LINGUISTIC DISTANCE AS A DETERMINANT OF BILATERAL TRADE

by

William K. Hutchinson

Working Paper No. 01-W30R

December 2001
Revised October 2003

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

"Linguistic Distance" as a Determinant of Bilateral Trade

William K. Hutchinson
Visiting Professor
Department of Economics
Vanderbilt University
October, 2003
(revised)
e-mail: william.hutchinson@vanderbilt.edu
phone: (615) 322-3428

I would like to thank Barry Chiswick, William Collins, James Dunlevy, Andre Souza, and members of the Brown Bag Seminar at Vanderbilt University for helpful comments on an earlier draft. Special thanks to Yanqin Fan for helpful econometric advice. As usual, I take responsibility for remaining errors.

JEL Codes: F1, J1
Key Words: Trade, Information costs, Immigration, Gravity

Abstract

We introduce a measure of language difficulty called "linguistic distance" into a modified gravity model to determine whether the fact that a language is further away from English affects the level of trade. Our sample of 36 non-English speaking countries includes Japan and South Korea, which we argue are special cases due to World War II, the Korean War, and subsequent close political and economic ties with the United States. Presence of a stock of immigrants in the home country has been shown to enhance trade, both exports and imports, with the country of origin. Controlling for network and information attributes provided by the presence of a stock of immigrants, the special relationship with Japan and Korea, and the standard gravity model variables, we find that the further a country's primary language is from English, the lower trade will be between the United States and that country. These results hold for aggregate exports and imports as well as for exports and imports of consumer manufactures and producer manufactures.

Variables accounting for the transactions costs associated with international trade have been included in gravity models since the early application of this model to explain the volume of bilateral international trade.[1] Although Bergstrand (1985 and 1989) limits transactions costs variables to adjacency and membership in a preferential trading area, Gould (1994) extends Bergstrand (1985) to explicitly model "costs associated with gaining foreign market information." Gould focuses on the trade enhancing effects arising from the presence of a stock of foreign immigrants residing in the home country.

Several papers have extended Gould (1994) and all of these have included a variable to account for the effect that a common language between the trading partners would have on transactions costs.[2] Frankel (1997) discusses the role of common language in the gravity framework and has included this variable in subsequent papers.[3] Boisso and Ferrantino (1997) attempt to capture the language effect by constructing a measure of the likelihood that an exporter from one country would encounter an importer in another country who spoke the same language. Their measure of linguistic similarity had no identifiable effect on the volume of trade between pairs of countries. That is, a higher probability that an exporter from one country would encounter an importer in a second country who spoke the same language had no significant effect on the volume of trade between the two countries.

Wagner, Head and Ries (2002) attempt to provide a measure of the degree of commonality of language between two countries that is very similar to that derived by Boisso

---

[1]Bergstand (1985) provides a review of the earlier papers as does Frankel (1997).
[2]See Head and Ries (1998), Dunlevy and Hutchinson (1999), Rauch (1999) Hutchinson and Dunlevy (2002) and Wagner, Head, and Ries (2002) which consider the effects on bilateral trade resulting from the formation of information linkages due to immigration and other network formation. Rauch and Trindade (2002) find that the larger the stock of Chinese immigrants in two countries, then the larger the volume of trade between the two countries.
[3]For example, see Frankel and Rose (2002) and many of the papers cited therein.

and Ferrantino, i.e., it is the probability that a randomly chosen person from a Canadian province would speak the same language as a randomly chosen person in the trading partner. These measures are a modification of the practice of using a dummy variable for a common language between trading partners. Hutchinson (2002) demonstrates that the greater the proportion of the population that speaks English as either a first or a second language, the higher the volume of trade (both exports and imports) between the U.S. and that country. [4]

We attempt to capture the trade effect imposed on residents of a country, native or immigrant, of learning a second language. We examine the effect on international trade that results from what Chiswick and Miller (1998) call "linguistic distance." How "distant" from English is a particular language, in the sense of how difficult is it to learn the foreign language. Using a measure of the relative difficulty that a native English speaking person encounters in learning particular languages[5], Chiswick and Miller (1998) study the effect of "linguistic distance" on English language fluency for immigrants in the United States. Fluency in English is a determining factor in the ability of an immigrant to realize the potential benefits from networking and to effectively use knowledge of their home country tastes and markets to promote trade and commerce between their host country and their country of origin. Thus, in addition to all the other factors that affect the volume of bilateral trade, "linguistic distance" is hypothesized to impact the volume of bilateral trade by increasing transactions costs and reducing the volume of trade between the U.S. and trading partners.

Greater difficulty in attaining English language fluency diminishes the ability of residents of foreign countries to develop networks and take advantage of specific knowledge of their home

---

[4]Jacques Melitz (2002) examines various issues related to language and communication: importance of particular languages and whether there might be negative effects resulting from groups of similar languages.

[5]See the data descriptions for an explanation of this variable.

country tastes and markets in promoting trade with the U.S.. Similarly, the greater the difficulty for a native English speaker to become fluent in a foreign language, the lower will be the volume of trade one would expect to observe between the U.S. and the country where the foreign language is spoken. Both effects work in the same direction: the greater difficulty that a native English speaker (non-English speaker) has in obtaining fluency in a foreign language (English), the lower the volume of bilateral trade one expects to find between the U.S. and the country in which the particular foreign language is spoken.

