ABSTRACT

What Did You Say? The Effect of Language Distance on International Service Trade

Cristin K. Hulyk, M.S.

Mentor: Van Pham, Ph.D.

Service trade is more heavily dependent on communication than commodity trade because its production and consumption occur simultaneously. “Linguistic distance,” a measure of how closely the majority languages of two nations are related, is negatively correlated with the quality of communication between countries. I show that linguistic distance has a significant nonlinear effect on the level of service trade. For example, a change from communication between majority languages Russian and Danish to Russian and Czech, a decrease in linguistic distance, corresponds to a 17.1% increase in service exports. Additionally, the linguistic distance to English has a significant effect on service trade and points to the use of English as a vehicle language for trade. The data show that linguistic distance has a larger effect on service trade while physical distance has a larger effect on commodity trade.

What Did You Say? The Effect of Language Distance on International Service Trade

by

Cristin K. Hulyk, B.A.

A Thesis

Approved by the Department of Economics

_____
H. Stephen Gardner, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Master of Science in Economics

Approved by the Thesis Committee

_____
Van Pham, Ph.D., Chairperson

_____
Steven L. Green, Ph.D.

_____
Joseph A. McKinney, Ph.D.

_____
Lydia A. Grebenyova, Ph.D.

Accepted by the Graduate School
May 2012

_____
J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank Dr. Van Pham for his assistance in making this thesis the finished product that it is. I would also like to thank the rest of my committee, Dr. Steve Green, Dr. Joseph McKinney, and Dr. Lydia Grebenyova, for their support and input throughout this project. To Colin, my amazing husband who was my encourager throughout graduate school, thank you for your listening ear and tireless encouragements with this thesis. And finally, I would like to thank my Lord and Savior, Jesus Christ, for giving me the grace to be a part of this program and complete the project presented here.

CHAPTER ONE

Introduction

An article in the USA Today from January 2011 headlined with the fact that the Philippines has passed India as the main overseas location for call center jobs (Yun and Chu 2011). Another recent article from Bloomberg Businessweek details how other front office jobs such as accounting and engineering are moving to Latin America and Eastern Europe (Helyar and Srivastava 2012). Whether one notices it, or complains about it, or not, the service industry is becoming an increasingly global operation. Due to improving technology, services that would never have been considered tradable are now eligible for delivery around the globe (Amiti and Wei 2009, 203). Services have previously suffered from what is referred to as the proximity burden- the requirement that the supplier and consumer must be in the same location. However, now that technological advances have allowed for more flexible service delivery in person or electronically, the proximity burden has eased and allowed for a greater range of flexibility in service trade (Francois and Hoekman 2010, 648).

The service sector has become an increasingly important component of global economic activity. A service can be simply defined as "a product that is not embodied in a physical good and that typically effects some change in another product, person, or institution" (Deardorff 2010). According to WTO statistics, the total value of global cross-border service exports was $3.8 trillion in 2008, approximately 20% of the global trade in goods and services (WTO 2011; Francois and Hoekman 2010, 643). The growth of global service trade is shown in figure 1. The level of service trade increased by

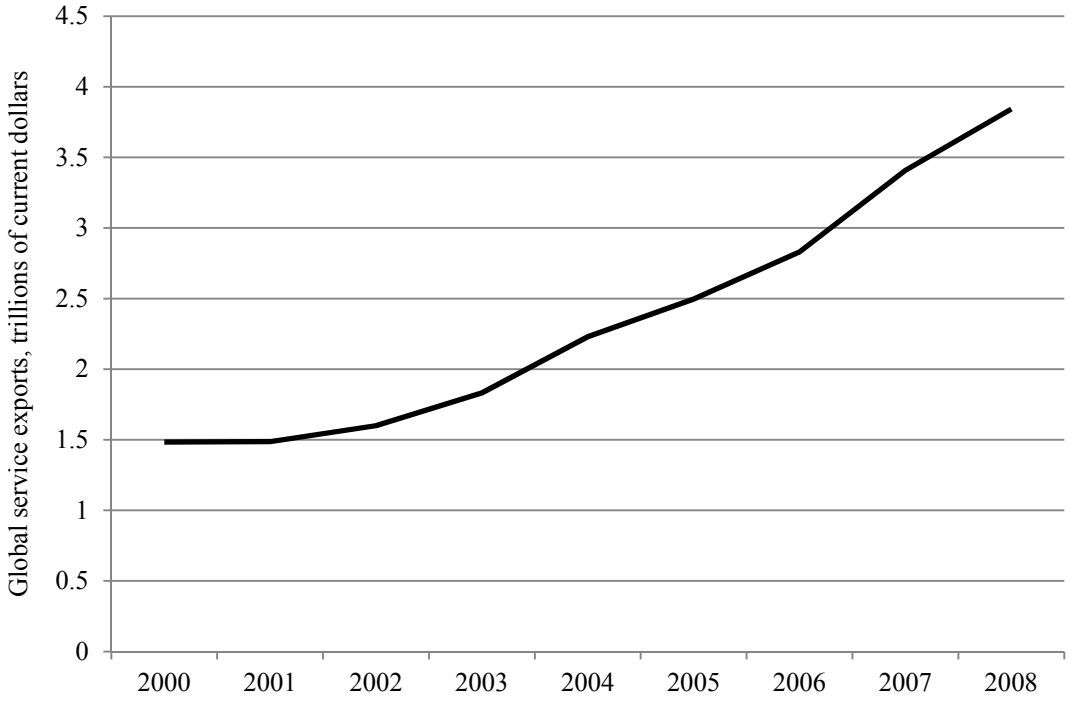

Figure 1. The growth of global service exports, 2000-2008. *Source*: WTO 2011, table A8

approximately 160% from 2000 to 2008 (WTO 2011). The service sector within the US makes up about 85% of nonagricultural employment (Jenson 2011, 14). Around the world, the service sector makes up 70% of employment and value added in OECD economies (OECD 2012). Services not only play an important role in developed economies, but in developing nations as well. Developing economies contributed 21% of global service exports in 2008, up from 11% in 1990. Service exports in developing countries are growing independently and, in some cases, at a faster rate than exports of goods. Some authors have suggested that service trade could even become an alternative path for economic growth and poverty alleviation (Goswami, Mattoo, and Saez 2011, 1).

Trade in services and trade in goods are fundamentally different from one another. With trade in goods, the production of the goods is often separate from the later transaction and trade of the good. Goods can be produced without the knowledge of

whether they will be sold in the domestic or a foreign economy. On the other hand, the production of services is closely linked with the trade of services since an arrangement is made for trade before the production begins (IMF 1993, 51). In other words, the production and consumption of services are simultaneous. This fact is the basis for the proximity burden; since production and consumption occur at the same time, the producer and consumer must therefore be located in the same vicinity. Many times a good can be confused for a service and vice versa since it is possible for a productive activity to have aspects of both.

I propose that because the production and consumption of services are coupled together, service trade is more heavily dependent on communication because it is more difficult to go through intermediaries. Commodity trade can move through contract manufacturers such as FoxConn Technology, the company who handles all of Apple's production facilities in China, so direct contact between the producer and ultimate consumer never occurs (Beales 2012). On the other hand, services flow directly from the producer to the consumer, so direct communication of some form is vital. In the above example with the Philippines, one of the main reasons cited for the growth of the service industry there has been the ability of the Philippines to speak an accent-neutral English that can more readily be understood by international clients. In a period marked by globalization, the language skills of a nation have the ability to shape its trade horizon in ways never predicted before. And this effect is all the more prominent due to the increasing importance of service trade in the global economy.

In the subsequent chapters, I will examine the effect that language has on service trade. I propose that the effect of language can be better measured by the use of linguistic

distance, a variable based on language acquisition theory that measures the relative distances between languages rather than just using a standard dummy variable for a shared language. I propose that linguistic distance affects communication between nations by influencing how costly language acquisition will be. Also, I offer an English distance variable as a way to measure the effect that English as a common language has on service trade. These exogenous language variables are used in place of a language dummy variable in a gravity equation to estimate patterns of international trade. Also, by making comparisons between service and commodity trade, I intend to show that service trade is more dependent on linguistic distance while commodity trade is more dependent on physical distance. Finally, I use the language variables to examine the different categories of service trade to look for fundamental differences in the types of activities included within the service sector and propose a new aggregate measure of business and personal services based on Jensen (2011).

Chapter two contains a literature review of gravity models of trade as well as the use of language variables and linguistic concepts within that area of study. Chapters three and four detail the data and methodology used in the paper. Chapter five presents the results from the regressions and, finally, chapter six is a discussion of the results and conclusion.

CHAPTER TWO

Literature Review

Treatment of Language in Gravity Models of Trade

Language variables are typically included along with a shared border dummy as a control variable in standard gravity models (Frankel 1997, 54). Usually a simple dummy variable is used that is equal to one if the two nations in the trading pair share a language, and zero otherwise (Anderson and van Wincoop 2004, 711). However, which languages are used for the presence of a shared language varies between papers. For example, Frankel (1997) only uses English, Spanish, Chinese, Arabic, French, German, Japanese, Dutch, and Portuguese for evaluation of a shared language, following the assumption that only certain languages have a significant effect on trade (74). Mayer and Zignago (2005), on the other hand, use a dummy variable based on whether countries share a language spoken by at least 9% of the population (17).

Melitz (2008) points out that if language minorities are included with other representative languages for a country, it is possible to have a very low probability that two random people from the nations' populations would actually speak the same language. In his opinion, languages spoken by a majority of people in a nation affect trade through a different mechanism than minority languages and need to be treated separately. He differentiates between direct communication and what he terms as open circuit communication, which reflects the presence of an established network of translation through major common languages. The dummy for direct communication was 1 if at least 4% of the population in both countries spoke the language. The dummy for

open circuit communication was 1 if the two nations shared an official language or a language spoken by at least 20% of the population (670). He found that common languages promoted trade through both translation and direct communication, but direct communication had a larger effect on trade.

One study that attempted to examine language outside of the use of a binary dummy variable was Boisso and Ferrantino (1997). The authors developed a linguistic distance index based on the likelihood that two random people selected from both nations in a trading country pair would speak the same native language. They used data from Katzner (1986) for native languages spoken by at least 1% of the population. Rauch and Trindade (2002) also used a similar language variable to measure common birth language in an attempt to proxy common tastes between countries. Their language data came from the Ethnologue Index, 12th edition from 1992, though they also use the linguistic distance index from Boisso and Ferrantino (1997) as well as a dummy variable for shared language spoken by at least 10% of the population to test their results for robustness.

There has only been one previous study that has attempted to measure the effect of linguistic distance on international trade. Lohmann (2011) used the World Atlas of Language Structures (Haspelmath et al. 2008) to develop an index of linguistic dissimilarity that captures the percentage of shared linguistic features between two languages out of a list of 139 possible features. The features are set up as main categories of linguistic information. For example, one feature is consonant inventory that has five main categories, namely small, moderately small, average, moderately large, and large. The features describe various linguistic information dealing with phonology, morphology, syntax, the lexicon, and writing systems. His index, much like the one used

in this paper, is designed to pick up the differences in the relative costs of the language barrier.

Estimates of the Language Barrier

Commodity Trade

Numerous studies have attempted to estimate the effect of a language barrier on the level of trade or in terms of a tariff equivalent. Most of these studies focus on language differences as a trade barrier with commodities trade. Frankel (1997) included a common language variable in what he termed as his "full" gravity model (54). Frankel found that sharing a common language led to 55% more trade, similar to the effect for sharing a border (75). A later study by Frankel and Rose (2002) that examined the effect of common currencies on trade reported that nations that share a language trade 80% more than those who do not, which again was similar in magnitude to the effect from a shared border (441). In their study, they found that a common language was a parameter of the border barrier effect as well as other factors such as a common currency, an FTA agreement, and a colonial relationship. Similarly, Boisso and Ferrantino (1997) found that a nation that is the completely dissimilar linguistically experienced 52-72% less trade than a nation completely similar linguistically. Rauch and Trindade (2002), however, found their language variable to only be significant for certain years and certain commodity groups that they used to split up the aggregate numbers.

Another approach to examining the effect of language is to estimate how much a common language affects the costs of trade in the form of tax equivalents. Eaton and Kortum (2002) in their estimated Ricardian trade model included a dummy variable for

sharing a language as one of several proxies for geographic distance (1760). They found that although increased geographic distance inhibits trade, the presence of a shared language lowers this effect (1761). Specifically, a shared language lowers costs by 6% while a shared border only lowers cost by 4% (1764-5). However, this study only looked at 19 OECD countries and only allowed English, German and French to be the possible shared languages. Anderson and van Wincoop (2004) point out that the interpretation of estimates from studies such as Eaton and Kortum (2002) depend on the estimated elasticity of substitution between imported and domestic goods. A review of the literature points to an elasticity of substitution of between 5 and 10 (716). When the authors use an intermediate value of 8 for the elasticity of substitution, the estimates from the Eaton and Kortum (2002) study and from another study, Hummels (1999, 2001), show that language costs in international trade caused a 7% mark-up on prices of imported goods while tariffs and non-tariff barriers to trade counted for about an 8% mark-up (Anderson and van Wincoop 2004, 693, 719).

The index of linguistic similarity, the Language Barrier Index, in Lohmann (2011) had a negative and significant effect on bilateral trade. The coefficient on his common language variable indicates that a 10% increase in the Language Barrier Index will cause a 7-10% decrease in trade flows between the countries. While this study uses a similar linguistic distance variable as I do, it mainly focuses on commodity trade flows rather than how language differences affect service trade. The following section reviews the use of a language variable in gravity equations for service trade.

Service Trade

While most papers dealing with a gravity model for trade look specifically at commodity trade, there are a few papers that examine service trade patterns and include some kind of language variable, though results from these studies are varied and sometimes contradicting. For example, Park (2002) uses a gravity model equation with service trade flow data from the Global Trade Analysis Project database from 1997. The results show that physical distance has less of an impact in service trade than in goods, and language has a positive and significant effect on service trade. This study divided the service trade data into 7 categories: construction, trade, transport, communication, financial, business, and others. The common language dummy was significant for all categories except communication, and the effect ranged from 11% more trade in transport services to 70% more trade in business services.

Another study by Lejour and de Paiva Verheijden (2004) examined service trade using a gravity model, however only between provinces in Canada and member states of the European Union. For the EU regressions, this study used of the OECD dataset on service trade and the language dummy variable used was equal to 0 if two regions share the same language and 0.2 if they did not. They found that the language dummy was negative and significant but did not have a large explanatory ability. Similar to the Park (2002) study, they also found that physical distance had less of an impact on service trade than on trade in commodities. Also, after dividing services into four categories, travel, transport, government, and other commercial services, the coefficient on the language dummy was negative and significant in all categories except for transportation. However,

their sample only included the EU15 countries with only 9 of those countries reporting trading partners.

Kimura and Lee (2006) also used the OECD dataset but with a wider sample of 10 OECD reporting countries and 47 partner countries that were OECD members and non-members. He used a dummy variable for whether two nations use the same language, pulling data from The CIA World Factbook. He found that the language dummy variable was positive and significant at the 1% level in all equations and had around the same magnitude in both the exports and imports equations. Specifically, the magnitude of the coefficient showed that countries that use the same language have 40-50% more service exports, holding all other factors constant. In the comparison between goods and service trade, the language variable was not significant for goods trade but was significantly larger in service exports than in goods exports. Also, they found that in their specification, physical distance was more important for service trade than for trade in goods, an opposite result than the findings of Park (2002) and Lejour and de Paiva Verheijden (2004).

Walsh (2008) also examined service trade using a gravity model approach, specifically a Hausman Taylor estimator for panel data. He used the OECD dataset with 27 OECD countries reporting service trade within the OECD and with up to 55 nonmember countries. He found that the wealth of the importing nation and a common language were the most important determinants of service imports (315). The common language dummy had a positive coefficient and therefore had a positive effect of service trade, though no specific interpretation of the coefficient was offered. Also, like the Lejour and de Paiva Verheijden (2004) study, the author divided the total service trade

aggregate into four categories to see how the effects of borders and distance changed with the different types of trade. Out of travel, transport, government, and other commercial services, only the other commercial services category showed language to have a significant effect on service imports. This is likely due to the fact that this category of services is more dependent on person-to-person communication. Interestingly, physical distance was not significant in the main service trade equation and was only significant for transport services in the disaggregated equations.

English as a Common Language for International Trade

Throughout the 1990s, English began to receive a special status as corporate language among the transnational corporations that were emerging (Truchot 2002, 14). Today, it seems to be the only language associated with international and global business (20). English is the most widely spoken foreign language, with 38% of EU citizens able to have a conversation in it (European Commission 2006, 12). Furthermore, 31% of those who know English indicate that they use it every day, making it the most widely used foreign language in the EU as well (16).

While the use of English as a global common language is oftentimes taken as a given, a lack of proper studies causes English to be often viewed as it seems to be rather than as it actually is (Truchot 2002, 20). Therefore, further studies that look into the use of English or its economic effect are beneficial and needed in order to better understand the global standing of the language. One such study that looks for evidence of English as a common language for international trade is Ku and Zussman (2010). In this study, the authors use the mean scores from the Test of English as a Foreign Language (TOEFL) exam to see if proficiency in English has an effect on trade above and beyond the effect

of a common language dummy and the use of English as a native language. They argue that for countries that do not speak English natively, improved English skills can lead to more foreign trade through facilitating communication with potential trading partners. Their OLS results show that trade increases by 31% if two countries share a common language, and by 1.51% for every 1% increase in mean TOEFL score. The authors also use linguistic distance as an instrument for TOEFL score since there may be reverse causality between English proficiency and international trade. The instrumental variable approach results in an even larger effect for English proficiency. This study provides strong evidence that a common language can have a facilitating effect on trade and that English is used as a vehicle language.

