Economic Challenges of Multilingual Societies[∇]

Jan Fidrmuc[*]

Victor Ginsburgh[†]

Shlomo Weber[‡]

March 2006

Abstract

We analyze challenges encountered by multilingual societies: linguistic standardization, linguistic disenfranchisement, and the optimal choice of linguistic regime. While the analysis is conceived generally to apply to any multilingual society, we pay particular attention to linguistic policies in the European Union. We analyze the optimal choice of linguistic regime, taking into account possible externalities between languages and dynamic effects of language proficiency. We also discuss the feasibility of linguistic reform in the EU.

JEL Codes: D70, O52, Z13.
Keywords: Languages, Disenfranchisement, the European Union, Linguistic standardization.

[∇] This paper was prepared for the April 2006 Panel Meeting of *Economic Policy* in Vienna. We are grateful to Paul Seabright and two anonymous referees for their very useful comments and suggestions that lead to a substantial improvement of our paper.

[*] Economics and Finance, and Center for Economic Development and Institutions (CEDI), Brunel University; CEPR, London; and WDI, University of Michigan. Contact information: Economics and Finance, Brunel University, Uxbridge, UB8 3PH, United Kingdom. Email: Jan.Fidrmuc@brunel.ac.uk or jan@fidrmuc.net. Phone: +44-1895-266-528, Fax: +44-1895-203-384. Web: http://www.fidrmuc.net/.

[†] ECARES, Université Libre de Bruxelles, and CORE, Université Catholique de Louvain., Contact information: ECARES, Université Libre de Bruxelles, 50 Avenue F.D. Roosevelt, CP 114, 1050 Brussels, Belgium. Email: vginsburg@ulb.ac.be. Phone: +32-2-650-3846, Fax: +32-2-650-4012.

[‡] CORE, Université Catholique de Louvain, Belgium, Southern Methodist University, Dallas, USA, and CEPR.

> "Languages are raised as flags by dominated people." Claude Hagège (2000, p. 13).
>
> "This world, which puts a banquet for all, then slams the door in the noses of so many, is simultaneously equalizing and unequal: equalizing in the ideas and habits it imposes and unequal in the opportunities it offers." Eduardo Galeano, cited by Tonkin (2003, p. 321).

1 Introduction

Public policies concerning linguistic diversity in various countries and international organizations increasingly appear at the forefront of public debate. Linguistic issues and, in particular, the treatment of minority languages are almost unparalleled in terms of their explosiveness and emotional appeal. As Bretton (1976, p. 447) points out: "Language may be the most explosive issue universally and over time. This is mainly because language alone, unlike all other concerns associated with nationalism and ethnocentrism … is so closely tied to the individual self. Fear of being deprived of communicating skills seems to raise political passion to a fever pitch." The challenges of multi-lingual societies are, of course, not new at all. A well-known example (besides the notoriously famous story of the Tower of Babel) is the *Rosetta Stone*: a royal degree inscribed into a large stone slab in 196 BC to announce the royal cult of King Ptolemy V on the first anniversary of his accession to the Egyptian throne. As Egypt was ruled by the (Greek-speaking) Ptolemaic dynasty at that time, the inscription was rendered both in Greek and Egyptian (the latter being in two versions, the traditional hieroglyphic script and the simpler contemporaneous *demotic* script). After the stone was rediscovered in 1799, its bilingual rendition allowed Jean-François Champollion to decipher the hieroglyphic script.

At present, there are 6,912 distinct languages spoken in 271 countries all over the world. About one third of the world's nation-states have official language provisions in their constitutions. Multilingualism therefore is undoubtedly an important part of the current political debate almost everywhere. One must recognize that attempts to maintain multilingual societies (and to avoid the fate of Babylonia) require willingness on behalf of the participating linguistic groups to make compromises and to accept some sort of *language standardization*. In an opening speech to the European Parliament, Queen

Beatrix of the Netherlands said: "… I am addressing you today in Dutch. At the same time, I am convinced that cooperation in Europe will increasingly demand concessions of us in this field. Unless we want to turn the European Union into a Tower of Babel, we shall have to make every effort to understand each other as clearly as possible."[1] However, "… like religion, language does not lend itself easily to compromise" (Laponce, 1992, p. 599-600).

The main objective of this paper is to present the challenges faced by multilingual societies including the issues of linguistic standardization, promotion or prohibition of regional and minority languages, and political and economic impact of such policies. An important element of our analysis is the trade-off between language standardization and language disenfranchisement. While linguistic standardization may deliver important benefits in terms of greater ease of communication, increased trade, improved economic performance and administrative efficiency, it also causes disenfranchisement when groups of individuals are prevented from using their own language. It may restrict basic linguistic rights, lead citizens to lose their ability to communicate in the language of their choice and even threaten cultural survival.

