
# Economics of Literary Translation. A Simple Theory and Evidence

Victor Ginsburgh, Shlomo Weber
and Sheila Weyers

NOTA DI LAVORO 12.2008

**FEBRUARY 2008**

KTHC – Knowledge, Technology, Human Capital

Victor Ginsburgh, *ECARES, Université Libre de Bruxelles and CORE, Université catholique de Louvain*
Shlomo Weber, *Southern Methodist University, Dallas and CEPR*
Sheila *Weyers, Université catholique de Louvain*

# Economics of Literary Translation. A Simple Theory and Evidence


## Summary

Books are an important factor of cultural transmission, but need, in most cases, to be translated. According to some authors, this may lead to a form of cultural domination of English. The population speaking English as a first language is, with the exception of Mandarin, the largest in the world. It is therefore not surprising that English produces more fiction (and much more scientific literature, as scientists from all countries write in English with increasing frequency) than any other language. We develop a theoretical model of translation, which is estimated on the basis of UNESCO translation data. We show that translations from English are dominated by translations from other languages, including Scandinavian ones and French.



**Keywords:** Languages, Translations, Cultural and Linguistic Distances

**JEL Classification:** L82, Z11

*We are grateful to Georg Kirchsteiger and Abdul Noury for comments. We are also grateful to Alain Brion and Mauro Rosi from UNESCO, who helped us to retrieve the data.*



*Address for correspondence*:

Shlomo Weber
Department of Economics Southern Methodist
University Dallas
TX 75275-0496
USA
E-mail: sweber@mail.smu.edu

## 1. Introduction

The literature on cultural transmission deals essentially with the media industries, and especially with movies and television programs.[1] Much less is written on music, that does not need translation, dubbing or subtitles. But as pointed out by Frith (1996, p. 157) well before the explosion of Internet, MP3 and iTunes, "the point is not that a new technology enabled--determined?--a new music international, but, rather, that the music's own essential mobility enabled the new technology to flourish, and shaped the way it worked."

People do not only watch television, movies, or listen to music, they also read. As of May 2006, sixty million copies of Dan Brown's *Da Vinci Code*, published in early 2003 were in print or sold. [2] Books are an important factor of transmission, but need in most cases to be translated. In fact, *Da Vinci Code*, was translated into 44 languages, and by October 2004, it had generated some sixteen titles supporting or debunking the code.

Though television and broadcasting have changed considerably the way "culture" is transmitted, books (and more generally written material, including the web) remain essential. As Susan Sontag pointed out while receiving the Peace Prize at the German Book Trade at the Frankfurt Book Fair in 2003:

> "[W]hat saved me as a schoolchild in Arizona, waiting to grow up, waiting to escape into larger reality, was reading books, books in translation as well as those written in English. To have access to literature, world literature, was to escape the prison of national vanity, of philistinism, of compulsory provincialism, of inane schooling, of imperfect destinies and bad luck. Literature was the passport to enter a larger life; that is, the zone of freedom."

However, translations are sometimes accused of leading to a form of cultural domination by some languages. According to Mélitz (2007), "if one language is sufficiently larger than others in the sales of original-language works, it will tend to crowd out the rest in translations…[and] those writing in the dominant language are privileged." A similar opinion was recently expressed in one of the important French literary bi-monthly, *La Quinzaine Littéraire* (2006), claiming that translations from English into French dominate in France.[3] The title of the article is unequivocal: "Fiction is American." Ganne and Minon (1992) show that France, Italy, Spain and Germany translate much more (18, 25, 26 and 15%) than the United Kingdom (3.3%). They attribute this to the "abundance of books that originate in the United States," and that need no translation in the UK. They also show that English is the

---

1 See Hoskins, McFadyen and Finn (1997) and the list of references therein.

2 http://en.wikipedia.org/wiki/The_Da_Vinci_Code (accessed November 26, 2005).

3 This is reminiscent of the literature on the "American hegemony" following the popularity of the television series *Dallas* in Europe in the seventies and eighties (see Bilteryst, 1991).

language that generates the largest number of translations in France, Italy, Spain and Germany. Heilbron (1999) describes the system as accounting for uneven flows between languages groups: On the European continent, 50 to 70 percent of the published translations are being made from English.

To simplify the discussion, let us make the following two rough and "all other things equal" (such as literacy rates) assumptions: (a) the number of books written in a language is proportional to the population that speak it, and (b) every reader reads the same number of books. Then it will be the case that (a) more books, available for translation, will be written in large countries (or languages that are native to many speakers) and (b) large countries will need less translations than small ones, because more books are available in their own language. If this were true, then the previous discussion seems to underestimate the fact that English is the mother tongue of some 400 million people (Crystal, 2001). It is therefore not surprising that English produces more fiction (and science) than any other European language, and that more books are translated from English. Table 1, which tabulates the number of literary works translated from some European languages illustrates that this is (roughly) the case. Indeed, even though a large number of books is translated from English, and fewer are translated into English, there are almost as many that are translated from French, and many more from Danish, Norwegian and Swedish. Likewise, "small" languages such as Danish, Estonian, Finnish, Icelandic, Norwegian or Slovene generate a lot of translations, while languages spoken by large populations, such as English, yield as little as 0.043 translated books per thousand readers.

These considerations ignore another important factor in comparing the number of translations: the role of cultural proximities. Except for the sake of exoticism, a thriller that features New York is more likely to be translated from English into French than a Chinese or an Estonian thriller that features Shanghai or Tallinn. Just think of how hard it is to read Dostoiewski or Tolstoi, before trying to get accustomed to Chinese or Estonian names of characters and streets. Our theoretical and empirical investigations are in line with the view that, since more novels are produced in English, which is culturally closer to other Indo-European languages than Mandarin, Arabic or Hindi, more should be translated from English than from other languages, even though these languages have large numbers of native speakers (e.g., Mandarin Chinese is spoken by 1.2 billion people, Crystal, 2001). This argument is reinforced by the costs aspect of translation. De Swaan (2001, p. 45) estimates them at 30 percent of the price of a 300-page book of which 2,000 copies ("certainly not a too conservative estimate," according to De Swaan) are circulated. Publishers will seldom translate books that have a low probability of being read. The role of cultural distances makes it necessary to take into account bilateral translations, rather than the marginal totals presented in Table 1.