The paper will proceed in the following manner. In the next section we present the model and discuss the data in section 3. Empirical results are presented and discussed in section 4 and these results are related to the trade literature in the final section.

2 Model

We need a model that, controlling for other determinants of trade, will allow us to determine the effect that "linguistic distance" has on the volume of trade. We use a gravity model based on Gould (1994) which was an extension of Bergstrand (1985). Gould introduces transactions costs into the model. He argues that these costs arise from the lack of knowledge that a person from one country has regarding the laws, customs, tastes, and riskiness of markets in another country. Gould attempts to model the effect that a stock of immigrants from a country will have on the cost of obtaining such information. Obtaining such information will result in larger trade volumes between the host country and the immigrant's country of origin. We argue that independent of the effect of immigrants on trade flows, the difficulty of obtaining information and conducting trade depends on the difficulty of learning the foreign country's language as well as the difficulty for someone from the foreign country learning the home

country's language. We measure this degree of difficulty with the Chiswick and Miller index discussed in the previous section.

Gould's model assumes identical technologies among N countries that produce products which are differentiated by country of destination. In the supply side of the model labor is allocated across industries for each country according to a constant elasticity of transformation (CET) technology. Labor is transformed at a constant elasticity but the rate is different when labor is transformed between foreign products than it is when labor is transformed between foreign and domestic products. Transactions costs are introduced as a wedge between the price that a good sells for in a foreign market and the price the domestic producer receives for the good. Gould models three types of transactions costs: tariffs and other trade barriers; transport costs; and information costs. As equation (1) describes, the price received, P^{*}_{ik}, by i for selling her product in the kth market is less than the price the product sells for in the foreign market, P_{ik}, because of these three types of transaction costs. The greater are the transactions costs the lower will be the supply of goods to foreign markets. We shall return to these costs in a moment.

(1) $P^{*}_{ik} = \frac{P_{ik}}{T_{ik}C_{ik}Z_{ik}}$

Consumers in all countries have the same CES utility function in which the elasticity of substitution differs when comparing foreign and domestic goods as opposed to comparing two foreign goods. There are N(N-1) import demand equations to match the N(N-1) export supply functions, which allows one to solve for equilibrium conditions. Assuming that countries are small allows one to derive price and trade flow quantity equations, which when multiplied yield value of aggregate trade flow equations. The value of aggregate trade flows is described in

equation (2) where trade flows are determined by country incomes, Y_i, Y_j, transactions costs, C_{ij}, T_{ij},[6] and Z_{ij}, export prices, P^*_{ik}, import prices, P_{kj}, a domestic price index, $(P^*_{ik} + P_{ii})$, and a foreign price index, $(P_{kj} + P_{jj})$.[7]

$$(2)\ P_{ij}X_{ij} = Y_i^{\alpha} Y_j^{\beta} T_{ij}^{-\gamma} C_{ij}^{-\gamma} Z_{ij}^{-\gamma} P_{ik}^{*\theta} P_{kj}^{\psi} [P^*_{ik} + P_{ii}]^{\xi} [P_{kj} + P_{jj}]^{\mu}$$

Gould offers a specific formulation for Z to capture the information effect of a stock of immigrants, which, as Wagner, Head and Ries (2002) point out, is not derived from theory.[8] We offer an alternative formulation that we believe captures the essence of the effects one would attribute to the presence of immigrants as well as the language difficulty encountered when attempting to take advantage of specific information about tastes, culture, laws, and market risk. We model Z_{ij} as described in equation (3), where M_{ij} is the stock of immigrants in country i from country j and ld_{ij} is the index of the difficulty for a person whose native tongue is the language of country i to learn the language of country j. This specification captures the positive trade effect resulting from a stock of immigrants and the negative effect resulting from greater difficulty in learning the foreign language.

$$(3)\ Z = \frac{ld_{ij}}{M_{ij}}$$

$$\frac{\partial X}{\partial M} > 0 \qquad \frac{\partial X}{\partial ld} < 0$$

---

[6]We do not consider tariffs in this paper, however, due to the limited availability of data on bilateral tariffs for many of the countries in the sample. If one believes that language commonality is corelated with tariffs and other trade barriers, then omitting tariffs could bias our results. However, we believe that the effects of successive GATT rounds have reduced the variation of tariffs for the U.S. among countries which will minimize the possible bias from omitting tariff data.

[7]We have already made the substitutions of price indexes for the theoretically derived price indexes which would have excluded the price of the commodity under consideration. Such indexes are not available.

[8]The particular form used by Gould forces him to use a specific estimation procedure due to the nonlinearities that result.

The advantage of specifying the model as in equations (1), (2), and (3) is that it can be estimated in the standard double logarithmic form that is typically used for the gravity model. Using the double logarithmic specification facilitates comparison with the results of other investigators who have examined trade in the gravity framework.

