LANGUAGE AND FOREIGN TRADE

Jacques Melitz

University of Strathclyde, CREST-INSEE, and CEPR

Earlier draft: CEPR Working Paper no. 3590
September 2002

This draft: July 2006

Abstract: While language plays a big role in gravity models, there has been little attention to the channels through which a common language promotes bilateral trade. This work proposes separate series for a common language when ease of communication in trade depends heavily on translation and where it depends strictly on the ability to communicate directly. The series related to direct communication is far more important in explaining bilateral trade, but the other series, where translation is essential, makes a distinct contribution as well. Either measure of a common language outperforms the measure in popular use, which is basically related to translation, and a linear combination of the two does far better. In addition, the paper examines the effect of literacy and linguistic diversity at home. Both of these country-specific linguistic influences promote foreign relative to domestic trade. Finally, the article studies the separate importance of English, the substitution effects of a common language and the separate role of network externalities.

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

Jacques Melitz
Department of Economics, William Duncan Building
130 Rottenrow
Glasgow G4 0LN, U.K.
Tel.: 44 141 548 3581
Fax: 44 141 552 5589
E-mail: jacques.melitz@strath.ac.uk

Many thanks for valuable comments to Francis Kramarz and Éric Maurin at CREST-INSEE in Paris, and, most of all, to Jeffrey Bergstrand and Marc Melitz.

Gravity models provide ample evidence that a common language has a significant impact on bilateral trade but studies of these models leave open the question of the channels through which the effect takes place. (For bibliography and examples, see Frankel (1997), Frankel and Rose (2002), and Anderson and van Wincoop (2004).) Based on the usual binary measure of a common language as zero or one, direct communication cannot be the answer. The usual measure is too little sensitive to the actual percentage of people in a country pair who can communicate directly and often considers a common language as present when only a minority in either country in a pair could communicate directly with the people in the other. For example, India and Tanzania are supposed to be an English pair while Niger and Senegal are a French one. But if we consider the probability that a random pair of people from India and Tanzania both speak English, or a similar random pair from Niger and Senegal both speak French, the figure is less than 10 percent in both cases. Implicitly the actual numbers who can communicate directly does not matter. An adequate system of translation will do as well. In this paper, I propose to construct separate series for a common language that puts the emphasis on translation, or along existing lines, and a different series that depends strictly on direct communication. This will then provide information about the channels through which ease of communication through language promotes bilateral trade. Better measures of a common language will result as well.

The usual reliance on binary variables as indicators of a common language largely reflects the difficulty of quantifying the numbers of speakers of different languages in a country. Yet considerable headway is possible by relying on Grimes (2000), now in its 14th edition since first appearing in 1951. This work is the result of a massive effort to condense the information supplied by the entire profession of ethnologists about world languages. There have indeed been at least three recent efforts to use this work to construct a general quantitative index of language in economic research: Hall and Jones (1999), Wagner (2000) and Rauch and Trindade (2002). But the aim has never been to provide a general index of the ability to communicate directly in foreign trade. Hall and Jones focus on language out of a concern with certain institutional/legal features. In close connection, they limit their attention to a few major languages. Though concerned with communication, Wagner deals strictly with the trade of

Canadian provinces, and chooses his languages accordingly. Rauch and Trindade focus on ethnic ties. Consequently, they collect data strictly on common *native* languages, whereas, of course, bilingualism is of the essence in regard to communication. There has been one earlier effort to construct a general quantitative index of a common language in order to analyze world trade from a similar perspective as mine, by Boisso and Ferrantino (1997). But the authors proceed, in this pioneering work, like Rauch and Trindade, to attribute only a single language to each person (not necessarily the mother tongue). They also rely on a far more summary treatment than Grimes', by Katzner (1986).

A few other important questions about language and trade remain unanswered. One is whether the world's dominant language, English, is more effective than the rest in promoting trade. Another is the impact of linguistic diversity at home and literacy on foreign trade. In principle, both of these country-specific aspects of language should promote foreign trade. Besides these questions, I will devote an appendix to the substitution effects of a common language and the network externalities of a common language.

The basic results can be summarized as follows. First, a common language promotes international trade both through direct communication and via translation. Direct communication appears to be far more important. But an established network facilitating access plays a basic role too. Next, despite the dominant position of English as a world language, English is no more effective in promoting trade than other major European languages. On the other hand, the major European languages as a group (including English) are more efficient than other languages in promoting trade. Further, both literacy and a diversity of tongues at home do indeed boost foreign trade, in accordance with theory. Finally, I find no evidence of substitution effects of language on third countries or of network externalities of language. However, the search for this evidence yields one important result: a scarcity of advantages of a common language intensifies the exploitation of those advantages of a common language that do exist.

The discussion will begin by developing the gravity equation that serves 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 depending on translation. Section III will present the elementary econometric results. The issue of the separate importance of different

world languages, including English, will come up at this stage. Section IV will then treat the significance of literacy and linguistic diversity at home. Since these two influences are country-specific, their treatment will require some adaptation of the earlier econometric analysis. The final section will contain some general discussion and suggestions for future research. An appendix will then explore the issues of substitution effects and network externalities. All the raw linguistic series, including those I constructed, follow in a separate data appendix.

I. The gravity model

The gravity model is particularly fitting in studying the relation of language to trade since it focuses on the general barriers to trade. 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, if we accept the usual assumptions of gravity models that all countries specialize in the production of separate goods or separate varieties, and we assume that utility functions are identical, homothetic, and CES everywhere, balanced bilateral trade results. There is then nothing except aggregate trade to investigate. Anderson and van Wincoop (2003) show that under the preceding assumptions, a particularly simple form of the gravity equation follows. It is:

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

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 $1+x_{ij}$ where x_{ij} stands for the percentage of the costs attributable to *foreign* trade frictions in relation to the export price p (regardless whether this price (fob) is p_i or p_j), and P_i and P_j are the respective Dixit-Stiglitz price levels (based on utility maximiza-

tion) in the two countries. In the case of P_j:

$$(2)\ P_i = \left[\sum_j \left(\eta_j p_i t_{ij} \right)^{1-\sigma} \right]^{1/(1-\sigma)}$$

where the summation sign embraces all i prices inclusive of p_j (in which case, exceptionally, $t_{jj}=1$) and η_j is the distribution parameter of the utility function (for good j or the varieties coming from country j). The corresponding equation holds for P_j. Evidently, σ must be greater than one, as empirical work tells us is predominantly the case, if there is to be a negative effect of t_{ij} on trade. Since language plainly affects trade partly by modifying costs of distribution, it may be noted that distribution costs can be easily incorporated in the preceding model (see Baier and Bergstrand (2001)).[1]

An increasingly popular way of treating the theoretical price indices, P_i and P_j is to introduce a separate fixed effect for each individual country. This procedure has the obvious merit of simplicity and comprehensiveness. By encompassing unobserved variables in the analysis, the method also controls for the bias induced by the possibility that the unobserved influences are correlated with the bilateral trade frictions t_{ij}. But the procedure suffers from the important disadvantage at present of making the country-specific factors difficult to analyze separately. This matters, as two of the major linguistic influences in the study, literacy and linguistic diversity, are country-specific. My response will be to deal with the country-specific influences at a later stage by examining their impact on the country fixed effects themselves.