The arguments raisedd above do not support the claim that the number of translations from English is *overall dominating* and *disproportionately* large. In fact, some authors even

suggest that the share of English in printed and electronic media is declining. Melitz (2007, Table 1, and p. 212) shows data suggesting that the role of English in literary works is decreasing over the last 30 years, while Pfanner (2007) indicates that only 36% of all blog postings on the Internet are in English, while 37 percent are written in Japanese and the share of Chinese and Spanish blogs is rapidly increasing.

The economic literature on translations including Melitz's (2007) seminal paper is quite small. The word "translation" appears nowhere in *Books* (Coser, Kadushin and Powell, 1982), or in the very comprehensive survey on the book industry by Canoy, van der Ploeg and van Ours (2006). Caves (2000) discusses books at great length without dwelling on translations. The closest to our model is the framework of Hjorth-Andersen (2001), who estimates a three-equation model of translations, where the first equation identifies the total number of titles in a given country, the second determines the aggregate propensity to translate, and the third disaggregates this total into single languages.

The paper is organized as follows. The theoretical model is discussed in Section 2. It leads to demand equations for translations by a representative reader that lend themselves to econometric estimation. Empirical results, described in Section 3, shed some light on the determinants of translations of fiction and other publications, and show that the hypothesis of the dominance of English is not well founded. Section 4 is devoted to some concluding comments.

## 2. The Theoretical Model

Consider a society (world) where citizens speak languages from the set $Q = \{1, 2, ..., k, ..., q\}$. Suppose that every citizen knows one and only one native language in $Q$. Denote by $P_j$ the population of those citizens whose native language is $j$, and by $L_j$ and $W_j$ their literacy rate and average income, respectively. Citizens who know language $j$ have access to books translated from other (foreign) languages $i = 1, 2, j-1, j+1, ..., q$. Subscript $i$ will denote a source language from which a book is translated into $j$, the so-called destination language. We assume that for every language $j$ there is a representative (average) reader $A_j$ who speaks $j$ and spends $R_j$ hours reading translated books. Let $t_{ij}$ be the number of titles translated from language $i$ that $A_j$ reads. We assume that:

*Assumption 1*: $R_j$ is an increasing function of the average literacy rate $L_j$ and average income $W_j$ of population $P_j$.

It is often more difficult to adjust to novels entrenched in different cultures. Therefore, we assume that it takes longer to read books translated from languages that are culturally more distant. Denote the cultural distance between languages $i$ and $j$ by $D_{ij}$. Then:

*Assumption 2*: There is a positive constant $r$ such that it takes $A_j$ $r(1 + D_{ij})$ hours to read a book translated from $i$ to $j$.

Notice that $A_j$'s reading time constraint can be formulated as:

$$r(1+D_{1j})t_{1j} + ...+ r(1+D_{j-1,j})t_{j-1,j} + r(1+D_{j+1,j})t_{j+1,j} ... + r(1+D_{qj})t_{qj} = R_j. \quad (1)$$

*Assumption 3*: $A_j$'s preferences for foreign books translated from languages different from $j$ are represented by the following Cobb-Douglas[4] utility function with $q-1$ variables:

$$u(t_{1j},t_{2j},...,t_{j-1,j},t_{j+1,j},...,t_{qj}) = t_{1j}^{\gamma_1} \times t_{2j}^{\gamma_2} \times ... \times t_{j-1,j}^{\gamma_{j-1}} \times t_{j+1,j}^{\gamma_{j+1}} \times ... \times t_{qj}^{\gamma_q} . \quad (2)$$

*Assumption 4*: For every language $i$ , the value $\gamma_i = \gamma(P_i)$, where $\gamma$ is an increasing function of the population $P_i$.

Assumption 3 offers a simple functional form for the utility derived from translated books. The representative reader has access to translations from all languages, including those from distant cultures giving her the possibility to learn about all possible cultures. Assumption 4 is meant to represent the influence of the source language $i$. If the number of writers per head in population $P_i$ is equal across languages, then the number of books written in a language is an increasing function of the number of its speakers.

For every pair of languages, $i$ and $j$, the average reader $A_j$ chooses the number $t_{ij}^*$ of foreign titles in $i$ that she will read by maximizing her utility $u_j(.)$ under her reading time constraint (1). This leads to the following demand functions:

$$t_{ij}^* = \frac{R_j}{r(1+D_{ij})}\Gamma_i , \quad (3)$$

where $\Gamma_i = \gamma_i / \Sigma_{k \neq j}^q \gamma_k$ .[5]

---

[4] Note that this assumption can easily be relaxed.

[5] Note that in this formulation $t_{ij}$ is decreasing in the distance between languages. One could argue that when languages are very close, there is no need for translation as both populations can read each other's books in the native language. The number of translated titles would then increase with the linguistic distance up to a point, and eventually decline when the linguistic gap between two languages becomes large. It is easy to develop a model that would reproduce such an inverted-U shape for the relation between number of titles translated and distance. However, we could find no evidence for this in the empirical results of Section 3, and did not pursue the idea. The reason is probably due to the fact that the distance for which the inverted-U curve peaks is quite low. The closest languages in our database are Slovak and Czech, and there are books translated between these two languages.