Gould used a country fixed effects model to determine how important the presence of a stock of immigrants was for trade between the U.S. and the home country of the immigrants. To determine the effect of "linguistic distance" we cannot use a country fixed effects model because, unlike the stock of immigrants, "linguistic distance" is fixed for a country. Therefore, in addition to the linguistic distance index, we include the standard country specific gravity model variable, distance. Distance, as measured by the great circle distances between national capitals, will be used as a proxy for transport cost. We also include a time trend to control for possible bias introduced by growth over time in the size of trade and income.[9]

The gravity model hypothesizes that the distance between two countries and their size determines the volume of bilateral trade, exports or imports. Although an imperfect measure, distance reflects the cost of shipping goods between countries, C_{ij}. Thus, the greater the distance the lower the volume of trade. The larger two countries are the greater the volume of trade that one would expect to observe because larger countries account for a larger share of world income and consumption. Following Gould, we use income and population to capture the size of trading partners.[10] We also follow Gould by including various characteristics of the immigrant stock, e.g., length of stay and the ratio of skilled to unskilled in the stock of immigrants from a particular country. Equation (4) presents the estimating equation to be used.

---

[9]See Anderson (1979), Baier and Bergstrand (2000), and Bergstrand (1985, 1989) for theoretical derivations of gravity models that result in estimating models similar to Gould (1994) and the one we employ.

[10]See Frankel (1997) for a discussion of how these variables account for size in gravity models.

(4) $V_{ijt} = F(\text{US Income}_t, \text{Foreign Income}_{jt}, \text{US Population}_t, \text{Foreign Population}_{jt}, \text{US Price Deflator}_t, \text{Foreign Price Deflator}_{jt}, \text{Export/Import Unit Value}_{ijt}, \text{Distance}_{ij}, \text{Immigrant Stock}_{ij}, \text{Skilled/Unskilled Immigrants}_{ijt}, \text{Period Stayed}_{ijt}, \text{Linguistic Distance}_{ij}, \text{year}_t) + \text{random error}_{ijt}$

V_{ijt} is the volume of trade, exports or imports, between i, the U.S., and country j in year t and, except for unit value, the other variables are as defined for each regression. "Unit value" is the ratio of U.S. export unit value to the foreign import unit value when estimating the export equation and the ratio of the U.S. import unit value to the foreign export unit value when estimating the import equation. All other variables appear in both the export and import equations. The variable year accounts for any general trend not otherwise captured by other included variables.

3 Data

Except for the linguistic distance index, the data are those graciously provided by David Gould and are the data that he used in Gould (1994). Thirty-six of the 47 countries in Gould's data set are classified as non-English speaking.[11] Data for the 36 countries are for all years between 1970 and 1986, except in the cases of: El Salvador (1970-1984); Ehtiopia (1970-1980); Morocco (1970-1985); Nicaragua (1970-1980); Singapore (1972-1980); Trinidad (1970-1985); Turkey (1970-1984); and Yugoslavia (1970-1983). Trade data were gathered from the *International Monetary Fund Direction of Trade Statistics* and consumer and producer manufactured exports and imports came from OECD statistics on trade in manufactured goods and were sorted on the basis of the International Standard Industrial Classification codes.[12] The

[11]Classification was based on data from the website for international trade resources maintained by Raymond Robertson. http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/TradeData.html#Gravity

[12]Gould (1994) indicates that an attempt was made to exclude ambiguous categories.

stock of immigrants and data relating to them were derived from the Immigration and Naturalization Service's public-use data on yearly immigration using the 1980 Census as a benchmark. Average length of stay was calculated as the average length of stay for immigrants who arrived between 1970 and 1986. Data for U.S. and foreign import and export unit value indexes, population, GDP, and GDP deflators were taken from the *IMF Financial Statistics.*[13] A list of countries and consumer and producer manufactures are provided in Appendix A and Appendix B, respectively.

The linguistic distance index is from Chiswick and Miller (1998) Appendix B. The index is based on language test scores (LS) that range from 1.0 to 3.0 by increments of 0.25, with a value of 1.0 assigned to a country with a language that was extremely difficult for a native English speaker to learn.[14] In our case, Japanese and Korean were the only two languages assigned the value of 1.0. Chiswick and Miller argue that foreign language fluency is part of an immigrant's human capital which allows him or her to realize higher earnings and to more efficiently search for jobs and consumer goods. We extend this to encompass the ability of a person to communicate with those from a foreign country to gain knowledge of tastes and market information to promote trade between the two countries. Thus, a language that is further from English results in greater transactions costs and lower trade volume between the U.S. and the country where the language is spoken.

The linguistic distance index, derived as (1/LS), is used in the estimation of equation (4), resulting in a variable that varies between 0.33 and 1.0. Thus, higher index values are associated

---

[13]The reader is referred to Gould (1994) for a more detailed description of data collection or one is available from this author upon request.

[14]Chiswick and Miller (1998) use the language scores reported in Hart-Gonzalez and Linderman (1993) from the U.S. Department of State, School of Language Studies. As they do, we assume that the difficulty encountered for a non-English speaker learning English is symmetric with the difficulty for an English speaker to learn a foreign language.

with more difficult languages, i.e., greater linguistic distance. The indexes appear along with the list of countries in Appendix A.

