

ABSTRACT

English as a global language

Jacques Melitz*

CREST and CEPII, Paris, and CEPR, London

How far has English already spread? How much further can we expect it to go? In response to the first question, this chapter tries to identify the areas of life where English already serves as a lingua franca in the world (more or less) and those where the language faces sharp competition and does not threaten to marginalize the other major languages. The former areas of life are international safety, the internal business of international organizations, internal communication within the international news industry, international sports and science. The latter areas are the press, television, the internet, publishing and international trade. As to the second question, about the future prospects of English, the chapter argues that the advance of English will depend heavily on the motives to learn the other major languages in the world as well. Based on the empirical evidence, the same model applies to the incentives to learn English and these other languages. On the important topic of welfare, the cultural market is the *single* one where it is arguable that the progress of English has gone too far. English dominance in the song, the cinema and the best-seller is indeed extraordinary and difficult to reconcile with the evidence of strong popular attachments to home languages, which is otherwise apparent.

JEL classification: I00, F10, Z1

Keywords: English, language, trade, publishing, welfare

Jacques Melitz
Département de la Recherche
CREST-INSEE
15 Bd. Gabriel Péri
92245 Malakoff CEDEX, FRANCE

Email: j.melitz@hw.ac.uk

*This paper is intended to be a chapter in *The Palgrave Handbook of Economics and Language*, 2015, edited by Victor Ginsburgh and Shlomo Weber, in preparation. It reports on a lot of concurrent work in Melitz and Toubal (2014) and Ginsburgh, Melitz and Toubal (2014). As a result, some passages in sub-section 3.5 and section 4 occur in one of these papers or the other and should be considered co-authored.

English as a global language

Jacques Melitz

CREST and CEPII, Paris and CEPR, London

1. Introduction

There has never been in the past a language spoken more widely in the world than English is today. How far has English already spread? How much further can we expect it to go? What are the welfare implications? These are the three central questions in this chapter. The first two are deeply intertwined. We must have a clear idea how far the spread of English has already gone in order to assess how much further it can be expected to go. The role of the language as a lingua franca in some areas is of particular interest. On all these matters, a popular book by the linguist Crystal (2003a) under the same title as this chapter is extremely useful. So are two works by the sociologist Graddol (1997, 2006) that were commissioned by the British Council. In dealing with both questions, I will try to move beyond these two authors most of all in connection with publishing, foreign trade and language learning. As regards trade, we know from economic research that common languages promote bilateral trade between countries. We also know from many sources, including survey evidence of exporting firms, that trade stimulates language learning. There is recent econometric support. To what extent does the expected growth of world trade in the future imply the further spread of English? To what extent does it instead imply limits to the expected spread of English because of the similar inducement to learn other languages as well? Do other factors besides trade also play an important role? These will all be subject-matter.

The third central question, regarding the welfare aspect, clearly raises distinct issues. There is much concern today about the welfare impact of the spread of English on native speakers of the other major world languages. One example, of course, is the well-known French support of “l’exception culturelle” [the cultural exception] in the field of trade. It is indeed surprising to what extent English dominates in certain cultural areas, including the song, the film and the best-seller, not necessarily publishing in general. Is there possibly too much English in some areas for the good of mankind?

2. The status of English as a global language

2.1 The world distribution of native and spoken languages

We may begin with approximate numbers for native and total speakers of English and the other principal languages in the world. Table 1 provides such figures for the 12 largest languages in terms of total speakers. Turkish/Azerbaijani/Turkmen, Italian, and Dutch/Afrikaans are added to these 12 languages even though they are not exactly the next 3 largest because they enter in a study of language learning on which I shall report prominently later on. The table shows two sets of estimates, one from *Ethnologue* and the other from Melitz and Toubal (2014) (hereafter MT), the source of the data for the relevant study of language learning. The figures from *Ethnologue* (obtained on the internet in May 2013) apply to different dates. To give an idea, they go back to 1998 for Arabic and 2001 for English, and refer to census data for 2000 for Chinese, 2006 for French and data as recent as 2010 for Spanish. The data from MT (which omit 2 of the 12 largest world languages, Hindi/Urdu and Bengali), all collected in 2010, are mainly bunched around 2001-2007. For the moment, it is not possible to produce such a table for a uniform year. All such tables rely on the assumption that language is a slow-moving variable, which is of course more reliable for native than for spoken language.

There are some large discrepancies between *Ethnologue* and MT calling for explanation. Regarding spoken English, the principal cause is that *Ethnologue* draws figures from an incomplete table in Crystal (2003a, Table 1) that concerns only ‘territories [75 of them] where English has held and continues to hold a special place’ (Crystal, p. 60), by which Crystal evidently means, almost without exception, territories that were under the administrative control of English-speaking countries at some time in living memory or else where the language is official or both. Consequently, those figures do not include spoken English in places like the Netherlands, Germany and the Scandinavian countries, where the language is widely spoken but has never been either that of the ruling political power in the country or official. The MT data largely replicates the data in Crystal’s table for the same 75 territories he mentions (except for the tiny territories that are not in their study), but adds data for spoken English in other parts of the world wherever the authors could find it. MT generally draw no distinction between speakers in countries where the language has “a special place” and other countries.¹ One important source of their data is an EC 2005 survey of (self-reported) ability to hold a conversation in foreign languages in the 28 current (2014) members of the EU plus Turkey (Eurobarometer, 2006). *Ethnologue* generally tends

to collect data for spoken language by non-natives (termed L2) from a selection of territories in Crystal's sense (though without being explicit). This explains the wide discrepancies between *Ethnologue* and MT for spoken French and German as well as for English. In the case of Turkish/Azerbaijani/Turkmen, Italian and Dutch/ Afrikaans, *Ethnologue* offers no L2 data. See the note to Table 1 about Chinese and Arabic.

[Table 1 here]

There are more native speakers of Chinese than any other language by far. However, based on Table 1, English is neck and neck with Chinese as world leader in total speakers. All other languages lag far behind. In fact, it would be easy to propose figures for English speakers far exceeding those for Chinese by extrapolating on the basis of attendance in English classes and/or some ability to understand. On such grounds, one could readily add an extra 300 million speakers of English for India and China alone (see for example Kachru, 2010, pp. 205-11, and for India see also Crystal 2003a, pp. 46-49). The MT data, like its basic sources (prominently including *Ethnologue*) is more conservative, and at least in the authors' minds, reflects an ability to converse.

Most important of all, English is way ahead of all other languages as a learned language by non-native speakers, and it is the only one to be well represented in all five continents. In terms of geographical dispersion, only French comes anywhere close to English but is still far behind.

2.2 Areas where English serves as a lingua franca

There are situations where reliance on interpreters and translations may even be dangerous for safety. Control towers must be able to communicate with air pilots instantly. Commanders of modern vessels at sea must be able to communicate with one another rapidly. Accordingly, in recent decades, active steps have been taken toward a single world language in the field of international safety. An English vocabulary named "airspeak" has been progressively adopted by over 180 countries based on the recommendations of the International Civil Aviation Organization. There is also a limited vocabulary, "seaspeak," based on English that has been adopted by the International Marine Organization.

In some other cases, multiple languages are not lethal but prohibitively expensive. Meetings of organizations with large international memberships could take place with simultaneous translation or, perhaps in the future through automatic translation based on voice

recognition. For the moment, though, this would be at excessive cost. In many instances, the mere publication of all the information that international organizations generate in the respective native tongues of the members is almost unimaginable. Thus, nearly by necessity, international political organizations tend to choose a limited number of official languages. Under the UN charter, there are five: Chinese, English, French, Russian, and Spanish; Arabic was added as a sixth in 1973. The official language of the IMF and the World Bank, both in Washington, is English. French is a second official language of the OECD (located in Paris) besides English, but it is possible to get along in the organization without French but not without English. Unsurprisingly, the only official language of the *Francophonie* is French. Similarly, the official languages of Mercosur are Spanish and Portuguese, not English. Yet these are exceptions. Generally, English tends to be at least one of the official languages of international political organizations. Interestingly, English is even the *only* official language of a couple of major regional political associations outside of Europe or North America: namely, OPEC and the South Asian Association for Regional Cooperation. There is only one international political organization that recognizes numerous official languages but at a notoriously high cost, the European Union (Fidrmuc and Ginsburgh, 2007). Even there, the organization adopts English, French and German as the only “working languages.”

In the case of private rather than public international associations, where the biblical problem of Babel rears up just as well, information is less readily available. Based on some sleuth work, Crystal (2003a) calculates that in 1995-1996, 85 per cent of such organizations made some official use of English. French was second with 49 per cent, and only Arabic, Spanish and German scored over 10%. Quite tellingly, he also finds that for European associations as such, English is an ‘official or working language’ in 99 per cent of the cases, and English plus French plus German is ‘the most popular European combination.’

There are other areas where English serves largely as a lingua franca in the world. The international press and international sports are two (see Graddol, 1997, 2006). People in the business of diffusing international news, or the firms active in the diffusion process itself, must obtain their information quickly. As a result, they have veered heavily toward English in data transmission among themselves. In close connection, there is a heavy concentration of providers of international news in English-speaking countries, including Reuters, the Associated Press, the BBC, NBC, and the *New York Times*. As relevant too, Crystal (2003a, pp. 91-93) traces the early development of the news industry largely to the English-speaking world. International sports also require a lingua franca for the minimal

communication necessary when competing athletes and an umpire meet on a playing field.