The link between the total number of titles translated from $i$ to $j$ and the preferences of the average reader is determined by geographical and cultural diversity within population $P_j$. Indeed, if the population is perfectly homogeneous, all readers will read the same titles, and the total number of titles translated from $i$ to $j$ will be equal to $t_{ij}^*$. This will happen if the population $P_j$ is small with a small number of literary critics who recommend the same books. Word of mouth between the few readers who do not live far away from each other will go in the same direction. In the other extreme case of a completely heterogeneous (and large) population $P_j$, in which each reader lives on an "island" and reads different titles, the total number of titles translated from $i$ to $j$ will be equal to $P_j\ t_{ij}^*$. In other words, when $P_j$ is small, the number of titles read there will be small. When $P_j$ is large and diverse (as in the case of English or Spanish whose speakers are scattered across countries and continents), there will be less information flowing between sub-regions, local populations will be more isolated from each other, the number of newspapers carrying literary criticisms will be larger, and the number of translated titles will be relatively large. Therefore, it is reasonable to represent the size and diversity of population $P_j$ by a heterogeneity index $H_j$,[6] which is an increasing function of the population size $P_j$. We then assume:

*Assumption 5*: The total number of titles translated from $i$ to $j$ is given by $t_{ij} = H_j t_{ij}^*$.

Demand functions for all languages $i \neq j$ can now be fully specified as

$$t_{ij} = H_j(P_j)\frac{R_j(L_j, W_j)}{r(1+D_{ij})}\Gamma_i, \tag{4}$$

where $\Gamma_i = \gamma_i(P_i) / \Sigma_{k\neq j}^{q}\gamma_k(P_k)$. It is easy to show that they satisfy the following properties.

*Proposition*: Under Assumptions 1 to 5, the number of titles translated from $i$ to $j$ is

(a) increasing in $P_j$, the population whose native language is $j$,
(b) increasing in $P_i$, the population whose native language is $i$,
(c) decreasing in $D_{ij}$, the linguistic (or cultural) distance between languages $i$ and $j$,
(d) increasing in $L_j$, the literacy level of the population $P_j$,
(e) increasing in $W_j$, the income level of the population $P_j$.

## 3. Empirical Results

The theoretical model leads us to estimate the following equation:

---

[6] For details, see Alesina *et al.* (2003), Bossert, d'Ambrosio, and La Ferrara. (2006) and Desmet, Ortuno-Ortin and Weber. (2007).

$$\ln t_{ij} = \alpha_1 \ln P_i + \alpha_2 \ln P_j + \alpha_3 \ln D_{ij} + \alpha_4 \ln L_j + \alpha_5 \ln W_j + \alpha_6 + v_{ij}, \quad (5)$$

where $t_{ij}$ is the number of translations from language $i$ to language $j$, $P_i$ and $P_j$ are the sizes of the populations that speak $i$ and $j$ as first language,[7] $D_{ij}$ is the distance between $i$ and $j$, $L_j$ and $W_j$ represent the literacy rate and the average income of the population speaking the destination language $j$, and $v_{ij}$ is an error term. The $\alpha$ are parameters to be estimated; they can easily be obtained from combinations of the parameters of (4). The variables are the same as those of the theoretical model. The Proposition implies that $\alpha_1$ and $\alpha_2$ should be positive, while $\alpha_3$ should be negative. Literacy and income of the population speaking the destination language are expected to have a positive influence, since more books will be read (and thus translated from other languages) in more literate and richer regions; $\alpha_4$ and $\alpha_5$ should therefore be positive.

The data on distances between languages that we take to represent distances between cultures[8] requires some explanation. The measure we use is based on so-called cognate data, established as follows. For each meaning from a list of 200 basic meanings (such as father, mother, digits, etc.) selected by Swadesh (1952), Dyen, Kruskal and Black (1992) collected the words used in 95 Indo-European speech varieties (i.e., languages and dialects) and classified these into cognate classes. For a given meaning, such a class contains all the words from different speech varieties that have an unbroken history of descent from a common ancestral word.[9] The distance between two languages $i$ and $j$ is then equal to the percentage of words in the two languages which *do not* descend from a common word. This distance will thus be close to 1 if the two languages have completely different roots (say English and Finnish, a non Indo-European language) and close to 0 otherwise (Slovak and Czech). In our context, this distance is not meant to measure the difficulty of translating from $i$ to $j$, which is more complex than just the relative proximities of vocabularies.[10] We rather follow Cavalli-Sforza (2000) and assume that linguistic distances are a proxy of cultural distances.

---

[7] We assume that all writers (and readers) write (and read) in their mother tongue. Writers such as Conrad, Nabokov, or Becket who emigrated and switched from their native language to the one of the country of immigration remain the exception.

[8] Though cultural distances are available to some extent (see Geert Hofstede, 1980, 1991, as well as Hofstede's websites http://spitswww.uvt.nl/web/iric/hofstede/page3.htm and http://geert-hofstede.international-business-center.com/index.shtml), we will use the linguistic distances computed by Dyen, Kruskal and Black (1992). The reason is twofold: (a) cultural distances are available for countries, not for languages, while UNESCO data are for languages; (b) cultural distances exist only for a small number of countries, and certainly not all those that are in our sample of translations between languages.

[9] Words borrowed from an other language are thus excluded.

[10] See e.g. Catford (1967) and Nida and Taber (1969). Nida (1975, p. 98) is very explicit about the two questions translation is confronted with: "The first concerns translation as an art rather than a science and the second raises the issue as to whether translation is even possible." The motto "traduttore, traditore" is well-known. The German poet Heine claimed that his poems translated into French, were just "moonlight stuffed with straw," and Nabokov who used to write indifferently in English and Russian notes (in 'On translating Eugen Onegin', one of his poems) that translation is "On a platter a poet's pale and glaring head, a parrot's screech, a monkey's chatter, and profanation of the dead." In ancient times, translation (of God's words) was blasphemy. The *Roll of Fasting* (first century A.D.) "records the belief that three days of utter darkness fell on the world when the Law was translated into Greek." See Steiner (1992, pp. 251-252)

Note the model described by Equation (5) is similar to the one that has been successfully used to describe international trade flows (and migrations) between countries, where $t_{ij}$ then represents trade flows, $P_i$ and $P_j$ are total exports originating in $i$ and total imports by $j$, while distances $D_{ij}$ are measured in several ways, including geographic distances and other closeness indicators between countries (such as free trade zones, common languages, etc.).[11]

*Data*

The sources for the data used are as follows. The number of translated titles is taken from the UNESCO database. Crystal (2001) provides population data for each language. Distances between languages are borrowed from Dyen, Kruskal and Black (1992). Finally, literacy rates can be found in UNESCO, Institute for Statistics (2002) and as income variable, we used per head gross national products in the various destination countries (World Bank, 2005). Further details are given in appendix.