The general study procedure will be as follows. Let the t_{ij} term in equation (1) be:

$$(3)\ t_{ij} = \prod_{k=1}^{m} (u_{ij,k})^{\gamma_k} \times \exp\left(\sum_{k=m+1}^{n} \gamma_k u_{ij,k} \right)$$

[1] Baier and Bergstrand (2001) point out that injecting distribution costs 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 relevant elasticity of transformation into sales (see Baier and Bergstrand, sec. 2.3.4). From the standpoint of estimation, however, these differences are minor.

where $[u_{ij,k}]_{k=1\ldots n}$ represents a vector of n symmetric bilateral trade frictions between countries i and j, and $[\gamma_k]_{k=1\ldots n}$ represents the vector of the associated elasticities or semi-elasticities. For $k \leq m$, the u_{ij} bilateral variables (for example, distance) are continuous (and the γ_k terms elasticities), while for $k > m$, they are either 0-1 dummies (like membership in a currency union) or percentage values (and the γ_k terms semi-elasticities). Since the data in the study consist of a panel with a time dimension, t subscripts will serve below to signify variables that move over time. But I leave out the language variables for later treatment. In the first stage of the analysis, the estimated form of the gravity model will then be:

$$(4)\ \log T_{ijt} = \alpha_o + \alpha_t + \delta_c Z_c + \alpha_y \log (Y_{it}Y_{jt}) + \beta_1 \log \text{distance}_{ij} + \beta_2 \text{adjacency}_{ij} + \beta_3 \text{currency-union}_{ijt} + \beta_4 \text{common-country}_{ij} + \beta_5 \text{free-trade-area}_{ijt} + \beta_6 \text{Excol}_{ij} + \beta_7 \text{Comcol}_{ij} + \beta_8 \text{currency-union/outsider}_{ijt} + \beta_9 \text{common-country/outsider}_{ij} + \beta_{10} \text{free-trade-area/outsider}_{ijt} + \beta_{11} \text{Col-relationship/outsider}_{ij} + \varepsilon_{ijt}$$

α_o in this listing is a constant. α_t is a time fixed effect for all periods except one (and embraces the varying levels of world income Y_t^W over time). $\delta_c Z_c$, in turn, relates to the country fixed effects (which concern $P_i P_j$). δ_c represents these effects themselves while Z_c is a vector of indicator variables (one per country), where $Z_c = 1$ if c = i or j and is 0 otherwise. The coefficient α_y of the log $(Y_{it}Y_{jt})$ term (which can only be estimated in the light of the variation of $Y_{it}Y_{jt}$ over time) should evidently equal one.

The β_k coefficients, $k = 1 \ldots 11$, $\beta_k = (1-\sigma)\,\gamma_k$, represent product terms relating to the overall elasticities of trade in the case of the continuous variables and the overall semi-elasticities in the case of the rest, which are either dummies or percentage terms. The variables associated with β_3 through β_7 relate to indices of political association that Frankel and Rose (2002) have successfully used before. Of these variables, the only ones whose meanings are not immediately apparent are Excol, which refers to a relation between an ex-colonizer and an ex-colony, and Comcol, which refers to a relation between two former colonies of the same colonizer. It is important to admit ex-colonial relationships and an ex-common colonizer, since former colonial attachments have strong linguistic consequences, and if colonial variables were left out of the analysis, any significance of language could be the result of 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 are available only for a much narrower sample of countries. Time subscripts are present for the dummy variables for currency union and free trade agreement because of entries and exits into both arrangements. The coefficients β_8-β_{11}, in turn, relate to third-country influences of the political variables. The associated variables are dummies with a value of one in bilateral trades between one member of a relevant political association and a non-member. In the case of an ex-colonial relationship, a single dummy serves to cover bilateral trade between a country that is either an ex-colonial power or an ex-colony and a country that is neither. (Any effort to distinguish the third-country effects of an ex-colonial relationship and an ex-common colonizer would be futile.) Some of these dummies have served before, though only one at a time, depending on the focus of interest (see Frankel (1997) and Rose (2000)). The term ε_{ijt} is white noise associated with the dependent variable, bilateral trade.

Following the estimation of equation (4) in the first stage of the investigation, the focus will shift on the factors determining the country fixed effects δ_c. Literacy and linguistic diversity will then enter, along with a number of other country-specific variables: namely, average output, average population, land area, landlocked, membership in ex-Soviet Union. Y_iY_j was already mentioned as present in equation (4), but it can only reflect variations in output over time there. However, regardless of such variations, countries with larger output trade more with everybody, including foreigners. Hence, average output should enter separately as an influence on the country fixed effects. Average population should similarly enter but for totally different reasons. The more people there are at home, the wider the opportunities to trade domestically and therefore without bearing the costs of foreign trade. On this ground, the variable should have a negative impact on bilateral trade (compare Frankel and Romer (1999)). True, population might have been included earlier in equation (4) along with output since it varies over time just as output does. However, population is much more sluggish than output and does not emerge as significant in my data set when country fixed effects are present. Countries that cover large land surfaces face a comparable situation to highly populous ones in some respects. Because of large internal distances (compare Nitsch (2000)),

the people in those countries find foreigners to be further away than those who live in small countries (all the more so if we control for remoteness). Correspondingly, large countries tend to be less open. In regard to the dummy for the members of the ex-Soviet Union, trade outside the Soviet bloc was notoriously small in this political system.

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.

A sensible question to raise, therefore, is how there can possibly be any language barrier when translation can link up all world languages to one another. To this query, there are two important answers, but their relative significance is important and unknown. The first answer regards the costs of translation – both the social overhead costs of preparing the ground for the wide distribution of translation services to large groups of people and the costs of dealing out such services to them individually. Even if true, in line with the previous paragraph, that once an appropriate foundation has been laid for widespread translation, the services can be meted out to people individually at a negligible cost, the social overhead costs can still be very important. Indeed, the usual treatment of a common language in foreign trade

says that they are since this treatment allows for only two common languages at most. The treatment therefore implicitly says that 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 can exist between any pair of countries. The second answer to the question how there can be a language problem when it is possible to resort to translation is that direct communication is important. We have some clear indication of the importance of direct communication from results of gravity models showing that immigrants promote trade with their country of origin (see Gould (1994), Head and Ries (1998), and Dunlevy and Hutchinson (1999), Wagner, Head and Ries (2002), and Rauch and Trindade (2001)). One likely reason for this impact of immigrants is their ability to speak their native language (as well as their capacity to translate this language into the primary one(s) in the host country). Accordingly, I will develop a separate measure of a common language pertaining to each of these two arguments individually. One will relate to a communication network and the other to direct communication. In doing so, I will use only one source in addition to Grimes (2000) – the *CIA country factbook*, a frequent reference among economists. The *factbook* will serve either to cover gaps in Grimes about official languages, or else in the case of literacy rates, as the primary source. The reason for switching to the *factbook* as the main source for literacy is that Grimes pays too little attention to the uniformity of dating this variable. I decided to stick to Grimes and the *CIA factbook*, following a broader search, mainly to facilitate reproduction.

