

The quantitative analysis of translation flows in the age of an international language*

Anthony Pym and Grzegorz Chrupała

1. Introduction

Arguments about the effects of hegemonic globalisation on cultural diversity have been used to suggest that translation can liberate us from the domination of just one international language, namely English. More specifically, some claim there should be fewer translations from English and more translations into English. However there is some empirical evidence that the percentages of translation from and to international languages cannot tell us how open or hegemonic a culture is, nor whether there should be more or fewer translations into or from an international language.

2. The debate

One consequence of globalisation, by whatever definition, would seem to be that translations account for only 2 to 4 percent of books published in the United States or the United Kingdom. This general proportion is much lower than the percentages often cited for other countries: 15 to 18 percent for France, 11 to 14 for Germany, some 25 for Italy, 25 to 26 for Spain, to bring together reports for years between 1985 and 1992 (Ganne & Minon 1992). Nor can one really doubt that translations from English account for a good deal of the movements into other languages. UNESCO figures indicate that English was the source language for an average of 41 percent of all translations in 1978–1980 (the years that will interest us here) and this proportion may have been as high as 49 percent in 1987 (Venuti 1998: 160). We must accept that the disparity between what is translated into English (not much) and what is rendered from it (a lot) is great. But what does this disparity actually mean?

It is easy enough to argue that globalisation, in linguistic terms, is a process whereby international languages become increasingly concentrated and dominant, to the point where just one language – currently English – has a clearly hegemonic position in cross-cultural communications. This general observation then produces some kind of binary opposition: when communicating between cultures, we may either translate or use the dominant trade language, English. The growth of one option – international English – must thus mean a decline in the other – translation –, which is why there are apparently so few translations into English. In this view, globalisation would have very negative results, not only for the lot of translators and the diversity of the world's languages and cultures, but also for English-language cultures themselves. The latter would be using globalisation to promote an inward-looking cultural autarky, oblivious to the damage they are causing to the rest of the world. This kind of argument is common enough among writers on translation. But can it be backed up by strong statistics?

We suspect not. Or better, some pernicious exploitation no doubt exists, but the available data show the process to be rather more complicated than many of the arguments would have it. In what follows, we shall try to explain our doubts in two ways: first in terms of a mental warm-up, second with some actual statistics on the numbers of book-titles translated into various languages.

3. Two simple mind games and some algebra

The first warm-up exercise is just common sense. Imagine, if you will, that there is a language with a huge number of books in it, and another with just a few books. Now, which of these languages is going to have the greater number of books translated *from* it? All else being equal, the bigger one, of course. So we might imagine that there are many translations *from English* just because there are many books published *in English*. And this need not imply any global conspiracy on the part of publishing companies or any other pernicious agent.

Second game: Let us suggest, as a general idea, that a language in which many books are published will have a lower translation percentage (i.e. translations as a proportion of all books published) than one in which fewer books are published (Pym 1996, 1998). Here we are talking about any language whatsoever. All we have to picture is the same scene as the one we have just thought about: Language A, with many books, has more things that could be translated than does Language B, which has fewer books. For example, Language A has 100 books, and Language B has 10.

Now, let us imagine an ideal world where translation percentages apply to all books independently of their origin. Let us say 10 percent of all books in this imaginary world are worth translating and are translated. That means 10 percent of

all A's books are translated into B, and 10 percent of all B's books are translated into A. Universal justice! But look at what happens: Language A gets just one book from B and finishes up with a translation rate of 0.99 percent (one translation for the 101 books published), and B receives 10 translations from A, so its translation rate is actually 50 percent (translations now account for half of all books published).

From this game we must conclude that a low translation percentage *in* a language may be due to no more than a relatively high number of books published in that language. This seems as obvious as the idea that the more books there are in a language, the more translations there are likely to be *from* that language (the finding of our first game). It would thus be normal for international English, which has numerous books published in it, to have a relatively low percentage of translations into it. The percentages would be a result of sheer size, and nothing else.