*Results*

Results for literary translations between 1979 and 2002 appear in Table 1. The first equation gives the results of (5), ignoring literacy rate and income in the destination language population. All parameters are significant at the one percent level, and carry the expected signs. Since they can be interpreted as elasticities, they show that a one percent increase in population of the source language increases the number of translations by 0.76 percent. The elasticity with respect to the destination language is much smaller (0.35). This is due to the fact that more books are written in languages that are spoken by many. Therefore, populations that speak these languages are more self-sufficient and less affected by translations from other languages. On the other hand, fewer titles are translated into languages spoken by smaller and more homogeneous populations. Returns to scale are strongly decreasing in the population of destination. The elasticity of the number of translations with respect to distance is not significantly different from -1. In the second equation we add literacy rate and income per head in the country of destination. As can be seen, both are significantly positive, as expected, and the elasticity with respect to the literacy rate is quite important (3.65), while the other parameters remain similar to those of the previous equation.

Both equations explain a little over 40 percent of the total variance of (the log of) the number of translations. The residual variance can be reduced in a significant way if one is ready to distinguish the effect of the various source languages. This is done in the third

[11] See Fotheringham (1981, 1984) for a discussion of the model in this context. See also Anderson and van Wincoop (2004).

equation where each of the 19 source languages is represented by a dummy variable (which takes the value one if a book is translated from that specific language, and zero otherwise). Each dummy is multiplied by (the log of) the population that speaks the language. This changes neither the specification, nor the values of the other parameters, but it frees the parameter picked by the dummy from the effect of the population that speaks the language, and makes the values of the parameters directly comparable.[12] The parameters can be thought of as elasticities. Thus, a one percent increase in the source-population that writes in Norwegian increases by 1.4 percent the translations from Norwegian, whereas this percentage is 0.17 in the case of Serbo-Croatian. On average, the elasticity does not differ that much from the value 0.76 that appears in the first two equations. The interesting point here is that it allows ranking source languages, and as can be seen, English is far from being the first, and is inferior to Norwegian, Danish, Swedish and (even) by French in this regard. If all languages had the same number of speakers, there would, for example, be 1.40/1.09 = 1.28 times more books translated from Norwegian than from English. There are also relatively more books translated from French than from English (though the difference between the two parameters is statistically not significant). Quite surprisingly, this refutes the often-stated claim of dominance of translations from the English language. Though Russian, Spanish and Portuguese are spoken by large populations, their role is often dwarfed by languages spoken in smaller countries, such as Finnish, Czech and Hungarian. The model however does not offer an explanation of this effect.[13]

In the last equation, we partition languages into groups, so as to keep the residual variance little or not affected (note that the adjusted R-squared even slightly increases between the third and the fourth equation). Scandinavian languages form the first group; English and French are on equal foot in the second group, etc.[14]

Table 2 exhibits the results of the same approach applied to translations 1979-2002 for "all other" (non-literary) books. While one has to be cautious concerning homogeneity of the data (see Appendix), results are comparable to the previous ones on literature, except that literacy rates are no longer significant. The first group consists of the cluster formed by German, French and English, but English is not first either. This is because English is more and more used as the language of science, and does not need to be translated, since most scientists can and do communicate in English.

---

[12] An alternative method that would give exactly the same results would be to run the regression with dummies, and then divide the coefficients by (the log of) the populations.

[13] Georg Kirchsteiger suggested that this may be due to the fact that some writers in smaller countries become very famous, get heavily translated during some periods (including the one under review), and disappear after some time. This can hardly happen in larger countries in which the number of authors is larger, since they get replaced more easily by others, given that the pool of writers is larger. Unfortunately, the data are aggregate (number of titles translated) and do not allow to check whether the number of translations is due to many authors or whether it is the consequence of few prolific authors getting translated because they are well known.

[14] Note that a much simpler calculation that does not take into account distances, literacy rates and incomes can be performed, by merely dividing the number of literary books translated by the number of speakers in the source language. For English, this amounts to 950 books per million speakers of English. For French, this number is 883.

A second question considered here is whether the situation changed during the 23 years under review. We examine three periods: 1979-1987, 1988-1997, and 1998-2002 and performed the same analysis as above on each of the sub-periods. The results that appear in Table 3 are rankings based on the parameters of dummy variables representing source languages. The first column ranks the 19 source languages according to increasing order of rank over the whole period, while the three following columns give the rank in each of the sub-periods. As an example, for literature, French is ranked fourth over the whole period, and this hardly changes over time (4, 4 and 4.5 in each of the sub-periods).[15] Roughly speaking, for literature, there are no changes: the six first languages (Norwegian, Danish, Swedish, French, English, and German) during the early 1980s are still so in the early 2000s. Finnish, Italian and Dutch seem to have gained somewhat, while Eastern European languages, including Russian, have lost some ground. For "all other" books, German, French and English were ranked 1, 2, 4 in the early 1980s and are tied for the second place. The influence of Slavic languages declined dramatically over time, following the loss of influence of former Socialist countries.