Previous Studies Using Language Distance

One of the primary uses of a quantitative measure of the distance between two languages has been in regards to ethnolinguistic diversity within a country. Greenberg (1965) suggested multiple measures for linguistic diversity, including a method for weighting a linguistic diversity measure for distance between languages[1]. This method uses a resemblance factor from the discipline of lexicostatistics equal to the number of shared cognates out of a list of 200 words.[2] Greenberg, a linguist at Columbia University, wanted to develop more precise methods for describing linguistic diversity in order make better comparisons of geographic areas as well as to be able to relate linguistic diversity to "political, economic, geographic, historic, and other non-linguistic factors" (109).

[1] The linguistic diversity measure described by Greenberg (1956) is the same as what is commonly referred to as ethnolinguistic fractionalization.

[2] For a more thorough treatment of quantitative methods for linguistic distance, see Data chapter.

Laitin (2000) was an early attempt to apply the methods laid out by Greenberg (1956) by computing and comparing each index for six former Soviet republics. However, due to concerns over the lexicostatistical method, Laitin introduced another method for calculating a resemblance factor that takes advantage of familial relationships among languages, often represented as family trees. He suggests assigning a score for the level at which two languages diverge from each other. Once the score is normalized, it comes out to be the number of shared branches divided by the maximum number of shared branches.

This idea was later developed by Fearon (2003) specifying the equation for linguistic distance used in this paper, adding that since cleavages farther back in history are likely to have greater effects than recent ones, the equation should be concave and therefore raised to a power of α, where α is less than one. Fearon (2003) used α equal to 0.5 while another study, Desmet, Ortuño-Ortín, and Weber (2009) used α equal to 0.05 to increase the relative distance between languages in the same family and languages in different families. These studies look at the resemblance factor as picking up linguistic as well as cultural similarities, assuming the two are correlated based on when two ethnic groups split apart from one another. Fearon and Laitin were primarily concerned with how their cultural diversity measure affects political economic outcomes within countries, such as governmental stability and civil wars. Desmet, Ortuño-Ortín, and Weber (2009) used the weighted linguistic diversity measure to see how ethnic differences affect redistribution within a nation. The study concluded that once distance between languages is accounted for, linguistic diversity has a significant negative effect on redistribution. Baldwin and Huber (2010) used the standard ethnolinguistic

fractionalization measure along with one weighted for linguistic distance and another weighted for income differences between ethnic groups to examine their differing effects on public good provision. They found that their income inequality weighted measure had a negative significant effect on public good provision, but the others did not.

A few studies have used the lexicostatistical approach described by Greenberg (1956). Ginsburgh, Ortuño-Ortín, and Weber (2005) used a list of distances for Indo-European languages based on lexicostatistics published by Dyen, Kruskal and Black (1992) to form a disenfranchisement index. This index was intended to show how someone who does not speak the official language and speaks a native language that is more distant from the official language is more disenfranchised linguistically than those with a less distant, non-official native language. A later study by Ginsburgh, Ortuño-Ortín, and Weber (2007) also used the Dyen, Kruskal, and Black (1992) distances in conjunction with a language acquisition model in an attempt to estimate demand functions for the major languages in Europe. In their language acquisition model, distance affected language learning costs: the larger the distance between two languages, the smaller the proportion of people who want to learn the foreign language. The language demand function also depended on size of the native and target language populations. This assumes that the individual who belongs to a large native language population will have less incentive to learn another language. Also, this assumes that if the target language population is large, there are more benefits to be gained from learning that language. In their empirical estimation, they use trade shares to represent possible gains from learning a language, underscoring the link between language and trade. Their theory holds well for estimating demand for English and German, but variables were less

significant for Spanish and French, indicating that other historical variables might be at play.

CHAPTER THREE

Data

Linguistic Distance Measurements

Linguistic distance is significant for the process of language learning through what is termed language transfer. Also referred to as cross-linguistic influence, this phenomenon relates to "the influence resulting from similarities or differences between the target language and any other language that has been previously (and perhaps imperfectly) learned" (Odlin 1989, 27). As the previous definition indicates, language transfer can be both positive and negative, where similarities can help learners and differences can hurt them. Negative transfer is easier to identify in research since it pertains to actual mistakes that language learners make; on the other hand, positive transfer can frequently only be detected by comparing the relative success of groups with different native languages (Odlin 1989, 36). Negative transfer can be manifested in numerous ways, such as underproduction of foreign structures, overproduction of more familiar or simpler features, or production errors such as substitutions of native words, alterations of structures, hypercorrection, or other errors that closely imitate a native structure (Odlin 1989, 36-38).

Studies that examine groups from different native languages often find evidence of language distances having an effect on language acquisition. For example, Ringbom (1987, 1992) found that native Swedish speakers in Finland had an easier time learning English than native Finnish speakers although teaching styles did not differ for the two groups. Whether it is the presence of negative transfer harming or positive transfer

aiding, language distance affects language acquisition. Language distance can be both an objective and a subjective measure depending on the learner. Odlin (1989) describes the two forces as follows:

> Objective measures of the distance between languages can be established through careful comparisons of structural similarities, which would show, for example, that the patterns of noun phrases in Spanish are more like those of French than those of English. While the cumulative similarities might be quantified in an objective manner, the *subjective* judgments of language distance by learners can matter considerably (32).

A lot of language transfer occurs because of learners' personal judgments over whether cross-linguistic similarities exist (Odlin 2003, 443). For example, a language learner who speaks multiple languages may draw more heavily upon knowledge of the language deemed more similar for aid in learning a new language (Odlin 1989, 141). Language distance greatly depends on the learner's perception of how distant a language is. Therefore, a second language that is perceived to be more related to another target language would be used more heavily for language transfer than the native language. Although specific predictions of language transfer may be difficult to make due to subjectivity among learners, an objective measure can be useful in predicting the overall, relative ease of learning a language. Even when subjectivity does exist, many times it is related to background characteristics such as age, literacy, social class, and motivation (Odlin 2003, 443).

Language transfer can exist in all linguistic subsystems such as orthography, phonetics, phonology, syntax, morphology, semantics, and pragmatics (Odlin 2003, 437). However, there is not a reliable way to weigh the relative effect that each subsystem will have on language transfer (Odlin 2003, 439). This is due to differing degrees of frequency among the subsystems. For instance, the use of phonemes occurs much more

frequently than the use of particular syntactic structures, but that does not reflect the relative importance of the two types of transfer (Odlin 2003, 440). An attempt to quantify language distance, and through that language transfer, should look to compare structural features in any of the linguistic subsystems as one subsystem does not necessarily have an advantage over any of the others.

Majority Language

The first step in quantifying linguistic distance is to choose which language or group of languages will be used to represent the linguistic community of a nation. For some nations, this is an easy process because there is one major language that is the official language and is spoken by a majority of the people in the nation. This would be countries such as the Czech Republic and Mexico, where over 90% of the population speak the official language (The World Factbook 2011). However, some countries have multiple official languages, such as Belgium where Dutch, French and German are all official languages but are spoken by approximately 60%, 40%, and less than 1% of the population, respectively. A possible solution to choosing a representative language in the presence of multiple official languages would be to choose the official language with the largest percentage of speakers, such as Dutch in Belgium.

The process of choosing a representative language is further complicated by the fact that sometimes an official language is merely a political distinction and the language itself is not spoken by a majority of the population. For example, the official language of Mauritius is English, but less than 1% of the nation speaks that language; most of the country speaks a French-based creole called Morisyen (The World Factbook 2011; Lewis 2009). Furthermore, the official language of Belize is English, but only 3.9% of the

population speaks it while 46% of the population speaks Spanish (Lewis 2009). Another large group of nations, mostly in Sub-Saharan Africa, do not have a language that is spoken by a majority of the nation. Cameroon, for example, has 281 languages with known speakers while the official languages are English and French (Lewis 2009). Many Pacific islands do not have a majority language, such as Papua New Guinea, who boasts 830 active languages for its population of around six million (Lewis 2009). In nations such as these, there is typically a lingua franca, or common language, used for commercial or governmental purposes. These may or may not be the official language of a country and it can be difficult to determine which lingua franca may be the most popular since most databases with language statistics, such as the Ethnologue, only report data on native speakers of languages. These statistics leave out valuable information on the common languages that are typically spoken as second or additional languages in multilingual communities.

A final problem with choosing a representative language is that presented by creole languages. Creoles evolve over time from pidgin languages, which are simplified amalgamations of area languages that frequently use one language as its base. As a pidgin is acquired by children as a native language, it becomes a creole language (Carstairs-McCarthy2001, 15). There are many nations, especially in the Caribbean and Pacific regions, that list a mainstream language as their official language, but the majority of the population actually speaks a creole language. Examples of such nations include Haiti, Jamaica, Seychelles, Sierra Leone, and Vanuatu (The World Factbook 2011). Also, Mauritius, mentioned above, is an example of a nation where the base of the creole spoken by a majority of the population is different from the official language (Lewis

2009; The World Factbook 2011). Sometimes the creole language also exists as a national lingua franca, such as Tok Pisin in Papua New Guinea (Lewis 2009). Classification of pidgins and creoles create problems, since the standard and pidgin/creole varieties are related along a continuum and the creole language and the base language may not be mutually intelligible (Crystal 2000, 108).

Multiple and irrelevant official languages, the lack of a majority language, and the existence of creoles are just some of the issues that exist when trying to choose a representative language for a country. One can attempt to collect data on the particular uses of languages in a nation[1] but comprehensive worldwide statistics on the uses of languages do not exist to my knowledge. Another approach is to use multiple languages for each country. For example, it is possible to use every language in a country such as with measures of ethnolinguistic fractionalization. However, the fact that there are 6,909 living languages today makes it restrictively complex to calculate a distance between every living language and a project of that magnitude is outside the scope of this paper. Additionally, many minority ethnolinguistic groups are marginalized from society and would not be likely to participate in international trade, so the suggestion of using every language is inappropriate for the focus of this paper. It is possible to use most spoken languages, but a specific cutoff across countries is likely to be arbitrary considering the wide range of linguistic contexts that exist country to country. Also, trying to determine the proper weights for languages only spoken by a portion of the community, especially with the lack of data on bilingualism, will carry its own problems.

While any method for measuring linguistic distance will inherently be noisy, in this paper I have chosen a representative language for each nation, a list of which can be

[1] For example, the language of commerce or of education can be a specific use of a language.

found in appendix table A. The main source of language data that I used in determining what I call the majority language of a nation is the CIA World Factbook. For country listings with percentages of speakers included, I used the language that is spoken by more than 50% of the population. In nations where a creole was the most spoken national language, I used the language upon which it is based, due to the lack of detailed information on characteristics of creole languages.

With nations that did not have a majority language, I used the official or national language with the most speakers. For entries that did not list percentages, I compared the number of native speakers for each language as listed in the Ethnologue 16th Edition Web Version. In addition to official languages, the Ethnologue lists national languages, which they define to be "languages spoken by a large portion of the population of a nation" (Lewis 2009). In nations where it was not clear what was the most widely spoken official or national language, mainly because of only having statistics for native speakers, I used the language with the most international influence. In Sub Saharan Africa, where the majority of the multilingual nations are located, this often was the colonial language. This is in conjunction with the theory in multiple language acquisition that learners draw more heavily upon second languages that are deemed more related to the target language (Odlin 1989, 141). So in Sub Saharan Africa, for example, the international official languages are likely to influence language acquisition more than the native languages.

Linguistic Distance

There have been various types of measurements employed throughout the literature to measure linguistic distance. While most linguists recognize that languages can be related to each other by various degrees, there have not been extensive measures

employed to quantify the distance in that discipline except in the more applied fields, such as second language acquisition. Many methods at quantifying linguistic distance have been used in the discipline of economics, as listed above in the literature review.

In this study, I use the historical method for the measure of the distance between two majority languages, referred to simply as linguistic distance. This method uses the approach in linguistics that characterizes relationships among languages as genetic, represented by a family tree diagram. The measure is similar to that used in Fearon (2003) which measures resemblance between languages as an increasing function of the number of shared branches between two languages, which is also concave in order to reflect that early divergence between two languages will have a larger impact than later divergence. Specifically, Fearon uses the following equation:

(1) $r_{ij} = (l / m)^{\alpha}$

Here, r_{ij} is linguistic resemblance between language *i* and language *j*, where *l* is the number of shared branches between *i* and *j*, *m* is the highest number of branches out of the languages in the data set, and α is a positive number less than one, specifically 0.5 in this study (220). I alter his equation slightly by setting *m* equal to the number of branches of the language of the exporting nation. This is in order to not skew measurements for language groupings that have a smaller number of divergences than others. For example, Russian has four branches while Portuguese, French, and Spanish have the largest number of classifications with a total of 10 branches. Sharing three branches with Russian has a different significance than only sharing three branches with Spanish. Linguistic tree diagrams are taken from the Ethnologue 16th Edition Web Version, which are based on the relationships listed in the *International Encyclopedia of Linguistics*

(Frawley 2003). The measure for linguistic distance is then one minus linguistic resemblance (r_{ij}). This measure for linguistic distance has also been recognized as a proxy for cultural distance as well, since language change occurs as ethnic groups become isolated from one another. Therefore, the farther away one language is from another may indicate how far away the two cultures have drifted as well. Descriptive statistics of linguistic distance, along with the other variables of interest, are found in table 1.

English Distance

The other language variable used in this study is English distance. This is the averaged distance of each majority language in a country pair to English. In order to better measure relative distance outside the Indo-European language family, alternative measures of language distance are used. Since language transfer can occur in any linguistic subsystem, I averaged measures of lexical, phonemic, and syntactic distance, explained below. In order to account for when a country pair might be likely to use English as a vehicle language, I interact the average distance to English with a dummy variable for no common languages between the countries. The dummy variable is one minus the dummy variable for common ethnic language from the GeoDist database (Mayer and Zignago 2011) which records a one if the two nations share a language spoken by at least 9% of the population. The English distance variable can be interpreted as a proxy for the ease of learning English and therefore an exogenous measure for the possible level of English skills in a linguistic community when English is necessary for trade.

In order to get a finer tuned language distance measure between the majority language of each country and English, I used the average of lexical, phonemic, and

Table 1. Descriptive statistics of main variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Linguistic distance	0.700	0.271	0	1
No common language * Combined English distance	0.401	0.136	0	0.714
Log GDP country 1	25.468	1.542	21.321	29.268
Log GDP country 2	25.256	1.939	16.665	29.268
Log population country 1	16.282	1.551	12.929	18.771
Log population country 2	16.373	2.031	9.189	21.004
Log GDP per capita country 1	9.299	1.066	6.382	10.930
Log GDP per capita country 2	8.949	1.347	4.599	11.329
Border	0.063	0.244	0	1
Avg total years of education	9.576	1.343	5.646	13.091
Log physical distance	7.941	1.042	4.842	9.842

Note: GDP and GDP per capita are in constant 2000 US dollars

syntactic distance. The lexical distance measure comes from Dyen, Kruskal, and Black (1992), who computed a percentage of lexical similarity between Indo-European languages based on the number of cognates from a list of 200 basic words. This method of comparing cognates is called lexicostastics and was introduced by Morris Swadesh in 1950. (Embleton 2003). It compares a list of 200 basic words that are common to all cultures, referred to as the Swadesh list, for two languages and counts the number of common cognates. The basic vocabulary of a language is compared because it is the most resistant to borrowing (Petroni 2008, 3). This measure can reflect the similarity between

the vocabularies of two languages as well as how related historically the two languages are, since cognates are defined as two words that have descended from the same ancestor word and do not include borrowed words (Dyan, Kruskal, and Black 1992, 95). The one drawback to this approach is that, as in the historical method, the similarity between two languages in different language families is by definition zero and the distance is equal to one. Therefore, I developed two additional measures of linguistic distance which I call phonemic distance and syntactic distance using data from the Global typological database developed by Merritt Ruhlen (2008).

Phonemic distance equals one minus the percentage of shared consonantal phonemes between English and the majority language of a nation, using the 24 consonantal phonemes of English as the base of comparison. A phoneme is the basic sounds of a language which can take on different variants of phonemes, called allophones, depending on the phonetic context. In the Ruhlen database, both standard phonemes and marginal phonemes mainly from loanwords are reported and both are included in the distance measure here. Also, modified phonemes, or phonemes with a slightly different articulation, were included if the modified phoneme existed as an allophone in English. For example, aspirated voiceless stops (/p^h/, /t^h/, /k^h/) are phonemic in English, but phonemic unaspirated voiceless stops were considered to be a shared phoneme since the unaspirated variety exists in English in certain consonant clusters, such as the word 'stop.'[2]

Syntactic distance was calculated by comparing six word orders between English and the majority language: Subject-Verb-Object, Adjective-Noun, Genitive-Noun,

[2] A consonant is aspirated when a brief puff of air follows the consonant, such as in 'tie' or 'pie' (Ruhlen 2008, 7).

