STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

LANDUAGE AND FOREIGN TRADE

BY

JACQUES MÉLITZ

UNIVERSITY OF STRATHCLYDE, CREST-INSEE & CEPR

NO. 03-10

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

LANGUAGE AND FOREIGN TRADE

Jacques Melitz

University of Strathclyde, CREST-INSEE, and CEPR

First draft: July 2002

Second draft: September 2002
(CEPR Working Paper no. 3590)

This draft: February 2004

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

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

Jacques Melitz
Department of Economics, Curran Building
100 Cathedral Street
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.

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

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

Our typical use of binary variables in economics as indicators of a common language largely reflects the difficulty of quantifying the numbers of speakers of different languages in a country. But some headway is possible. Perhaps the most extensive basis for progress today is Grimes (2000), now in its 14th edition since first appearing in 1951. This work is a massive effort to condense the information supplied by the entire profession of ethnologists about world languages. There have 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). I shall do the same but with a somewhat different purpose: that of providing a general index of the ability to communicate directly in worldwide trade. This is not the usual aim. Hall and Jones focus on language because of a concern with certain institutional/legal features. Relatedly, 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. For their part, 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 their pioneering work, like Rauch and Trindade, to attribute only a single language to each person. They also rely on a far more summary treatment than Grimes', by Katzner (1986).

As an added interest, the latest edition of Grimes (2000) contains an index of linguistic diversity for all countries present in the study for the first time. The index concerns "the probability that any two people in the country picked at random will have different mother tongues" (Grimes (2000), p. x). The higher the index – the closer to one – the higher the probability that a random pair of individuals will have different mother tongues (see Lieberson (1981)). It would have served me better to have an index of the probability that any two people at random 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. Mauro (1995) uses a similar index: namely, Taylor and Hudson's (1972) index of ethnolinguistic fractionalization (which derives entirely from a detailed Soviet linguistic study dating to 1964). Despite the sociological emphasis of the title, "ethnolinguistic fractionalization," Taylor and Hudson's index pertains to the identical issue as Grimes': whether two randomly chosen individuals in a country will share the same maternal language. Since Mauro (1995), Alesina and Wacziag (1998), Easterly and Levine (1997) and La Porta et al. (1999) have also

exploited the Taylor-Hudson index. On the other hand, these authors use the Taylor-Hudson index to treat issues of societal division and government behavior,[1] whereas I shall use the Grimes index strictly in connection with communication. 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.

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

The discussion will begin by developing the specific gravity equation that will serve in the empirical analysis (Section I). Next, I will explain my two indices of a common language:

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

the one relating to direct communication and the other concerning translation. Section III will present the elementary econometric results. Section IV will then extend the analysis in three directions: first, to cover the effects of different languages; second, to treat the distinction between the scale (trade-creation) and the substitution (trade-diversion) effects of a common language; and third, to examine the contribution of network externalities of language to bilateral trade. The final section will contain some general discussion and suggestions for future research. All the raw linguistic series, including those I constructed, are in the data appendix.

I. The gravity model

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

In using the gravity model, I shall consider output as exogenous. I shall also 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. Under the preceding assumptions, Anderson and van Wincoop (2003) show that 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 consumer-based price levels in the two

countries. In the case of P_j:

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

where the summation sign embraces all i prices inclusive of p_j (in which case, exceptionally, $p_j t_{jj} = p_j$) and β_i is the distribution parameter of the utility function (for good i or varieties coming from country i). The corresponding equation holds for P_i. 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.

Distribution costs can be easily incorporated in the preceding model, and this is important since language plainly affects trade partly by modifying costs of distribution. As Baier and Bergstrand (2001) recently show (in line with Bergstrand (1985)), injecting distribution costs is no problem: it merely requires supposing that a CES relationship applies to the "transformation" of goods into sales between different national markets. So long as the elasticity of "transformation" is the same whether goods move one way or the other, the gravity equation retains the same structural form as equation (1). The only differences are that the parenthetical expression $(t_{ij}/P_i P_j)$ 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.[2]

One fundamental issue in estimating equation (1) is the proper treatment of the theoretical price indices, P_i and P_j. A simple way of handling the problem is to introduce a separate fixed effect for each individual country. Several authors have recently chosen this route, and Feenstra (2004, ch. 5) provides a lucid explanation of the justification for the choice. Yet country fixed-effects do bear one important inconvenience with major consequences below. They interfere with the use of other explanatory variables that happen to be country-specific and constant over time, since those variables will be perfectly correlated with the country dummy. In the present instance, this problem would rule out two of the basic linguistic vari-

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

ables in the analysis, literacy and linguistic diversity. As a result, I will consider P_iP_j in two forms: one containing individual influences on P_iP_j but no country fixed effects, and the other with fewer or no individual influences on P_iP_j but country fixed effects. The aforementioned separate influences on P_iP_j that appear will take one of two forms. They will be either third-country effects of influences on t_{ij} (and thus T_{ij}), or reflections of the opposite pull of domestic trade on foreign trade (associated with the condition $t_{jj} < t_{ij}$). The second sorts of influences, concerning the opposite pull of domestic trade, are generally country-specific and therefore incompatible with country fixed effects.

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

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

where the u terms are continuous variables, and the v ones are 0-1 dummies. But, in the case of P_iP_j, I will use either:

$$(4)\ P_iP_j = \prod_{k=1}^{k=r} y_k^{\rho_k} \times \exp\left(\sum_{k=1}^{k=s} \omega_k z_k\right) \text{ or}$$

$$(5)\ P_iP_j = \prod_{k=1}^{k=r} y_k^{\rho_k} \times \exp\left(\sum_{k=1}^{k=s} \omega_k z_k\right) \times \exp\left(\sum_{k=1}^{k=c} \delta_k C_k\right) \qquad \underline{r} \le r \quad \underline{s} \le s$$

In this case, the y terms are continuous variables while the z ones are dummies, and the C_k variables in equation (5) refer to the country fixed effects. Because of the presence of these fixed effects, some of the influences y and z in equation (4) drop out of equation (5) ($\underline{r} \le r$ and $\underline{s} \le s$).

