DISCUSSION PAPER SERIES

No. 3590

LANGUAGE AND FOREIGN TRADE

Jacques Mélitz

INTERNATIONAL MACROECONOMICS and INTERNATIONAL TRADE

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP3590.asp**

ISSN 0265-8003

LANGUAGE AND FOREIGN TRADE

Jacques Mélitz, Center for Research in Economics and Statistics (CREST), Strathclyde University, and CEPR

Discussion Paper No. 3590
October 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3590

October 2002

ABSTRACT

Language and Foreign Trade*

The significance of a common language in foreign trade hinges on translation as well as the ability to communicate directly. In fact, without admitting the facility of translation from one or two selected languages, it is impossible to explain adequately the impact of a common language on foreign trade. Linguistic diversity at home also promotes foreign trade. But the most significant linguistic influence of all on foreign trade is the ability to read and write any language whatever. Besides these basic results, the study examines three special issues: whether English and other European languages are more effective than other languages in promoting foreign trade; whether there are substitution effects of a common language in foreign trade; and whether network externalities of language impinge on foreign trade.

JEL Classification: F10 and F40
Keywords: foreign trade, gravity model and language

Jacques Mélitz
Department of Economics
University of Strathclyde
100 Cathedral Street
Glasgow
G4 0LN
Tel: (44 141) 548 3581
Fax: (44 141) 552 5589
Email: jacques.melitz@strath.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100052

* Many thanks for valuable comments to Francis Kramarz and Eric Maurin at CREST-INSEE in Paris.

Submitted 09 September 2002

Gravity models provide ample evidence that a common language has a significant impact on bilateral trade. In the typical tests, the flow of bilateral trade between two countries is the dependent variable, while the explanatory variables comprise the respective output levels of the two countries. Other explanatory variables consist of indicators of barriers or aids to trade, prominently including geographical distance. When a dummy variable for language is added, and this variable receives a value of one when there is a common language between the trading pair and zero otherwise, the dummy is generally highly significant. (For bibliography and examples, see Frankel (1997); Frankel and Rose (2002) offer fresh results.) Yet even if a common language thus clearly helps foreign trade, numerous questions remain. The previous dummy variable is itself a point of interrogation. This variable is often scored as one for a pair of countries when a minority of the population in one can understand the people in the other (sometimes less than 5%, as in the instance of Niger-Burkina Faso or Pakistan-Kenya). But it is scored the same way if over 95% of the population in both places can communicate directly (Germany-Austria). The implication is that the actual number of people who can communicate directly does not matter: if only some other criteria are met, a small proportion of bilingual individuals can make the desired market information available to everyone. In other words, an adequate system of translation will do just as well as direct communication in fostering foreign trade. But is this true?

Other basic questions abound. Is the world's dominant language, English, more effective than the rest in promoting trade? Does a common language boost trade between some countries at the expense of others or increase trade in general? Are the frequent references to the network externalities of a common language correct? Further still, how important is a common language inside a country? Does linguistic diversity at home reduce domestic trade? If so, is there compensation through foreign trade? Last but not least, does literacy promote trade? This whole host of questions about language and trade forms the subject of this inquiry.

Our typical use of binary variables in economics as indicators of a common language largely reflects the difficulty of quantifying the numbers of speakers of different languages in a country. But some headway is possible. Perhaps the most extensive basis for progress today is Grimes (2000), now in its 14th edition since first appearing in 1951. This work is a massive

effort to condense the information supplied by the entire profession of ethnologists about world languages. Although the study has served in recent economic research (see Hall and Jones (1999) and Rauch (1999)), it has never been used to construct a general quantitative index of the ability to communicate directly in foreign trade. Boisso and Ferrantino (1990) are probably the only ones to have attempted to construct such an index thus far. But their pioneering work relies on a summary treatment by Katzner (1986).

As an added interest, the latest edition of Grimes (2000) contains an index of linguistic diversity for all countries present in the study for the first time. The index concerns "the probability that any two people in the country picked at random will have different mother tongues" (Grimes (2000), p. x). The higher the index – the closer to one – the higher the probability that a random pair of individuals will have different mother tongues (see Lieberson (1981)). It would have served me better to have an index of the probability that any two people at random will not be able to communicate through a common language, since two people with different mother tongues may evidently both be fluent in a third language. But the linguistic diversity index in Grimes is a good reflection of my preferred one. Mauro (1995) uses a similar index: namely, Taylor and Hudson's (1972) index of ethnolinguistic fractionalization (which derives entirely from a detailed Soviet linguistic study dating 1964). Despite the sociological emphasis of the title, "ethnolinguistic fractionalization," Taylor and Hudson's index pertains to the identical issue whether two randomly chosen individuals in a country will share the same maternal language. Since Mauro (1995), Alesina and Wacziag (1998), Easterly and Levine (1997) and La Porta et al. (1999) have also exploited the Taylor-Hudson index. However, these authors use the Taylor-Hudson index to treat issues of societal division and government behavior,[1] while I shall use the Grimes index strictly in connection with communication. In regard to the 108 "countries" which both Grimes and Taylor-Hudson score for linguistic diversity, the correlation coefficient between the two indices is .85.

In order to infer percentages of speakers of different languages, I shall employ Grimes

[1] This practice has aroused criticism from Collier (2001). Easterly and Levine (1997) provide some detailed discussion of the Taylor-Hudson index (derived from the Soviet study) and other closely related indices (which they have used and La Porta et al. (1999) have followed them).

(2000) predominantly for all the countries in my sample: 185 altogether. These are the same 185 that serve in a recent study by Frankel and Rose (2002), from which I draw most of my other data. My only supplementary source will be the *CIA country factbook* – a frequent reference among economists. But I will rely on the *CIA factbook* mainly for literacy rates (though this will have repercussions, as I will explain). My decision to stick to Grimes and the *CIA factbook*, following a broader search, is largely aimed to facilitate reproduction.

The basic results can be summarized as follows. First, a common language promotes international trade both directly and via translation, that is, both as a result of sheer numbers who can communicate person to person, and as a result of an established network of translation. One cannot say whether direct communication or a network of translation is more important. Second, a common language increases foreign trade in the aggregate. There is some substitution: that is, an increase in foreign trade between two sets of nationals who share a common language (directly or through translation) does come at the expense of trade with different foreigners; but the overall effect on foreign trade is positive. Third, a common language exerts positive network externalities on foreign trade. But these externalities issue from within the trading countries themselves. There is no evidence that the broad usage of a language elsewhere affects bilateral trade between two countries. Thus, despite the dominant position of English as a world language, English is no more effective in promoting trade than other major European languages. Still, as a fourth basic conclusion, the major European languages as a group (including English) are more efficient than other languages in promoting trade. Fifth, a diversity of tongues at home does indeed boost foreign trade, and therefore, if only by implication, diminishes domestic trade. Sixth and last, literacy increases foreign trade. Indeed, literacy emerges as by far the most important linguistic influence in the study (with Student *t*s of around 20). As an indication, the impact of a one-percent increase in literacy on foreign trade compensates for as much as a three-percent drop in language diversity inside a country.

The last result concerning literacy depends on treating the ratio of bilateral trade to joint output rather than bilateral trade as the dependent variable. But the emphasis is fully warranted. Typical use of the level of bilateral trade as the dependent variable in gravity models is

associated with the treatment of joint output as an exogenous influence on trade. Yet this treatment is inconsistent with the data. Furthermore, literacy and output per capita are highly correlated in the data as well (not surprisingly, in light of the influence of literacy on output per capita: see, for example, Barro (1991)). Consequently, when literacy, output and population serve side by side in explaining bilateral trade, the respective influences of the three become confused. I will return to the issue below.

The discussion will begin by developing the specific gravity equation that will serve in the empirical analysis (Section I). Next, I will explain my two indices of a common language, the one relating to direct communication and the other concerning translation. Section III will present the elementary econometric results. Section IV will then extend the analysis to three specific issues: first, the effects of different languages; second, the distinction between the scale (trade-creation) and the substitution (trade-diversion) effects of a common language; and third, the contribution of network externalities of language to bilateral trade. The final section will contain some general discussion and suggestions for future research. All the raw linguistic series, including those I constructed, are in the data appendix.

I. The gravity model

The gravity model is particularly fitting here because it focuses on the general barriers to trade, apart from the linguistic ones. Without controlling for other obstacles and aids to trade besides language – distance, political association, ex-colonial relationships, and the rest – it would be difficult, if not impossible, to draw inferences about linguistic effects, as such.

