance between French and other languages.

A further interesting distance technique that has been used is one based on linguistic trees, which computes these distances by dividing language into branches as represented by a tree diagram. The closer the branches are from one language to another, the lesser the distance between them. Adserà and Pytlikova (2015) investigate how linguistic distance influences migration choices by using this particular measure as a proxy for how easy or difficult it is to learn the language of the destination country. They created an index ranging from 0 to 1 in value that reflects the total number of levels of the linguistic family tree that the languages of the destination and the source country share in common. The estimating sample was on immigrants in 30 OECD destinations coming from 223 source countries during the years 1980-2009. Their results suggest that migration rates tend to increase with linguistic proximity, and that it is more important for migrants moving to non-English-speaking destinations than to English-speaking countries. They conjecture that this pattern could be due to the likely higher English proficiency level of the average migrant compared to proficiency levels in other languages.

In a recent study, Adserà and Ferrer (2015) contributed to the analysis of the labour market integration of Canadian immigrants. They used the confidential micro-data of the Canadian censuses of 1991, 1996, 2001 and 2006. Their sample is restricted to males aged 18 to 60 years with the exclusion of aboriginals. They selected a 25 percent random sample of Canadian-born individuals from each census plus all immigrants who arrived in Canada at age 18 or older. Using the measure of linguistic tree proximity of the immigrant mother

tongue to the host country language developed in Adserà and Pytlikova (2015), they examined the role that it plays in the labour market performance of immigrants compared to their native-born counterparts. They estimate a Mincerian earnings equation that is augmented with indicators for job skill requirements and for linguistic proximity. Their results show that immigrants with closer linguistic proximity receive higher weekly wages than those with more distant proximity. In fact, immigrants whose languages share no branch with either English or French earn weekly wages that are 32 percent lower than otherwise similar Canadian-born workers. Another interesting result is that immigrants whose mother tongue is close to French and who are settling in Quebec have similar or better labour market outcomes than immigrants whose mother tongue is close to English, and who are settling outside Quebec.

Our work expands on the above-cited literature by first investigating the effect of immigrants' linguistic distances on the choice of the language utilised at work. Specifically, the outcome variables are the relative intensities of using each of the two official languages given the worker's host language. Furthermore, this research aims to compare the economic returns of using French and English at work for immigrants in the Montreal metropolitan area based on Wage equations. This facet of our paper extends the work of Grenier and Nadeau (2016) by including the linguistic distance variable.

3 Conceptual framwework

Learning a foreign language (and using it at work) is an investment in human capital that depends on the benefits and costs (Grenier and Nadeau (2016); Chiswick and Miller (2015)). The benefits of obtaining this skill can be directly related to the wage premium received from applying it, and the costs are related to the difficulty of learning that language. The linguistic distance from an immigrant mother tongue and an official language is an important element of the cost of learning and using the languages. The first hypothesis that we will attempt to test in our empirical analysis is that immigrants whose mother tongue is close to French/English

will have relatively low difficulty in learning and eventually using French/English in the workplace.

The second component of our empirical analysis involves the economic returns to using these languages at work. Conditional on learning the language, immigrants are expected to use it at work, and the choice of using French, English or both should have a direct impact on their earnings. In a multilingual community such as Montreal, one language might yield a higher economic return than another, an effect which could be represented by the supply and demand framework. On the supply side, there are those workers who are able and willing to use at least one of the official languages at work, and on the demand side there are the employers who are looking to fill positions requiring these skills. A decrease in supply and/or an increase in demand of labour will ceteris paribus raise wages; and a decrease in demand and/or an increase in supply of labour will "ceteris paribus" reduce wages. As an illustration, consider a group of workers called "X" who can only use French at work, and another group of workers called "Y" who can only use English at work. Other things held constant, if a change such as the implementation of a new immigration policy increases the supply of workers in group X and reduces the supply of workers in group Y, we would expect a decrease in the wage rate for group X and an increase in the wage rate of group Y.

In addition to direct impacts on worker productivity in the local labour market, wage differentials among language groups could be generated related to the relative values of languages in the world. In the twenty-first century, English is perceived across the world as the lingua franca. With the increasing globalization of the world's economy and the intercommunication that it involves, English is now commonly used as the principal vehicle of communication for facilitating international trade, business relations, scientific research, and tourism interactions. This special status and value of the English language over French could also be a potential explanation of any estimated difference in economic returns to their utilization in the Montreal area. Our econometric specifications do not allow us to determine specifically the relative explanatory powers of these two behavioral mechanisms.

4 Data and summary statistics

4.1 Data and main indicators

The data sets used for this paper are the micro-data master files of the 2001 and 2006 Canadian Censuses and the 2011 National Household Survey (NHS) from Statistics Canada. Censuses prior to 2001 are not exploited in this analysis because the variables on languages used at work are not available. Those databases provide a rich source of information on labour market characteristics, immigrant status, and language features. They also contain detailed information about the economic, social and demographic characteristics of the Canadian population. Another important advantage is the large sample size covering 20 percent of the Canadian population and containing more than 200 variables. To compute the measure of linguistic distance between languages, we use the database of the ASJP developed by linguists at the Max Planck Institute for Evolutionary Anthropology. From these sources of data, two key indicators are employed in the analysis: Linguistic distance and Relative intensity of use of English and French at work.

In regards to the measure of Linguistic distance, we draw on the database of the ASJP developed by the Max Planck Institute for Evolutionary Anthropology, and which is based on the Levenshtein distance. The measure is constructed from a list of words in pairs of languages with similar meanings and from the number of edits that are necessary to transform a word from one language to the other. It is normalized to account for differences in lengths of words and for word lexical similarities resulting from coincidence. The details of the calculations are provided in the appendix. We obtain the linguistic distance between both of Canada's official languages (French and English) and all other foreign mother tongues reported by the respondents in the data sets (about 150 different languages). The linguistic distances from French and English have values ranging from 0 (for the same language) to approximately 104 for the greatest distance between two languages.

In regards to the measures of the relative intensity of use of English and French at work, we borrow an index developed by Grenier and Nadeau (2016). This measure is derived from a main question and sub-question posed in the censuses and in the NHS regarding the languages used at work, which are worded in the following way: 1) In this job, what language did this person use most often? 2) Did this person use any other languages on a regular basis in this job? To those questions the respondent can answer French, English, both, or another language. Since this research focuses on the use of both of the official languages, we remove immigrants who reported using a non-official language at work in either the main question and the sub-question; this restriction left us with approximately 90 percent of our original sample of immigrants. Based on the various combinations of answers to the above two questions, we define the English-to-French intensity (EtoF) index, which assumes values between zero and one-hundred. Relatively low (high) values are associated with a pattern of English (French) predominance. Specifically, the values refer to typical patterns and are assigned as follows:

- 0 if English is used most often and French is never used on a regular basis
- 25 if English is used most often, but French is also used on a regular basis
- 50 if both official languages are equally used most often
- 75 if French is used most often, but English is also used on a regular basis
- 100 if French is used most often and English is never used on a regular basis.

An intuitive interpretation of this index is that it approximates the proportion of time an immigrant uses French at work as opposed to English in the context of the dual linguistic nature of the labour market in the Montreal metropolitan area. Given the somewhat arbitrary assignment of values to this index, a qualitative version that considers only the order of the choices will also be used in this study.

4.2 Summary statistics

We restrict our analysis to immigrants aged between 25 and 65 years old who are employed as full-time, full-year workers (30 hours or more per week, 48 weeks or more per year) and who became landed immigrant after the age of 18. We follow previous studies by focusing on individuals who reported having worked full-time in order to obtain consistent annual earnings estimates that are not affected by workers who work fewer hours.