Thus, the two estimated forms of the gravity model will be:

$$(6)\ \log T_{ij} = \text{constant} + \alpha \log (Y_iY_j) + (1-\sigma)\,\gamma_1 \log u_1 \ldots + (1-\sigma)\,\gamma_m \log u_m$$
$$+ (1-\sigma)\,\eta_1 v_1 \ldots + (1-\sigma)\,\eta_n v_n + (1-\sigma)\,\lambda_1 \log w_1 \ldots + (1-\sigma)\,\lambda_q \log w_q$$
$$+ (\sigma-1)\,\rho_1 \log y_1 \ldots + (\sigma-1)\,\rho_1 \log y_r + (\sigma-1)\,\omega_1 z_1 \ldots + (\sigma-1)\,\omega_s z_s + e_{ij}$$

and

$$(7)\ \log T_{ij} = \text{constant} + + \alpha \log (Y_iY_j) + (1-\sigma)\,\gamma_1 \log u_1 \ldots + (1-\sigma)\,\gamma_m \log u_m$$
$$+ (1-\sigma)\,\eta_1 v_1 \ldots + (1-\sigma)\,\eta_n v_n + (1-\sigma)\,\lambda_1 \log w_1 \ldots + (1-\sigma)\,\lambda_q \log w_q$$
$$+ (\sigma-1)\,\rho_1 \log y_1 \ldots + (\sigma-1)\,\rho_{\underline{r}} \log y_{\underline{r}} + (\sigma-1)\,\omega_1 z_1 \ldots + (\sigma-1)\,\omega_{\underline{s}} z_{\underline{s}}$$
$$+ (\sigma-1)\,\delta_i C_i + (\sigma-1)\,\delta_j C_j \ldots + (\sigma-1)\,\delta_c C_c\, e_{ij} \qquad \underline{r} \le r \text{ and } \underline{s} \le s$$

The constants in equations (6) and (7) refer to Y^W. The coefficient α of the $\log (Y_iY_j)$ term

in both equations (whose its presence in equation (7) depends on some temporal dimension) should evidently equal one. The country dummies C_i and C_j equal one while the rest of the C_k dummies in equation (7) ($k \neq i, j$) are zero. The $\underline{r}$ members of y_k in equation (7) refer to the y_k variables in equation (6) that are not defined by country but by country pair ($\underline{r} \leq r$) while the $\underline{s}$ members of z_k in equation (7) are the corresponding z_k variables in equation (6) ($\underline{s} \leq s$). The new w_k terms refer to some possible influences on balanced-bilateral-trade that are consistent with the assumptions of the model but are not reflected in any other way. A fitting example of w_k would be differences in climate. If large differences in climate between country pairs imply better opportunities for trade between them than with other countries that have more similar climates, the variable should be added. It would not be reflected otherwise – in consumer tastes, general price levels, or trade frictions – but would be consistent with exogenous output and balanced bilateral trade. On the other hand, the exchange rate would be ineligible as a w_k term, since this variable necessarily has opposite implications for two trading partners. The term e_{ij} is white noise associated with the dependent variable, bilateral trade.

The bulk of the empirical work will concern equation (6). However, the test of equation (7) will serve as an essential check of robustness. In general, equation (7) has the important merit of adding missing variables. But it has the obvious drawback of mixing up many influences by country and thereby hiding some separate effects on P_iP_j that may be of primary interest.

The principal u, v, w, y and z terms in the estimates are as follows. I leave out the language variables for later treatment.

$\log u_1$ = log of (product of) relative distance.

$v_1, \ldots, v_n$ = adjacency, currency union, political union, free trade area, ex-colonial relationship, ex-common colonizer.

$\log w_1$ = North-South difference.

$\log y_1$, $\log y_2$, $\log y_3$ = log of (product of) remoteness, log of (product of) population, log of (product of) land area.

z_1 = number of landlocked in pair.

As regards the sole u variable, u_1, relative distance refers to the distance between two

countries (d_{ij}) divided by the square root of the product of their remoteness from the rest ($(R_iR_j)^{0.5}$). The "remoteness" of a country (R_i) refers to the straight-line average distance of the country (i) from all of the others in the sample (including country j). This is not the ordinary definition of "remoteness". But I adopt it for want of a fitting synonym, even though the use of "remoteness," in the usual sense, in testing gravity equations has come under serious criticism by Anderson and van Wincoop (2003). In usual parlance, the term refers to a *weighted*-average distance of a country in a trading pair to all third-countries where the weights rest on output. In light of those weights, the variable provides, or is supposed to provide, a broad measure of alternative trading opportunities, and thus may serve as a nearly general surrogate for P_iP_j. It is this usual treatment of "remoteness" as a broad reflection of P_iP_j that has been the subject of Anderson and van Wincoop's criticism. The same usage will not occur here. At present, "remoteness" will emphatically refer strictly to geography. There are no output weights; and "remoteness" is only one of a number of separate reflections of "multilateral trade resistance" (P_iP_j) in the study. The decomposition of distance between relative distance and remoteness ($\log (d_{ij})/((R_iR_j)^{0.5}) + \log (R_iR_j)^{0.5} = \log d_{ij}$) basically serves to permit a separate estimate of the effect of distance in shifting trade away from more distant foreign trade partners toward closer ones (an element of bilateral trade resistance) and between all foreigners and fellow countrymen (an element of multilateral trade resistance) (see Melitz (2004a, b)). This is my sole reason for the decomposition.