Perhaps the most intriguing area of English as a lingua franca is science and scholarship since in this case the basic mechanism at work is the self-interest of the individual scientist and scholar. There have been studies of publications in the natural sciences from 1880 to 1996 based on American, French, German and Russian bibliographies. The results, which are drawn from original work by Tsunoda (1983) and Ammon (1998), are summarized in Hamel (2007). They show only 36 per cent of publications in English in 1880, followed by a rise to around 50 per cent in 1940-50 to 75 per cent in 1980 and 91 per cent in 1996. The 91 per cent average for 1996 covers a top of 94-95 per cent for physics and mathematics and a low of 83 per cent for chemistry. The information for the social sciences and humanities starts in 1974 and goes up to 1995 and the trend proceeds from 67 per cent in 1974 to around 70 per cent in 1980 and on to 83 per cent in 1995. Therefore, by 1995 the social sciences and humanities attained the same level as chemistry in 1996. The decentralized nature of the incentives to publish scientific work in English is clear. Truchot (1996) spells out the tendency for natural scientists in France to switch from publishing in French to English since the mid-seventies in the face of political pressures to stick to the home language. In addition, all the top journals in sociology, political science and economics are in English and the social scientist confronts the same incentives to publish in English as the natural scientist in much of the world. Even if one publishes occasionally in one's home language, those who seek an international reputation mostly try to publish in English, especially their best work.

2.3 Areas of puzzling English supremacy

There are, however, some puzzling areas of English supremacy. The song, the motion picture, and the best-seller are three. Native speakers of foreign languages who consume English songs do not even necessarily understand the lyrics. As for the film and the best-seller, the content must be dubbed or translated (sub-titled) from the English when it is sold to foreign-language speakers. Thus, it is futile even to try to explain the extraordinary supremacy of English in these three areas on the basis of benefits of a lingua franca.

The facts are striking. The list of best-selling songs since 1942 includes only four singles that are not sung in English out of the 126 that sold over five million copies (one Portuguese, one Japanese, one French, one Italian) and only five not sung in English out of the 98 that did the same (one Portuguese, two Japanese, to Spanish) (personal count based on web search). As regards *Europe*, of the current (late 2013) 100 best-sellers, all of the top

20 are sung in English and only six out of the top 40 are sung in different languages. Only for the rest of the top 100 does the count begin to even up between English and all other languages together. This is not true for Latin American countries where Spanish and Portuguese stand up well to English even over the top 20 or so. There are also some countries in the rest of the world, especially in Asia like Japan, where the home language dominates in the top ten.

In the case of the top-grossing foreign-language films at the box office all-time, not a single non-English-language film shows up in the top 500 (Box Office Mojo). The non-English-language film that grossed the most (*Crouching Tiger, Hidden Dragon*) earned 62 per cent of the receipts of the 500th on the all-time list. For each of the five most recent years (2008-2012) not one non-English-language film shows up in the top 20 biggest box-office hits. In the case of books, the best-sellers of all time include exactly two out of the top 100 that were originally written in a non-English language (web search). One is the Swedish trilogy by Stieg Larsson, published posthumously in English, *The Girl with the Dragon Tattoo*, *The Girl who Played with Fire*, and *The Girl who Kicked the Hornets' Nest*. The other is Paulo Coelho's *The Alchemist*, originally written in Portuguese. (Specialists in literature may choke.)

We will come back to the arresting dominance of English in the previous cultural areas in connection with welfare.

3. Areas where English faces sharp limits

It is important, next, to discuss areas where English, though in the lead, faces sharp competition. The relevant areas are wide and cover the daily or weekly press, television, the web, publishing and trade. When it comes to trade, the spread of English encounters limitations not only for consumption but investment goods and therefore in strict communication between firms. These areas of activity serve to underline the exceptional situation for the song, the film and the best-seller, where the goods must also meet the market test of individual consumer satisfaction and where the pressure to meet this market test is not even necessarily weaker.

3.1 The press

The English press has a wider international presence than other presses, but the daily

presses of the world plainly reflect the native languages well. Japan boasts five or six of the ten newspapers with the top world circulation, all in Japanese; the enormous Chinese press is predominantly in Chinese; the German press is in German; the Italian one in Italian, the Czech one in Czech, etc. The newspaper with the single largest world circulation, published in India, *The Times of India*, is in English. Yet Hindi is also well represented in the Indian press. Admittedly, there are many English-language newspapers that are published outside the major English-speaking countries, like the *China Daily* (published in Hong Kong), and the single major world newspaper essentially intended for international distribution is in English: namely, the *International New York Times* (only recently the *International Herald Tribune*). Still, there can be no claim of English dominance outside of native-English countries as regards the press.

3.2 Television

The story for television is more interesting. There had been a scarcity of broadcasting space for transmitting TV signals in the air in the sixties when technological innovations revolutionized the industry. First, cable TV, next the launching of satellite TV, following the cabling of satellite TV, and most recently digital compression, led to a wealth of television channels. The US developed TV earlier than the rest of the world, partly because of its huge home market. Therefore the country was able at first to furnish viewing content to foreign TV channels, hungry for material, at lower cost than those channels were able to produce the material for themselves. There resulted an enormous trade surplus in audiovisual material for the US in the seventies and eighties (*The Economist*, 1997). A highly influential UNESCO study in 1983, updating a similar study in 1973, showed that the earlier one-way flow of programming material from the US to the rest of the world (Tapis 1984) had continued. The programming material consisted heavily of entertainment. Overall, a third or more of the total programming time in the rest of the world came from the US in 1983, with important differences by country and region. For example, 75 per cent of the viewing content in Latin America, 44 percent in Western Europe, and 21 per cent in the Arab countries came from the US. The US, in turn, hardly imported from anyone. There was talk at this time of US dominance and “hegemony” (Tracey 1985, Biltreyst 1991). (This was the time of the world success of the US television series *Dallas*.) However, studies soon showed that TV audiences everywhere preferred home-made material, and as TV matured and foreign TV audiences grew, home-made material progressively replaced imported one and several observers correctly predicted that the trend would continue in many parts of the world (Berwanger 1987 and Hoskins and McFadyen 1991).

Graddol (1997) has extended the narrative since the development of satellite TV. Once again, the early development was largely American and fed fears of massive spread of English and US culture at the expense of other major languages and other cultures. However, the eighties and nineties saw the arrival of large non-US participants in Satellite TV, like *Arte* and *Euronews* (which now broadcasts in 15 languages). More tellingly still, the English providers started to broadcast in other languages and to adapt the language to local preferences. Thus, after beginning broadcasting from Hong Kong exclusively in English and Chinese, *Star TV* began to promote local Asian languages. Likewise, *CNN International* launched a Spanish program in Latin America. Writing in 1996, Graddol predicted (p. 60): ‘National networks in English-speaking countries will continue to establish operations in other parts of the world, but their programming policies will emphasize local languages.’ On the other hand, he says (2006, p. 46): ‘English, however, remains the preferred language for global reach.’ The several efforts under way now to broadcast internationally by non-English channels have adopted the model of the German *Deutsche Welle*, which decided right from the start to broadcast both in German and English.

3.3 The internet

The story of the internet largely follows the same script as for TV. The internet began predominantly in the US and there was a strong sense in its early days in the eighties that it would spur the learning of English. There is probably some truth to this, but what we have seen mostly since is a progressive tendency to make the internet available in other languages. Until the nineties the only script available on the internet was ASCII, the American Standard Code for Information Interchange, which allowed only 95 printable characters (plus 33 controls such as plus signs, etc.). No foreign accents or diacritical marks were possible. However, Unicode followed, remedying the problem, and by September 1999 allowed 38 different scripts and a total of 49,259 characters. In its most recent version (6: 2), dating 2010, Unicode allows 100 scripts, almost all those in current use, and contains 110,182 characters. Crystal (2006) points out nicely that far from contributing to the death of languages, the internet helps to preserve them by permitting more of them to be stored permanently.

[Table 2 here]

The Internet World Stats website brings together world data about internet usage and population statistics. Table 2, based on this site, summarizes the essential failure of English to crowd out other languages on the internet since the year 2000. The first column of the table shows the top nine languages on the internet. The second column gives the estimated number of users for those nine languages in 2011. The results say that the top nine languages account for 80.2 per cent of all users. (I counted a total of 163 languages on the internet (including Latin and Esperanto) from one of the original sources, Web Technologies Surveys.) The third column gives the growth of the number of users of the internet 2000-2011. There we see the massive catch-up of English by Arabic, Russian and Chinese since 2000. Column 4 reproduces the information in column 1 as ratios of the world total of users (2.1 billion). Column 5 reproduces the same ratios instead on the basis of the last column of Table 1. This is done by dividing the numbers for the 9 languages in the column by their sum, 5090 (in millions), and then multiplying all of them by 80.2 per cent, so that the percentage total is the same as in the preceding column, 4. The ratios *for the individual languages* in columns 4 and 5 are remarkably close. Therefore the relative order of the 9 largest languages on the internet now resembles their position in terms of relative numbers of speakers, even if Spanish, Arabic, French and Russian trail somewhat and English is disproportionately ahead by 4.7 per cent. Obviously a massive adjustment of relative positions must have taken place in 2000-2011 to permit this result (see column 3) but it has already happened.²

3.4 Publishing

Walk into any book store almost anywhere on earth and you will find before you predominantly titles in the main local language. However, regardless where you are, there may well be an English section, maybe prominently in sight. (Alert: the subject now is the number of titles, not the number of best-sellers, which depends on sales of individual titles.) Since people like to read in their home language for pleasure, just as they like to read newspapers in their home language and they like to hear their home language on television, the book store will be stacked mostly with titles in the home language. Of considerable note, though, the clientele for books is also smaller than the one for newspaper, television, and internet; it is more educated; and it is more likely to read a foreign language in the original both for profit and pleasure. Since English is the world's most prominent lingua franca, we might therefore expect English to be disproportionately well represented in edition.