Table 1 presents the mean values of the key variables. The Consumer Price Index (CPI) has been utilized to convert 2001 and 2006 wages into 2011 constant dollars. The summary statistics indicate that immigrants in our sample earn on average $47,640 a year. Two thirds of them are married, and the average age is 45 years. For education, we divided our sample into five categories based on the highest diploma and degree attained. Immigrants in Montreal have high levels of education, as the proportion without a high school degree is the lowest for any category. Approximately a third of them obtained some postsecondary college education, and two out of five received a university bachelor's degree or a graduate degree. Table 1 also shows that there is a higher proportion of men than women, and that immigrants have been in Canada for an average of 24 years.

A main variable of interest the linguistic distance from English and French. We see that on average immigrants in the greater Montreal area have a mother tongue that is significantly more distant from English (85.9) than from French (71.6). To have a better idea of the distribution of mother tongues in our sample, the bottom of table 1 reports the five most common mother tongues. Besides French and English, Arabic, Spanish and Haitian Creole are the main immigrant mother tongues, and that they are all closer to French than to English. Those five mother tongues account for almost six out of ten immigrants in our total sample.

The other linguistic variable is the English-to-French intensity index defined above. Figure 1 consists of a histogram of the proportion of immigrants corresponding to each level

of intensity, and it shows a fairly high degree of dispersion. The distribution confirms the presence and the importance of both official languages at work in the Montreal metropolitan area. In fact, about half of the workers used both languages at work.

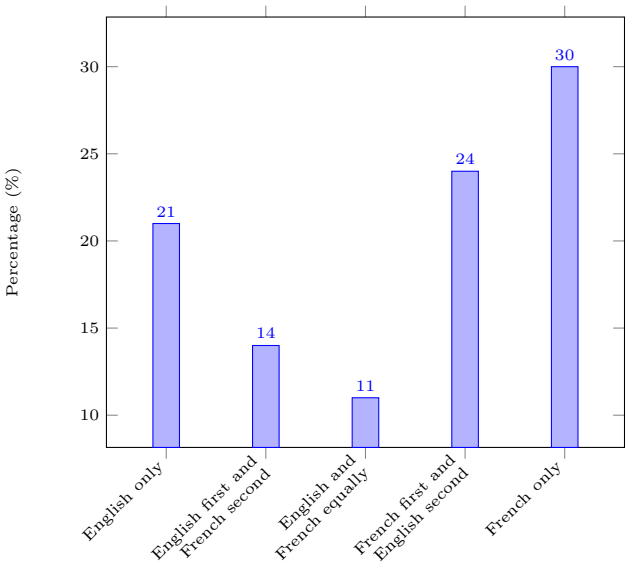


Figure 1: Proportion of immigrants using English and French at work, Montreal CMA, 2001, 2006 and 2011

5 Empirical methodology

5.1 Model specifications

Our model consists of two regression equations, the first of which estimates the effect of linguistic distance and other regressors on the relative intensity of use of the two official languages at work. The second one estimates the economic return of that relative intensity. The model consists of the following equations:

$$EtoF_{i,t} = \beta_0 + \beta_1 LDF_i + \beta_2 LDE_i + X_{i,t}\delta + \lambda_t + [Z_{i,t}\theta + W_i\varsigma] + \varepsilon_{i,t} \quad (1)$$

$$ln(wage)_{i,t} = \rho_0 + \rho_1 EtoF_{i,t} + X_{i,t}\eta + \lambda_t + [Z_{i,t}\phi + W_i\vartheta] + \nu_{i,t} \quad (2)$$

In equation (1), the (limited) dependent variable, $EtoF_{i,t}$,is the measure of the intensity of using French as opposed to English at work for individual (i) in period (t). As explained above, it increases (decreases) with the use of French (English). The two key independent variables for this model are the linguistic distances between the immigrant's mother tongue and French (LDF_i) and English (LDE_i) respectively. The sign of the estimated coefficient of β_1 is expected to be negative, since the further the language is from French, the less likely it is to be used at work. The sign of the estimated coefficient of β_2 is expected to be positive, since the further the language is from English, the less (more) likely that English (French) is used at work. The matrix $X_{i,t}$ contains the socioeconomic variables, which include age, age squared, gender, education, marital status, years since migration, years since migration squared, and an indicator variable for the presence of at least one child. The vector λ_t includes year-specific dummies to control for aggregate trends. An augmented specification includes the matrix $Z_{i,t}$ containing 15 industry-specific dummies, a set of dummies to control for the location of work within the Montreal region (working in the west, center, east and other sub-divisions), and an indicator for immigrants whose previous residence was outside Quebec one or five years earlier.

A potential problem with this regression equation is that the identification of the impact of linguistic barriers on the language use intensity might not be reliable due to the correlation between unobservable cultural differences and the linguistic distance variables (Isphording and Otten (2014)). To address this source of potential bias, we add a specification that includes 240 detailed country-of-birth fixed effects W_i.[2] The additional terms are written in brackets. The last term labeled $\varepsilon_{i,t}$ is the error term, which varies across individuals and time periods.

Equation (2) is a standard human capital earnings equation that is augmented to capture

[2]We examined the correspondence between mother tongues and country of birth, and we found that 89 percent of the countries in our sample have more than one mother tongue reported. Given the absence of a one-to-one correspondence, we also conducted the regression analysis including region-of-birth indicators and additional variables designed to capture cultural differences. The results are pretty similar to those generated from our primary specifications.

the impact of the intensity of using French as opposed to English at work on earnings. The dependent variable is the logarithm of annual wages, and the independent variables are the official language intensity $EtoF_{i,t}$, the socioeconomic controls $X_{i,t}$ and the year dummies λ_t. The additional controls specified as $Z_{i,t}$ and W_i (defined above) are included in some specifications. The term labeled $\nu_{i,t}$ is the error term. The sign of the coefficient estimate of the variable $EtoF_{i,t}$ is our primary focus. If working in French (English) is more highly remunerated in the labour market than is working in English (French), we expect a positive (negative) sign.

One concern with equation (2) is the common endogeneity problem, which in our case can arise from measurement errors or from omitted variables. The intensity index is derived from self-reported responses to a primary question and a sub-question that could reflect reporting errors. Certain individuals could take more or less time answering the sub-question, consequently introducing some unknown influences in the reported value of the index. In addition, similarly to the well-known ability bias that affects estimates of the economic return to schooling, our ordinary least squares estimates might overestimate the value of the true return on earnings of the intensity of using one of the official languages at work. For instance, immigrants with higher levels of innate ability tend to be those who perform better in the labour market and earn higher wages. Without controlling for individual ability, the estimated coefficient of the language-of-work variable could in part capture part this effect. To address this problem, we employ the instrumental variable methodology (IV), which will be discussed in section 6.3.

5.2 Robustness checks

We provide a set of sensitivity analyses to test the robustness of our results. The values of our English-to-French intensity index were arbitrarily assigned to 0, 25, 50, 75 and 100 according to the answers to the two relevant survey questions. Given the ordered nature of the language intensity index, as a first check, we verify the robustness of our results by

estimating an ordered probit model as an alternative specification to equation (1). Let $EtoF^*$ denote the latent variable of the "real intensity of use of French as opposed to English at work", with higher (lower) values representing a higher intensity of use of French (English) at work. This variable is then modelled as a multi-variate function of the same set of regressors as those included in equation (1). This model is expressed as follows:

$$EtoF^* = \alpha_1 LDF_i + \alpha_2 LDE_i + X_{i,t}\tau + \lambda_t + [Z_{i,t}\varpi + W_i\chi] + \xi_{i,t} \tag{3}$$

Where α_1 and α_2 are the coefficients associated with the effects of linguistic distance on the intensity of using an official language at work. The latent dependent variable $EtoF^*$, corresponds to the observed variable $EtoF$, which must be recoded such that the values of the language of work index are 0, 1, 2, 3, and 4. [3] It is assumed that (EtoF* is related to the observable, ordinal $EtoF$ variable as follows:

$EtoF_i = 0 \quad \text{if} \quad \infty < EtoF_i^* < \mu_1 \quad$ (English only)

$EtoF_i = 1 \quad \text{if} \quad \mu_1 < EtoF_i^* < \mu_2 \quad$ (English first and French second)

$EtoF_i = 2 \quad \text{if} \quad \mu_2 < EtoF_i^* < \mu_3 \quad$ (English and French equally)

$EtoF_i = 3 \quad \text{if} \quad \mu_3 < EtoF_i^* < \mu_4 \quad$ (French first and English second)

$EtoF_i = 4 \quad \text{if} \quad \mu_4 < EtoF_i^* < \infty \quad$ (French only)

Where μ_i are threshold parameters ($\mu_1 < \mu_2 < \mu_3 < \mu_4$) to be estimated along with the coefficients of equation (3). After obtaining the parameter values that maximize the likelihood functions, we derive the marginal effects showing the changes in probabilities of immigrants of being at different intensity levels occasioned by one unit changes in the linguistic distance variables.