Besides adjacency, the v variables are the indices of political association that Frankel and Rose (2002) have successfully used before. The concern with ex-colonial relationships and an ex-common colonizer is particularly important here, 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. The only w variable in the study is the North-South difference between trading partners, as measured by the absolute difference

in the latitudes between the two.[3] This variable reflects the fact that a country's latitude affects the length of its days, its sunlight, its temperatures and seasons, and thus may alter not only its plant and animal life and the yield of its lands and waters, but also its required insulation, energy and equipment, and its optimal production techniques. As I have argued before (Melitz (2004a, b)), the North-South difference may therefore foster trade based on comparative advantage.

In regard to multilateral trade resistance, discussion of the first indicator, remoteness (as measured by log $(R_iR_j)^{0.5}$), has already preceded. Frankel and Romer (1999) provide a particularly cogent explanation for the next two, population and land area. As they observe, the two reflect the negative pull of domestic trade on foreign trade. The more people there are at home, the wider the opportunities to trade domestically, and therefore without bearing the costs of foreign trade. Land area, in turn, is an indicator of internal distance (which has sometimes been measured, perhaps more intuitively, as a line between two domestic locations, as in Wei (1996)). Other things equal, including the home population, a big country will have fewer foreigners to trade with at close range. This will tend to favor domestic trade. I consider landlocked as a z variable because it is defined by country and thus relates to the general choice between domestic and foreign trade, independently of trade partner. All of these last y and z terms drop out when country fixed effects are admitted. But I will subsequently add some y and z terms reflecting third-country effects, and these terms, relating to P_iP_j, will potentially coexist with the country-fixed effects.

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

[3] Let lat1 and lat2 stand for the respective latitudes of country 1 and country 2 in a trading pair (with Northern latitudes positive and Southern ones negative). Then the North-South difference is |lat1 – lat2|. A related variable that now appears often in studies of economic growth is Distance from the Tropics, |lat1| + |lat2|. But the variable proves insignificant in my tests regarding trade.

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

Indeed, it is possible to turn the whole issue of a common language on its head, and to ask how there can possibly be any language barrier when translation can link up all world languages to one another. To this query, there are two answers, whose joint significance is clear, but whose relative significance is not. The first 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. According to this treatment, the problems of mounting and maintaining a language network are so large that, regardless of population size and number of languages, only two common languages exist between any pair of countries at most. The second answer to the query about the possible significance of language barriers is the importance of direct communication. We have some clear indication of the importance of such 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 on trade is the immigrants' 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 separately. One will relate to a communication network and the other to direct communication. In doing so, I will use only one supplementary source besides Grimes (2000): the *CIA country factbook* – a frequent reference among economists. The decision to stick to Grimes and the *CIA factbook*, following a broader search, is mainly aimed 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.[4] 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, 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. 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 185 "countries" in the study (including some overseas departments and territories). Lowering the re-

[4] 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.

quired 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.[5] 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 I relied on this source for the literacy rates, and only used Grimes to fill in missing values. My reason for switching to the *factbook* for literacy rates lies in the much wider discrepancies in the dating of this particular variable in Grimes. Direct Communication obtains by summing up the products of the respective percentages of speakers over all the relevant languages (at least 4%) in the two trading countries. In principle, those values could have exceeded one because of bilingualism. But very few such cases arose. In those cases, I set the numbers equal to one, which was equivalent to a general normalization.

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. In addition, linguistic diversity at home should do the same since in its presence linguistic barriers even exist at home. Finally, linguistic obstacles must be more severe for people who cannot read and write, whereas alternative forms of communication – through speech and action – are easier at home. Therefore, those people should be especially handicapped in foreign trade. Thus, literacy should promote foreign relative to home trade.

---

[5] 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. He deals exclusively with bilateral trade with the United States. Paradoxically, English as a second language emerges, if anything, as more significant. However, judging from the results, the distinction between English as a first or a second language would not pass any statistical test of significance.

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. 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. (This assignment follows corrections for a few slips in the working paper version of their article that happened to catch my attention in an earlier study (Melitz (2004a)), most of which Frankel and Rose have since repaired). Table 2 shows the correlation matrix relating to my two indices of a common language and their single one over the (approximately) 30,000 observations of bilateral trade in the statistical analysis. The 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 – 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, 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 cor-

relation coefficient between my two language indices, Open-circuit and Direct Communication, is high, 0.73, but not so high as to undermine the distinction – especially if we consider that both indices are equally zero in nearly three-quarters of the cases.

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. The 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 corrections concerning language). First, I measure geographical distance differently. Whereas he locates countries at their geographical center (in conformity with the CIA), I place them wherever their most populous city stands (as found in the CD-Rom *encarta*). Second, I consider all departments and territories of a country as automatically belonging to a free trade zone in the mother country. As a result, my dummies for a common country and free trade area are mutually exclusive.

The first column of Table 3 shows the result of the test of equation (4) in the usual form, with (the log of) bilateral trade (nominal imports plus exports in dollars deflated by the U.S. GDP chain price index and divided by two) as the dependent variable. As in Frankel and Rose, the test covers observations for six separate years at five-year intervals, starting with 1970 and ending in 1995, and therefore includes controls for the individual years. Robust standard errors are shown (after correction for clustering of data for individual trading pairs). All of the earlier hypotheses are confirmed. The results are basically the same as those reported in Melitz (2004a,b), which in turn do not differ from Frankel and Rose (2002) in regard to the common variables in both studies (that is, all of them except the decomposition of distance between relative distance and remoteness, North-South, and the language variables).

The rest of the discussion focuses on language.