The first measure, titled "open-circuit communication," will require no set 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.[2] 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 that have the widest international currency. This will essentially mean sticking to Arabic,

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

English and French in some African examples where Swahili, Hausa or Fulfulde could have served instead. Open-circuit Communication has a value of one if the required condition is met (in any combination of the two alternatives for the two countries) and zero otherwise. It cannot be overemphasized that this measure is the result of an effort to make sense of the usual indices of a common language. 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 who can communicate directly. Fortunately, those percentages needed to be calculated only when they were large enough to make any statistical difference 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 157 "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 different dialects, creole and pidgin versions of a language as equivalent. I also deliberately cumulated speakers of a language without regard to any second languages they may possess.[3] 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* served significantly, since as indicated earlier, I relied primarily on this source for the literacy rates and only used Grimes to fill in missing values. 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,

[3] Hutchinson (2002) tests the difference between English as a first language and as a second language (based on data in Crystal (1997)) using a gravity specification. But he does so exclusively with regard to bilateral trade with the United States. Paradoxically, English as a second language emerges, if anything, as more significant (which tends to support my emphasis on language as a tool of communication). However, judging from the results, the distinction between English as a first or a second language would not pass any statistical test of significance.

which was equivalent to a general normalization.

As regards the country-specific linguistic variables that enter the study, the latest edition of Grimes (2000) provides an index of linguistic diversity for the first time. This index, covering all the countries in the study, 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 in a country 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 may serve as a reflection of my preferred one.[4]

The underlying hypotheses about the signs of the influences of the language variables in the study are fairly obvious. Different languages are impediments to communication, therefore trade. In addition, there exists more uniformity of language within a country than between a country and the rest of the world. Thus, linguistic obstacles must interfere more with foreign than domestic trade. It follows that higher levels of Open-circuit and Direct Communication should both raise foreign trade. Linguistic diversity will also favor foreign trade. When such diversity is considerable at home, domestic trade will be less effective in avoiding problems of communication. For example, if we compare South Africa, a country with many large language groups, with Venezuela, a country with only one, linguistic barriers favor domestic trade more in Venezuela than South Africa. Of course, South Africans consequently

[4] A similar index, drawn from Taylor and Hudson (1972) and resting entirely on a detailed Soviet linguistic study dating to 1964, with the sociological title "ethnolinguistic fractionalization," pertains to the identical issue as Grimes': whether two randomly chosen individuals in a country will share the same maternal language. There have been a number of recent uses of the Taylor-Hudson index in economics: Mauro (1995), Alesina and Wacziag (1998), Easterly and Levine (1997) and La Porta et al. (1999). All of these authors employ the index to treat issues of societal division and government behavior, whereas I shall use the Grimes index strictly in connection with communication. The broad sociological and political usage of the Taylor-Hudson index by these authors has also aroused criticism from Collier (2001). For some detailed discussion of this index and other closely related indices, see Easterly and Levine (1997). In regard to the 108 "countries" that both Grimes and Taylor-Hudson score for linguistic diversity, the correlation coefficient between the two indices is high: .85.

have more incentive to acquire multiple languages. But this merely reinforces the basic point that linguistic diversity reduces the tendency of the language factor to work in favor of home relative to foreign trade. In sum, linguistic diversity at home should promote foreign trade. Finally, linguistic obstacles must be more severe for people who cannot read and write. Since alternative forms of communication – through speech and action – are easier at home, illiterates should be especially handicapped in foreign trade. Thus, literacy should promote foreign relative to home trade.

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 Communi-cation, the cited languages are strictly "source" languages. The "destination" languages – those in which the messages are received – constitute a considerable number, probably over a hundred; the number 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. However, 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 their strict importance in domestic trade. Japanese is an example. They could also be missing because of lack of trade data. Some striking examples come from the ex-Soviet Union and 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), whose database I use for the other variables besides language. 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 correlation matrix also covers the indices of literacy and language diversity in the study. As can be seen, the correlation between Open-circuit Communication and Frankel and Rose's (FR) Common Language variable is only .81. This is rather low if we consider that the two variables are basically meant to signify exactly 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 (Farsi), 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.). As also noteworthy, the correlation coefficient between my two language indices, Open-circuit and Direct Communication, is 0.73. While this ratio is high, we must remember that both indices are zero in the majority of cases, which raises the correlation. If we limit ourselves to the 23 percent of the test observations where one or the other of the two indices is positive, the correlation between them is only .19. From this other and relevant perspective, the difference between the two indices is very wide.[5]

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

III. Basic results

Virtually all of the data for the variables besides language in the study come from Frankel and Rose (2002) and are described in their data appendix. I owe a considerable 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 differences 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. 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

[5] In the cases where at least one of the three language indices in question is positive, the FR index is more highly correlated with direct communication than open-circuit communication (see table 2). The reason is that in the frequent cases where the FR index is zero and the index of open-circuit communication is one, the index of direct communication is mostly somewhere in between.

and free trade area are mutually exclusive.

The first column of Table 3 shows the result of the test of equation (4) 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 FR, 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. The Student *t*s that are shown are based on robust standard errors (after correction for clustering of data for individual trading pairs). However, the estimates of the time and country fixed effects are omitted. The results notably confirm the hypothesis of a unitary elasticity of bilateral trade with respect to output. Thus, the treatment of output as exogenous is inessential, as I have confirmed. When we impose unitary elasticity with respect to output (more exactly, $Y_{it}Y_{jt}$), the results are unaffected (this is true throughout). The coefficients and significance of distance and adjacency are also of the usual order. So are the coefficients and the significance of the five political variables, as a comparison with FR (2002) will show (though there are some differences concerning Common Currency and Ex-common Colonizer). Free Trade Area has a positive effect on third-country trade, Political Union only marginally so, and Common ex-Colonizer has a negative effect on third-party trade. These results for the political variables will vary only mildly below. The rest of the discussion focuses on language.

First and foremost, Direct Communication is extremely important with a coefficient around one while Open-circuit Communication is not so. Is the usual implicit emphasis on indirect communication via translation therefore mistaken? Further probing shows otherwise. As seen in column 2, once we eliminate the non-European languages (Arabic, Chinese, Hindi, Malay, Persian and Turkish) from Open-circuit Communication and we limit ourselves to the rest, Open-circuit Communication is highly significant with a Student *t* close to 4. Thus, at least following adjustment, Open-circuit Communication emerges as very important even when Direct Communication is present. Facility of translation and direct intercourse are two separate influences on foreign trade. Notwithstanding, Direct Communication is distinctly the more effective of the two.

Column 3 focuses on the merits of these two linguistic variables as opposed to Frankel

and Rose's. In this column, I inject the FR index of a common language into the equation. Its presence hardly even alters the estimates and Student *t*s of Direct Communication and Open-circuit Communication in European languages while it enters (insignificantly) with the wrong sign. In addition, I experimented with the FR index as an alternative to open-circuit communication (with or without non-European languages) together with Direct Communication. In this case, the index has a coefficient close to zero. Thus, my measure succeeds in displaying the separate significance of translation, as distinct from Direct Communication, whereas the FR index does not.