4 Results

The results from estimating equation (4) in double log form, except for the year dummy, appear in Table 1, where the standard gravity variables have the expected coefficients. Foreign income is positive and significant but less than one, which indicates that a ten percent increase in the income of a foreign country results in less than a ten percent increase in U.S. exports to that country. Holding income constant, an increase in the population of a foreign country results in lower exports to that country which is consistent with the implied reduction in the standard of living. Higher U.S. population, holding income constant, results in lower exports, which is consistent with the capacity pressure literature.[15]

Since the price deflators reflect a weighted combination of traded and nontraded goods, a positive sign on the U.S. price deflator and a negative sign on the foreign price deflator may well reflect the fact that trade comprises a much smaller share of GDP in the U.S. than it does in most of the countries in the sample. The negative sign for the ratio of the export unit value to import unit value indicates that if prices of U.S. export goods rise relative to the prices of all goods imported by the foreign country, then one would expect lower exports from the U.S. to that country. Similarly, if prices of U.S. imports generally rise faster than the prices of exports from a foreign country, one would expect the U.S. to import more from that foreign country, i.e., the sign is positive as it appears in the import equation results.

Distance is negative but not significantly different from zero for both exports and imports. The ratio of skilled to unskilled immigrants is negative for both exports and imports,

[15]For information on the capacity pressure hypothesis see Dunlevy (1980) and Raynold and Dunlevy (1998).

but is only significant for imports. Thus, the U.S. imports less from countries that have a higher ratio of skilled to unskilled immigrants in the U.S. Moreover, the longer the immigrants stay in the U.S. the lower is trade with their country of origin.

The stock of immigrants has a positive and significant effect on the volume of exports and imports. A ten percent larger stock of immigrants results in approximately a one percent greater volume of both exports and imports. The relatively larger coefficient for imports may reflect the added immigrant taste effect for imports that does not affect exports. Linguistic distance is positive but insignificant for exports and negative and significant for imports. Thus, the more difficult it is for a native English speaker to learn a language the lower will be the imports from that country. Exports are not significantly affected by linguistic distance.

Given the sample of countries (See Appendix A), one might ask if any countries are outliers or may dominate the results, given that we cannot control directly for country fixed effects and determine the effect of linguistic distance. Figures 1 and 2 contain the log of the ratio of average exports to average GNP and the log of the ratio of average imports to average GNP plotted against the country's linguistic index, respectively. One quickly sees that Japan and South Korea are outliers relative to the other countries.[16] If one fits a least squares line to the group of countries, other than Japan and South Korea, then it is easy to see that these two are significantly different from the others. Thus, we estimate equation (4) controlling for the specific intercept shift characteristics of Japan and South Korea. We argue that this relationship exists because of the unique ties between these two countries and the U.S. as a result of wars and the political situation that has prevailed for more than a half century.

[16]Singapore is an additional outlier for exports and Trinidad is an additional outlier for imports. Neither country lies far from a regression line fit to the remaining group of country ratios. One might ask about the special relationship between the U.S. and European countries. Germany has been the U.S.'s second or third largest trading partner, but its trade is not inordinately larger relative to its GNP.

Columns 3 and 4 of Table 1, Exports(D) and Imports(D), contain the results from estimating equation (4) using an intercept shift dummy variable for Japan and South Korea. The coefficients, not shown, were large and significantly greater than zero, as one would expect given the evidence in Figures 1 and 2.[17] Results for the standard gravity model variables are little changed from those in columns 1 and 2 of Table 1. The immigrant stock coefficients remain positive and significantly greater than zero. However, the linguistic distance coefficients for both exports and imports are now negative and significant. The coefficient for imports is nearly four times more negative than the value in column 2 of Table 1. Results that were only marginally evident become quite stark when one controls for the special characteristics associated with Japan and Korea. We shall include this control in the remainder of our analysis.

Serial correlation resulting from persistence effects in trade are common problems with time series or panel data. We follow Eichengreen and Irwin (1998) in using the Hatanaka 2-Step process to correct for serial correlation and to account for any persistence that may occur as a result of prior trade relations between the U.S. and the other countries in the sample. The Hatanaka process requires that one obtain estimates of the dependent variable (exports or imports) using an instrumental variable approach. The predicted value for the dependent variable is used as a lagged dependent variable in estimating the model to generate a serial correlation coefficient that is used to quasi-difference the data. The final model contains all of the quasi-differenced variables, including the lagged dependent variable, plus the lagged error term from the estimation that generated the serial correlation coefficient. Thus, we can account for persistence effects and the impact of a shock last period on exports or imports in the current period.[18]

---

[17]The shift variable for Japan and South Korea was 2.1 for total exports and 6.2 for total imports.
[18]See Greene (1993) for a more detailed explanation of this process.

Results from applying the Hatanaka 2-Step correction for serial correlation and persistence are presented in Table 2. U.S. income and population are not significantly different from zero once persistence and lagged shocks are accounted for in the model. Foreign income and population have the same effect as in Table 1. That is, foreign income has a positive effect on U.S. exports and the U.S. tends to import more from countries with higher income. The negative coefficient on foreign population indicates that, holding income constant an increase in population results in lower U.S. exports and imports. Price deflators and export and import unit value ratios have the same effect on the volume of trade as in Table 1. Immigrant stock has a positive and significant effect on both exports and imports, which is slightly larger than these coefficients were in Table 1. Linguistic distance has a slightly smaller, but statistically significant negative effect on both exports and imports.

Lagged exports and imports have significant negative coefficients, which indicates that the U.S. tended to trade less with countries that it had traded more with in the past or a negative persistence effect. The lagged error term generated a negative and significant effect for trade. That is, larger than expected exports from the U.S. to a country in one year resulted in a reduction in exports to that country relative to what was expected in the subsequent year. The effect was analogous for imports.