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

The one unsatisfactory feature of the previous tests is the 1.4 output-elasticity of trade instead of one. However, this coefficient is impossible to interpret independently of the negative coefficient of population in the equation, since output and population are closely correlated (compare Frankel (1997, pp. 57-61)). Yet, at least on my interpretation, theoretically the two influences are distinct. One of them concerns the feature of homothetic preferences, whereas the other reflects the opportunities to trade at home and free of any of the obstacles to foreign trade. In addition, joint output is not really independent of bilateral trade, and very significantly too, literacy interacts with per capita output (not surprisingly, in light of the influence of literacy on output: see, for example, Barro (1991)). If we regress (the log of) joint output (Y_iY_j) on (the log of) bilateral trade, (the log of) joint population, and joint literacy,

control for individual years, and correct the standard errors for clustering, we get the following estimate:

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

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

In response to this problem, the next two estimates of equation (4) impose a unitary elasticity of influence of joint output on trade.[6] The variable on the left hand side is then the ratio of bilateral trade to the product of the two countries' outputs. Column 4 replicates the estimate in column 1 following this change. The R-square jumps up from 65 to 89 percent (to which I ascribe little importance). The product of the two countries' populations remains significant while its negative coefficient drops. But this negative coefficient now can be more properly interpreted as reflecting the impact of opportunities to trade with fellow citizens in diminishing foreign trade. The influence of remoteness also rises and becomes roughly equivalent to that of relative distance, though the coefficient of relative distance is still estimated far more precisely. (I have no explanation for this last alteration.) However, the most notable changes of all relate to the four linguistic variables. Those variables retain their signs and significance, but the impact of literacy shoots up. The most powerful linguistic influence on foreign trade in the study – by far – now emerges as the ability to read and write. According to the estimates, much more can be done to increase trade between two countries by promoting an extra percentage point of literacy than an extra percentage point of a common language. Of course, this impact of literacy could stem partly from production skills as well as the ability to communicate (especially since output is no longer separately present as an explanatory variable). Still, either way literacy is essential.[7]

---

[6] Compare Harrigan (1994), Frankel and Rose (1998), and Anderson and van Wincoop (2003).

[7] More troubling would be any associated impact of literacy on tastes and income elasticities of demand. I cannot explore the issue here, as doing so would require a more flexible version of the gravity model.

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

IV. Further Tests

(a) *Different languages*

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

I begin in Table 4 by repeating the estimate in column 4 of the previous table for convenience. The second column shows the result of the extreme assumption that English provides the only effective circuit of communication, or, in other words, that the other languages only affect trade through person-to-person communication. The general fit of the equation worsens somewhat and both the coefficient and the significance of Direct Communication rise. The third column is even more to the point: it shows what happens when we introduce Open-circuit Communication in English as a separate variable side by side with Open-circuit Communication in the other European languages (that is, after excluding Arabic, Chinese,

Hindi, Malay, Persian and Turkish from Open-circuit Communication). Open-circuit Communication in English emerges as totally insignificant, and the other Open-circuit Communication variable remains highly so. There is thus no evidence of the special advantage of English as opposed to the other European tongues in open-circuit communication.[8] I experimented next with English, Spanish, and French – or the three most important languages in the study – as the only languages providing an effective circuit of communication. The results do not justify the conjecture. On the other hand, as already intimated, the idea that the European languages as a group are the only ones to serve as effective circuits of communication in trade does receive notable support. Column 4 shows the outcome of the relevant experiment. In comparing with the wider definition of Open-circuit Communication in column 1, the new one clearly performs better, and this improvement holds up in all my subsequent experiments. I will therefore only report further results with Open-circuit Communication in the European languages, and in summarizing, I will use a straight-line average of this Open-circuit variable and Direct Communication, termed Common Language*. Column 5 shows the result of repeating the estimate in the last column of table 1 with Common Language* instead of Common Language.

(b) *Substitution or scale effects*

Does a linguistic tie raise trade between two countries at the expense of trade with third countries or increase foreign trade in the aggregate? This next sort of question has arisen before in connection with some of the other variables in the analysis: namely, the political variables: free trade agreement, currency union, and so forth. The usual approach is to introduce a dummy variable for cases where one or both countries in a trade belong to the relevant

[8] 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 as his using more countries, 63 of them, Frankel, Stein and Wei (1998) report no difference in the significance of English as opposed to Spanish, French or German.

political association with some third country but not with one another. However, this method cannot serve here – 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 is possible.

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

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

try member of a relevant political association trades with a country outside the association and 0 otherwise. Once those other dummy variables are added, the proposed 0-1 index for lack of linguistic ties drops substantially in significance while the continuous index (with positive values heavily concentrated in the 0.4-1 range) is much less affected. Therefore, I retained the continuous index.

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

(c) *Network externalities*

What about the presence of "network externalities" of a common language? Advertisers pay more to publicize at peak times and in well-frequented places. It is entirely plausible that speakers of a language would benefit from larger numbers of other speakers of the same tongue in the market. Perhaps the outstanding formal development of the idea comes from Church and King (1993) (who wrote with the issue of French and English in Canada in mind). A related development is in Lazear (1999), whose presentation has the added interest of centering the external benefits precisely on trade (though he does not focus on externalities himself). Random encounters take place between people. If they speak the same language, a trade occurs. Otherwise, it doesn't. The more individuals in a surrounding who speak the same language, the higher the probability that random encounters will result in trades. It should be noted that both Church and King and Lazear consider the issue to be strictly direct communication, as is not the case here. Numerous passing references to network externalities of a

common language 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.

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.

Accordingly, I constructed a separate measure of income for the previous sort of external effects. In this case, nine languages drop out (including one of the open-circuit languages, Swedish, which strictly functions between Sweden and Finland). This leaves twenty. For 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, based on a previous calculation founded on the data in the language appendix and 1980 real income.[9] The third language is simply ignored. Next, the respective contributions of the trading pair to the preceding English and Arabic income aggregates are deducted. Finally, a coefficient of .5 is applied to the remainder for English, one of .2 to the remainder for Arabic, and the two totals are added up. The log of this sum of world income serves in the estimates.

The results confirm the hypothesis of external effects coming from inside the trading countries but deny any similar external effects coming from the rest of the world. The confirmation of the external effects from within is shown in the second column of Table 5. Both the index of external effects and Common Language* can be seen to be significant at the 5% confidence level. The root mean square error also drops in this estimate. (Note that all 29 "source" languages function in this case.) In addition, the coefficient of Common Language* is lower than before. But this is only to be expected, since the impact of this last variable now comes partly from a joint effect with income. The coefficient of Common Language* is also estimated with much less precision (even though Common Language* and this index are uncorrelated over the pertinent observations, which are those where both variables are positive rather than zero by construction).