Such is the supremacy of English as a world language today that there is also cause to wonder whether this language bears a greater impact than the rest. Is English more effective in promoting trade than the other languages? It is difficult to see how English could 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 and independently of any externalities. The question then is whether English exerts a greater influence as a source language on trade through indirect communication (via translation) than the other European open-circuit languages do.

To begin, the inclusion of country fixed effects in the study has considerable significance in answering this question. Suppose the country fixed effects were absent and all European languages were equally effective in open-circuit communication. Then the gravity model would say that English has a *smaller* effect on bilateral trade than the other European languages. The point is simple: two English-speaking (or English-connected) countries have many more alternatives to trading with one another than, say, two Dutch-speaking (or Dutch-connected) ones have. Linguistic facility therefore is a weaker inducement to trade between any *specific* English pair than any *specific* Dutch one. In terms of the model, if English and Dutch are equally effective in communication, English reduces t_{ij}/P_iP_j (in equation (1)) far less than Dutch in Open-circuit Communication. But, in principle, the use of country fixed effects in the estimate should repair this problem by taking separate account of P_iP_j. Thus, in the presence of these effects, the greater effectiveness of English in promoting trade should show

up in a higher coefficient for English in Open-circuit Communication than for the other European languages.

The fourth column in Table 3 contains the relevant test. Following the introduction of Open-circuit Communication in English, the estimates for Direct Communication and Open-circuit Communication in the European languages stay essentially the same while Open-circuit Communication in English appears as insignificant and even with the wrong (a negative) sign. Therefore, the hypothesis fails. There is no evidence that English possesses any special advantage over the other European tongues in open-circuit communication. To all indication, the role of English is already adequately captured by the earlier two linguistic indices, where the language looms large.[6] [7]

As regards the relative influence of direct and indirect communication, column 2 would indicate that the impact of Direct Communication is about three times larger than that of Open-Circuit Communication in European languages. Of course, if we drop Direct Communication from the estimate and only keep Open-circuit Communication in European languages, the latter's importance climbs. As seen in column 5, the coefficient of Open-circuit Communication in European languages then rises to .63, or three-quarters of the level of Direct Communication in column 2 and its significance also shoots up. But this simply reflects the fact that the index no longer stands for indirect communication alone but for direct communication as well. The correlation between Direct Communication and Open-circuit Com-

[6] Helliwell (1999) gets seemingly conflicting results. He finds English to be far more important than German, French or Spanish in a related study treating a common language as a binary variable (explicitly based on status as an official language). But the conflict could come from the fact that he does not use Direct Communication (in which English plays a prominent role) as a control variable, whereas I do. In other words, as opposed to him, I test only for the separate significance of English as an Open-circuit Communication language. In addition, Helliwell's sample covers a much narrower group of countries, only 33, 22 of them from the OECD. As he notes himself, in a similar study to his using more countries, 63, Frankel, Stein and Wei (1998) report no difference in the significance of English as opposed to Spanish, French or German. Like other earlier students of language, Helliwell also does not introduce country fixed effects (which should have worked against English, as I have explained).

[7] I experimented too with the hypothesis that English, Spanish, and French – or the three most important languages in the study – are the only languages providing an effective circuit of communication. This next conjecture is impossible to reject. However, the hypothesis that all of the European languages serve in Open-circuit Communication stands up as well, and I will stick to this last hypothesis.

munication in European languages is .67 (somewhat below the correlation between Direct Communication and Open-circuit Communication in all language, which is .73).

Column 6 adds important information. It shows what happens when we use a straight-line average of Direct Communication and Open-circuit Communication in European languages. The coefficient of the combined variable exceeds that of Direct Communication alone (the one in column 2). This corroborates the earlier conclusion that translation makes a distinct and separate contribution. The last column of table 3 provides an important additional test based on the combined index in the previous column, or Common Language, as I will now refer to it. There I introduce separate country fixed effects for different years and correspondingly drop the preceding year and country fixed effects (as well as $Y_{it}Y_{jt}$). (At this point, the $\delta_c Z_c$ terms of equation (4) are then indexed by t while α_t and α_y drop out.) Following, the size and statistical significance of the coefficient of Common Language stays the same. (This is equally true if the dependent variable is the ratio of trade to output.) I have also experimented with separate year-by-year estimates of the last equation (not shown). The impact of Common Language shows no trend while the annual coefficients are all highly significant.

Column 6 displays a coefficient of Common Language of .97, which implies an elasticity of influence of 1.64 (exp(.97) = 2.64 – 1). Of course, that influence largely reflects the elasticity of substitution between goods σ (since $\beta_{comlang} = (1-\sigma)\gamma_{comlang}$ given equations (1) and (3)). Usual estimates of this next elasticity go from 6 to 10 (based on Obstfeld and Rogoff (2000) and Anderson and van Wincoop (2004)). On this basis, the elasticity of influence of a common language, per se ($\gamma_{comlang}$), on bilateral trade is between .35 and .18 (1.64 divided by 6–1 or 10–1). These are larger numbers than the .05-.12 figures that Anderson and van Wincoop (2004) retain in their recent survey of trade costs (based on Eaton and Kortum (2002) and Hummels (1999)). But their numbers are somewhat on the low side as compared with the general literature (see, for example, FR (2002) or Frankel (1997)). In addition, those numbers rest strictly on dummy variables, while, as I have shown, such reliance leads to underestimates of the impact of a common language. As a final observation, my proposed .18 to .35 range for the influence of a common language is of the general order of the figures in Table 3 for the impacts of the political aids to trade (which are quite ordinary in terms of the general

literature).

IV. Literacy and linguistic diversity

The impacts of literacy and linguistic diversity at home lie buried in the country fixed effects in the previous estimates. The first column of Table 4 shows what happens when we try to bring these variables into the open by merely sacrificing the country fixed effects. (The year dummies remain.) In this case, the other country-specific influences that were mentioned in the theoretical discussion – average population size, land area and landlocked – can also be added. All of the new variables emerge as highly significant with the right signs, while the earlier variables remain so as well (but the coefficient of output goes up suspiciously high). Note, in particular, that if anything, the influence of Common Language emerges *less* strongly than before. However, this new estimate ignores many nationally specific influences of P_iP_j, which need not be orthogonal to the explanatory variables. Hence, the estimate may be biased. Since Anderson and van Wincoop's (2003) persuasive demonstration of the pitfalls of brushing over multilateral trade resistance (P_iP_j), many would now even argue that the estimate has little to do with the gravity model.

