Journal of International Economics 93 (2014) 351–363

Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

Native language, spoken language, translation and trade ☆

Jacques Melitz [a,c,d,e,*], Farid Toubal [b,e,1]

[a] Department of Economics, Mary Burton Building, Heriot-Watt University, Edinburgh EH14 4AS, UK
[b] Ecole Normale Supérieure de Cachan, Paris School of Economics, France
[c] CEPR, UK
[d] CREST, France
[e] CEPII, France

ARTICLE INFO

Article history:
Received 23 January 2013
Received in revised form 1 April 2014
Accepted 1 April 2014
Available online 13 April 2014

JEL classification:
F10
F40

Keywords:
Language
Bilateral trade
Gravity models

ABSTRACT

We construct new series for common native language and common spoken language for 195 countries, which we use together with series for common official language and linguistic proximity in order to draw inferences about (1) the aggregate impact of all linguistic factors on bilateral trade, (2) the separate role of ease of communication as distinct from ethnicity and trust, and (3) the contribution of translation and interpreters to ease of communication. The results show that the impact of linguistic factors, all together, is at least twice as great as the usual dummy variable for common language, resting on official language, would say. In addition, ease of communication plays a distinct role, apart from ethnicity and trust, and so far as ease of communication enters, translation and interpreters are significant. Finally, emigrants have much to do with the role of ethnicity and trust in linguistic influence.

1. Introduction

It is now customary to control for common language in the study of any influence on bilateral trade, whatever the influence may be. The usual measure of common language is a binary one based on official status. However, it is not obvious that such a measure of common language can adequately reflect the diverse sources of linguistic influence on trade, including ethnic ties and trust, ability to communicate directly, and ability to communicate indirectly through interpreters and translation. In this study we try to estimate the impact of language on bilateral trade from all the likely sources by constructing separate measures of common native language CNL, common spoken language CSL, common official language COL, and linguistic proximity LP between different native languages. The interest of this combination of measures is easy to see. If CSL is significant in the presence of CNL, the significance of CSL would say that ease of communication acts separately beyond ethnicity and probably trust. The additional importance of COL, in the joint presence of CSL and CNL, would indicate the contribution of institutionalized support for translation from a chosen language into the others that are spoken at home. If LP proves significant while all three previous measures of a common language are present, this would reflect the ease of obtaining translations and interpreters when native languages differ and without any public support, and perhaps also the influence of ethnic rapport between groups when their native languages differ. We base our measures of CSL and CNL on the products of the percentages of speakers in a country pair. The product would then represent the probability that two people at random from a pair of countries understand one another in some language in the case of CSL and in their native language in the case of CNL. Evidently, CSL is equal to or greater than CNL and both go from 0 to 1. COL is the usual binary (0, 1) measure. Our LP measure comes from an international project by ethnolinguists and ethnostatisticians, the Automated Similarity Judgment Program or ASJP (see Brown et al., 2008), that provides an index of similarities of words with identical meanings for a limited vocabulary of words between different language pairs based on expert judgments.[2]

Our results show that all 4 measures are jointly important. It is indeed difficult to capture the varied sources of linguistic influence along any single dimension, whether the dimension be the ability to speak, native speech, or official status. The popular measure, COL,

☆ The authors would like to thank Paul Bergin, Mathieu Crozet, Ronald Davies, Peter Egger, Victor Ginsburgh, Thierry Mayer, Marc Melitz, Giovanni Peri, the members of the economics seminars at CES-Ifo, ETR Zurich, Heriot-Watt University, the Paris School of Economics, the University of California at Davis, UCLA, and University College Dublin, and two anonymous referees for valuable comments.

* Corresponding author at: Department of Economics, Mary Burton Building, Heriot-Watt University, Edinburgh EH14 4AS, UK.
E-mail addresses: j.melitz@hw.ac.uk (J. Melitz), ftoubal@ens-cachan.fr (F. Toubal).
[1] 61, avenue du Président Wilson, Bat Cournot, Office 503. 94235 Cachan cedex, France.

[2] For an earlier use of the ASJP databank in a trade study that centers on four particular languages, English, French, Spanish and Arabic, see Selmier and Oh (2012).

http://dx.doi.org/10.1016/j.jinteco.2014.04.004
0022-1996/

underestimates the total impact of language at least on the order of one-half. This reinforces the earlier conclusion of Melitz (2008), which, however, had rested on far poorer data. Further, Melitz had merely taken for granted that the influence of language depends on ease of communication without paying separate attention to common native language and the associated roles of ethnicity and trust. We also push the analysis forward in three directions. First, we control for some factors that could have correlated effects on affinities and trust but are not always taken into account in studying language: namely, common religion, common law and the history of wars since 1823. Second, we investigate the impact of our linguistic variables for the Rauch classification between homogenous, listed and differentiated goods. Finally, we study the separate role of immigrants.

Of course, once we allow CSL to enter in explaining bilateral trade, we open the door to simultaneity bias. In response, we propose a measure of common language resting strictly on exogenous factors for use as a control for language in studies of bilateral trade when the focus is not on language but elsewhere. This measure depends strictly on CNL, COL and LP, and not CSL. However, when the subject is language itself, for example, the trade benefit of acquiring second languages or else the case for promoting second languages through public schooling in order to promote trade, a joint determination of bilateral trade and common language will be required. It will then be necessary to go beyond our work. Notwithstanding, we believe our work to be an essential preliminary for such later investigation. Any effort to determine bilateral trade and common language jointly must capture the main linguistic influences on trade and be able to measure those influences. In addition, the large role of acquired languages, interpreters and translation in trade that we bring to light matters both for empirical analysis and public policy. Empirically, it means that firms can expand their foreign trade by training labor in foreign languages and hiring people with foreign language skills who are not necessarily native speakers. As regards public policy, the study supports the value of foreign languages in school curricula. In the closing Section, we will return to the empirical and normative implications of our study.

The next Section contains the basic gravity model of bilateral trade that we will use, where we shall explain our controls in order to study language. In the following Section, we will discuss our data and our measures. Section 4 shall concern the econometric specification. All of our results depend strictly on the cross-sectional evidence in the ten years 1998 through 2007. We shall use panel estimates for 1998–2007 to summarize the results but only in the presence of country-year fixed effects so that the results depend strictly on the cross-sections. Though we shall base our quantitative estimates on these panel estimates, the yearly evidence will always be a point of reference and we shall expose any doubts that arise based on this evidence.[3] Section 5 contains our baseline results, resting on OLS. Since our main analysis depends on the positive values for trade, we will also entertain the issue of the zeros in the trade data in the next Section (6). Section 7 will then study separately each of the three Rauch classifications. Section 8 will propose our aforementioned aggregate index of a common language based on exogenous factors. According to this measure, on a scale of 1 to 100 a one-point increase in common language from all the previous sources increases bilateral trade by 1.15%. Estimates based on official status alone would be around 0.5%. In terms of the literature, 0.5 corresponds precisely to the estimate in Frankel and Rose (2002) and in Melitz (2008). Two recent meta-analyses, by Egger and Lassmann (2012) and Head and Mayer (2013), which cover many studies, respectively report coefficients of around 0.44 and 0.5. Section 8 introduces cross-migrants. As will be seen, cross-migrants have a clear impact on bilateral trade, though one that is difficult to assess exactly because of simultaneity bias. Perhaps part of migrants' influence is independent of language. But isolating this part would be a separate project. According to our analysis, the influence of cross-migrants may account for a high proportion of the role of ethnicity and trust in explaining linguistic effects on bilateral trade. In addition, since our work assumes that the particular language does not matter for the results, Section 9 will examine this assumption for English. We find no separate role for this language, nor for any of the other major world ones. Section 10 will contain our concluding assessment. There we will also return to the wider implications of our study.

[3] The yearly evidence itself is available in online Appendix A as well as largely in the earlier working paper version, Melitz and Toubal (2012).

2. Theory

We shall use the gravity model in our study with a single minor adaptation: namely, to treat the differences in prices on delivery (cif) from different countries as stemming either from trade frictions, as is usually done, or else from Armington (1969) preferences for trade with different countries. This will allow for the possibility that the influence of common language reflects a choice of trade partners as such rather than trade frictions. The basic equation, which remains founded on CES preferences in all countries, is:

$$M_{ij} = \left(\frac{t_{ij}}{P_i P_j}\right)^{1-\beta} \frac{Y_i Y_j}{Y_W} \tag{1}$$

M_{ij} is the trade flow from country j to country i. Y_i and Y_j are the respective incomes of the importing and exporting countries and Y_W is world output. β is the elasticity of substitution between different goods and greater than 1. P_i is the multilateral trade resistance of the importing country and P_j is the multilateral trade resistance of the exporting country. t_{ij} is a set of trade frictions or aids to trade, where the aids can take the form of discounts that the firms allow out of ethnic ties or trust. Those t_{ij} terms also depend on a combination of fixed costs or aids, affecting the number of firms, and variable costs or aids, affecting the production by firm. The M_{ji} equation is the same with t_{ji}/P_iP_j instead, but t_{ij} need not equal t_{ji}, thereby admitting unbalanced trade.

We shall not be interested in the decomposition of t_{ij} (or t_{ji}) between fixed and variable components, and therefore, quite specifically, we shall only be interested in the sum impact of language on trade. Otherwise, the instances of zero bilateral trade would have special significance, as Helpman et al. (2008) have shown. We will also not concern ourselves with the symmetry of the respective impacts of linguistic influences on imports in the two opposite directions for a country pair. Recent work would imply that the linguistic effects reflecting trust between country pairs are notably asymmetric (see Guiso et al., 2009; Felbermayr and Toubal, 2010). We shall disregard the point.

