FIW – Working Paper

FIW Working Paper N°23
January 2009

Talking Trade: Language Barriers in Intra-Canadian Commerce

Nicolas Sauter

Abstract

This paper tests for one mechanism that can explain the existence of a language barrier to trade. Specifically, I ask if those industries that require more cross-border communication in order to export their products trade more between Canadian provinces that know the other's language(s). I find that trade in industries with a need to communicate directly (orally) with importers increases with the probability that people in another province speak the same language. This finding can fill a missing link in the empirical trade literature, which lacked convincing arguments for the observed correlation between language commonality and the total volume of trade.

JEL classification: F10, F15, F40;
Keywords: Language, Trade cost, Gravity model, Canada;

The author/s

Nicolas Sauter is researcher at the Munich Graduate School of Economics, Germany. E-mail: nicolas.sauter@lrz.uni-muenchen.de

1 Introduction

In the light of falling tariffs and transport cost, the importance of institutional barriers to trade has captured much attention in recent research.[1] The existence of a language barrier in trade has been documented in numerous empirical studies. Rose (2000) finds that countries sharing a common language trade 1.5 times more with each other. Anderson and van Windcoop (2004) estimate that the tax equivalent of the language barrier amounts to seven percent.

While gravity models of aggregate trade flows find robust evidence for the language barrier, these models remain silent on the question of the channel through which language affects trade. From a logical point of view, it is even questionable if language should affect international trade at all, given that international trade flows consist mainly of manufactures. For instance, in order to trade two manufacturing goods between the US and China only one translator is required, whose services are unlikely to affect total trading cost. Also the fact that with China and Japan two countries with relatively few fluent English speakers are among the top five trading nations contradicts the importance of language for manufacturing trade. Services trade, on the other hand, often requires the ability of both the service provider and his customer to communicate directly with each other.

A second shortcoming of the studies mentioned above is their opaque measurement of the language barrier. It is typically represented as either a binary indicator for countries that share a common official language (e.g. Frankel and Rose, 2002), or as the probability that two randomly chosen people from two countries share a common mother tongue (e.g. Melitz, 2008). Alternatively, Hutchinson (2002) and Ku and Zussman (2008) suggest to use the fluency in English - the lingua franca of international trade - as a proxy for the ability of natives from two countries to communicate in a common third language. These measures do indeed all reflect some aspect of the potential language barrier. Yet what is really required for trade is that there is a sufficient number of people in both countries who are proficient in at least one of the other's language(s), irrespective of whether they speak a lingua franca, the same official, native, or second language. Also, the common proxies might take up all other kinds of bilateral institutional similarities, thereby imposing an upward bias on the estimate for the language barrier in gravity models.

This paper wants to provide one way to resolve the missing motivation of the language barrier and to reduce measurement bias of the effect of language on trade. In particular, I test if communication-intensive industries trade more between Canadian provinces with a good knowledge of the other's language(s) compared to those industries that require less communication with the trading partner. Such a finding could justify the alleged role of language as a trade barrier. Though it is less general than conventional gravity models, this

[1] Recent examples are Rauch (2002), Nunn (2007), and Levchenko (2007).

simple approach has two advantages: First, it tests for one specific mechanism through which language affects trade. Second, it corrects for other institutional factors that could bias the estimates via fixed-bilateral effects between Canadian provinces.

Previous work that comes closest to this paper is from Fink et al. (2005), who show that trade is significantly lower between countries with high bilateral international calling prices. They find that the price effect is larger for trade in differentiated products compared to goods that are traded over organized exchanges, which corroborates the hypothesis that trade in communication-intensive goods is more sensitive to deficiencies in direct communication. However, they estimate that halving the importer's calling prices would boost aggregate trade by 42.5%, which seems unreasonably high. Melitz (2008) uses several measures of language commonality to estimate the effect of language on international trade flows. In contrast to this paper, Melitz's variables on language commonality do not measure the knowledge of second languages. He only proxies the probability that two randomly chosen persons from two different countries have the same native language. My measure incorporates the two-sided knowledge of English, French, and Chinese as first or second languages between Canadian provinces. The empirical evidence presented in section 3 indicates that this measure is a better proxy for the language-trade link. Moreover, Melitz focuses on the total volume of trade, not on specific industries. So he cannot attribute the estimated language effects to a channel through which language erects a trade barrier.

My results suggest that commerce in industries that require direct communication for trade increases with the probability that people in another Canadian province speak the same language. I cannot find evidence for an impact of indirect communication via mail on intra-Canadian trade flows.

The remainder of the paper is structured as follows: In section 2 the estimation equation is motivated. Section 3 describes the data. In section 4, the baseline results are discussed. In section 5, I control for potential endogeneity. Section 6 consists of robustness checks. Finally, section 7 wraps up the discussion. A detailed description of the variable labels is provided in the appendix.

2 Empirical Model

While there is strong support for the language barrier in empirical research, hardly any theoretical work has analyzed this issue, probably because it seems self-evident that people can only trade if they are able to communicate with each other. Yet the case for language in trade is not clear-cut: For instance, while rice or oil can be bought at the merchandise exchange without the need to learn any Asian or Arabic languages, a buyer of a laptop in

Quebec will require explanations, software, and support services in French. To see exactly how language can affect trade patterns, imagine the following scenario: There are two regions, whose populations speak different languages. Translation is costly. If some products require more communication between buyer and seller for trade to proceed, translation cost will more adversely affect trade in those products. If more people learn the other's language, total translation cost will fall. Hence, I propose the following hypothesis: ceteris paribus, a high language commonality between two regions should disproportionately help communication-intensive industries to trade.

With respect to the type of communication used, I distinguish between direct or spoken communication and indirect or written communication. I expect direct communication to have a larger effect on the volume of trade than indirect communication: Assume there are two communication-intensive industries, one of which requires spoken communication, whereas the other industry has a large need to exchange written documents in order to trade. While direct communication has to occur simultaneously between two individuals, indirect communication often requires not more than a one-time service like translating catalogs, advertisements, or letter templates. Thus total translation cost will be higher in the case of direct communication. Hence, direct communication-intensive industries are more likely to be affected by the language barrier.

In order to thoroughly test my hypotheses, I diverge from the standard gravity framework in two important ways: First, I study the impact of language differences on trade within one country only, which limits bias of the language estimates from other institutional differences, such as legal, historical or social ties. This downweights the possibility that my language estimates capture some home bias (or border) effects that are well-known to the international trade literature.[2] Second, the estimation equation includes fixed-effects for each bilateral pairing of Canadian provinces and territories as well as fixed-industry effects. Thus the estimates will only pick-up variation between industries and province pairings, thereby taking out heterogeneity that is unrelated to the potential language-trade channel I seek to identify.

This paper focuses on Canada, which is the only OECD country with more than one official language for which detailed inter-regional data on trade flows is available. While this choice limits the scope of the study and the number of potential sources of language variation, it offers at least three advantages. Firstly, the relative uniformity of Canada's legal and social system alleviates institutional bias that is possibly present in studies of international trade. As communication-intensive industries are often contract-intensive as well, estimates from cross-country regressions would be likely to incorporate effects of comparative advantage in regions with sound legal institutions (Nunn, 2007). Secondly, recent research points out that the caveat of taking up border effects is significantly lower in intra-national trade.

[2]See McCullum (1995), Helliwell (1996), Anderson and van Windcoop (2003).