Controlling for all of the standard gravity model variables, including the presence of a stock of immigrants, we have shown that trade will be lower with countries where the language is more distant from English, i.e., more difficult for an English speaker to learn. It is frequently the case that different types of products are differentially affected by gravity model variables.[19]

---

[19]Gould (1994), Dunlevy and Hutchinson (1999) and Hutchinson and Dunlevy (2002) have found that the presence of a stock of immigrants differentially impacts particular product categories.

We shall examine the differential effect of linguistic distance on consumer manufactures and producer manufactures.

Consumer Goods Trade

We observe in columns 1 and 2 of Table 3 that the U.S. imports relatively more consumer manufactured goods from countries that have higher incomes. That is, a country with 10 percent higher income would be expected to have approximately 12 percent higher exports of consumer manufactures to the U.S.. The immigrant stock coefficients are larger for both exports and imports, indicating that immigrants enhanced trade in consumer manufactures to a greater extent than was the case for total trade.

Linguistic distance had a larger effect on exports of consumer manufactures and a slightly smaller effect on imports. Other things the same, the difficulty with learning a foreign language had a larger negative effect on exports of consumer manufactures than it did for total exports. Linguistic distance had a smaller effect on imports of consumer manufactures than it did for total imports. Other model coefficients were similar to those for total exports and imports in Table 2.

Producer Manufactures

Columns 3 and 4 of Table 3 contain regression results for U.S. exports and imports of producer manufactures. The results for all of the standard variables are similar to those for consumer manufactures. Presence of a stock of immigrants has a lower effect on exports of producer manufactures to their country of origin than was the case for either exports of consumer manufactures or total exports. The effect of a stock of immigrants on U.S. imports of producer manufactures from their country of origin is less than was true for imports of consumer manufactures but larger than was the case for total imports. Although the immigrant taste effect

results in a larger migrant stock effect on imports than exports, one would expect that this taste effect might be smaller for producer manufactures than for consumer manufactures.

Linguistic distance has a larger impact on exports of producer manufactures than on exports of consumer manufactures, but a smaller effect on imports of producer manufactures than on imports of consumer manufactures. However, the language barrier continues to impact imports more than exports. Thus, the language barrier appears to be of greater importance for producer manufactures exports than other types of exports but relatively less important for imports of producer manufactures than other types of imports.

One can argue that the language barrier imposes greater costs when one is considering imports of either producer or consumer manufactures than is the case for exports. An examination of the list of products in producer manufactures found in Appendix B supports this argument. If U.S. goods are better known through brand name recognition or experience, then language differences create less of a barrier for exports of these goods. The preciseness of information necessary for importing many of these products would result in much higher costs associated with a language barrier, i.e., linguistic distance.

5 Final Comments

Nearly every gravity model of bilateral trade has employed dummy variables to account for a common language between bilateral trading partners. Gould (1994) showed that immigrants from non-English speaking countries had a larger pro-trade effect than did immigrants from English speaking countries when examining trade between the U.S. and other countries. Using a gravity model for trade, which in addition to the traditional variables, accounts for transactions costs, we find that linguistic distance reduces the volume of U.S. trade with a country even in the presence of a stock of immigrants from that country. Thus,

immigrants may provide information that enhances trade, but linguistic distance diminishes the volume of trade.[20] We show that, among non-English speaking countries, one observes lower trade with countries whose language is more distant from English.

We show that trade with Japan and South Korea is different than trade with the other countries in the sample, i.e., trade is much greater relative to GNP than it is for the other countries. Controlling for this special trade relationship results in a much larger and significantly negative effect of linguistic distance on trade. The Hatanaka 2-Step process is employed to correct for serial correlation and to capture any persistence effects in trade between the U.S. and these 36 countries.[21]

The linguistic distance results hold for total exports and total imports as well as consumer manufactures and producer manufactures exports and imports. We find that information costs are more important for imports than exports and especially for imports of producer manufactures. Linguistic distance is also more important for exports of producer manufactures than for either consumer manufactures exports or total exports.

The results in this paper add to the growing body of information regarding the importance of transactions cost in international trade. We show that, in addition to the English versus non-English distinction, the "distance" between a language and English makes a significant difference in the volume of trade between non-English speaking countries and the U.S.

---

[20]In regression not reported, linguistic distance was interacted with immigrant stock, but the results were very insignificant. The coefficients on immigrant stock and linguistic distance remained the same as reported in this paper.

[21]Regression were also run for equation (4) including time fixed effects and the results were unchanged.

References

Anderson, J.E., (1979) "A Theoretical Foundation for the Gravity Equation." *American Economic Review* 69 (1), 106-116.

Baier, Scott L. and Jeffrey H. Bergstrand (2001) "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity." *Journal of International Economics* 53 (1) pp. 1-27.

Bergstrand, Jeffrey H. (1985) "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence." *Review of Economics and Statistics* 67 (3), 474-481.

Bergstrand, Jeffrey H. (1989) "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory of International Trade." *Review of Economics and Statistics* 71 (1), 143-153.

Bergstrand, Jeffrey H. (1990) "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis, and the Determinants of Bilateral Intra-industry Trade. *Economic Journal* 100 (4), 1216-1229.

