The Gravity of Ideas: How distance affects translations*

Isabelle Sin

Motu Economic and Public Policy Research

June 2012

Abstract

I use newly digitized data on over two million book translations for the period 1949-2000 to study the factors that affect the international diffusion of ideas carried by books. I find the elasticity of translations with respect to physical distance was between -0.3 and -0.5 in the 1990s, and higher in earlier decades when globalisation was less and international communication more difficult. Translations decrease in distance especially in less developed countries, suggesting that countries further from the world knowledge frontier, which face the greatest potential benefit from adopting ideas that already exist elsewhere, are actually less able to access these ideas. Results also suggest that a large population of bilingual speakers is a substitute for proximity to the original country in terms of facilitating translation flows. Finally, I augment a standard gravity model to include measures of linguistic, religious, and cultural distance. These all reduce translations, but together account for only a quarter of the correlation between translations and physical distance.

---

* An early version of this paper appeared in my dissertation, "Insights from Book Translations on the International Diffusion of Knowledge." I am grateful to my advisors, Ran Abramitzky, Avner Greif, Nick Bloom, Jim Fearon, and Kalina Manova for extremely helpful discussion and suggestions. I owe special thanks to the Index Translationum team, especially Alain Brion, Mauro Rosi, and Marius Tukaj for providing me with the translation data. I gratefully acknowledge financial support from the Ric Weiland Fellowship, the Rovensky Fellowship, and B.F. Haley and E.S. Shaw Fellowship.

1. Introduction

Economists and economic historians recognize the importance of knowledge and ideas for growth and development. (See, for example, Kuznets, 1966, Mokyr, 2002, 2009, 2010, and Romer, 1990, 1993.) Furthermore, the international sharing of ideas plays a huge role in growth over the long term: for instance, Klenow and Rodríguez-Clare (2005) estimate world GDP would be just 6% of its current level if countries did not share ideas. However, international idea flows have received relatively little attention in the empirical economics literature, largely because they are challenging to measure.

Translations of books are one measure of international flow of ideas, especially in the pre-internet era. Language barriers impede the spread of ideas stored in books, and translations are an important means by which such ideas diffuse across these barriers; flows of book translations are thus a tractable empirical measure of the flow of written ideas between linguistic groups. Book translations provide an attractive way to quantify idea flows because they are both non-rival and disembodied; they are a pure measure of idea flows rather than a by-product of a process such as trade or migration, and their key purpose is to make the ideas contained in the book accessible to speakers of another language.

The objective of this paper is to shed light on the determinants of translation flows between countries, in order to gain insight into the factors that encourage or inhibit the international diffusion of ideas. Although the ideas contained in books are non-rival, geographic and cultural distance and other types of dissimilarities between countries may inhibit translation flows, for instance by increasing search and transaction costs. By studying the relationship between measures of distance between countries and translation flows, I shed light on an important type of impediment to the free international diffusion of ideas.

I use data on the universe of translations published in a large number of countries for the period 1949 to 2000, sourced from Unesco's Index Translationum. Data for the period 1979-2000 are available in digital form; in addition, I digitized data for every fifth year from 1949 to 1974. The Index Translationum contains bibliographic information for each translation published in a participating country, including the country, city, and year in the which the translation was published, the language of the original title and the target language into which it was translated, the field of the title, the number of pages or volumes of the title, the author, and the original and translated titles of the book. I used these entries to construct data on the magnitude of translation flows between pairs of languages by publishing country, by field and over time.

I study the relationship between physical distance and translation flows within an augmented gravity framework, where the translation flow between a pair of countries varies with characteristics of the two countries (such as GDP) and the distance between them. Although translations have transportation costs that are close to zero, they may decrease with distance for a number of reasons. Supply frictions such as the search and information costs of identifying titles worth translating are likely to increase with distance, as are the various costs of negotiating contracts.[1] Translations may also fall off with distance because distance is correlated with tastes, meaning closer countries cater better than more distant countries to local tastes in books.[2]

Indeed, I find translation flows decrease significantly with distance, with an elasticity with respect to distance of between -0.3 and -0.5 during the 1990s. Note however that this elasticity is considerably smaller than the equivalent elasticity for trade found in the literature, which usually ranges from -1.08 to -1.24.[3]

With the increase in globalization and decrease in international communication and information costs that have occurred over time, we would expect the relationship between translations and distance to have decreased over time. I next examine how the elasticity of translations with respect to distance changed over the latter half of the twentieth century. I find the elasticity fell significantly over the period 1949-1999, especially in the last two decades. This result contrasts with the puzzling finding that the relationship between distance and trade in goods did not decrease over this period,

[1] That is, translations are subject to the costs that are related to forming a contract between parties in different countries. These costs may vary with distance, and include the time and legal costs of negotiating and enforcing the contract, direct and indirect costs related to transacting between currencies, and the costs of overcoming any language barriers that exist between the parties.

[2] Blum and Goldfarb (2006) suggest this factor plays a significant role in the distance effect for taste-dependent digital goods.

[3] Disdier and Head (2008).

instead remaining strong.[4] However, it is consistent with the finding that the effect of distance on patent citations has fallen over time.[5]

In comparisons across fields of the effect of distance on translations, I find the distance effect is larger for titles in exact and applied science than for titles in fields where taste ought to play a larger role, such as philosophy, arts and literature. The stronger correlation between distance and translations in fields where we expect taste to be less important suggests a greater importance for contracting or search and information costs relative to consumer tastes in driving the distance effect. These results also demonstrate that distance matters even for translations of "economically useful" titles (in the sense of Mokyr's (2002) "useful knowledge"), not just for titles that may be considered largely consumption goods.

One potential reason translations decrease with distance could be because physical distance is correlated with cultural distance, which may increase transaction costs or decrease the similarity of taste in books. To explore this possibility, I add controls for religious distance, linguistic distance, genetic distance, and survey measures of differences in cultural values, all of which are expected to capture some element of cultural differences. Religious and linguistic distance and Hofstede's (1980, 2001) survey measure of cultural distance are negatively related to translations, but physical distance remains important even when these controls are added. Furthermore, adding these controls reduces the elasticity of translations with respect to physical distance by at most a quarter. This suggests that cultural differences contribute to distance-varying transaction costs or to demand that prefers titles written in nearby countries, but that other distance-varying costs play a larger role.

The extent to which translations decrease with the distance may vary by translating country if some countries face higher distance-related transaction costs than others. I next allow the elasticity of translations with respect to distance to vary by the level of development of the translating country, as measured by GDP per capita or urbanization. I find a strong differential effect: the effect of distance is 89% weaker for a translating country on the 75th percentile of GDP per capita among countries in my data

[4] Disdier and Head (2008).
[5] MacGarvie (2005) and Griffith et al. (2007).

than for a country on the 25th percentile. This suggests that countries further from the world knowledge frontier, which face the greatest potential benefit from adopting ideas that already exist elsewhere, are actually less able to access these ideas.

In addition, I allow the elasticity of translations with respect to distance to vary with the proportion of the population in the translating country that speaks both the original and target languages. Bilingual speakers may facilitate translation flows by reducing transactions costs or providing a source of knowledge about foreign original titles.[6] I find that having a high proportion of bilingual speakers is a substitute for proximity to the original country: my results suggest that a one-standard-deviation increase in bilingualism (11% of the population) reduces the distance effect by 12 percent.

2. Translation data

In this section I describe the data on translations. Descriptions of the construction of the other variables can be found in the Data Appendix. The data on translations I use are derived from Unesco's Index Translationum (IT), an international bibliography of the translations published in a wide range of countries over the periods 1932 to 1940 and 1948 to the present. In the majority of cases, these bibliographical entries are acquired by Unesco from the central depository of the translating country, which, under the law of legal deposit, receives copies of every book published in the country and intended for circulation.[7]

Titles are categorized into fields according to the nine main categories of the Universal Decimal Classification (UDC) system: General; Philosophy (including Psychology); Religion and Theology; Law, Social Sciences, Education; Natural and Exact Sciences; Applied Sciences; Arts, Games, Sports; Literature (including books for children)[8]; History Geography, Biography (including memoirs and autobiographies).

---

[6] Note however that causality may alternatively run in the opposite direction: many people may have chosen to learn the original language because of the relevance of the ideas produced in it.

[7] Note that although there may be a delay of several years between the national depository of a country receiving a translation and Unesco listing the translation in the IT, the IT reports the year in which such translations were published, not just the year in which they were reported. I attribute them to the former and disregard the latter.

[8] Philology and Linguistics were a separate (and very small) category prior to 1970, and then were combined with Literature. I group them with Literature for all years for consistency.

The bibliographic entry for each translation includes information on the country, city, and year in the which the translation was published, the language of the original title and the target language into which it was translated, the field (UDC class) of the title, the number of pages or volumes of the title, the author, and the original and translated titles of the book.[9]

2.1. Digital translation data: 1979 to 2000

For approximately the period 1979 to 2000, I acquired the IT from Unesco in digital format. Prior to 1979, these data do not exist in digital form. Beyond 2000, there are still translations reported for some countries, but in many cases reporting of the translations published in these years is clearly still incomplete. I do not use data from countries in years where translations are incompletely reported. The digital record for each translation includes the full bibliographic record for the translation, and usually bibliographic details of the original title.

In the regressions that aim to capture contemporary translation patterns, I use translation data from two points in time, the first being the annual average for 1993 to 1995, and second being the annual average for 1998 to 2000. This period is short enough to likely have a relatively constant relationship between translations and distances throughout. In addition, the two points in time fit into the pattern of every fifth year that I use for examining historical trends in translations, as described below. Finally, including two periods as opposed to just one allows more precise estimates of the relationships of interest. The averaging process reduces noise in the data, while limiting the number of time-varying fixed effects required, which is necessary to be able to feasibly estimate the PML model I use, as described in section 3.[10] At the same time, this maximizes the number of countries in the sample: if data are available for a country for only one or two years in either of the three-year periods, I use the average translations for those one or two years.

[9] In a few instances, the IT reports that a title was translated from its original language via an intermediate language. In these cases, I consider the idea flow to be from the original language to the final language, with the intermediate language just part of the mechanism. I thus count these as translations from the original language to the target language, and disregard the intermediate language.

[10] Using just the years 1994 and 1999, instead of the averages as described here, does not substantially alter the results, though it increases the standard errors on the estimates (results not presented).