Next, we propose to model t_{ij} in a convenient log-linear form, namely

$$t_{ij} = D^{\gamma_1} \times \exp\left(\sum_{k=2}^{n} \gamma_k v_{ij,k}\right) \tag{2}$$

where D is bilateral distance and the v_{ij} terms are bilateral frictions or aids to trade. Accordingly, γ_1 is an elasticity and $[\gamma_k]_{k=2,\ldots,n}$ is a vector of semi-elasticities. Except for 2 cases that we will explain in due course, all of the v_{ij} terms are either 0, 1 dummies or else continuous 0–1 values going from 0 to 1.

COL, CSL, CNL, and LP will be separate v_{ij} terms. Melitz (2008) interprets the dummy or 0,1 character of COL as implying that status as an official language means that all messages in the language are received by everyone in the country at no marginal cost, regardless what language they speak. There is an overhead social cost of establishing an official language and therefore a maximum of two languages with official status in accord with the literature. But once a language is official, receiving messages that originate in this language requires no private cost, overhead or otherwise: everyone is "hooked up." Here we shall follow this view except on one important point. For reasons that will

emerge later, we will consider the presence of a *private* once-and-for-all overhead cost of getting "hooked up". This leads us to abandon the reference in Melitz to "open-circuit communication". As always, if COL equals 1 a country pair shares an official language and otherwise COL equals 0.

As mentioned in the introduction, CSL is a probability (0–1) that a pair of people at random from two countries understands one another in some language and CNL is the 0–1 probability that a random pair from two countries speak the same native language. LP refers to the closeness of two different native languages based on the similarity of words with identical meanings, where a rise in LP means greater closeness. As a fundamental point, LP is therefore irrelevant when two native languages are identical. For that reason, we never entertain LP as a factor when CNL is 1 and assign it a value of 0 in this case as well as when two languages bear no resemblance to one another whatever. In principle, we might have assigned LP a value of 1 rather than 0 when CNL is 1 and simply constructed a combined 0–1 CNL + LP variable with LP adding something to the probability of communication in encounters between people when their native languages differ. However, our measure of LP rests on a completely different scale than the one for CNL. Furthermore, we wanted to distinguish the issue of translation and ability to interpret from that of direct communication so far as we could. For these reasons, we prefer to estimate the two influences separately (in a manner that we shall discuss) and assign separate coefficients to them though we shall try to combine them eventually.[4]

The additional v_{ij} terms are required controls in order to discern the impact of linguistic ties on bilateral trade. Countries with a common border often share a common language. Pre-WWII colonial history in the twentieth century and earlier is also highly important. People in ex-colonies of an ex-colonizer often know the language of the ex-colonizer and, as a result, people in two ex-colonies of the same ex-colonizer will also tend to know the ex-colonizer's language. We therefore use dummies for common border, relations between ex-colonies and ex-colonizer and relations between pairs of ex-colonies of the same ex-colonizer as additional v_{ij} terms and we base ex-colonial relationships on the situation in 1939, at the start of WWII.[5]

In addition, we wanted to reflect some additional variables that have entered the gravity literature more recently and could well interact with the linguistic variables. These are common legal system, common religion, and trust (apart from whatever indication of trust a common language provides). A common legal system affects the costs of engaging in contracts, a consideration not unlike the costs of misunderstanding that result from different languages. A common religion creates affinities and trust between people just as CNL might. On such reasoning, we added a 0,1 dummy for common legal system, and created a continuous 0–1 variable for common religion that reflects the probability that two people at random from two countries will share the same religion. To reflect trust as distinct from native language was a particular problem. Guiso et al. (2009) had exploited survey evidence about trust as such in an EU survey of EU members. We have no such possibility in our worldwide sample. They also used genetic distance and somatic distance to reflect ancestral links between people. However, no one has yet converted these indices into worldwide ones for all country pairs.[6] The only measure of ancestral links of theirs that we were able to use readily is the history of wars; or at least we could do so by limiting ourselves to wars since 1823 rather than 1500 as they had. This more limited measure of ancestral conflicts, it should be noted, has already proven useful in related work concerning civil wars by Sarkees and Wayman (2010) (to say nothing of related work by Martin et al. (2008) where the civil war data starts only in 1950).

We assume that all of the previous controls are exogenous. We also experimented with two controls that are clearly endogenous and are prominent in the literature, free trade agreements and common currency areas. As neither had any effect on the results for language, and they have no special interest here, we decided to drop them. On the other hand, as already indicated, we experimented widely with another endogenous variable that is clearly eminently related to language: namely, cross-migration.[7] This next variable only figures prominently in work on gravity models when it is itself the primary subject of investigation. Therefore, we decided to estimate the impact of linguistic influences in its absence in our main investigation and to deal with it separately later. So doing also provides us with an estimate of linguistic effects in our baseline investigation where the only endogenous variable is CSL.[8]

3. Data and measures

Obviously crucial for our work was an ability to construct separate series for CSL, CNL, COL and LP. Of the four, the only easy series to construct is COL. CNL was the easiest one to build of the other three. In principle, we could have done so based on a single source, *Ethnologue*, or perhaps *Encyclopedia Britannica* (which contains less detailed information) as Alesina et al. (2003) did, though we proceeded differently. However, constructing series for CSL and LP was a considerable challenge. We shall open our discussion of the data series with the language variables.

3.1. Common official language

There are quite a few countries with many official languages (see the Wikipedia "list of official languages by state"). However, work on the gravity model generally admits only two. If we interpret COL in our way as implying that the relevant official language(s) is (are) available to anyone in the country in a language the person understands, this choice seems entirely reasonable and we shall follow it. Regarding the choice of the two official languages, we shall rely on the usual source, the CIA World Factbook, but we considered the broader evidence.[9] In cases where the two-language limit as such posed an issue, we kept the two most important in total world trade. This meant keeping English and Chinese in Singapore but dropping Malay, which is rather important in the region (a problematic case). As a result of this exercise, all in all, we have 19 official languages (only 19 since a language must be official in at least 2 countries in order to count). These languages are listed in Table 1.

[4] When we do combine the two, we also render the series for LP comparable (at the means) to the one for COL, the other linguistic series that refers to translation.

[5] Common country also sometimes enters as a variable in gravity models because of separate entries for overseas territories of countries (e.g., France and Guadeloupe). Our database does not include these overseas regions separately (e.g., Guadeloupe is included in France).

[6] In a related study to that of Guiso et al. (2009), Giuliano et al. (2006) also limited their use of genetic and somatic indices to Europe.

[7] It is clear from earlier studies that cross-migration hinges partly on bilateral trade even though the work thus far has tended to concentrate on the impact the other way, that is, that of emigrants on trade.

[8] Of course, the influence of cross-migration means that native languages are not fully exogenous, as is mostly neglected, especially outside of studies of long time series, and we do the same here.

[9] As an example of the insufficiency of the *Factbook*, English was adopted as an official language in Sudan only in 2005, during our study period, while Russian was adopted officially in Tajikistan in 2009, since our study period. However, in Tajikistan, Russian had continued to be widely used uninterruptedly in government and the media since the breakdown of the Soviet Union in 1990, whereas there is no reason to believe that the decision of Sudan to adopt English was independent of trade in our study period. Similarly, in some countries, though the language of the former colonial ruler was dropped officially after national independence, it remained in wide use in government and the media throughout. This pertains to French in Algeria, Morocco and Tunisia. Other issues arose. Thus, Lebanon has a law specifying situations where French may be used officially. German is official in some neighboring regions of Denmark. In the case of all such questions, we tended toward a liberal interpretation on the grounds that the basic issue was public support for the language through government auspices. Thus, we accepted German in Denmark, Russian in Tajikistan, French in Lebanon, Algeria, Morocco and Tunisia.

Table 1
Common languages.

Official, spoken and native languages	Other spoken and native languages
Arabic	Albanian
Bulgarian	Armenian
Chinese	Bengali
Danish	Bosnian
Dutch	Croatian
English	Czech
French	Fang
German	Finnish
Greek	Fulfulde
Italian	Hausa
Malay	Hindi
Persian (Farsi)	Hungarian
Portuguese	Javanese
Romanian	Lingala
Russian	Nepali
Spanish	Pashto
Swahili	Polish
Swedish	Quechua
Turkish	Serbian
	Tamil
	Ukrainian
	Urdu
	Uzbek

3.2. Common spoken language and common native language

CSL and CNL are best discussed together since we constructed them jointly. Our point of departure was the data from the EU survey in November–December 2005 (Special Eurobarometer 243, 2006), which covers the current 28 EU members (which only numbered 25 at the time) plus Turkey, a current applicant. The survey includes 32 languages. For spoken language, we summed the percentage responses to the question "Which languages do you speak well enough in order to be able to have a conversation, excluding your mother tongue (… multiple answers possible)" and for native language we recorded the percentage responses to "What is your maternal language." The rest of our data for spoken and native language by country was assembled from a variety of sources. We explain these sources in a separate online Appendix (Appendix A) where we include all of our raw linguistic data per country. As an important point, in collecting this data, we relied on information from the identical source for native and spoken language wherever possible, and when this could not be done, we gave preference to closer dates. By necessity, our figures range over the years 2001–2008.

In addition, because of our particular interests, we required all languages to be spoken by at least 4% of the population in two different countries in our world sample (as in Melitz, 2008). Lower ratios would have expanded the work greatly without affecting the results. The outcome is a total of 42 CSL languages, including all the 19 COL ones (but only 21 of the 32 in the EU survey).[10] The additional 23 CSL languages besides the COL ones are also listed in Table 1. Every CSL language is a CNL one.[11]

After the data collection, it was necessary to go from the national data to country pair data. This meant calculating the sums of the products of the population shares that speak identical languages by country pair. Some double-counting took place. Consider simply the fact that the 2005 survey allows respondents to quote as many as 3 languages besides their native ones in which they can converse. A Dutch and a Belgian pair who can communicate in Dutch or German and perhaps also French may then count 2 or 3 times in our summation. There are indeed 34 cases of values greater than 1 following the summation or the first step in the construction of CSL from the national language data.