Let us try to formalize this simple model and see whether our intuitions and finger-counting games lead to valid generalizations. In order to this we need to switch from arithmetic to algebra and derive the appropriate equation describing the relations we are interested in. As before, we will assume a constant rate of translation *from* a language, so that if this rate were 10 percent (0.1) then 10 percent of texts written originally in a given language are going to be translated into all other languages. But we want our model to be valid for all possible translation rates and for any number of languages. So let's have:

- w_i = number of texts written in a language i (originals)
- p_i = total number of texts published in a language i (originals+translations)
- k = constant translation rate from each language (from 0 to 1)
- t_i = number of translations made *from* any given language i ; equals $k \times w_i$
- T_i = number of translations made *into* any given language i

Now, the rate of translation *into* a language i (let's call it r_i) is the following:

$$r_i = T_i / p_i$$

T_i , the number of translations going into language i , is a sum of translations made from all languages (sum of all t_i s), minus those made from i itself (normally one doesn't translate from, say, Kashubian into Kashubian).

The p_i , the total of publications in language i , is the number of originals w_i plus the above T_i . So, for n languages we get:

$$r_i = \frac{k w_1 + k w_2 + \dots + k w_n - k w_i}{w_i + (k w_1 + k w_2 + \dots + k w_n - k w_i)}$$

$$r_i = \frac{k(w_1 + w_2 + \dots + w_n - w_i)}{w_i + k(w_1 + w_2 + \dots + w_n - w_i)}$$

At this point let's put S (number of all originals written in all languages) instead of the unwieldy " $w_1 + w_2 + \dots + w_n$ ". So we get:

$$r_i = \frac{k(S - w_i)}{w_i + k(S - w_i)}$$

Here we have an equation that describes the rate of translation into language i as a function of S (the sum of all original texts written in all languages) and w_i (number of originals written in language i). However, it would be more convenient to have a single variable that would combine S and w_i . We can use

$$m_i = \frac{w_i}{S}$$

where m_i is language i 's 'market share', that is, the ratio of the number of originals written in that language to the number of originals written in all languages. From this we get:

$$w_i = m_i \times S$$

When we feed this into the equation we can get rid of S and finally we get:

$$r_i = \frac{k(1 - m_i)}{m_i + k(1 - m_i)}$$

For low (i.e. realistic) values of k , this relation is far from linear. Let's see a plot for this function $f(m) = k(1 - m) / (m + k(1 - m))$ for k at 0.1, the level we adopted in our mind game (see Figure 1).

The non-linearity of the function makes our original point even stronger. Not only do translation rates decrease as languages become bigger, they decrease faster than a simple linear relation would suggest. Our equation gives a fully linear plot only at $k = 1$, which stands for the highly implausible scenario where 100 percent of the texts written in any language are translated into all other languages, and for $k = 0$ where no translations at all are done.

What are the possible objections to this simple model?

First, our basic assumption of a constant rate of translation from every language is probably not very realistic. That may be so, but the assumption is an acceptable first approximation. By simplifying matters it allows us to construct a manageable model in order to demonstrate the influence of the size of a language on its translation rate.

Second, some might object, the ideal world should have a 10 percent proportion or whatever going *into* each language, such that Language A with 100 books would be obliged to translate all the 10 books from Language B, and Language B with just 10 books would be obliged to translate no more than one book from Language A. That is a possible scenario. However, if each language takes its translations in accordance with the distribution of books available to be translated, the biggest