2.2. Hand-collected translation data: 1949 to 1979

For translations prior to 1979, the IT exists only in hard copy format. The total number of translations listed in the IT in one year is often thirty or forty thousand; to make digitization manageable, I restrict my digitization effort to every fifth year from 1949 to 1979. I choose to begin my sample period with 1949 because Unesco only began systematic data collection in 1948. Specifically, Unesco did not compile translations for the period 1941-47, and the pre-war data (1932 to 1940) were collected by a different institution and are not entirely comparable.[11] Because some countries do not report their translations to Unesco every year, and in order to maximize the geographical coverage of my historical translation data, where the exact year of interest was not available for a country but the preceding or following year was, I substitute that year instead.[12] The years for which I have translation data for each country are listed in Appendix B.

Within each country, year and field, I take a 100, 50, 20, 10 or 5 percent sample of entries.[13] This amounts to approximately 100,000 records in total. I choose the percentage to give me approximately 100 titles (or collect data on all translations where the total number is fewer than 100) in total for each country-year-field group. In all subsequent work I weight observations according to the inverse of their probability of being sampled. For each entry I sample, I record the reporting country, original and target languages, UDC category, year of publication and number of pages of the book.

For the historical translation series, I combine these newly-digitized data with digitized data provided by Unesco for every fifth year from 1979 to 1999.

---

[11] In future work, I plan to expand my sample to include the pre-WWII period.

[12] Where data exist for consecutive years, they are very highly correlated, so this approximation is unlikely to have a significant effect on the results.

[13] My sample within each country-year-category group is pseudo-random in the following sense. Entries in the IT are identified by an entry number which starts at one each volume and either counts up throughout the whole volume, or restarts from one at the start of each new country entry. If I am taking a one in n sample, I sample every title whose identification number is a multiple of n. I do this instead of taking a genuinely random sample for speed of data entry, and because the ordering within each group of titles alphabetically by author means this method is unlikely to bias my sample with respect the original or target language, the main dimensions of interest that vary within such a group.

3. Empirical strategy

To shed light on the impediments to translation flows, in section 4 I investigate the relationship between geographic distance and translations. To decompose the distance effect I find, I add further controls for various types of distance or dissimilarity between the countries.

The basic specification is a gravity model in multiplicative form estimated using the pseudo-maximum-likelihood (PML) estimator recommended by Santos Silva and Tenreyro (2006) for the gravity equation specifically and constant-elasticity models in general. Santos Silva and Tenreyro consider a constant elasticity model of the form

$$T_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \eta_{ij},$$

where η_{ij} is a multiplicative error term with $E\left(\eta_{ij} \mid Y_i,\ Y_j,\ D_{ij}\right) = 1$ and where η_{ij} is assumed to be statistically independent of the regressors. They show that if η_{ij} is heteroskedastic in a manner that depends on the regressors, then $\ln \eta_{ij}$ is not independent of the regressors, and thus linearizing equation (x) and estimating it by OLS leads to inconsistent estimates. In fact, they demonstrate that this heteroskedasticity is usually present and substantial in gravity models of trade, so estimating the relationship in multiplicative form using their PML estimator is preferred. The nature of translation data suggests such heteroskedasticity is also likely to be present here, so I use their PML estimation technique. A further advantage of this method is that it has no difficulty with observations where the value of the dependent variable is zero.

The equation I thus estimate is:

$$trans_{ijt} = \alpha dist_{ij}^{\phi} e^{\beta_{it}} e^{\gamma_{jt}} \varepsilon_{ijt}, \qquad (1a)$$

the more familiar linearized form of which is

$$\ln trans_{ijt} = \alpha' + \phi \ln dist_{ij} + \beta_{it} + \gamma_{jt} + \nu_{ijt}, \qquad (1b)$$

where $\alpha' \equiv \ln \alpha$ and $v_{ijt} \equiv \ln \varepsilon_{ijt}$. Here $trans_{ijt}$ is the number of translations into language-in-country pair *i*, from language *j*, in year *t*, $dist_{ij}$ is the geographic distance between the main country of language *j* and the country denoted by *i*, the βs are time-varying fixed effects for target language-in-country, the γs are time-varying fixed effects for original language, and ε is a error term with mean 1. The coefficient of interest is ϕ, which is the elasticity of translation flows with respect to geographic distance.

I run specifications where I control for the population and GDP per capita of the original and translating countries instead of including time-varying origin and target fixed effects. However, Anderson and van Wincoop (2004) show such a specification is likely to result in a biased estimate of the coefficient on distance because it suffers omitted variable bias. In the translation context, it does not account for the average barriers to translation from all possible original languages faced by a country. These are likely to be correlated with average distance from potential original countries, and thus with the distance from any one original country, so failing to account for them introduces bias. Thus my preferred specification includes time-varying origin and target fixed effects.

I also augment this model by including measures of non-physical distance between the countries, such as differences in culture.

3.1. Original languages and target languages and countries for gravity model

An observation in the gravity equation I estimate is an original language, a target language in a country, and a year. Two questions then arise. First, from what set of original languages should translations be included, and should this vary by translating country? Second, into which target language or languages in each country should translations be included?

I do not allow the set of original languages to vary by translating country. That is, each target language in a translating country in a year contributes the same number of observations to the regression, one for each language in a set of original languages that does not vary by country. The advantage of this method is that it does not impose any priors about which countries will translate from which languages. However, it does mean many measured translations flows are zero. The set of original languages out of which I consider translations in my primary specification is the set of the most widely spoken 100 languages as listed by Ethnologue. From these languages, I drop those out of which translations are never published.

One option for target languages in translating countries would be to include an observation for translations into each possible language in every country. However, most of these flows will be zero. In fact, most would also not signify a relevant transfer of ideas even if they were non-zero. For instance, Peru has no reason to translate into Czech, and, if it were to publish such translations, they could not be read by more than a trivial number of Peruvians anyway. In my primary specification, I thus include translations into a language in a country only if the language is (de facto) official in the whole of the translating country, suggesting such translations indicate an inflow of ideas to a settled and non-trivial group in the country. This means translations into the regional or immigrant languages of a country are generally not considered. However, it means I do include translations into some languages that don't have a considerable body of native speakers who communicate primarily in that language in the country. This is usually because the language is official for heritage purposes and out of respect for a minority language group that has largely been assimilated into the dominant language group.

For the purpose of both original and target languages, I group dialects of a language together, and I aggregate languages to the macrolanguage level as they are coded by the ISO 639-3 classification system.[14]

A macrolanguage is defined by Ethnologue as "multiple, closely related individual languages that are deemed in some usage contexts to be a single language." For example, Bosnian, Croatian, and Serbian are the three languages that constitute the macrolanguage Serbo-Croatian. One such usage context as described by Ethnologue is the literature or the writing system; thus individual languages within a macrolanguage may be identical or virtually identical in written form, and if they are not they are likely to be highly mutually intelligible in written form. Consequently, it seems appropriate to group them together as a single language for translation purposes. In addition, in some cases the IT does not distinguish between translations from (or into) individual languages of a macrolanguage, so separating them would not be possible.

[14] Available from http://www.sil.org/iso639-3/.

3.2. Matching original languages to countries

Identifying which translations flows are occurring between which populations is complicated by the lack of a one-to-one mapping between languages and countries. Although the bibliographic entries in the Index Translationum identify the country in which the translation was published, one of their limitations is that the country in which the original title was published is not given, only the language in which it was originally written. In order to examine the relationship between physical distance and translation flows, I need to attribute translations to original countries.

Conceptually, there are several ways to think about what this attribution might want to capture. One option is that titles diffuse through linkages with the country in which they are written. Thus ideally the relevant "distance" for an original language that is spoken in multiple countries is some weighted average of the distances to those countries. However, if most of the titles published in the language originate in one of those countries, the distance to that country seems a sensible approximation. An alternative is that all countries where a particular language is spoken widely (even if they do not generate many original works in the language) act as distributors of the titles written in that language. Thus to be close to English, for example, a country need not be close to the US or the UK, but may be close to a smaller English-speaking nation such as Australia or New Zealand.

I thus use two alternative strategies to attribute translations to original countries. In my central specification, I attribute all translations from a language to the "main" country of the language as listed by Ethnologue, with the exception of English, which I attribute to the USA rather than the UK in my main specification, based on the much greater GDP of the USA relative to the UK. However, I note that results are robust to attributing English to the UK, or to dropping it entirely. In my alternative specification, for each original language I compile a list of major countries of the language. A language is classified as major in a country based on how widespread its native speakers are in the country, the population of native speakers in the country relative to worldwide, and whether the language is national or official in the country. Very small countries (e.g. Monaco) are not counted as major unless they are the main country of the language. The major countries for each original language are listed in Appendix A. I then set the original country of translations in country C from language L to be language L's major country that is physically closest to country C.

4. How distances affect translations

In this section, I study how bilateral translation flows are affected by distance between countries, thus shedding light on an important impediment to the international diffusion of ides.

Note also that this is a setting with many commonalities with trade in goods, but here transportation, time, and much of the distribution costs are negligible. Specifically, because only a single copy of the title must travel between the countries in order for a translation to occur, translations have effectively zero transportation costs, both direct and indirect. They are also largely free from border-related costs, policy barriers such as tariffs and quotas, and many legal and regulatory costs. That is, translations face near zero costs related to the physical movement of goods. Translations are, however, expected to be subject to all the costs of contracting between parties in different countries, plus search and information costs, that trade in goods face and that may vary with distance. In addition, both trade and translations may occur more between closer countries because consumers in these countries have more similar tastes, thus more demand for each other's books or products. Studying how distance affects translations thus sheds light on the factors beyond transportation costs that contribute to the negative relationship between distance and trade.

My estimation framework is an augmented gravity model, in which (directional) translations between two countries depend on the economic sizes of the countries, and the physical distance between them. I assume a constant elasticity functional form, and estimate the model by PML as described in section 3. I add controls such as the cultural distance between the countries to measure the extent to which countries translate more from their neighbors because they are more culturally similar to them.

4.1. The negative distance effect: Neighboring countries translate more from each other than from distant countries

In a basic gravity specification with physical distance as the only distance measure, presented in Table 1, I find a strong negative correlation between the number of titles translated and the physical distance between the original and translating countries. Appendix Table A presents the same specifications, but uses OLS and predicts the natural log of the number of translations plus 1. The strong negative correlation is again present, though the coefficients on distance are smaller in magnitude. The data used in these regressions are a panel of the years 1994 and 1999, a short enough period that we expect the relationship between distance and translations to have remained relatively constant. Column 1 presents the basic gravity specification where the number of translations flowing from one country to another depends on the populations of the two countries, their GDP per capita, and the distance between them. For each target language in a translating country, we consider translations from the same set of original languages, namely those of the 100 languages most widely spoken in the world that are ever translated. For each of the 56 countries with translation data in at least one of the two years, we consider translations into each of the languages that are official in the whole of the country. To generate distance measures, I assign each original language to its main country as described in section 3.2.