In earlier related research, covering data from 1971 to 1991 (Melitz, 2007), I found support for the preceding view, and in a renewed study of the same question, based on the latest information, the support appears greater. According to the UNESCO annual Statistical Yearbook, the principal data source, the ratio of published English titles to total world titles around 1971 was 24 per cent (1973 Yearbook). Total world population at the time was 3.8 billion. 24 per cent makes 900 million, which is far above the then-current English speaking population. Based on Table 1, the world ratio of English-speakers attained around 17 per cent in the vicinity of 2005 and must have been lower in 1971. Therefore English was disproportionately represented. An interesting development took place in the next two decades. As English penetrated worldwide, the ratio of English titles to the total fell to 21 per cent around 1981 (1984 Yearbook) and 20 per cent around 1991 (1995 Yearbook). This was probably a reflection of an improvement in education and literacy rates in the non-native-English part of the world, together with faster population growth in this part of the world than the native English one. However, according to latest results, for circa 2009, the forces enhancing English relative to other languages have taken over in publishing since the eighties. The ratio of publications of English titles to the world total rose to around 30 per cent in 2009, well up from up from 20 per cent in 1991, and quite high relative to the ratio of around 17 per cent for English speakers relative to total world population circa 2005.³ Therefore, English does indeed now occupy a highly disproportionate place in world edition though other languages are still well represented.

The prominence of English for translations is more pronounced than for the rest. Already for 1960 to 1987, the UNESCO yearbooks show that the share of English in translations as a source language was 41 per cent in 1960 and then rose to 56 per cent in 1987⁴ even though the percentage of English in total titles fell from 24 to 20-21 per cent, as we saw. Correspondingly, the UK and the US publishing industry took little interest in translation. As the world ratio of translations to total titles went down from 9 per cent to 7 per cent during the 1960-87 stretch, the British publishing of translations rose from a bare 2.1 to 3.7 per cent and the US one dropped from 6.6 (an exceptional postwar high) to 3 per cent (see Table 2 in Melitz, 2007). In striking contrast, European countries were and remain unusually interested in translations. A study sponsored by the Commission of European Communities (BIPE, 1993) shows a ratio of 17 per cent for translations for the then-members (incorporating a 3 per cent figure the UK) in 1991 (see Table 3 in Melitz, 2007). For the literature classification of titles, the ratio of translations to total titles was almost twice as high, 32 per cent (including 4 per cent for the UK). Ginsburgh, Weber and Weyers (2011) provide highly confirmatory findings in a study exploiting the UNESCO web

source, *Index Translationum*. They report translations out of English 22.09 times higher than translations into English for publications in the literature classification covering 19 Indo-European languages in 1979-2002. The next highest ratio in their sample, for Swedish, is 1.17 to one (and clearly reflects an unusually high level of translations of Swedish into Danish and Norwegian). Russian and French are the only other two languages besides Swedish for which the ratio is even close to one rather than well below it. See the opening chapter by Allen (2004) in a collective work sponsored by the international literary association PEN, for an interesting and compelling account of the worries of specialists in literature about the low level of translations into English in the literature classification.

Nevertheless, I include translation as an area where English faces sharp limits. The reason is the visible interest in translations of original works in non-English languages outside of the native-English countries. In the aforementioned BIPE (1993) study, about 43 to 49 percent of the translations in the literature classification in France, Italy, Spain and Greece are from other source languages besides English and the figures are in the range of one-third for Germany, the Netherlands and Denmark.⁵ To all indications, therefore, it is not translations that are nearly closed to authors who write in other languages than English but translations into English.

A partial explanation which is inspired by Ginsburgh, Weber and Weyers (2011) may be offered. Based on familiar reasoning in economics, suppose that all readers like a variety of titles for reading pleasure but exhibit home preference, which means a preference for reading works that were originally written in their home language. This agrees with the estimate of only 7 per cent for translations relative to total titles for the world as a whole in 1991, 17 per cent for the European countries in the BIPE sample in 1993 and the high of 31 per cent in this last sample for titles belonging to the literature classification. Suppose, in addition, that the degree of home preference rises with the size of a language community. In other words, large language communities are more insular than small ones: the small ones are more interested in different cultures and the outside world. Then the arithmetic goes in the right direction. For example, suppose that home preference is twice as high in a language community that is twice as large as another. Then 90 per cent home preference in the larger language community would mean only 45 per cent home preference in the smaller one. That makes for 10 per cent translations relative to total titles in the larger community as opposed to 55 per cent in the smaller one. It is easy to see that this hypothesis will go a long way toward reconciling the data with consumer sovereignty. If so, then in terms of standard economic reasoning, there is no market failure.

Notice, though, that the reasoning does not even begin to address the evidence for best sellers. In the previous example of one language community twice as large as the other, if on average home sales are the same for works in the home language that are translated into the foreign language as they are for other works written in the home language and also, separately, for those that are translated into the home language, then *on average* the total world sales *of a translated work* originally written in the large language and the small language would be *exactly the same*. In both cases, the total of home sales plus foreign sales would be the sum of the same two numbers. The issue of the best-seller concerns the distribution of sales of individual works, and more specifically the upper tail of the distribution. The extraordinary dominance of English in best-sellers remains as puzzling as before.⁶

3.5 Trade

It is intuitive that a common language would boost trade, especially for goods that are not perfectly homogeneous and demand investigation. A series of questionnaire surveys of investment in language skills by exporting firms in Europe confirms the intuition. The series began in 1996 and culminated in a large study in 2005 commissioned by the European Commission from CILT (a British organization focused on foreign language skills in business). This study, by Hagen et al (2006), covers a sample of 2000 small and medium-sized exporting enterprises (SMEs) in 29 European countries, including Turkey, and 30 large multinationals (MNEs), all home-based in France (see Annex 4 of the study). To the question, has your company undertaken foreign language training in the last 3 years, 35 per cent of the 2000 SMEs answer yes. The percentage of these 2000 firms foreseeing a need to acquire additional expertise in foreign languages in the next 3 years is higher, 42 per cent. If anything, the MNEs are more conscious of the importance of investing in linguistic skills than the SMEs. 60 per cent of the MNEs recognize deficiencies. This is below the 75 per cent figure in a previous study of similar inspiration covering 151 multinationals with a broader international distribution of home bases, including the UK, Germany and a sprinkling of other countries besides France. See Feely and Winslow (2005), another CILT publication, and also Bel Habib (2011).

The previous survey evidence, however, does not tell us much about the impact of a common language on bilateral trade. For such further knowledge, the gravity model of international trade is better suited. Somewhere in the early nineties, researchers at the

World Bank began introducing common language into this model – which has since become very popular (only since) – as a factor promoting trade in merchandise between country pairs along with the other two prominent factors reducing trade frictions in the model, namely, geographical proximity and a common border (see Havrylyshyn and Pritchett 1991, Foroutan and Pritchett 1993, and Frankel 1997). (Studies of bilateral trade in services remain a serious empirical challenge because of inadequate statistics.) This was done by introducing a binary variable equal 1 if a country pair possesses a common language and 0 otherwise. The early uses effectively depended on peer acceptance and simply avoided ambiguous cases like France/Switzerland. The variable was highly successful. At first this was interpreted to reflect the importance of cultural proximity in trade with some acknowledged confusion between language and ex-colonial relationships. But the confusion was soon resolved by introducing ex-colonial relationships concurrently (through one or two additional binary (0 or 1) variables). Since 2000 the use of the common language variable has become pervasive because of the statistical significance of the variable and the intuitive sense of its importance. Systematic use of the variable was also made possible, almost regardless of country sample, by the adoption of a common official language as the criterion. Prominent estimates of the coefficient of common language based on this criterion are near 0.5. There are already two meta-analyses of results of studies using common language (most of which rest on the preceding index but not all of them): Egger and Lassmann 2012 resting on 81 studies, and Head and Mayer 2013 resting on 159. Both papers come up with estimates of around 0.4-0.5 for common language. Since the dependent variable in the estimates is the log of bilateral trade, 0.4 corresponds to about 50 per cent more trade between a country pair with a common language than a pair without one ($\exp(0.4) \cong 1.5-1$). This is a large effect. Yet upon reflection, the estimate could be too low since it depends on a flawed measure of common language, and measurement error tends to bias estimates downward.