As indicated earlier, I propose therefore to test the impact of the preceding country-specific variables on the country fixed effects themselves, as they were estimated before. The focus will be on the estimate in column 6 or the next-to-last column of Table 3 where Common Language is present with constant country fixed effects over time, or the next-to-last one. But focus on column 2 instead, where Open-circuit Communication and Direct Communication appear as separate variables, would make no difference. Nor would it matter if we used the estimate that followed the setting of the elasticity of output equal to one. There are 157 observations. The explanatory variables are the aforementioned country-specific influences.[8] Since average output and population are highly positively correlated, I present results with average output per head and population rather than aggregate output and population. This will

[8] I experimented with remoteness too, taken either as a straight-line (unweighted) average or an output-weighted average of distance from all other countries. Remoteness turns out to be totally insignificant. Thus, I ignore the variable here.

permit a better assessment of the negative influence of population.[9]

The result is in columns 2 and 3. The difference between the two is that column 3 drops 13 political units accounting for exceptionally few observations in the earlier estimates. These 13 units are also all extremely small (both in terms of population and output) and consist predominantly of dependencies with little or no independent political status. The 13, as well as all the remaining 144, are listed in Table 5. Examination shows that the basic difference in column 3 where the 13 are out is a rise in the coefficient for output per head and a fall in the one for literacy. I favor column 3.

As can be seen, all 7 explanatory variables enter significantly at the 5% confidence level with the right sign. Of special interest, output per head shows up with a coefficient of .26 and statistically significantly less than one (in column 3). Thus, while the hypothesis of a unitary output-elasticity of trade works well in regard to changes over time, it does not do so at all in comparing different countries. (This result is equally true if we use aggregate output instead of output per head.)

But attention properly centers on literacy and linguistic diversity. Of the two, literacy shows up with the larger coefficient, though the same significance level. However, since the two confidence intervals are wide, there is only a .61 probability that the effect of literacy exceeds that of linguistic diversity at all (based on an F test). Quite significantly, the joint presence of output per head and population means that we can interpret the influence of literacy in the estimate as resulting from the ability to communicate. As is well known from the growth literature (Barro (1991)), literacy has a marked impact on per capita output. Therefore, if per capita output were absent, the effect of literacy would reflect largely a positive impact of productivity rather than anything to do with the ability to trade. I have indeed verified that if this variable is omitted, both the coefficient and the Student *t* of literacy shoot up even if population remains. Of further interest, the coefficient of literacy greatly exceeds that of per

[9] The estimates with output and population instead of output per head and population are perfectly identical *in all respects* except that when per capita output serves, the size and *t* value of the negative coefficient of population are both much smaller.

capita output (even in column 3 as opposed to 2).[10]

The fourth and last column of Table 5 shows the result of dropping the index of membership in the ex-Soviet Union. Following, the significance of literacy falls. It only holds up at the 10% confidence level. But this is easy to interpret. The members of the ex-Soviet Union displayed both exceptionally high literacy rates and exceptionally low levels of foreign trade. Failure to control for this particularity of the political arrangement, which does not generalize at all, obscures the influence of literacy on trade. It is interesting also to compare the estimates in column 1, where the country fixed effects are omitted, with the rest. The results in this column resemble the others as concerns population, land area, and landlocked. However, literacy is less influential and linguistic diversity is more so than in the other columns.

I conclude that literacy and linguistic diversity both exert positive effects on foreign trade, in accordance with theory.

VI. Discussion and conclusion

We knew beforehand that a common language promotes trade. This study sheds light on the channels through which language exerts this influence and introduces a number of effects of language that had gone unnoticed before, or at least had not been studied. Apart from direct communication, translation is important in surmounting linguistic obstacles. Though few people may speak Portuguese in a country, they may be "tuned" to the language all the same. Thus, the impact of a common language on foreign trade can issue from translation as well as the ability to communicate directly. Notwithstanding, direct communication appears about three times more effective than indirect communication in promoting trade. Two general linguistic influences on foreign trade also emerge in the study. If people can read and write in any language, they can cope better with the problems posed by foreign languages in general. In addition, if they face linguistic obstacles at home, they trade more with foreigners. Accordingly, based on identical scores for Common Language (notably including zero), the pull of linguistic factors in favor of domestic trade is uneven and is lower for country pairs sharing high literacy rates or sharing high linguistic diversity or both. Further, English seems

[10] An F test (in column 3) shows that there is a .8 probability that the impact of literacy on trade exceeds that of output.

to offer no particular advantage in foreign trade, while the European languages as a whole do. These languages emerge as better instruments of Open-channel Communication than other ones. In addition, as shown in an appendix, linguistic links between countries boost bilateral trade without damaging trade with third countries at all. As shown in the same appendix, there is also no evidence that network externalities of language contribute to trade between countries. The importance of those externalities has thus been exaggerated. The reason may have to do with the fact that a lot of market information gets transmitted without a common language (through limited vocabularies, small numbers of fluent speakers, dictionaries, and means of non-verbal communication). The separate significance of literacy and linguistic diversity has a similar interpretation: communication takes place without a common language.[11]

The study suggests a number of avenues of further research. First, it would be interesting to see if the relative significance of Open-circuit and Direct Communication differs for trade in different sorts of goods. As a possibility, rudimentary communication might suffice for trade in homogeneous goods while trade in heterogeneous goods might require more sophisticated intercourse. If so, open-circuit Communication might be particularly important in trade in homogenous goods while Direct Communication is unusually so in trade in heterogeneous goods. Rauch (1999) suggests a hypothesis of this sort. He argues that close personal relations between exporters and importers are much more significant in trade in heterogeneous goods than perfectly homogeneous ones. Because of issues of information, private "networks" are necessary in one case whereas "markets" suffice in the other. The proposed hypothesis about the separate roles of Open-Circuit Communication and Direct Communication in different markets follows easily.[12] [13]

---

[11] The focus on trade is important. A common language may be more important in production than in trade. In fact, to all indication, even varying shades in the ability to communicate *directly* matter a lot in the labor market. Empirical work regularly 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).

[12] In the case of certain homogeneous goods, including crude oil and many primary products, the relevant information could even be so small and easy to get (in part, if not entirely, because of organized exchanges) that there are no language hurdles at all. Translation might resolve the communication problem altogether at no mentionable cost.

[13] Interestingly, Rauch and Trindade (2001) apply the proposed hypothesis to the distinction between a common language and an ethnic network rather than between Open-Circuit and

Another basic extension of the study would be to admit additional variables reflecting culture or ethnicity. 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, in the context of worldwide trade, the only features of the language variables that apply generally regard communication. Other associated features are not always present, as it would be easy to document. This is obvious in the case of Open-circuit Communication where translation is the central mechanism. But even as regards Direct Communication, I have taken steps to assure the primacy of communication by including second-language and non-native speakers. Based on other work in the literature, one missing cultural variable that easily comes to mind is the stock of immigrants. As the analysis here stands, immigrants are implicit since they affect the index of linguistic diversity. But this index also reflects long-standing, sometimes ancestral, linguistic divisions inside national boundaries (as in the case of India or Switzerland, for example). The interest of a separate consideration of immigrants would lie in setting apart the element of ethnic ties to other communities abroad (which affects tastes, skills, trust and information) from strict issues of communication. A separate treatment of immigrants would generally promote the interpretation of the linguistic variables as relating exclusively to communication (compare note 13).