Hummels and Hillberry (2003) showed that intra-US trade is unlikely to suffer from intra-national border effects.[3] Similarly, Combes et al. (2003) estimate that in France more than 60% of the potential intra-national home bias can be explained by internal migration and cultural networks. Such network effects between Canadian provinces and territories are likely to be primarily determined by linguistic differences, since Helliwell (1997) already pointed out that internal migration has little trade creating effect within Canada. Finally, the arguments presented above for the existence of a language-trade channel should be mainly relevant for service-intensive industries. Therefore I refrain from studying intra-European trade (which otherwise would make a perfect case for the language-trade link), because services are not sufficiently liberalized across EU members (e.g. Kox and Lejour, 2005; Kox and Lejour, 2006).

This paper introduces a new way to thinking about the gravity model of trade that rests on the work by Rajan and Zingales (1998), Romalis (2004), and Nunn (2007). These papers use industry- and cross-country-variation to identify sources of a country's comparative advantage across industries. I adapt their approach for a single country setting, where I exploit trade variation across industries and bilateral province pairings. Specifically, I eliminate any variation, which is not needed to test the main hypothesis: trade in communication-intensive industries is higher between provinces with a higher language commonality. The model I estimate is then:

$$\ln trade_{ijk} = \delta_{ij} + \delta_k + \beta_1 \ln(trans_k dist_{ij}) + \beta_2 prod_{ijk} + \beta_3 c_k lang_{ij} + \epsilon_{ijk}, \qquad (1)$$

where $\ln trade_{ijk}$ is the natural logarithm of the bilateral trade flow from province i to province j in industry k. The fixed-bilateral effects δ_{ij} take up all trade variation for each country pairing that is constant across industries. Similarly, δ_k are industry fixed-effects that are constant across bilateral trade flows. Compared to the traditional gravity model, the bilateral fixed-effects do not allow using variables that are constant across country pairings. Thus, the impact of distance on trade is proxied with the log of the interaction between the transport-intensity of a sector, $trans_k$, and distance, $dist_{ij}$. In order to capture differences in comparative advantage, $prod_{ijk} = \frac{production_{ik}}{GDP_i} - \frac{production_{jk}}{GDP_j}$ reflects differences in the structure of production between two provinces. The main variable of interest is the interaction of c_k and $lang_{ij}$, in which c_k reflects the need for communication and $lang_{ij}$ stands for the language commonality between two provinces. ϵ_{ijk} is a random error. As common in the literature (e.g. Anderson and van Windcoop, 2003; Melitz, 2008), I assume that imports and exports are affected symmetrically by the interaction effects.

The approach here is conceptually different from industry-level gravity models that estimate the semi-elasticity of the language commonality with respect to trade (e.g. Deardorff,

[3] Hummels and Hillberry (2003) showed that Wolf's (2000) dataset does not properly account for intra-US trade distances and wholesale trade flows.

1998; Hummels, 2001). The bilateral fixed-effects capture the direct effect of the language commonality on the volume of trade in my estimation equation. Hence, the coefficient of interest β_3 only captures the effect that language commonality has on the pattern of trade and provides no direct interpretation as a semi-elasticity of the language barrier.

The estimates of (1) should not be regarded as conclusive evidence for the language-trade channel. First, there may be determinants of trade that are omitted from (1). As a matter of fact Canada's English speaking provinces tend to be richer and domicile more Protestants than Catholics compared to their French speaking counterparts. Therefore, a primary concern is that $c_k lang_{ij}$ may be simply capturing the fact that wealth and religion shape intra-Canadian trade patterns. I carefully control for these alternative determinants of the language-trade channel. Second, the direction of causality implied by equation (1) may be wrong. If trade fosters the adoption of the other's language, causality might run from trade to language. In consequence, estimates of β_3 may be biased. In section 5, I instrument for language variation that is unaffected by this feedback effect. Finally, this paper concentrates on the analysis of positive exports and imports. Thus the interpretation of the estimates is conditional on a province trading in an industry, thereby disregarding the effect of language on the decision to enter an industry. I check for the sensitivity of my results to the inclusion of zero trade in section 6.

3 The Data

The most disaggregated inter-provincial trade data available for Canada are at the 2-digit industry level. The data comprise all recorded (non-zero) inter-regional trade flows of Canada's ten provinces and three territories for the year 2001.[4] The final data classify in 38 industries that comprise agriculture, manufacturing and service industries. I study trade across all sectors, which is different from other studies that solely focus on manufacturing trade (e.g. Nunn, 2007; Romalis, 2004; Hummels, 2001). I have numerous reasons for this approach: The first reason is that Canada's internal trade differs from international trade, where services trade is negligible compared to manufactures. In 2001, service trade accounted for 56.7% of total intra-Canadian trade flows. The second reason is that the language channel should be present across all sectors of the economy. Particularly, the service sector is likely to be more language-sensitive than manufactures. So leaving out one of the sectors would narrow the scope of this study.

[4]Although industry-level trade data is available from 1997 to 2004, Census data is only available for the years 1996 and 2001. For a discussion of the derivation of inter-provincial trade flows from IO-tables, see Généreux and Langen (2002).

Provincial gross domestic products in current prices as well as population estimates have been retrieved from the Statistics Canada home-page. The distance variable is from Feenstra (2004), who provides distances between the capitals of Canadian provinces. I added distances for each pairing that involves trade with the three territories (Northwest Territory, Nunavut, and Yukon Territory), using the respective longitudes and latitudes.

3.1 Language Variables

In contrast to the language proxies used in previous studies, this paper measures language commonality as the probability that any two people from different provinces picked at random will be able to communicate with each other.[5] Compared to previous measures that reflect some aspect of language commonality, this variable is more in line with theory, because trade requires a sufficient knowledge of the trading partner's language in order to reduce translation cost.

The measure of language commonality between provinces is constructed from the Census survey. The survey asks for mother tongue, knowledge of official languages, and use of languages at work. Table 1 depicts the percentage of speakers of English and French as a mother tongue in Canadian provinces. While English is the dominant mother tongue (59.5%), 22.7% of the Canadian population are native French speakers. French mother tongue speakers, are mainly concentrated in Quebec (81.2% French mother tongue speakers) and New Brunswick (32.9%).

The statistics show clearly that the language barrier cannot be represented by the distribution of mother tongues within the population: 17.6% of all Canadians have a mother tongue different from the two official languages. Yet only 1.5% of all Canadians are unable to speak at least one of the two official languages, which measures the real ability of people to communicate with each other. Across provinces, only the Inuit population in the Nunavut territories constitutes an exception with 13.1% of the population knowing neither English nor French. While in most provinces more than 97% of the population speaks English as first or second language, Quebec (45.4%) and New Brunswick (90.7%) are the two exceptions with relatively few English speakers.[6]

In addition, I control for potential Chinese networks within Canada that have been shown to be relevant for international trade flows (Rauch, 2002). With 2.7%, Canada's Chinese minority supplies the third largest language group of Canada's working population. Different

[5] For an excellent overview of language measures used in previous research, see Melitz (2008).

[6] Unfortunately, the data do not allow to draw explicit conclusions on the fluency of language knowledge. Yet Hutchinson (2002) cannot find a statistically significant difference between speakers of English as a mother tongue or second language, when analyzing the volume of US exports and imports.

from other minorities, 42% of speakers with a $'Chinese'$ mother tongue also use non-official languages often or sometimes at work, as table 2 shows. No other language group uses non-native languages at work so frequently. As I analyze the language knowledge across the total population, I use the population share with Chinese origin as a proxy for knowledge of $'Chinese'$.