Consider, for example, the set of 28 countries covered in the aforementioned Eurobarometer (2006) survey of language skills. In this sample, English is the official language of a single pair: the UK and Ireland. Therefore based on the measure, common English boosts bilateral trade strictly for this one pair in the entire sample whereas, of course, knowledge of English is widespread in the sample and it is also uneven. A similar problem arises for common German, which, according to the measure (or at least the standard application of it), exists strictly between Germany and Austria whereas German is widely spoken in Denmark (where it could also be considered official) and the parts of Eastern Europe in the sample. There are other reasons for skepticism. Official languages are

about as common across countries as national flags. As nearly inevitable, official languages sometimes arise despite moderate or low levels of speakers if linguistic diversity is high.⁷ Thus, French and English are official in a good number of African countries where the percentage levels of speakers of either language are below one-quarter and not infrequently below 10 per cent. Spoken English is also only around 20 per cent in some Pacific Islands where the language is official, like Fiji, Kiribati, and Mauritius. Quite apart from statistical bias, the earlier issue I raised, if we consider the many reasons why a common language might promote bilateral trade, ranging from ethnicity, tastes and trust to ease of communication and facility of obtaining written or spoken translations, one can easily question that a measure based on official status alone would cover the subject. A recent study, born of such doubts, uses four separate measures of a common language simultaneously: common native language, common spoken language and linguistic proximity together with common official language (MT, 2014).⁸ The resulting estimate of the impact of a common language goes up from around 0.5 to about 1.1. On this last estimate, a common language raises bilateral trade by 200 per cent ($\exp(1.1)-1 \cong 2$).⁹ All four measures are also simultaneously important.

The natural question to ask here is whether English promotes bilateral trade more effectively than other languages. Table 3 focuses on this question based on the MT study. The relevant gravity equation covers 195 countries over the ten years 1998 through 2007 and the dependent variable is (the log of) bilateral trade. The estimates also employ separate exporter-year and importer-year fixed effects so that the results depend entirely on the cross-sections. There are bilateral controls for distance, common border, ex-colonizer/colony relationship, ex-common-colonizer relationship, common religion, common legal system and years at war since 1815. All the controls enter highly significantly different from zero in the estimates and carry the expected signs. But in the interests of space, Table 3 reproduces strictly the coefficients and *t* values of the linguistic variables. As evident from column 1, all four linguistic influences are highly significantly different from zero, all but common native language at the 99 percent significance level. In addition, the sum of their effects adds up approximately to 1.1, as mentioned above. One might worry that these estimates are subject to simultaneity bias because of the reciprocal effect of bilateral trade on common spoken language. For this reason, the paper also provides estimates founded strictly on common official language, common native language and linguistic distance (in which case those variables partly reflect common spoken language, with which they are highly correlated). The estimates are about the same.

[Table 3 here]

Columns 2 and 3 concern the separate impact of English. Column 2 does so by dropping all the languages except English from the analysis. To see the interest, suppose that the results of column 1 depended on English alone. In that case, the measures of the linguistic variables in column 2 would be superior. They would simply remove errors of measurement and should yield higher and better estimated coefficients. However, if instead the measures of a common language in column 1 are the right ones, then to the contrary, the measures of linguistic influence in column 2 would be noisy and should yield lower and less well estimated coefficients than the previous ones. Indeed, in this last case – that is, if the broad measures of a common language are the appropriate ones – there are two reasons why the English-based measures of the linguistic variables might perform particularly badly. In the first place, an English-speaking country has a great many solutions for skirting the language barrier altogether. There are lots of other English-speaking countries with which it could trade. Therefore, common English could be an especially weak spur to trade with any single common-language partner.¹⁰ Alternatively, a country speaking Portuguese, for example, would have far fewer alternative partners with which to trade in order to avoid the language barrier and therefore might exploit those opportunities more intensely.¹¹ On this ground, the coefficients of the linguistic variables based on English alone might be exceptionally low apart from measurement error. The second problem could be equally serious. Relying on English alone means drawing numerous distinctions between country pairs who share a common language other than English based upon their English, and proposing a quantitative ordering of linguistic ties between these non-English pairs based on their common English alone. Especially large distortions might arise. The results in column 2 confirm the broad suspicion from these considerations that basing common language on common English alone may lead to particularly poor estimates. The effect of common English does not even show up, except curiously for linguistic proximity.

Column 3 proceeds, probably more reasonably, by examining whether adding the linguistic effects of English separately improves the estimate. Once again, the result is negative. But in this case, we get very confusing outcomes except for linguistic distance. The effects of a common official language, a common spoken language, and a common native language for the broad measures go in opposite directions to those for English alone, and in addition the coefficients of the previous three measures are largely implausible, especially for spoken language where the negative effect exceeds the positive one. Furthermore, in the case of official and spoken language the signs for English alone are negative, which would imply a

negative correction for English, whereas for native language the sign for English is positive, implying a positive correction for English. None of this makes sense.¹² All in all, the clearest conclusion from column 3 is that the effort to estimate the impact of English separately fails, just as it fails in column 2. Column 1 offers the best estimate.

Of course, this is not to deny that English probably has a larger effect on *multilateral* trade or trade with everyone than common Russian, common Spanish, or any other common language. This is true simply by virtue of the language's size. World trade by native English speakers is around 23 per cent of total world trade.¹³ The next highest ratio, for Chinese, is 11 per cent, followed by Spanish 10 per cent and German 9 per cent. For this reason alone, English ought to contribute more to the reduction of linguistic trade frictions with the rest of the world than any other language. However, even on this reasoning, English contributes most to multilateral trade only on average. For the Portuguese-speaking Brazil, situated in South America, there may well be a greater incentive to learn Spanish than English to promote multilateral trade. The same goes for Russian in Eastern Europe or around Kazakhstan, and for Chinese or Malay in Southeast Asia, etc.

The earlier survey evidence supports these last inferences. Of the 2000 small and medium-sized exporting enterprises (SMEs) in the Hagen et al (2006) study that report foreign language training in the last three years (constituting 35 per cent of the sample, we may recall), only a quarter provided it in English. 18 per cent provided the training in German, 15 per cent in French. In addition, three small languages figure in the top ten: Czech (5 per cent), Danish (3 per cent), and Estonian (3 per cent). The following quote (p. 19) is especially noteworthy:

When companies were asked to identify the languages they used in their major export markets it was apparent there is widespread use of intermediary languages for third markets. For example, English is used to trade in over 20 different markets, including the four Anglophone countries, UK, USA, Canada and Ireland. German is used for exporting to 15 markets (including Germany and Austria), Russian is used to trade in the Baltic States, Poland and Bulgaria and French is used in 8 markets (including France, Belgium and Luxembourg).

Thus, even when it is a question of a third language as an intermediary in trade, English is not necessarily the choice.

Unsurprisingly, the special role of English in international commerce emerges best in the

survey evidence of the multinational enterprises (MNEs). These firms do indeed have a striking preference for English in their communication with customers and their own subsidiary companies. 63 per cent of them prefer English and another 20 per cent a mix of English and a home language, basically French in the particular sample (pp. 42-43). This is easy to interpret. These firms have the greatest use for a lingua franca as a coordination device. They face a language problem internally and not only in dealings with customers. In addition, they confront the language problem on a world scale and therefore have a broader world view of the multilateral aspects than the SMEs. Because of their world concerns, English is their most common choice. Interestingly, though, the multinationals' other target languages for future learning besides English also reflect their broader international concerns than the SMEs'. Spanish, Chinese and Arabic figure far more heavily than for the SMEs. French is of lesser interest and German and Italian disappear (see Hagen et al, 2006, p. 45). (The same shift of orientation toward Asia, Latin America and the Middle East in language learning emerged in earlier questionnaires of MNEs.) Notwithstanding, only 29 per cent of the language deficiency that the MNEs wish to repair concerns English, close to the 26 per cent figure for the SMEs.

4. Language learning

Suppose we turn next to language learning and consider the subject in the same spirit as trade, that is, first in terms of learning foreign languages in general and later in terms of the difference it makes if the language is English. Economists possess a theoretical model for dealing with language learning that stems from Selten and Pool (1991) and Church and King (1993) and that was recently extended by Gabszewicz, Ginsburgh and Weber (2011) (GGW). The model uses game-theoretical reasoning but can serve to study the learning decision independently of such reasoning. It admits three fundamental influences on the learning decision: (1) total world speakers of the target language; (2) total world speakers of the home-country language; and (3) the cost of learning. A recent econometric study by Ginsburgh, Melitz and Toubal (2014) (GMT) extends this model beyond GGW by admitting trade with speakers of the target language as a fourth influence on learning. In this case, the trade variable largely absorbs the commercial inducements to learn the target language so that in its presence, the total number of world speakers of the target language reflects essentially the other inducements to learn it, for example, ease of social interaction with people from different cultures and benefits of access to their cultures and their literary and artistic heritages. GMT also introduce linguistic distance (or the inverse of the previous measure of linguistic proximity) and the literacy rate as indicators of the cost of learning

(where trade can be understood as also serving this purpose). All five of the expected signs in the model are intuitive. (1) A larger world population of speakers of a target language should make the language more attractive to learn. (2) Larger trade with speakers of this language should do the same. (3) A larger world population of speakers of the home language should reduce the incentive to learn the foreign one. (4) Linguistic distance should tend to raise the cost of learning the language and (5) literacy should help to learn it.

GMT (2014) apply this model to the learning of 13 important world languages over 193 countries. The 13 languages are, in alphabetical order, Arabic, Chinese, Dutch, English, French, German, Italian, Japanese, Malay, Portuguese, Russian, Spanish and Turkish. The database is the same one as in MT (2014). The study centers on 2005, though the values of the language variables refer to different years over 2001-2007. Importantly, the authors do not study the decision of people to learn the primary language of their country of residence on the ground that those who do not already know this language need to learn it for daily living. There is therefore no concern with the learning of German in Germany or English in the US. But the study does consider the learning of Turkish in Germany and Spanish in the US, for example, though there are native Turkish speakers in Germany and native Spanish speakers in the US.