## 4. Conclusions

We construct and estimate a model that offers some insight into the determinants of literary and other translations. Though the estimated resulting equations are very close to the well-known gravity model used in many international trade applications, their theoretical roots are derived from a simple demand for books equation. We show that the model fits well the data and that conclusions of English (American) language hegemony in literature are based on incomplete reasoning. In this respect, the number of books that are translated from one language to another is not necessarily an accurate indicator of the power of a language. The model should take into account the number of books written in the source language, as well as the cultural distances between languages. It is obvious that the more titles are written in a language, the more will be translated into other languages, as long as cultural traits are similar. If they are not, cultural distances will also play a role: the smaller the distance, the larger the number of translations. Once the number of titles translated between languages takes the two factors into account, the English language hegemony hypothesis ceases to hold.

[15] The ranks are approximations resulting from four different regressions, and nothing ensures that they are fully consistent. English is for instance no. 5 over the whole period, and is no. 4, 4 and 4.5 in each sub-period. But the results are usually consistent, and if they are not, the inconsistency is minor.

## 5. References


Alesina, Alberto, Arnaud Devleeschawer, 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.

Biltereyst, Daniel (1991), Resisting American hegemony: A comparative analysis of the reception of domestic and US fiction, *European Journal of Communication* 6, 469-497.

Bossert, Walter, Conchita d'Ambrosio and Eliana La Ferrara (2006), A generalized index of fractionalization, mimeo.

Canoy, Marcel, Rick van der Ploeg and Jan van Ours (2006), The economics of books, in Victor Ginsburgh and David Throsby, eds., *Handbook of the Economics of Arts and Culture*, Amsterdam: Elsevier.

Catford, J. C. (1967), *A Linguistic Theory of Translation*, Oxford: Oxford University Press.

Cavalli-Sforza, Luigi Luca (2000), *Genes, Peoples and Languages*, Berkeley: University of California Press.

Caves, Richard (2000), *Creative Industries, Contracts between Art and Commerce*, Cambridge, MA: Harvard University Press.

Coser, Lewis, Charles Kadushin and Walter Powell (1982), *Books, The Culture and Commerce of Publishing*, Chicago: The University of Chicago Press.

Crystal, David (2001), *A Dictionary of Language*, Chicago: Chicago University Press.

De Swaan, Abram (1993), *Words of the World*, Polity Press.

Desmet, Klaus, Ignacio Ortuno-Ortin and Shlomo Weber (2007), Income, diversity and redistribution, mimeo.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black (1992), An Indo-European classification: A lexicostatistical experiment, *Transactions of the American Philosophical Society* 82, Philadelphia: American Philosophical Society.

Fotheringham, A. Stewart (1981), Spatial structure and distance-decay parameters, *Annals of the Association of American Geographers* 71, 425-436.

Fotheringham, A. Stewart (1984), Spatial flows and spatial patterns, *Environment and Planning A* 16, 529-543.

Frith, Simon, Does music cross boundaries?, in Annemoon van Hemel, Hans Mommaas and Cas Smithuijsen, eds., *Trading Culture*, Amsterdam: Boekman Foundation.

Ganne, Valérie and Marc Minon (1992), Géographies de la traduction, in Françoise Barret-Ducrocq, ed., *Traduire l'Europe*, Paris: Payot.

Heilbron, Johan (1999), Towards a sociology of translation. Book translations as a cultural world system, *European Journal of Social Theory* 2, 429-444.

Hjorth-Andersen, Christian (2001), A model of translations, *Journal of Cultural Economics* 25, 203-217.

Hofstede, Geert (1980), *Culture's Consequences*, Beverly Hills, California: Sage.

Hofstede, Geert (1991), *Culture and Organizations*, London: McGraw-Hill.

Hoskins, Colin, Stuart McFadyen and Adam Finn (1997), *Global Television and Film*, Oxford: Clarendon Press.

Mélitz, Jacques (2007), The impact of English dominance on literature and welfare, *Journal of Economic Behavior and Organization* 64, 193-215.

Nida, Eugene (1975), *Language Structure and Translation*, Stanford, CA: Stanford University Press (selected and introduced by S. Dil).

Nida, Eugene and Charles Taber (1969), *The Theory and Practice of Translation*, Leiden: E. J. Brill.

Pfanner Eric (2007), When brands go global, *Herald Tribune,* July 23.

Quinzaine Littéraire (2006), La Fiction, c'est l'Amérique, 16-30 June.

Steiner, George (1992), *After Babel. Aspects of Language and Translation*, Oxford: Oxford University Press.

Swadesh, Morris (1952), Lexicostatistic dating of prehistoric ethnic contacts, *Proceedings of the American Philosophical Society 96*, 452-463.

UNESCO, Institute for Statistics (2002), Literacy and Non Formal Education Sector, Estimates and Projections of Adult Illiteracy for Population aged 15 years and above, by country and gender 1970-2025, July 2002 assessment.

World Bank Development Indicators Database (2005), GNI per capita 2004, Atlas method and PPP, 15 July 2005.

**Table 1 Number of Translations**
(per 1,000 speakers)

| Language | From | To |
|---|---|---|
| Bulgarian | 0.158 | 1.182 |
| Czech | 0.333 | 1.702 |
| Danish | 1.329 | 5.730 |
| Dutch | 0.276 | 2.159 |
| English | 0.950 | 0.043 |
| Estonian | - | 5.480 |
| Finnish | 0.312 | 3.330 |
| French | 0.884 | 1.069 |
| German | 0.451 | 1.220 |
| Greek | - | 0.500 |
| Hungarian | 0.201 | 0.667 |
| Icelandic | - | 14.300 |
| Italian | 0.270 | 0.204 |
| Latvian | - | 1.652 |
| Lithuanian | - | 1.083 |
| Norwegian | 1.319 | 5.586 |
| Polish | 0.113 | 0.482 |
| Portuguese | 0.023 | 0.139 |
| Romanian | 0.073 | 0.279 |
| Russian | 0.133 | 0.156 |
| Serbo-Croatian | 0.038 | 0.291 |
| Slovak | 0.237 | 1.248 |
| Slovene | - | 2.180 |
| Spanish | 0.054 | 0.277 |
| Swedish | 1.759 | 1.500 |
| Ukrainian | - | 0.041 |