Demonstrative Pronoun-Noun, Number-Noun, and Possessive Pronoun-Noun.[3] Specifically, it was calculated as one minus the fraction of shared word orders between English and the majority language. In some cases, such as the genitive in English, two forms exist for word order (GN/NG). Here, a similarity is marked if the majority language also has both constructions possible. The word order characteristics in the Ruhlen database refer more to surface level word order, which lacks the sophistication of some of the more modern syntactic theories of generative grammar. A preferable approach would be to use specific parameters of syntax, such as the head first/head last parameter, to compare the syntax of different languages, but to my knowledge a comprehensive database of those specific characteristics does not exist, though the World Atlas of Language Structures dose begin to try to accumulate information of this kind.

The final English distance variable in theory would range from 0 to 1, however the maximum value as seen in table 1 is only 0.714. This is due to the inclusion of syntactic and phonemic distance, which in all cases are less than one. Since at least some phonemes or syntactic aspects are shared between the majority languages and English, English distance never reaches the maximum theoretical distance of one. Also, this variable is likely to be biased towards one since it is the interaction between the no common language dummy and the average English distance for the country pair.

[3] Genitive refers to a noun case or prepositional phrase that indicates possession. Two constructions for the genitive exist in English, such as 'the student's thesis' or 'the thesis of the student' (Ruhlen 2008, 20). A demonstrative pronoun is one that points out a specific noun that it is referencing, such as 'this' or 'these' (Loberger and Welsh 2002, 24). A possessive pronoun would be one such as 'my' or 'our.'

Test of Language Variables

In order to check for the use of linguistic distance as a proxy for the ease of learning a language, and therefore the predicted level of use of that language in a foreign nation, I regressed English distance against average Internet-based TOEFL scores from September 2005 to December 2008 (Education Testing Services 2007, 2008, 2009). The TOEFL is a standard English proficiency exam used as a part of college admissions, visa applications, and professional licensing accreditation processes in over 130 countries (Education Testing Services 2012). The test contains four sections covering reading, listening, speaking and writing skills, each worth 30 points for a total of 120 possible points. Since average scores are dependent on the population that takes the exam, I use average years of schooling in 2005 and GDP in 2008 as well as regional dummies to control for systematic differences in the testing population between countries. I have data for 119 countries, excluding native English speaking nations. The results of the OLS regression are found in table 2. In this data, the average TOEFL scores ranged from 55 to 102.33, with a mean of 79.07. The significant coefficient on English distance suggests that a one standard deviation (0.24) increase in English distance results in a 4.15 point drop in average score. Once regional dummies are added, the magnitude of the coefficient on English distance increases and remains highly statistically significant.

Additionally, table 3 shows OLS regression results for combined English distance and combined averaged TOEFL scores within country pairs. Here, the variables are combined by averaging the values for the two countries. Given the available TOEFL score data, 1,619 country pairs are used in the regressions. Again, controls for average years of schooling in 2005 and GDP in 2008 are included as well as regional dummy

Table 2. OLS estimates of the effect of English distance on average TOEFL scores, 2005-2008

Avg TOEFL score	1	2
English distance	-15.17***	-16.67***
	[3.616]	[4.967]
Avg years of education	2.030***	1.103***
	[0.284]	[0.365]
Log GDP	1.529***	1.053**
	[0.424]	[0.433]
Regional dummies	No	Yes
Constant	34.86***	57.33***
	[9.351]	[11.16]
Observations	119	119
R-squared	0.52	0.593

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Table 3. OLS estimates of the effect of combined English distance on combined average TOEFL scores within country pairs, 2005-2008

Combined average TOEFL score	1	2
Combined English distance	-26.43***	-28.14***
	[1.404]	[1.793]
Combined avg years of schooling	1.436***	0.753***
	[0.0982]	[0.108]
Log GDP country 1	-0.307***	-0.652***
	[0.0841]	[0.0811]
Log GDP country 2	0.289***	-0.201***
	[0.0777]	[0.0770]
Regional dummies	No	Yes
Constant	86.21***	112.1***
	[3.063]	[3.171]
Observations	1,619	1,619
R-squared	0.327	0.445

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

variables for each country. English distance is significant at the 1% level for both regressions. The magnitude of the coefficient for combined English distance suggests that a standard deviation increase in the average distance from English for the two majority languages corresponds to a 3.0 point decrease in the averaged TOEFL scores.

While the above regressions show that linguistic distance has a significant effect on language ability, there are other factors that affect language acquisition. Linguistic distance is likely to have an effect on the developmental path of language learning through language transfer, but individuals will probably exhibit varying degrees of success in second language acquisition due to various cognitive and affective factors that create differences among individual learners (Mitchell and Myles 2004, 24-27). Cognitive factors may include different degrees of intelligence and language aptitude or different language learning strategies that may be more successful for one learner over another (Gardner and MacIntyre 1992). Affective, or emotional, factors include the learner's attitude towards the target language, level of motivation, or the presence of language anxiety versus the willingness to communicate (Gardner and MacIntyre 1993). Most of these factors are unobservable in a population and may create endogeneity between the number of speakers of a second language and the level of trade in a nation, since the level of trade or other economic factors may affect the motivation or attitude of a learner. Language learning is a complex process which is difficult to be summarized in one variable. However, linguistic distance is an exogenous proxy for language learning outcomes so is useful although imperfect in this study.

Trade Data

The service trade data in this study come from the UN Service Trade database. Data is a cross-section from 2008 with 41 countries reporting exports to 207 partner countries, for a total of 2,323 country pair observations. The primary regressions use the 200 Total EBOPS Services aggregate measure. Disaggregated service trade by categories also comes from the UN service trade database. Commodity trade data comes from the UN comtrade database. Specifically, the H1 total trade variable is used. In order to make comparisons between services and commodity trade, the same subset of reporting and partner countries was used although a larger amount of data is available for commodity trade.

Other Control Variables

The basic controls used in the regressions are GDP and population, or alternatively GDP per capita, which are all obtained from the World Development Indicators at the World Bank. These variables are included for each country in the country pair. However, due to the smaller number of reporting countries, there is less variation in these variables as compared to partner countries. As in shown in table 1, the means for these variables are similar between reporting and partner countries, however the standard deviations are greater for the partner countries, referred to as country 2 in the table.

The variable for total years of schooling comes from the Barro Lee dataset update within the Education Statistics dataset at the World Bank. The physical distance measure used in the paper comes from Mayer and Zignago (2005), whose data was made available through the GeoDist database through CEPII (GeoDist 2011). The measure weighs

bilateral distances according to population of major cities (Mayer and Zignago 2011). Additional controls taken from the GeoDist database include a dummy for whether two countries share a border as well as dummies for whether the countries have ever had a colonial link and whether they have had a common colonizer after 1945. The common language dummy mentioned also came from this database.

Geography controls come from the geography data from Gallup and Sachs (1999) provided by the Center for International Development at Harvard (CID 2007). Specifically, I am using a measure for the proportion of land within 100 kilometers of the ocean or an ocean-navigable river, a dummy for whether a country is landlocked, and land area in square kilometers.[4] I control for institutional quality by using the average score from the Corruption Perceptions Index (Transparency International 2008). Also, I use the political stability measure from the Worldwide Governance Indicators, which represents the "perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means" (WGI 2011). Since the variable ranges from approximately -2.5 to 2.5, I added 2.5 to make the range from 0 to 5 and averaged the measures for the two countries. The regional dummies I use come from Fearon (2003) and are for the following regions: Asia including the Pacific, the Middle East and North Africa, Latin America, Sub-Saharan Africa, Eastern Europe and Western countries (U.S., Canada, and Western Europe).

The additional variables besides GDP, population, and GDP per capita are combined for the country pair. Specifically, continuous variables are averaged, and dummy variables that relate to a specific country, rather than the country pair, are added

[4] The propotion of land within a hundred kilometers of the ocean or an ocean navigable river measure excludes coastline or rivers leading to coastline that is frozen most of the year. The dummy for landlocked countries excludes countries in Western and Central Europe (Gallup and Sachs 1999, 35-36).

together so that the variable takes on the value of 0, 1 or 2. Regressions including separate variables for each country in the country pair were also run, however there was no evidence that the inclusion of separate variables significantly changed the coefficients on the variables of interest.

CHAPTER FOUR

Methodology

The main regression model used in this paper is the gravity equation for trade. This is the standard model for estimating international trade flows that is based on the assumption that trade patterns depends positively on the size of the countries and negatively on physical distance (Frankel 1997, 53). The gravity equation has typically been estimated using OLS but lately other estimation techniques have been used to more accurately estimate panel data. Frankel (1997) describes the basic formation of the gravity equation as only income and distance, while variables for common border, common language, and per capita income are frequently added for the standard "full" gravity equation (54). In this paper, I will be using OLS with cross sectional data for 2008. My main gravity equation, with variables of interest added, is as follows:

(2) $Service\ trade_{ij} = \alpha + \beta_1\ Language\ distance_{ij} + \beta_2\ Log\ physical\ distance_{ij}$

$+ \mathbf{X'\gamma} + \varepsilon_{ij},$

where $Service\ trade_{ij}$ is either Log service exports from country i to country j or, in some specifications, service exports as a share of the exporting nation GDP. The main variable of interest for this study is $Language\ distance_{ij}$, which is either linguistic distance between the country pair or the average of English distance for the country pair multiplied by the no common language dummy variable. Additionally, $Log\ physical\ distance_{ij}$ is also a variable of interest, specifically for measuring the effect of the proximity burden in service trade as compared to commodity trade. $\mathbf{X'}$ is a vector of control variables, including GDP and population for both countries in the country pair,

the dummy variable for a shared boundary, and the mean total years of schooling averaged for the two countries. α_{ij} is a constant, and ε_{ij} is the error term, which is assumed to be independently and identically distributed normal.

Frankel (1997) describes GDP and population as two measures of the size of an economy. Generally, while holding GDP constant the sign on population is negative, since countries tend to trade less if their country is larger. Trade may be larger in absolute terms but is oftentimes smaller as a proportion of GDP if the domestic economy is larger and therefore more self-sufficient. Also, there is obviously a close relationship between GDP, population, and GDP per capita since GDP per capita is equal to GDP divided by population. Since these terms are in log form in the gravity equations, Frankel (1997) claims that the equation with GDP and GDP per capita is mathematically equivalent to the equation with GDP and population (57). He furthers claims that when in the form of the gravity model with GDP and GDP per capita, a coefficient of less than one on GDP demonstrates that trade rises less than proportionately with size. However, in the equation with population, the same phenomenon would be reflected with negative coefficients on population. Additionally, GDP per capita can be viewed to have a possible effect on trade, since oftentimes a higher level of development can be indicative of more innovative products that would be demanded by other nations as exports, of a more advanced infrastructure that is conducive to trade, or a great level of trade liberalization (58).

Another variable that is commonly included in gravity models but is excluded here is a dummy variable for whether two nations belong to the same regional trade agreement or have a bilateral trade agreement. I believe the exclusion of a variable such

as this will not have a major impact here since most regional trade agreements and other bilateral agreements have not been effective in liberalizing service trade and most advances in that area have come from unilateral policy reforms (Francois and Hoekman 2010, 674). Also, Grunfeld and Moxnes (2003) included a variable for regional trade agreements in their gravity model for service trade but did not find it to be statistically significant.

Additional controls are added to the main gravity equation for robustness checks. Four groups of additional controls are added for supplementary regressions. The geography controls are a dummy for whether the countries are landlocked, the logged average land area of the countries, and a proportion for the amount of land within 100 kilometers of the ocean or an ocean navigable river. The next group of controls is for institutional quality and includes the average CPI score and the political stability measure. Following that, variables are added to control for whether one nation in the country pair has ever been a colony of the other or whether the two nations have shared a common colonizer. The last group of controls is the regional dummy variables listed above. Since these controls are merely for robustness, the results for these variables are not reported; however, the presence of different sets of controls is marked in the tables.

CHAPTER FIVE

Results

The Effect of Language on Service Trade

Language is proposed to have an effect on service trade because the production of the services occurs at the same time as consumption, providing less opportunity for trade to flow through intermediaries that can have specialized knowledge of the local language and culture. Due to the integrated and simultaneous nature of service trade, more direct contact and communication between the producer and consumer are required. Although improving technology is easing the proximity burden that traditionally hindered service trade, language barriers are still likely to be a major hindrance to communication and, therefore, trade.

This section will examine the nature of the effect of language on service trade by using exogenous linguistic distance variables that approximate the relative ease of one speaker learning another language. This will affect both the quality of language acquisition and resulting language ability as well as the likelihood of finding a worker with the necessary language skills for a particular export market. If languages are easier to learn, finding staff or external translators with specific language skills or training in-house workers in that language will be less costly overall, leading to higher levels of trade.

The first question of relevance for this study relates to whether language even has an effect on the level of trade. To see if there is a difference in level of trade between countries that share a language and those that do not, I ran a simple trade gravity equation

regression using a dummy for no common language. The results shown in table 4 suggest that the absence of a common language does have a significant effect on service trade, corresponding to a decrease in the level of exports of 64.6% if two countries do not have a common language, holding all other factors constant.[1] Once physical distance is added, the effect from lack of a common language increase to a 68.9% decrease in service exports.

Table 4. OLS estimates of the effect of a common language on service exports

Log service exports	1	2
No common language	-1.038***	-1.169***
	[0.158]	[0.131]
Log GDP country 1	1.235***	1.253***
	[0.0430]	[0.0356]
Log GDP country 2	1.103***	1.046***
	[0.0348]	[0.0288]
Log population country 1	-0.626***	-0.498***
	[0.0467]	[0.0388]
Log population country 2	-0.594***	-0.391***
	[0.0350]	[0.0297]
Physical distance		-1.097***
		[0.0343]
Constant	-20.63***	-16.17***
	[0.907]	[0.764]
Observations	2,251	2,249
R-squared	0.566	0.702

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Linguistic Distance between Trading Partners

While many studies mention that there is a language barrier in regards to trade, many do not examine the relative language barrier due to different language distances

[1] The coefficients for dummy variables are interpreted as exp(β) – 1 since the dependent variable is in the log form.

(Frankel 1997; Frankel and Rose 2002; Anderson and van Wincoop 2004; Mayer and Zignago 2005; Rauch and Trindade 2002; Eaton and Kortum 2002; Hummels 1999, 2001). Therefore, I use the linguistic distance variable described above to measure the effect of the relative distance between the majority languages of two countries. Additionally, I assume that linguistic distance will not have a constant effect as distance increases due to two reasons. First, initial increases in linguistic distance are likely to have a larger effect compared to similar increases at the other end of the continuum. Because a linguistic distance of zero means that two nations share a majority language, any initial increase to a nonzero value such as 0.1 will dramatically decrease the level of service trade since foreign languages are introduced between the trading partners. At the same time, an increase in linguistic distance that is equivalent to movement from one distant language to another will not increase language learning costs to the same degree.

The second reason for a nonlinear effect of linguistic distance on service trade is the nature of the linguistic distance variable itself. As explained in chapter three, linguistic distance is equal to one minus the resemblance factor between two languages. The resemblance factor computation used for linguistic distance is the equation defined by Fearon (2003), specifically that the resemblance factor is equal to the number of shared branches divided by the maximum number of shared branches between the two languages, all raised to the one-half power. While Fearon allows that any increasing function of the number of shared classifications between the two languages would work, he also mentions that the function should be concave to reflect the fact that earlier divergences will produce greater differences than later divergences.[2] Because the

[2] It is important to note that Fearon (2003) uses linguistic differences as a proxy for cultural differences while in this study I am specifically interested in the linguistic aspects.

resemblance factor is concave to the number of shared classifications, the measure of linguistic distance will also be concave and will therefore be likely to have a nonlinear effect. In order to capture the nonlinear effect on service trade, I use both linguistic distance and linguistic distance squared in the regressions.

Table 5 presents the regression results for the effect of linguistic distance on service trade. In column 1, a basic gravity model specification, the coefficients on the linguistic distance variables are significant at the 1% level and the signs on the coefficients are as expected, indicating that as linguistic distance increases, service trade is decreasing but by a diminishing amount each time. In columns 2-4, additional controls are added which increase the R-squared values of the regression but do not change the significance for the linguistic distance variables, though the magnitudes fluctuate. Specifically, the shared border dummy is added in column 2 and the average total years of education for the two countries is added in column 3.