In using the gravity model, I shall limit attention to influences on total bilateral trade, without regard to differences between exports and imports. This opens the way for an important simplifying assumption: namely, that trade frictions raise the price to the importer above the exporter's price by the same percentage, regardless whether the goods move one way or the other. On that supposition, balanced trade results, given the other assumptions of gravity models: that all countries specialize in the production of a separate good, and that utility functions are identical, homothetic, and CES everywhere. There is then nothing except aggregate trade to investigate. With trade costs the same whichever direction goods move, Anderson and van Wincoop (2000) also show that a particularly simple form of the gravity equation follows. It

is:

$$(1)\ T_{ij} = \frac{Y_i Y_j}{Y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma}$$

where T_{ij} is the trade flow in either direction between countries i and j, Y_i and Y_j are the respective incomes of the two countries, Y^W is world income, σ is the elasticity of substitution between different goods, t_{ij} is one plus the ratio of the costs attributable to trade frictions in relation to the export price (regardless whether this price (fob) is p_i or p_j), and P_i and P_j are the respective Dixit-Stiglitz consumer-based price levels in the two countries. To take the case of P_j:

$$(2)\ P_j = \left[\sum_i \left(\beta_i p_i t_{ij}\right)^{1-\sigma}\right]^{1/(1-\sigma)}$$

where the summation sign embraces all i prices inclusive of p_j, and β_i is the distribution parameter of the utility function (and the corresponding equation holds for P_i). Evidently, σ must be greater than one, as empirical work tells us is overwhelmingly the case, if there is to be a negative effect of t_{ij} on trade.

Distribution costs can be easily incorporated in the preceding model, and this is important since language plainly affects trade partly by modifying costs of distribution. As Baier and Bergstrand (2001) recently show (in line with Bergstrand (1985)), injecting distribution costs is no problem: it merely requires supposing that a CES relationship applies to the "transformation" of goods into sales between different national markets. So long as the elasticity of "transformation" is the same whether goods move one way or the other, the gravity equation retains the same structural form as equation (1). The only differences are that the parenthetical expression (t_{ij}/P_iP_j) in this equation will comprise more terms and the entire expression will be raised to the power $\gamma/(\gamma+\sigma)$ rather than $1-\sigma$, where γ is the elasticity of transformation. From the standpoint of estimation, however, these differences are minor. At least, they are so if only the prices are constant and set equal to one, two conditions that can be easily met if we stick to the cross-sectional dimension (see Baier and Bergstrand, sec. 2.3.4).[2]

[2] Of course, Bergstrand himself has always favored a more complicated version with distinct export and import behavior in his considerable work on gravity models (including the paper with Baier).

I shall construe the critical t_{ij} term in equation (1) to be:

$$(3)\quad t_{ij} = \prod_{k=1}^{k=m} u_k^{\gamma_k} \times \exp\left(\prod_{k=1}^{k=n} \eta_k v_k \right)$$

where the u terms are continuous variables, and the v ones are 0-1 dummies. Specifically, the estimated gravity model will then take the form:

$$\begin{aligned}(4)\ \log T_{ij} = \text{constant} &+ \alpha \log (Y_i Y_j) + (1-\sigma)\, \gamma_1 \log u_1 \ldots + (1-\sigma)\, \gamma_m \log u_m \\ &+ (1-\sigma)\, \eta_1 v_1 + \ldots + (1-\sigma)\, \eta_n v_n \\ &+ (1-\sigma)\, \lambda_1 \log w_1 + (1-\sigma)\, \lambda_r \log w_r + e_{ij}\end{aligned}$$

The coefficient α of the log $(Y_i Y_j)$ term should evidently equal one. The new w_r terms refer to possible influences on balanced-bilateral-trade that are consistent with the assumptions of the model but are not reflected in any other way. A fitting example of w_r would be differences in climate. If large differences in climate between country pairs imply better opportunities for trade between them than with other countries that have more similar climates, the variable should be added. It would not be reflected otherwise – in consumer tastes, general price levels, or trade frictions – but would be consistent with balanced trade. On the other hand, the exchange rate would be ineligible as a w_r term, since this variable necessarily has opposite implications for two trading partners. As I indicated before, the failure to pay attention to the p_i and p_j components of the $P_i P_j$ term of equation (1) (see equation (2)) depends on the idea that they may be considered constant and set equal to one (compare Anderson (1979) and Deardorff (1996)). The term e_{ij} is white noise associated with the dependent variable, bilateral trade.

Apart from the language variables, the precise u, v and w terms in the estimates will be:

log u_1, …., log u_m = log of (product of) relative distance, log of (product of) remoteness, log of (product of) population, log of (product of) land area.

v_1, …. , v_n = adjacency, number of landlocked in pair, currency union, political union, free trade area, ex-colonial relationship, ex-common colonizer.

log w_1 = North-South difference.

The first two u variables decompose bilateral distance into the distance divided by re-

moteness (more exactly, the distance squared divided by the remoteness of one trading partner times that of the other), and remoteness itself. The remoteness of a particular country is the straight-line average of the distances of the country from all the others in the sample. This decomposition permits a separate estimate of the effect of distance in shifting trade away from more distant foreign trade partners toward closer ones, and between all foreigners and fellow countrymen. Frankel and Romer (1999) provide a particularly cogent explanation for the next two variables, population and land area. As they observe, the two reflect the negative pull of domestic trade on foreign trade. The more people there are at home, the wider the opportunities to trade domestically, and therefore without bearing the costs of foreign trade. Land area, in turn, reflects the costs of internal trade associated with transport and distance. This variable can be interpreted as a measure of internal distance (which has sometimes been measured, perhaps more intuitively, as a line between two domestic locations, as in Wei (1996)). Imagine the hypothetical enlargement of a country at a given population size. As the country grows, foreigners – or at least some of them – move further away from the national population. Therefore, domestic trade should benefit at the expense of foreign trade.

The v variables assume the values of zero or one, except for the number of landlocked in a pair, which can be two. Adjacency and the number of landlocked evidently relate to transport and distance generally. The last four v variables are indices of political association that Frankel and Rose have successfully used. The concern with ex-colonial relationships and an ex-common colonizer is particularly important, since former colonial attachments have strong linguistic consequences, and if colonial variables were left out of the analysis, any significance of language could be attributed to earlier colonial attachments. The dummy variables for the political associations also provide some reflection of protectionism, which is not otherwise taken into account. There exist detailed indices of trade protection, but these apply to a much narrower sample of countries. The only w variable in the study is the North-South difference between trading partners, as measured by the absolute difference in the latitudes of the two. This variable reflects differences in climate and seasons, which, as I have argued before (Melitz (2001)), may foster trade based on comparative advantage.

The next section turns to the principal concern, language.

II. The language variables

The meaning of a common language between two countries ceases to be obvious once we admit translation. In principle, a small group of bilinguals could make all market information available to each person in his or her preferred tongue in both countries. This could be done on-line via an electronic system. Or alternatively, one could imagine wholesalers employing translators to provide all market information to everyone further down the distribution chain in their own language. In either case, the marginal cost of the translation services could be zero to the final users. This last condition is essential. To see why, consider the analogy with money. National currencies can also be converted through banks. But the purchaser of foreign currency pays the exchange cost in every transaction. If it were necessary to pay for the translation of every separate message in a foreign tongue, there would be no question of a common language. But if new foreign-language messages can be gotten in one's home tongue for free, then the linguistic barrier may be non-existent or weak. As I have indicated before, the usual treatment of a common language as an all-or-nothing condition suggests that translation is a free good at the margin.

Indeed, it is possible to turn the whole issue of a common language on its head, and to ask how there can possibly be any language barrier if translation can link up all world languages to one another. There are two obvious answers; but while their joint significance is clear, their relative importance is not. The first regards the costs of translation – the social overhead costs of enabling the widespread distribution of translation services as well as the costs of dealing out such services to people individually. Even if true, based on the previous paragraph, that once an appropriate foundation for translation has been laid, the services can be provided to individuals at a negligible cost, the social overhead costs can be very important, and the usual treatment of a common language in foreign trade says that they are. According to this treatment, the problems of mounting and maintaining a language network are so large that, regardless of population size and number of languages, only two common languages exist between any pair of countries at most. The second answer to the query why language barriers are significant is the need for direct communication. Perhaps the clearest indication we have of the separate importance of direct communication in trade comes from appli-

cations of gravity models which show that immigrants increase trade with their country of origin (see Gould (1994), Head and Ries (1998), and Dunlevy and Hutchinson (1999)). One reason for this effect may well be the immigrants' ability to speak their native language (as well as their capacity to translate). Accordingly, I will develop separate measures of a common language pertaining to a communication network and direct communication.

The first measure, titled "open-circuit communication," will require no particular number of speakers for a common language, but simply demand that the language be either official or widely spoken in both countries (in any combination). By "widely spoken," I will mean that 20 percent or more of the population possesses the language.[3] In accordance with previous work, I will also recognize only two "open-circuit" languages at most in any country. Seldom will this last limitation make any difference, and where it does, I will retain the two languages with the widest international currency. This will essentially mean sticking to Arabic, English and French in some African examples where one or another of these languages could have been dropped in favor of Swahili, Hausa or Fulfulde, as the case may be. Open-circuit Communication has a value of one if the required condition is met (in any combination of the two branches) and zero otherwise. It cannot be overemphasized that this measure results from an effort to make sense of the indices of a common language that have served thus far. Fifteen open-circuit languages result from the criterion; all are listed in Table 1.