[3] This mapping is positive and monotonic such that higher (lower) levels represent greater use of French at work and vice versa.

As a second check of the robustness of the results of our first equation, we investigate whether immigrants who have one of the official languages as a mother tongue are the group that drives the findings. We explore this possibility by omitting from our sample immigrants with a linguistic distance of zero from French or English. These individuals account for about 28 percent of our original sample.

For our third robustness check, we verify if our results are sensitive to the specification of the linguistic distance variable by using an alternative indicator, namely a dummy variable that is equal to 0 if the distance between an immigrant's mother tongue is closer to English than to French, and equal to 1 if the distance is closer to French than to English. The motivation behind this specification is the assertion that immigrants will tend to choose the official language which is closer to their mother tongue. This set of regressions also form the first stage of the instrumental variables earnings regression that we will present below.

The last robustness check investigates whether the main findings of our earnings regression are affected by using an alternative measure of the official language intensity variable. We repeat the regression analysis by defining the indexes as a group of indicator variables, allowing us to derive separate estimates for the impact for each group of language use intensity. This is similar to our first robustness check described above, but in this instance the intensity index enters as an independent variable rather than the dependent variable. The results of all of these sensitivity analyses are presented below along with our main findings.

6 The results

6.1 Impact of linguistic distance on the use of official languages at work

The OLS results for the equation modeling the impact of the linguistic distance variables on the official language used at work are presented in table 2. The first three columns display the

results for our full sample, and results in the last column are generated from a sample that excludes immigrants whose mother tongue is English or French (which refers to the second robustness check described above). The specification listed in column (1) includes the basic controls for age, age squared, gender, education, marital status, years since migration, years since migration squared, the presence of children, and year dummies. The specification in column (2) adds controls for the location of work within the census metropolitan area of Montreal, an indicator for the region of residence outside the province of Quebec one year or five years earlier, and industry fixed effects. [4] The results for the fullest specification, which includes the detailed country-of-birth effects, are listed in column (3).

For all of the specifications, the sign of the estimated coefficient for the distance from French variable are negative and highly statistically significant, and thus our key finding is that the relative intensity of using French at work as opposed to English decreases with the linguistic distance from French. Similarly, the results for the estimates related to the linguistic distance from English variable are positive and highly statistically significant, indicating that the relative intensity of using French at work as opposed to English increases with the linguistic distance from English. [5] We note, however, that the coefficients' estimates decline in magnitude in the third specification when we add the detailed country-of-birth effects. This pattern suggests that in the first and second specifications, from which those controls are omitted, these estimates of the impact of linguistic distance captured the effect of cultural differences. Column (4), in which the results based on omitting immigrants with a linguistic distance of zero from one of the official language are listed, shows similar empirical patterns for both linguistic distance variables.

We repeated the analysis by replacing the linear equations with an ordered probit model

[4]The census contains two questions regarding past residences. One is a flag for residence outside of Quebec one year ago or earlier. Another is a flag for residence outside of Quebec five years ago or earlier. We include a single indicator that is the union of these two variables.

[5]In order to address multicollinearity concerns, we calculated the coefficient of correlation between two variables: the linguistic distance between French and a given mother tongue and the linguistic distance between English and the same mother tongue. The value of this correlation is -0.097, indicating that there is a very weak linear relationship between the variables.

(for the first robustness check noted above), and the results are presented in table 3. The estimated coefficients for the linguistic distance variables are all statistically significant and follow the same empirical pattern as the prior model for all four specifications, confirming that immigrants whose mother tongue is more distant from French/English have a lower probability of using French/English at work. Table 4 presents the corresponding marginal effects of a unit change in the linguistic distance from French and English (respectively) on the probability of being in one of the language-at-work intensity levels, holding the value of other covariates constant at their means. For all specifications, we discern similar patterns: a unit increase in the distance from French (English) decreases the probability for the event of the intensity levels of using French (English) more often at work.

We also extended our analysis by estimating the predicted probabilities of the trait of having different intensity levels of using an official language at work corresponding to the range of values of the variables of the distances from French and English. The results are estimated separately by specification and are presented in Figures 2 and 3, respectively, for the distances from French and English. The results from these graphs support the previous findings. For instance, the graphs in Figure 2 show that immigrants with a closer linguistic distance to French have a higher predicted probability of realizing the outcome of using French only at work, and a lower predicted probability of realizing the outcome of not using French at work. [6] The pattern is similar for the effects of the linguistic distance from English on the intensity of using the two official languages at work (Figure 3). Note that the slope of the outcomes profile decreases in the case of our fullest specification that includes the country-of-birth fixed effects. This again shows the importance of controlling for these indicators, and omitting them can overestimate the true impact of the linguistic distance variables.

Table 5 presents the results of the effect on English to French intensity of the linguistic

[6]Moving from left to right, the distance of the native tongue from French increases, the predicted probability of the outcome of using French only at work decreases, and the predicted probability of the outcome of not using French at work increases.

distance variable when it is specified as a single binary regressor (the third robustness check). The variable "Closer to French" assumes a value of 1 when French is closer to the immigrant's mother tongue than English and a value of 0 otherwise. The estimated coefficients presented in table 5 are consistent with prior expectations. The positive coefficient estimates indicate that immigrants whose linguistic distance is closer to French are more likely to use French at work. Again, this effect decreases in magnitude in our fullest specification, but the estimates remain statistically significant.

The results for the socioeconomic control variables are consistent across all of our models (tables 2,3,5). Age has a positive effect on the use of French at work, with the square of age indicating a diminishing effect as age increases. Male immigrants are less likely to use French at work than female immigrants, and the effect of marital status is not consistent across specifications. The positive coefficient for the presence of children is positive in all specifications, suggesting that immigrants with children are more likely to use French. The results for the years- since-migration variable are negative across all specifications but are not statistically significant. The negative signs of the estimated coefficients for the educational indicators are highly statistically significant in all specifications, suggesting that immigrants with higher levels of education are more likely to use English than French at work. A similar result was discerned by Grenier and Nadeau (2016).

6.2 The impact of the languages used at work on earnings

Table 6 presents the regression results of the impact on earnings of the intensity of working in an official language. The coefficient estimates of the English-to-French intensity index are negative in all specifications. They are not statistically significant in the first two specifications, but they are in our fuller specifications (columns 3 and 4) that include countries of birth fixed effects. This interesting result suggests that immigrants in Greater Montreal gain a higher payoff from using English at work than from using French, despite the fact that French is the sole official language of Quebec. For instance, in the third column, the annual

earnings disadvantage associated with a change from 0 to 100 in our English to French index (i.e. from working only in English to working only in French) is estimated at 7.3 percent. The corresponding wage penalties in the specifications whose results are contained in the first two columns are much lower in magnitude. This empirical pattern could suggest that the wage premium for using English in the workplace is partly correlated with immigrants' foreign qualifications, such as the education, experience, and skill sets obtained from their source countries. Holding these characteristics fixed across immigrants from different source countries increases the expected return of using the lingua franca relative to French. Altogether, those results provide evidence that immigrants in the metropolitan area of Montreal gain higher economic returns from using English at work than from using French.