A further extension of the study would be to examine 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 prod-

---

Direct Communication. In their reasoning, a common language works the same way in all trade, whereas an ethnic network promotes trade in differentiated goods as such. The ethnic network in their study is the Chinese diaspora across the world. They also treat a common language as a common *mother tongue* (measured as a continuous variable), which they consider – questionably, I think – as essentially reflecting common tastes. In regard to world trade, even a common *native* language may mostly reflect the ability to communicate. I will return to this issue in the text.

ucts?[14] Evidently, the question cannot be studied within the confines of the current version of the gravity model, which applies strictly to two-way trade, but would require a more intricate version of the model that distinguishes between exports and imports (and is often associated with Bergstrand (1985)).[15]

Finally, it would be interesting also to move beyond the idea that the language variables are constants. 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. In treating the data furnished in Grimes (2000) as roughly contemporaneous and steady, I have followed the practice in the field of trade of viewing all indicators of a common language and linguistic diversity in foreign trade as slow-moving variables that can be regarded as fixed.[16] Grimes (2000) also corroborates this practice by furnishing unique figures for linguistic variables despite wide discrepancies in dating when dates even 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. Doing so would make possible the study of the evolution of all three variables and their influence on trade.

[14] For data and further discussion, see Melitz (2006).

[15] See Gould (1994), Head and Ries (1998), Dunlevy and Hutchinson (1999), Wagner, Head and Ries (2002) and Hutchinson (2002), all of whom distinguish between the influence of immigrant links on imports and exports and all of whom also rely on specifications that are related to Bergstrand.

[16] The same practice does not hold in labor. Labor studies of the impact of linguistic skills on wages make prominent use of time series evidence (see the references in note 11).

APPENDIX 1

Substitution Effects and Network Externalities

This appendix will consider whether the positive impact of a common language on bilateral trade comes at the expense of trade with third-countries. It will also treat the question whether the frequent references to the network externalities of a common language are correct. The answer to both questions will be negative. I find no evidence of either substitution effects or network externalities. But something different emerges from the inquiry: a positive effect of small GDP on bilateral trade between countries with a common language.

(b) *Substitution effects*

A number of studies have looked at the possibility that the positive effects of political variables on bilateral trade between country pairs come at the expense of trade with third countries. The usual approach, which I follow in the text, is to introduce a separate dummy variable to signify trades where one or both of the parties to a trade belong to a relevant political association with some third party but not with one another. However, the method cannot serve in connection with common language – at least, not without modification. There are too few instances of trade between country pairs where neither partner scores positively for Common Language with any third country. As a result, any dummy of the previous sort, if used together with Common Language, would cover too much of the field. In cases where one of the two indices is zero, the other will be nearly always positive, and inversely. It would then prove impossible to distinguish statistically between the influences of the two. But a certain adaptation, assuring a large percentage of cases where both indices are zero, is possible.

Common Language varies between values of 0 and 1 with disproportionately few examples of values between 0.1 and 0.5. Any pair of countries scoring 1 for Open-circuit Communication in a European language must register at least 0.5 for Common Language. As for the rest, those that share only one language in direct communication can only score at most 0.1 for Common Language if the language is European (since they must score less than 0.2 for Direct Communication). Thus, among the rest only those sharing several European languages and obtaining scores of less than 0.2 for each of them or else sharing a non-European language in Direct Communication can possibly score between 0.1 and 0.5. This opens the way

for an index of country pairs with a value below 0.1 for Common Language where one *but not the other* member has a value of 0.1 or over for this variable in trade with some third country. Such an index will focus on trade between country pairs where one member is generally "isolated" linguistically (the one with a value of under 0.1) and the other is predominantly not (the one who displays values 0.1 or over since these values will, in fact, predominantly go up to 0.5 and over in third-party trade). The index will then score as zero the numerous instances of pairs where *both* members have some significant linguistic ties with someone but not with one another, as well as the limited cases where both countries have no significant linguistic ties to anyone. Consequently, there will remain a large proportion of observations in the sample where both this indicator and Common Language will be zero. The point of putting those observations to the side 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. The first two columns of Table 6 show the outcome of the relevant experiment. Neither column shows any sign of a substitution effect of a common language on trade with third parties. On this evidence, the increase in foreign trade issuing from a common language is not at the expense of foreign trade with the rest.

(b) *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 who speak the same language, the higher the probability that random encounter 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 are also 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 language.

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 the intensity of trade, or the ratio of trade to income, 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 (in logs). 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. This first measure amounts to looking for a scale effect of the aggregate income of a trading pair with a common language, with the aggregate conditioned on the language weights.

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 would be that 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. An externality is particularly clear.

Accordingly, I constructed a separate measure of income for the next sort of external effects. In this instance, nine languages drop out (including one of the open-circuit languages, Swedish, which strictly functions between Sweden and Finland). This leaves twenty. For every bilateral pair, I constructed a different total of these twenty world aggregates of income. 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, where the totals rest on a previous calculation based on the data in the language appendix and 1990 real income.[17] 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 two measures of external effects are only negligibly positively correlated in the relevant instances (7%), namely, those where Common Language is positive (as otherwise both indices are zero by definition). This is easily explained. The measure of external effects coming from within trading pairs tends to be low for small country pairs and high for large country pairs (where small and large refer to income). As a result, in the cases where Com-

[17] The previous calculation goes as follows in the example of English. For each country, in turn, start with one or zero 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.

mon Language is positive, this measure tends to be uncorrelated with Common Language, which varies independently of country size (the correlation is zero). On the other hand, in those same cases, the measure of external effects coming from outside the trading pair tends to be high whenever the common language is large in the rest of the world (concerning English in particular). Thus, the correlation with Common language is substantial (it is 0.56). This explains why the two measures of external effects are little correlated with one another in the cases where Common Language is positive.

Columns 3 and 4 of Table 6 show what happens when the two indices of external effects enter one at a time and column 5 shows what happens when the two are included together. When the two enter separately, they both show up with a highly significant *negative* sign (while the coefficient of Common Language rises in compensation and everything else stays the same).[18] When the two enter together, the external effect coming from within the country pair still shows up as significantly negative while the other effect vanishes. Therefore, the former effect is clearly the more important of the two.

These results are plainly at odds with the hypothesis of positive externalities. We are evidently not picking up any positive effect of crowding and size. What we find instead is a positive effect of smallness for pairs with a common language. In the case of column 3, this effect would seem to reflect the fewer opportunities for profitable trade at home in small countries than large ones. Small countries with a common language would then trade more with one another than large countries with a common language. As regards column 4, the positive effect appears to reflect, instead, the tendency for greater trade between country pairs who share a small language than country pairs who share a large one. The latter pairs have more ways to benefit from linguistic ease in trade than by trading together. Both conclusions make sense. Both of them say that a scarcity of advantages of a common language intensifies the exploitation of the advantages of a common language that do exist. But both put in question the idea of a positive effect of size on ratios of trade to output between partners with a

[18] It is easy to understand why the estimate for the product of the GDPs is not affected since the new variables relate strictly to joint effects of common language and a constant output level (for 1990) while the coefficient for the product of the GDPs picks up nothing but an effect of output movement over time.

common language.[19]

[19] In the light of the results, of course, both of the relevant variables could be relabeled without reference to network externalities and the present nomenclature reflects the hypotheses that led to their construction.