As expected, the population and GDP per capita of the translating country are positively and significantly correlated with translation flows with elasticities of 0.72 and 0.75 respectively. The elasticity of translations with respect to the population of the original country is 1.1; the elasticity with respect to the GDP per capita of the original country is 3.3. This strong relationship between wealth of the original country and translations suggests the creation of ideas with international relevance is very much concentrated in rich countries, whereas less rich countries tend to consume ideas created elsewhere. The OLS version of this regression, presented in column 1 of Appendix Table A, shows these basic covariates have moderate explanatory power: the R-squared in this regression is 0.15.

Column 2 of Table 1 adds controls for colonization relationships between the original country and the translating country in either direction. There are relatively few of these in the data, particularly because the translating country must have at least one official language that differs from the language of the colonizer in order for the pair to appear in the data, and neither direction of colonizing relationship is significantly correlated with translation flows.

In these first two specifications, the elasticity of translations with respect to geographic distance is -0.9, suggesting a 10% increase in the distance between two countries corresponds to a 9% decrease in translation flows between them. However, Anderson and van Wincoop (2004) show such a specification is likely to suffer omitted variable bias as explained in section 3, so in column 3 I add in time-varying fixed effects for original language and target language-translating country pairs. The elasticity falls in magnitude to -0.47, but remains significant.

Next I add controls for the original and translating countries being contiguous, and the original language being widely spoken in the translating country. Both are associated with significantly higher translations, but their inclusion doesn't eliminate the relationship between distance and translations. The interpretation of these two effects is similar. Sharing a land border with a country suggests the populations will both interact more, implying lower search and transaction costs of translating from each other, and have more similar tastes, implying greater demand for translations. Similarly, mixing geographically with a group that speaks a foreign language can be expected to stimulate translations from both the demand and the supply sides.

Columns 5 to 10 of Table 1 run the same specification as column 4, but vary the sample of original languages and translating countries in a number of ways. Column 5 restricts the original languages to those in the top 100 that can be unambiguously attributed to a single country, which eliminates many of the large original languages such as English, German and Spanish. The elasticity of translations with respect to distance increases in magnitude to -1.1 in this specification. Column 6 restricts original languages to the four main "research languages", namely English, French, German and Japanese. The magnitude of the correlation is similar, though significance decreases because of the much smaller sample size. Column 7 uses all of the top 100 original languages, but attributes each to the country in which it is widely spoken that is geographically closest to the translating country (as explained in section 3.2), instead of to its main country. The coefficient on distance falls slightly in magnitude, which suggests geographic proximity to a secondary country of a language may be a less-than-perfect substitute for geographic proximity to the main country of the language for the purpose of enhancing idea flows. Column 8 differs from column 4 in that it restricts the sample of target languages in translating countries to those where the translating country is the main country of the target language. For example, it includes translations into German in Germany, but excludes translations into German in Switzerland. The coefficient of interest is unaffected. Column 9 instead includes all the target languages for each translating country that are (i) official in at least part of the translating country, (ii) spoken natively by at least 500,000 people in the country, and (iii) spoken by at least 5% of the country's population. Results are again largely unaffected. Finally, column 10 looks at translation flows only within Europe. That is, it includes original languages in the top 100 that have a European country as their main country, and translating countries that are European. The coefficient on distance increases slightly.

Overall, it seems that in the 1990s a 10 percent increase in distance corresponded to roughly a 3 to 5 percent decrease in translations, despite translations having zero transportation costs. This suggests there are significant distance-varying costs involved in translation, which may relate to search and information, or to the costs of forming contracts. Geographic correlation of tastes that causes demand to decrease with distance may also contribute to the distance effect.

This elasticity of -0.3 to -0.5 is significantly lower than those found in the literature on trade in goods, which generally range from -1.08 to -1.24.[15] Under the (admittedly strong) assumption that the non-transportation costs faced by translations vary with distance in the same way as the equivalent costs for trade, the magnitudes of these coefficients suggest that roughly half to three quarters of the elasticity of trade with respect to distance is the result of transportation costs. However, this comparison may be confounded by the use of PML estimation in this paper. Where Santos Silva and Tenreyro (2006) use PML as opposed to OLS to estimate the distance effect on trade in a model with importer and exporter fixed effects, their coefficient falls from -1.3 to -0.75. This lower elasticity estimate for trade suggests a third to three fifths of the distance effect in trade is due to transportation costs. These estimates are in the same range as the value of a half found by Feyrer (2011) using a very different approach.

[15] Disdier and Head (2008)

4.2. The negative correlation between physical distance and translations decreased over time

The latter half of the twentieth century saw many changes, including increases in globalisation and the ease of international communication. These are expected to have weakened the observed relationship between distance and translations. I thus next estimate how the elasticity of translations with respect to distance has changed over time. Figure 1 illustrates how the correlation between physical distance and translations changed over time. These correlations are coefficients from regressions of translations on geographic distance, origin and target fixed effects, and controls as in column 4 of Table 1, run separately for each fifth year from 1949 to 1999. The figure presents the 95% confidence interval of the coefficient on translations for two different sets of translating countries: the solid blue lines are for the consistent set of 9 countries for which data are available every year; the dashed red lines are for all the countries for which data are available in the particular year. In each case, the magnitude of the negative correlation decreased significantly over the period 1949 to 1999, particularly over the last two decades.

This contrasts with the changes seen in the distance effect in trade, which, according to Disdier and Head's (2008) meta-analysis of the results from many papers, rose mid-century and has remained persistently high since. The decrease in the inhibitory effect of physical distance on translations over time is consistent with several causal mechanisms. For instance, the ease of international travel and communication decreased over this period, and their costs fell. This could have both weakened the relationship between distance and the search, information, and transaction costs of translation, and stimulated interest in geographically distant cultures. If search and information costs are higher on average for books than for goods, this could explain why the distance effect decreased for translations but not for trade. Note MacGarvie (2005) similarly finds a decrease in the effect of distance on patent citations over the period 1980-1995, which is also consistent with such a change in information costs.

4.3. Translations of different types of books are affected differently by physical distance

There are a number of reasons to expect translations of titles in different fields to be affected differently by physical distance. On the demand side, fields differ in the extent to which their ideas are region-specific. For instance, history titles frequently focus on a particular region of the world, thus are likely to be of more interest to countries in that region. Similarly, religion titles tend to relate to a specific religion, and thus will be of more interest in countries where that religion is widespread, which tend to be geographically clustered. Conversely, many natural science ideas (such as ideas in physics and chemistry) are equally relevant anywhere in the world. In addition, the degree to which titles written in different languages are substitutes for each other varies by field. In fields with high substitutability, there may be no reason to translate from very distant languages because nearby languages are sufficient to meet demand, thus if costs rise with distance translations may fall off quickly with distance. In fields with low substitutability, a specific idea can only be sourced from one language, so distance is likely to play a lesser role in determining translation flows.

Figure 4 shows the coefficients and 95% confidence intervals of the coefficient on physical distance when translations in each field are regressed on physical distance and other controls as in column 4 of Table 6. Physical distance and translations are negatively correlated for all fields of translation, though the magnitude of the correlation varies across fields. Perhaps surprisingly, physical distance has the largest inhibitory effect on translations in the fields of natural science and applied science, and the smallest in philosophy and arts.

These results by field demonstrate that distance matters even for translations of "economically useful" titles such as titles in natural and applied science, not just for titles that may be considered largely consumption goods, such as many philosophy, arts, or fiction titles. Furthermore, the fact distance has a greater effect for titles with less of a cultural or taste component suggests taste differences that increase with distance may have a lesser role in driving the distance effect on translations relative to supply-side frictions.

4.4. Countries with similar physical environments translate more from each other

The negative relationship between translation flows and physical distance could be driven by several factors, all of which apply to trade in goods to some extent: search and information costs involved in identifying foreign titles worth translating; costs of negotiating rights to translate a title; and tastes for ideas that differ more widely between more distant countries.

One reason tastes for ideas may be more similar in neighboring regions is that physical environment (such as climate, terrain, the types of plants that will grow etc) tends to be more similar in neighboring regions, and the physical environment in which a society lives might affect the types of ideas that are relevant or interesting to its members. To estimate the importance of this effect, I augment the basic gravity model with the difference between countries in altitude profile, biome region profile, and climate region profile. Column 1 of Table 2 presents the results from this regression. Differences in altitude profile and biome region profile significantly inhibit translation flows, but together these three differences explain only a modest 6.5% of the negative correlation between physical distance and translations. The coefficient on altitude profile difference suggests that, relative to two countries with the same altitude profiles, two countries with altitude profiles that are only 90% similar will translate 8% less from each other. However, much of this correlation can be shown to be attributable to the cultural differences that are correlated with altitude profile differences (columns 2 to 6). A similar increase in the difference in biome region profiles corresponds to a 3% decrease in translations.

4.5. Countries with similar cultures translate more from each other

Cultural differences that are correlated with physical distance could cause translations to decrease with distance for two main reasons. On the demand side, cultural similarity could imply similar preferences, meaning the countries have higher demand for each other's books. On the supply side, cultural similarity could lead to greater trust and understanding, which reduce transaction costs. In columns 2 to 8 of Table 2, I thus add controls for various measures of cultural distance: religious distance, linguistic distance, genetic distance, and survey-based measures of cultural differences. With the exception of genetic distance, these cultural distance measures each significantly inhibit translation flows; the effects are particularly strong for linguistic and religious distance. A 10 percentage point decrease in the probability a randomly chosen individual from the translating country has the same religion as a randomly chosen individual from the original country corresponds to a 10 to 18 percent decrease in translations. A 10 percentage point increase in distance between the languages corresponds to a 2 to 7 percent decrease in translations. Finally, a one-standard-deviation increase in cultural distance as measured by Hofstede's (1980) survey measures corresponds to an 8 to 16 percent decrease in translations.

Note however that adding these controls reduces the elasticity of translations with respect to physical distance by at most a quarter. This suggests that cultural differences contribute to distance-varying contracting costs or to demand that prefers titles written in nearby countries, but that other distance-varying costs, such as search and information costs, play a large role in the distance effect.