Log physical distance is added as a control in column 4. Although it is possible to expect linguistic distance to only be a proxy for physical distance, the correlation coefficient between linguistic distance and physical distance is only 0.2772. Therefore, these two variables are likely to represent the separate effects of physical distance and linguistic distance on service trade. Logged physical distance is significant in columns 4-8 and the magnitude is always greater than negative one. This suggests a 1% increase in physical distance will result in a slightly greater than 1% decrease in level of service exports. It is important to note that before physical distance in column 4 was added, the magnitude on the border dummy is incredibly large and significant. This would suggest that there is a large proximity burden, since sharing a border is correlated with 10 to 13

Table 5. OLS estimates of the effect of linguistic distance on service exports

Log service exports	1	2	3	4
Linguistic distance	-4.655***	-2.690***	-2.452***	-3.839***
	[0.651]	[0.647]	[0.626]	[0.533]
Linguistic distance squared	2.977***	1.740***	1.600***	3.014***
	[0.493]	[0.486]	[0.470]	[0.401]
Log GDP country 1	1.243***	1.296***	1.277***	1.287***
	[0.0417]	[0.0404]	[0.0399]	[0.0337]
Log GDP country 2	1.056***	1.076***	0.906***	0.936***
	[0.0338]	[0.0332]	[0.0403]	[0.0340]
Log population country 1	-0.645***	-0.698***	-0.649***	-0.518***
	[0.0456]	[0.0444]	[0.0439]	[0.0373]
Log population country 2	-0.520***	-0.566***	-0.396***	-0.258***
	[0.0336]	[0.0335]	[0.0415]	[0.0355]
Border		2.623***	2.404***	0.369**
		[0.178]	[0.174]	[0.163]
Avg total years of education			0.264***	0.172***
			[0.0365]	[0.0310]
Log physical distance				-1.050***
				[0.0364]
Geography controls				
Institutional controls				
Colonial controls				
Regional dummies				
Constant	-19.97***	-21.03***	-22.41***	-18.26***
	[0.882]	[0.873]	[0.882]	[0.760]
Observations	2,323	2,251	2,095	2,093
R-squared	0.564	0.607	0.622	0.73

Table 5 continued

Log service exports	5	6	7	8
Linguistic distance	-4.223***	-3.678***	-2.816***	-2.293***
	[0.548]	[0.539]	[0.535]	[0.540]
Linguistic distance squared	3.385***	2.955***	2.300***	1.559***
	[0.416]	[0.410]	[0.407]	[0.429]
Log GDP country 1	1.252***	0.822***	0.841***	0.880***
	[0.0356]	[0.0572]	[0.0560]	[0.0607]
Log GDP country 2	0.898***	0.590***	0.623***	0.704***
	[0.0363]	[0.0491]	[0.0483]	[0.0518]
Log population country 1	-0.385***	0.0971	0.0873	-0.00383
	[0.0476]	[0.0699]	[0.0683]	[0.0733]
Log population country 2	-0.175***	0.165***	0.147**	-0.0114
	[0.0448]	[0.0592]	[0.0581]	[0.0620]
Border	0.402**	0.411**	0.133	0.125
	[0.166]	[0.162]	[0.161]	[0.161]
Avg total years of education	0.209***	0.158***	0.152***	0.164***
	[0.0340]	[0.0340]	[0.0332]	[0.0356]
Log physical distance	-1.091***	-1.142***	-1.114***	-1.145***
	[0.0414]	[0.0409]	[0.0401]	[0.0499]
Geography controls	Yes	Yes	Yes	Yes
Institutional controls		Yes	Yes	Yes
Colonial controls			Yes	Yes
Regional dummies				Yes
Constant	-20.39***	-17.69***	-18.98***	-17.70***
	[0.893]	[0.982]	[0.974]	[1.019]
Observations	1,798	1,798	1,798	1,798
R-squared	0.749	0.761	0.772	0.782

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

times more service trade. But once physical distance is added, the magnitude and significance of the border dummy drop, though the border dummy is still significant at the 5% level and suggestive of an increase in service exports of about 45% when there is

a shared border. Also, the border dummy and physical distance are likely to have a combined effect, since oftentimes a change in the border dummy would be accompanied by an increase in physical distance. Since border and physical distance have an effect on service trade, this suggests that there is still a proximity burden of some kind despite increasing communications technology, though not as large as the one suggested by the border dummy alone.

The regression in column 4 is the main specification for this study. In interpreting the magnitude of the coefficients for linguistic distance, it is important to know the beginning value as well as the size of the change being considered. Because the standard deviation for linguistic distance is 0.271, an example of a decrease in standard deviation in linguistic distance would be a change from Russian learning Danish to Russian learning Czech, another Slavic language. Using the estimates from the main gravity equation, this approximate standard deviation decrease in linguistic distance would correspond to a 17.1% increase in service trade, holding all other factors constant. Similarly, the change from Finnish learning Hungarian, two languages that only share one level of common classification, to Finnish learning Estonian, which share two common classifications, represents an approximate standard deviation decrease in linguistic distance. This change would correspond to a 30.9% increase in service trade.[3]

As for the other control variables, the variables for GDP and population control for the size of an economy and nation and the coefficients are all significant at the 1% level. The positive coefficients on the GDP variables are as expected since the larger two countries are, the more likely they are to trade with one another. However, the change in

[3] For reference, the linguistic distance between Russian and Danish is 0.5 and between Russian and Czech is 0.293. The linguistic distance between Finnish and Hungarian is 0.423 and between Finnish and Estonian is 0.184.

magnitude as more control variables are added show the effect from GDP changing from elastic to inelastic, so these regressions do not have robust enough results to comment on the elasticity of GDP on service exports. Population is also a control for the size of an economy, since a large domestic economy by itself can lead to a smaller amount of trade due to the country being more self-sufficient. Therefore, the negative coefficients on population are as expected and explicitly show that an increase in population while keeping GDP constant will have a negative effect on trade. This suggests that service trade rises less than proportionately with an increase in the size of a nation. Also, since an increase in GDP while holding population constant relates to an increase in income level of a country, GDP per capita is also a variable of interest for service trade. Table 6 shows the results for GDP per capita along with GDP, an alternative specification suggested by Frankel (1997). The results show that GDP per capita is positive and significant at the 1% level in the main gravity equation in column 1, leading to the conclusion that countries with a higher level of income trade more in services. However, these results are not robust as the additional controls are added in columns 2-5.

The education variable in table 5 suggests that a higher average level of education between two partners leads to a greater level of service trade. The positive significant coefficient on average years of total schooling fluctuates between 0.152 and 0.264 depending on the controls included in the regression. This translates to an increase of approximately 20-35% in the level of service trade as the average level of education increases by a standard deviation.

While the examples of changes above show that linguistic distance has a negative effect on service trade, the coefficients suggest that eventually the negative effect of

Table 6. OLS estimates of the effect of linguistic distance on service exports with GDP per capita as a control

Log services trade	1	2	3	4	5
Linguistic distance	-3.846***	-4.216***	-3.670***	-2.806***	-2.289***
	[0.533]	[0.548]	[0.539]	[0.535]	[0.540]
Linguistic distance squared	3.024***	3.377***	2.945***	2.288***	1.554***
	[0.401]	[0.416]	[0.410]	[0.407]	[0.429]
Log GDP country 1	0.773***	0.867***	0.919***	0.928***	0.875***
	[0.0221]	[0.0325]	[0.0336]	[0.0329]	[0.0331]
Log GDP country 2	0.678***	0.723***	0.755***	0.771***	0.693***
	[0.0209]	[0.0301]	[0.0306]	[0.0299]	[0.0308]
Log GDP per capita country 1	0.518***	0.387***	-0.0971	-0.0869	0.00587
	[0.0369]	[0.0477]	[0.0700]	[0.0684]	[0.0734]
Log GDP per capita country 2	0.264***	0.175***	-0.166***	-0.149**	0.0111
	[0.0353]	[0.0449]	[0.0594]	[0.0583]	[0.0621]
Border	0.375**	0.404**	0.413**	0.135	0.127
	[0.162]	[0.166]	[0.162]	[0.162]	[0.161]
Avg total years of education	0.171***	0.208***	0.158***	0.153***	0.164***
	[0.0310]	[0.0341]	[0.0340]	[0.0333]	[0.0356]
Log physical distance	-1.050***	-1.090***	-1.141***	-1.113***	-1.145***
	[0.0364]	[0.0414]	[0.0410]	[0.0401]	[0.0499]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-18.40***	-20.41***	-17.69***	-18.99***	-17.71***
	[0.759]	[0.894]	[0.983]	[0.975]	[1.019]
Observations	2,092	1,797	1,797	1,797	1,797
R-squared	0.731	0.749	0.761	0.773	0.782

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

linguistic distance will level out and even turn positive. For example, using the main specification in column 4, a standard deviation increase in linguistic distance at the mean

will lead to a 10.3% increase in service trade. This is the opposite effect as expected, since a larger linguistic distance should lead to less service trade due to communication having higher costs in the form of language learning. In column 4, a minimum occurs at 0.637, so service exports would be increasing with respect to linguistic distance at the mean. While it can be expected that the influence from linguistic distance would level off, the increase in service trade at the end is a surprising result. This would suggest that countries that have a linguistic distance of one could possibly have a higher level of trade than nations with a medium linguistic distance.[4]

One reason for the interesting result in regard to linguistic distance could be due to the distribution of the variable. Linguistic distance has a range from 0 to about 0.684 and then a discontinuity until 1. The amount of observations at 1 causes the mean of the variable to be 0.700, so that the only observations in the top half of the distribution are equal to 1. The reason that linguistic distance begins to have a positive influence on service trade could be due to all the observations where linguistic distance equals one causing the best fit line to begin to have a positive slope. In order to correct for this discontinuity, a dummy variable for the two languages belonging to different language families was created. Then the linguistic distance variable was recalculated to only record a value when the dummy variable equaled zero.[5] The regression results using this correction are shown in table 7.

With this correction, linguistic distance and linguistic distance squared were both still significant at either the 1% or 5% level, except for linguistic distance squared in the

[4] A linguistic distance of one for this variable means that the two languages do not share any branches and are from different language families.

[5] Formally, the new linguistic distance is equal to the old linguistic distance times the value of one minus the different language family dummy variable.

Table 7. OLS estimates of corrected linguistic distance on level of service exports

Log service exports	1	2	3	4	5
Linguistic distance	-4.332***	-4.565***	-3.846***	-2.308**	-2.529**
	[1.006]	[1.039]	[1.018]	[1.014]	[1.000]
Linguistic distance squared	3.745***	3.896***	3.208**	1.549	1.911
	[1.322]	[1.381]	[1.351]	[1.340]	[1.322]
Different language family dummy	-0.881***	-0.877***	-0.744***	-0.464**	-0.762***
	[0.193]	[0.198]	[0.194]	[0.193]	[0.195]
Log GDP country 1	1.283***	1.250***	0.820***	0.846***	0.879***
	[0.0350]	[0.0369]	[0.0579]	[0.0567]	[0.0611]
Log GDP country 2	0.937***	0.898***	0.589***	0.622***	0.705***
	[0.0341]	[0.0364]	[0.0492]	[0.0484]	[0.0520]
Log population country 1	-0.515***	-0.384***	0.0985	0.0838	-0.0048
	[0.0383]	[0.0482]	[0.0703]	[0.0687]	[0.0735]
Log population country 2	-0.258***	-0.175***	0.167***	0.149**	-0.012
	[0.0355]	[0.0449]	[0.0594]	[0.0583]	[0.0623]
Border	0.366**	0.402**	0.412**	0.137	0.125
	[0.163]	[0.166]	[0.162]	[0.162]	[0.161]
Avg total years of education	0.173***	0.209***	0.159***	0.152***	0.165***
	[0.0311]	[0.0341]	[0.0340]	[0.0333]	[0.0358]
Log physical distance	-1.049***	-1.090***	-1.141***	-1.114***	-1.145***
	[0.0365]	[0.0414]	[0.0410]	[0.0401]	[0.0500]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-18.18***	-20.34***	-17.66***	-19.08***	-17.67***
	[0.778]	[0.910]	[0.997]	[0.989]	[1.031]
Observations	2,092	1,797	1,797	1,797	1,797
R-squared	0.731	0.749	0.761	0.773	0.782

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

last two regressions. The dummy variable for different language families was also significant at either the 1% or 5% level for all specifications. The negative coefficient on this variable is as expected, since countries that have a linguistic distance of one have an overall lower level of service trade. Furthermore, the coefficients in column 1 show that the minimum for the nonlinear effect occurs at 0.578, below the maximum value. Even with the correction for the discontinuity in the linguistic distance variable, the data still indicate that the change in service trade with respect to linguistic distance turns positive with the larger values of linguistic distance. Therefore, a simple explanation of distance between majority languages of two nations is not sufficient to explain why, after a certain threshold, linguistic distance begins to have a positive effect on service trade.

Service Trade and English Distance

The results from the nonlinearity with linguistic distance indicate that another phenomenon is occurring in relation to language and service trade. In Ku and Zussman (2010) the authors go beyond the traditional treatment of the language barrier and look at the possibility that trading partners can communicate in a non-native language to facilitate trade (251). They specifically examined the role of English in trade and found that it promoted trade even among non-native speakers. The use of a vehicle language lowers trading costs by designating only one language that workers must learn and master rather than an almost infinite number as trading partners increase.

In order to test for the use of English as a vehicle language with service trade, I use the constructed English distance variable to test for an effect on level of service trade. Although the variable is simply referred to as English distance, it is the product of the no common language dummy and average distance to English for the majority languages in

the country pair, as described in the Data chapter. The variable is designed to show the effect that distance to English has for service trade between a country pair when the countries lack a common language with which to communicate. Table 8 shows regression results for English distance on service trade.

Table 8. OLS estimates of the effect of English distance on service exports

Log service exports	1	2	3	4	5
No common language * Combined English distance	-1.527***	-1.795***	-1.253***	-0.829***	-1.121***
	[0.221]	[0.242]	[0.246]	[0.245]	[0.251]
Log GDP country 1	1.264***	1.217***	0.832***	0.848***	0.871***
	[0.0343]	[0.0367]	[0.0577]	[0.0563]	[0.0606]
Log GDP country 2	0.933***	0.891***	0.615***	0.639***	0.714***
	[0.0341]	[0.0364]	[0.0498]	[0.0489]	[0.0518]
Log population country 1	-0.515***	-0.335***	0.0886	0.0827	0.00918
	[0.0372]	[0.0487]	[0.0705]	[0.0688]	[0.0732]
Log population country 2	-0.250***	-0.118***	0.168***	0.153***	-0.0149
	[0.0355]	[0.0450]	[0.0594]	[0.0582]	[0.0620]
Border	0.471***	0.577***	0.587***	0.263	0.21
	[0.160]	[0.163]	[0.160]	[0.160]	[0.159]
Avg total years of education	0.140***	0.175***	0.141***	0.141***	0.158***
	[0.0314]	[0.0344]	[0.0343]	[0.0335]	[0.0355]
Log physical distance	-1.037***	-1.032***	-1.091***	-1.072***	-1.142***
	[0.0361]	[0.0401]	[0.0401]	[0.0392]	[0.0492]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-17.97***	-20.20***	-17.77***	-19.17***	-17.48***
	[0.771]	[0.906]	[0.994]	[0.985]	[1.032]
Observations	2,093	1,798	1,798	1,798	1,798
R-squared	0.729	0.747	0.758	0.77	0.781

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

The other control variables within the regressions have similar magnitudes and levels of significance as in the regressions for linguistic distance. The English distance measure is significant at the 1% level and is robust to the addition of further controls. The coefficient on English distance in column 1 indicates that a standard deviation decrease in the averaged distance to English for the two nations results in a 20.8% increase in service trade.[6] A change in the average distance to English could be one or both countries in a country pair changing so that their majority languages are closer to or farther away from English. For example, German and Hungarian are two languages which have the mean English distance, specifically the average of how far those two languages are from English. A standard deviation decrease in English distance would be like switching to German and Swedish, a language more closely related to English. So here, English distance would be a determinant in the different amount of trade between Germany and Hungary, and Germany and Sweden. Specifically, the change in English distance corresponds to a 19.1% increase in service exports.

The magnitude of the coefficient on English distance fluctuates, however, as the additional controls are added in. The inclusion of more geography controls causes the magnitude to increase on English distance, but as the other controls are added, the magnitude decreases to -1.121. This corresponds with only a 15.3% increase for a standard deviation decrease in English distance. Overall, the coefficients suggest that the decrease in English distance is correlated with a range of 11.3-24.5% increase in service trade.

[6] The standard deviation of English distance is 0.1364541.

Service Trade as a Share of GDP

Language matters in regards to the level of service trade, but does it also have a strong significance for the importance of service trade in an economy? I measure the importance of service exports with a particular country in an economy as a share of GDP of the exporting nation. The results for the main gravity equations using service trade as share of GDP are in table 9.

As before, linguistic distance and linguistic distance squared are both significant, though linguistic distance squared is only significant at the 5% level. The coefficients show that a standard deviation decrease at the mean in linguistic distance would lead to a 0.17 percentage point increase in service exports as a share of GDP. That would translate to an approximate $941 million increase in the level of service trade for the mean GDP[7]. The coefficients also show that the minimum from the nonlinearity occurs at 0.959, so at the end of our range of observation. Therefore, linguistic distance has a negative effect on services exports as a share of GDP until linguistic distance reaches its upper reaches. Again, this would only affect the languages that belong to different language families, as their linguistic distance is equal to one. So as a nation has an increasingly dissimilar language from the exporting nation, the language barrier will cause service exports to have less importance in an economy until the languages have a distance of one, where the language barrier stops having a negative effect. Again, this suggests the use of a vehicle language such as English to lower language costs. The significance of the linguistic distance variables remains steady as the additional controls are added. The full tables for the share of GDP results are located in appendix table B1.

[7] The exporting nation mean GDP is $554 billion with a standard deviation of $30.1 billion.