On the other hand, the index of external effects stemming from third countries is never

[9] The earlier calculation goes as follows, given the example of English. For each country, in turn, take 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. Do the same for Arabic, etc.

significant except when used alone, that is, in the absence of either Common Language* or the other index of external influence. As shown in column 3 of Table 5, if this variable is merely added in the previous equation, Common Language* is the only one of the three that remains significant. The earlier index of external effects – the one that would concern external effects coming from within the two countries – also becomes unimportant. The significance of the index of external influence stemming from third countries when included alone has no particular importance, but must be attributed to the index's positive correlation of .6 with Common Language* (over the relevant observations).[10] There is independent reason to think so. If the external effects of language coming from third countries were significant, we might have expected English to show up as a separate influence on bilateral trade in the previous test without any control for external effects (table 4). Such was not the case. This reinforces the conclusion that the only externalities affecting trade are those coming from inside the borders of the trading parties themselves.

(d) *Country-specific fixed effects*

Finally, consider the result of testing equation (7) rather than equation (6), or including fixed effects for all "countries" (185 fixed effects in all, 25 of which drop out because of insufficient observations). In this case, all the earlier variables that depend on country i or country j alone are removed. Those that drop out are remoteness, population, land area, landlocked, language diversity and literacy (population being the only one of the six that varies at all over time). However, in principle, the performance of the other variables – those whose values depend on a country's trade partner – should be unaffected. Accordingly, I retained these last variables in the next test: namely, relative distance, North-South, adjacency, Open-circuit Communication, Direct Communication, the political variables, and the indicators of

[10] It may be important to explain why this .6 correlation is consistent with the absence of any correlation between the other index of external effects – those coming from within the trading pairs – and Common Language*. The world income of English speakers is many times greater than the world income of speakers of any other language. English also contributes heavily to Common Language*. This combination of factors underlies the .6 correlation. On the other hand, there are many cases of tiny English-speaking countries trading with one another, while the index of the effects *within* the trading pairs depends strictly on the income of the trading parties themselves. This then explains the lack of any correlation of the earlier index of external effects with Common Language*.

substitution or complementarity. (But I also dropped the two indicators of the external effects of language, which depend on a unique country-specific value for income.) The results are shown in the last column of Table 5.

Except for the indices of substitution or complementarity and two of the political variables (namely, free trade agreement and a common ex-colonizer), none of the relevant coefficients or standard errors is notably affected. The significance of a common language even goes up. Specifically, while the coefficient and significance of Open-circuit Communication in a European tongue drop (though the Student *t* remains above 3), the coefficient and significance of Direct Communication double. (When used instead of Open-circuit and Direct Communication, Common Language* has a coefficient of .91 and standard error of .07.) The only deterioration associated with language in this next test is that the impact of a common language on third parties ceases to emerge. On the whole, the test is thus plainly corroborative. If we compare the results of the test – or the last column of Table 5 – with the earlier estimates, we find only a moderate increase in R^2 and decrease in RMSE associated with the country fixed effects. Hence, to all appearance, remoteness, population, land area, access to sea, language diversity and literacy capture many of the relevant features of countries (therefore, P_iP_j) that are covered by the country fixed effects.

V. Discussion and conclusion

We knew beforehand that a common language promotes trade. This study sheds light on the particular channels through which language exerts its influence, and introduces a number of influences 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. Unless we allow for special facilities for obtaining information issuing from *chosen* source languages, we miss a good deal of the ease of communication through language in foreign trade. Moreover, judging from the results, dictionaries, pictures, signs, numbers, and minimum vocabularies, etc., probably matter too in overcoming linguistic barriers. Why else would literacy make an important contribution? If people can read and write in any language, they can cope better with the problems posed by foreign languages in general. As further evi-

dence, the incentive to overcome language barriers is a factor. If people face linguistic obstacles at home, they somehow manage to trade more with foreigners. In sum, possessing foreigner languages is only one of several ways of overcoming the linguistic hurdles in trade with foreigners. Thus, no single measure adequately covers the topic of language in foreign trade. But every one of the linguistic influences in the study speaks to the importance of resolving linguistic problems in foreign trade.

The separate and distinct effects of Open-circuit and Direct Communication on trade may also reflect some specialization. The two channels of communication could operate largely in trade in different sorts of goods. Open-circuit Communication might be especially important in trade in homogeneous goods, where rudimentary communication may suffice, whereas Direct Communication may serve heavily in trade in heterogeneous goods, which may require more sophisticated intercourse. Rauch (1999) implicitly suggests a hypothesis of this sort. He argues that close personal relations between exporters and importers are not necessarily significant in trade of perfectly homogeneous goods, whereas they are so in trade in heterogeneous goods. Because of issues of information, "markets" suffice in one case, while private "networks" are necessary in the other. The hypothesis of the separate roles of Open-Circuit and Direct Communication in the two cases follows easily.[11] [12]

A few other results of the study probe further into specific aspects of language. English seems to offer no particular advantage in foreign trade, while the European languages as a whole do. These emerge as better instruments of Open-channel Communication than other languages. In addition, linguistic links between countries seem to boost their bilateral trade

[11] In the case of certain homogeneous goods, including crude oil and many primary products, the relevant information may 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 may then resolve the communication problem altogether at no notable cost.