The process of globalization over the past half century has made the world smaller in many ways; international travel has become cheaper and faster, and global communications have improved beyond measure. The forces that have allowed distant cultures to mingle more easily may have decreased cultural barriers to the flow of ideas. Also plausible is that globalization has caused a reactionary increase in nationalism that may have actually decreased receptiveness to foreign ideas. It is thus unclear theoretically how the relationship between cultural distances and translation flows will have changed over time. Appendix Figure A shows the negative correlation between religious distance and translations tended to increase between 1949 and 1999, while the correlation between linguistic distance and translations tended to decrease.[16]

[16] In column 9 of Table 7, I add controls for trade flows in each direction between the original and translating countries, in order to see descriptively how trade in ideas (translations) are correlated with trade in goods. Note the coefficients on these variables in particular should not be interpreted causally because of reverse causality and unobserved heterogeneity. For instance, trade flows between countries may cause an increase in idea flows and thus an increase in translations, but translations may increase understanding and decrease transaction costs, thus increasing trade. Trade in manufactured goods and translations may also be complements. The coefficients on both imports and exports are positive and significant, and are similar in magnitude: a 10% higher flow of trade in either direction corresponds to a 2.4% higher translation flow. The direction of this effect is consistent with the causality stories running in either direction. One interesting point to note is that inclusion of these two trade variables eliminates the negative correlation

4.6. Translations published in more developed countries decrease less with physical distance

Because of differences such as the infrastructure or technology available them or the structure of their publishing industries, some countries may face higher distance-related costs of translation than others. In Table 3, I thus allow the effect of distance to differ by various characteristics of the original or translating country. In column 1, I allow the effect of distance to differ by the wealth of the translating country. I find the effect of distance is 89% weaker for a translating country on the 75th percentile of GDP per capita among countries in my data relative to the 25th percentile, or 70% weaker for a country with urbanization rate on the 75th percentile relative to the 25th percentile (column 3). More developed countries can differ from less developed countries across a multitude of dimensions, making it difficult to establish the causal mechanism behind these results. For instance, communication technologies tend to be more advanced, reliable, and widespread in richer countries, which could reduce search and information costs.

However, the fact that translations published in poorer countries are more affected by distance has potentially important implications for the international diffusion of knowledge. Specifically, it suggests that countries that are further from the world knowledge frontier, and thus that can benefit most from adopting ideas that already exist elsewhere, are actually less able to access these ideas.

Similarly, distance is significantly less important for translations of titles originating in wealthier countries. The effect of distance is 49% weaker for translations from original countries on the 75th percentile of GDP per capita relative to the 25th percentile (column 2); it is 52% weaker for a country with urbanization rate on the 75th percentile relative to the 25th percentile (column 4).

Distance is also significantly less important for translating countries that are more democratic, as shown in column 5. The effect of distance is 44% weaker for a translating country on the 75th percentile of democracy relative to a country on the 25th percentile.

---

between distance and translations. However, as Appendix Tables B1 and B2 show, this was not the case prior to the 1990s.

Such a relationship could be observed if more democratic countries were less threatened by ideas that differed more from their own than were less democratic countries. The level of democracy in the original country is not significantly correlated with the strength of the relationship between physical distance and translations (column 6).

4.7. A large bilingual population increases translations and reduces the relationship between distance and translations

There are a number of reasons to expect bilingualism in the original and target languages to matter for translation flows. On the one hand, having a higher proportion of the population in the translating country that speaks both the original and target languages is likely to decrease the costs of translation: the supply of potential translators will be greater, and editors in publishing houses will both be more likely to be able to evaluate the original titles for themselves, and better understand the cultural barriers that add cost to transactions. Bilingual individuals may also facilitate the international flow of information about foreign titles worth translating.[17] On the other hand, if bilingual individuals view the original titles as substitutes for translations, this may result in lower demand for translations, causing translation flows to decrease.

In column 7 of Table 3, I present regression results that test which of these effects dominates, by including in the augmented gravity regression variables for the proportion of the population in the translating country a) that is native in the target language and conversational in the original language, and b) that is native in the original language and conversational in the target language.[18] I find both types of bilingualism are significantly correlated with higher translation flows. Specifically, a one-standard-deviation higher fraction of the population that is native in the target language and speaks the original language corresponds to an 8 percent higher translation flow, and a one-standard-deviation higher fraction of the population that is native in the original language and speaks the target language corresponds to a 22 percent higher translation flow.

[17] Note that causality may also run in the opposite direction. That is, individuals may choose to learn a foreign language because that language produces a lot of ideas that they value, or the presence of a large migrant population may stimulate interest in translations from the original language of the migrants.

[18] Individuals of type (a) are likely to be natives to the country who have chosen to learn the original language; individuals of type (b) are likely to be migrants from a country that speaks the original language, who have learned the local language.

A bilingual population may also mitigate distance-related costs of translations, for example, by facilitating the transmission of information about foreign titles to translate. I next test whether the relationship between physical distance and translations varies with the proportion of the population that is bilingual in the original and target languages. Column 8 of Table 3 presents the results from an augmented gravity regression where the distance effect is allowed to vary with both types of bilingualism. It shows that the proportion of bilinguals who are native in the target language is negatively correlated with the strength of the distance effect. A one-standard-deviation increase in the proportion of the population with this type of bilingualism (11 percent of the population) corresponds to a distance effect that is 12 percent weaker. This suggests that a moderately-sized bilingual population reduces but does not eliminate distance-related costs of translation.

5. Conclusions

In this paper I investigate how flows of book translations between countries are correlated with the physical distance between the countries, and thus shed light on some important impediments to the international diffusion of ideas.

My analysis may also be informative about the underlying causes of the negative relationship between distance and trade in goods, a robust finding in international economics, the driving factors behind which remain imperfectly understood.[21] The most obvious contributing factor to the relationship is transportation costs, but an increasing literature demonstrates that transportation costs account for only a fraction of the total distance effect.[22] For example, in a recent paper, Feyrer (2011) uses time-varying exogenous variation in effective distance generated by the temporary closure of the Suez Canal to estimate that only half the elasticity of trade with respect to distance is driven by transportation costs.

Here I consider a setting in which transportation, time, and much of the distribution costs are negligible by studying how distance affects the translation of

[21] e.g., Disdier and Head (2008), Blum and Goldfarb (2006), Feyrer (2011).
[22] Anderson and van Wincoop (2004).

books.[23] Specifically, only a single copy of the title, in digital or hard copy form, must travel between the countries in order for a translation to occur. Translations thus have effectively zero transportation costs, both direct (freight, insurance) and indirect (e.g. holding cost for the goods in transit). Translations are also largely free from several other trade costs (discussed in Anderson and van Wincoop, 2004): they avoid border-related costs, policy barriers such as tariffs and quotas, and many legal and regulatory costs. In general, none of the costs related to the physical movement of goods apply to book translations. Studying the determinants of translation flows is thus informative on the drivers beyond transportation costs of the negative relationship between trade in goods and distance.

I estimate a gravity-type model in which translation flows are affected by characteristics of the original and translating countries (such as GDP per capita) and the distance between them. I estimate the elasticity of translations with respect to distance to be -0.3 to -0.5 during the 1990s, which is considerably smaller than the equivalent elasticity for trade found in the literature, which usually ranges from -1.08 to -1.24.[24] Under the assumption that the non-transportation costs faced by translations vary with distance in the same way as the equivalent costs for trade, the magnitudes of these coefficients suggest that roughly half to three quarters of the elasticity of trade with respect to distance is the result of transportation costs; comparisons with results using a more similar estimation method, from Santos Silva and Tenreyro (2006), decrease this range to a third to three fifths. Although my method is very different, these results are comparable to Feyrer's (2011) estimate that half the elasticity of trade with respect to distance is the result of transportation costs.

Several pieces of more refined analysis of the relationship between translations and distance are consistent with an important role for search and information costs and a lesser role for demand factors in the negative relationship between translations and distance. First, the distance effect is larger in the fields of natural and applied science, where tastes are less important, than in the fields of arts, literature and philosophy, which have a higher cultural component. This is the opposite to what we would expect if

[23] This approach is similar to that of Blum and Goldfarb (2006), who study how distance affects trade in digital goods consumed over the internet, which have no transportation, time, or distribution costs.
[24] Disdier and Head (2008).

geographically correlated tastes were the main driving factor behind the distance effect. Second, cultural distance between countries does inhibit translation flows, but accounts for relatively little of the overall distance effect, suggesting non-cultural factors play a large role.

My results also have implications for the international diffusion of ideas. The augmented gravity regressions that include geographic and cultural distances suggest that, idea flows are hindered both by geographic distance and cultural distance between countries. Furthermore, idea flows into less developed countries are hindered more by distance than idea flows into more developed countries. This relationship is a force against income convergence between rich and poor countries: the countries that can benefit most from catch-up growth by adopting foreign ideas seem to face greater frictions in accessing these ideas. However, the inhibiting effect of distance has decreased over time, which suggests that even the barriers surrounding less developed countries may be lower in the future.

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth*, 8: 155-194.

Anderson, James, and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature*, XLII: 691-751.

Blum, Bernardo, and Avi Goldfarb. 2006. "Does the Internet Defy the Law of Gravity?" *Journal of International Economics*, 70: 384-405.

Disdier, Anne-Célia, and Keith Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *Review of Economics and Statistics*, 90(1): 37-48.

Fearon, James. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8: 195-222.

Feyrer, James. 2011. "Distance, Trade, and Income - The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment." Working Paper.

Griffth, Rachel, Sokbae Lee, and John van Reenen. 2007. "Is Distance Dying at Last? Falling Home Bias in Fixed Effects Models of Patent Citations." NBER Working Paper No. 13338.

Hofstede, Geert. 1979. "Value systems in forty countries: Interpretation, validation, and consequences for theory." In *Cross-Cultural Contributions to Psychology*, ed. L. H. Eckensberger, W. J. Lonner and Y. H. Poortinga. Lisse, Netherlands: Swets and Zeitlinger.

Hofstede, Geert. 1980. *Culture's Consequences: International Differences in Work-related Values*. Beverly Hills, CA: Sage.

Hofstede, Geert. 1982. "Dimensions of National Cultures." In *Diversity and unity in cross-cultural psychology*, ed. R. Rath, H. S. Asthana, D. Sinha and J. B. H. Sinha. Lisse, Netherlands: Swets and Zeitlinger.

Hofstede, Geert. 1983. "Dimensions of national cultures in fifty countries and three regions." In *Explications in Cross-cultural Psychology*, ed. J.B. Deregowski, S. Dziurawiec and R.C. Annis, 335-355. Lisse, Netherlands: Swets and Zeitlinger.

Hofstede, Geert. 2001. *Culture's Consequences*. 2nd ed., Thousand Oaks, CA: Sage.

Hofstede, Geert, and Michael Bond. 1984. "Hofstede's Cultural Dimensions: An Independent Validation Using Rokeach's Value Survey." *Journal of Cross-Cultural Psychology*, 15(4): 417-433.

Klenow, Pete, and Andrés Rodríguez-Clare. 2005. "Externalities and Growth." In *Handbook of Economic Growth*. Vol. 1A, ed. Philippe Aghion and Steven N. Durlauf, Chapter 11. Elsevier.

Kogut, Bruce, and Harbir Singh. 1988. "The Effect of National Culture on the Choice of Entry Mode." *Journal of International Business Studies*, 19(3): 411-432.