In order to correct for this problem, we applied a uniform algorithm to all of the data in constructing CSL. Let the aforementioned sum of products or the unadjusted value of a common spoken language be α_{ij} where $\alpha_{ij} = \sum_1^n L_{1i}L_{1j}$ for country pair ij, L_1 is the percentage of speakers of a specific language and n is the number of spoken languages the countries share. The algorithm requires first identifying the language that contributes most to α_{ij}, recording its contribution, or $\max(\alpha_{ij})$, which is necessarily equal to or less than 1, and then calculating

$$\mathrm{CSL} = \max(\alpha) + (\alpha - \max(\alpha))(1 - \max(\alpha))$$

(where we drop the country subscripts without ambiguity). CSL is now the adjusted value of α that we will use. In the aforementioned 34 cases of α greater than 1 (whose maximum value is 1.645 for the Netherlands and Belgium-Luxembourg), $\alpha - \max(\alpha)$ is always less than 1. Therefore the algorithm assures that CSL is 1 and below.[12] In the other cases, whenever α is close to $\max(\alpha)$, the adjustment is negligible and CSL virtually equals $\max(\alpha)$. However, if α is notably above $\max(\alpha)$, there can be a non-negligible downward adjustment and this adjustment will be all the higher if the values of $\max(\alpha)$ are higher or closer to 1. This makes sense since values of $\max(\alpha)$ closer to 1 leave less room for 2 people from 2 different countries to understand each other *only* in a different language than the one already included in $\max(\alpha)$. We checked and found that the estimates of the influence of CSL on bilateral trade following the application of the algorithm raise the coefficient of CSL notably without changing the standard error in our estimates. This is exactly the desired result since it signifies that the adjustment eliminates a part of α that has no effect on bilateral trade (double-counting). We see no simpler way of making the adjustment.

Since we summed the products of the percentages of native speakers of common languages by country pair in constructing CNL in the same manner as for CSL, values greater than one could have arisen for CNL as well because the EU survey invites respondents to mention more than one maternal language if they consider that right. However, no such cases arose. In general, double-counting appears negligible in our calculation of CNL and no adjustment was needed.

3.3. Linguistic proximity

The LP measure raises distinct issues. In this case, the native language is at the heart of the matter regardless whether the language has any role outside the country. The problem is to correct for ease of communication between two countries if they have no common language, whether official, native or spoken. Thus, Japanese and Korean count even though they do not figure in CNL (as mentioned in note 10) and, for example, Tagalog is more relevant than English in the Philippines. In this case, 89 native languages matter. There would have been more except that in order to simplify, we only admitted 2 native languages at most in calculating LP. When there are 2, we adjusted their relative percentages in the country to sum to 1, the same score

[10] In identifying these 42 languages, we equated Tajik and Persian (Farsi); Hindi and Hindustani; Afrikaans and Dutch; Macedonian and Bulgarian; Turkmen, Azerbaijani, and Turkish; and Belarusian and Russian. In light of the 4% minimum, some large world languages fall out of our list, including Japanese, which is not spoken by 4% of the people anywhere outside of Japan, and including Korean (since we neglected North and South). Wherever languages qualified, we also recorded data down to 1% where we found it (though this does not affect our results). Of separate note, native speakers of Mandarin, the largest form of Chinese with 0.71 of the total native speakers, do not necessarily understand some of the other Chinese dialects, like Wu or Shanghainese (0.065) and Yue or Cantonese (0.052). Our treatment of Chinese as a single language follows *Ethnologue*, which terms it a *macrolanguage* on the ground of custom and the tendency of native speakers to identify themselves with the label. But in addition, we tested and found that excluding Chinese from our common languages has no impact on the results.

[11] This need not have happened. If any CSL language had failed to be a native language in more than a single country (even at the 1 percent level), it would have fallen out of the CNL group. No such case arose.

[12] The *lowest* value of CSL in these 34 cases is .75 and relates to Switzerland and Denmark, for which the unadjusted value α is 1.01. This CSL value implies 1 chance out of 4 that a Dane and a Swiss at random will not understand each other in any language and about the same chance (since α − CSL is .26) that they will understand each other in 2 languages or more.

we ascribed to a single native language. Thus, Switzerland shows 0.74 for German and 0.26 for French, Bolivia 0.54 for Spanish and 0.46 for Quechua. The minimum percentage we recorded for a native language was 0.13 for Russian in Israel. Very significantly too, we assigned 31 zeros. Those are cases of countries with a high index of linguistic diversity (in *Ethnologue*) and where no native language concerns a majority of the population. The underlying logic is clear. When languages are widely dispersed at home, the linguistic benefit of trading at home rather than abroad is muddy to begin with. Therefore, it is questionable to make fine distinctions about the distances of the 2 principal native languages to foreign languages.[13]

Next, we constructed two separate measures of LP, LP1 and LP2. LP1 is inspired by an idea in Laitin (2000) and Fearon (2003) (jointly and earlier in unpublished work), which since has been taken up in studies of various topics (see Guiso et al., 2009; Desmet et al., 2009a,b; Ginsburgh and Weber, 2011). The idea was to base calculations of linguistic proximities on the *Ethnologue* classification of language trees between trees, branches and sub-branches. We allowed 4 possibilities, 0 for 2 languages belonging to separate family trees, 0.25 for 2 languages belonging to different branches of the same family tree (English and French), 0.50 for 2 languages belonging to the same branch (English and German), and 0.75 for 2 languages belonging to the same sub-branch (German and Dutch) (Fearon, 2003 suggests a more sophisticated use of sub-divisions.). However, this methodology is problematic in comparing languages belonging to different trees. Not only does the methodology always score LP as zero in these cases, but it assumes that 0.5 means the same in the Indo-European group as in the Altaic, Turkic one. LP2 overcomes this problem. It rests instead on the aforementioned ASPJ scoring of similarity between 200 words (sometimes 100) in a list (or two lists) that was (were) first compiled by Swadesh (1952). The members of the ASJP project have found that a selection of 40 of these words is fully adequate (See the list in Bakker et al., 2009).

We obtained our matrix of 89 by 88 linguistic distances from Dik Bakker (in October 2010), and decided to use the ASJP group's preferred measure which makes an adjustment for noise (the fact that words with identical meaning can resemble each other by chance). The adjusted series go from 0 to 105 rather than 0 to 1. So we multiplied all the data by 100/105 to normalize the data at 0 to 100. The original series also signify linguistic distance instead of linguistic proximity, while we prefer the latter, if nothing else because we want all the expected signs of the linguistic variables in the estimates to be the same. Therefore, we took the reciprocal of each figure and we multiplied it by the lowest number in the original series (9.92 for Serbo-Croatian and Croatian, or the 2 closest languages in the series). This then inverted the order of the numbers without touching the sign while converting the series from 0–100 to 0–1.

Once we had our two respective 89 by 88 bilateral matrices for linguistic proximity by language (following the aforementioned adjustments for the ASJP matrix), we needed to convert the two into country by country matrices. This was no mean task since it required the consideration of 195 countries; but it did not demand any further research.[14] LP1 and LP2 followed from the conversion. In a final step, we normalized both series once more so that their averages for *the positive values of LP2* in our sample estimates would equal exactly 1. This last normalization makes the estimated values of their coefficients exactly comparable to one another and exactly comparable to the coefficient of COL. Making the coefficients of LP comparable to those of COL makes sense since both variables concern translation. The normalization also means that individual values of LP1 and LP2 now go from 0 to more than 1.

3.4. Bilateral trade and distance

We turn next to the rest of the variables that enter into our gravity equation and begin with bilateral trade and distance. Our source for bilateral trade is the BACI database of CEPII, which corrects for various inconsistencies (see Gaulier and Zignago, 2010). The series concerns 224 countries in 1998 to 2007 inclusively, of which 29 (mostly tiny islands) drop out because of missing information on religion, legal framework and/or the share of native and spoken languages. Eventually, we also dropped all observations that do not fit into Rauch's tripartite classification (as the BACI database permits us to do). This last limitation meant losing only a minor additional percentage of the remaining observations, less than 0.5 of 1%. Our measure of distance rests on the 2 most populated cities and comes from the CEPII database as well.

3.5. The controls

The controls in the gravity equation demand our attention next. Both of our colonial variables come from Head et al. (2010). For common legal system, we went to JuriGlobe, which classifies legal systems worldwide between Civil Law, Common Law, Muslim law and Customary Law and indicates instances of mixed systems (mixes of the 4). Then we assigned 1 to all country pairs that shared Civil law, Common law or Muslim law and 0 to all the rest. Thus, we treated all countries with either Customary Law or a mixed legal system as not sharing a legal system with anyone.

With respect to common religion, our starting point was the CIA *World Factbook*, which reports population shares for Buddhist, Christian, Hindu, Jewish and Muslim, and a residual population share of "atheists." Next, we broke down the Christian and Muslim shares into finer distinctions. For Christians, we distinguished between Roman Catholic, Catholic Orthodox, and Protestants, as the CIA *Factbook* allows except for 15 countries in our sample, mostly African ones and also China. In these cases, we retrieved the added information either from the International Religious Freedom Report (2007) or the World Christian Database (2005). For Muslim, we distinguished between Shia and Sunni. To do so, we used the Pew Forum (2009) whenever the CIA *Factbook* did not suffice. In order to construct common religion in the final step, we went ahead exactly as we had for CNL and summed the products of population shares with the same religion. Ours is a more detailed measure of common religion than we have seen elsewhere.[15]

As regards the years at war since 1823, we relied on the Correlates of War Project (COW, v4.0), the data for which is available at http://www.correlatesofwar.org/ and goes up to 2003. This meant identifying former states of Germany with Germany, identifying the Kingdom of Naples and Sicily with Italy, and substituting Russia for USSR. The series for the number of years at war goes from 0 to 17.