We proceed with the examination of selected current and past linguistic policies in multilingual countries where the historical path plays a crucial role in the evolution of linguistic policies. For example, most Western European countries became linguistically unified through a combination of conquest and nation-building, whereas Central and Eastern European countries, and many others outside of Europe, remain linguistically fractionalized. Often, an imperial power imposed its control over a country along with its language, leaving behind a linguistically (and often also religiously, or racially) split society. The new national borders established through conquest frequently cut across linguistic boundaries. In other instances, linguistic divisions arose through immigration (see Huntington, 2004, who presents a cautionary tale of the rise of the Hispanic population in the U.S. and the ensuing social and linguistic divisions). We also discuss the linguistic practices espoused by the United Nations and other international organizations.

[1] Address by Her Majesty the Queen of the Netherlands to the European Parliament in Strasbourg, 26 October 2004. See http://www.koninklijkhuis.nl/content.jsp?objectid=4096 for the original (Dutch) version and http://www.koninklijkhuis.nl/content.jsp?objectid=4099 for the English translation of the speech.

While our analysis can be applied to a wide range of multilingual (multiethnic) nations, multi-lateral federations as well as multi-national organizations, specific emphasis is given to the question of linguistic arrangements (and the prospect for a linguistic reform) in the European Union. The EU plays an increasingly important role in determining the legal framework and resolving legal conflicts in its member countries. The EU also stands out with its commitment to extensive multilingualism that is largely unparalleled elsewhere, which makes it a particularly suitable object for an analysis of linguistic policies.

We assess the relative importance of European languages, considering the *power* of languages as measured by the Shapley value and by the disenfranchisement rates that would prevail if the set of EU official languages were limited to that particular language or a combination of languages. We use the insights generated by that analysis to motivate our discussion of optimal sets of official languages. The issues of sequencing and linguistic distance are crucial in this context: because some languages are more or less widely spoken outside of their countries and because some pairs or groups of languages are relatively similar while others are very different. We formulate a procedure for selecting the optimal *reduced* sub-set of m official languages from among all *eligible* languages so as to minimize the resulting disenfranchisement rate. We present a sequence of such optimal sets for every value of m between 1 and 20 official languages, and an alternative sequence that, besides considering disenfranchisement, also takes account of linguistic distance. Both sequences suggest that an intermediate scenario, such as six official languages (English, French, German, Polish, Italian and Spanish) may be the best solution to the linguistic challenges that the EU faces. Such a solution would result in both relatively modest disenfranchisement and low overall costs. Last but not least, we argue that based on the distribution of linguistic skills across different age cohorts at present, it is reasonable to expect the disenfranchisement rates to fall further in the future. Hence, even if a reform results in excessive disenfranchisement at present, its adverse effects will diminish over time simply because of changing patterns of language acquisition.

We complement our analysis with a theoretical model of linguistic-regime choice. We show that in the presence of externalities due to delays and miscommunication, it may be optimal to reduce the number of official languages and/or to scale down the

extent of translation for some languages. We also show that a coordinated decision on the optimal extent of translation leads to lower externalities than a non-cooperative equilibrium where each country decides its own translation regime, ignoring the externalities that this imposes on others.

Languages and linguistic regimes can have important economic implications, both at the individual and aggregate levels. We therefore conclude our analysis with a discussion of the costs of multilingualism. Besides the direct financial costs and externalities due to delays and miscommunications, we argue that there are also substantial opportunity costs. In particular, languages play an important role in shaping the flows of international trade, investment and migration. Multilingualism is therefore costly also because of the foregone potential gains from increased economic interactions between countries. In the face of linguistic heterogeneity, individuals who possess linguistic skills instead serve as facilitators of international transactions and therefore they often enjoy considerable private labor-market returns.

The paper is organized as follows. Section 2 presents a broad view of language use and linguistic policies in Europe and elsewhere. In Section 3, we analyze multilingualism and disenfranchisement in the EU and formulate criteria for finding an optimal *reduced* set of official languages, or a sequence of such optimal sets. Section 4 presents a theoretical analysis of optimal choice of linguistic regime in the presence of linguistic externalities. Section 5 discusses the various costs of multilingualism in the EU. Finally, Section 6 is devoted to concluding remarks.

2 Multilingualism and Linguistic Standardization

2.1 Multilingualism in Europe and in the World

According to the Ethnologue database, there are 6,912 living languages on the face of the Earth. Since there are only 271 nations, dependencies and other entities (according to the CIA World Factbook[2]), a large number of countries, if not most, should be in fact multi-ethnic and multilingual. Of course, many of these nearly seven thousand languages are spoken in small and often remote and isolated communities. For example, the Ethnologue database identifies 820 living languages in Papua New Guinea alone (the population of

[2] See http://www.ethnologue.com/ and http://www.cia.gov/cia/publications/factbook/, respectively.