Kuznets, Simon. 1966. *Modern Economic Growth: Rate, Structure and Spread*. New Haven, CT: Yale University Press.

MacGarvie, Megan. 2005. "The Determinants of International Knowledge Diffusion as Measured by Patent Citations." *Economic Letters*, 87: 121-126.

Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.

Mokyr, Joel. 2009. "The Contribution of Economic History to the Study of Innovation and Technical Change: 1750-1914." In *Handbook of Technological Change*. Vol. 1, ed. Bronwyn H. Hall and Nathan Rosenberg, Chapter 2. Elsevier.

Mokyr, Joel. 2010. *The Enlightened Economy: an Economic History of Britain 1800-1950*. Yale University Press.

Ng, Siew Imm, Julie Ann Lee, and Geoffrey Soutar. 2007. "Are Hofstede's and Schwartz's value frameworks congruent?" *International Marketing Review*, 24(2): 164-180.

Romer, Paul. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5): S71-S102.

Romer, Paul. 1993. "Idea Gaps and Object Gaps in Economic Development." *Journal of Monetary Economics*, 32(3): 543-573.

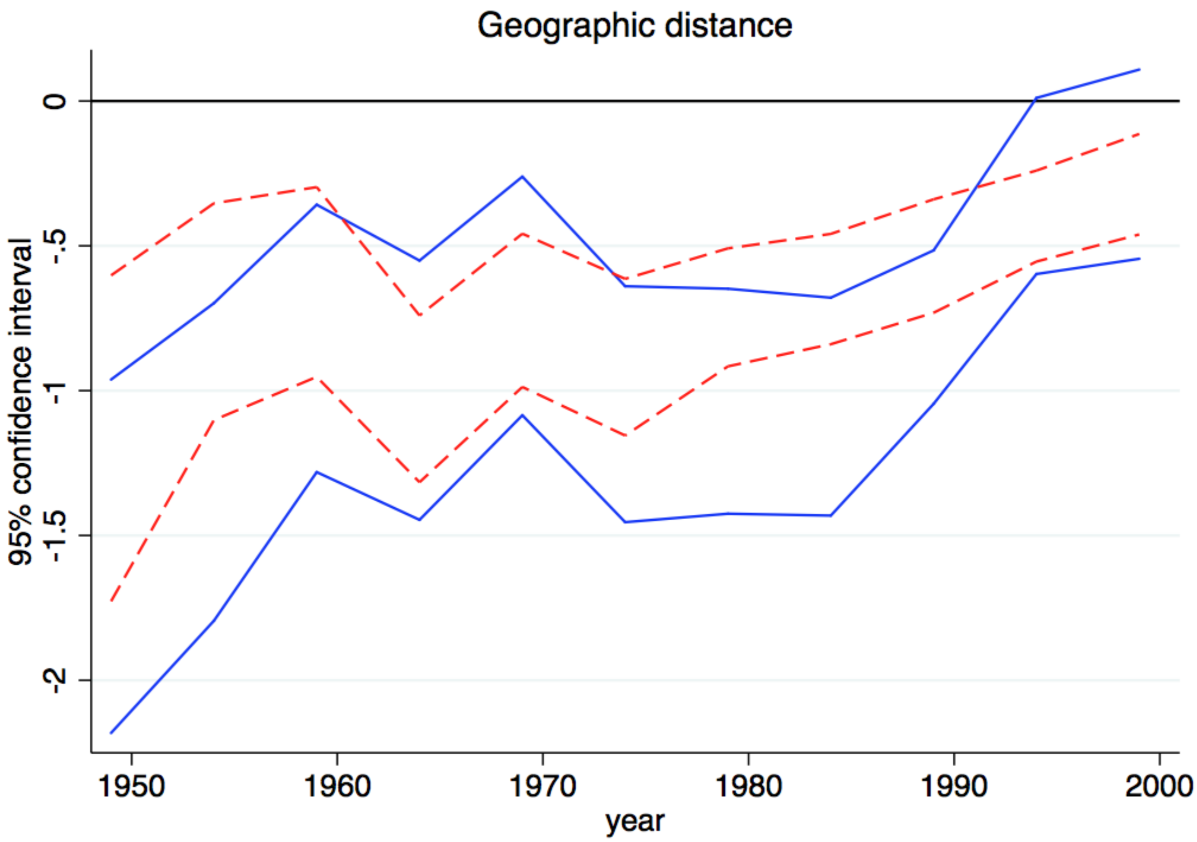
Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641-658.

Schwartz, S.H. 1994. "Beyond individualism/collectivism: new cultural dimensions of values." In *Individualism and Collectivism: Theory, Method, and Applications*, ed. U. Kim, H.C. Triandis, C. Kagitcibasi, S.C. Choi and G. Yoon, 85-119. Thousand Oaks, CA: Sage.

Schwartz, S.H. 1999. "Cultural value differences: some implications for work." *Applied Psychology: An International Review*, 48: 23-48.

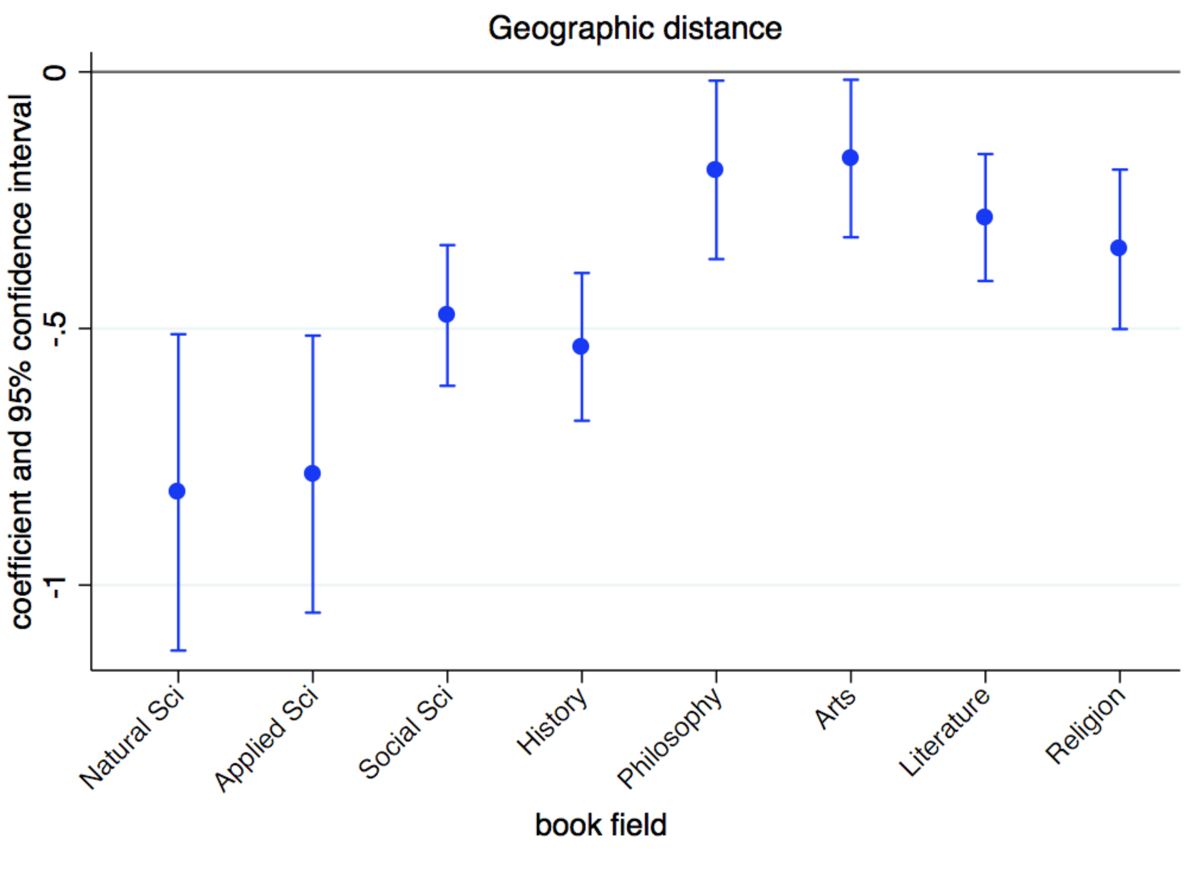
Spolaore, Enrico, and Romain Wacziarg. 2009. "The Diffusion of Development." *Quarterly Journal of Economics*, 124(2): 469-529.

Figure 1: The negative correlation between geographic distance and translations decreased over time



This figure shows the 95% confidence interval of the coefficient on geographic distance in regressions of the number of translations (ln) on distance (ln) and other controls as in column (4) of Table 1, run separately by year. The solid blue line is for the consistent set of 9 countries for which data are available each year; the dashed red line is for all the countries for which data are available in any one year.

Figure 2: The negative correlation between geographic distance and translations by field



This figure shows the point estimate and 95% confidence interval of the coefficient on geographic distance in regressions of the number of translations (ln) on distance (ln) and other controls as in column (4) of Table 1, run separately by book field. Data are for the years 1994 and 1999.

Table 1: Closer countries translate more from each other

Dependent variable: number of translations (ln)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Original languages:	top 100	top 100	top 100	top 100	unambiguous original country	main research languages	top 100, assigned to nearest major country	top 100	top 100	European of top 100
Target languages for each translating country:	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country, translating country is main country	official and widespread in country	official in whole country, European country
Variable										
Physical distance between original and translating countries (ln)	-0.888***	-0.882***	-0.473***	-0.341***	-1.088***	-0.326*	-0.254***	-0.311***	-0.360***	-0.564***
	(0.111)	(0.110)	(0.071)	(0.062)	(0.112)	(0.183)	(0.050)	(0.065)	(0.061)	(0.084)
Population of translating country (ln)	0.722***	0.732***								
	(0.079)	(0.082)								
GDP per capita of translating country (ln)	0.745***	0.762***								
	(0.116)	(0.123)								
Population of original country (ln)	1.099***	1.090***								
	(0.064)	(0.064)								
GDP per capita of original country (ln)	3.324***	3.309***								
	(0.150)	(0.148)								
Original country colonised translating country		0.048								
		(0.443)								
Translating country colonised original country		-0.595								
		(0.503)								
Original and translating countries are contiguous				0.531***	0.479***	0.339***	0.462***	0.537***	0.530***	0.389***
				(0.101)	(0.153)	(0.118)	(0.093)	(0.106)	(0.100)	(0.108)
Original language is widespread in translating country				1.095***		1.084***	1.122***	-0.133	1.165***	0.865***
				(0.157)		(0.181)	(0.176)	(0.354)	(0.169)	(0.181)
Original country is translating country	-2.688***	-2.694***	3.564***	2.846***	2.170***		3.011***	3.655***	1.874***	2.056***
	(0.624)	(0.624)	(0.294)	(0.317)	(0.336)		(0.278)	(0.422)	(0.290)	(0.367)
Time-varying target language/country fixed effects			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying original language fixed effects			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,434	12,434	13,262	13,262	9,205	517	13,262	6,637	15,187	2,316
Translating countries	56	56	58	58	58	58	58	37	56	31