For each bilateral pairing ij, the variable for language commonality is constructed as follows :

$$lang_{ij} = \sum_{l=1}^{L} (knowledge^{l})_i (knowledge^{l})_j, \tag{2}$$

where $l = \{English, French, 'Chinese'\}$. $English$ is the sum of people knowing English and people knowing English and French, divided by the total population of the province. The $French$ and $'Chinese'$ measures are constructed similarly. $lang_{ij}$ is not bound at one, since people may be fluent in several languages. However, I restrict the probability that two randomly chosen people are able to communicate with each other to one in cases, where I calculate values slightly larger than one. Based on equation (2), I also construct a measure for religious (denominational) commonality, where l=\{*Anglicans, Baptists, Buddhists, Catholics, Hindus, Jews, Muslims, United church*\}.

Table 3 depicts the resulting language commonality for all bilateral country pairings. The pairings range from Quebec and Nunavut, where the probability that two randomly chosen people understand each other is 43.1%, to Saskatchewan and Prince Edward Island, where everybody speaks the same language. Virtually all variation in language knowledge comes from the two French speaking provinces and the territories, whereas the pairing Ontario/British Columbia exhibits the lowest language commonality (0.962) among the English speaking provinces.

3.2 Communication-intensities of Industries

My argument hinges on a careful choice of c_k, the proxy for differences in the industry-specific need of direct and indirect communication between importer and exporter. Rauch (1999) classifies manufacturing goods on whether they are traded on an exchange, reference priced or neither of both. However, I refrain from using his classification for two reasons. Firstly, it only captures manufacturing goods, yet the services sector is an important pillar of intra-Canadian trade that accounts for much of the language-related variation, as will be seen below. Secondly, it is not possible to extend his classification to services, because services are typically neither reference priced nor traded on exchanges. Nonetheless, it is of particular interest to have a measure that proxies the language-intensity of services trade, because service provision often depends on face-to-face communication with the importer, which bears high translation cost if agents speak different languages.[7]

Therefore, I construct a new measure for the communication-intensity of industries that takes advantage of detailed input output (IO) tables. IO data are available for the manufacturing as well as the service sector and allow me to rank all industries according to their need for communication between trading partners. Thereby I implicitly assume that the input structure of communication services proxies the need for direct and indirect communication between exporter and importer. Given the relatively high level of aggregation of the trade data, all this assumption postulates is that if the printing industry needs a larger share of communication inputs than the paper industry relative to its total inputs, trading printing products also requires more communication for trade. I measure the direct communication-intensity by the share of telecommunications services in total inputs for each industry. The Indirect communication-intensity is measured by the input share of postal services. Since the IO tables at M-level aggregation (2-digit level) exist only for Canada as a whole, I assume that the average Canadian input structure persists across provinces. This strong assumption is less problematic in the Canadian case, where production structures are relatively similar, than in cross-country studies (e.g. Rajan and Zingales, 1998; Nunn, 2007).

The resulting ranking seems reasonable by common sense, as can be seen in table 4. The share of telecommunication inputs ranges from 0.05% (Fishery) to 5.3% (Professional services). As expected, table 4 shows that service industries are more communication-intensive than manufacturing industries. Among manufacturing industries, more complex products are generally ranked higher, which is consistent with the Rauch (1999) classification, where more complex manufactures such as electronic equipment are rarely reference-priced or traded on an exchange. With respect to postal service inputs, the basic pattern persists that services are more communication-intensive that manufactures. Yet there is variation between the relative ranking of industries within each sector.

Finally, industries are ranked according to their relative cost of transportation, presupposing that distance has a larger trade diverting impact on industries with higher transport cost. I calculate the share of transport margins in total inputs, which is defined as the charges paid to a third part in order to deliver a product from the producer to the (intermediate or final) purchaser. The ranking of the transport variable in table 4 shows that services generally have lower transportation cost than manufactures. Particularly heavy industries rank high, e.g. metal, mineral products, chemical, and motor vehicle industries.

[7] Experiments with the Rauch data proved inconclusive. I manually matched classifications and calculated the percentage of goods that are neither reference priced nor traded on public exchanges for each manufacturing industry. However, the estimated effects are only significant if trade flows in the (language-insensitive) petroleum and coal industries are included in the sample. This result supports the idea that translation cost are negligible in Canadian manufacturing trade. Similarly, Rauch (2002) finds no significant trade-deterring effect of language on trade in differentiated manufacturing goods at the international level.

Summary statistics are presented in table 5. As all interactions are obtained from the multiplication of two shares, their actual values are very low. It is therefore impossible to interpret the estimates of β_3 directly.

Table 6 reports cross-correlations of the interaction terms. The high correlation between most variables imposes a potential multicollinearity problem, which could inflate t-statistics. To alleviate such problems, I avoid to lump all variables together in one regression, and run separate regressions for each variable of interest in section 6.

If not otherwise indicated, I drop the fuel as well as the petroleum and coal industries from the sample. As these industries are unlikely to be very language-sensitive, the high trade volumes in both industries could bias the estimates downward.

4 Empirical Results

Before the regression results are discussed, I will present some graphical evidence that supports the choice of the functional form. Figure 1 displays non-parametric regressions of equation (1), using the multivariate scatter plot smoother by Royston and Cox (2005). The Royston and Cox algorithm smoothes non-parametric estimates for each independent variable conditional on the other independent variables. Because these locally smoothed estimates require no assumptions about the functional form of the relationship, this method provides a check for the specification of equation (1).

The left part of figure 1 plots the partial effect of $telecom_k lang_{ij}$ against the logarithm of the volume of trade. The estimates are conditional on production differentials, the distance measure, bilateral and industry-fixed-effects as well as the interaction between language and postal services. The figure indicates that the non-parametric estimates are sufficiently linear to justify the choice of the functional form in (1). Also, the predictive power of the partial estimate is relatively high, as the tight confidence intervals show. The figure shows that trade increases in the interaction between language commonality and the direct communication measure. This is evidence in favor of the hypothesis that common language has a more pronounced effect on communication-intensive industries. Since all observations to the right of 0.0083 on the horizontal axis belong to services industries, the graph also indicates that most language-related variation in Canadian trade stems from service industries.

Similarly, the right part of figure 1 displays non-parametric estimates of the postal services interaction from the same regression. The graph shows that the volume of trade is straightly decreasing in the interaction of language with the indirect communication proxy. Because these estimates are conditional on the inclusion of the telecommunication-language interaction, the figure indicates that the partial effect of indirect communication on trade is potentially negative.

Parametric estimates of equation (1) are reported in table 7. Each model has a R^2 of more than 0.40. The high explanatory power of the models is also reflected in large F-statistics. As expected, the transport-distance interaction has a significant negative impact on inter-provincial trade in all models. The intuition for this estimate is that trade with distant provinces is particularly low for transport-intensive industries.

In accordance with standard trade theory I find that specialization affects trade positively. The estimate implies an average impact of $prod_{ijk}$ on trade of 39%, given the standard deviation of 0.045 of the production differential within an industry.[8]

In column (1), the interaction between language commonality and direct communication is statistically significant at the 5% level. When shifting from the 25^{th} to the 75^{th} percentile of the distribution of $lang_{ij}$, this implies an increase of trade volume by 2.64% for an average communication-intensive industry. For a service-intensive industry such as health, this effect would correspond to an increase of trade by 6.91%.[9] There is, however, less evidence for the presence of an indirect communication channel. The $post_k lang_{ij}$ variable in column (2) is statistically and economically insignificant. Only when both interactions are included together as in column (3), $post_k lang_{ij}$ becomes significant. This could be due the high correlation between $post_k$ and $telecom_k$ that blows up test statistics. The estimated partial impact of $telecom_k lang_{ij}$ nearly doubles. This indicates that industries that rely on direct interaction in order export their products trade more between areas with a high language commonality. On the other hand, I now estimate a significantly negative partial effect of the postal services interaction on trade. While the significance of this effect should be interpreted with great care due to potential multicollinearity problems, one can say that the evidence for an indirect communication channel via mail is less conclusive.