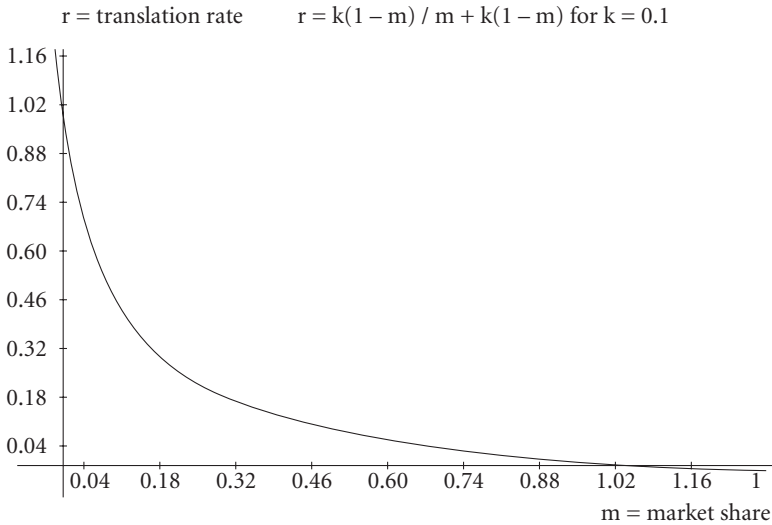


Figure 1. Hypothetical relation between translation rate and market share

one will still have the most books translated from it. So even this does not entirely solve the problem of global asymmetry, because the rates are equal but the absolute numbers are far from equal. Further, few cultural theorists would be entirely happy with a command economy that were really so indifferent to the relative qualities of books. Would we really want to judge and translate texts solely in terms of their origin?

4. Using UNESCO data

The games only give us models and hypotheses. To test them we need data on the translations and nontranslations published in a fairly wide range of languages. The numbers most readily available are those in the UNESCO yearbooks, although the yearbooks become less complete and appear less trustworthy from 1986 (thanks in no small measure to the withdrawal of the United States). We have thus decided to look at the data published in 1985, which actually includes figures for 1978–1983. Some notes on these numbers are necessary before proceeding:

- This database is rather less than ideal: key countries are missing in many of the tables; the figures given in different tables sometimes do not agree with each other; the notes on the language breakdowns are incomplete.
- This is also a precarious database because different countries have different definitions of the basic categories (e.g. what counts as a “book”; what counts

as “literature”; whether we are counting first editions or all editions). Care must thus be taken to ensure that the raw data for each country are compared only with other raw data for the same country (e.g. translations in Germany vs. non-translations in Germany), so that the more blatant differences in definition can be overcome. In search of the least perilous path we have looked at the numbers of first-edition titles in all categories of books (by whatever definition) for 21 languages (which were actually all the languages available for comparison with any degree of certitude).

- This is an intriguing database because it gives information on books published in non-national languages (e.g. books in French published in Sweden). This enables us to estimate the *total* number of books published in English without being limited to national categories such as the United States or the United Kingdom. It also gives us some kind of measure as to how non-nationalist certain countries are about their publishing (on which, more below).
- The database gives figures for several years, so some anomalies can be ironed out by taking the means. It also covers a fair range of languages, both big and small.
- The period concerned was one of relative stability in political terms, perhaps free of the high volatility that marks the translational development of languages with few books. For example, we find from another source (Vallverdú 1978) that translation percentages into Catalan were 55 percent in 1965, 8.3 percent in 1973, and 16.5 percent in 1977. Our test numbers should try to avoid such rapid shifts, which are due more to local developmental factors than to the general principles we are interested in testing.
- Perhaps because of the above reasons, the percentages of translations found in the UNESCO data are generally lower than those given in other sources for more recent dates. We find only 13 percent (instead of 25 percent) for Spanish; only 11 percent for German and French apiece. However, since the rate for English is just over 2 percent both here and in other sources, the fundamental difference we are interested in is not quite obscured by the uncertainties of UNESCO.

5. Two principles tested

Our first mind game suggested that the more books were published in a language, the more translations there would be from that language. When we test this principle on our data, we get the results in Figure 2.

English is clearly in an anomalous position. And about twelve languages are huddled together in the bottom left corner, with not many books published and