Notes: This table presents the results of PML regressions (as described in section 3) of the number of translations from an original language to a target language in a translating country in a year. The same original languages are included for every target language; zero values are included in the estimation, as allowed by the PML procedure. The set of original languages and target language/countries included vary by column. The years included are 1994 and 1999.
The original languages in columns 1-4, 8 and 9 are all those languages in the 100 most widely spoken languages worldwide that are ever translated (top 100 languages). The original languages in column 4 are those of the top 100 languages than can unambiguously be assigned to a single original country. The original languages in column 5 are the four major "research languages", namely English, French, German and Japanese. The original languages in column 7 are the top 100 languages, but the original country used for each language is the geographically closest country where the language is widespread, rather than the main country of the language. The original languages in column 10 are those in the top 100
The target languages in columns 1-7 are the languages that are official in the whole of the translating country. The target languages in column 8 are the languages that are official in the whole of the translating country, and for which the translating country is the main country of the language (e.g. German in Germany, but not German in Switzerland). The target languages in column 9 are the languages that are 1) official in at least part of the translating country, 2) spoken natively by at least 500,000 people in the country, and 3) spoken by at least 5% of the country's population. The target languages in column 10 are the languages that are official in the whole of the translating country, for European countries
Standard errors are robust. Asterisks denote significance at: * p<0.10, ** p<0.05, *** p<0.01.

Table 2: Countries with more similar physical environments and cultures translate more from each other

Dependent variable: number of translations (ln)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Physical distance between original and translating countries (ln)	-0.319***	-0.272***	-0.247***	-0.288***	-0.243***	-0.287***	-0.163	-0.335***	0.076
	(0.087)	(0.074)	(0.067)	(0.083)	(0.085)	(0.092)	(0.114)	(0.110)	(0.069)
Difference between altitude profiles of original and translating countries	-0.788***	-0.720***	-0.369*	-0.353	-0.069	-0.266	-0.168	-0.782***	-0.144
	(0.213)	(0.203)	(0.222)	(0.224)	(0.180)	(0.269)	(0.218)	(0.245)	(0.194)
Difference between climate region profiles of original and translating countrie	0.188	0.112	0.085	0.107	0.140	0.128	0.103	0.270	0.200
	(0.201)	(0.168)	(0.155)	(0.162)	(0.186)	(0.181)	(0.327)	(0.240)	(0.146)
Difference between biome region profiles of original and translating countries	-0.307***	-0.290***	-0.258***	-0.234***	-0.192*	-0.200**	-0.390***	-0.264**	-0.442***
	(0.101)	(0.081)	(0.081)	(0.086)	(0.104)	(0.098)	(0.138)	(0.113)	(0.077)
Religious distance		-1.275***	-1.029***	-1.040***	-1.833***	-1.098***			-0.794***
		(0.211)	(0.199)	(0.198)	(0.164)	(0.236)			(0.179)
Linguistic distance			-0.680***	-0.689***	-0.232**	-0.631***			-0.442***
			(0.144)	(0.148)	(0.106)	(0.145)			(0.141)
Genetic distance				0.130	0.461*	0.069			0.076
				(0.129)	(0.272)	(0.140)			(0.125)
Hofstede's cultural distance					-0.075***		-0.163***		
					(0.017)		(0.027)		
Schwartz's cultural distance						0.002		-0.020	
						(0.028)		(0.029)	
Imports into target country from original country (ln)									0.245***
									(0.039)
Exports from target country into original country (ln)									0.240***
									(0.044)
Original and translating countries are contiguous	0.528***	0.377***	0.342***	0.332***	0.206***	0.335***	0.568***	0.512***	0.183**
	(0.096)	(0.100)	(0.100)	(0.100)	(0.068)	(0.111)	(0.077)	(0.105)	(0.078)
Original language is widespread in translating country	1.168***	1.084***	1.246***	1.247***	1.165***	1.289***	1.262***	1.128***	1.221***
	(0.157)	(0.152)	(0.142)	(0.143)	(0.165)	(0.169)	(0.155)	(0.197)	(0.140)
Original country is translating country	2.340***	1.971***	1.976***	1.940***	1.900***	2.035***	2.359***	2.471***	6.048***
	(0.341)	(0.290)	(0.292)	(0.295)	(0.281)	(0.332)	(0.331)	(0.393)	(0.749)
Time-varying target language/country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying original language fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummy variables for imports are zero and for exports are zero									Yes
Observations	13,262	13,262	13,124	12,988	5,318	4,095	5,403	4,095	12,641
Translating countries	58	58	58	58	36	33	36	33	57

Notes: This table presents the results of PML regressions (as described in section 3) of the number of translations from an original language to a target language in a translating country in a year. The same original languages are included for every target language; zero values are included in the estimation, as allowed by the PML procedure. The original languages are all those languages in the most widely spoken 100 languages worldwide that are ever translated. The target language/countries included are all the languages that are official in the whole of the translating country. The years included are 1994 and 1999.
The altitude profile, climate region profile, and biome region profile difference variables are all constructed to vary between 0 (no overlap in profiles) and 1 (identical profiles). Religious distance is the probability a randomly chosen person from the translating country and a randomly chosen person from the original country have the same religion. Standard errors are robust. Asterisks denote significance at: * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Translations into and out of more developed countries decrease less with physical distance

Dependent variable: number of translations (ln)								
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Physical distance (ln) * GDP per capita of translating country (ln)	0.696***							
	(0.083)							
Physical distance (ln) * GDP per capita of original country (ln)		0.306***						
		(0.060)						
Physical distance (ln) * urbanization of translating country (fraction)			1.894***					
			(0.330)					
Physical distance (ln) * urbanization of original country (fraction)				0.990***				
				(0.340)				
Physical distance (ln) * level of democracy of translating country					0.049*			
					(0.025)			
Physical distance (ln) * level of democracy of original country						0.026		
						(0.016)		
Physical distance (ln) * proportion of translating country bilingual, native in target								0.690***
								(0.188)
Physical distance (ln) * proportion of translating country bilingual, native in original								-0.758
								(1.688)
Proportion of translating country bilingual, native in target							0.712*	-3.842***
							(0.408)	(1.282)
Proportion of translating country bilingual, native in original							6.359***	10.828
							(0.893)	(8.835)
Additional controls as described in the notes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying target language/country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying original language fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,706	12,438	12,988	12,988	12,611	12,428	883	883
Translating countries	56	58	58	58	56	58	28	28

Notes: This table presents the results of PML regressions (as described in section 3) of the number of translations from an original language to a target language in a translating country in a year. The same original languages are included for every target language; zero values are included in the estimation, as allowed by the PML procedure. The original languages are all those languages in the most widely spoken 100 languages worldwide that are ever translated. The target language/countries included are all the languages that are official in the whole of the translating country. The years included are 1994 and 1999. The variable *proportion of translating country bilingual, native in target* is the proportion of the adult population of the translating country that is native in the target language and at least conversational in the original language. *Proportion of translating country bilingual, native in original* is defined similarly.

All regressions include controls for the physical distance between original and translating countries (ln), altitude, biome, and climate differences, religious distance, linguistic distance, genetic distance, a dummy for the original and translating countries being contiguous and a dummy for the original country being the translating country. Columns 1 to 6 also include a dummy for the original language being widespread in the translating country. Standard errors are robust. Asterisks denote significance at: * p<0.10, ** p<0.05, *** p<0.01.

Data Appendix

A.1. Physical environment data

I generate three measures of difference in physical environment based on differences in the altitude profiles of original and translating countries, the biome region profiles of the countries, and the climate region profiles of the countries. All three measures are generated from the Center for International Earth Science Information Network's data set "National Aggregates of Geospatial Data Collection: Population, Landscape, and Climate Estimates, Version 2 (PLACE II)".[25]

For biome region and climate region, if information is missing for some fraction of the land area of the country, I rescale the non-missing data to sum to 100%.[26] Altitude data are effectively never missing.

A.1.1. Altitude profile distance

For each country, I calculate the proportion of the total land area that falls into each of three altitude zones: sea level to 100 meters above sea level, 100 meters to 800 meters, and 800 or more meters above sea level. These cutoffs were chosen to give a wide distribution across countries of proportion in each of the three regions. Globally, on average countries fall 35 percent into the lowest zone, 46 percent into the intermediate zone, and 19 percent into the highest zone. Denote the proportions of the translating country in each of the three altitude zones by alt_0^{trans}, alt_{100}^{trans}, and alt_{800}^{trans} respectively, and the equivalent proportions in the original country by alt_0^{orig}, alt_{100}^{orig}, and alt_{800}^{orig} respectively. Then I define the altitude profile distance between the countries to be:

$$AltitudeDist = 1 - \left[\min\left(alt_0^{trans},\ alt_0^{orig}\right) + \min\left(alt_{100}^{trans},\ alt_{100}^{orig}\right) + \min\left(alt_{800}^{trans},\ alt_{800}^{orig}\right)\right].$$

Note this distance measure takes the value 0 if the original and translating country both lie entirely in the same altitude zone (e.g., below 100 meters above sea level). Conversely, if the two countries lie entirely in different altitude zones (e.g., the original country lies entirely below 100 meters above sea level, and the translating country lies

---

[25] Available online at http://sedac.ciesin.columbia.edu/place/.

[26] Note the fraction of the country with missing data is never greater than 5%.

entirely between 100 and 800 meters), the distance between them is 1. In general, the measure captures the proportion of the two countries that lie in the same zone. To illustrate, the altitude profile distance between the Netherlands (almost entirely low-lying) and France (largely intermediate altitude, with some high and some low)) is 0.75, whereas the distance between the Netherlands and Switzerland (largely high altitude) is 0.99.