Chiswick, Barry R. and Paul W. Miller (1998) "English Language Fluency Among Immigrants in the United States." *Research in Labor Economics* 17, 151-200.

Dunlevy, James A. and William K. Hutchinson (1999) "The Impact of Immigration on American Import Trade in the Late Nineteenth and Early Twentieth Centuries." *Journal of Economic History* 59 (4), 1043-1062.

Dunlevy, James A. (1980) "A Test of the Capacity Pressure Hypothesis Within a Simultaneous Equations Model of Export Performance," *Review of Economics and Statistics* 62(1) February. pp 131-135.

Eichengreen, Barry and Douglas A. Irwin (1998) "The Role of History in Bilateral Trade Flows," in Jeffrey A. Frankel (ed.) *The Regionalization of the World Economy* Chicago: University of Chicago Press. pp. 33-62.

Frankel, Jeffrey A. (1997) *Regional Trading Blocs in the World Economic System*. Institute for International Economics, Washington, D.C.

Frankel, Jeffrey A. and Andrew K. Rose (2002) "An Estimate of the Effect of Common Currencies on Trade and Income" *Quarterly Journal of Economics* 117:2 pp. 437-466.

Gould, David M. (1994) Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows." *Review of Economics and Statistics* 76, 302-316.

Greene, William H. (1993) *Econometric Analysis* (2nd. ed.) New York: Macmillan Publishing Co.

Hart-Gonzalez, L. and S. Linderman (1993) "Expected Achievement in Speaking Proficiency, 1993." Mimeographed, School of Language Studies, Foreign Services Institute, U.S. Department of State.

Head, Keith and Ries, John. (1998) "Immigration and Trade Creation: Econometric Evidence from Canada." *Canadian Journal of Economics* 31, 47-62.

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

Hutchinson, William K. and James A. Dunlevy (2001) "The Pro-Trade Effect of Immigration on American Exports During the Classical Liberal Period" Vanderbilt University, Department of Economics *Working Paper* 01-W25.

Melitz, Jacques (2002) "Language and Foreign Trade" CEPR Working Paper

Rauch, James E. (1999) "Networks Versus Markets in International Trade." *Journal of International Economics* 48, 7-35.

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

Raynold, Prosper and James A. Dunlevy (1998) "Aggregate Shocks and the Relationship Between U.S. Business Cycle Fluctuations and Export Performance," *Journal of Economic Integration* 13(1) March pp. 163-98.

Wagner, Don, Keith Head, and John Ries (2002) "Immigration and the Trade of Provinces" *Scottish Journal of Political Economy*, forthcoming.

Table 1 Results for Total Bilateral Exports and Total Bilateral Imports: The U.S. and 36 Trading Partners, 1971-1986

Variables	Exports	Imports	Exports(D)	Imports(D)
Constant	1097.07***	1304.70***	865.3***	625.24***
	(5.38)	(4.32)	(4.48)	(3.03)
US Income	2.55*	5.07**	2.15	3.90***
	(1.77)	(2.35)	(1.59)	(2.61)
Foreign Income	0.87***	1.12***	0.67***	0.51***
	(21.00)	(18.96)	(15.09)	(11.97)
U.S. Population	-211.49***	-258.19***	-167.01***	-127.78***
	(5.63)	(4.65)	(4.68)	(3.36)
Foreign Population	-0.11***	-0.19***	-0.01	0.12***
	(2.91)	(3.58)	(0.02)	(3.05)
U.S. Price Deflator	7.16***	3.30	0.14***	0.91
	(4.54)	(1.41)	(7.12)	(0.57)
Foreign Price Deflator	-0.16***	-0.14***	-0.16***	-0.13***
	(7.66)	(5.75)	(7.48)	(5.58)
US Export/Foreign	-0.15**		-0.18***	
Import Unit Value	(2.59)		(3.30)	
US Import/Foreign		0.27***		0.19***
Export Unit Value		(3.83)		(3.85)
Distance	-0.09	-0.20	-0.13	-0.30***
	(0.81)	(1.17)	(1.33)	(2.91)
Immigrant Stock	0.09***	0.12***	0.08***	0.10***
	(2.82)	(2.67)	(2.72)	(3.38)
Skilled/Unskilled	-0.02	-0.16*	0.01	-0.07
Immigrants	(0.33)	(1.67)	(0.13)	(1.15)
Period Stayed	-1.75***	-2.44***	-1.24***	-0.96***
	(8.84)	(8.92)	(6.20)	(4.99)
Linguistic Distance	0.13	-1.18***	-1.37***	-5.60***
	(0.85)	(3.86)	(5.99)	(20.37)
Year	1.64***	2.23	1.25***	1.07***
	(4.52)	(4.12)	(3.62)	(2.90)
Number of Observations	547	547	547	547
R-Squared	0.79	0.72	0.82	0.86

t-statistics are in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent level respectively. t-values are based on the robust standard errors correction procedure for heteroskedasticity using the STATA software package. The coefficients for the South Korea and Japan dummy variable in Export(D) and Import(D) were not included but were positive and highly significant as the evidence in Figures 1 and 2 would indicate.