In testing the model, simultaneity is a problem. Learning has a reciprocal influence on three of the preceding five influences upon it: namely, the world number of speakers of the target language, the world number of speakers of the home language, and trade with speakers of the target language. As regards the first two of these influences, it is possible to handle the difficulty by using native speakers to measure total speakers (since native and total speakers are strongly correlated). But this remedy will not work for the third influence, or trade with speakers of the target language, since learning a language will affect trade with native speakers as well as other speakers of the language. Thus, in order to estimate the impact of this third influence on learning, the study uses an instrument for trade. The instrument is the percentage of the domestic output of the trade-partner countries that is attributable to the native speakers of the target language. Learning the target language will affect this percentage negligibly if at all,¹⁴ whereas the percentage is likely to affect trade with native speakers of the target language. Therefore, the variable is appropriate as an instrument.

One problematic aspect remains. The database contains 2125 zeros for differences between spoken and native language, or learned languages, and only 240 positive values. There would be fewer zeros and more positive values if positive values of learning had been

recorded below one percent of the population. But one percent is a lower threshold in the database.¹⁵ In light of this feature, the authors offer separate estimates for the zeros and the positive values. In the case of the zeros, they substitute values of one for all the positive values in the database and merely study the decision to learn (a value of one) or not to learn (a value of zero). In the other case, where they focus strictly on the 240 observations of learning and keep those observations as they are (as positive percentages), the authors effectively study the percentage of people who learn a foreign language conditional on the presence of some positive learning. In the first case, they apply probit, while in the second case, they apply ordinary least squares. In both cases, they instrument for trade. Effectively, therefore, they use probit with instrumentation in the first case and 2SLS in the other.

Table 4 summarizes the main results. The probit estimates are the marginal effects evaluated at the sample means of the variables. The results look quite favorable on the whole. In the absence of any correction for the endogeneity of trade, all five explanatory variables come out significantly with the right signs (columns 1 and 4) except for literacy in the OLS result (column 4). The results after correction for the endogeneity of trade are obviously more important. In this case, columns 2 and 5 show that in the first-stage estimates (for the correction), the output instrument is strong. After the correction (columns 3 and 6), the size of world speakers of the target language ceases to be significant but otherwise the results remain confirmatory, particularly in the full sample estimate (column 3). In the case of the positive-sample estimate (column 6), linguistic distance and literacy cease to be significant at the conventional 90 per cent confidence level but they remain so at the 88 per cent level. Generally, trade is the most powerful influence on learning in the estimates.¹⁶ However, in its presence, it becomes impossible to detect the positive influence of the world population of the target language on learning. This is the only basic problem in the results. The other four explanatory variables enter significantly (with minor qualification for linguistic distance and literacy) with the right sign.

[Table 4 here]

How does the model fare for English? More specifically, do the results call for dealing with English separately as a special case? The best way to answer this question would be to introduce as many of the languages as possible simultaneously or at least a large number of them and to see if English performs differently than the rest.¹⁷ In such tests, English is always insignificant and so are the rest of the languages. Therefore, Table 5 displays a different set of results. The table shows the errors in the estimates in columns (3) and (6) of

Table 4 after correcting for endogenous trade. More specifically, it shows the means and the standard errors (as well as the *t*-statistics) of the residuals for both sets of results, language by language. This gives an idea of the direction of the errors and how statistically significant they are.

The general impression from Table 5, in conformity with the broad evidence with the use of controls for individual languages (mentioned above), is that the model performs little differently for the separate languages.¹⁸ Perhaps the model performs worst for English if we judge from the mean error of -0.647 for this language in the probit estimate, which is the highest in the sample in absolute terms. This error is also negative (under-prediction), which can be interpreted to mean that English is a world lingua franca, since there is more learning of the language than the model predicts in-sample. However, the standard deviation for English in this estimate is also the highest and it denotes a significant percentage of cases of predicted positive learning when there is none (the *t*-statistic is low, 0.41). Furthermore, in the positive-sample estimate, the mean error for English is almost zero and the lowest of all. This goes entirely against the idea of status as a lingua franca.¹⁹ In the final analysis, the study says that learning English is subject to the same principles as learning other languages. If so, it is wrong to try to assess the future of English in isolation, without allowing for the similar incentives to learn other major languages.

[Table 5 here]

What about the future of English? According to the model, the evolution of trade will have a profound effect. But the effects of trade are notably symmetric. Growth in Chinese/English trade should promote the learning of Chinese in native-English countries just as it should promote the learning of English in native-Chinese countries. Whether it will raise the importance of English relative to Chinese in the world will therefore depend heavily on the evolution of the share of trade with English speakers on the Chinese side relative to the evolution of the share of trade with Chinese speakers on the English side. That is what the econometric model says.²⁰ The influence of demographic changes is simpler to analyze. Suppose for example that the Arabic and Spanish-speaking populations grow fast while numbers in the rest of the world remain constant. Then the Arabic and Spanish-speaking populations will wish to learn fewer foreign languages while speakers of other languages will not wish to learn either more or less Arabic or Spanish. Thus, Arabic and Spanish will become relatively more important. In theory, of course, these same demographic assumptions should mean more learning of Arabic and Spanish in absolute terms, which

would simply reinforce the rise in the relative size of those two languages. According to the results of the tests of the GMT model, however, this reinforcing effect depends entirely on a rise in the share of Arabic and Spanish trade in non-Arabic and non-Spanish-speaking countries and therefore may not materialize. But in any event, the basic demographic assumptions do not favor English. These remarks lend general support to Graddol (1997, 2006) who predicts a significant growth of Chinese, Arabic and Spanish in the future based on trade and demographic trends. In close connection, he also questions how much further the spread of English relative to other languages can be expected to go.

5. Welfare implications

There are areas of international encounter where the welfare benefits of a lingua franca are evident. For example, a lingua franca is of mutual benefit to everyone in travel and inside the world industry of news diffusion. The spread of multinational organizations promotes a lingua franca too. The political clout of the US may be important in explaining the convergence on English as the preferred choice of lingua franca in these cases. But by and large, no other choice would bring the world population as much benefit. True, as Van Parijs (2011) stresses, from a welfare perspective, the distribution of the gains raises issues. Those who learn English as their native language obtain a certain rent. On the other hand, they are also more likely to be unilingual than hundreds of millions of other people. Consider the child born in a small Eastern European country who, by virtue of circumstance, by the age of 15 is fluent in Russian and German, two major world languages, as well as the home language, and perhaps a fourth one because of a monolingual grand-parent. From the language perspective, perhaps this sort of person should be deemed especially lucky. If so, one could easily argue that native English speakers suffer a liability – a winner’s curse. The adoption of their native language as the primary lingua franca in the world only worsens the curse: it reduces their incentives to learn a second language still further. The British Chambers of Commerce (2003-2004) and some other British organizations, like the Nuffield Foundation (see the Nuffield Report, 2000), can be understood to lean toward such a view since they deplore the lower possession of foreign languages and the lower attention to foreign languages in school curricula in the UK than most elsewhere in the EU.²¹

It is indisputable, though, that the advantages of native English speakers receive the most attention. True, if English is adopted as the lingua franca in an international assembly or conference room, native speakers of this language are typically in a better position than

those for whom English is an inferior choice. In this respect, considerations of fairness will occasionally suggest improvements in the operating rules of organizations. Pool (1991), for example, shows that it is possible to reconcile efficiency and fairness in the distribution of the costs of translation in international organizations. There could also be derogations to the use of English (or any lingua franca, for that matter) based on axioms of fairness as a function of the circumstances (for example, who the principal interested parties are).

What about the normative aspects of the dominance of English in science and the academic world? In 2006 a group organized by the American Council of Learned Societies (ACLS) to report on the translation of social science texts issued “A plea for social scientists to write in their own languages” at the annual meeting of the American Council of Learned Societies (see Tymowski, 2006 and also Allen, 2007, p. 29). The group sees no harm if mathematicians and natural scientists worldwide write in English, but does so if social scientists do the same. The idea, therefore, is that the content of social science investigation can only be fully stated in the investigator’s language. If true this would mean that the broad switch to English in social science investigation is indeed a collective drawback and, furthermore, quite unfortunately, that social sciences depend heavily on polyglots and the skills of translators. Yet the only evidence is a reference to the early nineteenth-century writing of Humboldt about the relation of language to empirical perception. In embracing the same position, Hamel (2007, p. 65) refers to the earlier and related philosophical writing of Herder. Many of us would prefer better evidence that the social sciences are more reliant on the diversity of languages in the world in communicating their work than the natural sciences, and that the strong personal inducements social scientists face to express themselves in English impoverish their contributions to knowledge. In a communication to the same Congress to which Truchot (1996) addressed his paper on the (then-current) political pressures facing French natural scientists to write in French, Walter (1996) notably points out that the tendency of scientists to veer toward a single language in writing traces back more than two millennia with the language of choice shifting from Sumerian to Greek to Arabic to Latin. From this perspective, the turn to English in the course of the 20th century was a return to a longer term trend and the 19th-century condition where German, English and French shared the top spot as the preferred languages in science was an aberration. Perhaps the return to the longer time trend is for the better: information travels faster.