Table 9. OLS estimates of the effect of linguistic and English distance on service exports as a share of GDP

Service exports as share of GDP	1	2
Linguistic distance	-0.0236***	
	[0.00710]	
Linguistic distance squared	0.0123**	
	[0.00534]	
No common language * Combined English distance		-0.0180***
		[0.00292]
Log GDP country 1	0.000147	-0.000391
	[0.000447]	[0.000454]
Log GDP country 2	0.00379***	0.00376***
	[0.000453]	[0.000452]
Log population country 1	-0.00279***	-0.00228***
	[0.000497]	[0.000494]
Log population country 2	-0.000909*	-0.000893*
	[0.000471]	[0.000470]
Border	0.0137***	0.0139***
	[0.00217]	[0.00213]
Avg total years of education	-0.00122***	-0.00152***
	[0.000412]	[0.000415]
Log physical distance	-0.00364***	-0.00406***
	[0.000485]	[0.000478]
Constant	0.0151	0.0246**
	[0.0101]	[0.0102]
Observations	2,103	2,103
R-squared	0.189	0.193

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

The variable for English distance is consistently significant at the 1% level both in the main gravity equation in table 9, column 2 as well as when additional controls are added for robustness.[8] This shows that English impacts services exports as a share of GDP through its use as a vehicle language. Specifically it shows that as two nations have

[8] See appendix table B2 for full regression results.

easier access to English through language transfer, service trade will have a greater importance in the economy. The coefficient on the English distance variable suggests that a standard deviation decrease in English distance corresponds to a 0.25% point increase in services exports to a particular country as a share of GDP. Using the mean GDP, this would translate to an increase of service exports by $1.4 billion.

The different control variables in the main gravity equation have varying degrees of significance throughout the different regression specifications. Only the border dummy and physical distance are consistently significant at the 1% level through all the regressions. In both equations, the positive coefficient on the border dummy affirms the common assumption that service exports to a country that shares a border will be more important than to countries which do not. Holding physical distance and other factors constant, sharing a border leads to greater level of trade equal to 1.4% of GDP for the exporting nation. Also, a standard deviation decrease in physical distance would lead to a 0.38 percentage point increase in services exports as a share of GDP. While both the level of trade regressions and the share of GDP regressions show that physical distance and the border effect matter for service trade, does it have a different effect than in commodity trade? Also, does language have the same type of effect in services and commodity trade? The next section will compare how language and physical distance affect trade levels for goods and services.

Comparison of Service and Commodity Trade

In general, we can predict that service trade is likely to be more dependent on communication because production and consumption cannot be separated as in commodity trade. Since there is a larger dependence on communication, the variables for

linguistic distance between countries and English distance would have a larger effect on service trade, as the relative difficulty of learning a new language would hinder trade more in this type of trade. Also, service trade is likely to be impacted less by physical distance, as compared to commodity trade where goods have to be shipped. The first comparison for how language affects trade in services versus in goods is to look at the simple gravity equation with the no common language dummy used in table 4. The results for both types of trade are listed in table 10. Overall, the coefficients on the no common language dummy in all the regressions and on the physical distance variable in columns 2 and 4 are significant at the 1% level. The absolute value of magnitudes of the coefficients on the no common language variable are significantly larger for the service trade regressions at the 10% level in columns 1 and 3 and at the 5% level in columns 2 and 4.[9] In columns 2 and 4, the lack of a common language is correlated with 68.9% less trade in services and 58.1% less trade in commodities. In columns 2 and 4, physical distance has a larger effect on commodity trade, which is statistically larger than the effect on service trade at the 1% level. While these regressions are lacking many of the control variables used in this study, table 10 shows the basic pattern of the difference between services and commodities trade.

The results to the main gravity equation with linguistic distance and linguistic distance squared are presented in table 11. Both linguistic distance variables and the physical distance variable are all significant, allowing for comparisons between the two types of trade. The regressions with the additional controls for robustness checks also

[9] The equation I use to test the null hypothesis that $\beta_1 = \beta_2$ is $Z = (\beta_1 - \beta_2)/\sqrt{(\sigma_{\beta 1}^2 + \sigma_{\beta 2}^2)}$ from Paternoster et al. (1998).

Table 10. OLS estimates of the effect of no common language on service and commodity exports

Log exports	Services		Commodities	
	1	2	3	4
No common language	-1.037***	-1.169***	-0.731***	-0.871***
	[0.158]	[0.131]	[0.153]	[0.114]
Log GDP country 1	1.237***	1.253***	0.947***	0.969***
	[0.0430]	[0.0356]	[0.0415]	[0.0307]
Log GDP country 2	1.100***	1.046***	0.857***	0.815***
	[0.0347]	[0.0288]	[0.0343]	[0.0254]
Log population country 1	-0.627***	-0.498***	-0.0071	0.128***
	[0.0467]	[0.0388]	[0.0453]	[0.0337]
Log population country 2	-0.592***	-0.391***	-0.190***	0.0496*
	[0.0350]	[0.0297]	[0.0343]	[0.0261]
Physical distance		-1.097***		-1.255***
		[0.0343]		[0.0301]
Constant	-20.62***	-16.17***	-22.81***	-18.25***
	[0.907]	[0.764]	[0.886]	[0.665]
Observations	2,251	2,249	2,191	2,189
R-squared	0.566	0.702	0.56	0.757

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

show the same significance levels and general trends.[10] The coefficients for physical distance suggest that a 1% decrease in physical distance is correlated with a 1.05% increase in service trade and a 1.17% increase in commodities trade. Also, there is a larger border effect with commodities trade than with services, amounting to a 56.5% difference in commodities trade and only 36.9% with services.

A simple comparison with the linguistic distance variables is a little more difficult considering the nonlinearity. The best mode of comparison is through again using the example of a standard deviation decrease in linguistic distance from Russian learning

[10] For the full results of the comparison of the effect of linguistic distance on services and commodities trade, see appendix table B3.

Table 11. OLS estimates of the effect of linguistic distance on service and commodity exports

Log exports	Services	Commodities
Linguistic distance	-3.839***	-2.905***
	[0.533]	[0.474]
Linguistic distance squared	3.014***	2.672***
	[0.401]	[0.356]
Log GDP country 1	1.287***	0.986***
	[0.0337]	[0.0292]
Log GDP country 2	0.936***	0.637***
	[0.0340]	[0.0302]
Log population country 1	-0.518***	0.130***
	[0.0373]	[0.0324]
Log population country 2	-0.258***	0.198***
	[0.0355]	[0.0313]
Border	0.369**	0.565***
	[0.163]	[0.142]
Avg total years of education	0.172***	0.264***
	[0.0310]	[0.0277]
Log physical distance	-1.050***	-1.167***
	[0.0364]	[0.0319]
Constant	-18.26***	-20.22***
	[0.760]	[0.666]
Observations	2,093	2,048
R-squared	0.73	0.789

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Danish to Russian learning Czech. This would lead to a 17.1% increase in service trade and only a 4.8% increase in commodity trade. Also, an approximate standard deviation decrease in linguistic distance from Finnish learning Hungarian to Finnish learning Estonian would cause a 30.9% increase in service trade and a 15.5% increase in commodity trade. These two examples are evidence that linguistic distance has a larger effect on service trade.

The coefficients for linguistic distance for both types of trade again point out the fact that the effect from linguistic distance levels out and turns positive as linguistic distance increases. This phenomenon is explained in this paper by the occurrence of a vehicle language for trade, namely English. Both types of trade have a minimum that occurs within the range of linguistic distance. In service trade, the minimum occurs when linguistic distance is equal to 0.637 and in commodities trade when it is equal to 0.544. This result suggests that service trade and commodity trade both experience some intervening factor, such as a vehicle language, that facilitates trade when languages are extremely distant.

Table 12 compares the effect of English distance and physical distance on services and commodities trade. Again, physical distance is significant at the 1% level for all regressions. The coefficients on physical trade show that physical distance has a greater effect on commodities trade which is about 10-15% points higher than on service trade. As for English distance, it is significant at the 1% level for service trade, but has varied significance for commodities trade. In the regressions shown in table 12, the coefficient in column 3 is barely insignificant with a P value of 0.106. However, in column 4 the coefficient switches signs and has a P value of 0.944. The variable only has significance when the geography controls are added to the main gravity equation. [11] The English distance variable has a relatively large effect for service trade; a standard deviation decrease from the mean would lead to a 20.8% increase in service trade for regression 1 and a 15.3% increase for regression 2. A comparable amount for commodity trade does not exist since the coefficients are not significant. However, using a Z test to

[11] See appendix table B4 for the full results of the comparison of the effect of English distance on service and commodity exports.

check that the coefficients in column 1 and 3 do not equal each other, we can say that at the 1% level the effect of English distance on service trade is significantly higher than the effect on commodity trade.

Table 12. OLS estimates of the effect of English distance on service and commodity exports

Log exports	Services		Commodities	
	1	2	3	4
No common language *	-1.527***	-1.121***	-0.317	0.0145
Combined English distance	[0.221]	[0.251]	[0.196]	[0.208]
Log GDP country 1	1.264***	0.871***	1.004***	0.780***
	[0.0343]	[0.0606]	[0.0303]	[0.0504]
Log GDP country 2	0.933***	0.714***	0.638***	0.566***
	[0.0341]	[0.0518]	[0.0308]	[0.0432]
Log population country 1	-0.515***	0.00918	0.0959***	0.300***
	[0.0372]	[0.0732]	[0.0330]	[0.0609]
Log population country 2	-0.250***	-0.0149	0.202***	0.270***
	[0.0355]	[0.0620]	[0.0319]	[0.0516]
Border	0.471***	0.21	0.692***	0.345***
	[0.160]	[0.159]	[0.142]	[0.132]
Avg total years of education	0.140***	0.158***	0.252***	0.253***
	[0.0314]	[0.0355]	[0.0285]	[0.0304]
Log physical distance	-1.037***	-1.142***	-1.118***	-1.279***
	[0.0361]	[0.0492]	[0.0321]	[0.0412]
Geography controls		Yes		Yes
Institutional controls		Yes		Yes
Colonial controls		Yes		Yes
Regional dummies		Yes		Yes
Constant	-17.97***	-17.48***	-20.91***	-19.88***
	[0.771]	[1.032]	[0.687]	[0.855]
Observations	2,093	1,798	2,048	1,759
R-squared	0.729	0.781	0.78	0.81

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

The lack of a significant effect on commodities trade could reflect relative lack of importance of language in regards to that type of trade. For example, companies performing commodities trade with other nations oftentimes go through intermediaries for specific business ventures, such as the use of external translators or contract manufacturers like FoxConn. For example, in a survey of small to medium sized firms across the European Union, 45% of firms employed external translators or interpreters to support business with other nations (CILT 2006, 31). While there are numerous methods for getting around language barriers, it is more difficult in service trade because of the greater weight on communication due to production and trade being coupled together. In commodities trade where the production process and trading process can be completely separate, language oftentimes just represents an extra cost with trading. However, in service trade, I propose, a lack of language skills is an inhibiting factor to production and trade in services occurring between two nations.

Comparison of Service Trade Aggregates

While linguistic and English distance have had a significant effect on service trade, the magnitude of the variables have varied greatly depending on which control variables are added to the regression. This could be due to aggregation issues, since the aggregated service trade measure from the UN Service trade database combines a wide range of services. One of the standard ways for classifying services has been the Extended Balance of Payments Services classification (EBOPS 2002), developed in 1996 by the OECD and Eurostat (United Nations 2011a). The Total EBOPS Services category is divided into 11 subcategories: transportation; travel; communication services; construction services; insurance services; financial services; computer and information

services; royalties and license fees; other business services; personal, cultural, and recreational services; and government services not included elsewhere (United Nations 2011b). These listed subcategories make up a wide range of economic activities that differ greatly from one another.

Another system for categorizing services was described by Jenson (2011) based upon the North American Industry Classification System (NAICS). He divided all economic activity into five broad categories: goods production, trade activities, business services, personal services, and public services. The main two categories that he focused on in his book on global service trade were business services and personal services. The category for business services included "information; finance and insurance; real estate; professional, scientific, and technical services; management of companies and enterprises; and administrative and waste remediation services (NAICS sectors 51 through 56)" (16). Personal services included "educational services; health care and social assistance; arts, entertainment, and recreation services; and accommodation and food services (NAICS sectors 61, 62, 71, 72, and 81)" (16). Interestingly, construction was considered a part of goods production and transportation would likely be considered to be a part of trade activities. Government services would likely fall under Jensen's category of public services and there is not a clear location for the royalties and license fees, though these more likely would be included in a balance of payments classification and not in an economic activity classification like the one in Jensen (2011). It is clear from the different classifications that there is some disagreement over how to classify services. Do these differences in classification stem from different personal opinions or from systematic differences between these services and the type of activity they describe?

In table 13 and table 14, I report results for the regression with all controls for each category of service trade according to the EBOPS classification. Only the coefficients for the language variables and physical distance are reported, though the full results are reported in the appendix tables B5 and B6 respectively. Fewer observations were available due to the smaller number of countries reporting disaggregated service trade data for every category.

Table 13 shows the results of linguistic distance, linguistic distance squared, and physical distance on categories of service trade. Physical distance is significant for all categories of service trade at the 1% level. Linguistic distance is not significant only for transportation and construction, suggesting that these two types of services rely less on communication than the others. That is definitely a credible result, especially considering those two services are not even included as services under Jensen's classification.

Table 14 shows the results for English distance and physical distance. Again, physical distance is significant for all types of service trade. The significance of English distance could signify the services that may more heavily depend on English as a vehicle language for trade. These would be services that are more heavily dependent on communication as well as may require a greater mastery of the language, which would increase the costs of language acquisition. The only categories of services for which English distance is not significant are transportation, construction, royalties and license fees, and government services. Again, the lack of significance on the transportation coefficient makes sense because although transportation is a service, it deals with transporting goods that can be produced independently of the service performed. Also, the lack of a significant construction coefficient is expected because it can also be

Table 13. OLS estimates of the effect of linguistic distance on categories of services exports

Log categories of services exports	Transportation	Travel	Communications services	Construction	Insurance services	Financial services
Linguistic distance	0.279	-2.360***	-2.633***	-1.923	-3.697***	-5.175***
	[0.763]	[0.629]	[0.920]	[1.453]	[1.111]	[1.112]
Linguistic distance squared	-0.531	1.400***	1.172	1.3	1.809**	3.188***
	[0.604]	[0.506]	[0.712]	[1.112]	[0.890]	[0.880]
Physical distance	-1.202***	-1.229***	-1.407***	-1.273***	-0.908***	-1.254***
	[0.0700]	[0.0607]	[0.0808]	[0.131]	[0.107]	[0.107]
Observations	1,241	1,238	963	657	797	927
R-squared	0.7	0.729	0.731	0.416	0.6	0.68

Log categories of services trade	Computer services	Royalties and license fees	Other business services	Personal, etc. services	Government services
Linguistic distance	-4.396***	-3.271***	-3.516***	-4.106***	-2.761**
	[0.981]	[0.972]	[0.862]	[1.028]	[1.245]
Linguistic distance squared	2.809***	2.856***	2.519***	2.885***	2.982***
	[0.779]	[0.767]	[0.663]	[0.806]	[1.000]
Physical distance	-1.185***	-0.812***	-1.236***	-0.995***	-1.095***
	[0.0926]	[0.0922]	[0.0747]	[0.0973]	[0.124]
Observations	936	766	1,073	714	670
R-squared	0.647	0.732	0.734	0.608	0.551

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1; regressions also control for GDP, population, shared borders, average years of schooling, and additional controls for robustness

Table 14. OLS estimates of the effect of English distance on categories of services exports

Log categories of services exports	Transportation	Travel	Communications services	Construction	Insurance services	Financial services
No common language *	-0.414	-2.363***	-1.681***	0.0696	-3.518***	-4.414***
Combined English distance	[0.353]	[0.299]	[0.426]	[0.676]	[0.491]	[0.503]
Physical distance	-1.222***	-1.221***	-1.437***	-1.287***	-0.942***	-1.215***
	[0.0692]	[0.0589]	[0.0803]	[0.129]	[0.103]	[0.102]
Observations	1,241	1,238	963	657	797	927
R-squared	0.699	0.737	0.727	0.415	0.61	0.694

Log categories of services trade	Computer services	Royalties and license fees	Other business services	Personal, etc. services	Government services
No common language *	-2.722***	0.54	-1.535***	-1.401***	0.457
Combined English distance	[0.465]	[0.481]	[0.386]	[0.528]	[0.629]
Physical distance	-1.155***	-0.778***	-1.219***	-0.994***	-1.005***
	[0.0911]	[0.0918]	[0.0737]	[0.0956]	[0.123]
Observations	936	766	1,073	714	670
R-squared	0.649	0.728	0.733	0.602	0.541

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1; regressions also control for GDP, population, shared borders, average years of schooling, and additional controls for robustness

thought of as commodity trade given that a good is being produced. However, since the production of a good is tied to the trade of it, it can also be thought of as a service. I would suggest that construction and transportation are hybrid economic activities that do not fit purely into commodity or service classifications. Regardless of what they are called, they appear to be systematically different than many of the other categories of services in the area of communication needed. The lack of significance for royalties and license fees as well as for governmental services is also believable since these may be services that require less face to face interaction, in the case of royalties and license fees, or are performed more often in official languages for political reasons, in the case of governmental services.