Figure 1 Plot of Exports to GDP Ratio to Language Index

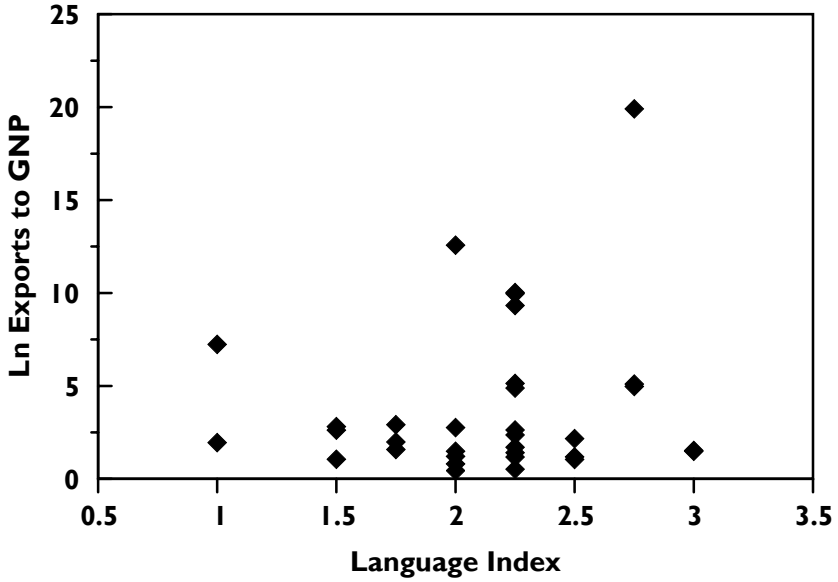

Figure 2 Plot of Imports to GNP Ratio to Language Index

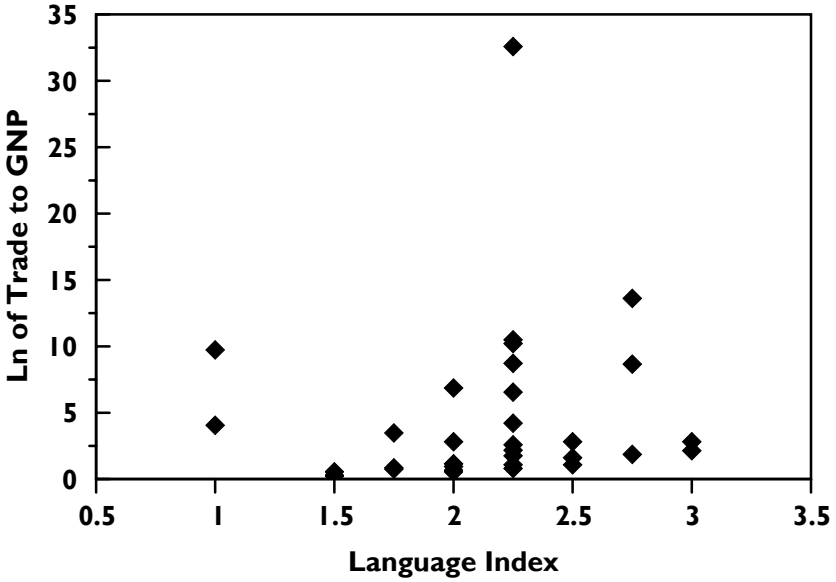

Table 2 Results for Total Exports and Total Imports With Hatanaka 2-Step Correction
The U.S. and 36 Non-English Speaking Trading Partners, 1972-1986

Variables	Exports	Imports
Constant	21.63	67.60
	(0.54)	(0.89)
US Income	-1.76	-5.17
	(0.37)	(0.67)
Foreign Income	0.86***	0.96***
	(20.43)	(15.32)
U.S. Population	-17.73	-47.51
	(0.36)	(0.63)
Foreign Population	-0.14***	-0.16***
	(4.37)	(3.11)
US Price Deflator	5.68	7.85
	(1.05)	(0.84)
Foreign Price Deflator	-0.14***	-0.15***
	(7.27)	5.80)
U.S.Export/Foreign Import Unit Values	-0.16*** (3.05)	
U.S. Import/Foreign Export Unit Values		0.28*** (3.81)
Distance	-0.16*	-0.33**
	(1.91)	(2.55)
Immigrant Stock	0.09***	0.14***
	(3.53)	(3.84)
Skilled/Unskilled Immigrants	-0.02 (0.38)	-0.06 (0.85)
Period Stayed	-2.53***	-2.66***
	(12.02)	(8.89)
Linguistic Distance	-0.44***	-2.59***
	(2.68)	(8.89)
Lagged Exports	-0.07*** (3.71)	
Lagged Imports		-0.07*** (3.01)
Lagged Error Term	-0.82***	-0.80***
	(16.25)	(11.73)
Number of Observations	476	476
R-Squared	0.88	0.82

t-statistics are in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels , respectively. t-values are based on the robust standard errors correction procedure for heteroskedasticity using the STATA software package. The coefficients for the South Korea and Japan dummy variable were not included but were positive and highly significant. The time trend coefficient is not reported in the table and it was only marginally positive for imports.