The second measure, "direct communication," depends on the percentage of speakers in both countries. Fortunately, those percentages needed to be calculated only when they were large enough to make any statistical difference in the analysis, or in explaining bilateral trade. I eventually retained 4% as the minimum figure for a language to count for Direct Communication. This limited the number of relevant languages to 29 out of the total of over 5,000 in the 185 "countries" in the study (including some overseas departments and territories). Lowering the required percentage to 3 would have increased the relevant number of languages by a dozen or so without affecting the estimates perceptibly. In constructing the figures, I treated

[3] Any figure in the 10% to 30% range would have made only modest difference, in light of the importance of official status. de Swaan (2000) contains some interesting discussion of the frequent assignment of official status to a minority language in multilingual societies.

different dialects, creole and pidgin versions of a language as equivalent. Where numbers of speakers could not be inferred from Grimes in any other way, I used literacy as a guide, while paying attention to the alternative language(s) to which the literacy rates might refer. In those instances, the *CIA country factbook* came into significant play, since, as mentioned before, I relied on this source for the literacy rates, and only used Grimes to fill in missing values. (My reason for switching to the *factbook* for literacy rates lies in the much wider discrepancies in the dating of the variable in Grimes.) The figure for Direct Communication obtains by summing up the products of the respective percentages of speakers over all the relevant languages (at least 4%) in the two trading countries. In principle, those values could have exceeded one because of bilingualism. But very few such cases arose. In those cases, I set the numbers equal to one, which was equivalent to a general normalization.

Table 1 lists all 29 languages figuring in the Direct Communication index and shows all 15 of those that also serve as open-circuit languages. In the case of Open-circuit Communication, the cited languages are strictly "source" languages. The "destination" languages – those in which the messages are received – represent a large number, probably over a hundred, which remains undetermined here. (There are about 250 languages in the world with over a million speakers, and some small languages, like Maltese, receive strong government support.) It is noteworthy that a few of the 15 open-circuit languages are small, whereas many big languages are missing from the 29 direct-communication ones. Big languages may be missing because of strict importance in domestic trade. Japanese is an example. They could also be missing because of lack of trade data. Striking examples come from the ex-Soviet Union and they include Russian.

The appendix contains my scoring of the linguistic information underlying the construction of the language variables. There, I also display the assignment of languages by country by Frankel and Rose (2002). The latter assignment follows corrections for a few slips in the working paper version of the article which happened to catch my attention in an earlier study (Melitz (2001)) (most of which Frankel and Rose have since repaired). Table 2 shows the correlation matrix relating to my two indices of a common language and their single one over the (approximately) 30,000 observations of bilateral trade in the statistical analysis. The

indices of literacy and language diversity in the study are also included in the correlation matrix. As can be seen, the correlation between Open-circuit Communication and Frankel and Rose's (FR) Common Language is only .75 – rather low if we consider that the two are probably meant to signify the same thing. This imperfect correlation stems entirely from my more frequent assignment of "open-circuit" languages to countries than theirs (with the single exception of Mauritania, which they list as French-speaking and I do not). In the first place, there are six open-circuit languages in my work that FR do not recognize at all: Danish, Greek, Turkish, Persian, Hindi and Malay. In addition, I assign the other nine open-circuit languages many more times than they do (for example, Spanish in Gibraltar, Dutch in the Dutch Antilles, English in St. Helena, French in Algeria, Morocco and Tunisia, etc.). Also noteworthy is the fact that the correlation coefficient between my two language indices, Open-circuit and Direct Communication, is high, 0.73, but not so high as to undermine the distinction – especially when we consider that both indices are zero in nearly three-quarters of the cases.

Table 2 further displays the lack of any correlation of my two indices of linguistic ties with literacy and linguistic diversity. But unsurprisingly, these last two variables are significantly negatively correlated with one another (–.43). We would expect so since a large number of major languages in a country will often be a sign of low levels of market integration (not always, if only because of immigration). Consequently, such survival tends to go together with poverty and illiteracy.

III. The basic results

Virtually all of the data for the variables besides language in the study come from Frankel and Rose (2002). I owe an enormous debt to Rose for making this data public on his website. There are only two changes here in the relevant Rose database (apart from the aforementioned corrections concerning language). First, I measure geographical distance differently. Whereas he locates countries at their geographical center (in conformity with the CIA), I place them wherever their most populous city stands (as found in the CD-rom *encarta*). Second, I consider all departments and territories of a country as automatically belonging to a free trade zone in the mother country. As a result, my dummies for a common country and free trade area are mutually exclusive.

The first column of Table 3 shows the result of the test of equation (4) in the usual form, with (the log of) bilateral trade (nominal imports plus exports in dollars deflated by the U.S. GDP chain price index and divided by two) as the dependent variable. As in Frankel and Rose, the test covers observations for six separate years at five-year intervals, starting with 1970 and ending in 1995, and therefore includes controls for the individual years. Robust standard errors are shown (after correction for clustering of data for individual trading pairs). All of the earlier hypotheses are confirmed. The results are basically the same as those reported in Melitz (2001), which in turn do not differ from Frankel and Rose (2002) as regards the common variables in both studies (that is, all of them except the decomposition of distance between relative distance and remoteness, North-South, and the language variables). The rest of the discussion focuses on language.

First and foremost, all four linguistic variables enter significantly with the right positive signs. Of the four, the outstanding influence is the diversity of languages at home. But all other three linguistic variables have Student *t*s above 3. The next two columns probe more deeply into the influences of Open-circuit and Direct Communication. These two influences partly overlap, and there could be some negative interaction between them. In order to see, I replaced them by their average – their sum divided by two – in column 2, and I termed this average "Common Language" (a variable of the same dimension as either one of them alone). The coefficient of Common Language shows up as essentially the sum of the two separate influences. On this evidence, the two influences bear distinct, additive impacts. Column 3 injects the FR index of a common language into the equation, and thereby focuses on the relative merits of FR's linguistic variable as opposed to my two. The FR index is totally unimportant. Its presence does not even alter the estimates and Student *t*s of Open-circuit and Direct Communication. I experimented with the FR index in combination with either one of my two linguistic variables alone or their straight-line average. Either one of them totally dominates the FR index, with Student *t*s more than twice as high, and the use of the average, Common Language, completely removes any significance of the FR index. To all evidence, the systematic and separate measurement of Open-circuit and Direct Communication is worthwhile, and facility of translation and direct intercourse represent distinct influences on

foreign trade.

There is one unsatisfactory feature of the previous tests, however: the 1.4 output-elasticity of trade instead of one. This coefficient is impossible to interpret independently of the negative coefficient of population in the equation, since output and population are closely correlated (compare Frankel (1997, pp. 57-61)). But, at least on my interpretation, theoretically the two influences are distinct. One concerns the feature of homothetic preferences, whereas the other reflects the opportunities to trade at home and free of any of the obstacles to foreign trade. In addition, joint output is not really independent of bilateral trade, and most important of all at present, literacy interacts with per capita output. If we regress (the log of) joint output (Y_iY_j) on (the log of) bilateral trade, (the log of) joint population, and joint literacy, control for individual years, and correct the standard errors for clustering, we get the following estimate:

$$Y_iY_j = .19 \text{ bilateral trade} + .84 \text{ product of populations} + 2.76 \text{ product of literacy rates}$$
$$(.01) \qquad (.002) \qquad (0.03)$$
$$R^2 = .98 \qquad n = 31249 \qquad \text{number of clusters} = 8202$$

(robust standard errors in parentheses). Evidently, therefore, it is impossible to distinguish the respective impacts of output, population and literacy on bilateral trade in the previous equations.

The next two estimates of equation (4) impose a unitary elasticity of influence of joint output on trade.[4] The variable on the left hand side is then the ratio of bilateral trade to the product of the two countries' outputs. Column 4 replicates the estimate in column 1 following this change. The R-square jumps up from 65 to 89 percent. The product of the two countries' populations remains significant while its negative coefficient drops. However, this negative coefficient now can be properly interpreted, for the first time, as reflecting the impact of opportunities to trade with fellow citizens in diminishing foreign trade. The influence of remoteness also rises and becomes roughly equivalent to that of relative distance, though the coefficient of relative distance is still estimated far more precisely. (I have no explanation for this last alteration.) However, the most notable changes of all relate to the four linguistic variables.

[4] Compare Harrigan (1994) and Frankel and Rose (1998).

Those variables retain their signs and significance, but the impact of literacy shoots up. The most powerful linguistic influence on foreign trade in the study – by far – now emerges as the ability to read and write. According to the estimates, if we compare percentage-point for percentage-point, much more can be done to increase trade between two countries by promoting literacy than a common language.

The last column of Table 3 facilitates interpretation of the relative influence of the other linguistic variables besides literacy. Once again, I substitute Common Language for Open-circuit and Direct Communication. Once more, this single variable has a coefficient that matches the sum of the coefficients of the previous two. But the impact of Common Language now emerges as, if anything, higher than that of linguistic diversity. Its statistical significance is higher too. Thus, common linguistic ground with foreigners seems to be at least as important as linguistic diversity at home in fostering foreign trade. This is not necessarily surprising. Linguistic diversity encourages trade with all foreigners indiscriminately, whereas a common tongue promotes trade with those foreigners with whom communication is especially easy. From this perspective, the surprising thing may be that Common Language does not dominate language diversity even more.

IV. Further Tests

(a) *Different languages*

Such is the supremacy of English as a world language today that there is cause to inquire whether English is more effective in promoting trade than other languages. It is difficult to see how English could be able to convey information better than other languages in person-to-person communication – that is, except through external effects, which I will consider separately. But in the case of open-circuit communication, the greater effectiveness of English could come from economies of scale in the diffusion of messages, independently of any externalities. The question thus arises whether English exerts a greater influence as a source language on trade than the other open-circuit languages do.