For the stock of migrants, we utilized the World Bank International Bilateral Migration Stock database which is available for 226 countries and territories. The database is described in detail in Parsons et al. (2007).

[13] The 31 countries to which we assigned zeros notably include India (where linguistic diversity scores 0.94 out of 1). The other examples are mostly African ones: South Africa is an outstanding case. Following this exercise, the 89 languages we have to deal with exclude 5 of the 42 CSL languages (Fang, Fulfulde, Hausa, Lingala and Urdu) for various reasons (an insufficient percentage of native speakers, excessive linguistic diversity or both).

[14] Basically, for each country pair, we had either 1, 2 or 4 linguistic proximities to consider. When there were 2 or 4, we needed to construct an appropriate weighted average, which we based on the products of the population ratios in both countries. Remember that a LP of 0 between 2 countries can mean *either* that the 2 countries speak the same language – and therefore LP is irrelevant – or that their languages are so different that there is no proximity between them.

[15] There are two recent studies that analyze the effects of adherence to different major world religions (e.g., Muslim) on bilateral trade and that contain some sophisticated measures of common religion as well: Helble (2007) and Lewer and Van den Berg (2007). In both articles, the authors control for common language with a binary variable (based on one of the usual sources, the popular Havemann website in Helble's case, the CIA *Factbook* in Lewer and Van den Berg's).

4. The econometric form

We estimate cross-sections in the individual years 1998 through 2007 with country fixed effects and present a panel estimate over the ten years with country-year fixed effects as a basic summary. After log-linearizing Eq. (1) (following substitution of Eq. (2) for t_{ij}), the form for the individual-year cross-sections is:

$$\begin{aligned}\log M_{ij} = {} & \alpha_o + \delta_c Z_c + \alpha_1 COL_{ij} + \alpha_2 CSL_{ij} + \alpha_3 CNL_{ij} + \alpha_4 LP_{ij} \\ & + \alpha_5 \log Dist + \alpha_6 Adjacency_{ij} + \alpha_7 Excol_{ij} + \alpha_8 Comcol_{ij} \\ & + \alpha_9 Comleg_{ij} + \alpha_{10} Comrel_{ij} + \alpha_{11} Histwars_{ij} + \varepsilon_{ij}.\end{aligned}$$

α_o is a constant that encompasses Y_W. $\delta_c Z_c$ is a set of country fixed effects which will reflect all country-specific unobserved characteristics in addition to Y_i, Y_j, P_i and P_j. δ_c represents the effects themselves while Z_c is a vector of indicator variables (one per country) where Z_c equals one if $c = i$ or j and is 0 otherwise. The coefficients α_i, $i = 1, \ldots, 11$, are products of separate bilateral influences on t_{ij}, on the one hand, and $1 - \beta$, on the other, where $1 - \beta$ is the common negative effect of the elasticity of substitution between goods (since $\beta > 1$). The disturbance term, ε_{ij}, is assumed to be log-normally distributed.

As a result of the logarithmic specification, we lose all observations of zero bilateral trade. The principal problem with this elimination of the zeros is a possible selection bias. Imagine that linguistic factors had no role in explaining the cases of the zeros and operated only in the instances of positive trade. Then we might find important linguistic influences in our estimates strictly because of our automatic dropping of the zeros resulting from our choice of equation form. We focus on this issue in a subsequent section.

There are some instances of zero trade in one direction but not the other in our sample. Except for these cases, we have two separate positive observations for imports by individual country pair. Therefore we adjust the standard errors upward for clustering by country pairs in the panel estimates.

5. The results for total trade

We turn to the results and begin with the correlation matrix for the separate COL, CSL, CNL and LP series over the 209,276 observations in 1998–2007 in the panel estimates (The matrices for the individual years can only differ because of minor sample differences and they are virtually identical.). As seen from Table 2, the correlation between COL and either CSL or CNL is well below 1 and only moderately above 0.5. The outstanding reason is that there are many countries where domestic linguistic diversity is high and the official language (or both of them if there are 2) is (are) not widely spoken. In addition, the correlation between CSL and CNL is only 0.68 and significantly below 1. In this case the reason is that European languages and Arabic are important as second languages in the world, especially English. LP1 (language tree) and LP2 (ASJP) are highly correlated with one another at 0.84, just as we would expect. They are also both moderately negatively correlated with CNL and positively correlated with CSL. Their negative correlation with CNL is probably due essentially to the fact that their positive values depend on positive values of 1 — CNL. Their positive – and more interesting – correlation with CSL probably reflects the fact that higher values of either make a foreign language easier to learn. If we put the two previous opposite correlations together, we can deduce from Table 2 that there is a 0.25 positive correlation between spoken non-native languages and LP1 and a 0.28 positive correlation between spoken non-native languages and LP2.

In the first 3 columns of Table 3 we show what happens when we introduce COL, CSL or CNL alternatively by itself. Each of the three performs extremely well. But the coefficient of COL is substantially lower than the other two. In addition, since CSL incorporates CNL and we can hardly suppose that a common learned second-language damages bilateral trade, the lower coefficient of CSL than CNL probably results from simultaneity bias. Column 4 of Table 3 proceeds to include COL, CSL and CNL all at once. The coefficients of the 3 notably drop below their earlier values in columns 1–3, a clear indication that each variable, if standing alone, partly reflects the other 2. However, while COL and CSL remain extremely important in column 5, CNL becomes totally insignificant. Instead of pausing on this last result, let us move on to columns 5 and 6 where we introduce LP1 and LP2 as alternatives. Both indicators of LP have identical coefficients of 0.07/0.08 and both are precisely estimated, LP1 more so than LP2. However, when either indicator is present, the coefficient of CNL rises and becomes significant at the 95% confidence level. On this evidence, the importance of native language only emerges once we recognize gradations in linguistic proximity between different native languages and we cease to suppose a sharp cleavage between the presence and absence of a CNL. In addition, based on columns 5 and 6, all four aspects of common language appear as simultaneously important. Furthermore, the importance of spoken language clearly dominates that of native language. Last, official status matters independently of anything else.

For the remainder of our study, we will stick to LP2 even though the estimate of LP1 is more precise than LP2 in Table 3. This greater precision is not robust. In earlier experiments with minor differences in the sample, we found the relative precision of LP1 and LP2 to vary and to go sometimes in favor of LP2. Fundamentally, LP2 seems to us better founded and a better basis for reasoning and our later experiments. We shall skip the discussion of column 7 until an appropriate later point. All of these results for language emerge clearly in the individual years. The only notable difference is that the performance of CNL in combination with the other linguistic variables (columns (5) and (6)) is uneven (as the online Appendix B and the earlier working paper version show).

Of some interest as well, common religion, common legal system and history of wars are all significant and with the expected signs both in the full sample and in the individual years. Their coefficients are also fairly stable from year to year. There may be some qualification for history of wars, but that is all.

6. The zeros for trade

One possible problem in our study, as indicated before, is selection bias. Suppose that the influence of language in our estimates depended on our automatic exclusion of the zeros through our choice of a log-linear specification. A popular way to deal with this problem since Santos Silva and Tenreyro (2006) is Poisson pseudo maximum likelihood (PPML). In a detailed discussion of PPML, Head and Mayer

Table 2
Correlation Table (195 countries and 209,276 observations).

	Common official language	Common spoken language	Common native language	Linguistic proximity (tree)	Linguistic proximity (ASPJ)
Common official language	1.0000				
Common spoken language	0.5587	1.0000			
Common native language	0.5399	0.6791	1.0000		
Linguistic proximity (tree)	−0.1634	0.1489	−0.0980	1.0000	
Linguistic proximity (ASPJ)	−0.2284	0.1173	−0.1586	0.8384	1.0000

Table 3
Common language. Regressand: log of bilateral trade (Total).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Common official language	0.514			0.316	0.360	0.351	0.431
	(13.518)			(6.864)	(7.716)	(7.561)	(9.740)
Common spoken language		0.775		0.503	0.399	0.396	
		(14.651)		(6.578)	(5.104)	(4.910)	
Common native language			0.856	0.062	0.294	0.284	0.639
			(11.227)	(0.573)	(2.588)	(2.344)	(6.755)
Linguistic proximity (Tree)					0.073		
					(6.170)		
Linguistic proximity (ASJP)						0.078	0.105
						(4.253)	(6.048)
Distance (log)	−1.394	−1.379	−1.385	−1.375	−1.364	−1.365	−1.366
	(−90.272)	(−87.949)	(−88.075)	(−87.679)	(−86.392)	(−86.420)	(−86.458)
Contiguity	0.722	0.671	0.719	0.679	0.662	0.670	0.690
	(8.413)	(7.766)	(8.345)	(7.885)	(7.723)	(7.817)	(8.077)
Ex colonizer/colony	1.484	1.579	1.653	1.472	1.500	1.484	1.501
	(14.347)	(15.297)	(15.757)	(14.329)	(14.588)	(14.426)	(14.506)
Common colonizer	0.754	0.851	0.909	0.780	0.775	0.779	0.785
	(16.687)	(19.461)	(20.636)	(17.085)	(16.957)	(17.045)	(17.102)
Common religion	0.429	0.329	0.416	0.325	0.264	0.289	0.319
	(8.664)	(6.475)	(8.293)	(6.383)	(5.087)	(5.589)	(6.210)
Common legal system	0.244	0.311	0.274	0.240	0.209	0.217	0.189
	(6.817)	(9.029)	(7.695)	(6.544)	(5.666)	(5.866)	(5.202)
History of wars	−0.398	−0.417	−0.385	−0.397	−0.382	−0.382	−0.365
	(−2.388)	(−2.501)	(−2.357)	(−2.382)	(−2.272)	(−2.283)	(−2.188)
Observations	209,276	209,276	209,276	209,276	209,276	209,276	209,276
Adjusted R^2	0.756	0.756	0.756	0.757	0.757	0.757	0.757

All regressions contain exporter/year and importer/year fixed effects. Student *ts* are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair.