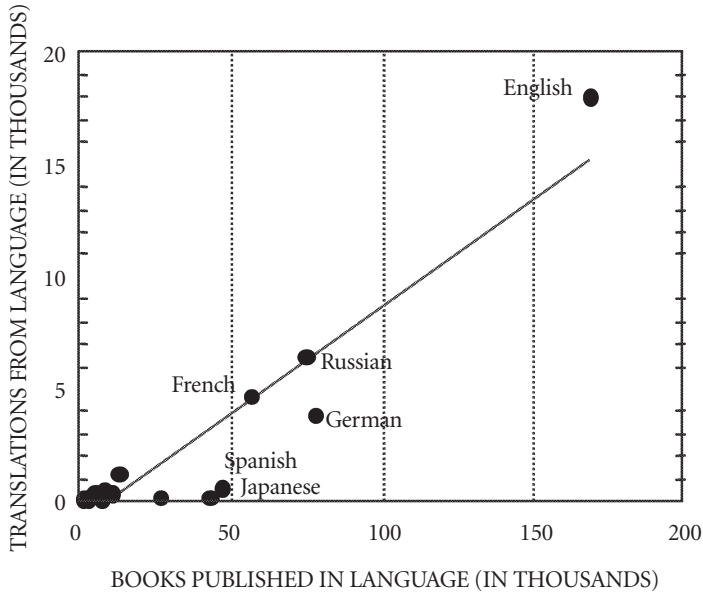


Figure 2. Books translated from language, by books published in language. UNESCO data for 1979–1983

thus not many that they could have had translated. But the regression line here averages all of that out and tells us that the general hypothesis holds. In fact the correlation is quite good ($p < .0001$), as might be expected for such a banal hypothesis. The only slight hitch is that English is clearly above the regression line, which might suggest that more books are indeed translated from it than is the norm for this field. We shall return to this problem later. For the moment, let us simply note that the fact that 41 percent or so of all translations are from English no longer looks quite so abnormal: it is more or less in keeping with what the regression line would predict.

Our second game and the equation derived from it suggested that the more books were published in a language, the lower the translation rate would be in that language. In other words, the bigger you become, the smaller the share of your cultural energy you put into intranlations (i.e. translations into your home language). When we test this on our data we obtain the following picture (see Figure 3).

Once again English is out on a limb, at a point that is difficult to compare with other languages. We might also note that the highest translation percentage in the sample was for Albanian, which in 1979–1981 was (falsely?) perceived as one of the most closed cultures in the world. And yet here we find that 25 percent of its books

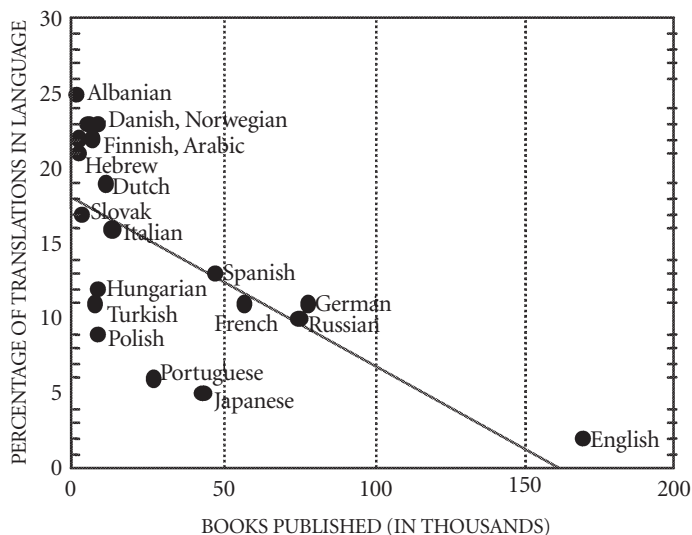


Figure 3. Percentages of translations, by books published in different languages. UNESCO data for 1979–1983

were translations. Could this be true? As the scatter plot suggests, Albanian had a high translation percentage not just because of any cultural openness but perhaps also because it had so few books published.

Unfortunately the general relation here is not as strong as one might have hoped for ($p = .0009$), which is just under the 0.5 we would like for a convincing correlation). Yet the picture seems clear enough in general terms, especially for the languages with the larger numbers of publications. As the number of books grows, the percentage of translations tends to decline.

Figure 3 seems to show quite different behaviour for a group of small languages, for a group of large national languages, and for English out there all by itself. This means there may be groups of languages obeying quite different dynamics, such that further analysis would have to consider each group on its own terms. We would then have even less reason to compare translations into English with translations into Albanian.