A.1.2. Biome region profile

Biome data are originally from the World Wildlife Fund (WWF) Terrestrial Ecoregions of the World dataset, which capture global terrestrial vegetation biodiversity patterns. The data classify land into one of 14 biome types:

1. tropical and subtropical moist broadleaf forests
2. tropical and subtropical dry broadleaf forests
3. tropical and subtropical coniferous forests
4. temperate broadleaf and mixed forests
5. temperate conifer forests
6. boreal forests/taiga
7. tropical and subtropical grasslands, savannas and shrublands
8. temperate grasslands, savannas and shrublands
9. flooded grasslands and savannas
10. montane grasslands and shrublands
11. tundra
12. Mediterranean forests, woodlands and scrub
13. deserts and xeric shrublands
14. mangroves

Biome region profile distance is defined analogously to altitude region profile distance. Specifically, denote the fraction of the translating country that falls into the i^{th} biome region by $biome_i^{trans}$ and the fraction of the original country that falls into the i^{th} biome region by $biome_i^{orig}$. Then the biome region distance between the original and translating countries is given by:

$$BiomeDist = 1 - \sum_{i=1}^{14} \min\left(biome_i^{trans},\ biome_i^{orig}\right).$$

A.1.3. Climate region profile

I use an aggregate version of the Köppen Climate Classification, in which land is classified as falling into one of five climate regions:

1. tropical
2. polar
3. temperate
4. cold
5. dry

I then define climate region profile distance analogously to biome region profile distance, but summing over these five categories instead of the 14 for biome region.

A.2. Cultural distance data

A.2.1. Religious distance

My data on religious distance are generated from the data on the religious distribution of the population of each country used by Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003). Note these distributions are for one point in time only, so the measures of religious distance I use do not vary by year. I aggregate up religions to the following categories:

1. Atheist
2. Anglican
3. African Christian
4. East Asian religions
5. Eastern Orthodox
6. Indian religions
7. Jewish
8. Muslim
9. Oriental Orthodox
10. Protestant

11. Roman Catholic
12. Christian not elsewhere classified
13. Numerous "traditional" religions that I regard as distinct from each other.

The primary measure of religious distance I use is the probability a randomly chosen person from the original country and a randomly chosen person from the translating country have a different religion, as classified into the categories above. For some countries, a (usually small) proportion of the population has missing religion. I assume alternatively that a randomly chosen person with missing religion a) is always of a different religion to any person from the other country (main specification) or b) has the same religion as a person from the other country with missing religion (results not presented). The two distance measures are highly correlated and regression results are unaffected. As a second alternative measure, I use an indicator variable for the most widespread religion in the two countries being the same. Because this last measure uses less of the variation in the data, regression results using it tend to be weaker statistically, but point in the same direction.

A.2.2. Linguistic distance

My primary measure of linguistic distance is based on the linguistic tree measure used by Fearon (2003)[27]. This measure of linguistic distance is intended to capture how long ago the two languages split from each other, which proxies for both the degree of dissimilarity of the languages, and the cultural distance that has evolved between the speakers of the languages.

My primary distance measure is generated as follows. First, each language is classified as in the 16th edition of Ethnologue. For example, Spanish is classified as follows:

- Indo-European
    - Italic
        - Romance
            - Italo-Western
                - Western

[27] Whom I thank for kindly sharing his data with me.

- Gallo-Iberian
- Ibero-Romance
- West-Iberian
- Castilian
- Spanish

and French is classified as:

- Indo-European
- Italic
- Romance
- Italo-Western
- Western
- Gallo-Iberian
- Gallo-Romance
- Gallo-Rhaetian
- Oïl
- French

Each of these categories (e.g., Gallo-Iberian) is considered a node on the language tree. I define the distance between two languages as

$$LinguisticDist_{ij} = 1 - \frac{2 \times CommonNodes_{ij}}{Nodes_i + Nodes_j}$$

where i and j denote the two languages, *CommonNodes* is the number of nodes they have in common (e.g., 6 in the case of Spanish and French), and *Nodes* is the number of nodes the individual language has (e.g., 10 in the case of Spanish). The distance between a language and itself is thus 0, and two entirely unrelated languages are distance 1 apart. In general two languages are further apart the smaller is their common ancestry relative to their overall evolution. French and Spanish, for instance, are somewhat related with a distance of 0.4 (= 1 – 12/20).

As an alternative measure of linguistic distance, I use the exact measure used in Fearon (2003), which is given by

$$LinguisticDist_{ij}^{Fearon} = \sqrt{\frac{15 - CommonNodes_{ij}}{15}}.$$

The significance of 15 here is that this is the maximum number of nodes any one language has in Ethnologue's classification scheme. The main difference between the two measures is that related languages with relatively few nodes in the language tree, such as Czech and Slovak, are considered relatively close according to my primary measure (0.2 for Czech and Slovak), but less close according to Fearon's measure (0.86 for Czech and Slovak). According to both measures, 80 percent of language pairs worldwide are distance 1 from each other. The two measures yield similar results in the regressions.

A.2.3. Genetic distance

I use Spolaore and Wacziarg's (2009) measure of genetic distance. This distance is defined at the country-pair level and captures the time elapsed since the two populations' last common ancestors. Where the population of a country consists of more than one genetically distinct group, the population-weighted average over the different groups is used.

A.2.4. Hofstede's (1980, 2001) cultural distance measure

My first survey-based measure of cultural distance is the variance-adjusted average of Hofstede's (1979, 1980, 1982, 1983, 2001) four cultural dimension measures: power distance, uncertainty avoidance, individualism, and masculinity.[28] These four dimensions were generated from surveys of 88,000 IBM employees in 53 different countries. They relate especially to values in the workplace, but are closely tied in to basic anthropological and societal issues (Hofstede and Bond, 1984).

The first dimension is "power distance", defined as "the extent to which less powerful members of institutions and organizations accept that power is distributed unequally." The second dimension is "uncertainty avoidance", or "the extent to which

[28] This method of combining Hofstede's dimensions was used previously by studies including Kogut and Singh (1988) and Ng, Lee, and Soutar (2007).

people feel threatened by ambiguous situations, and have created beliefs and institutions that try to avoid these." The third dimension is a continuum that ranges from "individualism", or "a situation in which people are supposed to look after themselves and their immediate family only," to "collectivism", or "a situation in which people belong to in-groups or collectivities which are supposed to look after them in exchange for loyalty." The fourth dimension is a continuum between "masculinity", or "a situation in which the dominant values in society are success, money, and things," and "femininity", or "a situation in which the dominant values in society are caring for others and the quality of life."

My measure of cultural distance based on Hofstede's cultural dimensions is given by

$$HofstedeDist_{ij} = \frac{1}{4}\sum_{k=1}^{4}\left[\frac{\left(I_i^k - I_j^k\right)^2}{Var_k}\right]$$

where i and j denote countries, k denotes the dimension, I_i^k is country i's value for dimension k, and Var_k is the variance across countries of the index for dimension k.

Differences between countries in these dimensions reflect differences in values, priorities, and accepted norms. Such differences may hinder translation flows from the supply side. Furthermore, they may mean original titles written in the countries are likely to encompass more different world views, which may make them more demanded in translation because they have no domestic substitutes, or less demanded because the ideas they contain are less acceptable.

A.2.5. Schwartz's (1994, 1999) cultural distance measure

My second survey-based measure of cultural distance is based on Schwartz's (1994, 1999) seven cultural value dimensions. Schwartz's framework is theory-driven, with elements derived from earlier work in the social sciences. The first of Schwartz's dimensions is "conservatism", defined as "a cultural emphasis on maintenance of the status quo, propriety, and restraint of actions or inclinations that might disrupt the solidary group or the traditional." Conservatism stands in opposition to two types of autonomy: autonomy in ideas and thought, called "intellectual autonomy", and autonomy in feelings and emotions, called "affective autonomy". Intellectual autonomy is defined as "a cultural emphasis on the desirability of individuals independently pursuing their own ideas and intellectual directions." Affective autonomy is "a cultural emphasis on the desirability of individuals independently pursuing affectively positive experience," such as pleasure, or an exciting or varied life.

The next dimension is "hierarchy", or "a cultural emphasis on the legitimacy of an unequal distribution of power, roles and resources," which has clear commonalities with Hofstede's power distance dimension. Hierarchy stands in opposition to "egalitarianism", or "a cultural emphasis on transcendence of selfish interests in favor of voluntary commitment to promoting the welfare of others."

The next is "mastery", meaning "a cultural emphasis on getting ahead through active self-assertion," which opposes "harmony", defined as "a cultural emphasis on fitting harmoniously into the environment."

I follow the approach of Ng, Lee and Soutar (2007), and construct an average distance on Schwartz's dimensions analogously to my average using Hofstede's dimensions, but where the sum is instead over the seven dimensions.

A.3. Bilingualism data

The variables on bilingualism are constructed using data from four Eurobarometer surveys.[29] Because the languages about which ability in was asked differ by survey, the original languages I include vary by country and year:

1. Eurobarometer 44.0, conducted in 1995 (translating countries: Austria, Belguim, Germany, Denmark, Spain, Finland, France, Italy, Netherlands, Portugal, Sweden; original languages: Arabic, German, Greek, English, French, Italian, Japanese, Dutch, Portuguese, Russian, Spanish, Swedish, Chinese);
2. Central and Eastern Eurobarometer (CEEB) 6, conducted in 1995 (translating countries: Albania, Belarus, Bulgaria, Czech Republic,

[29] All Eurobarometer data were downloaded from http://zacat.gesis.org.

Estonia, Hungary, Kazakhstan, Lithuania, Latvia, Macedonia, Poland, Romania, Slovakia, Slovenia, Russia, Ukraine; original languages: Azerbaijani, Belarussian, Bulgarian, Czech, German, Greek, English, Estonian, French, Hungarian, Italian, Kazakh, Latvian, Lithuanian, Macedonian, Polish, Romanian/Moldovan, Slovenian, Spanish, Swedish, Tartar, Turkish, Ukrainian, Uzbek);

3. Eurobarometer 54LAN, conducted in 2000 (translating countries: Austria, Belgium, Germany, Denmark, Spain, Finland, France, Italy, Netherlands; original languages: Arabic, German, Greek, English, French, Irish, Dutch, Portuguese, Spanish, Swedish, Turkish, Chinese); and
4. Candidate Countries Eurobarometer 2001.1, conducted in 2001 (translating countries: Bulgaria, Czech Republic, Estonia, Hungary, Lithuania, Latvia, Poland, Romania, Slovakia, Slovenia, Turkey; original languages: Bulgarian, Czech, Greek, Estonian, Latvian, Lithuanian, Maltese, Polish, Romanian/Moldovan, Russian, Slovak, Turkish).

The first two data sources are used to construct variables for use with the 1994 translation data, and the second two with the 1999 translation data. In each case, I calculate my variables using the language ability of the population aged 18 and older, and weight observations within the country using the survey weights that match the sample to the population in terms of demographics.

I construct the bilingualism variables using two questions from each of the surveys: "What is your mother tongue?" and "Which [of these] languages can you speak well enough to take part in a conversation, except your mother tongue?"[30] An individual is defined as bilingual and native in the target language if he specifies the target language as his mother tongue, and the original language as a second mother tongue (where permitted) or as an additional language in which he is conversational. Bilingual and native in the original language is defined similarly.