I begin in Table 4 by repeating the estimate in column 4 of the previous table for convenience. The second column shows the result of the extreme assumption that English provides the only effective circuit of communication, or, in other words, that the other languages

only affect trade through person-to-person communication. The general fit of the equation worsens somewhat and both the coefficient and the significance of Direct Communication rise. The third column is even more to the point: it shows what happens when we introduce Open-circuit Communication in English as a separate variable side by side with Open-circuit Communication. The latter emerges as the larger and more significant influence of the two. In addition, the difference between the two is patently not statistically significant. Hence, there is no evidence that English bears any impact on trade beyond its heavy contribution to Open-circuit and Direct Communication. I experimented next with English, Spanish, and French – or the three most important European languages in the study – as the only languages providing an effective circuit of communication. Insufficient justification arises. On the other hand, the idea that the European languages as a group are the only ones to serve as source languages in trade does receive notable support. Column 4 shows the result of the experiment (which excludes Arabic, Chinese, Hindi, Malay, Persian and Turkish from Open-circuit Communication). The estimate clearly improves, and this last improvement holds up in all my subsequent experiments. I will therefore only report further results with Open-circuit Communication in the European languages, and in summarizing, I will use a straight-line average of this Open-circuit variable and Direct Communication, termed Common Language*. The estimate of column 1 with Common Language* instead of Common Language is shown in column 5.

(b) *Substitution or scale effects*

Does a linguistic tie raise trade between two countries at the expense of trade with third countries or increase foreign trade in the aggregate? This sort of question has arisen before in connection with some of the other variables in the analysis: namely, the political associations: free trade agreement, currency union, and so forth. The basic approach has been to introduce a dummy variable for bilateral trade between two countries if one or both of them belong to the relevant political association with a third country but not with one another. The sign and statistical significance of the dummy variable tells us whether a substitution or a scale effect is present or whether neither one nor the other is true. However, this method cannot serve here – at least, not without modification. There are too few cases of bilateral trade between countries where neither one scores positively for Common Language* with anyone.

As a result, any dummy of the previous sort, if used together with Common Language*, would cover too much of the field (whenever one is positive, the other is predominantly zero, and inversely). It would then prove impossible to distinguish statistically between the influence of the two. But a certain adaptation is possible.

Common Language* varies between values of 0 and 1 with disproportionately few examples of values between .1 and .5, since any pair of countries scoring 1 for Open-circuit Communication in a European language must register at least .5 for Common Language*, and except for country pairs with a common open-circuit language in a non-European language, all others can score at most 0.1 for Common Language* (since they must score less than 0.2 for Direct Communication). It is then possible to construct an index of trade between country pairs that have a value below .1 for Common Language* where one *but not the other* member has a value of .1 or over for this variable in trade with some third country. This focuses attention on trade between countries that are generally "isolated" linguistically and countries that are predominantly not. As a result, there is still a large number of observations, constituting a sizable proportion of the sample, where both this indicator and Common Language* will be zero. These observations regard pairs of countries that both have some significant linguistic ties with someone but not with one another. The point of putting those cases aside is simply to make possible a test relating to the other, smaller, and distinct set of cases.

In instances where the previous specialized index of a linguistic barrier should be scored as positive, there is also a choice between scoring it 1 or the absolute difference between the two maximal values for Common Language* in trade with third parties. I experimented with both forms: the 0-1 and the continuous varieties. Both forms yield identical results in the foregoing equations in Table 4. But if substitution or complementary effects are at issue, there is reason to ask whether currency unions, political unions, free trade agreements, and ex-colonial-relationships too yield substitution or complementary effects. This further exploration can be done in the usual manner: by scoring a dummy variable as 1 when a country member of a relevant political association trades with a country outside the association and 0 otherwise. Once those other dummy variables are added, the proposed 0-1 index for lack of linguistic ties drops substantially in significance while the continuous index (with positive

values heavily concentrated in the 0.4-1 range) is much less affected. Therefore, I retained the continuous index.

The first column of Table 5 shows the outcome of the relevant experiment. The equation contains the relevant five new variables for trade with third parties, the one relating to a common language, and the four relating to the political associations. In the case of the political relationships, the results conform to earlier studies, and will receive no further comment (see Melitz (2001)). But in the case of language, our main concern, linguistic ties now appear to have a highly significant negative effect on trade with third parties. Still, this substitution effect is smaller than the positive effect of Common Language* on trade between countries with linguistic ties, and the difference between the two coefficients is very significant statistically (according to a Hausman test). Hence, there remains a net positive influence of linguistic ties on aggregate foreign trade (which could be entirely at the expense of domestic trade).

(c) *Network externalities*

What about the presence of "network externalities" of a common language? Advertisers pay more to publicize at peak times and in well-frequented places. It is entirely plausible that speakers of a language would benefit from larger numbers of other speakers of the same tongue in the market. Perhaps the outstanding formal development of the idea comes from Church and King (1993) (who wrote with the issue of French and English in Canada in mind). A related development is in Lazear (1999), whose presentation has the added interest of centering the external benefits precisely on trade (though he does not focus on externalities himself). Random encounters take place between people. If they speak the same language, a trade occurs. Otherwise, it doesn't. The more individuals in a surrounding that speak the same language, the higher the probability that random encounters will result in trades. It should be noted that both Church and King and Lazear consider the issue to be strictly direct communication, as is not the case here. Numerous passing references to network externalities of a common language can also be found in print. (One outstanding, oft-cited case is Sabourin (1985).) Interestingly, Dowd and Greenaway (1993) close an article on the network externalities of money with a paragraph suggesting the application of their reasoning to language. Yet I know of no previous attempt to test the hypothesis of external benefits of a common lan-

guage.

In trying to do so here, I will entertain two separate but complementary meanings (or manifestations) of external benefits of a common language. The first is intrinsic, the second is not but often intimated. Any network externality of a common language says that the impact of the language depends on the numbers of people who are connected (either directly or through a circuit) rather than merely the percentages. Thus, a given percentage of English speakers in a small community should have a smaller impact on trade than the same percentage in a huge community if network externalities really matter. Beyond numbers, though, what must be relevant is the aggregate income of the people in a language circuit (or who are hooked up). Accordingly, my first measure of a source of external effects of a common language is the aggregate real income of the people with linguistic ties: that is, the product of the percentage value, Common Language*, and the sum of real incomes of the two trading countries combined. Since Common Language* is a constant in the study, I shall use a constant for real income as well, namely, the value for 1990, or in the few cases where only earlier figures are available (at five-year intervals), the latest one.[5]

My second, and more conjectural, measure relates to the external effects of a common language stemming from third-countries. The hypothesis in this case is that worldwide usage is important. A simple interpretation goes as follows: so far as people have a choice of language because they are bilingual (multilingual) or live in a country which receives messages in two open-circuit languages, they will tend to gravitate toward the one with the widest international currency – even if (as concerns Direct Communication) that language is not their preferred one and they do not master it well. Under this hypothesis, a language with widespread usage in the world may wield a larger impact on bilateral trade than another language that is better-known within the two countries. Evidently, this last sort of external effect on bilateral trade cannot be captured by any aggregate of income *within* the two countries.

Accordingly, I constructed a separate measure of income for the previous sort of external effects. In this case, nine languages drop out (including one of the open-circuit languages, Swedish, which strictly functions between Sweden and Finland). This leaves twenty. For

[5] I see no advantage in averaging the (one to six) annual figures for individual country pairs.

every bilateral pair, a different total of these twenty world aggregates of income was constructed. An example will suffice. Consider a trading pair with a Common Language* of .8, composed of .5 for English, .2 for Arabic and .1 for a third language that does not operate elsewhere (or, effectively, is neither official nor spoken by as many as 4% of the population anywhere else). The construction of the index then begins with the separate totals for the world income of people who can receive messages sent in English and in Arabic, based on a previous calculation founded on the data in the language appendix and 1980 real income.[6] The third language is simply ignored. Next, the respective contributions of the trading pair to the preceding English and Arabic income aggregates are deducted. Finally, a coefficient of .5 is applied to the remainder for English, one of .2 to the remainder for Arabic, and the two totals are added up. The log of this sum of world income serves in the estimates.

The results confirm the hypothesis of external effects coming from inside the trading countries but deny any similar external effects coming from the rest of the world. The confirmation of the external effects from within is shown in the second column of Table 5. Both the index of external effects and Common Language* can be seen to be significant at the 5% confidence level. The root mean square error also drops in this estimate. (Note that all 29 "source" languages function in this case.) In addition, the coefficient of Common Language* is lower than before. But this is only to be expected, since the impact of this variable now comes partly from a joint effect with income. The coefficient of Common Language* is also estimated with much less precision. This next deterioration cannot be ascribed to any confusion with the index of external effects, since Common Language* and this index are uncorrelated over the pertinent observations (namely, those where both variables are positive (as in the other examples both of them are zero by construction)). Thus, the drop in precision must be the result of lower statistical variance.