[12] Interestingly, Rauch and Trindade (2001) apply the proposed hypothesis to the distinction between a common language and an ethnic network rather than two separate uses of language in communication. They reason that a common language should affect trade in all categories of goods the same way, whereas an ethnic network should promote trade in differentiated goods more than in homogeneous ones. In their study, the ethnic network 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. Even a common *native* language in a worldwide sample may mostly reflect the ability to communicate. I will return to the issue in the text.

partly at the expense of trade with third countries, but not entirely. Finally, there is some evidence that network externalities of language contribute to trade between countries. But so far as this is true, the external benefits depend on the aggregate income of the particular trading pairs. On the whole, network externalities are not as central in trade as they are sometimes imagined to be, and the reason may have to do with the earlier point that a lot of market information gets transmitted in the absence of a common language.[13]

In closing, one important limitation of the work deserves emphasis. 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. Furthermore, in the case of Direct Communication, I have taken steps to assure the primacy of communication by including second-language and non-native speakers. (In the case of Open-circuit Communication, assuring this primacy was never even an issue.) But this is not to deny that additional variables reflecting culture or ethnicity would be helpful. Based on previous work, one cultural variable that easily comes to mind is the stock of immigrants. As the analysis now stands, immigrants are implicit here since they affect the index of linguistic diversity. But this index also reflects long-standing, sometimes ancestral, linguistic divisions inside national boundaries. 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 in the analysis would generally aid in interpreting the linguistic variables as relating essentially to communication (compare Rauch and Trindade (2001) and the discussion in note 12).

Another useful extension would be to study trade in goods that are especially con-

[13] It may be worth noting as well that a common language is likely to be more important in production than in trade. To all indication, even varying shades in the ability to communicate *directly* matter a lot in the labor market. Empirical work uniformly shows a considerable impact of linguistic skill on wages: see McManus, Gould and Welch (1983), Chiswick and Miller (1995), and the references in a broad survey by Grin (1996).

nected to language, such as movies, television programs, books, and vocal music. In these cases, the world dominance of English and the production by the U.S. in particular, is notorious. Cultural products in English sell extremely well in many places where English is otherwise secondary in foreign as well as domestic trade. Broadly, how come English plays no special role in facilitating foreign trade but dominates the market for language-related products?[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 largely associated with Bergstrand (1985)).[15]

Finally, my treatment of the language variables as fixed effects can be questioned. Literacy rates have risen substantially in the last fifty years. Migrations accelerated in the nineties in many parts of the world. Spanish is now significant in the U.S., Russian in Israel, etc. I have treated the data furnished in Grimes (2000) as roughly contemporaneous. In doing so, I have followed the practice in economics of treating all indicators of a common language and linguistic diversity in foreign trade as slow-moving variables that can be regarded as constants.[16] Indeed, Grimes (2000) 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. In the process, there could be an effort to show the evolutions of all three variables. Apart from an improved ability to test earlier hypotheses, one outcome would be the opportunity to tackle various historical questions, such as the relationship of linguistic changes to the increase in the number of countries in the post-World War II period, a topic that Alesina and Spolaore (1997) have effectively raised.

---

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

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

[16] In fact, this generalization applies to studies of trade but not labor. Labor studies of the impact of linguistic skills on wages do make prominent use of time series evidence (see the references in note 13).

REFERENCES CITED

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

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

Anderson, James and Eric van Wincoop (2001). "Gravity with *gravitas*: A solution to the border problem," *American Economic Review*, 93, 170-192.

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

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

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

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

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

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

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

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

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.

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

Feenstra, Robert (2004). *Advanced International Trade: Theory and Evidence*, Princeton University Press, chapter 5.

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.

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

Head, Keith and John Ries (1998). "Immigration and trade creation: Econometric evidence

from Canada," *Canadian Journal of Economics*, 31, 46-62.

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.

Helliwell, John and Geneviève Verdier (2001). "Measuring international trade distances: A new method applied to estimate provincial border effects in Canada," *Canadian Journal of Economics*, 34, 1024-41.

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

Melitz, Jacques (2004a). "Geography, trade and currency union," in Volbert Alexander, Jacques Melitz, and George von Furstenberg, eds., *Monetary Unions and Hard Pegs: Effects on Trade, Financial Development and Stability*, Oxford University Press and CEPR, chapter 5.

Melitz, Jacques (2004a). "Distance, trade and Political Associations," University of Strathclyde Working Paper.

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.

Wei, Shang-Jin (1996). "Intra-national versus international trade: How stubborn are nation states in globalization?" *NBER Working Paper 5531*.

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	1			
Frankel-Rose	0.81	0.75	1		
Literacy	–0.10	0.10	–0.03	1	
Diversity	0.14	–0.08	0.07	–0.43	1

TABLE 3: THE BASIC RESULTS

REGRESSAND:	**Log of bilateral trade**			**Bilateral trade ÷ product of GDP**	
Log product of relative distance	−1.25 (.03)	−1.26 (.03)	−1.25 (.03)	−1.21 (.04)	−1.21 (.04)
Log product of remoteness	−.31 (.15)	−.29 (.15)	−.31 (.15)	−1.09 (.15)	−1.09 (.15)
Log North-South Difference	.26 (.02)	.26 (.02)	.26 (.02)	.25 (.02)	.25 (.02)
Log product of real GDP	1.4 (.02)	1.4 (.02)	1.4 (.02)		
Log product of Population	−.44 (.02)	−.44 (.02)	−.44 (.02)	−.06 (.01)	−.06 (.01)
Log product of land area	−.18 (.01)	−.18 (.01)	−.18 (.01)	−.17 (.01)	−.17 (.01)
Adjacency (0,1)	.72 (.12)	.72 (.12)	.72 (.12)	.73 (.12)	.73 (.12)
Number of landlocked in pair (0, 1, 2)	−.29 (.04)	−.29 (.04)	−.3 (.04)	−.42 (.04)	−.42 (.04)
Common language: Frankel-Rose (0,1)			**.05 (.1)**		
Open-circuit communication (0,1)	**.28 (.08)**		**.26 (.09)**	**.38 (.08)**	
Direct communication	**.46 (.13)**		**.43 (.14)**	**.35 (.13)**	
Common language		**.69 (.07)**			**.73 (.07)**
Product of linguistic diversity	**.93 (.11)**	**.91 (.11)**	**.93 (.11)**	**.57 (.11)**	**.58 (.11)**
Product of literacy rate	**.37 (.11)**	**38 (.11)**	**.37 (.11)**	**1.81 (.09)**	**1.81 (.09)**
Currency union (0,1)	1.38 (.18)	1.39 (.18)	1.37 (.18)	1.23 (.19)	1.23 (.19)
Political union (0,1)	.71 (.45)	.74 (.45)	.73 (.45)	.76 (.44)	.75 (.44)
Free trade area (0,1)	1 (.1)	1.02 (.1)	1 (.1)	1.12 (.09)	1.12 (.09)
Ex-colonial relationship (0,1)	1.97 (.13)	1.97 (.13)	1.97 (.13)	2.02 (.12)	2.02 (.12)
Ex-common-colonizer (0,1)	.43 (.09)	.41 (.08)	.42 (.09)	.21 (.09)	.21 (.08)
R^2	.65	.65	.65	.89	.89
RMSE	1.95	1.95	1.95	1.98	1.98