(2013) propose two varieties as well: gamma PML and multinomial PML. We tried all three by adding the zeros for all country pairs appearing in our previous panel estimates. This yields 80,224 additional observations constituting around 0.28 of the new total. The results with gamma PML and multinomial PML indicate no selection bias, whereas those with ordinary PPML leave the issue open. When COL, CSL and CNL serve separately, as in columns (1), (2) and (3) of Table 3, the three are all significant for all 10 individual years with gamma or multinomial PML (all but once at the 99% confidence level except for the combination of CNL and multinomial PML when the significance is sometimes only at the 90% confidence level). With ordinary PPML, however, COL is never significant, CSL is so only 2 years out of 10 at the 90% confidence level and CNL is so 7 years out of 10 at the same confidence level. When COL, CSL and CNL serve together along with LP2, as in column (6) of Table 3, gamma PML continues to yield good results for the linguistic variables, multinomial PML does not do as well as before but still tolerably: CSL continues to matter for all individual years while LP2 does so too. The results with ordinary PPML become even poorer than before.

It should be added, though, that ordinary PPML yields other problems. Not only do the linguistic variables cease to matter when it serves but so do both colonization variables and common religion while the significance of the history of wars becomes hap-hazard. On the other hand, the results with gamma PML correspond well to those in column (6) of Table 3 not only for language, but for the other variables, though they are notably less stable than the corresponding results for OLS from year to year and therefore less reliable (as shown in online Appendix B). The two colonization variables, common legal system and history of wars all remain significant with the same signs and orders of magnitude as before (to say nothing of distance and contiguity, which are always significant whatever the estimation method). Only common religion performs worse with some opposite and significant signs. Based on the results for gamma PML in particular we rule out selection bias.[16]

7. The results for the Rauch classification

We shall next try to exploit the Rauch decomposition of bilateral trade between homogeneous goods, listed goods and differentiated goods in Table 4 (Rauch, 1999). Homogeneous goods are quoted on organized exchanges and consist entirely of primary products like corn, oil, wheat, etc. Listed goods are not quoted on organized exchanges yet are still standard enough to be bought on the basis of price lists *without knowledge of the particular supplier*. Examples are many standardized sorts or grades of fertilizers, chemicals, and (certain) wired rods or plates of iron and steel.[17] In the case of differentiated goods, the purchaser buys from a specific supplier. Illustrations are automobiles, consumers' apparel, toys or cookware. Evidently we expect linguistic influences to become progressively more important as we go from homogeneous to listed to differentiated goods since the required information rises in this direction. For the same reason, we expect ethnic ties and trust to be more important as we move that way. The results for the three different categories support our hypotheses broadly; but there are some gray areas that we will not cover up.

The first column in Table 4 simply repeats the results in Table 3, column 6, for convenience. The next one provides the results for homogeneous goods. In this case, we omit CNL. If CNL serves as the sole linguistic variable (in estimates that we do not show), it is insignificant in half the individual years and has a low coefficient in the panel estimate over the period as a whole. Thus, it seems unimportant. However, when introduced jointly with CSL, the joint effect of CSL and CNL stays about the same but the coefficient of CSL rises and that of CNL turns negative in compensation, sometimes significantly so. It is difficult to make any sense of this last result. Furthermore, except for the change in the coefficient of CSL, CNL's absence has no effect on the rest of the estimate. This explains why we drop CNL. Following, the results can be read as suggesting that language is essentially important in conveying information — indeed so much so that the importance of language does not even require any public support through official status. COL is insignificant. The insignificance of common religion conforms broadly. It accords with the idea that the role of language owes little to personal affinities

[16] All of the results in this Section, beyond those in online Appendix B (concerning gamma PML), are available on request.

[17] We use Rauch's conservative definition of the classifications.

Table 4
Rauch categories. Regressand: log of bilateral trade.

	Total trade (1)	Homogeneous goods (2)	Listed goods (3)	Differentiated goods (4)
Common official language	0.351	0.027	0.193	0.420
	(7.561)	(0.404)	(3.581)	(9.298)
Common spoken language	0.396	0.676	0.643	0.453
	(4.910)	(7.037)	(7.076)	(5.812)
Common native language	0.284		0.052	0.248
	(2.344)		(0.389)	(2.056)
Linguistic proximity (ASJP)	0.078	0.097	0.096	0.055
	(4.253)	(3.968)	(4.545)	(2.984)
Distance (log)	−1.365	−1.189	−1.409	−1.409
	(−86.420)	(−51.295)	(−79.948)	(−90.849)
Contiguity	0.670	0.670	0.746	0.761
	(7.817)	(7.376)	(8.644)	(8.951)
Ex colonizer/colony	1.484	1.453	1.329	1.440
	(14.426)	(11.510)	(12.102)	(13.971)
Common colonizer	0.779	0.550	0.837	0.813
	(17.045)	(8.086)	(15.949)	(18.177)
Common religion	0.289	0.026	0.231	0.311
	(5.589)	(0.328)	(3.889)	(6.164)
Common legal system	0.217	0.474	0.223	0.020
	(5.866)	(8.401)	(5.398)	(0.555)
History of wars	−0.382	0.510	0.305	0.128
	(−2.283)	(2.673)	(1.795)	(0.760)
Observations	209,276	118,377	157,581	195,163
Adjusted R^2	0.757	0.576	0.710	0.782

All regressions contain exporter/year and importer/year fixed effects. Student *ts* are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair.

and trust. The main discomfort with this interpretation is the significance of LP, which only fits if LP can be regarded as reflecting strictly ease of translation or almost so. In that case, everything still hangs together and the results say that the importance of language for trade in homogeneous goods depends essentially on direct communication and ease of translation in a decentralized manner and without public support.

In the case of listed goods, CNL is not significant either but keeping it in the analysis raises no problem. CSL is not affected either way. COL, LP and common religion, as well as CSL, also retain the same coefficients regardless. They are all highly significant. The importance of COL in the presence of CSL and LP means that the support of translation through government auspices now matters. The relevance of religious ties is the only problematic aspect. If religious ties matter, why does CNL not matter as well? Perhaps the importance of religious ties may also be regarded as a sign that the significance of LP partly reflects ethnic rapport and trust rather than strictly ease of communication through translation.

In the case of differentiated goods, the coefficient of COL is both significant and almost as large as that of CSL. Translation is clearly important. For the first time, the significance of CNL is also difficult to deny even though CNL is not important every single year. However, we encountered various signs in our work that the significance of CSL and CNL are partly confused in the Rauch decomposition for differentiated goods. We accept its significance.[18]

The results for common legal system and history of wars in Table 4 are also interesting. Common legal system has a coefficient of 0.47 for homogeneous goods, a much lower coefficient of 0.22 which is still highly significant for listed goods, and a totally insignificant coefficient for differentiated goods. This would suggest some substitution between reliance on similar law and investment in information. Specifically, when little information is required, as for homogeneous goods, there is heavy reliance on similar law and when lots of information is required, there is enough investment in information to make similar law irrelevant. Note, finally, that the history of wars ceases to be uniformly significant and always bears the wrong sign when bilateral trade is divided by Rauch classification.

8. A proposed aggregate index of a common language

Is it possible to summarize the evidence about the linguistic influences in an index resting strictly on exogenous linguistic factors? That would be highly useful since we have many occasions to wish to control for such factors when our interest lies elsewhere. Moreover, on these occasions we sometimes work with small country samples when separate identification of several linguistic series may be extremely difficult. The answer to the question is yes. In other words, if we merely want to control for language in studying something else, a summary index of a common language can rest on COL, CNL and LP alone. Let us first go back to the last column of Table 3 where we drop CSL. As seen, the sum of the influences of COL, CNL and LP in this column stays about the same as the sum of those of COL, CNL, LP plus CSL in the previous column (it rises moderately). Thus, whatever contribution spoken language makes to the explanation of bilateral trade in column 6 of Table 3 is still present in column 7. Of course, it also follows that the coefficient of CNL in column 7 represents largely, if not predominantly, the role of spoken rather than native language.

We may then construct a 0–1 index of common language based on COL, CNL and LP. To do so, we decided to privilege CNL and strictly normalize COL + LP2, which we did by dividing the series by its highest value and next multiplying it by 1 − CNL. (Remember that LP2 had already been normalized to equal 1, like COL, at the sample mean of its positive values.) Then we equated common language with the sum of

[18] These results of the Rauch classification, taken as a whole, raise doubts about the view that a COL implies that everyone receives messages in an official language for free (as in Melitz, 2008). Far more significantly, they also give cause to think that CSL reflects translation as well as direct communication. LP is the clue in both cases. As regards COL, the results for homogeneous goods are central. The fact that LP matters for communicative ability whereas COL does not clearly does not agree with the idea that an official language means that all messages in the official language are available for free in one's own tongue (unless we also suppose that LP matters for all languages except official ones, which makes little sense). Consequently, even though we continue to consider the 0,1 character of COL to imply that there are no variable costs of receiving messages from an official language, we now recognize some private fixed cost of receiving the messages or getting "hooked up" in this (or these two) language(s). Next, and more importantly, Table 3 and the results in Table 4 when we remove LP clearly indicate that the introduction of LP reduces the coefficient of CSL (see online Appendix B). It does so not only for total trade but for all three Rauch categories separately (not shown). This would strongly suggest that CSL partly reflects bilingualism and translation and not only direct communication. COL and LP therefore are not alone in reflecting translation; CSL does so too.