REFERENCES CITED

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

Anderson, James and Eric van Wincoop (2003). "Gravity with *gravitas*: A solution to the border problem," *American Economic Review*, 93, 170-192.

Anderson, James and Eric van Wincoop (2004). "Trade costs," *Journal of Economic Literature* 42, 691-751.

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*, 67, 474-481.

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.

Crystal, David (1997). *English as a Global Language*, Cambridge University 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.

Eaton, Jonathan and Samuel Kortum (2002). "Technology, geography and trade," *Econometrica*, 70, 1741-1779.

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.

Frankel, Jeffrey, Ernesto Stein and Shang-Jin Wei, "Continental trading blocs: Are they natural or supernatural?" in Jeffrey Frankel, ed., *The regionalization of the world economy*, The University of Chicago Press, 91-113.

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.

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

Helliwell, John (1998). *How Much Do National Borders Matter?* Washington D.C.: Brookings Institution.

Helliwell, John (1999). "Language and Trade," in A. Breton, ed., *Exploring the Economics of Language*, Canadian Heritage.

Hummels, David (1999). "Have international transportation costs declined?," Working Paper, Purdue University.

Hutchinson, William (2002). "Does ease of communication increase trade? Commonality of language and bilateral trade," *Scottish Journal of Political Economy*, 49, 544-556.

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 (2006). "The impact of English dominance on literature and welfare," Working Paper, University of Strathclyde, Glasgow (earlier version, *CEPR Discussion Paper 2055*, 1999).

Nitsch, Volker (2000). "National borders and international trade: Evidence from the European Union," Canadian Journal of Economics 33, 1091-1105.

Obstfeld, Maurice and Kenneth Rogoff (2000). "The six major puzzles in international macroeconomics: Is there a common cause?" In Ben Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2000*, NBER, pp. 339-390.

Rose, Andrew (2000). "One money, one market: Estimating the effects of common currencies

on trade," *Economic Policy*, pp. 7-46.

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

Rauch, James and Vitor Trindade (2001). "Ethnic Chinese networks in international trade," *Review of Economics and Statistics*, 84, 116-130.

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..

Wagner, Don (2000). "Essays on the mobility of goods and people," Dissertation at the University of British Columbia, Canada.

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

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 (Farsi)	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 (0.19)*	1			
Frankel-Rose	0.81 (0.27)*	0.75 (0.42)*	1		
Literacy	–0.10	0.10	–0.03	1	
Diversity	0.14	–0.08	0.07	–0.43	1

*The parenthetical correlations relate to the 23 percent of the observations in the effective test sample where at least one of the three language indices is positive.

TABLE 3: COMMON LANGUAGE
Regressand: Log of bilateral trade

Log product of real GDP	1.04 (27)	1.04 (27)	1.04 (27)	1.04 (27)	1.04 (27)	1.04 (27)	
Log of distance	–1.24 (–46)	–1.24 (–46)	–1.24 (–46)	–1.23 (–45)	–1.26 (–47)	–1.24 (–46)	–1.25 (–45)
Adjacency (0,1)	.45 (3.54)	.45 (3.53)	.46 (3.56)	.45 (3.5)	.45 (3.52)	.45 (3.5)	.43 (3.32)
Common language: Frankel-Rose (0,1)			**–.14 (–1.54)**				
Open-circuit Communication (0,1)	**.11 (1.48)**						
Open-circuit Communication: European (0,1)		**.28 (3.84)**	**.34 (4.2)**	**.36 (4)**	**.63 (11.2)**		
Open-circuit Communication: English (0,1)				**–.15 (–1.43)**			
Direct Communication	**.99 (7.9)**	**.84 (7.3)**	**.93 (7)**	**.84 (7.3)**			
Common Language						**.97 (13)**	**.98 (13)**
Currency union (0,1)	1.1 (5)	1.08 (4.9)	1.1 (5)	1.06 (4.9)	1.07 (4.8)	1.06 (4.9)	1.19 (5.4)
Political union (0,1)	.81 (1.94)	.83 (2.01)	.79 (1.88)	0.82 (2)	1.16 (2.85)	0.95 (2.32)	0.83 (1.98)
Free trade area (0,1)	.48 (3.49)	.49 (3.53)	.49 (3.49)	.51 (3.65)	.56 (3.88)	.52 (3.69)	.52 (3.52)
Ex-colonial relationship (0,1)	1.69 (13.7)	1.63 (13.4)	1.65 (13.4)	1.63 (13.5)	1.67 (13.2)	1.61 (13.1)	1.62 (13)
Ex-common-colonizer (0,1)	.44 (2.44)	.34 (1.92)	.36 (2.01)	.36 (1.99)	.2 (1.14)	.25 (1.4)	.29 (1.67)
Currency union/ outsider (0,1)	.1 (1.31)	.09 (1.25)	.09 (1.22)	.09 (1.22)	.05 (.68)	.08 (1.05)	.07 (0.89)
Political union/ outsider (0,1)	.17 (1.58)	.17 (1.61)	.18 (1.64)	.18 (1.64)	.21 (1.91)	.19 (1.73)	.16 (1.47)
FTA/outsider (0,1)	.16 (2.73)	.16 (2.72)	.16 (2.69)	.16 (2.77)	.14 (2.39)	.15 (2.61)	.17 (2.79)
Ex-colony/colonizer/outsider (0,1)	–.16 (2.78)	–.18 (–1.98)	–.19 (–1.95)	–.18 (–1.99)	–.21 (–2.37)	–.2 (–2. 23)	–.19 (–2. 09)
R^2	.73	.73	.73	.73	.73	.73	.75
RMSE	1.73	1.72	1.72	1.72	1.73	1.73	1.68

Number of Observations: 31,010. Country-specific fixed effects and year-specific fixed effects (first 6 columns) and country-year fixed effects (last column) are not reported. Student *t*s are in parentheses. These are based on robust standard errors after corrections for clustering of country pairs.

TABLE 4: LITERACY AND LINGUISTIC DIVERSITY

REGRESSAND	Log of bilateral trade (1)	REGRESSAND	Country fixed effects (2)	Country fixed effects (3)	Country fixed effects (4)
Log product of real GDP	1.33 (67)	**Log average real GDP**	.16 (1.97)	.26 (3.3)	.27 (3.4)
Log of distance	−1.04 (–42)				
Log product of population	–.38 (–17)	**Log average population**	–.1 (–2.4)	–.12 (–2.5)	–.13 (–2.8)
Log product of land area	–.16 (–17)	**Log of land area**	–.12 (–3.4)	–.13 (–3.4)	–.12 (–3.3)
Adjacency (0,1)	.72 (5.7)				
Landlocked (0,1,2)	–.29 (–6.8)	**Landlocked (0,1)**	–.41 (–2.2)	–.36 (–2.1)	–.43 (–2.5)
Common Language	**.85 (11.8)**				
Product of linguistic diversity	**.95 (8.4)**	**Linguistic diversity**	**.51 (2.4)**	**.49 (2.4)**	**.55 (2.7)**
Product of literacy rate	**0.48 (4.1)**	**Literacy rate**	**1.04 (2.7)**	**.81 (2.2)**	**.65 (1.75)**
Currency union (0,1)	1.45 (7.68)	**Ex-Soviet Union (0,1)**	–.67 (–2.3)	–.66 (–2.4)	
Political union (0,1)	.99 (2.15)				.
Free trade area (0,1)	1.12 (10.5)				
Ex-colonial relationship (0,1)	1.57 (11.9)				
Ex-common-colonizer (0,1)	.24 (2.5)				
Currency union/ outsider (0,1)	.3 (6.9)				
Political union/outsider (0,1)	.29 (5.5)				
FTA/outsider (0,1)	.3 (6.1)				
Ex-colony/colonizer/outsider (0,1)	–.15 (–3.2)				
R^2	.65	**$\overline{R}^2$**	.48	.51	.49
RMSE	1.94	**RMSE**	.71	.66	.67
No. of observations	31,010	**No. of observations**	157	144	144