The regression results contained in Table 7 are based on a set of indicators for each level of the index (i.e. the fourth robustness check) for the following categories: English only (reference), English first and French second, English and French equally, French first and English second, and French only. The results are consistent with our previous findings, but they provide further insights. For all the specifications, immigrants using English first and French second (i.e. English approximately 75 percent of the time and French approximately 25 percent of the time) are the groups receiving the highest wage premium, and immigrants who use only French at work are the ones receiving the highest wage penalty. The results of our fullest specification show that immigrants who use English at work most of the time and French occasionally benefit from a wage premium of 4.3 percent compared to immigrants who use only English at work. In contrast, immigrants who use only French at work are the group receiving the largest wage penalties, earning 6 percent less than the reference group that uses only English. Those findings confirm that it pays more to use English at work, but that immigrants who are able to alternate to French enjoy the highest wage premium.

In regards to the impacts of the other socio-economic variables, the results are as expected. For instance, in all our specifications, education has an important effect on earnings; immigrants with higher education levels earn significantly higher wages. Age and years since

migration have positive effects on earnings, while both variables have negative estimated coefficients for the squared variables, indicating a concave pattern. Male immigrants earn significantly higher wages than their female counterparts, and the effect of marital status is positive.

6.3 Results of the instrumental variable estimation of the earnings equation

One important empirical challenge arising from our earnings equation is the potential endogeneity of the language-of-work variable, which could be affected by measurement errors and/or ability bias. To address this problem, we employ the instrumental variable technique. As is well-known, we must select an instrument that is correlated with the language-of-work variable but also uncorrelated with the error term of the wage equation.

Several variables have been used as instruments in the literature (see Chiswick and Miller (2015), for a brief survey). This research has shown that the IV estimates of the effects of language skills on earnings are usually higher than their OLS counterparts, suggesting that an attenuation bias stemming from measurement errors dominates the potential ability bias, which one would expect to work in the other direction. [7] However, these authors also note that in some regression equations estimated by IV, "the coefficient on the instrumented language variable is extremely large, far too large to be believed" (page 242). The large variability of results that are found in the literature are not uncommon for IV estimates.

In our context, candidates for instruments could be the linguistic distance indicators from English and French that we defined earlier. In our analysis presented above, we verified that those linguistic distances from French/English are correlated with the intensity of using French and English at work (i.e. the instrumented variable), which validates the first condition required for identification for the IV estimates.

[7] To the extent that able individuals are more likely to invest in learning a second language, the OLS estimates of the impact of the language-at-work variable would be upwardly biased.

The second required condition is that the instrument (the linguistic distance between an immigrant's mother tongue and a Canadian official language) has no direct impact on earnings. At first glance, this assumption seems to be satisfied, as we expect linguistic distance to impact immigrants' labour market outcome only indirectly through the channel of the language-of-work variable. However, this condition may not hold because of the potential correlation between linguistic distance variables and unobservable differences between the home and the host country characteristics that could remain in the error term. For example, immigrants from countries whose institutional environment and educational system are similar to those of the host country are relatively more likely to perform well in the labour market of the host country. In these circumstances, their measure of linguistic distance would be partially correlated with the error term of the earnings equation. With this possibility in mind, we attempted a series of different regressions using the distance from French and the distance from English as instruments, and we obtained erratic results across specifications. In an attempt to address this challenge, we employ as an instrument our previous indicator, which assumes a value of 1 if the mother tongue of an immigrant is closer to French than to English, and assumes a value of 0 otherwise. By regrouping over the range linguistic distances and collapsing them into a binary variable, we aim to eliminate as much as possible the effect that is specific to the home country that could remain in the error term and could directly impact wages.

These results are presented in Table 8. The coefficient estimates of the English-to- French intensity variable are all statistically significant and negative. The estimated magnitude of the impact on earnings for using French instead of English at work exceeds the magnitude of the one obtained from the OLS regression for all four specifications. Those results show a larger earnings disadvantage for the change from working only in French to working only in English (i.e. the full range), estimated at 66 percent in the specification that includes place-of-birth effects based on the entire sample. Because of the two potential sources of endogeneity bias mentioned above, our prior expectations of the magnitudes of the IV results

compared to the OLS ones are not obvious. If we expect that more able immigrants choose to learn English and work in it for whatever reason, the OLS coefficients estimates would overstate the returns to working in English. Since our IV estimates, however, indicate stronger effects than the OLS estimates, the ability bias hypothesis would apply only if the more able immigrants choose to learn and work in French. Our estimates militate more to the presence of measurement errors causing attenuation bias. This pattern is consistent with the findings of most of the earlier research - e.g., Dustmann and van Soest (2001), Bleakley and Chinn (2004), Ginsburgh and Prieto-Rodriguez (2011).

Several diagnostic statistics are presented at the bottom of Table 8. Because we only have one instrument, an over-identification test is not feasible. To confirm our suspicion that our language-of-work variable is endogenous, we carry out the Wooldridge's score test, which is designed to account for robust standard errors. The resulting test statistics are significant for all of our specifications, confirming that the language of work variable is indeed endogenous.

We investigate the possibility that our instrument is weak by reporting the F-statistic of the first stage regression. The values of the F-statistics across the first, second and third (our preferred) specifications are larger than the conventional benchmark of 10, verifying that the instrument is strong. However, when we omit francophone and anglophone immigrants in our fourth specification, the value of the F-statistic falls below 10. It is not surprising that the instrument loses its strength in this last specification, as those individuals comprise 28 percent of the estimating sample. The range of the instrument and its coefficient of variation are decreased as a result of the omission of individuals whose linguistic distances from one of the official languages are zero. We also see, nonetheless, that in the first stage regression (table 5), the instrument in the fourth specification is still positive and statistically significant despite its lower magnitude compared to other specifications. In summary of our analysis of the earnings equations, the OLS and IV results are qualitatively similar; immigrants in the Greater Montreal area reap a higher payoff of using English at work, even though French is the official language of the province of Quebec.

7 Conclusion

The linguistic dimension of immigrants' integration in the Canadian labour market is an important factor in determining their economic success, and their capacity to use the host country's languages plays an important role. The linguistic dynamics in the Greater Region of Montreal represent an interesting case of an environment where both Canadian official languages are widely used, and where immigrants are an increasingly important component of the labour force. Based on data drawn from the 2001 and 2006 Canadian censuses, from the 2011 National Household Survey, and from the ASJP database, we have investigated the relationships between linguistic distance and the intensity of use of English and French at work by immigrants in the Montreal metropolitan area. We found that the linguistic distances between immigrants' mother tongues and both English and French have an important impact on the language choices at work. We also investigated the role of the languages used at work on the earnings of immigrants. We found that these immigrants gain higher payoffs from using English at work than from using French, and that the higher payoff occurs for jobs where English is used most often and French is used on an occasional basis. The results confirm that the command of the international lingua franca is an important factor in the economic success of immigrants in Canada's largest primary French-speaking city, as reflected in their wage levels are presumably through their marginal products.

In regards to policy repercussions, favouring the selection of immigrants who have a good knowledge of French while also knowing some English could be a suitable strategy to ensure the sustainability of the French language, while at the same time facilitating their assimilation. Over the past decades, various legislations (Bill 63, Bill 22 and Bill 101) helped the province of Quebec promote the preservation of the French language, and many of those laws impacted the integration of immigrants directly (such as the mandatory schooling in

French). Nevertheless, the presence of English as an international language remains an important ingredient for success in the Montreal labour market. The challenge for the province is to find an appropriate good balance between the preservation of the French language and the economic benefits that its citizens can gain by being able to learn and use English.

8 Bibliography

References

Adserà, Alicia and Ana Ferrer (2015). "The effect of linguistic proximity on the occupational assimilation of immigrant men in Canada." CreAM Discussion Paper Series 1517, Centre for Research and Analysis of Migration (CreAM), Department of Economics, University College London.

Adserà, Alicia and Mariola Pytlikova (2015). "The Role of Language in Shaping International Migration." *Economic journal*, **125**(586):F49–F81.