The estimates of β_3 might be biased if determinants of trade have been omitted from (1) that are correlated with the explanatory variables. Warren (2003) argues that the economic development of the French speaking provinces was retarded. Within Quebec most businesses were in the hands of an English speaking minority before strong French-promoting legislation was passed in the 1970s. If Canada's English speaking population were more affluent, all I capture with the language interaction would be a wealth effect. Hence, I control for the interaction between telecommunication and joint provincial GDP per capita in column (4). The insignificance of the estimate and the fact that the estimated β_3 remains practically unchanged indicate that I am really capturing language effects.

Another reason for bias of β_3 could be that other institutional variables that are correlated

[8] I calculated the effect as $\%\Delta trade_{ijk} = 100 * \hat{\beta}_2 * 0.045 = 100 * 8.647 * 0.045 = 38.91\%$.

[9] These numbers have been calculated for the pairings NL-NU (25^{th} percentile) and ON-PE (75^{th} percentile), where the effect for an average industry is $\%\Delta trade_{ijk} = 100 * \hat{\beta}_3 * \overline{telecom}_{ijk} * (lang_{ij}^{75} - lang_{ij}^{25}) = 100 * 20.23 * 0.0107 * (0.989 - 0.868)$.

with language have been omitted from equation (1). It could be that the foremost Catholic population in French-speaking Canada distrusts Protestant business partners or exhibits different demand patterns. If this were the case, the alleged language effect would really capture religious affiliation. Although Lipset (1990) argues that religion has a smaller role in Canadian everyday life than in the US, religious commonality has been shown to affect international trade patterns (e.g. Lewer and van den Berg, 2007; Helble, 2007). Hence, I control for the probability that two randomly chosen people from two states have the same denomination. The religion measure is highly correlated with language commonality (0.72). Yet the estimate in column (5) is insignificant, while β_3 remains a significant determinant of trade. The fact that the estimate still is of similar magnitude is evidence in favor of the language channel.[10]

5 IV Results

Although the approach taken here reduces several potential sources of bias that are present in standard gravity models, the estimate of β_3 could still be subject to endogeneity. I deal with this issue using legal language status as an instrumental variable (IV).

While Canada's Official Language Act of 1969 guarantees equal legal status of both English and French with respect to federal administrative services, federal courts, and in Parliament, some provinces enacted additional language laws. Particularly, Quebec and New Brunswick passed own official language acts during the 1970s that promote the use of French at the work place, in educational institutions, and for administrative procedures. The *Official Languages of New Brunswick Act* was first enacted in 1973 and later on revised. Likewise, Amendment 16.1 of the Canadian constitution, which was enacted in 1993, reinforces the equal status of the French language in New Brunswick. Quebec passed the *Official Language Act* (Bill 22) in 1974 and the *Charter of the French Language* (Bill 101) in 1977. Warren (2003) argues that these laws triggered a revival of the French language in everyday life and also in business, where English was to become the primary language in the 1970s. Lazear (1999) shows that the protection of minority interests by the government reduces incentives to learn the majority language, implying lower knowledge of English in those regions that guarantee specific language rights.

Therefore I use the legal language status across provinces as an instrument for the probability that two people from two provinces speak the same language. In particular, I use the

[10]Another variable that could be correlated with language commonality is ethnic origin. However, the data do not allow to disentangle ethnic origin and language ties for French Canadians, because most Canadian Catholics are also Québecois. Similarly, constructing an aggregate measure along the lines of equation (2) will not yield a consistent proxy for ethnicity, because large ethnic groups within Canada have ethnic origins that are unlikely to affect trading patterns, e.g. English, Irish, Scottish or Welsh.

interaction $c_k legal_{ij}$ as an instrument for $c_k lang_{ij}$, where c_k is assumed to be exogenous. As legal language status is predetermined and unaffected by the trade flow in 2001, it is a suitable instrument to isolate exogenous variation in language commonality. The variable $legal_{ij}$ is a dummy, which is one if Quebec or New Brunswick are a trading partner in a bilateral pairing, two for trade flows between these two provinces, and zero otherwise.

The IV estimates are reported in table 8. I only report second stage estimates. The statistics from the first stage regressions indicate that the IV estimator can be used. Columns (1)-(4) report large F statistics and high partial R^2s of the first stage regressions. Also, the instrument $c_k legal_{ij}$ is significantly partially correlated with $c_k lang_{ij}$ in the first stage regression. The IV estimate of $lang_{ij} telecom_k$ in (1) is positive and statistically significant, supporting the hypothesis of a language-trade channel. However, the estimate is larger than the OLS estimate in table 7, not smaller, as the potential reverse causality suggests. This could indicate a weak instrument, yet neither the partial correlations, nor the t-statistic suggest presence of a weak correlation. The Cragg and Donald (1993) test for weak instruments rejects the hypothesis that the equation is only weakly identified. To test for local average treatment effects, I drop the three territories from the sample, since they are partly inhabited by natives. This reduces much of the language variation that cannot be attributed to laws affirming the use of French in business. The estimate in (2) decreases to a value 23.91 which is close to the OLS estimates. This corresponds to an increase of trade in the health industry by 0.26% if $lang_{ij}$ increases by 1%.[11] Column (3) reports that indirect communication is not significantly related to the volume of trade, which is in line with the estimates from the fixed-effects regressions. I also compare the importance of indirect and direct communication in column (4). Instrumenting for both interaction terms and dropping the territories, the estimates confirm that language affects trade only in those industries with a high share of direct communication inputs.

6 Robustness

The following section tests the sensitivity of my results to the choice of the sample and potential bias of the estimates due to the focus on positive trade flows. Table 9 reports the sensitivity of the estimated coefficients to the removal of influential observations and the choice of the language variable. The estimates of β_3 are obtained from separate regressions of equation (1), using one interaction term in each run. The upper half of table 9 reports coefficients for the interactions between direct communication-intensities in each industry and language commonalities. The bottom of the table reports interactions for indirect communi-

[11] The calculation of the partial derivative with respect to $lang_{ij}$ yields: $\frac{\partial \ln trade_{ijk}}{\partial lang_{ij}} = \hat{\beta}_3 * telecom_k = 23.91 * 0.0107 = 0.26$.

cation intensities.

As a first robustness check, I estimate equation (1) with the full sample, including the fuel and coal and petroleum industries. Unsurprisingly, this reduces the estimate slightly, because both industries are unlikely to be affected by language differences. Then I examine if the estimates change if I drop all trade flows with the three territories from the sample. Yet the estimate for inter-provincial trade is of similar size. As a final test, I consider if my results reflect a mere one-off effect for the year 2001. Using 1998 trade, GDP, IO data, as well as Census data on language and ethnicity variables from 1996, I can confirm that the language-trade effects are statistically significant and of similar size.