Table 5
Common language index. Regressand: log of bilateral trade.

	Total trade (1)	Homogeneous goods (2)	Listed goods (3)	Differentiated goods (4)
Common language	1.153	0.676	1.051	1.237
	(14.468)	(5.595)	(11.986)	(15.642)
Distance (log)	−1.362	−1.208	−1.412	−1.406
	(−85.788)	(−52.175)	(−80.128)	(−89.967)
Contiguity	0.689	0.702	0.777	0.780
	(8.074)	(7.725)	(9.032)	(9.201)
Ex colonizer/colony	1.624	1.507	1.424	1.622
	(15.574)	(12.097)	(12.790)	(15.514)
Common colonizer	0.868	0.584	0.903	0.919
	(19.737)	(8.709)	(17.613)	(21.319)
Common religion	0.314	0.106	0.280	0.338
	(6.116)	(1.334)	(4.712)	(6.738)
Common legal system	0.225	0.444	0.187	0.039
	(6.275)	(7.804)	(4.626)	(1.092)
History of wars	−0.365	0.528	0.331	0.147
	(−2.196)	(2.795)	(1.969)	(0.875)
Observations	209,276	118,377	157,581	195,163
Adjusted R^2	0.756	0.575	0.710	0.781

All regressions contain exporter/year and importer/year fixed effects. Student *ts* are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair.

CNL and this normalized sum of COL + LP2, which is equal to 1 − CNL at most.[19] Table 5 provides the resulting panel estimates for the same gravity equation as before for total bilateral trade and for the 3 separate Rauch classifications. Based on column 1, the coefficient of this common language index is only slightly higher than the sum of the coefficients of COL, CNL and LP in column 6 of Table 3. It is about 1.15 and very precisely estimated. The separate coefficients of the index for homogeneous, listed and differentiated goods show up in the next 3 successive columns. They go from 0.68 to 1.05 to 1.24. All 3 are also precisely estimated, the coefficient for homogeneous goods less so than the other two. The rest of the equation is not affected by our aggregation of the linguistic influences in a single index. In particular, the earlier pattern of estimates of common religion, common legal system and history of wars occurs for the three Rauch classifications. Specifically, common religion is not significant for homogeneous goods but highly so for the other two classifications. Common legal system is highly significant for homogeneous goods, less so yet still highly significant for listed goods and no longer significant at all for heterogeneous goods. The coefficient of history of wars is small, significant and with the right sign for the aggregate, but partly insignificant and always with the wrong sign for the Rauch decomposition. The complete year by year estimates of the 4 panel estimates in Table 5 (available on request) indicate that the annual estimates of the coefficients of common language are quite stable. Only for homogeneous goods is there a large movement from year to year.

9. The role of cross-migrants

Thus far we have included no endogenous influences but CSL in the gravity equation. As mentioned earlier, however, one excluded endogenous influence notably alters the linguistic effects: namely, the stock of cross-migrants. Suppose we now add this variable. We shall consider the stock of migrants as exporters and importers separately. Thus, for example, in the case of French imports from Germany, the Germans immigrants into France (importers) and the French immigrants into Germany (exporters) figure separately. Both our measures concern the stock of migrants in the year 2000. Since we take logs and zeros can exist in either series, we lose about 15% of the observations.

Table 6 shows the effect of introducing both series of migrants in our fundamental econometric specification. In line with much earlier work on the subject of the role of migrants in trade between the host and home country, they prove extremely important (Gould, 1994; Head and Ries, 1998; Dunlevy and Hutchinson, 1999; Wagner et al., 2002; Rauch and Trindade, 2002).[20] As we can see, in the first column, concerning bilateral trade in the aggregate, both series for migration (log) enter with precisely estimated coefficients of 0.14. If we assume that the effects of migration in this estimate had been simply reflected before in the linguistic variables, the total linguistic effects are now the sum of the coefficients of the previous linguistic variables plus these two migration ones. Therefore, we get a total influence of language of .93, somewhat lower than before (1.1) but still high. Both migration variables in this estimate are endogenous as well as CSL. If we substitute our common language index for the 4 linguistic ones to correct, at least, for the endogeneity of CSL, as is shown in the next column, we get a total influence of .98, moderately below the previous estimate of 1.15.

Of note, the introduction of migrants renders CNL completely insignificant in the first column. The last 3 columns pursue the analysis further by admitting the Rauch decomposition of trade. Four points stand out. First, CNL is no longer significant even for differentiated goods. Next, migrants are significant for all 3 categories of goods, including homogeneous ones, where their effect is substantial (.22). Third, their influence rises steadily as we go from these to listed (.28) to differentiated goods (.31). Thus, while migrants increase trade across-the-board, their influence becomes more marked as we move toward goods that are more information-intensive. Last, even after taking into account the separate influence of migrants, COL still has a large and highly significant influence for differentiated goods, implying an important role of translation and indirect communication for these goods apart from ethnicity and trust, just as before.

[19] This is not the only way to proceed but it is a simple one. A more sophisticated way would be to take into account the differences in the accuracy of the estimates of COL, CNL and LP. Yet the simplicity of our method is a recommendation (as otherwise the aggregate becomes a function of the estimates). It is especially so since the accuracies of the separate estimates of COL, CNL and LP are broadly comparable.

[20] Of some note as well, the most recent literature on the relation between language and migration includes some attempts to use several measures of linguistic influence at once. See Belot and Ederveen (2012) and Adsera and Pytlikova (2011).

Table 6
Migration. Regressand: log of bilateral trade.

	Total trade (1)	Total trade (2)	Homogeneous goods (3)	Listed goods (4)	Differentiated goods (5)
Common official language	0.222		−0.032	0.106	0.301
	(4.522)		(−0.456)	(1.910)	(6.470)
Common spoken language	0.288		0.492	0.389	0.347
	(3.411)		(4.907)	(4.220)	(4.349)
Common native language	0.078			−0.028	0.051
	(0.633)			(−0.207)	(0.417)
Linguistic proximity (ASJP)	0.056		0.085	0.072	0.033
	(3.050)		(3.536)	(3.607)	(1.887)
Common language		0.694			
		(8.364)			
Migration: importers (log)	0.144	0.146	0.134	0.149	0.165
	(18.726)	(19.065)	(11.832)	(17.641)	(22.093)
Migration: exporters (log)	0.141	0.143	0.086	0.131	0.145
	(18.183)	(18.507)	(7.830)	(15.554)	(19.772)
Distance (log)	−1.081	−1.075	−1.002	−1.155	−1.108
	(−55.581)	(−55.152)	(−36.162)	(−54.970)	(−59.021)
Contiguity	0.131	0.137	0.274	0.253	0.170
	(1.548)	(1.617)	(2.973)	(2.976)	(2.045)
Ex colonizer/colony	0.878	0.953	1.017	0.708	0.770
	(9.408)	(10.317)	(8.097)	(7.042)	(8.287)
Common colonizer	0.628	0.687	0.349	0.622	0.646
	(12.531)	(14.136)	(4.770)	(11.121)	(13.455)
Common religion	0.167	0.178	−0.084	0.071	0.151
	(3.102)	(3.331)	(−1.034)	(1.172)	(2.891)
Common legal system	0.233	0.228	0.462	0.231	0.032
	(6.005)	(6.063)	(7.911)	(5.454)	(0.866)
History of wars	−0.667	−0.654	0.367	0.047	−0.188
	(−4.471)	(−4.427)	(1.995)	(0.307)	(−1.311)
Observations	176,884	176,884	109,06	140,366	166,809
Adjusted R^2	0.773	0.772	0.589	0.730	0.804

All regressions contain exporter/year and importer/year fixed effects. Student *ts* are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair.

10. English as a separate language

The analysis thus far supposes that the particular language makes no difference. Many would question this assumption, for English in particular. We therefore tested the separate importance of English. Since we did so, we looked at the other major world languages too, and we summarize the results in Table 7, where we concentrate on English. The first test, column 1, is purely expository. It treats English as the only common language. Suppose that all of our results depended on English alone (a view that we encountered). Then the measures of COL, CSL, CNL and LP2 in this first column would remove errors of measurement and yield higher and better estimated coefficients. Suppose instead that our measures of a common language are the correct ones. Then the measures of linguistic influence in this column would be noisy and yield lower and less well estimated coefficients than the previous ones. In fact, in this last case – that is, if our measures of a common language are the appropriate ones – there are two reasons why the English-based measures of the linguistic variables might perform particularly badly. In the first place, an English-speaking country has a great many solutions for skirting the language barrier altogether. There are lots of other English-speaking countries with which it could trade. Therefore, common English can be expected to be an especially weak spur to trade with any single common-language partner. Alternatively, a country speaking Portuguese, for example, would have far fewer alternative partners with which to trade in order to avoid the language barrier and therefore might exploit those opportunities more intensely.[21] This is, of course, the identical point Anderson and van Wincoop (2003) make in explaining why national trade barriers form a far more powerful incentive for bilateral trade between two Canadian provinces than between two US states. On this ground, the coefficients of the linguistic variables based on English alone might be exceptionally low apart from measurement error. The second problem could be equally serious. Relying on English alone means drawing numerous distinctions between country pairs who share a common language other than English based upon their English, and proposing a quantitative ordering of linguistic ties between these non-English pairs based on their common English alone. Especially large distortions might arise.

The results in column 1 basically confirm the broad suspicion that measures of a common language resting on English alone would perform badly. COL, CSL and CNL for English are insignificant. Yet it is true that LP2 matters for English, a point to which we will return.