This is not to deny the possible uneven impact of the central role of English on the welfare of scientists and scholars in the humanities. How far do native-English scientists and

scholars in the humanities have an advantage over the rest (despite their smaller incentives to learn a second language, to their possible regret outside of their professional life and maybe inside of it too)? A study by Sandelin and Sarafoglou (2003) has some indirect bearing on the subject. The authors examine whether native-English countries (namely, Australia, Canada, Ireland, New Zealand, the UK and the US in their study) display a larger number of articles in professional publications by resident scientists or scholars per habitant than other countries in a sample of 30 countries, and the authors do so separately for social sciences, arts and the humanities, and the natural sciences. They also control for world size of the principal home language (total world native-Spanish population in the case of Spain, for example, etc.) and GDP. The results show that the native-English countries do indeed host a larger number of articles in professional publications by resident scientists and scholars per capita (that is, in relation to the total national population) than other countries. Interestingly, however, this is true for publications in the natural sciences as well as the social sciences and the arts and the humanities. The coefficient for the natural sciences is around a third as high as for the other two areas, but it is still statistically significantly different from zero at the 95 per cent confidence level. On this evidence, the native-English status of a country favors a higher ratio of home-produced publications to the national population in the natural sciences as well as in the social sciences (quite apart from the world population of native speakers of the home language), if only to a lesser extent. This would suggest, without implying it, that the difference in the welfare advantage of the prominence of English to a native-English scientist or scholar in the three different spheres is only a matter of degree.²² But I know no study bearing directly on the question.

Let us next return to the areas where multilingualism, interpreters and translations are able to overcome the problem of Babel and English is simply “the first among equals,” so to speak. I have argued that these areas are extremely wide and cover the press, television, the internet, publishing (including translation where I had some difficulty defending the theme), and trade, with three exceptions to which I will return. In these areas, the evidence would show that, broadly speaking, while English is in the lead, it faces sharp limits and all other major languages are safe.²³ The evidence concerning language learning is supportive. If English really threatened to marginalize other major languages in these broad areas of life, we would expect that a model of language learning that treated eight or so of the next ten largest languages exactly the same way as English would perform rather badly. Yet the model does fine. Does any general welfare problem arise from the eminent position of English in these cases? It would seem that the answer is negative. From the standpoint of welfare, there is no clear basis for advocating any special international effort to promote or

demote English in the press, television, the internet, publishing and trade.

It might perhaps be wise to encourage foreign language learning in general in the world to promote wages, trade and/or culture. But the place of English relative to other languages is a separate topic. Learning English comes at the expense of learning other languages. For example, the hypothetical argument (which does not seem exactly alien) that everyone should learn English as a second language is quite dubious. The argument would imply that some of the time that is currently spent on learning other second languages by people who do not know English would be better spent on learning English. Yet no one has yet made the case. An appeal to coordination failures would not suffice since such failures will often argue for promoting other languages as there are at least half a dozen of them that are internationally as widely spoken as English in large regions of the world. In addition, the principle of diminishing returns interferes. Even if English is now the most useful language to learn in one's homeland and occupation, once enough others learn it, the best language to acquire may well become another. Ginsburgh and Prieto (2011) offer labor market support. Interpreters and translations should also be kept in mind. Translation renders a lot of learning inefficiently costly (mercifully so), which may also argue for less learning of English.

The cultural areas of English supremacy provide, in my opinion, the only promising ground for the thesis of too much English. Suppose home preference worldwide in the sense that people everywhere, including those with a second language or a smattering of them, prefer to function in their own language in their private lives. The evidence is overwhelming: it comes from the newspapers people read (even if they know a second language), the television they watch, the internet they use, and many of the facts about edition and trade. Interestingly too, the supporting evidence covers the film and the best-seller as well if only we remember that non-native English people mostly view English-language films that are dubbed or, in the case of the majority only when there is no choice, sub-titled, and they mostly read foreign-language best-sellers in translation. How can it then be welfare-improving for English to dominate the song, the film, and the best-seller so much as it does?

The argument that it is welfare-improving must be that, somehow, foreign-language speakers obtain compensation for their sacrifice of home language benefits either through higher quality in other regards or else through lower social costs. As one possibility, the native-English countries are simply better at producing the relevant goods. As another,

benefits of market specialization are at work. Because of a mix of comparative advantage and economies of scale, all countries do best to narrow the range of their production activities. As one manifestation, native-English countries specialize in producing the relevant cultural products. In other words, just as Germany specializes in the high-quality end of many consumer durables or France does in some niches in jewelry and perfumes, the native-English countries occupy this particular turf. The argument has merit with respect to film where, in fact, the dominance is specifically American as such, and the British film industry is just as dominated by Hollywood as the French one. Hollywood also rose to world prominence under the silent film, before “talkies.” Thus, one can even make a case that Hollywood is the fundamental factor, not English, though some of us will remain skeptical because of the contribution of the French, German, Italian, Japanese, Russian, Spanish and Swedish film in the history of the cinema (see also François and van Ypersele, 2002). But in the case of the song and the best-seller, the argument seems wobbly. Where is the benefit in quality or the reduction in social costs? One particular ground for suspicion is that in other related cultural areas, like music without words, photography, painting, sculpture and architecture, the native-English contribution is closer to what we would expect on the basis of relative population size, income, wealth and tradition. There is something special going on in the cultural market where language is concerned.

I would argue that there is a possible long-term danger. Suppose that the budding author and popular singer of talent feel under the same pressure as the scientist and scholar in the humanities to make his/her name in English. Assume also, as seems to be true in science, that the outstanding talents are especially likely to succumb to this temptation. Their odds of success are higher but the lure is at least the same. In fact, we know of several cases of authors in the hall of literary fame who dropped their native language in favor of others with a larger and more impressive literary tradition (never any opposite examples to my knowledge since the move away from Latin and the printing press): Conrad, Kafka, and Ionesco, to mention three. Gogol evidently seriously hesitated between his native Ukrainian and Russian. This is crucially relevant. The threat is not that creative writing will generally dry up in other languages but that the best work will be done in English (a tendency which, if already under way, would help to resolve the paradox I underlined before). In the case of science, where such a tendency is manifest, there is no harm and indeed even a benefit. In literature, however, we cannot reason the same. Literature is an area where language as such is a source of pleasures and enjoyment, an end in itself, rather than a mere instrument for conveying information, or as an outstanding economic contribution by Church and King (1993) has it, a “communication technology.” True, under the previous scenario, English

might continue to evolve and to discover new literary veins. But how much comfort is that? If everyone wrote music for the violin, likewise music might still thrive and we might even hear new echoes of the viola and the guitar and who knows what else issuing from violin strings. Still, what about the keyboard, the winds and the percussion? It does not seem plausible to argue that there would be no loss.

References

- E. Allen (2007) 'Translation, globalization and English' In E. Allen, ed., *To Be Translated or not to Be* (Barcelona: Institut Ramon Llull).
- U. Ammon (1998) *Ist Deutsch noch Internationale Wissenschaftssprache? Englisch auch für die Lehre an den Deutschsprachigen Hochschulen* (Berlin: de Gruyter).
- U. Ammon (2006) 'Language planning for international scientific communication: An overview of questions and potential solutions' *Current Issues in Language Planning* 7, 1-30.
- J. Anderson and E. van Wincoop (2003) 'Gravity with gravitas: A solution to the border problem' *American Economic Review* 93, 170–192.
- D. Bakker, A. Müller, V. Velupillai, S. Wichmann, C. Brown, P. Brown, D. Egorov, R. Mailhammer, A. Grant and E. Holman (2009) 'Adding typology to lexicostatistics: A combined approach to language classification' *Linguistic Typology* 13, 167-179.
- D. Berwanger (1987) *Television and the Third World: New Technology and Social Change* (Bonn: Friedrich-Ebert-Stiftung).
- D. Biltereyst (1991) 'Resisting American hegemony: a comparative analysis of the reception of domestic and US fiction' *European Journal of Communication* 6, 469-497.
- BIPE [Bureau de l'information et de la prévision économique] Conseil (1993), *Approche Statistique de la Production Littéraire en Europe*, mimeo.
- C. Brown, E. Holman, S. Wichmann and V. Velupillai (2008) 'Automatic classification of the world's languages: A description of the method and preliminary results' *Language Typology and Universals* 61, 285-308.

- J. Church and I. King (1993) 'Bilingualism and network externalities' *Canadian Journal of Economics* 26, 337-345.
- D. Crystal (2003a) *English as a Global Language* (Cambridge, UK: Cambridge University Press, 2d edition).
- D. Crystal (2003b) *The Cambridge Encyclopaedia of the English Language* (Cambridge, UK: Cambridge University Press, 2d edition).
- D. Crystal (2006) *Language and the Internet* (Cambridge, UK: Cambridge University Press, 2d edition).
- A. Dalby (2002) *Language in Danger* (London: Allen Lane: The Penguin Press).
- R. Dearing (2006) 'The Dearing Report to the Secretary of State for Education & Skills' *The Languages Review* (available online).
- A. De Swaan (2001) *Words of the World* (New York: Wiley).
- J. Diamond (2012) *The World until Yesterday* (New York: Penguin Group).
- P. Egger and A. Lassmann (2012) 'The language effect in international trade: A meta-analysis' *Economics Letters* 116, 221-224.
- Ethnologue *Languages of the World* (Dallas, USA: Summer Institute of Linguistics, available online).
- Eurobarometer (2006) 'Europeans and their languages' *Special Eurobarometer 243* (Brussels: The European Commission).
- A. Feely and D. Winslow (2005) *Talking sense. A Research Study of Language Skills Management in Major Companies* (London: CILT, the National Center for Languages).
- J. Fidrmuc and V. Ginsburgh (2007) 'Languages in the EU: The quest for equality and its cost' *European Economic Review* 51, 1351-1369.
- F. Foroutan and L. Pritchett (1993) 'Intra-Sub-Saharan African trade: Is it too little?' *Journal of*

African Economics 2, 74-105.