Source: UNESCO, see Appendix

**Table 2 Estimation Results. Literature**
(dependent variable: (log of) number of translations, 1979-2002)

| Variable | (1) | | (2) | | (3) | | (4) | |
|---|---|---|---|---|---|---|---|---|
| | Coeff. | St.error | Coeff. | St.error | Coeff. | Sr.error | Coeff. | St.error |
| Source-language population | 0.76 | 0.05 | 0.76 | 0.05 | | | | |
| Destination-language population | 0.35 | 0.04 | 0.42 | 0.04 | 0.43 | 0.03 | 0.43 | 0.03 |
| Distance between languages | -1.04 | 0.15 | -1.00 | 0.15 | -1.05 | 0.10 | -1.05 | 0.10 |
| Destination-language literacy rate | | | 3.65 | 1.26 | 3.93 | 0.80 | 3.95 | 0.79 |
| Destination-language GNI/head | | | 0.45 | 0.14 | 0.52 | 0.09 | 0.52 | 0.09 |
| Intercept | 8.09 | 1.03 | 3.13 | 1.72 | 3.17 | 1.11 | 3.21 | 1.09 |
| **Source-language x Population** | | | | | | | | |
| Norwegian | | | | | 1.40 | 0.20 | | |
| Danish | | | | | 1.38 | 0.17 | | |
| Swedish | | | | | 1.29 | 0.13 | | |
| French | | | | | 1.13 | 0.07 | | |
| English | | | | | 1.09 | 0.05 | | |
| German | | | | | 0.99 | 0.06 | | |
| Finnish | | | | | 0.84 | 0.16 | | |
| Italian | | | | | 0.79 | 0.07 | | |
| Czech | | | | | 0.78 | 0.11 | | |
| Russian | | | | | 0.74 | 0.05 | | |
| Hungarian | | | | | 0.74 | 0.11 | | |
| Spanish | | | | | 0.57 | 0.05 | | |
| Polish | | | | | 0.55 | 0.07 | | |
| Dutch | | | | | 0.54 | 0.09 | | |
| Portuguese | | | | | 0.31 | 0.05 | | |
| Romanian | | | | | 0.27 | 0.09 | | |
| Bulgarian | | | | | 0.23 | 0.13 | | |
| Serbo-Croatian | | | | | 0.17 | 0.08 | | |
| Slovene | | | | | 0.00 | - | | |
| **Source languages x Population grouped** | | | | | | | | |
| Norwegian, Danish, Swedish | | | | | | | 1.31 | 0.10 |
| French, English | | | | | | | 1.10 | 0.04 |
| German | | | | | | | 0.98 | 0.05 |
| Finnish, Italian, Czech, Russian, Hungarian | | | | | | | 0.75 | 0.04 |
| Spanish, Polish, Dutch | | | | | | | 0.55 | 0.04 |
| Portuguese, Romanian | | | | | | | 0.29 | 0.04 |
| Bulgarian, Serbo-Croatian | | | | | | | 0.17 | 0.07 |
| Slovene | | | | | | | 0.00 | - |
| No. of observations | 471 | | 471 | | 471 | | 471 | |
| Adjusted R-squared | 0.411 | | 0.440 | | 0.774 | | 0.779 | |

In Eq. (3) and (4), Slovene is the omitted variable. All coefficients are different from zero at the 0.000 probability level, with the exception of the literacy rate in equations (2)--prob. level 0.044 and (3)--prob. level 0.804 and the intercept in equation (3)--prob. level 0.310 and (6)--prob. level 0.052.

**Table 3 Estimation Results. All Other**
(dependent variable: (log of) number of translations, 1979-2002)

| Variable | (1) | | (2) | | (3) | | (4) | |
|---|---|---|---|---|---|---|---|---|
| | Coeff. | St.error | Coeff. | St.error | Coeff. | St.error | Coeff. | St.error |
| Source-language population | 0.88 | 0.06 | 0.87 | 0.06 | | | | |
| Destination-language population | 0.55 | 0.05 | 0.62 | 0.05 | 0.63 | 0.04 | 0.63 | 0.04 |
| Distance between languages | -1.29 | 0.19 | -1.26 | 0.18 | -1.31 | 0.14 | -1.29 | 0.13 |
| Destination-language literacy rate | | | 0.38 | 1.52 | 0.88 | 1.07 | 0.93 | 1.06 |
| Destination-language GNI/head | | | 1.02 | 0.17 | 1.08 | 0.12 | 1.08 | 0.12 |
| Intercept | 8.19 | 1.27 | -2.11 | 2.07 | -1.74 | 1.50 | -1.56 | 1.44 |
| **Source-language x Population** | | | | | | | | |
| German | | | | | 1.16 | 0.08 | | |
| French | | | | | 1.15 | 0.09 | | |
| English | | | | | 1.13 | 0.06 | | |
| Danish | | | | | 1.07 | 0.22 | | |
| Swedish | | | | | 0.93 | 0.17 | | |
| Russian | | | | | 0.88 | 0.07 | | |
| Italian | | | | | 0.87 | 0.09 | | |
| Hungarian | | | | | 0.76 | 0.15 | | |
| Czech | | | | | 0.74 | 0.14 | | |
| Norwegian | | | | | 0.72 | 0.27 | | |
| Finnish | | | | | 0.71 | 0.22 | | |
| Dutch | | | | | 0.63 | 0.13 | | |
| Polish | | | | | 0.50 | 0.10 | | |
| Spanish | | | | | 0.50 | 0.07 | | |
| Serbo-Croatian | | | | | 0.31 | 0.11 | | |
| Bulgarian | | | | | 0.19 | 0.17 | | |
| Portuguese | | | | | 0.11 | 0.07 | | |
| Romanian | | | | | 0.10 | 0.12 | | |
| Slovene | | | | | 0.00 | - | | |
| **Source languages x Population grouped** | | | | | | | | |
| German, French, English | | | | | | | 1.08 | 0.04 |
| Danish, Swedish, Russian, Italian | | | | | | | 0.81 | 0.05 |
| Hungarian, Czech, Norwegian, Finnish, Dutch | | | | | | | 0.57 | 0.07 |
| Polish, Spanish | | | | | | | 0.43 | 0.05 |
| Serbo-Croatian | | | | | | | 0.22 | 0.08 |
| Bulgarian, Portuguese, Romanian, Slovene | | | | | | | 0.00 | - |
| No. of observations | 471 | | 471 | | 471 | | 471 | |
| Adjusted R-squared | 0.431 | | 0.474 | | 0.739 | | 0.745 | |