The index of external effects stemming from third countries is never significant except when used alone, that is, in the absence of Common Language* or the preceding index of ex-

[6] Regarding English, the earlier calculation goes as follows. Take one or zero for each country, considered one at a time, depending upon whether English is an Open-circuit Language, add the figure for English under Spoken Languages, divide by two, and multiply by the country's 1990 income. Then add up all the country totals. Repeat for Arabic, etc.

ternal influence. As shown in column 3 of Table 5, if this variable is merely added in the previous equation, Common Language* is the only one of the three that remains significant. The previous index of external effects also becomes unimportant. The fact that the new index of external influence is significant when included alone has no particular importance, but must be attributed to the index's positive correlation of .6 with Common Language* (over the relevant observations).[7] There is independent reason to think so. If the external effects of language coming from third countries were significant, we might have expected English to show up as a separate influence on bilateral trade in the previous test without any control for external effects. Such was not the case. This reinforces the conclusion that the only externalities affecting trade are those coming from inside the borders of the trading parties themselves.

(d) *Country-specific fixed effects*

As a final test, consider the result of adding fixed effects for all "countries" (185 fixed effects in all, 25 of which drop out). Two more factors then explain each trade observation, T_{ij}: a dummy variable for country i and another for country j. Such a test can only interfere with the significance of the variables whose values depend on the individual country alone: namely, remoteness, population, land area, landlocked, language diversity and literacy. Indeed, the only reason why the influence of these variables may still be possible to estimate at all in these tests is that there are a number of observations (1 to 6) for individual country pairs at five-year intervals. (This then opens up the possibility of different pairings at different dates, and different values for identical pairings). However, in principle, the performance of the other variables – namely, those whose values depend on the opposite country in the trading pair – should be unaffected. Accordingly, I have added fixed effects for countries when only the latter variables are retained: namely, relative distance, North-South, adjacency, Open-

[7] It may be important to explain why this .6 correlation is not inconsistent with the absence of any correlation between the other index of external effects – those coming from within the trading pairs – and Common Language*. The world income of English speakers is many times greater than the world income of speakers of any other language. English also contributes heavily to Common Language*. This combination of factors underlies the .6 correlation. On the other hand, there are many cases of tiny English-speaking countries trading with one another, while the index of the effects *within* the trading pairs depends strictly on the income of the trading parties themselves. This explains the lack of any correlation of this next index of external effects with Common Language*.

circuit Communication, Direct Communication, the political variables, and the indicators of substitution or complementarity. (I have also dropped the indicators of external effects, which depend on income, another variable that is defined by individual country). The results are shown in the last column of Table 5.

Except for the indices of substitution or complementarity and two of the political variables (namely, free trade agreement and a common ex-colonizer), none of the relevant coefficients or standard errors are notably affected. In the case of a common language, the significance even goes up. While the coefficient and significance of Open-circuit Communication in a European tongue drop (Student *t* above 3), the coefficient and significance of Direct Communication double. (When used instead of Open-circuit and Direct Communication, Common Language* has a coefficient of .91 and standard error of .07.) As one drawback, though, the impact of a common language on third parties ceases to emerge. On the whole, the test is plainly confirmatory. But the earlier estimates without country-specific fixed effects have the distinct advantage of permitting the identification of particular country-specific characteristics affecting their bilateral trade, and they may be preferred on this ground.

V. Discussion and conclusion

We knew beforehand that a common language promotes trade. This study digs more deeply and introduces many nuances. Difficulties of communicating at home evidently encourage people to trade more with foreigners, regardless of linguistic differences. If people are able to read and write in any language, they can cope better with the problems posed by foreign languages in general. Moreover, as regards the well-known impact of a common language on trade, translation is a factor. One basic result of the study is that much of the significance of a common language in foreign trade reflects ease of receiving translations from one or two selected foreign languages. Depending on the home country, people may be “tuned” to Portuguese, German or Dutch, whether they speak the language or not. Unless we allow for special facilities for obtaining information issuing from chosen source languages, we miss a good deal of the ease of communication through language in foreign trade.

The fact that Open-circuit and Direct Communication have separate and distinct effects on trade, according to the results, may also reflect some specialization. The two channels of

communication may operate largely in trade in different sorts of goods. Open-circuit Communication might be especially important in trade in homogeneous goods, where rudimentary communication may suffice, whereas Direct Communication may serve heavily in trade in heterogeneous goods, which may require more sophisticated intercourse. Rauch (1999) suggests a hypothesis of this sort. He argues that close personal relations between exporters and importers are not necessarily significant in trade of perfectly homogeneous goods, whereas close personal interaction between exporters and importers is required in trade in heterogeneous goods. Because of issues of information, "markets" suffice in one case, while private "networks" are necessary in the other. My hypothesis of the separate roles of Open-Circuit and Direct Communication in the two cases follows easily.

A few other results of the study probe further into the role of language. English seems to offer no particular advantage in foreign trade, but the European languages in toto do: they emerge as better instruments of Open-channel Communication than other languages. In addition, linguistic links between countries tend to boost their bilateral trade partly at the expense of trade with third countries, but not entirely. Finally, there is some evidence that network externalities of language contribute to trade between countries. But so far as this is true, the external benefits depend on the aggregate income of the particular trading pairs. In general, however, network externalities are not as central in trade as they are sometimes imagined to be, and the reason may have to do with the fact that a lot of market information gets transmitted even in the absence of a common language.[8]

In closing, a few extensions of the study may be suggested for future work. Language has been interpreted here strictly as a tool of communication, even though it obviously reflects many aspects of culture as well. The reason for this narrow interpretation is that the only fea-

[8] There are numerous ways of conveying information without a common language: interpreters, dictionaries, numbers, pictures, signs, minimal vocabularies, etc. The earlier evidence relating to the impact of literacy and language diversity on foreign trade underlines this point. Yet it is also worth noting that a common language may be more important in production than in trade. To all indication, even varying shades of ability to communicate *directly* matter a lot in the labor market. Empirical work uniformly shows a considerable impact of linguistic skill on wages: see McManus, Gould and Welch (1983), Chiswick and Miller (1995), and the references in a broad survey by Grin (1996).

tures of the language variables in the study that apply generally regard communication. Other associated features are not always present. However, additional variables reflecting culture or ethnicity would clearly be helpful. Based on previous work, one cultural variable that comes to mind is the stock of immigrants. As the analysis now stands, immigrants are already present since they affect the index of linguistic diversity, if nothing else. But this index marks long-standing, sometimes ancestral, linguistic divisions inside national boundaries as well. The interest of a separate consideration of immigrants lies in disentangling the element of ethnic ties to other communities abroad (which affects tastes, skills, and information) and setting off this element from strict issues of communication. Following this step, the index of linguistic diversity would be easier to assign to matters of communication as such. A separate treatment of immigrants would generally reinforce the idea that the linguistic variables relate strictly to communication.

Another useful extension would be to study trade in goods that are especially connected to language, such as movies, television programs, books, and vocal music. In these cases, the world dominance of English, and the production by the U.S. in particular, is notorious. Cultural products in English sell extremely well in many places where English is otherwise secondary in foreign as well as domestic trade. Broadly, how come English plays no special role in facilitating foreign trade but dominates the market for language-related products?[9] Evidently, the question cannot be studied within the confines of the current version of the gravity model, which applies strictly to two-way trade. Rather, the investigation would require a more intricate version of the model, which distinguishes between exports and imports and is largely associated with Bergstrand (1995).[10]

Finally, my treatment of the language variables as fixed effects can be questioned. Literacy rates have risen substantially in the last fifty years. Migrations accelerated in the nineties in many parts of the world. Spanish is now significant in the U.S., Russian in Israel, etc. I have treated the data furnished in Grimes (2000) as roughly contemporaneous. In doing so, I have

[9] For some of the data, and a limited effort to answer the question, see Melitz (2000).

[10] It may be noted that Gould (1994), Head and Ries (1998), and Dunlevy and Hutchinson (1999), who distinguish between the influence of immigrant links on imports and exports, all rely on specifications that relate to Bergstrand's.

followed the practice in economics of treating all indicators of a common language and linguistic diversity in foreign trade as slow-moving variables that can be regarded as constants.[11] Moreover, Grimes (2000) corroborates the previous practice by furnishing unique figures for linguistic variables, despite wide discrepancies in dating, when dates appear. Yet from the standpoint of trade analysis, there is little doubt that more coherent series for literacy, Open-circuit Communication and Direct Communication could be constructed from a broad variety of national sources. Something could be done, in the process, to show the evolutions of all three variables. Apart from an improved ability to test earlier hypotheses, one outcome would be the opportunity to tackle the question of the relationship between linguistic changes and the increase in the number of countries in the post-World War II period, a basic topic that Alesina and Spolaore (1997) have effectively raised.

[11] This description applies to trade, not labor. Labor studies of the impact of linguistic skills on wages notably make use of time series evidence (see the references in note 8).

REFERENCES CITED

Alesina, Alberto and Enrico Spolaore (1997). "On the number and size of nations," *Quarterly Journal of Economics*, 112, 1027-1056.

Alesina, Alberto and Romain Wacziarg (1998). "Openness, country size and government," *Journal of Public Economics* 69, 305-321.

Anderson, James (1979). "Theoretical foundations of the gravity equation," *American Economic Review*, 69, 106-116.