As a second robustness check, I run the above regressions with an alternative proxy for language commonality. While $lang_{ij}$ proxies the potential for communication in the total population of two provinces, I use the probability that two randomly chosen people from two province use the same official language at work as an alternative measure. This measure would be preferred if one wants to account for the importance of language in business networks. Due to data availability, the measure $work_{ij}$ can only be constructed for 2001. $work_{ij}$ and $lang_{ij}$ are highly correlated. Table 9 reports that the estimates are statistically significant and of similar size in all samples. Overall, the robustness checks support the preceding fixed-effects and IV results: The telecommunication interactions are statistically significant in all models, whereas a higher language commonality has no significant impact on trade in postal service-intensive sectors. Hence, intra-Canadian trade patterns reveal a language-trade channel in industries that depend on direct rather than indirect communication.

Finally, I test if the above analysis is sensitive to the exclusion of zero trade flows from the sample. In order to account for these, I set all observations for which trade flows are not reported to zero. A log-transformation of zero values is not possible, but several methods have been suggested to deal with this issue. Sample selection procedures would probably be the most elegant way to adjust the estimates for zero observations. However, the estimation of sample selection models requires that at least one independent variable explains the selection process but is not partially correlated with the dependent variable (in order not to rely on distributional assumptions). While such exclusion restrictions can be justified for international trade flows (Helpman et al., 2008), it is hardly possible to find such variables for the intra-country trade data used here.

To get around the selection problem, Silva and Tenreyro (2006) suggest the use of the Poisson model for gravity equations. The Poisson estimator uses all positive and zero observations in a way that allows to interpret the coefficients similar to gravity estimates. Although it is typically used for count data, the Poisson estimator is consistent as long as the mean function is correctly specified. Helpman et al. (2008) confirm that the Poisson method yields similar estimates as their generalized gravity equation, which employs a two-stage procedure to account for the intensive and extensive margin of trade.

While most Poisson estimates in table 10 are of similar size as the fixed-effects estimates, the inclusion of zero trade flows has a strong effect on the distance estimate. The estimated trade barrier of distance is more than twice as large. This indicates that the predominant reason not to enter a trade relationship with another province is transport cost. Columns (1) and (2) report regression results for all possible provinces-industries combinations. The estimates of the interaction terms are of similar magnitude as the fixed-effects estimates. In columns (2) and (4), I drop the fuel as well as the petroleum and coal industries from the sample. Now the magnitude of the language interactions resembles the IV estimates from table 9. Hence, the Poisson regressions indicate that language differences impact the intensive margin of trade between Canadian provinces and territories rather than the extensive margin. The decision to enter into trade relations with another province seems to be primarily driven by transport cost considerations.

7 Conclusion

In this paper, I identify one mechanism that could justify the empirical evidence for the language barrier to trade in gravity models. I argue that language commonality should disproportionately foster trade in industries requiring more direct communication with the importer in order to trade, if high translation cost erect trade barriers.

To test for the existence of such a mechanism, this paper breaks new ground in the empirical trade literature. Firstly, it focuses on a single country, Canada, to test for language-induced differences in intra-industry trade. This reduces bias from other institutions that is potentially present in cross-country studies. Secondly, I deviate from standard gravity models by introducing bilateral fixed-effects and industry fixed-effects, thereby eliminating all variation that does not contribute to identify the proposed language channel of trade. Finally, I am able to control for potential endogeneity within the language-trade channel, using exogenous variation of the legal language status within Canadian provinces.

The findings support the proposed language-trade mechanism. The language barrier is larger for industries, which require more direct communication in order to trade their products. Particularly service industries trade more between provinces with a high proportion of same-language speakers. Moreover, the language channel appears to depend on direct (spoken) communication rather than indirect communication by mail. This is in line with Fink et al. (2005), who demonstrate the importance of international calling prices for the volume of bilateral trade. Finally, the significant negative relationship between the volume of trade and the distance-transport cost interaction holds potential for future applications of this methodology.

Future research might study in how far the language-trade channel also applies to international trade flows. It is likely that language will prove to be mainly an impediment to services trade and trade in complex goods that require direct communication with the foreign importer. If this were the case, language could turn out to be a source of comparative advantage that allows countries with a higher language commonality to specialize in more advanced goods and services.

References

ANDERSON, JAMES E. and ERIC VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, **93:1**, 170-192.

ANDERSON, JAMES E. and ERIC VAN WINCOOP (2004): "Trade Costs," *Journal of Economic Literature*, **42:3**, 691-751.

COMBES, PIERRE-PHILIPPE, MIREN LAFOURCADE and THIERRY MAYER (2003): "Can Business and Social Networks Explain the Border Effect Puzzle?," CEPR Discussion Paper, 3750.

CRAGG, JOHN G. and STEPHEN G. DONALD (1993): "Testing identifiability and specification in instrumental variables models," *Econometric Theory*, **9:2**, 222-240.

DEARDORFF, ALAN V. (1998): "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?," in *The Regionalization of the World Economy*, Jeffrey A. Frankel. University of Chicago Press: Chicago, 7-32.

FINK, CARSTEN, AADITYA MATTOO and ILEANA CRISTINA NEAGU (2005): "Assessing the Impact of Communication Cost on International Trade," *Journal of International Economics*, **67**, 428-445.

FRANKEL, JEFFREY A. and ANDREW ROSE (2002): "An Estimate of the Effect of Common Currencies on Trade and Income," *Quarterly Journal of Economics*, **117:2**, 437-466.

GÉNÉREUX, PIERRE A. and BRENT LANGEN (2002): "The Derivation of Provincial (Interregional) Trade Flows: The Canadian Experience," Statistics Canada Working Paper, 98.

HELBLE, MATTHIAS (2007): "Is God Good for Trade?," *Kyklos*, **60:3**, 385-413.

HELLIWELL, JOHN F. (1996): "Do National Borders Matter for Quebec's Trade?," *Canadian Journal of Economics*, **29:3**, 507-522.

HELLIWELL, JOHN F. (1997): "National Borders, Trade and Migration," *Pacific Economic Review*, **2:3**, 165-185.

HELPMAN, ELHANAN, MARC MELITZ and YONA RUBINSTEIN (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, **73:2**, 441-487.

Hummels, David (2001): "Towards a Geography of Trade Costs," mimeo, University of Chicago.

Hummels, David and Russell Hillberry (2003): "Intranational Home Bias: Some Explanations," *The Review of Economics and Statistics*, **85:4**, 1089-1092.

Hutchinson, William K. (2002): "Does Ease of Communication Increase Trade? Commonality of Language and Bilateral Trade," *Scottish Journal of Political Economy*, **49:5**, 544-556.

Lipset, Seymour Martin (1990): *Continental Divide. The Values and Institutions of the United States and Canada.* Routledge: London.

McCallum, John (1995): "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, **83:3**, 615-623.

Kox, Henk and Arjan Lejour (2005): "Regulatory heterogeneity as obstacle for international services trade," CPB Discussion Paper, 49.

Kox, Henk and Arjan Lejour (2006): "The Effects of the Services Directive on Intra-EU Trade and FDI," *Revue Economique*, **57:4**, 747-769.

Ku, Hyejin and Asaf Zussman (2008): "Lingua Franca: The role of English in International Trade," mimeo, Cornell University.

Lazear, Edward (1999): "Culture and Language," *Journal of Political Economy*, **107:6**, 95-125.

Levchenko, Andrei (2007): "Institutional Quality and International Trade," *Review of Economic Studies*, **74:3**, 791-819.