6. Translation vs. foreign-language reading

There was no third game above, but we can play one here. Or you could play it in a bookshop next time you travel. Just walk around and try to estimate the percentage of translations on the shelves (and if you are not yet convinced that English is

dominant, take a close look at the business-management or computing sections). You might get estimates as high as 40 percent or 50 percent, or even 80 percent for some fields if the bookshop is at all academic or intellectual. But what you will also find, in many countries, are foreign books that are *not* translated, usually on sale in the major trade languages of our day, increasingly in English. That is, in many countries the tendency is to read directly in foreign languages, without translation. And this practice, unevenly distributed, must necessarily skew any attempt to associate “cultural openness” (or any such value) with a high percentage of translations. If the Swedes, for example, are all reading in English, they should not really need translations from English, and they will potentially have a very open culture with a rather low percentage of translations. So all the arguments that might be based on translation data, including ours, are at best limited in what they can say about the ways cultures react to globalisation.

The hypothesis to be tested, the one that seems quite logical, is that the more a country consumes foreign-language books without translation, the lower the translation percentage will tend to be in its national language or languages.

To test this hypothesis we really need reliable data on book exports and imports between a wide range of languages. We do not have any such data. But what we do have, embedded in the UNESCO tables, are numbers of non-national-language books published in numerous countries (e.g. books published in French in Germany). *Faute de mieux*, we might hope that these numbers indicate uses of nontranslation that are compatible with a certain degree of cultural openness. True, the numbers may concern the country’s foreign projection more than its internal consumption patterns. But that is no reason for not looking at the data.

Here is what happens to our hypothesis (Figure 4).

Despite severe limitations on the data (UNESCO does not tell us about non-national-language publications in the United States, the United Kingdom, France or Germany), the results are interesting. Our logical hypothesis is shown to be quite wrong. When countries publish many books in foreign languages, they *also* tend to translate many books from foreign languages. The R^2 here is a high 0.717 ($p < .0001$), which is a good correlation for a hypothesis that is not at all banal.

Exactly what this means requires more information. It could be that intranlation and extranlation are simply moving hand-in-hand, as complementary sides of increased cross-cultural exchange. Or it may just be that translations and non-national-language publications more actively help each other in raising public awareness about foreign books, with each practice stimulating more than its own narrow market, as was argued by Schlösser (1937:2) when observing similar phenomena in the German reception of English literary texts.

Such speculation goes beyond our present concerns. All we are really interested in showing is that if translation percentages are to indicate something like cultural openness, *percentages of nontranslations might be assumed to operate the same way*,

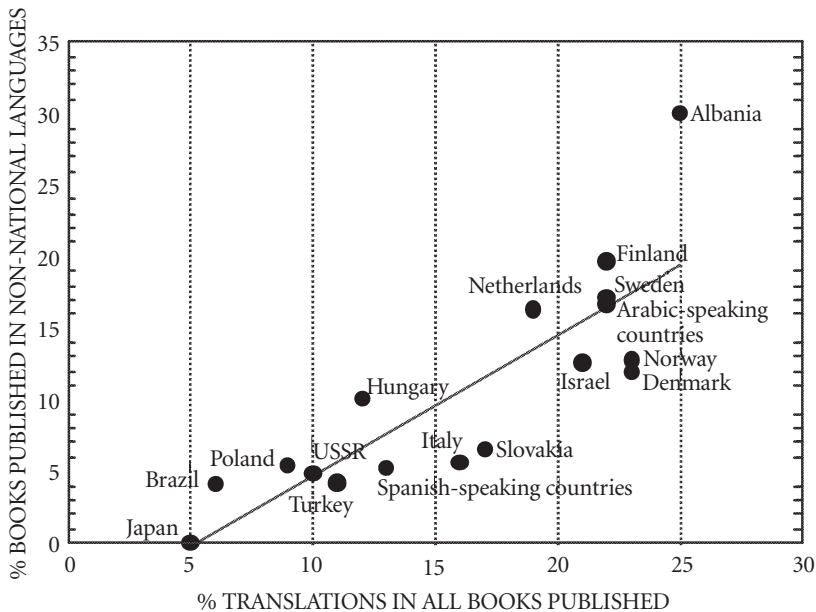


Figure 4. Percentages of translations, by percentages of books published in non-national languages. UNESCO data for 1979–1983

at least until we find good data able to prove that translation and foreign-language-reading are compensatory rather than complementary intercultural strategies.