Anderson, James and Eric van Wincoop (2000). "Gravity with *gravitas*: A solution to the border problem," Working Paper, Boston College.

Baier, Scott and Jeffrey Bergstrand (2001). "The growth of world trade: tariffs, transport costs, and income similarity," *Journal of International Economics*, 53, 1-27.

Barro, Robert (1991). "Economic growth in a cross section of countries," *Quarterly Journal of Economics*, 106, 407-443.

Bergstrand, Jeffrey (1985). "The gravity equation in international trade: Some microeconomic foundations and empirical evidence," *Review of Economics and Statistics*, 71, 143-153.

Boisso, Dale and Michael Ferrantino (1997). "Economic distance, cultural distance, and openness in international trade: Empirical puzzles," *Journal of Economic Integration*, 12, 456-484.

Central Intelligence Agency *Country Factbook*, U.S. government printing office, available online.

Chiswick, Barry and Paul Miller (1995). "The endogeneity between language and earnings: International analyses," *Journal of Labor Economics*, 13, 246-248.

Church, Jeffrey and Ian King (1993). "Bilingualism and network externalities," *Canadian Journal of Economics*, 26, 337-345.

Collier, Paul (2001). "Implications of ethnic diversity," *Economic Policy*, 32, 127-166.

Deardorff, Alan (1998). "Determinants of bilateral trade: Does gravity work in a neoclassical world?" in Jeffrey Frankel, ed., *The regionalization of the world economy*, The University of Chicago Press.

de Swaan, Abram (2001). *Words of the World*, Polity Press: Cambridge, UK.

Dowd, Kevin and David Greenaway (1993). "Currency competition, network externalities and switching costs: Towards an alternative view of optimum currency areas," *Economic Journal*, 103, 1180-1189.

Dunlevy, James and William Hutchinson (1999). "The impact of immigration on American import trade in the late nineteenth and early twentieth centuries," *Journal of Economic History*, 59, 1043-1062.

Easterly, William and Ross Levine (1997). "Africa's growth tragedy: Policies and ethnic divisions," *Quarterly Journal of Economics*, 112, 1203-1250.

Frankel, Jeffrey (1997). *Regional trading blocs in the world trading system*, Institute for International Economics.

Frankel, Jeffrey and David Romer (1999). "Does trade cause growth?" *American Economic Review*, 89, 379-399.

Frankel, Jeffrey and Andrew Rose (1998). "The endogeneity of the optimum currency area criteria," *Economic Journal*, 108, 1009-1025.

Frankel, Jeffrey and Andrew Rose (2002). "An estimate of the effect of common currencies on trade and income," *Quarterly Journal of Economics*, 117, 437-466.

Gould, David (1994). "Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows," *Review of Economics and Statistics*, 69, 301-316.

Grimes, Barbara, ed. (2000). *Ethnologue: Languages of the world*, Summer Institute of Linguistics, International Academic Bookstore, Dallas, Texas, 14th ed.

Grin, François (1996). "The economics of language: survey, assessment, and prospects," *International Journal of the Sociology of Language*, 121, 17-44.

Hall, Robert and Charles Jones (1999). "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics*, 114, 83-116.

Harrigan, James (1994). "Scale economies and the volume of trade," *Review of Economics and Statistics*, 76, 321-328.

Head, Keith and John Ries (1998). "Immigration and trade creation: Econometric evidence from Canada," *Canadian Journal of Economics*, 31, 46-62.

Katzner, Kenneth (1986). *Languages of the world*, Routledge and Kegan Paul.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Schleifer and Robert Vishny (1999). "The quality of government," *Journal of Law, Economics and Organization* 15, 222-279.

Lazear, Edward (1999). "Culture and language," *Journal of Political Economy*, 107, no.6, pt.2, S95-S125.

Lieberson, Stanley (1981). *Language diversity and language contact*, Stanford University Press.

Mauro, Paolo (1995). "Corruption and growth," *Quarterly Journal of Economics*, 110, 681-712.

McManus, Walter, William Gould and Finis Welch (1983). "Earnings of Hispanic men: The role of English language proficiency," *Journal of Labor Economics*, 1, 101-130.

Melitz, Jacques (2000). "English-language dominance, literature and welfare," Working Paper, University of Strathclyde, Glasgow (earlier version, *CEPR Discussion Paper 2055*, 1999).

Melitz, Jacques (2001). "Geography, trade and currency union," Working Paper, University of Strathclyde, Glasgow (earlier version, *CEPR Working Paper 2987*, 2001).

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

Sabourin, Conrad (1985). "La théorie des environnements linguistiques," in François Vaillancourt, *Économie et langue*, Conseil de la langue française, Québec.

Taylor, Charles and Michael Hudson (1972). *World handbook of political and social indicators*, Yale University Press, 2d. ed..

Wei, Shang-Jin (1996). "Intra-national versus international trade: How stubborn are nation states in globalization?" *NBER Working Paper 5331*.

TABLE 1

LANGUAGES

Open-Circuit Languages	**Other Direct-Communication Languages**
Arabic	Albanian
Chinese	Fang
Danish	Fulfulde
Dutch	Hausa
English	Hungarian
French	Italian
German	Javanese
Greek	Lingala
Hindi	Nepali
Malay	Pashto
Persian	Quechua
Portuguese	Swahili
Spanish	Tamil
Swedish	Urdu
Turkish	

TABLE 2

CORRELATION MATRIX

	Open-circuit Communication	Direct Communi-cation	Common Language Frankel-Rose	Literacy	Language Diversity
Open-circuit	1				
Direct	0.73	1			
Frankel-Rose	0.75	0.68	1		
Literacy	–0.10	0.08	–0.05	1	
Diversity	0.13	–0.07	0.09	–0.43	1

TABLE 3: THE BASIC RESULTS

REGRESSAND:	**Log of bilateral trade**			**Bilateral trade ÷ product of GDP**	
Log product of relative distance	–.63 (.02)	–.63 (.02)	–.63 (.02)	–.6 (.02)	–.6 (.02)
Log product of remoteness	–.15 (.07)	–.15 (.07)	–.15 (.07)	–.54 (.08)	–.55 (.07)
Log North-South Difference	.26 (.02)	.26 (.02)	.26 (.02)	.25 (.02)	.25 (.02)
Log product of real GDP	1.4 (.2)	1.4 (.2)	1.4 (.2)		
Log product of Population	–.44 (.02)	–.44 (.02)	–.44 (.02)	–.06 (.01)	–.06 (.01)
Log product of land area	–.18 (.01)	–.18 (.01)	–.18 (.01)	–.17 (.01)	–.17 (.01)
Adjacency (0,1)	.72 (.12)	.72 (.12)	.72 (.12)	.73 (.12)	.73 (.12)
Number of landlocked in pair (0, 1, 2)	–.29 (.04)	–.29 (.04)	–.3 (.04)	–.42 (.04)	–.42 (.04)
Common language: Frankel-Rose (0,1)			**.06** **(.1)**		
Open-circuit communication (0,1)	**.28** **(.08)**		**.26** **(.09)**	**.37** **(.08)**	
Direct communication	**.46** **(.13)**		**.43** **(.14)**	**.35** **(.13)**	
Common language		**.69** **(.07)**			**.73** **(.07)**
Product of linguistic diversity	**.93** **(.11)**	**.91** **(.11)**	**.93** **(.11)**	**.57** **(.11)**	**.58** **(.11)**
Product of literacy rate	**.37** **(.11)**	**38** **(.11)**	**.37** **(.11)**	**1.81** **(.09)**	**1.81** **(.09)**
Currency union (0,1)	1.38 (.18)	1.39 (.18)	1.37 (.18)	1.23 (.19)	1.23 (.19)
Political union (0,1)	.71 (.45)	.74 (.45)	.73 (.45)	.76 (.44)	.75 (.44)
Free trade area (0,1)	1 (.1)	1.02 (.1)	1 (.1)	1.12 (.09)	1.12 (.09)
Ex-colonial relationship (0,1)	1.98 (.13)	1.97 (.13)	1.97 (.13)	2.02 (.12)	2.02 (.12)
Ex-common-colonizer (0,1)	.43 (.09)	.41 (.08)	.42 (.09)	.21 (.09)	.21 (.08)
R^2	.65	.65	.65	.89	.89
RMSE	1.95	1.95	1.95	1.98	1.98

Number of Observations: 31,010. Year-specific fixed effects are not reported. Robust standard errors in parentheses; corrections for clustering of country pairs.