P. François and T. van Ypersele (2002) 'On the protection of cultural goods' *Journal of International Economics* 56, 359-369.

J. Frankel (1997) *Regional Trading Blocs in the World Trading System* (Washington DC: Institute for International Economics).

J. Gabszewicz, V. Ginsburgh, and S. Weber (2011) 'Bilingualism and communicative benefits' *Annals of Economics and Statistics* 101/102, 271-286.

V. Ginsburgh, J. Melitz and F. Toubal (2014) 'Foreign language learning: An econometric analysis' CEPR Working Paper, number to be assigned.

V. Ginsburgh, I. Ortuno-Ortin and S. Weber (2007) 'Learning foreign languages. Theoretical and empirical implications of the Selten and Pool model' *Journal of Economic Behavior and Organization* 64, 337-347.

V. Ginsburgh and J. Prieto (2011) 'Returns to foreign languages of native workers in the European Union' *Industrial and Labor Relations* 64, 599-618.

V. Ginsburgh, S. Weber and S. Weyers (2011) 'The economics of literary translation: Some theory and evidence' *Poetics* 39, 228-246.

D. Graddol (1997). *The future of English* (London: British Council).

D. Graddol (2006) *English Next* (London: British Council).

C. Hagège (2009) *On the Death and Life of Languages* (New Haven: Yale University Press).

S. Hagen with J. Foreman-Peck, S. Davila-Philippon, B. Nordgren and S. Hagen (2006) *ELAN: Effects on the European economy of shortages of foreign language skills in enterprise*, (London: CILT, The national center for languages).

R. Hamel (2006) 'The dominance of English in the international scientific periodical literature and the future of language use in science' *AILA Review* 20, 53-71.

- O. Havrylyshyn and L. Pritchett (1991) 'European trade patterns after the transition' Policy Research Working Paper 748, World Bank, Washington DC.
- K. Head and T. Mayer (2013) 'Gravity equations: Workhorse, toolkit and cookbook' CEPR Discussion Paper No. 9322, London.
- C. Hoskins and S. McFadyen (1991) 'The U.S. competitive advantage in the global television market: Is it sustainable in the new broadcasting environment?' *Canadian Journal of Communication* 16, 207-224.
- B. Kachru (2010) *Asian Englishes: Beyond the Canon* (Hong Kong: Hong Kong University Press).
- H. Ku and A. Zussman (2010) 'Lingua franca: The role of English in international trade' *Journal of Economic Behavior & Organization* 75, 250-260.
- J. Melitz (2007) 'The impact of English dominance on literature and welfare' *Journal of Economic Behavior & Organization* 64, 193-215.
- J. Melitz (2008) 'Language and foreign trade' *European Economic Review* 52, 667-699.
- J. Melitz and F. Toubal (2014) 'Native language, spoken language, translation and foreign trade' *Journal of International Economics* 93, 351-363.
- Nuffield Foundation (2000) *Languages: the next generation; the final report and recommendations of the Nuffield Languages Inquiry* (London: Nuffield Foundation).
- N. Ostler (2010) *The Last Lingua Franca. English Until the Return of Babel* (London: Allen Lane).
- J. Pool (1991) 'The official language problem' *American Political Science Review* 85, 495-514.
- B. Sandelin and N. Sarafoglou (2004) 'Language and scientific publication statistics' *Language Problems and Language Planning* 28, 1-10.
- G. Sapiro (2010) 'Globalization and cultural diversity in the book market: The case of literary

translations in the US and in France' *Poetics* 38, 419-439.

The British Chambers of Commerce (2003-2004) *BBC language survey. The impact of foreign languages on British business – Part 1*, 2003; *Part 2*, 2004 (London: The British Chambers of Commerce).

The Economist (1997) 'Schools brief: A world view' November 29-December 5, 91-92.

M. Tracey (1985) 'The poisoned chalice? International television and the idea of dominance' *Daedalus* 114 Fall, 17-56.

C. Truchot (1996) 'La langue française en sciences un cas de figure : la situation linguistique des langues en France' *Le français et les langues scientifiques de demain* (Québec : Colloque Université de Québec 19-21 mars).

M. Tsunoda (1983) 'Les langues internationales dans les publications scientifiques et techniques' *Sophia Linguistica* 13, 144-155.

A. Tymowski (2006) *Guidelines for the Translation of Social Science Texts* (New York: American Council of Learned Societies, Annual meeting, Appendix I).

P. Van Parijs (2011), *Linguistic Justice for Europe and the World* (Oxford: Oxford University Press).

T. Varis (1984) 'The international flow of television programs' *Journal of Communication* 34, 143-152.

H. Walter (1996) 'L'évolution des langues de la communication scientifique' *Le français et les langues scientifiques de demain* (Québec: Colloque Université de Québec 19-21 mars).

L. Wiesenfeld (1999) *The World Almanac and Book of Facts* (Mahwah, NJ, USA: Primedia Reference Inc.).

Z. Yong and K. Campbell (1995) 'English in China' *World Englishes* 3, 377-390.

Table 1

Worldwide speakers (millions)

Language	<i>Ethnologue</i> native speakers (1)	<i>Ethnologue</i> total speakers (2)	MT native speakers (3)	MT total speakers (4)
Chinese ¹	1197	>1197	1161	1165
English ²	335	>765	357	1123
Spanish	406	466	401	479
Hindi/Urdu ²	324	>387	–	–
Arabic ¹	206	246	244	272
Russian	162	272	184	267
French	68	118	69	260
Bengali	193	250	–	–
Portuguese	202	217	209	222
German	84	112	89	168
Malay ²	23	>163	22	158
Japanese	122	123	126	126
Turkish/Azerbaijani/Turkmen ²	83	>83	91	102
Italian ²	61	>61	64	77
Dutch/Afrikaans ²	28	>28	22	37

¹Chinese and Arabic are "macrolanguages" in *Ethnologue*'s terms, ones that group together native speakers of distinct and often mutually unintelligible dialects. These are single languages only on the basis of custom and the tendency of native speakers to identify themselves with the general label (Mandarin serving as the main reference point for Chinese, standard Arabic for Arabic). The 1197 figure for Chinese combines Mandarin (population 847.8), Gan Chinese, Hakko, Huizhou, Jinyu, Min Bei, Min Dong, Min Nan, Min Zhong, Pu-Xian, Wu (Shanghainese, 77.2), Xiang, and Yue (Cantonese, 62.2). While *Ethnologue* cites a figure of 178 million L2 speakers for Mandarin, the vast majority of these are native speakers of a separate Chinese dialect and I have no way of knowing how many are so rather than native speakers of a foreign language (though I believe the number is small). This explains the inscription >1197 in column 2. As regards Arabic, *Ethnologue* draws upon Wiesenfeld (1999), a world almanac.

²The ">" inscriptions in column 2 call for explanation. As regards the >387 for Hindi/Urdu, *Ethnologue* cites a precise figure of 120 million for L2 speakers of Hindi, but this figure includes an uncertain number of the 57 million Urdu speakers. That is why I have added a total of 63 million (120 minus 57) to 324 million in this column and the 387 figure therefore yields a minimum. As concerns the other five ">" signs, *Ethnologue* explicitly says "over 430 million" for L2 for English, "over 140 million" for Malay (cited under language of Indonesia rather than Malaysia), and does not provide any L2 figures for Turkish/Azerbaijani/Turkmen, Italian, and Dutch/Afrikaans.

Table 2

Top 9 languages used on the web

Top 9 languages in the internet (1)	Internet users by language (millions) (2)	2000-11 growth in internet (%) (3)	Internet users (% of total) (4)	MT speakers (% of total) (5)
English	565	301	26.8	22.1
Chinese	510	1,479	24.2	22.9
Spanish	165	807	7.8	9.4
Japanese	99	111	4.7	2.5
Portuguese	83	990	3.9	4.4
German	75	174	3.6	3.3
Arabic	65	2,501	3.6	5.3
French	60	398	3	5.1
Russian	60	1,826	3	5.2
Top 9 languages	1,682	421	80.2	80.2
Rest of languages	418	588	19.8	19.8
World total	2,100	482	100	100

Table 3
Linguistic influences on bilateral trade

	(1)	(2)	(3)
Common official language (0,1)	0.351*** (7.56)		0.405*** (5.64)
Common spoken language (0-1)	0.396*** (4.91)		1.244*** (8.55)
Common native language (0-1)	0.284** (2.34)		-0.379** (-2.24)
Linguistic proximity (0-1)	0.078** (4.26)		0.06*** (2.89)
Common official language: English (0,1)		0.084 (1.42)	-0.237*** (-2.66)
Common spoken language: English (0-1)		-0.034 (-0.35)	-1.447*** (-8.38)
Common native language: English (0-1)		-0.001 (-0.01)	0.763*** (3.17)
Linguistic proximity to English (0-1)		0.092*** (2.89)	0.083** (2.32)
No of observations	209,276	209,276	209,276
Adjusted R ²	0.757	0.755	0.758

Student *ts* in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Downward correction of standard errors for country pairs that appear twice the same year, with the opposite country as the exporter (2,850 clusters). 0-1 signifies continuous values between 0 and 1; 0,1 signifies binary values either 0 or 1. Numerous controls are not reported (see text).