Column 2 is the genuine test. It examines whether adding separate measures of a common language for English to the earlier measures in the tests supports a separate consideration of English. In this case, the results are entirely negative for COL, CSL and CNL. For all 3 measures, the sign of a common language without any separate notice of English and the one based on English alone go in opposite directions (the signs of COL and CSL becoming significantly negative for English). There is no sense in this. Given the high quality of the results for the linguistic variables in the absence of special attention to English, the only inference is that the separate consideration of the language is unfounded.[22] However, as regards LP2, English is still separately significant in column 2.

The similar tests for the 3 next largest languages in our database – French, Spanish and Arabic – yield similar results. As a summary indication, column 3 presents the results of a similar test to the previous one

[21] Of course, for that very reason, people in the Portuguese-speaking country would have stronger incentives to become multilingual. But this diminishes the weight of the point without denying it altogether. Note also that the higher multilateral trade barrier the Portuguese-speaking country faces because of language is independently captured by our country fixed effects.

[22] These last results are reminiscent of those we obtained when we introduced CNL together with CSL for homogeneous goods. In this case too the signs of CNL and CSL went in opposite directions (the sign of CNL becoming significantly negative) and we drew the same (or the corresponding) inference that CNL should not be introduced jointly with CSL.

Table 7
English as a separate common language. Regressand: log of bilateral trade (Total).

	(1)	(2)	(3)
Common official language		0.405	0.233
		(5.643)	(4.198)
Common spoken language		1.244	0.439
		(8.545)	(4.903)
Common native language		−0.379	0.350
		(−2.240)	(2.463)
Linguistic proximity (ASJP)		0.060	0.115
		(2.892)	(5.053)
Common official language: English or (only column 3) other major European	0.084	−0.237	0.449
	(1.416)	(−2.658)	(4.807)
Common spoken language: English or (only column 3) other major European	−0.034	−1.447	−0.656
	(−0.344)	(−8.377)	(−3.164)
Common native language: English or (only column 3) other major European	−0.001	0.763	0.085
	(−0.007)	(3.173)	(0.349)
Linguistic proximity (ASJP): English or (only column 3) other major European	0.092	0.083	−0.075
	(2.887)	(2.316)	(−3.038)
Distance (log)	−1.418	−1.344	−1.369
	(−91.968)	(−83.993)	(−84.907)
Common border	0.749	0.622	0.654
	(8.694)	(7.206)	(7.646)
Ex colonizer/colony	1.742	1.445	1.451
	(16.223)	(14.446)	(13.980)
Common colonizer	0.884	0.758	0.755
	(19.627)	(16.628)	(16.459)
Common religion	0.533	0.241	0.326
	(10.695)	(4.644)	(6.242)
Common legal system	0.422	0.338	0.267
	(10.427)	(8.172)	(6.954)
History of wars	−0.437	−0.402	−0.388
	(−2.615)	(−2.426)	(−2.336)
Observations	209,276	209,276	209,276
Adjusted R^2	0.755	0.758	0.757

All regressions contain exporter/year and importer/year fixed effects. Student *t*s are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair.

for English in column 2 that lumps together the major West and Central European world languages besides English, namely, French, Spanish, German and Portuguese. Quite specifically, the measures of COL, CSL and CNL for these 4 languages in column 3 follow from our method of construction after setting all values for all common languages in our database except these 4 equal to zero and recognizing no linguistic distances LP2 except to these 4. As can be seen, broadly speaking, this alternative set of languages as a group yields no better results than English does (though in the case of COL the combined measure does do better than English, as is true for French and Spanish separately). We also find, rather uncomfortably, that linguistic proximity harms bilateral trade for this combination of languages, which is possibly simply a reflection of the earlier result that native English helps exceptionally since English figures prominently in the separate measure of LP2 in the same estimate (whose effect is now correspondingly higher). In other estimates for individual languages, we also find that LP2 helps to interpret foreign languages for Spanish and is harmful for French and Arabic. All these results about the significance of separate native languages in interpreting foreign languages based on linguistic proximity remain a mystery to us.

With this caveat, we conclude that the distinction of English, or any other major language for that matter, is not warranted. Once we control for distance, contiguity, ex-colonialism, law, religion, the history of wars, and country/year fixed effects or multilateral trade resistance, all that really matters is a common language, whatever the language may be.

11. Discussion and conclusion

It is common practice in the trade literature to use a binary 0, 1 variable to control for a common language. We have shown that this practice takes us way off the mark in estimating the impact of linguistic factors on bilateral trade. Probably the most clear-cut basis for answering yes or no to the presence of a common language is a COL. Country samples of any size where, even as a rough approximation, every individual in all pairs has the same native language or else no one in all pairs shares a native language with anyone in the opposite country are either imaginary or highly unlikely. Yet it is precisely when official status serves as the basis for a dummy variable for a common language that the underestimate of common language is greatest, in the order of one-half in terms of semi-elasticities (0.5 instead of 1.1/1.15) (In terms of elasticities, the underestimate is higher, closer to one-third: $\exp(0.5) - 1 \cong 0.6$ and $\exp(1.1) - 1 \cong 2$).

In sum, there is no way to embrace the influence of language on bilateral trade by using a measure of common language along any single dimension. Only a measure embracing a broad range of the linguistic influences on bilateral trade will do. One source of linguistic influence that sometimes gets primary attention is ethnic ties. This is particularly true in studies that center on emigrants (e.g., Rauch and Trindade, 2002). Admittedly, the linguistic influences on trade stemming from immigrants probably owe much to ethnicity and trust. However, ease of communication is also of sizable importance.

In the first place, CSL is always significant in the presence of CNL though the opposite is not true. Therefore, ability to speak, as such, makes a difference, apart from native speech. In addition, COL always matters in the presence of CSL and CNL, which further says that institutional support for translating a language that parts of the population do not understand into one that they do makes a difference too.[23] Quite significantly, the separate importance of CSL and COL in the presence of

[23] Of considerable note, though, interpreters and translation are probably far less effective in production within a firm than in trade. Labor studies show a substantial positive return to command of the principal home language on the wages of immigrants. See, McManus et al. (1983), Chiswick and Miller (1995, 2002, 2007), Dustmann and van Soest (2002), and Dustmann and Fabbri (2003). We would conjecture that the wage return would be lower if translation and interpreters were as effective in production as they are in trade.

CNL holds for differentiated goods, where the information problem is greatest and the network effects, culture and personal affinities can be expected to be especially important. Ease of communication also helps to see why language matters for homogeneous goods. In this case, the required information is so small that we might even have expected linguistic barriers to pose no problems at all. However, the highly significant coefficient of the common language index of 0.68 (Table 5, column 2) in the relevant estimate disproves the hunch. Upon reflection, the ability to communicate in depth helps one to understand why. This ability is never irrelevant in trade since things can go wrong. Goods may arrive late or damaged; contracts may not be honored; there may need to be recourse to the small print. It is pertinent in this connection that a common legal system matters as well for homogeneous goods. It enters significantly with a semi-elasticity of 0.44, not that far below 0.68, whereas common religion is irrelevant. True, migration is also significant when it is admitted for these goods with a semi-elasticity of influence of 0.22 (Table 6, column 3, without correction for endogeneity). But this impact of migrants need not be independent of ease of communication.

Other evidence can be brought to bear. Surveys of firms engaged in foreign trade show that they are much concerned with communication skills without any indication of special preoccupation with native speech. In 2005 the European Commission financed a large study on foreign language skills in business from CILT, a British organization focusing on the subject: Hagen et al. (2006). The study covered a sample of 2000 small and medium-sized exporting enterprises (SMEs) in 29 European countries, including Turkey, and 30 large multinationals (MNEs), all home-based in France (see Annex 4 of the study). To the question, has your company undertaken foreign language training in the last 3 years, 35% of the 2000 SMEs answer yes. This is of course investment that is entirely devoted to improving communication skills of non-native speakers. The percentage of these 2000 firms foreseeing a need to acquire additional expertise in foreign languages in the next 3 years is higher, 42% (see also Bel Habib, 2011). If anything, the MNEs are more conscious of the importance of investing in linguistic skills than the SMEs without any evidence of a greater focus on native speech. 60% of the MNEs recognize deficiencies. This is below the 75% figure in a previous study of similar inspiration covering 151 multinationals with a broader international distribution of home bases, including the UK, Germany and a sprinkling of other countries besides France. See Feely and Winslow (2005), another CILT publication. The composition of the languages that these firms expect to require is also interesting. In the case of the SMEs, some small languages are in the top 10: Czech (5%), Danish (3%) and Estonian (3%). The MNEs are more heavily interested in English, which is easy to interpret since the MNEs can be expected to be more sensitive to multilateral trade resistance and learning English will reduce this variable or P_iP_j in Eq. (1) more than learning any other language. These firms even face a language problem internally. Notably, however, English is not their only concern. Their wider international interests also show up in a greater emphasis than the SMEs' on other languages with importance over large geographical surfaces and covering many countries in different parts of the earth like Spanish and Arabic.

A recent study by Egger and Lassmann (2013) is also to the point. These authors study a multilingual sample of people possessing German, French and Italian in Switzerland and are therefore able to distinguish trades between partners possessing the same native language and trades between partners possessing a common language that differs from their native one. They find that native language as such – thus apart from the ability to communicate – has a semi-elasticity of influence on bilateral trade of around 0.3. As they observe, this is well below the usual estimates of total linguistic effects, which, we would add, rest on official language alone.