TABLE 4: ENGLISH AND OTHER EUROPEAN LANGUAGES

REGRESSAND: **Bilateral trade ÷ product of GDP**					
Log product of relative distance	–.6 (.02)	–.62 (.02)	–.61 (.02)	–.61 (.02)	–.6 (.02)
Log product of remoteness	–.54 (.08)	–.56 (.08)	–.55 (.08)	–.58 (.08)	–.58 (.08)
Log North-South Difference	.25 (.02)	.26 (.02)	.25 (.02)	.25 (.02)	.25 (.02)
Log product of Population	–.06 (.01)	–.07 (.01)	–.07 (.01)	–.07 (.01)	–.06 (.01)
Log product of land area	–.17 (.01)	–.16 (.01)	–.17 (.01)	–.17 (.01)	–.17 (.01)
Adjacency (0,1)	.73 (.12)	.77 (.12)	.74 (.12)	.74 (.12)	.74 (.12)
Number of landlocked in pair (0, 1, 2)	–.42 (.04)	–.41 (.04)	–.41 (.04)	–.43 (.04)	–.43 (.04)
Open-circuit communication: English (0,1)		**.33 (.09)**	**.19 (.09)**		
Open-circuit communication: European (0,1)				**.44 (.08)**	
Open-circuit communication (0,1)	**.37 (.08)**		**.28 (.09)**		
Direct communication	**.35 (.13)**	**.61 (.1)**	**.36 (.13)**	**.33 (.12)**	
Common Language*					**.81 (.08)**
Product of linguistic diversity	**.57 (.11)**	**.58 (.11)**	**.55 (.11)**	**.57 (.11)**	**.57 (.11)**
Product of literacy rate	**1.81 (.09)**	**1.75 (.09)**	**1.79 (.09)**	**1.79 (.09)**	**1.79 (.09)**
Currency union (0,1)	1.23 (.19)	1.26 (.19)	1.26 (.19)	1.19 (.19)	1.19 (.19)
Political union (0,1)	.76 (.44)	.76 (.47)	.78 (.45)	.75 (.44)	.75 (.44)
Free trade area (0,1)	1.12 (.09)	1.02 (.09)	1.09 (.09)	1.1 (.09)	1.1 (.09)
Ex-colonial relationship (0,1)	2.02 (.12)	2.04 (.13)	2 (.13)	1.98 (.12)	1.98 (.12)
Ex-common-colonizer (0,1)	.21 (.09)	.23 (.09)	.18 (.09)	.18 (.09)	.18 (.09)
R^2	.89	.89	.89	.89	.89
RMSE	1.9825	1.9829	1.9821	1.9815	1.9815

Number of Observations: 31,010. Year-specific fixed effects are not reported. Robust standard errors in parentheses; corrections for clustering of country pairs.

TABLE 5: FURTHER TESTS

REGRESSAND: **Bilateral trade ÷ product of GDP**				
Log product of relative distance	–.61 (.02)	–.61 (.02)	–.61 (.02)	–.69 (.02)
Log product of remoteness	–.62 (.08)	–.62 (.08)	–.61 (.08)	
Log North-South Difference	.24 (.02)	.24 (.02)	.24 (.02)	.19 (.02)
Log product of Population	–.03 (.01)	–.03 (.01)	–.03 (.01)	
Log product of land area	–.17 (.01)	–.17 (.01)	–.17 (.01)	
Adjacency (0,1)	.79 (.12)	.79 (.12)	.76 (.12)	.54 (.13)
Number of landlocked in pair (0, 1, 2)	–.38 (.04)	–.38 (.04)	–.38 (.04)	
Open-circuit Communi-cation: European (0,1)				**.24 (0.7)**
Direct Communication				**.82 (.11)**
Common Language*	**.67 (.08)**	**.37 (.17)**	**.41 (.2)**	
Network externality from within		**.013 (.006)**	**.016 (.012)**	
Network externality from outside			**–.004 (.014)**	
Product of linguistic diversity	**.68 (.12)**	**.66 (.12)**	**.66 (.12)**	
Product of literacy rate	**1.67 (.1)**	**1.66 (.1)**	**1.66 (.1)**	
Currency union (0,1)	1.39 (.19)	1.46 (.19)	1.47 (.2)	1.11 (.22)
Political union (0,1)	1.18 (.45)	1.22 (.45)	1.2 (.45)	.86 (.41)
Free trade area (0,1)	1.3 (.1)	1.34 (.1)	1.35 (.1)	.55 (.14)
Ex-colonial relationship (0,1)	1.45 (.13)	1.42 (.13)	1.42 (.13)	1.52 (.12)
Ex-common-colonizer (0,1)	.31 (.09)	.29 (.09)	.3 (.09)	.66 (.09)
Common language*/ outsider	**–.2 (.05)**	**–.19 (.05)**	**–.19 (.05)**	**–.02 (.05)**
Currency union/ outsider (0,1)	.34 (.04)	.34 (.04)	.34 (.04)	.11 (0.8)
Political union/outsider (0,1)	.31 (.05)	.31 (.05)	.31 (.05)	.16 (.11)
FTA/outsider (0,1)	.4 (.05)	.4 (.05)	.4 (.05)	.15 (.06)

Ex-colony/colonizer/ outsider (0,1)	–.01 (.04)	–.01 (.04)	–.01 (.04)	–.05 (.04)
R^2	0.89	0.89	0.89	0.91
RMSE	1.9581	1.9578	1.9578	1.72

Number of Observations: 31,010. Year-specific fixed effects are not reported. Nor are country-specific fixed effects in the last column. Robust standard errors in parentheses; corrections for clustering of country pairs.

APPENDIX

THE LANGUAGE DATA*

COUNTRY	LIT-ERACY	DIVER-SITY	LANGUAGES (Frankel-Rose)	OPEN-CIRCUIT LANGUAGES	SPOKEN LANGUAGES
Afghanistan	.31	.7	-	Persian	Pashto .45 Persian .4
Albania	.93	.26	-	-	Albanian . 98
Algeria	.62	.31	Arabic	Arabic, French	Arabic .83 French .2
American Samoa	.97	.12	English	English	English . 97
Angola	.42	.76	Portuguese	Portuguese	Portuguese .42
Anguilla	.95	0	English	English	English 1
Antigua & Barbuda	.89	0	English	English	English 1
Argentina	.96	.21	Spanish	Spanish	Spanish .92 Italian .04
Aruba	.97	0	-	Dutch, Spanish	Spanish .74
Australia	.99	.13	English	English	English .99
Austria	.98	.14	German	German	German .98
Bahamas	.98	.01	English	English	English .98
Bahrain	.85	.53	Arabic	Arabic	Arabic .85
Bangladesh	.38	.31	-	-	-
Barbados	.97	.09	English	English	English .97
Belgium	.98	.65	Dutch, French	Dutch, French	Dutch .56 French .56
Belize	.70	.70	English	English, Spanish	English .93 Spanish .35
Benin	.37	.90	French	French	French .37
Bermuda	.98	0	-	English	English 1
Bhutan	.42	.82	-	-	Nepali .08
Bolivia	.83	.68	Spanish	Spanish	Spanish .44 Quechua .36
Brazil	.83	.03	Portuguese	Portuguese	Portuguese .95
Brit. Indian Ocean Terr's	.98	0	-	English	English 1
British Virgin Islands	.98	.24	English	English	English .98
Brunei	.88	.45	-	Malay, English	Malay .88 English .05
Bulgaria	.98	.22	-	Turkish	Turkish .09
Burkina Faso	.19	.76	French	French	French .19
Burundi	.35	0	French	French	French .35
Cambodia	.35	.31	-	-	-
Cameroon	.63	.97	French, English	French, English	French .42 Fulfulde .30 English .21 Fang .05
Canada	.97	.55	English, French	English, French	English .65 French .22
Cayman Islands	.98	.58	-	English	English .98
Central African Republic	.60	.96	French	French	French .6
Chad	.48	.95	French	Arabic, French	Arabic .5 French .48
Chile	.95	.60	Spanish	Spanish	Spanish .93
China	.81	.48	Chinese	Chinese	Chinese .84
Colombia	.91	.03	Spanish	Spanish	Spanish .84
Comoros	.57	.01	French, Arabic	French, Arabic	French .3 Arabic .3
Congo Democratic Re-public	.77	.92	French	French	French .58 Swahili .17 Lingala .12
Congo Republic	.75	.61	French	French	French .7 Lingala .12
Cook Islands	.93	.37	English	English	English .93
Costa Rica	.95	.04	Spanish	Spanish	Spanish .87
Côte d'Ivoire	.48	.91	French	French	French .48
Cuba	.96	0	-	Spanish	Spanish .91
Cyprus	.94	.37	-	Greek, Turkish	Greek .75 Turkish .20