Papua New Guinea is 5.4 million!), 737 in Indonesia, 415 in India, 291 in Mexico, 235 in China, 214 in the D.R. Congo, 188 in Brazil and 162 in the US.[3] While some of these languages are spoken only by a few mainly elderly people and as such are very likely to disappear soon, many appear viable: not counting the languages denoted as *nearly extinct*, the numbers are 797 in Papua New Guinea, 705 in Indonesia, 411 in India, 283 in Mexico, 235 in China, 214 in the D.R. Congo, 158 in Brazil and 94 in the US.

Although most of the aforementioned examples are less developed countries, linguistic and ethnic heterogeneity is certainly not exclusively a third-world phenomenon. In Europe, for example, despite a long tradition of the nation-state, every country – with the possible exception of Iceland – is multilingual (see Table A1 in the Appendix for further details), and there is a rich heritage of indigenous languages, such as Welsh in the UK, Catalan and Basque in Spain and Frisian in the Netherlands.

The challenges encountered by multilingual societies include the issues of linguistic standardization, promotion or prohibition of regional and minority languages, political and economic impact of such policies, and their fairness.[4] Linguistic standardization is often necessary to prevent communication from becoming excessively costly or outright impossible. A public policy entailing concessions and compromises, however, necessarily imposes important restrictions on the linguistic rights of some segments of the society. For example, the United States has implemented various restrictive language policies throughout its relatively short history: California rewrote its state constitution in 1879 to eliminate Spanish language rights, Pennsylvania made English proficiency a condition of employment in its coalmines in 1897 in a none-too-subtle way to exclude Italians and Slavs, and security concerns during World War I led to unprecedented bans on public use of the German language in schools, on the street, during religious services, and even over the telephone (Crawford, 1997).

While linguistic standardization may deliver important benefits in terms of increased trade, improved economic performance and reduced social conflict, it inevitably brings about the problem of linguistic disenfranchisement, introduced in

[3] Only languages of indigenous and established populations are included in these figures. Thus, the figure for the US includes, besides English and many Native American languages, also immigrant languages such as Pennsylvania German and Louisiana Creole French – but not the standard German or French spoken by more recent German and French immigrants to the US.

[4] See Van Parijs (2003).

Ginsburgh and Weber (2005). Disenfranchisement arises as a result of imposing restrictions on the linguistic rights of different linguistic groups. As suggested by Longman (2004), "provision should be given, in a polity that seeks democratic legitimacy and widest possible social acceptance, to facilitating participation in, and comprehension of, political deliberation in a language one understands and can communicate in effectively."

The issues of linguistic standardization and disenfranchisement often arise in the context of the choice of official languages. In a world of 6,912 living languages the designation of languages to be used for official purposes could be a difficult political issue that has far-reaching personal and social implications. As Pool (1991, p. 495) points out, "[t]hose whose languages are not official spend years learning others' languages and may still communicate with difficulty, compete unequally for unemployment and participation, and suffer from minority or peripheral status." Pool outlines the reasons for which it is difficult to find a stable and fair resolution of the issue of official languages. Those include the divisive, symbolic and contentious nature of language conflict, inherent incompatibilities between language communities, reluctance of the majority group to concede linguistic rights to minorities, the power of civil servants to protect their linguistic privileges, and the important and unpredictable material and symbolic consequences of linguistic choices.

Even the meaning of an official language varies enormously and crucially depends on the political context. For example, it would mean one thing for a small "unilingual" country like Iceland, to declare an official language as an emblem of national pride. It would have a different meaning for the United States to do so, since the political impact would be to restrict and denigrate minority tongues that already are subordinate to English. And it meant something else again when, in 1991, Puerto Rico ended its official bilingualism, imposed by military force in 1902, in favor of Spanish as its sole official language.[5] It is not uncommon for language standardization to be explicitly directed against minority languages. When this happens, it comes at an enormous cost to civil liberties. Turkey, for example, has outlawed minority language usage in many contexts. Others elevate a single national language for purely symbolic, ceremonial purposes. In

[5] The new policy lasted barely two years, an indication of its unique political subtext: combat between Puerto Rico's statehood and its "commonwealth" status.

modern times, constitutions often include explicit provisions for minority rights, giving official status to more than one language. In practice, some guarantees are meticulously observed; some are ignored. Language laws may also serve to facilitate communication, for example, in post-colonial nations that remain linguistically diverse, or to mediate ethnic rivalries (Crawford, 1997).