Table 4: Foreign language learning

	Full sample			Positive sample		
	Probit	IV Probit		OLS	IV 2SLS	
		First stage	Second stage		First stage	Second stage
	(1)	(2)	(3)	(4)	(5)	(6)
Speakers of acquired languages (log)	0.014*** (4.348)	0.001 (0.720)	-0.001 (-1.109)	0.024* (1.841)	-0.001 (-0.154)	-0.032 (-1.306)
Speakers of native languages (log)	-0.015*** (-3.992)	-0.000 (-0.720)	-0.003*** (-4.049)	-0.024*** (-4.412)	0.002 (0.754)	-0.029*** (-3.384)
Trade with acquired language countries (0-1)	0.465*** (9.243)		0.263*** (3.828)	0.788*** (4.688)		2.665*** (4.129)
Distance between native and acquired language (0-1)	-0.317*** (-6.966)	-0.079*** (-4.657)	-0.058*** (-5.293)	-0.355** (-2.197)	-0.062 (-1.295)	-0.279 (-1.633)
Literacy rate in learning countries	0.249*** (5.323)	0.010 (1.466)	0.041*** (3.292)	0.064 (0.570)	-0.109* (-1.852)	0.286 (1.536)
Instrument (GDP ratio)		0.524*** (11.570)			0.373*** (4.232)	
No. of observations	2,365	2,365	2,365	240	240	240
R-squared		0.202		0.236	0.156	
No. of countries	193	193	193	94	94	94

Student *t*s in parentheses. These are based on robust standard errors clustered at country level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Intercepts are not reported.

Table 5
Residuals of IV regressions by language

Language	Full sample			Positive sample		
	Mean ^(a)	Std. dev.	tstat	Mean ^(a)	Std. dev.	tstat
Arabic	-0,190	0,652	-0,291	0,140	0,229	0,611
Chinese	-0,244	0,473	-0,515	-0,218	0,011	-19,143
Dutch	-0,209	0,360	-0,580	0,049	0,221	0,220
English	-0,647	1,561	-0,414	0,015	0,434	0,035
French	0,005	0,795	0,007	0,035	0,208	0,168
German	-0,075	0,659	-0,114	-0,101	0,184	-0,550
Italian	-0,062	0,676	-0,092	-0,065	0,139	-0,466
Japanese ^(b)	-0,214	0,554	-0,387			
Malay	-0,070	0,253	-0,278	0,416	0,297	1,400
Portuguese ^(b)	-0,170	0,274	-0,618	-0,120		
Russian	0,079	0,597	0,131	0,050	0,252	0,199
Spanish	-0,207	1,196	-0,173	-0,021	0,226	-0,093
Turkish	-0,114	0,326	-0,352	0,045	0,186	0,239

(a) Estimates of the positive sample are based on Pearson residuals from the Probit regression in Table 4, column 3 and those of the positive sample are based on the IV regression in Table 4, column 6.

(b) Japanese is not acquired. Portuguese is acquired only in Spain (no standard deviation).

ENDNOTES

¹ Ostler (2010) does the same. For further detail about the MT data, see Melitz and Toubal (2014). Their data by individual country is available on Toubal's website.

² One source of a possible discrepancy between the last two columns, which is difficult to assess, is that the estimates of internet users assign one language to each user, while the ratios in the last column rest on figures that admit bilingualism so that any speaker may be counted two or more times.

³ The 30 per cent figure for 2009 calls for elucidation since the information from the UNESCO yearbooks started shrinking in the nineties and this source no longer permits calculating this ratio for any date later than 1991. Happily, though, Wikipedia currently compiles a table of "books published per country per year" based on remaining UNESCO data and an assortment of well annotated other sources. According to the Wikipedia table, publications by the US and UK alone bring up the total for titles in English to 24 per cent (as it is possible to infer because these two countries publish little in foreign languages). The main difficulty in assessing how much above 24 per cent the ratio is comes from the absence of the information that UNESCO used to provide about titles in English published by non-native-English countries. In reaching the 30 per cent figure, I allowed for the publications in English in the Wikipedia table by India, Pakistan, Hong Kong, Canada, Australia, South Africa and New Zealand besides the US and the UK. This brought the total up to around 27.4 per cent. I then made an allowance of 2.6 per cent more for published titles in English by the rest. I believe this 2.6 per cent estimate to be a reasonable one since at the time when detailed information was available in the UNESCO Yearbooks, 3.6 per cent of the publications of the rest were in English around 1971, 2.1 per cent around 1981, and 2.6 per cent around 1991.

⁴ The trend continued beyond 1987. Table 1 in Sapiro 2010, based on the UNESCO online source *Index Translationum*, shows a rise in the ratio of English in translations from 45 per cent in 1980-89 to 59 per cent in 1990-99.

⁵ It is also clear from Sapiro (2010) that French publishers, at least, are particularly concerned with translations of literary works from other source languages besides English.

⁶ Of course, on average home-sales of home-language works that are translated are many times larger than home-sales of the rest of home-language works, and this is indeed likely to lead to larger world sales of translated works that are originally written in the language of the large community. But this raises a distinct consideration. Moreover, this next consideration, by itself, has no hope of resolving the puzzle. Translated works that are originally written in the other large languages besides English also sell many more times at home than the rest.

⁷ For some interesting discussion of the varied motives for choosing an official language or a number of them when no choice is obvious, see De Swaan (2001, *passim*).

⁸ The measures of spoken and common native language concern the probability that two people at random for a country pair will share the same spoken language or the same native language, respectively, as the case may be. The measure of linguistic proximity refers instead to similarities of a limited list of words with identical meanings based on expert judgments, where these judgments come from the Automatic Similarity Judgment Program, an international project by ethnolinguists and ethnostatisticians. The measure of common official language is the usual binary one.

⁹ This is similar to the result in an earlier study of similar inspiration based on poorer data by Melitz (2008), which had used only two measures of a common language, one of them mostly based on official status and the other strictly on spoken language.

¹⁰ This is, of course, the same point that Anderson and van Wincoop (2003) make in explaining why national trade barriers form a far weaker incentive for bilateral trade between two US states than between two Canadian provinces.

¹¹ Of course, for that very reason, people in the Portuguese-speaking country would have stronger incentives to become multilingual. This diminishes the weight of the point without denying it altogether.

¹² On the other hand, the result for linguistic proximity is sensible: namely, linguistic proximity between languages, regardless of the pair of languages, boosts trade with a positive correction for linguistic proximity to English. The strong results for linguistic proximity for English in Table 3 (for which I have no basic explanation especially in light of the rest of the results) find an echo in a study by Ku and Zussman (2010) of the impact of a common language on bilateral trade that focuses on English alone. These authors implicitly rest the impact of common English on bilateral trade strictly on linguistic proximity since they rely on bilateral differences of scores in different countries on TOEFL, the Princeton-administered Test of English as a Foreign Language. These international differences in test scores will be correlated with difficulties of learning English for people with different native languages and it is difficult to ascribe the effect of the scores on bilateral trade to anything else.

¹³ If we multiply aggregate trade by the percentage of native English speakers, country by country, sum up the products over all countries, and divide by the sum of trade over all countries, we get around 23 per cent. This measure means double-counting but since the double-counting is both in the numerator and the denominator it cancels out.

¹⁴ Some effect of learning on the instrument might come from stimulating trade with countries with which there had been no trade before but this factor can be expected to be negligible since

trade is widespread between countries with no common native language based on third languages, interpreters and translations.

¹⁵ Greater precision might also have been questionable.

¹⁶ The quantitative impact of trade is also considerable. Suppose a doubling of the trade share with speakers of the destination language (a 100 percentage point increase). According to the probit result, there is a 26.3 rise in the probability of learning where none exists thus far. In the case of the positive-sample result, which is conditional on some positive learning, a single percentage point rise in the trade share of the destination language will increase learning of the language by 2.66 percent.

¹⁷ If instead one simply introduces a separate term for the world level of English speakers when English is the destination language (or else a dummy variable with a value of one when English is this particular language and zero otherwise), English turns up with a significant coefficient in one test or the other, the full-sample or the positive-value one. But the same is true for most of the other languages in similar tests. The right tests to perform are those I mention where as many languages as possible enter simultaneously.

¹⁸ The enormous *t*-statistic of the mean error for Chinese in the positive-sample estimate is arresting but essentially misleading. This result depends strictly on Malaysia and Singapore, the only two countries where there is any learning of Chinese outside of China (at the one per cent threshold in the study). Thus, the standard deviation rests on these two observations alone. The mean of the error for Chinese in the probit estimate, which depends on all the observations, is actually about the same as in the positive-value sample but with a much higher standard deviation.

¹⁹ However, the result is entirely consistent with the evidence of English's status as such in some limited areas like air traffic control, scientific writing and international sports.

²⁰ Of course, a spurt of teaching of English in school is well under way in China (see Yong and Campbell, 1995) whereas the teaching of Chinese in English-speaking countries remains retarded today. It would indeed be helpful to introduce school curricula in foreign languages in the model (with the appropriate lag) if it could be done (if the data was widely enough available). However, it is not a foregone conclusion that major revision would follow: instruction in a foreign language as a child need not mean ability to converse in the language in adult life. The factors present in the model *may* still be the critical ones.

²¹ See also the Dearing Report to the Secretary of State for Languages and Skills in the UK (Dearing, 2006).

²² This is not the authors' emphasis.

²³ Of course, thousands of minor languages are endangered. See Dalby (2002), Hagège (2009), and Diamond (2012), ch. 10. However, this is a separate point in my view, quite distinct from the spread of English and its implications.