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

TABLE 4: ENGLISH AND OTHER EUROPEAN LANGUAGES

REGRESSAND: **Bilateral trade ÷ product of GDP**					
Log product of relative distance	−1.21 (.04)	−1.23 (.04)	−1.22 (.04)	−1.21 (.04)	−1.21 (.04)
Log product of remoteness	−1.09 (.15)	−1.11 (.15)	−1.15 (.15)	−1.15 (.15)	−1.15 (.15)
Log North-South Difference	.25 (.02)	.26 (.02)	.25 (.02)	.25 (.02)	.25 (.02)
Log product of Population	−.06 (.01)	−.07 (.01)	−.07 (.01)	−.07 (.01)	−.06 (.01)
Log product of land area	−.17 (.01)	−.16 (.01)	−.17 (.01)	−.17 (.01)	−.17 (.01)
Adjacency (0,1)	.73 (.12)	.77 (.12)	.75 (.12)	.74 (.12)	.74 (.12)
Number of landlocked in pair (0, 1, 2)	−.42 (.04)	−.41 (.04)	−.43 (.04)	−.43 (.04)	−.43 (.04)
Open-circuit communi-cation: English (0,1)		**.33 (.09)**	**.10 (.10)**		
Open-circuit communi-cation: European (0,1)			**.39 (.09)**	**.44 (.08)**	
Open-circuit communication (0,1)	**.37 (.08)**				
Direct communication	**.35 (.13)**	**.61 (.1)**	**.33 (.12)**	**.33 (.12)**	
Common Language*					**.81 (.08)**
Product of linguistic diversity	**.57 (.11)**	**.58 (.11)**	**.55 (.11)**	**.57 (.11)**	**.58 (.11)**
Product of literacy rate	**1.81 (.09)**	**1.75 (.09)**	**1.79 (.09)**	**1.79 (.09)**	**1.79 (.09)**
Currency union (0,1)	1.23 (.19)	1.26 (.19)	1.21 (.19)	1.19 (.19)	1.19 (.19)
Political union (0,1)	.76 (.44)	.76 (.47)	.77 (.44)	.75 (.44)	.73 (.44)
Free trade area (0,1)	1.12 (.09)	1.03 (.09)	1.09 (.09)	1.1 (.09)	1.1 (.09)
Ex-colonial relationship (0,1)	2.02 (.12)	2.04 (.13)	2 (.12)	1.98 (.12)	1.98 (.12)
Ex-common-colonizer (0,1)	.21 (.09)	.23 (.09)	.17 (.09)	.18 (.09)	.19 (.08)
R^2	.89	.89	.89	.89	.89
RMSE	1.9824	1.9829	1.9814	1.9815	1.9815

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

TABLE 5: FURTHER TESTS

REGRESSAND: **Bilateral trade ÷ product of GDP**				
Log product of relative distance	−1.22 (.04)	−1.22 (.04)	−1.22 (.04)	−1.37 (.04)
Log product of remoteness	−1.23 (.15)	−1.21 (.15)	−1.21 (.15)	
Log North-South Difference	.24 (.02)	.24 (.02)	.24 (.02)	.19 (.02)
Log product of Population	−.03 (.01)	−.03 (.01)	−.03 (.01)	
Log product of land area	−.17 (.01)	−.17 (.01)	−.18 (.01)	
Adjacency (0,1)	.79 (.12)	.77 (.12)	.76 (.12)	.54 (.13)
Number of landlocked in pair (0, 1, 2)	−.38 (.04)	−.38 (.04)	−.38 (.04)	
Open-circuit Communi-cation: European (0,1)				**.24 (0.7)**
Direct Communication				**.82 (.11)**
Common Language*	**.67 (.08)**	**.37 (.17)**	**.41 (.2)**	
Network externality from within		**.013 (.006)**	**.017 (.012)**	
Network externality from outside			**−.005 (.014)**	
Product of linguistic diversity	**.68 (.12)**	**.66 (.12)**	**.66 (.12)**	
Product of literacy rate	**1.67 (.1)**	**1.66 (.1)**	**1.66 (.1)**	
Currency union (0,1)	1.39 (.19)	1.46 (.19)	1.47 (.2)	1.11 (.22)
Political union (0,1)	1.18 (.45)	1.22 (.45)	1.2 (.45)	.86 (.41)
Free trade area (0,1)	1.3 (.1)	1.34 (.1)	1.35 (.1)	.55 (.14)
Ex-colonial relationship (0,1)	1.45 (.13)	1.42 (.13)	1.42 (.13)	1.53 (.12)
Ex-common-colonizer (0,1)	.31 (.09)	.29 (.09)	.3 (.09)	.66 (.09)
Common language*/ outsider	**−.2 (.05)**	**−.19 (.05)**	**−.19 (.05)**	**−.02 (.05)**
Currency union/ outsider (0,1)	.34 (.04)	.34 (.04)	.34 (.04)	.11 (0.8)
Political union/outsider (0,1)	.31 (.05)	.31 (.05)	.31 (.05)	.16 (.11)
FTA/outsider (0,1)	.4 (.05)	.4 (.05)	.4 (.05)	.15 (.06)