Albouy, David (2008). "The wage gap between Francophones and Anglophones: a Canadian perspective, 1970-2000." *Canadian Journal of Economics*, **41**(4):1211–1238.

Bleakley, Hoyt and Aimee Chinn (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants." *The Review of Economics and Statistics*, **86**(2):481–496.

Borjas, George J (2003). "The labor demand curve is downward sloping: Reexamining the impact of immigration on the labour market." *The Quarterly Journal of Economics*, **118**(4):1335–1374.

Carliner, Geoffrey (1981). "Wage Differences by Language Group and the Market for Language Skills in Canada." *Journal of Human Resources*, **16**(3):384–399.

Chiswick, Barry R. and Paul W. Miller (2005). "Linguistic distance: A quantitative measure of the distance between English and other languages." *Journal of Multilingual and Multicultural Development*, **26**(1):1–11.

Chiswick, Barry R. and Paul W. Miller (2015). "International Migration and the Economics of Language." In Barry R. Chiswick and Paul W. Miller, Handbook of The Economics of

International Migration, Volume 1A, The Immigrants, Elsevier, Oxford, chapter 5, 211-269.

Christofides, Louis N. and Robert Swidinsky (2010). "The Economic Returns to the Knowledge and Use of a Second Official Language: English in Quebec and French in the Rest-of-Canada." *Canadian Public Policy*, **36**(2):137–158.

Dustmann, Christian. and Arthur van Soest (2001). "Language Fluency and Earnings: Estimation with Misclassified Language Indicators." *The Review of Economics and Statistics*, **83**(4):663–674.

Ginsburgh, Victor and Juan Prieto-Rodriguez (2011). "Returns to Foreign Languages of Native Workers in the EU." *Industrial and Labor Relations Review*, **64**(3):599–618.

Ginsburgh, Victor and Shlomo Weber (2016). "Linguistic Distances and Ethnolinguistic Fractionalization and Disenfranchisement Indices." In Victor Ginsburgh and Shlomo Weber, The Palgrave handbook of economics and language, Palgrave Macmillan, New York, chapter 5, 137-173.

Grenier, Gilles and Serge Nadeau (2016). "English as the Lingua Franca and the Economic Value of Other Languages: the Case of the Language of Work in the Montreal Labour Market. " In Michele Gazzola and Bengt-Arne Wickström, The Economics of Language Policy, The MIT Press, Cambridge, Mass, chapter 8, 267-312.

Gunduz, Seda (2017). "The Impact of Immigrant Language Skills on Canadian Wages" Doctoral dissertation, University of Ottawa, retrieved from https://ruor.uottawa.ca/bitstream/10393/36624/1/Gunduz_Seda_2017_thesis.pdf.

Isphording, Ingo E. and Sebastian Otten (2013). "The Costs of Babylon, Linguistic Distance in Applied Economics." *Review of International Economics*, **21**(2):354–369.

Isphording, Ingo E. and Sebastian Otten (2014). "Linguistic barriers in the destination language acquisition of immigrants." *Journal of Economic Behavior Organization*, **1**(105):30–50.

Li, Peter S. and Chunhong Dong (2007). "Earnings of Chinese Immigrants in the Enclave and Mainstream Economy." *Canadian Review of Sociology*, **44**(1):65–99.

OECD (2016). Canada a Economic forecast summary. http://www.oecd.org/canada/canada-economic-forecast-summary.htm.

Ottaviano, Gianmarco I.P. and Giovanni Peri (2012). "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association*, **10**(1):152–197.

Petroni, Filippo. and Maurizio Serva (2010). "Measures of lexical distance between languages." *Physica A: Statistical Mechanics and its Applications*, **389**(11):2280–2283.

Shapiro, Daniel and Morton Stelcner (1997). "Language and earnings in Quebec: trends over twenty years, 1970-1990." *Canadian Public Policy*, **23**(2):115–140.

Swadesh, Morris (1952). "Lexicostatistical Dating of Prehistoric Ethnic Contracts." *Proceedings of the American Philosophical Society*, **96**:121–137.

Vaillancourt, François (1980). "Differences in Earnings by Language Groups in Quebec, 1970: An Economic Analysis." International Center for Research and Bilingualism.

Figure 2: Predicted probabilities generated by the ordered probit model for the events of immigrants being at different intensity levels of using French relative to English as the linguistic distance varies from French (left to right means further from French)