Lewer, Joshua J. and Hendrik van den Berg (2007): "Estimating the Institutional and Network Effects of Religious Cultures on International Trade," *Kyklos*, **60:2**, 255-277.

Melitz, Jacques (2008): "Language and Foreign Trade," *European Economic Review*, **52:4**, 667-699.

Nunn, Nathan (2007): "Relationship-Specific, Incomplete Contracts, and the Pattern of Trade," *Quarterly Journal of Economics*, **122:2**, 569-600.

Rajan, Raghuram G. and Luigi Zingales (1998): "Financial Dependence and Growth," *American Economic Review*, **88:4**, 559-586.

Rauch, James E. (1999): "Networks versus markets in international trade," *Journal of International Economics*, **48:1**, 7-35.

Rauch, James E. and Vitor Trindade (2002): "Ethnic Chinese Networks in International Trade," *The Review of Economics and Statistics*, **84:1**, 116-130.

Rose, Andrew K. (2001): "One money, one market: the effect of common currencies on trade," *Economic Policy*, **30:1**, 7-46.

Royston, Patrick and Nicholas J.Cox (2005): "A multivariable scatterplot smoother," *Stata Journal*, **5:3**, 405-412.

Silva, J. M. C. Santos and Silvana Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, **88:4**, 641-658.

Wagner, Don and Keith Head John Ries (2002): "Immigration and the Trade of Provinces," *Scottish Journal of Political Economy*, **49:5**, 507-525.

Warren, Jean-Philippe (2003): "The History of Quebec in the Perspective of the French Language," in *Linguistic Conflict and Language Laws. Understanding the Quebec Question*, Larrivée, Pierre. Palgrave Macmillan: Houndmills, 57-86.

Wolf, Holger. C. (2000): "Intranational Home bias in Trade," *The Review of Economics and Statistics*, **82:4**, 555-563.

Figure 1: The graphs depict conditional running line smoothed scatter plots, using the Royston and Cox (2005) algorithm. The dark line is the non-parametric point estimate of the language-communication interaction on the log of trade, conditional on the other covariates from equation (1). The dotted line depicts the upper and lower confidence intervals at the 95% level. The gray dots are individual estimates for each observation.

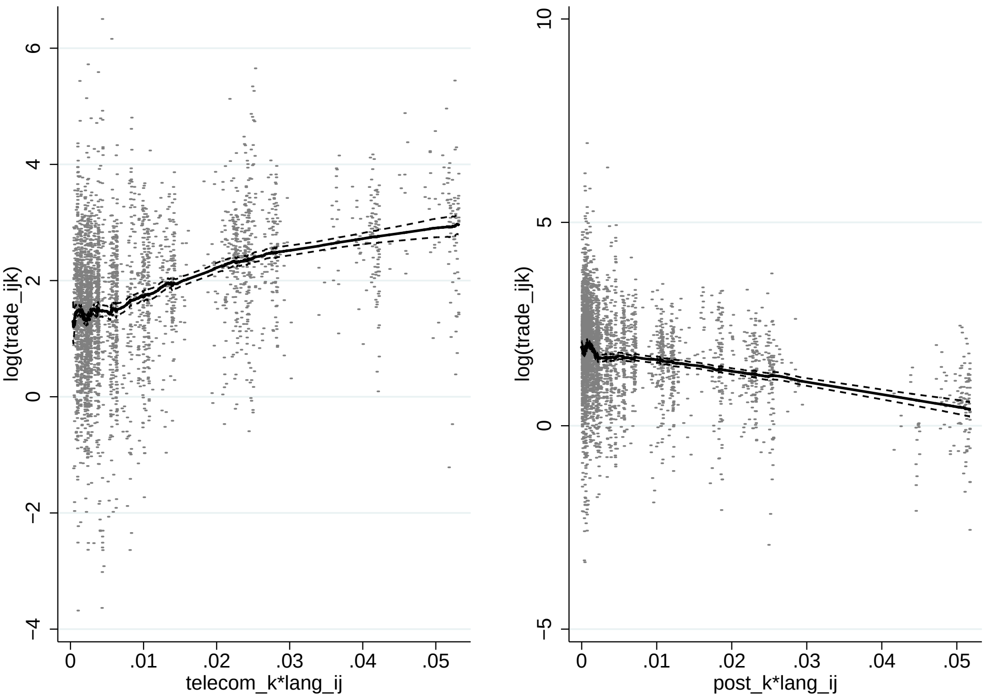

Table 1: Mother Tongues and Knowledge of Official Languages

	Mother Tongues				Knowledge of Official Languages				Total Population
	English	French	English and French	Other	English only	French only	English and French	None	
Canada	59.3%	22.7%	0.4%	17.6%	67.50%	13.30%	17.70%	1.50%	30,007,094
Newfoundland and Labrador (NL)	98.4%	0.4%	0.1%	1.1%	95.70%	0.00%	4.10%	0.10%	512,930
Prince Edward Island (PE)	93.9%	4.3%	0.3%	1.5%	87.90%	0.10%	12.00%	0.00%	135,294
Nova Scotia (NS)	93.0%	3.8%	0.3%	3.0%	89.70%	0.10%	10.10%	0.10%	908,007
New Brunswick (NB)	64.7%	32.9%	0.7%	1.7%	56.50%	9.20%	34.20%	0.10%	729,498
Quebec (QC)	8.0%	81.2%	0.8%	10.0%	4.60%	53.80%	40.80%	0.80%	7,237,479
Ontario (ON)	71.6%	4.4%	0.4%	23.7%	85.90%	0.40%	11.70%	2.10%	11,410,046
Manitoba (MB)	75.8%	4.1%	0.3%	19.9%	89.70%	0.10%	9.30%	0.80%	1,119,583
Saskatchewan (SK)	85.7%	1.9%	0.2%	12.2%	94.50%	0.00%	5.10%	0.30%	978,933
Alberta (AB)	81.8%	2.0%	0.2%	16.0%	92.00%	0.10%	6.90%	1.10%	2,974,807
British Columbia (BC)	74.1%	1.5%	0.2%	24.3%	90.30%	0.00%	7.00%	2.70%	3,907,738
Yukon Territory (YT)	87.1%	3.1%	0.3%	9.5%	89.40%	0.20%	10.10%	0.30%	28,674
Northwest Territories (NT)	78.1%	2.6%	0.2%	19.0%	90.40%	0.10%	8.40%	1.00%	37,360
Nunavut (NU)	27.6%	1.5%	0.1%	70.8%	83.00%	0.10%	3.80%	13.10%	26,745

Table 2: TOP 20 Non-Official Languages Used at Work, 2001

Mother tongue	Non-Official Language used	Share of working population
Total	4.6%	100.0%
English	0.7%	59.4%
French	0.5%	22.3%
'Chinese'	**42.0%**	**2.7%**
Italian	17.0%	1.6%
German	21.3%	1.4%
Chinese, nos	37.7%	1.4%
Cantonese	46.5%	1.0%
Punjabi	30.1%	0.9%
Spanish	26.2%	0.9%
Portuguese	26.0%	0.8%
Tagalog	11.7%	0.8%
Polish	17.2%	0.7%
Arabic	14.3%	0.6%
Greek	20.8%	0.4%
Dutch	8.5%	0.4%
Vietnamese	24.1%	0.4%
Ukrainian	11.5%	0.4%
Russian	20.8%	0.3%
Persian	16.1%	0.3%
Mandarin	46.1%	0.3%

Note: 'Chinese' is the sum of speakers with a mother tongue that is Cantonese, Mandarin, or Chinese (not otherwise specified). The data are from the 2001 census.