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

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

APPENDIX

THE LANGUAGE DATA*

COUNTRY	LIT-ERACY	DIVER-SITY	LANGUAGES (Frankel-Rose)	OPEN-CIRCUIT LANGUAGES	SPOKEN LANGUAGES
Afghanistan	.31	.7	-	Persian	Pashto .45 Persian .4
Albania	.93	.26	-	-	Albanian . 98
Algeria	.62	.31	Arabic	Arabic, French	Arabic .83 French .2
American Samoa	.97	.12	English	English	English . 97
Angola	.42	.76	Portuguese	Portuguese	Portuguese .42
Anguilla	.95	0	English	English	English 1
Antigua & Barbuda	.89	0	English	English	English 1
Argentina	.96	.21	Spanish	Spanish	Spanish .92 Italian .04
Aruba	.97	0	-	Dutch, Spanish	Spanish .74
Australia	.99	.13	English	English	English .99
Austria	.98	.14	German	German	German .98
Bahamas	.98	.01	English	English	English .98
Bahrain	.85	.53	Arabic	Arabic	Arabic .85
Bangladesh	.38	.31	-	-	-
Barbados	.97	.09	English	English	English .97
Belgium	.98	.65	Dutch, French	Dutch, French	Dutch .56 French .56
Belize	.70	.70	English	English, Spanish	English .93 Spanish .35
Benin	.37	.90	French	French	French .37
Bermuda	.98	0	-	English	English 1
Bhutan	.42	.82	-	-	Nepali .08
Bolivia	.83	.68	Spanish	Spanish	Spanish .44 Quechua .36
Brazil	.83	.03	Portuguese	Portuguese	Portuguese .95
Brit. Indian Ocean Terr's	.98	0	-	English	English 1
British Virgin Islands	.98	.24	English	English	English .98
Brunei	.88	.45	-	Malay, English	Malay .88 English .05
Bulgaria	.98	.22	-	Turkish	Turkish .09
Burkina Faso	.19	.76	French	French	French .19
Burundi	.35	0	French	French	French .35
Cambodia	.35	.31	-	-	-
Cameroon	.63	.97	French, English	French, English	French .42 Fulfulde .30 English .21 Fang .05
Canada	.97	.55	English, French	English, French	English .65 French .22
Cayman Islands	.98	.58	-	English	English .98
Central African Republic	.60	.96	French	French	French .6
Chad	.48	.95	French	Arabic, French	Arabic .5 French .48
Chile	.95	.60	Spanish	Spanish	Spanish .93
China	.81	.48	Chinese	Chinese	Chinese .84
Colombia	.91	.03	Spanish	Spanish	Spanish .84
Comoros	.57	.01	French, Arabic	French, Arabic	French .3 Arabic .3
Congo Democratic 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, all the data in this table is extracted from Grimes (2000) with ancillary use of the *CIA Country Factbook* (except in the one case of language diversity for Hong Kong, which is drawn from Taylor and Hudson (1972)). Literacy is from the *CIA Factbook* (with a few blanks filled in from Grimes). Language diversity is from Grimes. A zero for language diversity may mean that no calculation was made on the assumption that the number would be small (source: private correspondence). The percentage figures for the spoken languages are almost exclusively derived from Grimes. But some inferences depend on the literacy rates, in which case the *CIA Factbook* enters as well. Official languages come from Grimes (2000) (except for a few isolated entries drawn from the *CIA Factbook*). An Open-Circuit Language is either official or has at least 20% speakers, and the maximum number of Open-circuit Languages is two. A Spoken Language is spoken by at least 4%. Some major national languages are omitted because they are neither accepted as official nor spoken by as many as 4% of the population in any trading country outside of the home one. Other languages do not appear because of missing trade data. This is notoriously true in regard to the languages in the ex-Soviet Union, including Russian. Persian refers to Farsi. Fulfulde is also sometimes referred to as Fula or Fulani.

Strathclyde Discussion Paper Series: 2003

03-01 Mathias Hoffman and Ronald MacDonald
A Re-Examination of the Link Between Real Exchange Rates and Real Interest Rate Differentials

03-02 Ronald MacDonald and Cezary Wójcik
Catching Up: The Role of Demand, Supply and Regulated Price Effects on the Real Exchange Rates of Four Accession Countries

03-03 Richard Marsh and Iain McNicoll
Growth and Challenges in a Region's Knowledge Economy: A Decomposition Analysis.

03-04 Linda Ferguson, David Learmonth, Peter G. McGregor, J. Kim Swales, and Karen Turner
The Impact of the Barnett Formula on the Scottish Economy: A General Equilibrium Analysis.

03-05 Tamim Bayoumi, Giorgio Fazio, Manmohan Kumar, and Ronald MacDonald
Fatal Attraction: Using Distance to Measure Contagion in Good Times as Well as Bad

03-06 Eric McVittie and J. Kim Swales
Regional Policy Evaluation: Ignorance, Evidence and Influence

03-07 Nicholas Ford and Graham Roy
How Differences in the Expected Marginal Productivity of Capital can Explain the Principal Puzzles in International Macro-Economics

03-08 Linda Ferguson, Peter G. McGregor, J. Kim Swales, and Karen Turner
The Regional Distribution of Public Expenditures in the UK: An Exposition and Critique of the Barnett Formula

03-09 Karen Turner
The Additional Precision Provided by Regional-Specific Data: The Identification of Fuel-Use and Pollution Generation Coefficients in the Jersey Economy

03-10 Jacques Mélitz
Language & Foreign Trade

http://www.economics.strath.ac.uk/Research/Discussion_papers/discussion_papers.html