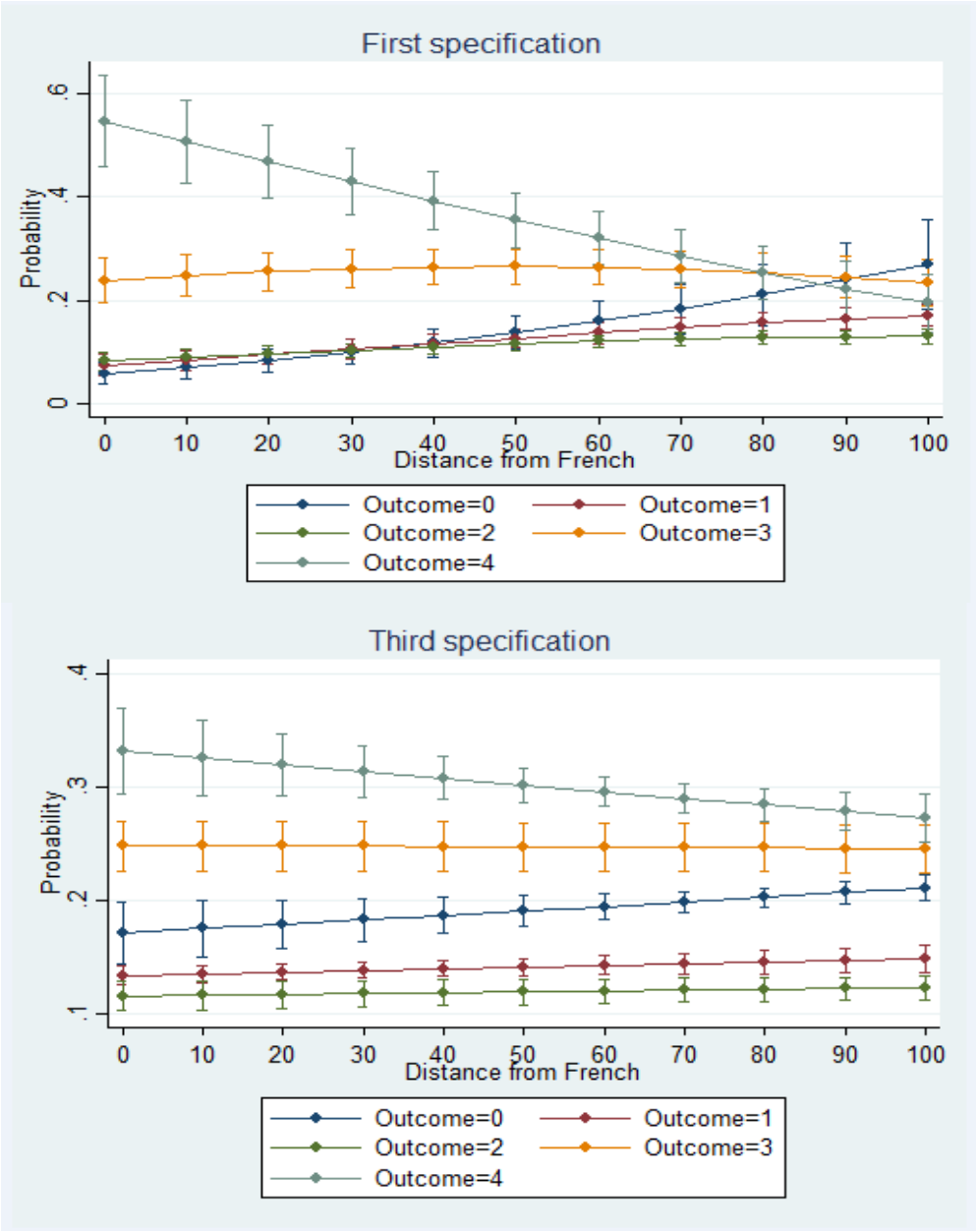
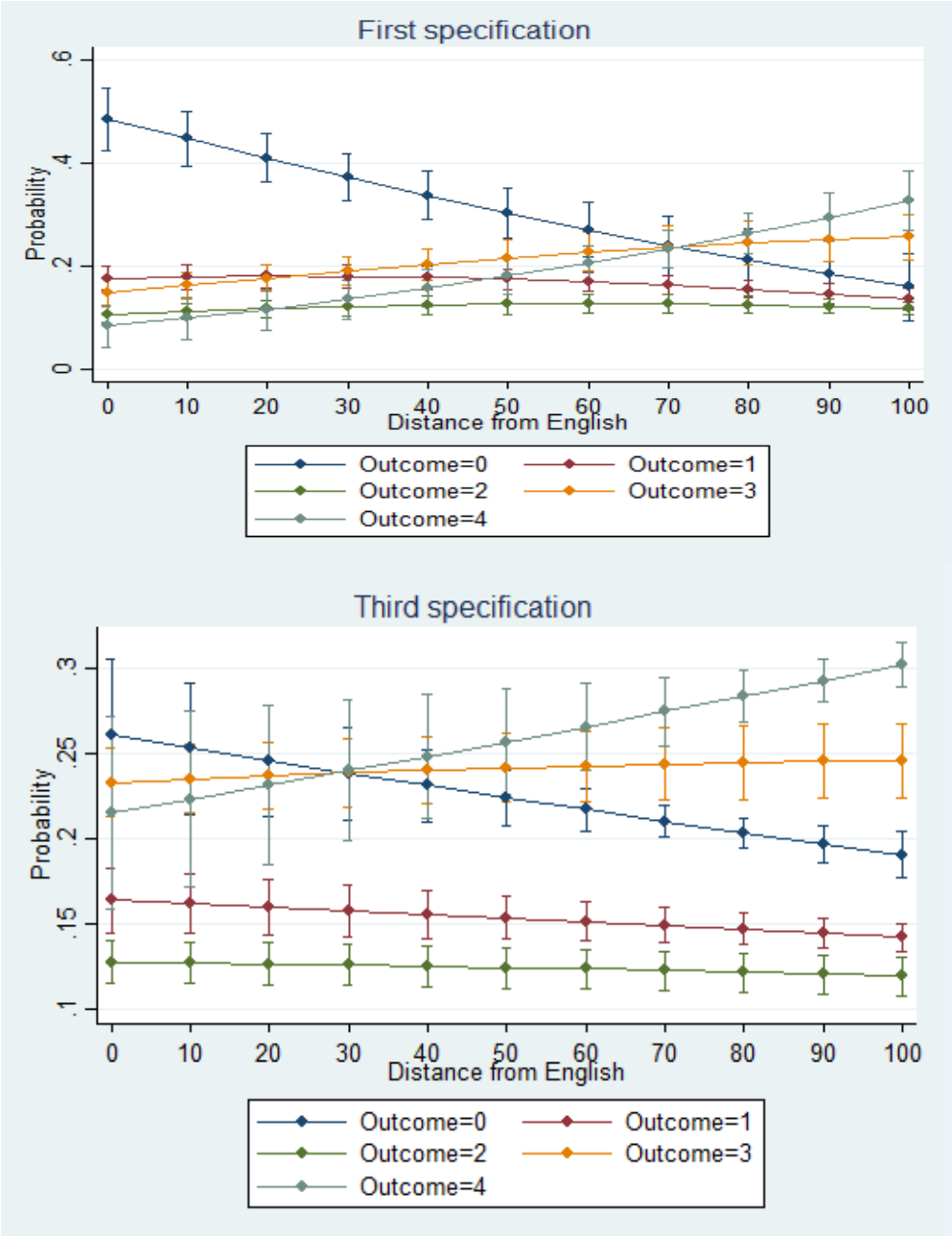


Figure 3: Predicted probabilities generated by the ordered probit model for the events of immigrants being at different intensity levels of using French relative to English as the linguistic distance varies from English (left to right means further from English)



A Tables

Table 1: Descriptive Statistics of major variables, Immigrants in Montreal CMA, 2001, 2006 and 2011

	Means (Std.dev.)
Linguistic distance from French	71.6
	(36.9)
Linguistic distance from English	85.9
	(28.1)
Age	45.2
	(9.7)
Annual wage	47640
	(41382)
Years since migration	24.5
	(11.6)
At least one child	0.65
Married	0.68
Male	0.59
Education:	
High school dropout	0.13
High school degree	0.16
Some postsecondary	0.33
Bachelor degree	0.19
Graduate degree	0.20
Top 5 mother tongue:	
French	0.19
Arabic	0.12
Spanish	0.10
English	0.09
Haitian Creole	0.08
Sample Size	67,440

Note: The Linguistic distance from English ranges from 0 to 104.2; the linguistic distance from French ranges from 0 to 104.0. The linguistic distances have been calculated using version 2.2 of the ASJP distance matrices. Annual earnings are converted to 2011 constant dollars using the CPI. Standard deviations are in parentheses.

Table 2: Ordinary least squares regression of the English-to-French (at work) intensity: Impact of linguistic distance variables

	(1)	(2)	(3)	(4)
Linguistic distance from French	-0.338***	-0.320***	-0.068**	-0.239***
	(0.0587)	(0.0538)	(0.0270)	(0.0904)
Linguistic distance from English	0.337***	0.320***	0.100***	0.295***
	(0.051)	(0.049)	(0.034)	(0.094)
Demographic and socioeconomic controls:				
Male	-2.246	-0.968	-3.163***	-3.876***
	(1.550)	(1.262)	(0.450)	(0.517)
Children	1.454*	1.023	0.877***	1.032***
	(0.838)	(0.713)	(0.320)	(0.356)
Age	1.717***	1.573***	0.763***	0.742***
	(0.413)	(0.401)	(0.186)	(0.218)
Age squared	-0.017***	-0.015***	-0.006***	-0.005**
	(0.004)	(0.004)	(0.002)	(0.002)
Years since migration	-0.053	-0.110	-0.133	-0.193
	(0.370)	(0.367)	(0.133)	(0.164)
Years since migration squared	0.00028	0.00059	0.00064	0.00096
	(0.006)	(0.006)	(0.002)	(0.003)
Married	-2.024	-1.508	0.024	0.671
	(1.366)	(1.254)	(0.422)	(0.471)
High school degree	-8.158***	-7.744***	-2.804***	-3.458***
	(2.471)	(2.325)	(0.986)	(0.920)
Some postsecondary	-4.639	-4.939	-2.725**	-3.085**
	(3.210)	(3.024)	(1.311)	(1.276)
Bachelor degree	-9.442**	-9.453**	-5.712***	-6.008***
	(3.827)	(3.668)	(1.671)	(1.490)
Graduate degree	-12.545***	-12.917***	-8.555***	-8.718***
	(3.905)	(3.583)	(1.632)	(1.654)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Place of Birth fixed effects	NO	NO	YES	YES
Observations	67,440	67,440	67,440	48,050
R-squared	0.202	0.241	0.45	0.414

Note: The dependent variable is the English-to-French language of work intensity. All regressions are weighted using Census weights. Columns (1), (2) and (3) are based on the entire sample. Column (4) excludes immigrants whose mother tongue is either French or English. Robust standard errors in parentheses are clustered at the country-of-birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 3: Ordered probit regression of the English-to-French intensity (at work): Impact of Linguistic Distance Variables