Year-specific fixed effects are not reported. Student *t*s are in parentheses. These are based on robust standard errors (after corrections for clustering of country pairs in column 1).

TABLE 5: THE COUNTRY FIXED EFFECTS
(of Table 4)

Algeria	Angola	Argentina	Australia
Austria	Bahamas	Bahrain	Bangladesh
Barbados	Belgium	Belize	Benin
Bhutan	Bolivia	Brazil	Bulgaria
Burkina Faso	Burundi	Cameroon	Canada
Central African Rep.	Chad	Chili	China
Colombia	Comoros	Congo Republic	Congo Demo. Rep.
Costa Rica	Cote d'Ivoire	Cyprus	Czechoslovakia
Denmark	Djibouti	Dominican Republic	Ecuador
Egypt	El Salvador	Ethiopia	Fiji
Finland	France	French So. Ant. Terr.	Gabon
Gambia	Germany, West	Germany, East	Ghana
Greece	Grenada	Guatemala	Guinea
Guinea-Bissau	Guyana	Haiti	Honduras
Hong Kong China	Hungary	Iceland	India
Indonesia	Iran	Iraq	Ireland
Israel	Italy	Jamaica	Japan
Jordan	Kenya	Korea, South	Kuwait
Lao	Lebanon	Liberia	Madagascar
Malawi	Malaysia	Maldives	Mali
Malta	Mauritania	Mauritius	Mexico
Mongolia	Morocco	Mozambique	Myanmar
Nepal	Netherlands	New Zealand	Nicaragua
Niger	Nigeria	Norway	Oman
Pakistan	Panama	Papua New Guinea	Paraguay
Peru	Philippines	Poland	Portugal
Qatar	Reunion	Romania	Rwanda
Saudi Arabia	Senegal	Seychelles	Sierra Leone
Singapore	Solomon Islands	Somalia	South Africa
Spain	Sri Lanka	St Kitts & Nevis	St Lucia
St Vincent & Gren.	Sudan	Suriname	Sweden
Switzerland	Syria	Taiwan	Tanzania
Thailand	Togo	Trinidad & Tobago	Tunisia
Turkey	USSR	Uganda	United Arab Emirates
United Kingdom	United States	Uruguay	Venezuela
Vietnam	Yugoslavia	Zambia	Zimbabwe
Antigua & Barbuda*	Bermuda*	Brit. Virgin Islands*	Cook Islands*
Dominica*	French Guiana*	Greenland*	Guadeloupe*
Kiribati*	Martinique*	Montserrat*	New Caledonia*
St Pierre & Miquelon*			

Number of Observations: 31,010. Even the estimates of column 2 in table 4 exclude 5 cases of political units that account for only a single observation or two in the entire sample (American Samoa, Aruba, Nauru, Samoa and Tonga).

TABLE 6: SUBSTITUTION EFFECTS AND NETWORK EXTERNALITIES
Regressand: Log of bilateral trade

Log product of real GDP	1.06 (28)	1.06 (28)	1.06 (28)	1.06 (28)	1.06 (28)
Log of distance	–1.24 (–46)	–1.25 (–46)	–1.25 (–46)	–1.25 (–45)	–1.25 (–46)
Adjacency (0,1)	.45 (3.5)	.45 (3.49)	.5 (3.89)	.47 (3.67)	.5 (3.9)
Common language	**.96 (12.8)**	**.97 (12.9)**	**1.61 (9.48)**	**1.64 (8.12)**	**1.6 (7.83)**
Network externality from within			**–.028 (–4.27)**		**–.029 (–2.16)**
Network externality from outside				**–.025 (–3.59)**	**0 (0.1)**
Currency union (0,1)	1.07 (4.8)	1.06 (4.8)	.98 (4.4)	1.07 (4.8)	.98 (4.3)
Political union (0,1)	.95 (2.32)	.95 (2.32)	.87 (2.12)	.78 (1.86)	.88 (2.12)
Free trade area (0,1)	.51 (3.62)	.51 (3.61)	.44 (3.11)	.47 (3.35)	.43 (3.09)
Ex-colonial relationship (0,1)	1.61 (13.1)	1.61 (13.1)	1.66 (13.5)	1.63 (13.3)	1.66 (13.3)
Ex-common-colonizer (0,1)	.25 (1.44)	.25 (1.42)	.32 (1.81)	.34 (1.89)	.32 (1.78)
Common language/outsider	**–0.02 (–0.47)**				
Common language/outsider (0,1)		**0 (0)**			
Currency union/ outsider (0,1)	.08 (1.05)	.08 (1.03)	.1 (1.26)	.1 (1.27)	.1 (1.25)
Political union/ outsider (0,1)	.19 (1.73)	.19 (1.73)	.18 (1.67)	.17 (1.6)	.18 (1.68)
FTA/outsider (0,1)	.15 (2.58)	.15 (2.61)	.14 (2.37)	.15 (2.5)	.14 (2.36)
Ex-colony/colonizer/outsider(0,1)	–.2 (–2.18)	–.2 (–2.2)	–.18 (–1.95)	–.18 (–1.98)	–.18 (–1.95)
R^2	.73	.73	.73	.73	.73
RMSE	1.73	1.73	1.72	1.72	1.72

Number of Observations: 31,010. Country-specific fixed effects and year-specific fixed effects are not reported. Student *t*s are in parentheses. These are based on robust standard errors after corrections for clustering of country pairs.

APPENDIX 2

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 Republic	.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 1
Falkland Islands	-	0	-	English	English 1
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, the data about languages (the last two rows) in this table is basically extracted from Grimes (2000) with ancillary use of the *CIA Country Factbook*. Literacy is from the *CIA Factbook* (with a few blanks filled in from Grimes). Language diversity is from Grimes (except in the one case of language diversity for Hong Kong, which is drawn from Taylor and Hudson (1972)). A zero for language diversity may mean that no calculation was made on the assumption that the number would be small (source: private correspondence). While the percentage figures for the spoken languages are almost exclusively derived from Grimes, 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. Persian refers to Farsi. Fulfulde is also sometimes referred to as Fula or Fulani.