In Eq. (3), Slovene is the omitted variable. In Eq. (4), the group consisting of Bulgarian, Portuguese, Romanian and Slovene is omitted. All coefficients are different from zero at the 0.000 probability level, with the exception of the literacy rate in equations (2)--prob. level 0.044 and (3)--prob. level 0.804 and the intercept in equation (3)--prob. level 0.310 and (6)--prob. level 0.052.

## Table 4 Ranking Languages 1979-2002

| | 1979-2002 | 1979-1987 | 1988-1997 | 1998-2002 |
|---|---|---|---|---|
| **Literature** | | | | |
| Norwegian | 1 | 5 | 5 | 1 |
| Danish | 2 | 1 | 2 | 2 |
| Swedish | 3 | 3 | 1 | 3 |
| French | 4 | 2 | 3 | 4.5 |
| English | 5 | 4 | 4 | 4.5 |
| German | 6 | 6 | 6 | 6 |
| Finnish | 7 | 10 | 10 | 8 |
| Italian | 8 | 11 | 7.5 | 7 |
| Czech | 9 | 9 | 7.5 | 10 |
| Russian | 10 | 7 | 9 | 11 |
| Hungarian | 11 | 8 | 11 | 12 |
| Spanish | 12 | 13 | 12 | 13 |
| Polish | 13 | 12 | 13 | 14 |
| Dutch | 14 | 14 | 14 | 9 |
| Portuguese | 15 | 18 | 16 | 15 |
| Rumanian | 16 | 16 | 17 | 17 |
| Bulgarian | 17 | 15 | 18 | 19 |
| Serbo-Croatian | 18 | 17 | 15 | 16 |
| Slovene | 19 | 19 | 19 | 18 |
| **All Other** | | | | |
| German | 1 | 1 | 1 | 2 |
| French | 2 | 2 | 3 | 2 |
| English | 3 | 4 | 2 | 2 |
| Danish | 4 | 8 | 8 | 1 |
| Swedish | 5 | 9 | 5 | 7 |
| Russian | 6 | 3 | 4 | 10 |
| Italian | 7 | 7 | 6 | 8 |
| Hungarian | 8 | 5 | 7 | 12 |
| Czech | 9 | 6 | 9 | 13 |
| Norwegian | 10 | 16 | 15 | 5 |
| Finnish | 11 | 15 | 10 | 6 |
| Dutch | 12 | 11.5 | 11 | 9 |
| Spanish | 13.5 | 13 | 12 | 11 |
| Polish | 13.5 | 10 | 13 | 14 |
| Serbo-Croatian | 15 | 11.5 | 13 | 15 |
| Bulgarian | 16 | 14 | 18 | 19 |
| Portuguese | 17 | 18 | 16 | 16 |
| Rumanian | 18 | 17 | 17 | 17 |
| Slovene | 19 | 19 | 19 | 18 |

**Appendix on Data**

*Number of titles translated*. Data are obtained from UNESCO's website http://databases.unesco.org/xtrans/stat/xTransList.a?lg=1 as well as http://databases.unesco.org/xtrans/stat/xTransXpert.a?lg=1. This is how the data are described by UNESCO: "The *Index Translationum* is a list of books translated in the world, i.e. an international bibliography of translations. The database contains cumulative bibliographical information on books translated and published in about one hundred of the UNESCO Member States since 1979 and totalling more than 1.700,000 entries in all disciplines: literature, social and human sciences, natural and exact sciences, art, history and so forth." UNESCO thus distinguishes several categories, but also provides aggregate numbers for all categories. We concentrated on one important subcategory, "literature" and bundled all the others under the title "all other." These include: general and bibliographies; philosophy and psychology; religion and theology; law, social sciences, and education; natural and exact sciences; applied sciences; art, games, and sports, history, geography and biographies. UNESCO receives the data from bibliography centres or national libraries in the participating countries.

The UNESCO database is often strongly criticized as being not very reliable, since (a) what is qualified as a book varies between countries (some countries include doctoral dissertations, governmental, parliamentary and administrative documents, annual reports from firms, others do not), and (b) show sharp fluctuations (Heilbron, 1999). We partly avoid both criticisms, since (a) we consider "literature" separately, and there is probably more agreement on this definition (though we also discuss briefly "all other" fields) and (b) we deal with groups of years, so that fluctuations are smoothed out. Note that Heilbron who makes these criticisms also writes that the UNESCO source is the only international source that is readily available.

We focused on the main European languages considered as official (thus excluding for instance Catalan), and chose to discard some languages for which the number of titles translated was too small (Albanian, Moldavian).

Table A1 provides an overview of the languages included in our data. As will be seen, some languages (Estonian, Greek, Icelandic, Latvian, Lithuanian, Slovene and Ukrainian) are included as destination languages only, since the total number of titles translated from these languages was very small. Our sample includes thus 19 source countries, and 26 destination countries, which leads to 475 (= 19*26-19) translation flows. We ignored 4 observations (translations from Finnish to Hungarian and Estonian, and from Hungarian to Finnish and Estonian) since there are no data on distances between these three languages. This leads to 471 observations both for "literature" and for "all other".