	(1)	(2)	(3)	(4)
Linguistic distance from French	-0.010***	-0.009***	-0.002*	-0.007*
	(0.001)	(0.001)	(0.0009)	(0.003)
Linguistic distance from English	0.010***	0.010***	0.003**	0.010**
	(0.002)	(0.002)	(0.001)	(0.003)
Demographic and socioeconomic controls:				
Male	-0.093	-0.049	-0.139***	-0.165***
	(0.052)	(0.042)	(0.019)	(0.022)
Children	0.054	0.040	0.042**	0.048***
	(0.029)	(0.026)	(0.014)	(0.013)
Age	0.053***	0.049***	0.027***	0.025**
	(0.013)	(0.012)	(0.007)	(0.008)
Age squared	-0.0005***	-0.0004***	-0.0002*	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Years since migration	-0.002	-0.004	-0.005	-0.007
	(0.012)	(0.012)	(0.005)	(0.006)
Years since migration squared	0.00001	0.00002	0.00003	0.00003
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Married	-0.065	-0.050	-0.0002	0.023
	(0.043)	(0.040)	(0.017)	(0.018)
High school degree	-0.281***	-0.272***	-0.133**	-0.159***
	(0.083)	(0.080)	(0.049)	(0.042)
Some postsecondary	-0.175	-0.191	-0.134*	-0.156*
	(0.109)	(0.106)	(0.066)	(0.061)
Bachelor degree	-0.333*	-0.346**	-0.253**	-0.268***
	(0.132)	(0.131)	(0.084)	(0.072)
Graduate degree	-0.440**	-0.473***	-0.375***	-0.375***
	(0.137)	(0.132)	(0.082)	(0.075)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Place of Birth fixed effects	NO	NO	YES	YES
Threshold 1	0.216	0.0025	-0.995***	-1.04*
	(0.305)	(0.315)	(0.178)	(0.436)
Threshold 2	0.710*	0.51	-0.353	-0.445
	(0.310)	(0.316)	(0.181)	(0.434)
Threshold 3	1.054***	0.866**	0.0844	-0.002
	(0.316)	(0.320)	(0.182)	(0.435)
Threshold 4	1.767***	1.603***	0.932***	0.792
	(0.326)	(0.321)	(0.190)	(0.442)
Observations	67440	67440	67440	48050
Pseudo R-squared	0.067	0.083	0.173	0.16

Note: Note: The dependent variable is the English- to-French language at work intensity, and all the regressions are weighted using Census weights. Columns (1), (2) and (3) are based on the entire sample. Column (4) excludes immigrants whose mother tongue is French or English. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 4: Ordered probit marginal effects of the different levels of language intensity (at work): impact of linguistic distance variables

	English Only	English first, French second	English and French equally	French first, English second	French only
Specification 1					
Distance from French	0.00264***	0.00106***	0.00036***	-0.00069***	-0.00337***
	(0.0006)	(0.0001)	(0.0001)	(0.0002)	(0.0005)
Distance from English	-0.00266***	-0.00107***	-0.00036**	0.00070***	0.00339***
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0007)
Specification 2					
Distance from French	0.00250***	0.00108***	0.00037***	-0.00071***	-0.00323***
	(0.0006)	(0.0001)	(0.0001)	(0.0002)	(0.0005)
Distance from English	-0.00254***	-0.00109***	-0.00037**	0.00073***	0.00328***
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0007)
Specification 3					
Distance from French	0.00045**	0.00031**	0.00010**	-0.00023**	-0.00063**
	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0003)
Distance from English	-0.00071***	-0.00049***	-0.00016***	0.00036**	0.00099***
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0004)
Specification 4					
Distance from French	0.00170**	0.00101**	0.00033**	-0.00075**	-0.00229**
	(0.0007)	(0.0004)	(0.0001)	(0.0003)	(0.0010)
Distance from English	-0.00228***	-0.00136***	-0.00044***	0.00101***	0.00307***
	(0.0008)	(0.0005)	(0.0002)	(0.0004)	(0.0010)

The first specification includes control for age, age-squared, marital status, gender, year-since-migration, years-since-migration squared, education, and census year. The second specification adds controls for location of work within Montreal, previous residence outside Quebec 1 or 5 year earlier, and industry-fixed effects. The third specification adds place-of-birth fixed effects, formers colonies indicators and the geographic distance. The fourth specification is similar to the third specification, but native speakers of English or French are omitted. The dependent variable is the English-to-French language of work intensity, and the regressions are weighted using Census weights. Robust standard errors in parentheses are clustered at the country-of-birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 5: Ordinary least squares regression of the English-to-French intensity (at work);linguistic distance measured as dummy variable

	(1)	(2)	(3)	(4)
Closer to French	28.942***	26.982***	9.342***	3.876*
	(5.329)	(5.113)	(2.774)	(2.332)
Demographic and socioeconomic controls:				
Male	-2.483	-1.057	-3.206***	-3.877***
	(1.940)	(1.522)	(0.449)	(0.518)
Children	1.841	1.401	0.878***	1.035***
	(1.287)	(1.081)	(0.318)	(0.358)
Age	1.514***	1.388***	0.742***	0.745***
	(0.524)	(0.507)	(0.183)	(0.219)
Age squared	-0.015***	-0.013***	-0.006***	-0.005**
	(0.005)	(0.005)	(0.002)	(0.002)
Years since migration	-0.362	-0.424	-0.129	-0.191
	(0.364)	(0.369)	(0.131)	(0.164)
Years since migration squared	0.005	0.005	0.0007	0.0009
	(0.006)	(0.006)	(0.002)	(0.003)
Married	-5.811***	-4.957**	-0.070	0.680
	(2.197)	(2.006)	(0.442)	(0.470)
High school degree	-7.652**	-7.422***	-2.659***	-3.429***
	(2.966)	(2.792)	(0.990)	(0.920)
Some postsecondary	-0.891	-1.779	-2.521*	-3.074**
	(3.875)	(3.667)	(1.317)	(1.274)
Bachelor degree	-7.304	-8.163*	-5.561***	-6.006***
	(4.545)	(4.411)	(1.672)	(1.489)
Graduate degree	-7.903	-9.535**	-8.311***	-8.720***
	(5.207)	(4.698)	(1.645)	(1.650)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Place of Birth fixed effects	NO	NO	YES	YES
Observations	67440	67440	67440	48050
R-squared	0.10	0.15	0.449	0.414

The dependent variable is the English-to-French language intensity (at work), and all regressions are weighted using Census weights.. Columns (1), (2) and (3) are based on the entire sample. Column (4) excludes immigrants whose mother tongue is either French or English. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 6: Ordinary least squares regression of the wage equation: Impact of the English-to-French Intensity (at Work) variable

	(1)	(2)	(3)	(4)
English to French intensity	-0.00027	-0.00014	-0.00073***	-0.00064***
	(0.0004)	(0.0004)	(0.0002)	(0.0002)
Demographic and socioeconomic controls:				
Male	0.236***	0.234***	0.231***	0.243***
	(0.015)	(0.013)	(0.012)	(0.013)
Children	-0.0057	-0.004	0.012*	0.009
	(0.009)	(0.009)	(0.007)	(0.008)
Age	0.041***	0.041***	0.043***	0.038***
	(0.004)	(0.004)	(0.005)	(0.004)
Age squared	-0.0004***	-0.0004***	-0.0004***	-0.0004***
	(0.000)	(0.000)	(0.000)	(0.000)
Years since migration	0.019***	0.014***	0.018***	0.019***
	(0.003)	(0.002)	(0.002)	(0.002)
Years since migration squared	-0.00007	-0.00002	-0.0001***	-0.0001**
	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.004	0.003	0.014*	0.015*
	(0.015)	(0.015)	(0.008)	(0.009)
High school degree	0.131***	0.112***	0.096***	0.094***
	(0.016)	(0.014)	(0.010)	(0.013)
Some postsecondary	0.339***	0.286***	0.245***	0.242***
	(0.022)	(0.023)	(0.013)	(0.015)
Bachelor degree	0.607***	0.492***	0.455***	0.440***
	(0.024)	(0.024)	(0.021)	(0.023)
Graduate degree	0.851***	0.699***	0.638***	0.621***
	(0.031)	(0.031)	(0.023)	(0.028)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Place of Birth fixed effects	NO	NO	YES	YES
Observations	67440	67440	67440	48050
R-squared	0.2	0.24	0.27	0.26