[30] Instead of the latter question, Eurobarometer 54LAN asks, "What other languages do you know?" and then asks how well the respondent speaks each mentioned language. I assume the mention of an additional language implies conversational ability in it.

Appendix Table A: Closer countries translate more from each other (OLS)

Dependent variable: ln number of translations + 1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Original languages:	top 100	top 100	top 100	top 100	unambiguous original country	main research languages	top 100, assigned to nearest major country	top 100	top 100	European of top 100
Target languages for each translating country:	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country	official in whole country, translating country is main country	official and widespread in country	official in whole country, European country
Variable										
Physical distance between original and translating countries (ln)	-0.116***	-0.114***	-0.129***	-0.110***	-0.039***	-0.374***	-0.110***	-0.145***	-0.120***	-0.203***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.005)	(0.080)	(0.008)	(0.013)	(0.007)	(0.035)
Population of translating country (ln)	0.043***	0.042***								
	(0.004)	(0.004)								
GDP per capita of translating country (ln)	0.072***	0.072***								
	(0.005)	(0.005)								
Population of original country (ln)	0.030***	0.030***								
	(0.002)	(0.002)								
GDP per capita of original country (ln)	0.133***	0.131***								
	(0.006)	(0.006)								
Original country colonised translating country		0.224***								
		(0.076)								
Translating country colonised original country		0.083								
		(0.072)								
Original and translating countries are contiguous				0.131***	0.127***	0.193	0.131***	0.171***	0.079***	0.186***
				(0.035)	(0.032)	(0.178)	(0.035)	(0.045)	(0.029)	(0.060)
Original language is widespread in translating country				0.027	.	0.070	0.027	-0.525***	0.191**	-0.447**
				(0.133)	.	(0.326)	(0.133)	(0.161)	(0.089)	(0.204)
Original country is translating country	-0.174***	-0.164**	-0.038	-0.002	0.233***	.	-0.002	0.636***	-0.231**	0.652**
	(0.064)	(0.064)	(0.060)	(0.145)	(0.073)	.	(0.145)	(0.226)	(0.093)	(0.308)
Time-varying target language/country fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying original language fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.150	0.153	0.568	0.570	0.372	0.862	0.570	0.695	0.524	0.746
Observations	12,434	12,434	13,262	13,262	9,205	517	13,262	6,637	15,187	2,316
Translating countries	56	56	58	58	58	58	58	37	56	31

Notes: This table duplicates the results in Table 1, but uses OLS and predicts ln(translations + 1) instead of using PML. See the notes to Table 1 for further details. Standard errors are robust. Asterisks denote significance at: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table B1: The effect of trade on translations over time: consistent countries

Dependent variable: number of translations (ln)								
Variable	1964	1969	1974	1979	1984	1989	1994	1999
Physical distance between original and translating countries (ln)	-0.723***	-0.550***	-0.694***	-0.630***	-0.634***	-0.442***	0.174	0.370***
	(0.148)	(0.131)	(0.115)	(0.153)	(0.111)	(0.114)	(0.154)	(0.137)
Difference between altitude profiles of original and translating countries	-0.896**	-0.671*	-0.273	-0.228	-0.632	-0.821**	-0.500	-0.451
	(0.398)	(0.401)	(0.413)	(0.535)	(0.408)	(0.359)	(0.423)	(0.426)
Difference between climate region profiles of original and translating countries	-0.124	0.255	0.208	0.067	0.112	-0.056	-0.016	0.122
	(0.284)	(0.240)	(0.207)	(0.233)	(0.257)	(0.221)	(0.363)	(0.306)
Difference between biome region profiles of original and translating countries	1.025***	0.302	0.485**	0.304	0.473**	0.491**	-0.437***	-0.354**
	(0.258)	(0.238)	(0.244)	(0.282)	(0.238)	(0.228)	(0.159)	(0.139)
Religious distance	-0.763**	-0.441	-0.488	-0.770**	-0.790**	-1.108***	-1.368***	-0.973***
	(0.310)	(0.322)	(0.331)	(0.327)	(0.345)	(0.313)	(0.342)	(0.330)
Linguistic distance	-1.347***	-2.111***	-1.695***	-1.478***	-1.297***	-1.251***	-0.606**	-0.899***
	(0.326)	(0.336)	(0.284)	(0.323)	(0.281)	(0.241)	(0.277)	(0.259)
Genetic distance	0.034	0.148	0.839**	1.171***	1.205***	0.994***	0.831**	0.215
	(0.324)	(0.333)	(0.335)	(0.444)	(0.336)	(0.262)	(0.400)	(0.236)
Imports into target country from original country	0.232*	0.269**	0.165*	0.115***	-0.013	-0.046	0.194*	0.207**
	(0.119)	(0.111)	(0.097)	(0.037)	(0.079)	(0.083)	(0.101)	(0.095)
Exports from target country into original country	0.025	0.023	0.169*	0.121	0.340***	0.298***	0.221**	0.382***
	(0.080)	(0.086)	(0.087)	(0.081)	(0.085)	(0.089)	(0.103)	(0.089)
Original and translating countries are contiguous	0.032	-0.144	0.013	0.043	-0.088	0.165	-0.094	-0.243**
	(0.221)	(0.246)	(0.191)	(0.192)	(0.170)	(0.153)	(0.133)	(0.102)
Original language is widespread in translating country	-0.359	0.408	0.795	-0.024	0.307	0.306	1.749*	1.564*
	(0.626)	(0.746)	(0.585)	(0.639)	(0.795)	(0.959)	(1.001)	(0.887)
Original country is translating country	-0.200	2.299	-0.207	-0.038	8.701***	-0.050	0.056	9.348***
	(1.576)	(1.728)	(1.642)	(1.797)	(1.447)	(1.677)	(2.057)	(1.770)
Target language/country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Original language fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummy variables for imports are zero and for exports are zero	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,769	1,769	1,769	1,769	1,769	1,769	1,769	1,769
Translating countries	17	17	17	17	17	17	17	17

Notes: This table presents the results of PML regressions (as described in section 3) of the number of translations from an original language to a target language in a translating country in a year. The same original languages are included for every target language; zero values are included in the estimation, as allowed by the PML procedure. The original languages are all those languages in the most widely spoken 100 languages worldwide that are ever translated. The target language/countries included are all the languages that are official in the whole of the translating country. Each column presents regression results for a different year, as given in the column header. The same target countries are used each year.

The altitude profile, climate region profile, and biome region profile difference variables are all constructed to vary between 0 (no overlap in profiles) and 1 (identical profiles). Religious distance is the probability a randomly chosen person from the translating country and a randomly chosen person from the original country have the same religion. Standard errors are robust. Asterisks denote significance at: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table B2: The effect of trade on translations over time: all available countries

Dependent variable: number of translations (ln)

Variable	1964	1969	1974	1979	1984	1989	1994	1999
Physical distance between original and translating countries (ln)	-0.882***	-0.430***	-0.598***	-0.510***	-0.447***	-0.417***	0.012	0.190*
	(0.191)	(0.106)	(0.115)	(0.101)	(0.088)	(0.082)	(0.093)	(0.107)
Difference between altitude profiles of original and translating countries	-0.031	-0.211	-0.124	0.046	0.374	-0.002	-0.264	-0.000
	(0.550)	(0.374)	(0.415)	(0.337)	(0.343)	(0.231)	(0.295)	(0.254)
Difference between climate region profiles of original and translating countries	0.159	0.232	0.109	0.180	0.175	-0.212	0.188	0.269
	(0.283)	(0.200)	(0.189)	(0.202)	(0.178)	(0.164)	(0.218)	(0.198)
Difference between biome region profiles of original and translating countries	0.223	0.312	0.382*	0.228	0.008	0.477***	-0.457***	-0.535***
	(0.232)	(0.210)	(0.198)	(0.223)	(0.203)	(0.178)	(0.116)	(0.102)
Religious distance	-0.803**	-0.614**	-0.400	-0.841***	-1.274***	-1.221***	-0.753***	-0.787***
	(0.319)	(0.299)	(0.304)	(0.285)	(0.252)	(0.224)	(0.275)	(0.217)
Linguistic distance	-1.765***	-1.486***	-1.816***	-1.153***	-0.771***	-0.517***	-0.459**	-0.386**
	(0.307)	(0.256)	(0.287)	(0.210)	(0.180)	(0.178)	(0.227)	(0.167)
Genetic distance	0.701**	0.076	0.798***	0.656***	0.413***	0.570***	0.044	0.053
	(0.346)	(0.238)	(0.187)	(0.203)	(0.140)	(0.165)	(0.186)	(0.151)
Imports into target country from original country	0.113	0.225***	0.071	0.112***	-0.077	-0.023	0.234***	0.364***
	(0.095)	(0.082)	(0.079)	(0.034)	(0.051)	(0.047)	(0.057)	(0.060)
Exports from target country into original country	0.122*	0.003	0.256***	0.086*	0.271***	0.208***	0.213***	0.216***
	(0.073)	(0.068)	(0.074)	(0.050)	(0.065)	(0.058)	(0.069)	(0.059)
Original and translating countries are contiguous	0.064	0.041	-0.022	0.235*	0.132	0.333***	0.210*	0.170*
	(0.163)	(0.149)	(0.156)	(0.121)	(0.107)	(0.093)	(0.126)	(0.092)
Original language is widespread in translating country	-0.731	0.557	0.857	0.472	0.742***	1.124***	1.166***	1.167***
	(0.607)	(0.614)	(0.548)	(0.287)	(0.264)	(0.217)	(0.205)	(0.236)
Original country is translating country	6.933***	7.619***	-0.087	4.210***	6.605***	5.704***	3.443*	7.576***
	(1.629)	(1.776)	(1.444)	(1.157)	(1.102)	(0.664)	(1.951)	(0.915)
Target language/country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Original language fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummy variables for imports are zero and for exports are zero	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,590	3,829	3,456	6,178	6,555	6,177	6,461	5,898
Translating countries	34	33	30	49	51	46	54	51

Notes: This table presents the results of PML regressions (as described in section 3) of the number of translations from an original language to a target language in a translating country in a year. The same original languages are included for every target language; zero values are included in the estimation, as allowed by the PML procedure. The original languages are all those languages in the most widely spoken 100 languages worldwide that are ever translated. The target language/countries included are all the languages that are official in the whole of the translating country. Each column presents regression results for a different year, as given in the column header. All available target countries are used each year.

The altitude profile, climate region profile, and biome region profile difference variables are all constructed to vary between 0 (no overlap in profiles) and 1 (identical profiles). Religious distance is the probability a randomly chosen person from the translating country and a randomly chosen person from the original country have the same religion. Standard errors are robust. Asterisks denote significance at: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.