COUNTRY	LIT-ERACY	DIVER-SITY	LANGUAGES (Frankel-Rose)	OPEN-CIRCUIT LANGUAGES	SPOKEN LANGUAGES
Czech Republic	.99	.06	-	-	-
Czechoslovakia	.99	.11	-	-	Hungarian .04
Denmark	.99	.05	-	Danish, German	Danish 1
Djibouti	.46	.58	French, Arabic	French, Arabic	French .46 Arabic .11
Dominica	.94	0	English	English, French	English 1 French .7
Dominican Republic	.82	.05	Spanish	Spanish	Spanish .87
Ecuador	.9	.26	Spanish	Spanish	Spanish .79 Quechua .12
Egypt Arab Republic	.51	.46	Arabic	Arabic	Arabic .97
El Salvador	.71	0	Spanish	Spanish	Spanish .92
Ethiopia	.35	.84	-	English	English .35
Faeroe Islands	.99	0	-	Danish	Danish 100
Falkland Islands	-	0	-	English	English 100
Fiji	.92	.60	English	English	English .92
Finland	.99	.14	Swedish	Swedish	Swedish .12
France	.99	.24	French	French	French .99
French Guiana	.83	.47	-	French	French .83
French So. Antarc.Terr's	1	0	French	French	French 1
Gabon	.63	.53	French	French	French .63 Fang .29
Gambia	.39	.73	English	English	English .39 Fulfulde .17
Germany, West	.99	.18	German	German	German .99
Germany, East	.99	.18	German	German	German .99
Ghana	.64	.79	English	English	English .48
Gibraltar	.80	.50	-	Spanish, English	Spanish .88 English .13
Greece	.95	.14	-	Greek	Greek .99
Greenland	.93	.27	-	Danish	Danish .93
Grenada	.98	0	English	English	English .98
Guadeloupe	.90	0	-	French	French 1
Guam	.99	.64	English	English	English .99
Guatemala	.56	.60	Spanish	Spanish	Spanish .44
Guinea	.35	.75	French	French	French .35 Fulfulde .04
Guinea-Bissau	.54	.85	Portuguese	Portuguese	Portuguese .54
Guyana	.98	.07	English	English	English .98
Haiti	.45	0	French	French	French 1
Honduras	.73	.05	Spanish	Spanish	Spanish .92
Hong Kong	.92	.48	Chinese, English	Chinese, English	Chinese .95
Hungary	.99	.14	-	-	Hungarian .98
Iceland	.99	0	-	-	-
India	.52	.93	English	Hindi, English	Hindi .50 Tamil .07 Urdu .05
Indonesia	.84	.83	-	-	Javanese .42 Malay .06
Iran	.72	.76	-	Persian	Persian .36
Iraq	.58	.65	Arabic	Arabic	Arabic .58
Ireland	.98	.17	English	English	English .99
Israel	.95	.65	-	English	Arabic .12
Italy	.98	.59	-	-	Italian .98
Jamaica	.85	.01	English	English	English 1
Japan	.99	.03	-	-	-
Jordan	.87	.48	Arabic	Arabic	Arabic .87
Kenya	.78	.90	-	English, Arabic	Swahili .78 English .40
Kiribati	.90	.03	English	English	English .90
Korea Democratic Rep.	.99	0	-	-	-
Korea Republic	.98	0	-	-	-
Kuwait	.79	.54	Arabic	Arabic	Arabic 1

COUNTRY	LIT-ERACY	DIVER-SITY	LANGUAGES (Frankel-Rose)	OPEN-CIRCUIT LANGUAGES	SPOKEN LANGUAGES
Laos	.57	.56	-	-	-
Lebanon	.86	.14	-	Arabic, French	Arabic .93 French .65
Liberia	.38	.91	English	English	English .64
Libya	.76	.35	-	Arabic	Arabic .96
Madagascar	.80	.50	French	French	French .80
Malawi	.58	.70	English	English	English .58
Malaysia	.83	.75	Chinese	Malay, Chinese	Malay .50 Chinese .2
Maldives	.93	.01	-	-	-
Mali	.31	.86	French	French	French .31 Fulfulde .11
Malta	.88	.02	English	English	English .88
Martinique	.93	0	French	French	French 1
Mauritania	.38	.19	French	Arabic	Arabic .38 Fulfulde .06
Mauritius	.83	.60	English	French, English	French .83 Urdu .06
Mexico	.90	.13	Spanish	Spanish	Spanish .88
Mongolia	.83	.30	-	-	-
Montserrat	.97	0	English	English	English 1
Morocco	.44	.47	Arabic	Arabic, French	Arabic .65
Mozambique	.40	.92	Portuguese	Portuguese	Portuguese .27
Myanmar	.83	.64	-	-	-
Nauru	.99	.57	-	English	English .65
Nepal	.27	.69	-	-	Nepali .7
Netherlands	.99	.20	Dutch	Dutch	Dutch .99
Netherlands Antilles	.98	.12	-	Dutch, Spanish	Dutch .98 Spanish .84
New Caledonia	.91	.84	-	French	French .91
New Zealand	.99	.10	English	English	English .99
Nicaragua	.66	.08	Spanish	Spanish	Spanish .92
Niger	.14	.64	French	French, Arabic	Hausa .50 French .14 Arabic .14 Fulfulde .08
Nigeria	.57	.88	English	English	Hausa .46
Niue	.95	0	English	English	English 1
Norway	.99	.08	-	-	-
Oman	.80	.68	Arabic	Arabic	Arabic .90
Pakistan	.38	.83	-	English	Urdu .7 Pashto .08
Panama	.91	.23	Spanish	Spanish	Spanish .77
Papua New Guinea	.72	.99	-	English	English .72
Paraguay	.92	.33	Spanish	Spanish	Spanish .92 Portuguese .12
Peru	.89	.35	Spanish	Spanish	Spanish .8 Quechua .17
Philippines	.95	.85	English	English	English .52
Poland	.99	.12	-	-	-
Portugal	.87	.02	Portuguese	Portuguese	Portuguese 1
Qatar	.79	.57	Arabic	Arabic, Persian	Arabic .79 Persian .23
Reunion	.79	.09	French	French	French 1 Tamil .18
Romania	.97	.20	-	-	Hungarian .11
Russia	.98	.27	-	-	-
Rwanda	.60	0	French, English	French, English	French .5 English .5
Samoa	.97	0	English	English	English .97
Saudi Arabia	.63	.56	Arabic	Arabic	Arabic .82
Senegal	.33	.77	French	French	French .3 Fulfulde .23
Seychelles	.58	.07	-	French, English	French .95 English .58
Sierra Leone	.31	.82	English	English	English .31
Singapore	.91	.74	Chinese, English	Chinese, English	Chinese .51 English .27 Malay .16
Solomon Islands	.32	.97	English	English	English .32
Somalia	.24	.2	Arabic	Arabic, English	English .18

COUNTRY	LIT-ERACY	DIVER-SITY	LANGUAGES (Frankel-Rose)	OPEN-CIRCUIT LANGUAGES	SPOKEN LANGUAGES
South Africa	.82	.87	English	English	English .3 Hindi .05
Spain	.97	.44	Spanish	Spanish	Spanish .97
Sri Lanka	.90	.31	-	-	Tamil .16
St Helena	.97	0	-	English	English 1
St Kitts & Nevis	.97	0	English	English	English 1
St Lucia	.54	.67	English	English, French	English 1 French .85
St Vincent & Grenadines	.96	0	English	English	English 1
St Pierre & Miquelon	.99	.07	French	French	French 1
Sudan	.46	.56	Arabic	Arabic	Arabic .51
Suriname	.93	.79	Dutch	Dutch, Hindi	Dutch .93 Hindi .38 Javanese .15
Sweden	.99	.37	Swedish	Swedish	Swedish .99
Switzerland	.99	.53	German, French	German, French	German .72 French .33 Italian .07
Syria	.71	.50	Arabic	Arabic	Arabic .8
Taiwan	.91	.49	Chinese	Chinese	Chinese .91
Tanzania	.68	.95	English	English, Arabic	Swahili .93 English .05
Thailand	.94	.75	-	-	Malay .05
Togo	.52	.89	French	French	French .52
Tonga	.98	.01	-	English	English .98
Trinidad & Tobago	.98	.47	English	English	English .98
Tunisia	.67	.01	Arabic	Arabic, French	Arabic .98
Turkey	.82	.25	-	Turkish	Turkish .9
Turks & Caicos Islands	.98	0	-	English	English 1
Tuvalu	.96	.17	English	English	-
U.S.S.R.	.98	.40	-	-	-
Uganda	.62	.93	English	English	English .62
United Arab Emirates	.79	.78	Arabic	Arabic	Arabic .89
United Kingdom	.99	.07	English	English	English .99
United States	.97	.35	English	English	English .97 Spanish .09
Uruguay	.97	.09	Spanish	Spanish	Spanish 1
Venezuela	.91	.02	Spanish	Spanish	Spanish .93
Vietnam	.94	.20	-	-	-
Virgin Islands (U.S.)	.92	.34	English	English	English 1
Yemen	.38	.56	-	Arabic	Arabic .95
Yugoslavia	.91	.32	-	-	Albanian .16
Zambia	.78	.9	English	English	English .85
Zimbabwe	.85	.56	English	English	English .62

*Languages (Frankel-Rose) is from Rose's database for Frankel and Rose (2002). Otherwise, all the data in this table is extracted from Grimes (2000) with ancillary use of the *CIA Country Factbook* (except in the one case of language diversity for Hong Kong, which is drawn from Taylor and Hudson (1972)). Literacy is from the *CIA Factbook* (with a few blanks filled in from Grimes). Language diversity is from Grimes. A zero for language diversity may mean that no calculation was made on the assumption that the number would be small (source: private correspondence). The percentage figures for the spoken languages are almost exclusively derived from Grimes. But some inferences depend on the literacy rates, in which case the *CIA Factbook* enters as well. Official languages come from Grimes (2000) (except for a few isolated entries drawn from the *CIA Factbook*). An Open-Circuit Language is either official or has at least 20% speakers, and the maximum number of Open-circuit Languages is two. A Spoken Language is spoken by at least 4%. Some major national languages are omitted because they are neither accepted as official nor spoken by as many as 4% of the population in any trading country outside of the home one. Other languages do not appear because of missing trade data. This is notoriously true in regard to the languages in the ex-Soviet Union, including Russian. Fulfulde is also sometimes referred to as Fula or Fulani.