The categories of service trade that did not show significance for the language variables were the same categories that did not fit in the two service categories designated by Jensen (2011). Travel was one service category in EBOPS that was hard to classify in Jensen's classification as either personal or business related. However, the larger travel category can be further divided into business and personal travel, which would then fit in the respective category. In order to solve for aggregation problems between different types of services, I added the services together that fit in Jensen's classification of business and personal services. It is believed that these services include more of what is thought of as service trade, and thus provides a better aggregation. The regression results from using business and personal services rather than EBOPS service aggregate can be seen in tables 15 and 16.

Table 15 shows the results of the effect of linguistic distance and linguistic distance squared on business and personal services. The linguistic distance variables are

Table 15. OLS estimates of the effect of linguistic distance on business and personal service export aggregate

Log business and personal service exports	1	2	3	4	5
Linguistic distance	-3.164***	-3.261***	-3.073***	-2.672***	-3.127***
	[0.734]	[0.749]	[0.742]	[0.722]	[0.751]
Linguistic distance squared	1.819***	2.043***	1.821***	1.462**	1.950***
	[0.574]	[0.585]	[0.584]	[0.569]	[0.602]
Log GDP country 1	0.683***	0.750***	0.458***	0.477***	0.425***
	[0.0546]	[0.0617]	[0.101]	[0.0979]	[0.113]
Log GDP country 2	0.593***	0.617***	0.379***	0.441***	0.410***
	[0.0587]	[0.0622]	[0.0920]	[0.0897]	[0.0985]
Log population country 1	0.023	-0.0328	0.295**	0.296**	0.317**
	[0.0649]	[0.0821]	[0.125]	[0.121]	[0.131]
Log population country 2	0.0189	0.00986	0.254**	0.220**	0.243**
	[0.0632]	[0.0774]	[0.108]	[0.105]	[0.110]
Border	0.341**	0.231	0.273	0.08	0.01
	[0.171]	[0.174]	[0.174]	[0.172]	[0.172]
Avg total years of education	0.140***	0.159***	0.124**	0.114**	0.0875
	[0.0513]	[0.0546]	[0.0558]	[0.0542]	[0.0580]
Log physical distance	-0.837***	-0.951***	-0.928***	-0.923***	-0.986***
	[0.0638]	[0.0729]	[0.0722]	[0.0699]	[0.0815]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-8.587***	-9.136***	-6.905***	-8.656***	-6.185***
	[1.194]	[1.293]	[1.464]	[1.456]	[1.618]
Observations	486	460	460	460	460
R-squared	0.669	0.679	0.689	0.71	0.721

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Table 16. OLS estimates of the effect of English distance on business and personal service export aggregate

Log business and personal service exports	1	2	3	4	5
No common language * Combined English distance	-2.998***	-2.914***	-2.773***	-2.422***	-2.470***
	[0.361]	[0.382]	[0.395]	[0.395]	[0.393]
Log GDP country 1	0.689***	0.719***	0.582***	0.588***	0.495***
	[0.0527]	[0.0599]	[0.0987]	[0.0967]	[0.110]
Log GDP country 2	0.545***	0.542***	0.440***	0.487***	0.435***
	[0.0568]	[0.0608]	[0.0893]	[0.0882]	[0.0962]
Log population country 1	0.0191	0.0162	0.161	0.174	0.231*
	[0.0625]	[0.0799]	[0.122]	[0.120]	[0.129]
Log population country 2	0.0531	0.083	0.182*	0.161	0.200*
	[0.0612]	[0.0755]	[0.106]	[0.104]	[0.108]
Border	0.311*	0.26	0.289*	0.13	0.064
	[0.160]	[0.162]	[0.162]	[0.163]	[0.163]
Avg total years of education	0.0495	0.0931*	0.0839	0.0762	0.0745
	[0.0506]	[0.0534]	[0.0547]	[0.0537]	[0.0563]
Log physical distance	-0.859***	-0.942***	-0.940***	-0.952***	-1.037***
	[0.0600]	[0.0667]	[0.0665]	[0.0652]	[0.0767]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-6.995***	-7.089***	-5.753***	-7.569***	-4.949***
	[1.186]	[1.301]	[1.455]	[1.478]	[1.612]
Observations	486	460	460	460	460
R-squared	0.69	0.698	0.702	0.716	0.73

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

significant at the 1% level in practically all specifications, so it appears that there is a nonlinear effect for business and personal services. In addition, the magnitudes of the variables do not fluctuate as greatly as in table 7. For column 1, a standard deviation decrease in linguistic distance at the mean is correlated with a 16.7% increase in service exports. In column 5, the standard deviation decrease at the mean corresponds to a 10.7% decrease in service exports. The effect of linguistic distance on service trade reaches a minimum and turns positive when linguistic distance is equal to 0.798-0.914 in the regressions in table 16. This further suggests that countries with majority languages from different language families, the extreme of linguistic distance, could potentially trade more than countries with moderately distant languages since they may be more likely to use a common language rather than learning each others' language.

Also important to note are the coefficients on GDP and population. While previously the coefficients for GDP were close to 1, in table 15 they are still significant but range from around 0.38 to 0.75. Where before one could have interpreted that the GDP has a more than proportionate effect on service exports, now it is clearly less than proportionate, or inelastic. I believe this interpretation is reasonable since changes in GDP occur within both the domestic and the export/import markets. Also, before the positive coefficients on GDP and the negative coefficients on population suggested that income level was a significant determinant of service exports, but now the coefficient on population is positive if it is significant, which depends on the specification. This pattern is consistent across tables 16 as well. These results suggest that income is not important for business and personal services once other factors are taken into account.

Table 16 shows the effect of English distance on business and personal services. The significance of the coefficients for the English distance variable show that English as a common language does have an effect on these services. The magnitudes of the coefficients are greater than those in table 8, suggesting that English has an even greater effect than indicated previously. Using column 5 estimates from tables 8 and 16, a standard deviation decrease in English distance corresponds with a 15.3% increase in the EBOPS service aggregate and a 31.3% increase in the business and personal services aggregate. This is a reasonable observation considering the excluded services were those that are less dependent on communication in their operation, particularly a common language. Overall, the use of the business and personal service aggregate in place of the EBOPS aggregate lessened the sensitivity of the coefficient magnitudes, since hybrid economic activities were excluded that could have been causing the coefficients to be more sensitive to control variables. Also, the general effects of the language variables were strengthened with this new aggregate measure as the dependent variable.

CHAPTER SIX

Discussion and Conclusion

Our results confirm that language and communication matter for service trade. By using a richer variable for linguistic distance, this paper indicates that the distance between languages has a strong effect on service trade. Previous dummy variables for a common language in essence assume that all distances between languages are constant and equal to one. It is important to notice that the coefficients in table 15 suggest that an increase in linguistic distance from 0 to 1, much like with a dummy variable, would result in 52-60% less trade. This amount is extremely similar to previous estimates for the effect of the language dummy variable in the literature (Kimura and Lee 2006; Francois and Hoekman 2010). However, our measures add to this calculation by indicating how smaller changes in linguistic distance also have a significant effect.

The significance of a nonlinear effect of linguistic distance on service and commodity trade is an interesting addition to the effect of relative language differences. While the significant coefficients showed that linguistic distance did have an effect on trade, it also indicated that as linguistic distance continued to increase, it eventually had a positive effect on service trade. The observation that countries with languages that are completely distant from one another might trade more than countries that have languages that are closer to the middle of the linguistic distance range is certainly intriguing. The diminishing effect of linguistic distance was expected; however, the positive upturn at the end was unexpected yet plausible.

One possible explanation for this effect is the occurrence of language costs and the use of a vehicle language. The initial decline in service trade as linguistic distance increases is reflective of the increasing costs to learning a foreign language for trade. Learning a language is largely thought of as a fixed cost for trade. As linguistic distance increases, the total cost of language acquisition increases as well. Firms will only engage in trade if the value of learning that language, and therefore trading, is more than the cost of learning the language. Firms have various strategies open to them to help communication, such as hiring native speakers or staff with language skills, using external translators, and doing in-house language training, all of which have different costs associated (CILT 2006, 5-6). Language acquisition can become more costly because of the difficulty of the learning process and also because of the likely accompanying shortage of individuals with adequate language skills who are therefore more costly to employ as staff or a translator. As costs increase more firms will elect to not engage in trade or will not have the ability to do so because of a lack of language skills. This causes service trade to decrease.

However, another language strategy for the companies would be to use a vehicle language. Using the target language is preferred because it allows for more contact with the customer and gives companies better information about the market (CILT 2006, 6). But as linguistic distance gets too large, using the target language becomes too costly and companies will switch to a vehicle language. Firms will switch to a vehicle language at different times based on their language resources and firm-specific language costs, causing the costs of increasing linguistic distance to diminish. After a certain point, the majority of companies have now switched to a vehicle language for trade. So as linguistic

distance increases, service trade begins to fall because the ability to communicate is weakening. But as languages reach the extreme of distance, countries switch to a vehicle language, communication is improved, and service trade begins to increase again.

Another way of looking at this would be that language skills are often a determinant of which export market a company targets (CILT 2006, 20). This is especially true for non-English speaking countries. Companies will target export markets where the expected value from business outweighs the cost of doing business there, including learning the native language. Therefore, controlling for other factors, more companies will choose to do business where linguistic distance is low because of lower language costs. Also, companies will do business at the other extreme because it is easier to just use a vehicle language. The vehicle language will still be costly, since if linguistic distance is high at least one or both of the nations has a large distance from English. However, there are efficiency effects with English since they can use English with a wide range of countries. Since using the native language is preferred for the long-run in doing business, the medium languages will have a higher cost for language learning because they might not be deemed "too foreign" to require a vehicle language. Therefore, companies will move forward with trying to do business there but at higher costs and possibly less success.

A key observation to remember in regards to the trade language decision is that although English as a common language can be more efficient for trade, the native language is preferred. In a survey of small and medium sized enterprises in Europe, the following observation was made: “Individual respondents mentioned that English might be used for initial market entry, but longer-term business partnerships depended upon

relationship-building and relationship management and, to achieve this, cultural and linguistic knowledge of the target country were essential" (CILT 2006, 6). As long as there are multiple languages in the world, there will always be a benefit to learning the native language of a target market. While the above is meant as a possible explanation for the nonlinear results, further evidence is needed in order to better understand the dynamics that occur in language acquisition for international trade.

The significant results for English distance corroborate these stories to explain the nonlinear effect of linguistic distance. Also, since good statistics on the use of English as a global business language are lacking, oftentimes the use of English is described as it seems rather than as it actually is (Truchot 2002, 20). So this paper presents further evidence for the use of English as a common language for trade. Because English distance in this paper is an exogenous measure, it avoids many endogeneity problems that are inherent with language learning statistics.[1] By using the interaction between the no common language dummy and the average English distance for the two countries, we get a more accurate depiction of when English is likely to be used as a common language and therefore the significance has all the more weight to it.

Physical distance was significant for both service trade and commodity trade, though it had a larger effect for commodity trade. Nevertheless, physical distance might be significant for different reasons. Physical distance obviously is important for commodity trade since goods must be transported from one location to another. However, for service trade that can be performed over the phone or electronically, the link is a little

[1] Because motivation is a major factor in second language acquisition, many times language statistics such as the number of second language English speakers will have reverse causality problems when applied to trade. It is difficult to determine if the number of English speakers affects trade or if the possibility for economic rewards in areas like trade affects the number of English speakers.

harder to make. One explanation for why physical distance matters for service trade comes from Portes and Rey (2005). In this paper, the authors estimated a gravity model to explain cross-border equity flows. To help explain why physical distance was significant, they also included a variable for time zone overlap and other variables related to the geography of information to see if physical distance was a proxy for information costs. Time zone overlap was significant, suggesting that for both cross-border equity flows, and possibly other forms of trade that deal with intangibles, physical distance matters more because of the information frictions it causes. For example, service trade may be more difficult for two countries on opposite sides of the world because it is difficult to find a time to communicate rather than the actual physical distance between them. We also see anecdotal evidence of this in the Bloomberg Businessweek article, where American companies found Latin America as a useful destination for outsourcing in order to have connections in a closer time zone than other locations such as India (Helyar and Srivastava 2012).

The comparison between service trade and commodities trade also showed that English distance was significant for service trade but had a varied effect on commodity trade. This could reflect the general lack of importance of language for commodity trade because of the use of intermediaries such as contract manufacturers. However, since the linguistic distance variables were significant for commodity trade, this could also reflect the lack of importance of English as the main common language for trade in goods. Survey evidence from Europe suggests that other common languages can be used besides English, such as Russian for trade between Latvia and Lithuania (CILT 2006, 25). Regardless of the cause for the lack of significance, the effect of English distance on

service trade is still significantly higher than the effect for commodity trade, so English distance is more important for service trade.

The estimates for the linguistic distance variables were originally very sensitive to regression specifications and would greatly shift in magnitude when control variables were added to the regression. Part of this sensitivity may be due to inappropriate aggregation between different kinds of services. By using Jensen's framework for classifying economic activity, the new aggregation of business and personal services provided a lot more consistent results. Also, Jensen's classification of services excluded the categories of services that were fundamentally different from the main core of services in regards to communication, as we could see from tables 13 and 14. This distinction in different types of services is useful when the object of study is to ascertain the determinants of service trade, such as in this paper. An aggregation including economic activity that is on the periphery of what is considered to be a typical service may skew results.

An area for possible further work involves examining the effect of language distance using different estimation techniques for robustness. This could include expanding the study here to a panel data set or using the Heckman selection model to account for the sample selection of whether trade even occurs between two nations or for the smaller number of reporting countries available with the UN dataset. The linguistic distance variable could be improved by adding further linguistic detail as in the English distance variable or designing a way to incorporate more than one language per country to cut down on error from limiting the majority languages of a country to one. Also, further research into the nonlinear effect of linguistic distance on trade is needed before

any firm conclusions can be made. It would be interesting to see if it were possible, using this information, to predict when a common language will be used in trade rather than the trading partners attempting to learn each others' languages. Another conceivable exercise would be to use the language variables as instrumental variables for examining the effect of service trade on other economic outcomes such as economic or productivity growth.

In conclusion, the study of trade in goods and services can be enhanced by looking into how language is used to assist trading activity. A lot of information is lost when only a dummy variable for common language is used. Relative language distances are significant for international trade and can also help point out the differences in different types of trade. Service trade is more dependent on communication since the production and consumption of the service occur at the same time. As globalization and liberalization of service trade continues, communication will become even more important and a greater understanding of its function in the global economy will be needed.

APPENDIX A

Table A. Countries and corresponding majority language in country pair regressions

Country	Language	Country	Language
Afghanistan	Dari	Cape Verde	Portuguese
Albania	Albanian	Cayman Isds	English
Algeria	Arabic	Central African Rep.	French
American Samoa	Samoan	Chad	French
Andorra	Catalan	Chile	Spanish
Angola	Portuguese	China	Mandarin
Antigua and Barbuda	English	China, Hong Kong SAR	Cantonese
Argentina	Spanish	China, Macao SAR	Cantonese
Armenia	Armenian	Colombia	Spanish
Aruba	Portuguese	Comoros	French
Australia	English	Congo	French
Austria	German	Costa Rica	Spanish
Azerbaijan	Azerbaijani	Cote d'Ivoire	French
Bahamas	English	Croatia	Serbo-Croatian
Bahrain	Arabic	Cuba	Spanish
Bangladesh	Bengali	Cyprus	Greek
Barbados	English	Czech Republic	Czech
Belarus	Russian	Dem. Rep. of the Congo	French
Belgium	Dutch	Denmark	Danish
Belize	Spanish	Djibouti	French
Benin	French	Dominica	French
Bermuda	English	Dominican Rep.	Spanish
Bhutan	Dzongkha	Ecuador	Spanish
Bolivia	Spanish	Egypt	Arabic
Bosnia Herzegovina	Bosnian	El Salvador	Spanish
Botswana	English	Equatorial Guinea	Spanish
Brazil	Portuguese	Eritrea	Arabic
Brunei Darussalam	Malay	Estonia	Estonian
Bulgaria	Bulgarian	Ethiopia	Amharic
Burkina Faso	French	Faeroe Isds	Faroese
Burundi	French	Fiji	Fijian
Cambodia	Khmer	Finland	Finnish
Cameroon	English	France	French
Canada	English	French Polynesia	French

Table A continued

Country	Language	Country	Language
Gabon	French	Libya	Arabic
Gambia	English	Liechtenstein	German, Swiss
Georgia	Georgian	Lithuania	Lithuanian
Germany	German	Luxembourg	Luxembourgish
Ghana	English	Madagascar	French
Gibraltar	English	Malawi	English
Greece	Greek	Malaysia	Malay
Greenland	Greenlandic	Maldives	Dhivehi (Maldivian)
Grenada	English	Mali	French
Guam (US)	English	Malta	Maltese
Guatemala	Spanish	Marshall Isds	Marshallese
Guinea	French	Mauritania	Arabic
Guinea-Bissau	Portuguese	Mauritius	French
Guyana	English	Mexico	Spanish
Haiti	French	Mongolia	Mongolian
Honduras	Spanish	Montenegro	Montenegrin
Hungary	Hungarian	Morocco	Arabic
Iceland	Icelandic	Mozambique	Portuguese
India	Hindi	Myanmar	Burmese
Indonesia	Indonesian	N. Mariana Isds	English
Iran	Persian	Namibia	English
Iraq	Arabic	Nepal	Nepali
Ireland	English	Netherlands	Dutch
Isle of Man	English	New Caledonia	French
Israel	Hebrew	New Zealand	English
Italy	Italian	Nicaragua	Spanish
Jamaica	English	Niger	French
Japan	Japanese	Nigeria	English
Jordan	Arabic	North Korea	Korean
Kazakhstan	Russian	Norway	Norwegian
Kenya	English	Oman	Arabic
Kiribati	Kiribati	Pakistan	Punjabi
Kuwait	Arabic	Panama	Spanish
Kyrgyzstan	Kyrgyz	Papua New Guinea	English
Palau	Palauan	Paraguay	Spanish
Laos	Lao	Peru	Spanish
Latvia	Latvian	Philippines	Filipino
Lebanon	Arabic	Poland	Polish
Lesotho	English	Portugal	Portguese
Liberia	English	Puerto Rico	Spanish