Table 3 Consumer Goods (C) and Producer Goods (P) Exports and Imports With Hatanaka 2-Step Correction: The US and 36 Non-English Speaking Trading Partners, 1972-1986

Variables	Exports(C)	Imports(C)	Exports(P)	Imports(P)
Constant	5.66	-68.56	-44.51	-7.17
	(0.16)	(1.51)	(0.88)	(0.13)
US Income	-2.91	-5.01	-1.43	2.04
	(0.51)	(0.48)	(0.34)	(0.30)
Foreign Income	0.97***	1.24***	0.94***	1.46***
	(21.37)	(19.32)	(22.64)	(27.58)
US Population	2.74	115.56	41.48	-0.44
	(0.06)	(1.39)	(0.93)	(0.01)
Foreign Population	-0.34***	-0.33***	-0.13***	-0.60***
	(8.76)	(5.58)	(3.60)	(12.05)
US Price Deflator	7.06	-6.09	5.85	-3.75
	(1.13)	(0.56)	(1.21)	(0.57)
Foreign Price Deflator	-0.13***	-0.22***	-0.15***	-0.05*
	(6.54)	(7.41)	(8.30)	(1.96)
US Export/Foreign Import Unit Values	-0.22***		-0.26***	
	(3.21)		(4.93)	
US Import/Foreign Export Unit Values		-0.07		-0.30***
		(0.74)		(2.70)
Distance	-0.33***	1.16***	-0.34***	-0.69***
	(3.31)	(8.26)	(3.52)	(4.56)
Immigrant Stock	0.12***	0.24***	0.08***	0.20***
	(4.22)	(6.02)	(2.87)	(5.12)
Skilled/Unskilled Immigrants	0.06	-0.25**	-0.03	0.28***
	(1.02)	(2.52)	(0.58)	(3.16)
Period Stayed	-3.48***	-2.71***	-2.66***	-2.62***
	(14.81)	(8.27)	(12.91)	(9.25)
Linguistic Distance	-0.52***	-2.26***	-0.94***	-1.50***
	(2.96)	(7.54)	(5.44)	(5.30)
Lagged Exports	-0.03		-0.03	
	(1.40)		(1.58)	
Lagged Imports		-0.09***		-0.05***
		(4.22)		(2.84)
Lagged Error	-0.79***	-0.91***	-0.74***	-0.70***
	(12.70)	(13.23)	(14.28)	(12.43)
Number of Observations	476	476	476	476
R-Squared	0.85	0.86	0.88	0.88

t-statistics are in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels , respectively. t-values are based on the robust standard errors correction procedure for heteroskedasticity using the STATA software package. The coefficients for the South Korea and Japan dummy variable were not included but were positive and highly significant. The time trend variable is not reported in the table as it was insignificant for all but exports of producer goods, where it was marginally significant.

Appendix A: Country List

Austria(2.25)	Brazil(2.5)	Colombia(2.25)	Cyprus(1.75)
Denmark(2.25)	El Salvador(2.25)	Ethiopia(2.0)	Finland(2.0)
France(2.5)	Greece(1.75)	Hungary(2.0)	Iceland(2.25)
Israel(2.0)	Italy(2.5)	Japan(1.0)	Jordan(1.5)
Malaysia(2.75)	Morocco(1.5)	Netherlands(2.75)	Nicaragua(2.25)
Norway(3.0)	Pakistan(1.75)	Philippines(2.25)	South Korea(1.0)
Singapore(2.75)	Spain(2.25)	Sri Lanka(1.75	) Sweden(3.0)
Switzerland(2.25)	Syria(1.5)	Thailand(2.0)	Trinidad(2.25)
Tunisia(1.5)	Turkey(2.0)	West Germany(2.25)	Yugoslavia(2.0)

Numbers in parentheses are the linguistic distance indexes which were inverted in the gravity regressions.

Appendix B: Product Lists

Consumer Manufactured Products

Preserved meat products
Canned fruits and vegetables
Bakery products
Wine products
Soft drinks and carbonated waters
Textiles excluding wearing apparel
Leather products excluding foot wear
Cane containers and small cane ware
Soap and cosmetics
Glass products
Metal furniture and fixtures
Radio and TV equipment and apparatus
Motor vehicles
Photographic and optical goods
Jewelry
Sporting goods
Dairy products
Canned and preserved fish
Cocoa and sugar confectionery
Malt liquors
Tobacco manufactures
Wearing apparel
Leather footwear
Wood furniture and fixtures
Pottery, china, and earthware
Cutlery, and general hardware
Metal products excl. machinery and equip.
Electric appliances and house wares
Motorcycles and bicycles
Watches and clocks
Musical instruments

Producer Manufactured Goods

Prepared animal feeds
Knitting mills
Fur dressing and dyeing industries
Pulp and paper
Basic industrial chemicals
Synthetic resins, and plastics
Drugs and medicines
Misc. petroleum and coal products
Cement, lime and plaster
Iron and steel basic industries
Structural metal products
Agricultural machinery
Aircraft
mach.
Electric industrial machinery
Shipbuilding and repairing
Scientific equipment
Weaving and finishing textiles
Rope and twine industries
Sawmills, planing, and other wood mills
Paper boxes, and paperboard
Fertilizers and pesticides
Paints, varnishes and lacquers
Petroleum products
Tire and tubes industries
Nonmetallic mineral products
Metal scrap
Engines and turbines
Metal and woodworking machinery
Industry mach., excl. metal and woodworking

Office, computing and accounting machinery
Transport equipment