As regards future research along our lines, crossing language barriers may be viewed as a separate topic as distinct from crossing national barriers. We know that only a small minority of firms export to as many as 5 foreign destinations and that these firms are unusually big and efficient (see Bernard et al., 2007; Eaton et al., 2011; Mayer and Ottaviano, 2007). However, usual evidence does not tell us whether these firms also share a common language with all their foreign sales destinations, while if they do they might still be less efficient than other firms that cross a language frontier. Mayer and Ottaviano (2007) provide evidence that this is so. They show, for France, that the percentage of individual firms who export to other French-speaking destinations is unusually large but also that the firms who exploit this linguistic advantage have lower average productivity than the rest of French exporting firms. In further work, it may also prove important to distinguish between fixed and variable costs, as we have not done. If fixed costs are important, foreign sales across 5 language barriers require more efficiency to be profitable than equal foreign sales across a single language barrier. In the case of language, the variable and fixed costs are also easy to interpret. The variable costs refer basically to hiring interpreters and buying translations. The fixed costs refer instead to hiring natives or others with linguistic skills or else providing language training to existing staff. All of these different aspects of linguistic policy feature in the survey evidence. Their relative importance may be of interest.

There may also be policy implications of our study about elementary education. Our results would say that foreign languages have a place in school curricula. But how large a place? At least in the UK, prominent voices have already been heard to say that this place is larger than the one that foreign language study is accorded (see the Nuffield Report, 2000; The British Chambers of Commerce, 2003–2004). These sources clearly manifest a particular British concern with the lower levels of foreign language training in elementary schools and language proficiency among adults in the UK than the rest of the EU (besides Ireland). The sources also assume that the impact of foreign languages on trade depends largely on communication. If ethnicity was the fundamental issue instead, immigration policy would be more to the point than language training in schools.

Appendix A. Data sources for spoken and native language and language data.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jinteco.2014.04.004.

Appendix B. Supplementary annual estimates of the baseline equation.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jinteco.2014.04.004.

References

Adsera, A., Pytlikova, M., 2011. The role of language in shaping international migration: Evidence from OECD countries 1985–2006. Mimeo.
Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., 2003. Fractionalization. J. Econ. Growth 8, 155–194.
Anderson, J., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border problem. Am. Econ. Rev. 93, 170–192.
Armington, P., 1969. A theory of demand for products distinguished by place of production. Int. Monet. Fund Staff Pap. 16, 159–176.
Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C., Brown, P., Egorov, D., Mailhammer, R., Grant, A., Holman, E., 2009. Adding typology to lexicostatistics: a combined approach to language classification. Linguist. Typology 13, 167–179.
Bel Habib, I., 2011. Multilingual skills provide export benefits and better access to new emerging markets. Int. Web J. 10, 1–27 (www.sens-public.org).
Belot, M., Ederveen, S., 2012. Cultural and institutional barriers in migration between OECD countries. J. Popul. Econ. 25, 1077–1105.
Bernard, A., Jensen, J.-B., Redding, S., Schott, P., 2007. Firms in international trade. J. Econ. Perspect. 21, 105–130.
Brown, C., Holman, E., Wichmann, S., Velupillai, V., 2008. Automatic classification of the world's languages: a description of the method and preliminary results. Lang. Typology Univers. 61 (4), 285–308.
Central Intelligence Agency, World Factbook, US Government Printing Office, available online.

Chiswick, B., Miller, P., 1995. The endogeneity between language and earnings: international analyses. J. Labor Econ. 13, 246–248.
Chiswick, B., Miller, P., 2002. Immigrant earnings: language skills, linguistic concentration and the business cycle. J. Popul. Econ. 15 (1), 312–357.
Chiswick, B., Miller, P., 2007. Computer usage, destination language proficiency and the earnings of natives and immigrants. Rev. Econ. Househ. 5 (2), 129–157.
Desmet, K., Ortuño-Ortin, I., Weber, S., 2009a. Linguistic diversity and redistribution. J. Eur. Econ. Assoc. 7 (6), 1291–1318.
Desmet, K., Ortuño-Ortin, I., Wacziarg, R., 2009b. The political economy of ethnolinguistic cleavages. CEPR Discussion Paper No. 7478.
Dunlevy, J., Hutchinson, W., 1999. The impact of immigration on American import trade in the late nineteenth and early twentieth centuries. J. Econ. Hist. 59, 1043–1062.
Dustmann, C., Fabbri, F., 2003. Language proficiency and labour market performance of immigrants in the UK. Econ. J. 113 (489), 695–717.
Dustmann, C., van Soest, A., 2002. Language and the earnings of immigrants. Ind. Labor Relat. Rev. 55 (3), 473–492.
Eaton, J., Kortum, S., Kramarz, F., 2011. An anatomy of international trade: evidence from French firms. Econometrica 79, 1453–1498.
Egger, P., Lassmann, A., 2012. The language effect in international trade: a meta-analysis. Econ. Lett. 116 (2), 221–224.
Egger, P., Lassmann, A., 2013. The causal impact of common native language on international trade: evidence from a spatial regression discontinuity design. ETH Zurich Working Paper.
Ethnologue: Languages of the World, 16th ed. Summer Institute of Linguistics, International Academic Bookstore, Dallas, TX (available online).
Fearon, J., 2003. Ethnic and cultural diversity by country. J. Econ. Growth 8, 195–222.
Feely, A., Winslow, D., 2005. Talking sense. A Research Study of Language Skills Management in Major CompaniesCILT, the National Center for Languages, London, UK.
Felbermayr, G., Toubal, F., 2010. Cultural proximity and trade. Eur. Econ. Rev. 54, 279–293.
Frankel, J., Rose, A., 2002. An estimate of the effect of common currencies on trade and income. Q. J. Econ. 117, 437–466.
Gaulier, G., Zignago, S., 2010. BACI: international trade database at the product-level: the 1994–2007 version. CEPII Working Paper, 2010–23.
Ginsburgh, V., Weber, S., 2011. How Many Languages Do We Need? The Economics of Linguistic Diversity. Princeton University Press.
Giuliano, P., Spilimbergo, A., Tonon, G., 2006. Genetic, cultural and geographical distances. CEPR Discussion Paper 5807.
Gould, D., 1994. Immigrant links to the home country: empirical implications for US bilateral trade flows. Rev. Econ. Stat. 69, 301–316.
Guiso, L., Sapienza, P., Zingales, L., 2009. Cultural biases in economic exchange. Q. J. Econ. 124, 1095–1131.
Hagen, Stephan with Foreman-Peck, J., Davila-Philippon, S., Nordgren, B., Hagen, Susanna, 2006. ELAN: Effects on the European economy of shortages of foreign language skills in enterprise, CILT, The national center for languages, London, UK.
Head, K., Mayer, T., 2013. Gravity equations: workhorse, toolkit and cookbook. CEPR Discussion Paper No. 9322.
Head, K., Ries, J., 1998. Immigration and trade creation: econometric evidence from Canada. Can. J. Econ. 31, 46–62.
Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. J. Int. Econ. 81 (1), 1–14.
Helble, M., 2007. Is God good for trade? Kyklos 60 (3), 385–413.
Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: trading partners and trading volumes. Q. J. Econ. 123, 441–488.
International Religious Freedom, 2007. http://www.state.gov/g/drl/rls/irf/2007/index.htm.
Laitin, D., 2000. What is a language community? Am. J. Polit. Sci. 44, 142–155.
Lewer, J., van den Berg, H., 2007. Estimating the institutional and network effects of religious cultures on bilateral trade. Kyklos 60 (2), 255–277.
Martin, P., Mayer, T., Thoenig, M., 2008. Make trade not war? Rev. Econ. Stud. 75 (3), 865–900.
Mayer, T., Ottaviano, G., 2007. The happy few: the internationalisation of European firms: new facts based on firm-level evidence. Bruegel Blueprint Series, vol. III.
McManus, W., Gould, W., Welch, F., 1983. Earnings of Hispanic men: the role of English language proficiency. J. Labor Econ. 1, 101–130.
Melitz, J., 2008. Language and foreign trade. Eur. Econ. Rev. 52, 667–699.
Melitz, J., Toubal, F., 2012. Native language, spoken language, translation and trade. CEPR Discussion Paper 9994.
Nuffield Foundation, 2000. Languages: The Next Generation; The Final Report and Recommendations of the Nuffield Languages Inquiry (London, UK).
Parsons, C., Skeldon, R., Walmsley, T., Winters, A., 2007. Quantifying international migration: a database of bilateral migrant stocks. World Bank Policy Research Working Paper No. 4165.
Pew Research Center's Forum on Religion & Public Life, 2009. Mapping the Global Muslim Population: A Report on the Size and Distribution of the World's Muslim Population. The Pew Research Center (October).
Rauch, J., 1999. Networks versus markets in international trade. J. Int. Econ. 48, 7–35.
Rauch, J., Trindade, V., 2002. Ethnic Chinese networks in international trade. Rev. Econ. Stat. 84, 116–130.
Santos Silva, J.M.C., Tenreyro, S., 2006. The log of gravity. Rev. Econ. Stat. 88 (4), 641–658.
Sarkees, M.R., Wayman, F., 2010. Resort to War: 1816–2007. CQ Press.
Selmier, T., Oh, C. Hoon, 2012. The power of major trade languages in trade and foreign direct investment. Rev. Int. Polit. Econ. 1, 1–29.
Special Eurobarometer 243, 2006. Europeans and Their Languages. The European Commission.
Swadesh, M., 1952. Lexico-statistic dating of prehistoric ethnic contacts. Proc. Am. Philos. Soc. 96, 121–137.
The British Chambers of Commerce, 2003–2004. BBC Language Survey. The Impact of Foreign Languages on British Business — Part 1, 2003, Part 2, 2004. The British Chambers of Commerce, London, UK.
Wagner, D., Head, K., Ries, J., 2002. Immigration and the trade of provinces. Scot. J. Polit. Econ. 49, 507–525.
World Christian Database, 2005. www.worldchristiandatabase.org.