Table A continued

Country	Language	Country	Language
Qatar	Arabic	Tunisia	Arabic
Rep. of Korea	Korean	Turkey	Turkish
Rep. of Moldova	Romanian	Turkmenistan	Turkmen
Romania	Romanian	Turks and Caicos	English
Russia	Russian	Tuvalu	Tuvaluan
Rwanda	English	Uganda	English
Saint Kitts and Nevis	English	Ukraine	Ukrainian
Saint Lucia	French	United Arab Emirates	Arabic
Saint Vincent and the Grenadines	English	United Kingdom	English
Samoa	Samoan	Uruguay	Spanish
San Marino	Italian	US Virgin Isds	English
Sao Tome and Principe	Portuguese	USA	English
Saudi Arabia	Arabic	Uzbekistan	Uzbek
Senegal	French	Vanuatu	French
Serbia	Serbo-Croatian	Venezuela	Spanish
Seychelles	French	Viet Nam	Vietnamese
Sierra Leone	English	Yemen	Arabic
Singapore	Mandarin	Zambia	English
Slovakia	Slovakian	Zimbabwe	English
Slovenia	Slovenian		
Solomon Isds	English		
Somalia	Somali		
South Africa	English		
Spain	Spanish		
Sri Lanka	Sinhala		
Sudan	Arabic		
Suriname	Dutch		
Swaziland	English		
Sweden	Swedish		
Switzerland	German		
Syria	Arabic		
Tajikistan	Tajiki		
Tanzania	English		
TFYR of Macedonia	Macedonian		
Thailand	Thai		
Timor-Leste	Portuguese		
Togo	French		
Tonga	Tongan		
Trinidad and Tobago	English		

APPENDIX B

Table B1. Full results for the OLS estimates of the effect of linguistic distance on service exports as share of GDP

Share of GDP	1	2	3	4	5
Linguistic distance	-0.0236***	-0.0233***	-0.0205***	-0.0174***	-0.0158***
	[0.00710]	[0.00359]	[0.00358]	[0.00362]	[0.00366]
Linguistic distance squared	0.0123**	0.0146***	0.0123***	0.0100***	0.00873***
	[0.00534]	[0.00272]	[0.00272]	[0.00275]	[0.00290]
Log GDP country 1	0.000147	-0.000647***	-0.00224***	-0.00222***	-0.00232***
	[0.000447]	[0.000233]	[0.000379]	[0.000378]	[0.000411]
Log GDP country 2	0.00379***	0.00249***	0.00150***	0.00154***	0.00207***
	[0.000453]	[0.000238]	[0.000326]	[0.000326]	[0.000351]
Log population country 1	-0.00279***	-0.000904***	0.000717	0.000712	0.000568
	[0.000497]	[0.000312]	[0.000464]	[0.000461]	[0.000497]
Log population country 2	-0.000909*	-0.000495*	0.000479	0.000495	-0.000173
	[0.000471]	[0.000293]	[0.000393]	[0.000393]	[0.000420]
Border	0.0137***	0.00743***	0.00759***	0.00656***	0.00562***
	[0.00217]	[0.00109]	[0.00108]	[0.00109]	[0.00109]
Avg total years of education	-0.00122***	-0.000428*	-0.000540**	-0.000530**	-8.65E-05
	[0.000412]	[0.000223]	[0.000225]	[0.000224]	[0.000241]
Log physical distance	-0.00364***	-0.00309***	-0.00335***	-0.00327***	-0.00453***
	[0.000485]	[0.000271]	[0.000272]	[0.000271]	[0.000338]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	0.0151	-0.00276	0.0139**	0.0113*	0.011
	[0.0101]	[0.00583]	[0.00650]	[0.00656]	[0.00689]
Observations	2,103	1,807	1,807	1,807	1,807
R-squared	0.189	0.311	0.325	0.334	0.356

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Table B2. Full results for the OLS estimates of the effect of English distance on service exports as share of GDP

Share of GDP	1	2	3	4	5
No common language * Combined English distance	-0.0180***	-0.0118***	-0.0102***	-0.00879***	-0.00657***
	[0.00292]	[0.00158]	[0.00163]	[0.00165]	[0.00171]
Log GDP country 1	-0.000391	-0.000952***	-0.00214***	-0.00214***	-0.00238***
	[0.000454]	[0.000239]	[0.000381]	[0.000380]	[0.000414]
Log GDP country 2	0.00376***	0.00250***	0.00179***	0.00179***	0.00213***
	[0.000452]	[0.000238]	[0.000329]	[0.000330]	[0.000354]
Log population country 1	-0.00228***	-0.000528*	0.000626	0.000635	0.0007
	[0.000494]	[0.000318]	[0.000466]	[0.000464]	[0.000500]
Log population country 2	-0.000893*	-0.000391	0.000223	0.000254	-0.000199
	[0.000470]	[0.000294]	[0.000393]	[0.000393]	[0.000423]
Border	0.0139***	0.00820***	0.00830***	0.00719***	0.00620***
	[0.00213]	[0.00106]	[0.00106]	[0.00108]	[0.00108]
Avg total years of education	-0.00152***	-0.000594***	-0.000628***	-0.000600***	-0.000164
	[0.000415]	[0.000224]	[0.000227]	[0.000226]	[0.000242]
Log physical distance	-0.00406***	-0.00318***	-0.00344***	-0.00339***	-0.00468***
	[0.000478]	[0.000262]	[0.000265]	[0.000264]	[0.000335]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	0.0246**	0.000172	0.0139**	0.0114*	0.0105
	[0.0102]	[0.00589]	[0.00655]	[0.00662]	[0.00701]
Observations	2,103	1,807	1,807	1,807	1,807
R-squared	0.193	0.311	0.32	0.328	0.347

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Table B3. Full regression results for the OLS estimates of the effect of linguistic distance on level of service and commodity exports

Log services exports	1	2	3	4	5
Linguistic distance	-3.839***	-4.223***	-3.678***	-2.816***	-2.293***
	[0.533]	[0.548]	[0.539]	[0.535]	[0.540]
Linguistic distance squared	3.014***	3.385***	2.955***	2.300***	1.559***
	[0.401]	[0.416]	[0.410]	[0.407]	[0.429]
Log GDP country 1	1.287***	1.252***	0.822***	0.841***	0.880***
	[0.0337]	[0.0356]	[0.0572]	[0.0560]	[0.0607]
Log GDP country 2	0.936***	0.898***	0.590***	0.623***	0.704***
	[0.0340]	[0.0363]	[0.0491]	[0.0483]	[0.0518]
Log population country 1	-0.518***	-0.385***	0.0971	0.0873	-0.00383
	[0.0373]	[0.0476]	[0.0699]	[0.0683]	[0.0733]
Log population country 2	-0.258***	-0.175***	0.165***	0.147**	-0.0114
	[0.0355]	[0.0448]	[0.0592]	[0.0581]	[0.0620]
Border	0.369**	0.402**	0.411**	0.133	0.125
	[0.163]	[0.166]	[0.162]	[0.161]	[0.161]
Avg total years of education	0.172***	0.209***	0.158***	0.152***	0.164***
	[0.0310]	[0.0340]	[0.0340]	[0.0332]	[0.0356]
Log physical distance	-1.050***	-1.091***	-1.142***	-1.114***	-1.145***
	[0.0364]	[0.0414]	[0.0409]	[0.0401]	[0.0499]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-18.26***	-20.39***	-17.69***	-18.98***	-17.70***
	[0.760]	[0.893]	[0.982]	[0.974]	[1.019]
Observations	2,093	1,798	1,798	1,798	1,798
R-squared	0.73	0.749	0.761	0.772	0.782

Table B3 continued

Log commodity exports	6	7	8	9	10
Linguistic distance	-2.905***	-3.330***	-3.139***	-2.574***	-1.784***
	[0.474]	[0.462]	[0.459]	[0.456]	[0.454]
Linguistic distance squared	2.672***	2.966***	2.851***	2.415***	1.445***
	[0.356]	[0.350]	[0.349]	[0.346]	[0.359]
Log GDP country 1	0.986***	0.963***	0.736***	0.760***	0.785***
	[0.0292]	[0.0294]	[0.0479]	[0.0469]	[0.0502]
Log GDP country 2	0.637***	0.655***	0.463***	0.502***	0.567***
	[0.0302]	[0.0307]	[0.0414]	[0.0408]	[0.0430]
Log population country 1	0.130***	0.0621	0.353***	0.339***	0.293***
	[0.0324]	[0.0394]	[0.0586]	[0.0573]	[0.0607]
Log population country 2	0.198***	0.160***	0.399***	0.371***	0.272***
	[0.0313]	[0.0375]	[0.0499]	[0.0490]	[0.0514]
Border	0.565***	0.583***	0.563***	0.388***	0.252*
	[0.142]	[0.138]	[0.136]	[0.135]	[0.133]
Avg total years of education	0.264***	0.205***	0.160***	0.150***	0.242***
	[0.0277]	[0.0289]	[0.0292]	[0.0285]	[0.0303]
Log physical distance	-1.167***	-1.143***	-1.154***	-1.133***	-1.293***
	[0.0319]	[0.0345]	[0.0345]	[0.0338]	[0.0416]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-20.22***	-17.44***	-17.36***	-18.66***	-19.08***
	[0.666]	[0.744]	[0.823]	[0.818]	[0.843]
Observations	2,048	1,759	1,759	1,759	1,759
R-squared	0.789	0.782	0.789	0.798	0.812

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Table B4. Full regression results for the OLS estimates of the effect of English distance on level of service and commodity exports

Log services exports	1	2	3	4	5
No common language * Combined English distance	-1.527***	-1.795***	-1.253***	-0.829***	-1.121***
	[0.221]	[0.242]	[0.246]	[0.245]	[0.251]
Log GDP country 1	1.264***	1.217***	0.832***	0.848***	0.871***
	[0.0343]	[0.0367]	[0.0577]	[0.0563]	[0.0606]
Log GDP country 2	0.933***	0.891***	0.615***	0.639***	0.714***
	[0.0341]	[0.0364]	[0.0498]	[0.0489]	[0.0518]
Log population country 1	-0.515***	-0.335***	0.0886	0.0827	0.00918
	[0.0372]	[0.0487]	[0.0705]	[0.0688]	[0.0732]
Log population country 2	-0.250***	-0.118***	0.168***	0.153***	-0.0149
	[0.0355]	[0.0450]	[0.0594]	[0.0582]	[0.0620]
Border	0.471***	0.577***	0.587***	0.263	0.21
	[0.160]	[0.163]	[0.160]	[0.160]	[0.159]
Avg total years of education	0.140***	0.175***	0.141***	0.141***	0.158***
	[0.0314]	[0.0344]	[0.0343]	[0.0335]	[0.0355]
Log physical distance	-1.037***	-1.032***	-1.091***	-1.072***	-1.142***
	[0.0361]	[0.0401]	[0.0401]	[0.0392]	[0.0492]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-17.97***	-20.20***	-17.77***	-19.17***	-17.48***
	[0.771]	[0.906]	[0.994]	[0.985]	[1.032]
Observations	2,093	1,798	1,798	1,798	1,798
R-squared	0.729	0.747	0.758	0.77	0.781

Table B4 continued

Log commodities exports	6	7	8	9	10
No common language * Combined English distance	-0.317	-0.365*	0.0107	0.329	0.0145
	[0.196]	[0.205]	[0.210]	[0.208]	[0.208]
Log GDP country 1	1.004***	0.981***	0.730***	0.753***	0.780***
	[0.0303]	[0.0310]	[0.0492]	[0.0479]	[0.0504]
Log GDP country 2	0.638***	0.658***	0.450***	0.486***	0.566***
	[0.0308]	[0.0314]	[0.0428]	[0.0419]	[0.0432]
Log population country 1	0.0959***	0.0574	0.366***	0.354***	0.300***
	[0.0330]	[0.0412]	[0.0602]	[0.0585]	[0.0609]
Log population country 2	0.202***	0.190***	0.434***	0.406***	0.270***
	[0.0319]	[0.0385]	[0.0510]	[0.0498]	[0.0516]
Border	0.692***	0.792***	0.785***	0.545***	0.345***
	[0.142]	[0.138]	[0.136]	[0.136]	[0.132]
Avg total years of education	0.252***	0.192***	0.154***	0.148***	0.253***
	[0.0285]	[0.0298]	[0.0300]	[0.0292]	[0.0304]
Log physical distance	-1.118***	-1.063***	-1.084***	-1.067***	-1.279***
	[0.0321]	[0.0341]	[0.0343]	[0.0334]	[0.0412]
Geography controls		Yes	Yes	Yes	Yes
Institutional controls			Yes	Yes	Yes
Colonial controls				Yes	Yes
Regional dummies					Yes
Constant	-20.91***	-18.61***	-18.32***	-19.80***	-19.88***
	[0.687]	[0.769]	[0.846]	[0.837]	[0.855]
Observations	2,048	1,759	1,759	1,759	1,759
R-squared	0.78	0.771	0.778	0.79	0.81

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Table B5. Full results for the OLS estimates of the effect of linguistic distance on categories of service exports

Log categories of service exports	Transportation	Travel	Communications services	Construction
Linguistic distance	0.279	-2.360***	-2.633***	-1.923
	[0.763]	[0.629]	[0.920]	[1.453]
Linguistic distance squared	-0.531	1.400***	1.172	1.3
	[0.604]	[0.506]	[0.712]	[1.112]
Physical distance	-1.202***	-1.229***	-1.407***	-1.273***
	[0.0700]	[0.0607]	[0.0808]	[0.131]
Log GDP country 1	0.655***	0.612***	0.578***	0.0304
	[0.1000]	[0.0891]	[0.115]	[0.181]
Log GDP country 2	0.817***	0.757***	0.750***	-0.0477
	[0.0803]	[0.0735]	[0.0913]	[0.147]
Log population country 1	0.169	0.317***	0.373***	0.923***
	[0.123]	[0.112]	[0.138]	[0.214]
Log population country 2	-0.192**	-0.132	0.0197	0.725***
	[0.0970]	[0.0900]	[0.108]	[0.169]
Border	0.215	0.25	-0.126	-0.0583
	[0.217]	[0.190]	[0.234]	[0.313]
Avg total years of education	0.0935*	0.0895*	0.165***	0.315***
	[0.0539]	[0.0476]	[0.0593]	[0.0942]
Geography controls	Yes	Yes	Yes	Yes
Institutional controls	Yes	Yes	Yes	Yes
Colonial controls	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes
Constant	-16.64***	-13.54***	-18.74***	-12.10***
	[1.416]	[1.284]	[1.610]	[2.614]
Observations	1,241	1,238	963	657
R-squared	0.7	0.729	0.731	0.416

Table B5 continued

Log categories of services exports	Insurance services	Financial services	Computer services	Royalties and license fees
Linguistic distance	-3.697***	-5.175***	-4.396***	-3.271***
	[1.111]	[1.112]	[0.981]	[0.972]
Linguistic distance squared	1.809**	3.188***	2.809***	2.856***
	[0.890]	[0.880]	[0.779]	[0.767]
Physical distance	-0.908***	-1.254***	-1.185***	-0.812***
	[0.107]	[0.107]	[0.0926]	[0.0922]
Log GDP country 1	1.266***	1.115***	0.612***	1.432***
	[0.156]	[0.147]	[0.133]	[0.142]
Log GDP country 2	0.834***	0.755***	0.280***	0.380***
	[0.120]	[0.118]	[0.106]	[0.107]
Log population country 1	-0.482***	-0.116	0.121	-0.149
	[0.183]	[0.179]	[0.162]	[0.167]
Log population country 2	-0.314**	-0.0259	0.609***	0.535***
	[0.144]	[0.140]	[0.125]	[0.127]
Border	0.606**	-0.626**	-0.609**	-0.35
	[0.291]	[0.294]	[0.261]	[0.249]
Avg total years of education	-0.0233	0.362***	0.428***	0.250***
	[0.0780]	[0.0776]	[0.0710]	[0.0713]
Geography controls	Yes	Yes	Yes	Yes
Institutional controls	Yes	Yes	Yes	Yes
Colonial controls	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes
Constant	-25.41***	-29.36***	-17.39***	-34.92***
	[2.188]	[2.016]	[1.872]	[2.035]
Observations	797	927	936	766
R-squared	0.6	0.68	0.647	0.732