7. Postscript: Minorities and the state of English

We are not arguing that there is no cultural hegemony. In fact, if one attempts a slightly different kind of calculation, translations into English might have been numerous enough to develop a certain kind of productive hegemony. For the period 1960–1986 the *Index Translationum* lists more than 2.5 times as many translations in Britain and the United States (1,872,050) as in France (688,720) or Italy (577,950). That is, the number of publications in English is now so great that readers can indeed find more translations there than they can in some languages with higher translation rates. This perhaps amounts to saying that imperialism can bring certain translational rewards; the victors carry off the trophies. Less provocatively, as we have noted, it means we are being less than astute whenever we try simple comparisons (the above lineal regressions) between the world's major trade language and other languages.

The sheer size of English means it has its own dynamics. It is not only the first language of some 320 million people worldwide, but also a second language for perhaps that number again. It is an official or dominant language in over 60 countries and “routinely in evidence” in a further 75 countries (Crystal 1997). It has multiple standardised varieties and is the matrix language for some 44 creoles and pidgins across the globe. This is a very big language, with countless variations churning within it.

Sheer size may bring with it a very high degree of inner diversity, visible in a wide range of distinct yet more or less mutually comprehensible varieties. Further, many of those different standards and non-standards gain access to printed communication, and do so in many different countries. Our data give some 24 percent of all books in English being published outside of the United States or the United Kingdom; the presence of writers born beyond those two main centres has become a regular feature of English-language literature. In fact, to complete our argument, the sheer size of English could mean that much of the diversity and new blood that other language groups seek through translation, English-language cultures may be receiving without translation (for the same argument in slightly different terms, see Constantin 1992: 126). Thus, in order to carry out a comparison between the United Kingdom and, say, Albania, we might want to take some of that 24 percent “extra-territorial” publication rate for English (i.e. of non-British or non-American origin) and add it on to the United Kingdom’s basic translation rate.

The point, once again, is that if we are trying to assess the effects of globalisation on cultural diversity, translation alone is neither a sufficient measure nor a sufficient remedy. Nontranslation may also be a measure of cultural diversity and openness. And quick statistics on the speed and dangers of globalisation must be handled with a great deal of care.

Note

* Briefer comments on the graphs have been published in Pym (2004: 40–45). The authors would like to thank Gideon Toury for corrections to a previous version of this report.

References

- Constantin, Jean-Paul (1992). “Les éditeurs”. In Françoise Barret-Ducrocq (Ed.), *Traduire l’Europe* (pp. 125–133). Paris: Payot.
- Ganne, Valérie & Minon, Marc (1992). “Géographies de la traduction”. In Françoise Barret-Ducrocq (Ed.), *Traduire l’Europe* (pp. 55–95). Paris: Payot.

- Crystal, David (1997). *English as a Global Language*. Cambridge: Cambridge University Press.
- Pym, Anthony (1996). "Venuti's Visibility". *Target*, 8(2), 165–177.
- Pym, Anthony (1998). *Method in Translation History*. Manchester: St. Jerome Publishing.
- Pym, Anthony (2004). *The Moving Text: Localization, translation, and distribution*. Amsterdam/Philadelphia: John Benjamins.
- Schlösser, Anselm (1937). *Die englische Literatur in Deutschland von 1895 bis 1934*. Jena: Biedermann.
- UNESCO (1985). *Statistical Yearbook / Annuaire statistique / Anuario estadístico*. Paris: Unesco.
- Vallverdú, Francesc (1987). "Els problemes de la traducció". In Carme Arnau et al. (Eds.), *Una aproximació a la literatura catalana i universal* (pp. 95–107). Barcelona: Fundació Caixa de Pensions.
- Venuti, Lawrence (1998). *The Scandals of Translation. Towards an Ethics of Difference*. London/New York: Routledge.