Table 3: Bilateral Language Commonality, 2001

$lang_{ij}$	Exporter	Importer	Mean Trade	Mean Exports	Mean Imports
0.431	QC	NU	3.56	4.25	0.95
0.492	QC	NL	24.57	30.49	18.31
0.501	SK	QC	29.64	21.05	37.96
0.509	QC	BC	91.80	111.13	71.30
0.515	QC	AB	112.26	122.93	101.59
0.530	QC	NT	5.64	6.79	2.09
0.539	QC	MB	42.85	40.31	45.72
0.549	QC	NS	41.98	51.03	32.09
0.550	YT	QC	1.20	0.60	1.41
0.557	QC	ON	838.45	774.08	901.12
0.568	QC	PE	7.26	8.69	5.45
0.804	NU	NB	0.49	0.10	0.62
0.822	QC	NB	64.18	64.05	64.32
0.847	NU	BC	1.35	0.68	1.55
0.851	ON	NU	4.13	4.86	1.70
0.861	NU	AB	3.21	0.95	3.94
0.861	NU	NT	2.99	1.03	4.95
0.863	NU	MB	0.81	0.50	0.92
0.867	SK	NU	0.33	0.34	0.24
0.868	YT	NU	0.95	1.78	0.13
0.868	NU	NL	0.34	0.25	0.37
0.870	NU	NS	1.01	0.34	1.28
0.872	PE	NU	0.30	0.20	0.60
0.913	NB	BC	5.38	5.17	5.59
0.924	NL	NB	17.05	20.65	13.79
0.926	SK	NB	1.83	1.65	2.03
0.927	NB	AB	5.83	4.92	6.73
0.934	NT	NB	1.23	0.20	1.43
0.938	ON	NB	61.90	86.32	35.27
0.939	NB	MB	3.11	3.00	3.22
0.948	YT	NB	0.58	0.15	0.65
0.949	NS	NB	32.15	31.90	32.38
0.958	PE	NB	8.84	6.42	11.05
0.962	ON	BC	276.05	373.43	175.81
0.968	NT	BC	8.71	15.03	5.65
0.970	BC	AB	230.30	226.31	234.39
0.971	MB	BC	30.89	34.11	27.77
0.974	SK	BC	32.15	28.39	35.91
0.974	NL	BC	3.73	2.47	4.55
0.975	ON	AB	410.28	463.42	355.57
0.975	ON	NT	18.92	13.35	34.68
0.977	YT	BC	6.02	7.51	5.07
0.978	NS	BC	7.86	6.94	8.77
0.978	ON	MB	120.77	135.62	104.65
0.979	SK	ON	108.92	92.42	126.39
0.979	ON	NL	52.53	68.39	35.14
0.980	PE	BC	1.13	0.62	1.46
0.984	NT	AB	9.57	2.61	13.49
0.984	YT	ON	4.02	2.40	4.60
0.986	ON	NS	81.28	109.05	50.91
0.986	MB	AB	78.38	67.17	88.96
0.987	NT	MB	1.50	0.82	1.76
0.989	SK	AB	110.14	85.05	135.23
0.989	PE	ON	14.67	9.64	18.17
0.989	SK	NT	0.83	0.85	0.78
0.990	NL	AB	8.42	2.70	12.30
0.991	NT	NL	0.39	0.23	0.44
0.992	SK	MB	36.57	35.92	37.31
0.992	YT	AB	2.90	1.28	3.60
0.993	NL	MB	1.91	1.37	2.20
0.993	YT	NT	1.21	2.25	0.28
0.994	NS	AB	11.27	10.11	12.36
0.995	NS	NT	1.84	2.64	0.34
0.996	YT	MB	0.42	0.41	0.42
0.996	PE	AB	1.29	1.14	1.39
0.997	SK	NL	1.39	1.45	1.27
0.997	YT	SK	0.29	0.33	0.28
0.998	PE	NT	0.23	0.23	0.20
0.998	NS	MB	4.12	3.86	4.38
0.998	YT	NL	0.10	0.05	0.11
1.000	SK	NS	3.07	2.97	3.18
1.000	PE	MB	0.95	1.27	0.74
1.000	PE	NL	1.98	2.58	1.43
1.000	NS	NL	15.29	19.88	10.84
1.000	YT	NS	0.29	0.13	0.36
1.000	PE	NS	7.54	5.20	9.80
1.000	YT	PE	0.20	-	0.20
1.000	SK	PE	0.52	0.63	0.36

Note: Trade is the average bilateral trade across all reported industries in million Canadian \$. $lang_{ij}$ is the probability that two randomly selected people from both regions are able to communicate with each other in English, French, or 'Chinese'. Own calculations.

Table 4: Input Shares by Sector, 2001

Industry	Trade *in million $*	Telecoms services *in %*	Postal Services *in %*	Transport Services *in %*
Fishery	6.78	0.05	-	0.68
Metal	90.31	0.07	0.02	2.18
Paper	60.41	0.09	0.05	3.94
Petroleum and Coal	102.60	0.09	0.01	0.59
Fuels	680.76	0.10	0.03	0.11
Lumber and Wood	35.25	0.14	0.04	2.14
Beverages and Tobacco	17.75	0.15	0.07	0.76
Residential Construction	74.65	0.15	0.07	0.76
Leather	29.74	0.19	0.19	0.94
Textiles	22.58	0.19	0.15	0.54
Hosiery	29.27	0.22	0.25	0.26
Fabricated Metal	50.42	0.22	0.11	1.48
Furniture	27.18	0.23	0.21	1.09
Mineral products	16.89	0.24	0.11	2.58
Minerals	8.48	0.24	0.08	0.70
Ores	77.29	0.24	0.08	0.70
Machinery	31.58	0.27	0.12	0.90
Motor vehicles, parts	93.38	0.27	0.14	1.84
Print	34.58	0.28	0.33	1.41
Manufactured Products	28.69	0.37	0.39	0.70
Accommodation and Meals	26.22	0.37	0.07	0.51
Chemical and Pharmaceutical	94.92	0.38	0.22	2.33
Forestry	26.02	0.40	0.04	0.20
Mining services	17.71	0.44	0.06	0.81
Electronic equipment	48.16	0.57	0.20	1.27
Grains	27.53	0.63	0.00	1.05
Fruits	82.46	0.63	0.00	1.05
Meat	88.13	0.63	0.00	1.05
other Agriculture	59.06	0.83	0.05	1.50
Retail	25.63	1.01	1.87	0.07
Finance and Insurance	143.51	1.07	0.56	0.02
Utilities	43.39	1.13	0.68	1.01
Educational services	3.32	1.41	0.71	0.05
Wholesale	140.05	2.27	1.08	0.13
Communication services	45.52	2.43	5.17	0.11
Health	3.70	2.52	0.45	0.05
other Services	51.25	2.82	2.32	0.39
Transport and Storage	74.00	4.21	1.22	0.88
Professional services	152.78	5.30	2.55	0.69

Note: Trade is the average trade in this industry across all reported bilateral flows. Telecommunication, postal, and transportation services inputs shares are calculated as the percentage of total inputs in that industry. Own calculations.