Note: The dependent variable is annual earnings converted to 2011 constant dollars (with the CPI), and all the regressions are weighted using Census weights. Robust standard errors in parentheses are clustered at the countries of birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 7: Ordinary least squares regression of wage equation: impact of language intensity (at work) specified as set of categorical variables

	(1)	(2)	(3)	(4)
Official Language used at work (ref: English only)				
English first and French second	0.083***	0.087***	0.043**	0.015
	(0.027)	(0.023)	(0.020)	(0.021)
English and French equally	-0.003	0.014	-0.017	-0.017
	(0.037)	(0.031)	(0.018)	(0.021)
French first and English second	0.03	0.053	-0.011	-0.023
	(0.045)	(0.039)	(0.019)	(0.021)
French only	-0.013	-0.004	-0.060***	-0.058***
	(0.040)	(0.036)	(0.018)	(0.019)
Demographic and socioeconomic controls:				
Male	0.235***	0.232***	0.231***	0.243***
	(0.015)	(0.013)	(0.012)	(0.013)
Children	-0.005	-0.003	0.013*	0.01
	(0.009)	(0.009)	(0.007)	(0.008)
Age	0.042***	0.042***	0.043***	0.039***
	(0.004)	(0.004)	(0.005)	(0.004)
Age squared	- 0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.000)	(0.000)	(0.000)	(0.000)
Years since migration	0.019***	0.014***	0.018***	0.019***
	(0.003)	(0.002)	(0.002)	(0.002)
Years since migration squared	-0.0001	-0.0001	-0.0001***	-0.0001**
	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.006	0.004	0.015*	0.015*
	(0.015)	(0.015)	(0.008)	(0.009)
High school degree	0.126***	0.106***	0.093***	0.092***
	(0.016)	(0.014)	(0.010)	(0.012)
Some postsecondary	0.332***	0.276***	0.239***	0.239***
	(0.022)	(0.022)	(0.014)	(0.015)
Bachelor degree	0.595***	0.477***	0.446***	0.436***
	(0.025)	(0.025)	(0.021)	(0.023)
Graduate degree	0.839***	0.682***	0.628***	0.616***
	(0.029)	(0.029)	(0.023)	(0.028)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Place of Birth fixed effects	NO	NO	YES	YES
Observations	67440	67440	67440	48050
R-squared	0.20	0.24	0.27	0.26

Note: The dependent variable is annual earnings converted to 2011 constant dollars using the CPI. All the regressions are weighted using Census weights. Robust standard errors in parentheses are clustered at the country-of-birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

Table 8: Instrumental variables regression of wage equation: impact of English-to-French (at work) intensity with "closer to French" indicator employed as instrument

	(1)	(2)	(3)	(4)
English to French intensity	-0.0031*	-0.0028*	-0.0066*	-0.0152*
	(0.002)	(0.002)	(0.004)	(0.009)
Demographic and socioeconomic controls:				
Male	0.232***	0.233***	0.212***	0.187***
	(0.0178)	(0.0139)	(0.0139)	(0.0340)
Children	0.0009	0.0005	0.018**	0.024**
	(0.008)	(0.008)	(0.008)	(0.012)
Age	0.045***	0.045***	0.047***	0.049***
	(0.004)	(0.004)	(0.004)	(0.007)
Age squared	-0.0005***	-0.0004***	-0.0005***	-0.0005***
	(0.000)	(0.000)	(0.000)	(0.000)
Years since migration	0.018***	0.014***	0.017***	0.016***
	(0.003)	(0.003)	(0.002)	(0.003)
Years since migration squared	-0.00006	-0.00001	-0.00010***	-0.00010*
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.010	-0.008	0.013	0.025**
	(0.021)	(0.019)	(0.009)	(0.013)
High school degree	0.105***	0.088***	0.080***	0.043
	(0.026)	(0.024)	(0.014)	(0.034)
Some postsecondary	0.332***	0.277***	0.229***	0.197***
	(0.031)	(0.031)	(0.014)	(0.035)
Bachelor degree	0.581***	0.466***	0.421***	0.352***
	(0.033)	(0.032)	(0.027)	(0.058)
Graduate degree	0.821***	0.668***	0.587***	0.493***
	(0.047)	(0.046)	(0.035)	(0.082)
Year fixed effects	YES	YES	YES	YES
Location of work within Montreal control	NO	YES	YES	YES
Previous residence outside Quebec control	NO	YES	YES	YES
Industry fixed effects	NO	YES	YES	YES
Place of Birth fixed effects	NO	NO	YES	YES
IV Test:				
Wooldridge's score test adjusted for clusters	3.74*	3.49*	5.55**	7.71***
F statistic from first stage adjusted for clusters	29.50***	27.85***	11.34***	2.76*
Observations	67440	67440	67440	48050
R-squared	0.18	0.22	0.22	.

Note: The dependent variable is annual earnings converted to 2011 constant dollars with the CPI; all the regressions are weighted using Census weights; Robust standard errors in parentheses are clustered at the country-of-birth level (186 clusters). Significance: *** at 1%, ** at 5%, * at 10%.

B Appendix

Calculation of the Levenshtein Linguistic Distance

The following explanation of the computation of the linguistic distance is based on the work of Petroni and Serva (2010). The ASJP linguistic distance is computed by using a list of 40 words in each language with similar meanings. The list includes, for example, words describing body parts, animals, plants, nature, verbs, adjectives, and pronouns that are used universally across languages. It was originally based on the 100-item Swadesh list (Swadesh (1952)), but was reduced to 40 items that were shown to suffice. To calculate the distances, the lexical similarities of all pairings of languages are compared using an algorithm called the Levenshtein distance (LD), which is calculated as the minimum number of edits (deletions, substitutions or insertions) required to transform a word from one language into another. To provide a very simple illustration, the Levenshtein distance between the French word "allo" to its corresponding English word "hello" is equal to two, the transformation of one word into the other cannot be effectuated with fewer than two edits.

1. allo hllo (substitution of "a" with "h")

2. hllo hello (insert "e" after "h")

A normalized measure of the Levenshtein distance (LDN) needs to be provided in order to account for the word lengths, because longer words inherently require more edits to be executed. The normalization is performed by dividing the LD between similar words in two different languages by the number of characters of the longer of the words in whichever language applies. The LDN between the words with meaning i in languages Q and W is equal to:

$$LDN(Q_i, W_i) = \frac{LD(Q_i, W_i)}{L(Q_i W_i)} \tag{4}$$

Where $LD(Q_i, W_i)$ is the Levenshtein distance between Q_i and W_i and $L(Q_i, W_i)$ is the number of characters of the longer word. The total linguistic distance (involving all words) between a pair of languages is then calculated by measuring the average distance of all n

words for those languages as follows.

$$LDN(Q,W) = \frac{1}{n}\sum_{i=1}^{n} LDN(Q_i, W_i) \tag{5}$$

Where Q_i and W_i correspond to the word i in languages Q and W. Finally, to account for word lexical similarity resulting from merely pure coincidence (as opposed to pure etymology), the program provides a further normalized measure labelled the Levenshtein distance normalized divided (LDND) between pair of languages. It is obtained by dividing $LDN(Q,W)$ the by the "global distance". The "global distance" is the average distance between two languages using only pairs of words with different meanings, which is given by:

$$GD(Q,W) = \frac{1}{n(n-1)}\sum_{i\neq k}^{n} LDN(Q_i, W_k) \tag{6}$$

The LDND is the final measure of linguistic distance, which is obtained by dividing the $LDN(Q,W)$ between pairs of languages with their respective values of $GD(Q,W)$. It is employed in our empirical analysis, and is written as:

$$LDND(Q,W) = \frac{LDN(Q,W)}{GD(Q,W)} \tag{7}$$