*Populations in source and destination languages*. These are taken from Crystal (2001). Only those who use the language as mother tongue are taken into account since we considered that

readers mostly read in their native tongue, and only very seldom in a foreign language, with the possible exception of scientists.

*Distances between languages*. See Dyen, Kruskal and Black (1992), which contains distances between all Indo-European languages. Distances between Finnish, Hungarian and Estonian and Indo-European languages were set to 1.

*Literacy rates in destination languages*. See UNESCO, Institute for Statistics (2002). We chose rates given for 1990 and computed population weighted rates for Portuguese (Brazil and Portugal), and Spanish (Argentina, Bolivia, Chile, Colombia, Dominican Republic, Ecuador, Guatemala, Honduras, Mexico, Peru, Spain, and Venezuela). For other languages, we took the rate of the country in which the language is native. UNESCO does not provide literacy rates for all Western European or North American countries. We assumed that these were equal to 100%.

*GNI per capita 2004 in destination languages*. See World Bank (2005). We used Purchasing Power Parity per capita GNI per head (international dollars) in 2004. No such data were available for 1990. We assumed that relative ranking did not drift too much apart between 1990 and 2005. Population weighted weighed GNIs are computed for Portuguese (Brazil and Portugal), Spanish (Argentina, Bolivia, Chile, Colombia, Dominican Republic, Ecuador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Salvador, Spain, Uruguay, and Venezuela), German (Germany and Austria) and English (Australia, Canada, New Zealand, United Kingdom, and United States). For other languages, we merely took the GNI of the country in which the language is native.

**Table A1**
**Languages included**

| Language | Source | Destination |
|---|---|---|
| Bulgarian | x | x |
| Czech | x | x |
| Danish | x | x |
| Dutch | x | x |
| English | x | x |
| Estonian | | x |
| Finnish | x | x |
| French | x | x |
| German | x | x |
| Greek | | x |
| Hungarian | x | x |
| Icelandic | | x |
| Italian | x | x |
| Latvian | | x |
| Lithuanian | | x |
| Norwegian | x | x |
| Polish | x | x |
| Portuguese | x | x |
| Romanian | x | x |
| Russian | x | x |
| Serbo-Croatian* | x | x |
| Slovak | x | x |
| Slovene | | x |
| Spanish | x | x |
| Swedish | x | x |
| Ukrainian | | x |

*UNESCO deals with Serbian, Croatian and Serbo-Croatian separately. The distance matrix between languages includes Serbo-Croatian only. We therefore aggregated the three languages.

**NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI**

**Fondazione Eni Enrico Mattei Working Paper Series**

**Our Note di Lavoro are available on the Internet at the following addresses:**

http://www.feem.it/Feem/Pub/Publications/WPapers/default.htm
http://www.ssrn.com/link/feem.html
http://www.repec.org
http://agecon.lib.umn.edu
http://www.bepress.com/feem/

**NOTE DI LAVORO PUBLISHED IN 2008**

| | | |
|---|---|---|
| CCMP | 1.2008 | *Valentina Bosetti, Carlo Carraro and Emanuele Massetti*: Banking Permits: Economic Efficiency and Distributional Effects |
| CCMP | 2.2008 | *Ruslana Palatnik and Mordechai Shechter*: Can Climate Change Mitigation Policy Benefit the Israeli Economy? A Computable General Equilibrium Analysis |
| KTHC | 3.2008 | *Lorenzo Casaburi, Valeria Gattai and G. Alfredo Minerva*: Firms' International Status and Heterogeneity in Performance: Evidence From Italy |
| KTHC | 4.2008 | *Fabio Sabatini*: Does Social Capital Mitigate Precariousness? |
| SIEV | 5.2008 | *Wisdom Akpalu*: On the Economics of Rational Self-Medication |
| CCMP | 6.2008 | *Carlo Carraro and Alessandra Sgobbi*: Climate Change Impacts and Adaptation Strategies In Italy. An Economic Assessment |
| ETA | 7.2008 | *Elodie Rouvière and Raphaël Soubeyran*: Collective Reputation, Entry and Minimum Quality Standard |
| IEM | 8.2008 | *Cristina Cattaneo, Matteo Manera and Elisa Scarpa*: Industrial Coal Demand in China: A Provincial Analysis |
| IEM | 9.2008 | *Massimiliano Serati, Matteo Manera and Michele Plotegher*: Econometric Models for Electricity Prices: A Critical Survey |
| CCMP | 10.2008 | *Bob van der Zwaan and Reyer Gerlagh*: The Economics of Geological $CO_2$ Storage and Leakage |
| KTHC | 11.2008 | *Maria Francesca Cracolici and Teodora Erika Uberti*: Geographical Distribution of Crime in Italian Provinces: A Spatial Econometric Analysis |
| KTHC | 12.2008 | *Victor Ginsburgh, Shlomo Weber and Sheila Weyers*: Economics of Literary Translation. A Simple Theory and Evidence |

**2008 SERIES**

| | |
|---|---|
| **CCMP** | *Climate Change Modelling and Policy* (Editor: Marzio Galeotti ) |
| **SIEV** | *Sustainability Indicators and Environmental Valuation* (Editor: Anil Markandya) |
| **NRM** | *Natural Resources Management* (Editor: Carlo Giupponi) |
| **KTHC** | *Knowledge, Technology, Human Capital* (Editor: Gianmarco Ottaviano) |
| **IEM** | *International Energy Markets* (Editor: Matteo Manera) |
| **CSRM** | *Corporate Social Responsibility and Sustainable Management* (Editor: Giulio Sapelli) |
| **PRCG** | *Privatisation Regulation Corporate Governance* (Editor: Bernardo Bortolotti) |
| **ETA** | *Economic Theory and Applications* (Editor: Carlo Carraro) |
| **CTN** | *Coalition Theory Network* |