Table 5: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
$trade_{ijk}$ (incl. zeros)	31.73644	170.27857	0	4628.5	6864
$\ln trade_{ijk}$	1.62792	2.37216	-2.30259	8.43999	3572
$\ln(trans_k dist_{ij})$	2.50666	1.39345	-3.5636	5.31714	6864
$prod_{ijk}$	0	0.04548	-0.34918	0.34918	5922
$telecom_k lang_{ij}$	0.00712	0.01038	0.0002	0.05304	6552
$post_k lang_{ij}$	0.00436	0.00868	0	0.05165	6396
$telecom_k work_{ij}$	0.00695	0.01026	0.00017	0.0528	6552
$post_k work_{ij}$	0.00426	0.00856	0	0.05142	6396
$telecom_k GDPpc_{ij}$	0.00001	0.00002	0	0.00021	6552
$telecom_k religion_{ij}$	0.00148	0.00258	0.00002	0.02806	6552

Table 6: Cross-correlation table

	Variables	1	2	3	4	5	6	7	8	9
1	$\ln(trade_{ijk})$	1.00								
2	$\ln(dist_{ij} * trans_k)$	-0.08	1.00							
3	$prod_{ijk}$	0.11	0.05	1.00						
4	$telecom_k * lang_{ij}$	0.02	-0.24	-0.02	1.00					
5	$post_k * lang_{ij}$	-0.03	-0.33	-0.03	0.65	1.00				
6	$telecom_k * work_{ij}$	0.01	-0.23	-0.02	0.99	0.65	1.00			
7	$post_k * work_{ij}$	-0.04	-0.32	-0.03	0.65	0.99	0.66	1.00		
8	$telecom_k * GDPpc_{ij}$	0.01	-0.18	-0.01	0.84	0.55	0.84	0.55	1.00	
9	$telecom_k * religion_{ij}$	0.02	-0.25	0.02	0.72	0.46	0.69	0.45	0.64	1.00

Table 7: Baseline estimates

	(1)	(2)	(3)	(4)	(5)
$\ln(trans_k dist_{ij})$	-0.409***	-0.415***	-0.413***	-0.409***	-0.409***
	(0.119)	(0.123)	(0.123)	(0.119)	(0.119)
$prod_{ijk}$	8.647***	7.854***	7.862***	8.646***	8.647***
	(0.965)	(0.933)	(0.932)	(0.967)	(0.965)
$telecom_k lang_{ij}$	20.23**		36.67***	20.12**	20.22*
	(8.943)		(10.98)	(8.840)	(10.96)
$post_k lang_{ij}$		2.654	-24.29***		
		(7.884)	(8.296)		
$telecom_k GDPpc_{ij}$				234.6	
				(3065)	
$telecom_k religion_{ij}$					-0.0379
					(20.12)
Observations	3330	3261	3261	3330	3330
F-statistic	60.71	63.91	62.75	59.18	59.14
R^2	0.412	0.411	0.413	0.412	0.412

Note: Note: The estimates are from fixed-effects regressions of equation (1). The dependent variable is the natural logarithm of bilateral trade between provinces i and j. All specifications include fixed-bilateral and -industry effects. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10, 5, and 1 percent levels.

Table 8: Instrumental Variables

	(1)	(2)	(3)	(4)
$\ln(trans_k dist_{ij})$	2.034**	1.259	-0.0801	0.903
	(0.993)	(1.020)	(1.024)	(1.150)
$prod_{ijk}$	8.657***	9.591***	7.858***	8.591***
	(0.895)	(0.896)	(0.865)	(0.852)
$telecom_k lang_{ij}$	38.73***	23.91**		32.26**
	(11.24)	(11.41)		(14.93)
$post_k lang_{ij}$			13.38	-13.89
			(12.62)	(18.61)
Observations	3327	2554	3258	2492
F-stat	62.76	80.94	63.18	79.13
Craag-Donald test (p-value)	0.000	0.000	0.000	0.000
R^2	0.411	0.492	0.410	0.501
1^{st}-stage F statistic ($telecom_k lang_{ij}$)	184.6	155.3		95.04
1^{st}-stage partial R^2 ($telecom_k lang_{ij}$)	0.435	0.423		0.424
1^{st}-stage F statistic ($post_k lang_{ij}$)			103.4	92.53
1^{st}-stage partial R^2 ($post_k lang_{ij}$)			0.434	0.423

Note: The estimates are from fixed-effects instrumental variables regressions of equation (1). The dependent variable is the bilateral trade between provinces i and j. All specifications include fixed-bilateral and -industry effects. In Columns (2) and (4), the sample is restricted to Canada's ten provinces. The Craag and Donald (1993) statistic tests the null hypothesis that the model is weakly identified. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10, 5, and 1 percent levels.

Table 9: Robustness and Sensitivity Analysis

	Full sample	Only Provinces	1998 sample
Telecommunication services			
$lang_{ij}$	18.17*	19.30**	24.44***
	(9.313)	(9.161)	(9.325)
	3409	2554	3072
$work_{ij}$	18.33**	17.03**	
	(7.519)	(7.333)	
	3409	2554	
Postal services			
$lang_{ij}$	0.833	3.039	6.658
	(8.204)	(8.795)	(7.057)
	3340	2492	3021
$work_{ij}$	2.509	4.205	
	(6.697)	(7.087)	
	3340	2492	

Note: The regressions are estimates of equation (1). The dependent variable is the natural logarithm of bilateral trade between provinces i and j. All specifications include fixed-bilateral and -industry effects. Each entry of the table reports the estimated coefficients for β_3 with robust standard errors reported in parentheses. Below this the number of observations in the regression is reported. *, **, and *** indicate significance at the 10, 5, and 1 percent levels. The full sample includes also the following sectors: Fuels, Petroleum and Coal.

Table 10: Robustness to zeros: Poisson estimates

	(1)	(2)	(3)	(4)
$\ln(trans_k dist_{ij})$	-0.974***	-0.972***	-0.979***	-0.969***
	(0.254)	(0.254)	(0.254)	(0.254)
$prod_{ijk}$	7.029***	6.916***	7.001***	6.831***
	(1.496)	(1.668)	(1.494)	(1.645)
$telecom_k lang_{ij}$	19.30***	23.36***		31.65***
	(6.062)	(5.930)		(7.622)
$post_k lang_{ij}$			11.82	-15.68*
			(7.356)	(8.586)
Observations	5610	5466	5454	5310
Pseudo-R^2	0.891	0.908	0.892	0.910

Note: The estimates are from poisson regressions of equation (1). The dependent variable is the bilateral trade between provinces i and j. Columns (1) and (3) use all available observations. All specifications include fixed-bilateral and -industry effects. In columns (2) and (4) the industries Fuels, Petroleum and Coal have been dropped. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10, 5, and 1 percent levels.

Table 11: Variable labels

Label	Explanation
$trade_{ijk}$	The trade volume between province or territory i and province or territory j in industry k in million Canadian dollars, including zero trade flows.
$\ln trade_{ijk}$	The natural logarithm of $trade_{ijk}$.
$trans_k$	The share of transport margins in total inputs of industry k.
$dist_{ij}$	The bilateral distance between two capital cities of provinces or territories.
$prod_{ijk}$	The industry differential in the total production of two provinces.
$telecom_k$	The share of telecommunication services in total inputs of industry k.
$post_k$	The share of postal services in total inputs of industry k.
$lang_{ij}$	The probability that two randomly chosen people from province i and j are able to communicate with each other.
$work_{ij}$	The probability that two randomly chosen people from province i and j use the same official language at work.
$GDPpc_{ij}$	The joint GDP per capita of provinces i and j.
$religion_{ij}$	The probability that two randomly chosen people from province i and j have the same religion or denomination.