Table B5 continued

Log categories of services exports	Other business services	Personal, etc. services	Government services
Linguistic distance	-3.516***	-4.106***	-2.761**
	[0.862]	[1.028]	[1.245]
Linguistic distance squared	2.519***	2.885***	2.982***
	[0.663]	[0.806]	[1.000]
Physical distance	-1.236***	-0.995***	-1.095***
	[0.0747]	[0.0973]	[0.124]
Log GDP country 1	1.095***	0.482***	0.324*
	[0.106]	[0.143]	[0.179]
Log GDP country 2	0.728***	0.667***	-0.0406
	[0.0842]	[0.114]	[0.147]
Log population country 1	-0.0143	0.161	0.470**
	[0.127]	[0.168]	[0.214]
Log population country 2	0.0578	-0.00052	0.635***
	[0.102]	[0.135]	[0.183]
Border	-0.428*	-0.262	-0.377
	[0.230]	[0.249]	[0.346]
Avg total years of education	0.139**	0.123*	-0.0983
	[0.0560]	[0.0739]	[0.0955]
Geography controls	Yes	Yes	Yes
Institutional controls	Yes	Yes	Yes
Colonial controls	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes
Constant	-22.75***	-14.53***	-10.26***
	[1.512]	[2.033]	[2.804]
Observations	1,073	714	670
R-squared	0.734	0.608	0.551

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

Table B6. Full results for the OLS estimates of the effect of English distance on categories of services exports

Log categories of services exports	Transportation	Travel	Communications services	Construction
No common language * Combined English distance	-0.414	-2.363***	-1.681***	0.0696
	[0.353]	[0.299]	[0.426]	[0.676]
Physical distance	-1.222***	-1.221***	-1.437***	-1.287***
	[0.0692]	[0.0589]	[0.0803]	[0.129]
Log GDP country 1	0.660***	0.581***	0.618***	0.039
	[0.100]	[0.0877]	[0.116]	[0.181]
Log GDP country 2	0.822***	0.764***	0.776***	-0.0335
	[0.0803]	[0.0722]	[0.0919]	[0.147]
Log population country 1	0.165	0.354***	0.336**	0.918***
	[0.123]	[0.110]	[0.139]	[0.214]
Log population country 2	-0.195**	-0.121	-7.54E-05	0.712***
	[0.0971]	[0.0886]	[0.109]	[0.169]
Border	0.187	0.321*	-0.0414	0.0393
	[0.214]	[0.184]	[0.230]	[0.306]
Avg total years of education	0.073	0.069	0.123**	0.315***
	[0.0532]	[0.0462]	[0.0592]	[0.0933]
Geography controls	Yes	Yes	Yes	Yes
Institutional controls	Yes	Yes	Yes	Yes
Colonial controls	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes
Constant	-16.06***	-11.89***	-18.12***	-12.94***
	[1.447]	[1.281]	[1.643]	[2.640]
Observations	1,241	1,238	963	657
R-squared	0.699	0.737	0.727	0.415

Table B6 continued

Log categories of services exports	Insurance services	Financial services	Computer services	Royalties and license fees
No common language * Combined English distance	-3.518***	-4.414***	-2.722***	0.54
	[0.491]	[0.503]	[0.465]	[0.481]
Physical distance	-0.942***	-1.215***	-1.155***	-0.778***
	[0.103]	[0.102]	[0.0911]	[0.0918]
Log GDP country 1	1.279***	1.086***	0.609***	1.453***
	[0.154]	[0.143]	[0.133]	[0.143]
Log GDP country 2	0.856***	0.748***	0.281***	0.389***
	[0.118]	[0.115]	[0.105]	[0.108]
Log population country 1	-0.470***	-0.0899	0.136	-0.154
	[0.181]	[0.175]	[0.161]	[0.169]
Log population country 2	-0.331**	-0.0212	0.600***	0.533***
	[0.142]	[0.137]	[0.125]	[0.128]
Border	0.662**	-0.476*	-0.445*	-0.135
	[0.279]	[0.280]	[0.254]	[0.245]
Avg total years of education	-0.125	0.305***	0.387***	0.297***
	[0.0771]	[0.0755]	[0.0704]	[0.0716]
Geography controls	Yes	Yes	Yes	Yes
Institutional controls	Yes	Yes	Yes	Yes
Colonial controls	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes
Constant	-22.89***	-25.92***	-15.87***	-37.57***
	[2.156]	[2.030]	[1.911]	[2.098]
Observations	797	927	936	766
R-squared	0.61	0.694	0.649	0.728

Table B6 continued

Log categories of services exports	Other business services	Personal, etc. services	Government services
No common language * Combined English distance	-1.535***	-1.401***	0.457
	[0.386]	[0.528]	[0.629]
Physical distance	-1.219***	-0.994***	-1.005***
	[0.0737]	[0.0956]	[0.123]
Log GDP country 1	1.101***	0.491***	0.274
	[0.106]	[0.143]	[0.181]
Log GDP country 2	0.736***	0.660***	-0.0684
	[0.0842]	[0.114]	[0.149]
Log population country 1	-0.0137	0.161	0.541**
	[0.127]	[0.168]	[0.216]
Log population country 2	0.0491	0.00476	0.655***
	[0.102]	[0.135]	[0.185]
Border	-0.293	-0.0902	-0.204
	[0.226]	[0.244]	[0.344]
Avg total years of education	0.134**	0.140*	-0.0356
	[0.0556]	[0.0732]	[0.0948]
Geography controls	Yes	Yes	Yes
Institutional controls	Yes	Yes	Yes
Colonial controls	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes
Constant	-22.55***	-14.73***	-12.05***
	[1.527]	[2.085]	[2.857]
Observations	1,073	714	670
R-squared	0.733	0.602	0.541

Note: Standard errors in brackets, *** p<.01, ** p<.05, * p<.1

REFERENCES

Amiti, Mary and Shang-Jin Wei. 2009. Service offshoring and productivity: Evidence from the US. *The World Economy* 32, no. 2: 203-220.

Anderson, James E. and Eric van Wincoop. 2004. Trade Costs. *Journal of Economic Literature* 42, no. 3: 691-751.

Baldwin, Kate and John D. Huber. 2010. Economic versus cultural differences: Forms of ethnic diversity and public goods provisions. *American Political Science Review* 104, no. 4: 644-662.

Beales, Richard. 2012. Apple needs good, not just better, supply chains. Breakingnews. Reuters. http://blogs.reuters.com/breakingviews/2012/03/07/apple-needs-good-not-just-better-supply-chain/ (accessed March 9, 2012).

Carstairs-McCarthy, Andrew. 2001. Origins of Language. In *The Handbook of Linguistics*, Ed. Mark Aronoff and Janie Rees-Miller, 1-18. Malden, MA: Blackwell Publishers.

CID. Research datasets. http://www.cid.harvard.edu/ciddata/ciddata.html (accessed January 13, 2012).

CILT. 2006. ELAN: Effects on the European economy of shortages of foreign language skills in enterprise. The National Centre for Languages. http://www.cilt.org.uk/home/research_and_statistics/research/cilt_activities/the_economic_case.aspx (accessed March 3, 2012).

Crystal, David. 2003. *The Cambridge Encyclopedia of the English Language*. 2nd ed. Cambridge: Cambridge University Press.

Deardorff, Alan V. 2010. Deardorff's Glossary of International Economics. University of Michigan. http://www-personal.umich.edu/~alandear/glossary/ (accessed March 23, 2012).

Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber. 2009. *Journal of the European Economic Association* 7, no. 6: 1291-1318.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black. 1992. An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82: iii-iv+1-132.

Eaton, Jonathan and Samuel Kortum. 2002. Technology, geography, and trade. *Econometrica* 70, no. 5: 1741-1779.

Education Testing Services. Who accepts TOEFL scores? http://www.ets.org/toefl/ibt/about/who_accepts_scores/ (accessed February 20, 2012).

———. 2007. Test and score data summary for TOEFL internet-based test: September 2005-December 2006 test data. http://www.ets.org/Media/Research/pdf/TOEFL-SUM-0506-iBT.pdf (accessed January 6, 2012).

———. 2008. Test and score data summary for TOEFL internet-based and paper-based tests: January 2007-December 2007 test data. http://www.ets.org/Media/Research/pdf/TOEFL-SUM-07.pdf (accessed January 6, 2012).

———. 2009. Test and score data summary for TOEFL internet-based and paper-based tests: January 2008-December 2008 test data. http://www.ets.org/Media/Tests/TOEFL/pdf/test_score_data_summary_2008.pdf (accessed January 6. 2012).

Embleton, Sheila. 2003. Historical linguistics: Methods: Mathematical methods. In *The International Encyclopedia of Linguistics*. 2nd ed. Ed. William J. Frawley. Oxford: Oxford University Press.

European Commission. 2006. Europeans and their languages. Special Eurobarometer 243. Brussels: European Commission. http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_sum_en.pdf (accessed January 12, 2012).

Fearon, James D. 2003. Ethnic and cultural diversity by country. *Journal of Economic Growth* 8, no. 2: 195-222.

Frankel, Jeffrey A. 1997. *Regional trading blocs in the world economic system.* Washington DC: Institute for International Economics.

Frankel, Jeffrey A. and Andrew Rose. 2002. An estimate of the effect of common currencies on trade and income. *The Quarterly Journal of Economics* 117, no. 2: 437-466.

Francois, Joseph and Bernard Hoekman. 2010. Service trade and policy. *Journal of Economic Literature* 48: 642-692.

Frawley, William J. ed. 2003. *The International Encyclopedia of Linguistics*. 2nd ed. Oxford: Oxford University Press.

Gallup, John Luke and Jeffrey D. Sachs. 1999. Geography and Economic Development. CID Working Paper No. 1. http://www.hks.harvard.edu/centers/cid/publications/faculty-working-papers/cid-working-paper-no.-001 (accessed January 13, 2012).

Gardner, Robert C. and Peter D. MacIntyre. 1992. A student's contributions to second language learning. Part I: Cognitive variables. *Language Teaching* 25, 211-220.

———. 1993. A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching* 26, 1-11.

GeoDist database. 2011. Centre D'Études Prospectives et D'Informations Internationales [CEPII]. http://www.cepii.fr/anglaisgraph/bdd/distances.htm (accessed January 14, 2012)

Ginsburgh, Victor, Ignacio Ortuño-Ortín, and Shlomo Weber. 2005. Disenfranchisement in linguistically diverse societies: The case of the European Union. *Journal of the European Economic Association* 3, no. 4: 946-965.

———. 2007. Learning foreign languages: Theoretical and empirical implications of the Selten and Pool model. *Journal of Economic Behavior and Organization* 64: 337-347.

Goswami, Arti Grover, Aaditya Mattoo, and Sebastian Saez, eds. 2012. *Exporting services: A developing country perspective*. Washington, DC: World Bank.

Greenberg, Joseph H. 1956. The measurement of linguistic distance. *Language* 32: 109-115.

Grunfeld, Leo A. and Andreas Moxnes. 2003. The intangible globalization: Explaining the patterns of international trade in services. Norwegian Institute of International Affairs Working Paper No. 657. http://english.nupi.no/content/download/1420/68964/version/5/file/657.pdf (accessed March 5, 2012).

Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie, eds. 2008. The World Atlas of Language Structures Online. Munich: Max Planck Digital Library. http://wals.info (accessed March 1, 2012).

Helyar, John and Mehul Srivastava. 2012. Outsourcing: A passage out of India. *Bloomberg Businessweek*, March 19- March 25.

Hummels, David. 1999, 2001. Toward a geography of trade costs. GTAP Working Paper No. 17, Purdue University.

IMF. 1993. Balance of Payments Manual. 5th ed. International Monetary Fund. http://www.imf.org/external/np/sta/bop/bopman.pdf (accessed March 7, 2012).

Jensen, Bradford J. 2011. *Global Trade in Services: Fear, Facts, and Offshoring.* Washington, DC: Peterson Institute for International Economics.

Katzner, Kenneth, 1986. *Languages of the World*. London: Routledge and Kegan Paul.

Kimura, Fukunari and Hyun-Hoon Lee. 2006. The gravity equation in international trade in services. *Review of World Economics* 142, no. 1: 92-121.

Ku, Hyejin and Asaf Zussman. 2010. Lingua franca: The role of English in international trade. *Journal of Economic Behavior & Organization* 75: 250-260.

Laitin, David D. 2000. What is a language community? *American Journal of Political Science* 44, no. 1: 142-155.

Lejour, Arjan and Jan-Willem de Paiva Verheijden. 2004. Service trade within Canada and the European Union: What do they have in common? CPB Netherlands Bureau for Economic Policy Analysis, CPB Discussion Paper No. 42. http://www.cpb.nl/en/publication/services-trade-within-canada-and-european-union-what-do-they-have-common (accessed March 5, 2012).

Lewis, M. Paul, ed. 2009 *Ethnologue: Languages of the World,* 16th ed. Dallas, TX: SIL International. Online version. http://www.ethnologue.com/ (accessed November 8, 2011-February 20, 2012).

Loberger, Gordon and Kate Shoup Welsh. 2002. *Webster's New World English grammar handbook: The ultimate desk reference.* Indianapolis: Wiley Publishing, Inc.

Lohmann, Johannes. 2011. Do language barriers affect trade? *Economic Letters* 110: 159-162.

Mayer, Thierry and Soledad Zignago. 2005. Market access in global and regional trade. CEPII Working Paper No 2005-2. http://www.cepii.fr/anglaisgraph/workpap/pdf/2005/wp05-02.pdf (accessed February 19, 2012).

———. 2011. Notes on CEPII's distance measures: The GeoDist database. CEPII Working Paper No 2011-25. http://www.cepii.fr/anglaisgraph/workpap/pdf/2011/wp2011-25.pdf (accessed February 22, 2012).

Melitz, Jacques. 2008. Language and foreign trade. *European Economic Review* 52: 667-699.

Mitchell, Rosamond and Florence Myles. 2004. *Second Language Learning Theories*. 2nd ed. New York: Oxford University Press Inc.

Odlin, Terence. 1989. *Language transfer: Cross-linguistic influence in language learning*. New York: Cambridge University Press.

———. 2003. Cross-linguistic influence. In *The handbook of second language acquisition*, Ed. Catherine J. Doughty and Michael H. Long, 436-487. Malden, MA: Blackwell Publishers.

OECD. Enhancing the Performance of the Services Sector. OECD, Directorate of Science, Technology, and Industry. http://www.oecd.org/document/2/0,3746,en_2649_33703_35026178_1_1_1_1,00.html (accessed March 9, 2012).

Park, Soon-Chan. 2002. Measuring tariff equivalents in cross-border trade in services. Korean Institute for International Economic Policy, Working Paper 02-15. http://www.kiep.go.kr/eng/publications/pub02.jsp?sCate=013002&sSubCate=005 (accessed March 5, 2012).

Paternoster, Raymond, Robert Brame, Paul Mazerole, and Alex Piquero. 1998. Using the correct statistical test for the equality of regression coefficients. *Criminology* 36, no. 4: 859-866.

Petroni, Filippo and Maurizio Serva. 2008. Language distance and tree reconstruction. *Journal of Statistical Mechanics: Theory and Experiment* P08012, 1-15.

Portes, Richard and Helene Rey. 2005. The determinants of cross-border equity flows. *Journal of International Economics* 65: 269-296.

Rauch, James E. and Vitor Trindade. 2002. Ethnic Chinese networks in international trade. *The Review of Economics and Statistics* 84, no. 1: 116-130.

Ringbom, Hakan. 1987. *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.

———. 1992. On L1 transfer in L2 comprehension and production. *Language Learning* 42, 85-112.

Ruhlen, Merritt. 2008. Global typological database. http://starling.rinet.ru/typology.pdf (accessed November 8, 2011).

Spolaore, Enrico and Romain Wacziarg. 2006. The diffusion of development. NBER Working Paper 12153. http://www.nber.org/papers/w12153 (accessed March 6, 2012).

Transparency International. CPI 2008 table. http://www.transparency.org/news_room/in_focus/2008/cpi2008/cpi_2008_table (accessed January 23, 2012).

Truchot, Claude. 2002. Key aspects of the use of English in Europe. Unpublished report, DGIV, Council of Europe. http://www.coe.int/t/dg4/linguistic/source/truchoten.pdf (accessed March 5, 2012).

United Nations. EBOPS 2002. United Nations International Trade Statistics Knowledgebase. http://unstats.un.org/unsd/tradekb/Knowledgebase/EBOPS-2002 (accessed March 7, 2012).

———. Extended Balance of Payments Services classification (EBOPS 2002). United Nations International Trade Statistics. http://unstats.un.org/unsd/tradeserv/EBOPS2002_eng.pdf (accessed November 1, 2012).

Walsh, Keith. 2008. Trade in services: Does gravity hold? *Journal of World Trade* 42, no. 2, 315-334.

WGI. 2011. The Worldwide Governance Indicators (WGI) project. http://info.worldbank.org/governance/wgi/index.asp (accessed January 25, 2012).

The World Factbook. Washington, DC: Central Intelligence Agency, 2009. https://www.cia.gov/library/publications/the-world-factbook/index.html (accessed November 27, 2011, February 10, 2012).

WTO. 2011. Table A8: World exports of commercial services by region and selected economy, 2000-2010, International Trade Statistics, World Trade Organization. http://www.wto.org/english/res_e/statis_e/its2011_e/its11_appendix_e.htm (accessed March 14, 2012).

Yun, Michelle and Kathy Chu. 2011. Philippines passes India in call-center jobs. *USA Today*, January 9.