Over the course of history, different countries developed a wide range of policies to deal with linguistically fractionalized societies. In this respect, the examples of practices adopted by linguistically diverse countries are instructive. Russia and China, for example, elevated the language of the dominant ethnic group to the status of a universally-used common language, while allowing extensive use of minority languages at the regional level. Similar practices used to be followed in European polities such as the United Kingdom and Spain. However, the extension of democracy typically translated into devolution and increased regional autonomy and, in turn, a greater use of minority languages. The contrast between linguistic policies in Spain under Franco and at present is particularly stark. India, a country where print media use 87 languages, radio broadcasts 71 languages and schools use 47 languages as medium of instruction,[6] undertook to formulate and implement a fair solution to linguistic diversity in the late sixties. The so-called Three-Language Formula was intended to strike a balance between the needs of individuals, communities and the country. Under this formula, which was a compromise between various pressure groups, it was envisaged that people from non-Hindi areas should learn Hindi, English and their regional language, whereas Hindi speakers would study Hindi, English and a third language. This approach was selected to accommodate the interests of group identity (mother tongues and regional languages), national pride and unity (Hindi), as well as administrative efficiency and technological progress (English).[7] However, the policy eventually failed in practice for the same reasons for which the formula had been introduced in the first place. Hindi-speaking states did not enforce the curriculum, and anti-Hindi government in Madras removed all teaching of Hindi in Tamil Nadu. Thus, even the best intentions and careful policy planning face formidable challenges when support of linguistic groups involved is absent.

[6] See http://www.ciil.org/calender/report.html.
[7] See Sridhar (1989).

European countries adopted a wide range of approaches to dealing with linguistic diversity. Some grant nation-wide recognition to two or more languages: Belgium, Finland, Ireland, Luxembourg, Malta, and Switzerland. Other countries have recognized (some) minority languages at the regional (provincial) or municipal level: Austria, Bulgaria, Croatia, Cyprus, Germany, Hungary, Italy, Lithuania, the Netherlands, Slovakia, Slovenia, Spain, and the United Kingdom. Usually, such recognition is granted in provinces, regions or municipalities where the linguistic minority constitutes a significant share of the population. Finally, the last group of countries includes those that only recognize a single official or national language: the Czech Republic, Denmark, Estonia, France, Greece, Iceland, Latvia, Norway, Poland, Portugal, Romania, Sweden, Turkey.[8]

The countries in Europe that do not endorse multilingualism are not necessarily less linguistically diverse than those that do. France is a case in point: despite the presence of a broad range of deeply-rooted linguistic minorities, the official practice is one of exclusive use of the French language at all levels of government, public administration and education.[9] Furthermore, even countries that extend regional official status to some languages do not grant the same privileges to all minority languages. Italy, for example, only conceded to recognize German in South Tyrol after a long and violent nationalist campaign. Its official recognition of Slovene at the municipal level was mandated by post World War II treaties that allowed Italy to retain ethnically mixed areas near Trieste rather than cede them to Yugoslavia. Other minority languages, including Sardinian and Ladin, which received some formal recognition, are hardly ever used in practice. Similarly, Germany accepted to recognize Sorbian (spoken in Brandenburg and Saxony in the former East Germany) after reunification – but continued in its practice of not extending the same privileges to Danish and Frisian.

It is notable also that only indigenous or long-established linguistic minorities benefit from linguistic recognition. In contrast, minority communities that arose through relatively recent immigration are typically not taken into account, even when they account for a non-negligible share of the total population. Examples are Italian and

[8] See Table A1 in the Appendix for more details on recognition of minority languages in Europe.

[9] Ironically, this sometimes results in an apparent discrimination of French citizens. For example, a Spanish Basque can testify in a court in Basque while a French Basque is expected to speak French.

Arabic speakers in Belgium and France, Turkish speakers in Germany and Austria, Russian speakers in the Baltic countries, and Bosnian, Croatian and Serbian speakers in Slovenia.[10] This practice stands in stark contrast with the experience of less developed countries which, after the end of colonial rule, typically continued to extend important privileges to the languages of their formal rulers.

The choice of linguistic regime seems to have important implications for linguistic development of minorities. Returning to the example of France, most members of linguistic minorities are currently either bilingual or monolingual in French with only partial knowledge of the minority language – this despite the fact that, at the time of the French Revolution, only some two-fifths of the French population were native French speakers (see Ortega and Tangeras, 2004). On the other hand, Irish was essentially brought back from the edge of extinction by the linguistic policy of Ireland following independence.

2.2 Multilingualism, Integration and Globalization

In the international arena, some languages have always played a more prominent role than others. After the decline of Latin and French, "globalization proceeds in English" (De Swaan, 2001, p. 186). It may be interesting that the dominance of English had been predicted more than 130 years ago by Alphonse de Candolle, a Paris-born Swiss botanist, who wrote what follows in his 1873 essay on languages: [11]

> "To understand [why English will become the dominating language], one has to think about the reasons for which a language is preferred. During the 17th and the 18th centuries, there were reasons for which French succeeded

[10] Rather than being motivated by racism or xenophobia, this reluctance to extend similar treatment to recently formed immigrant communities is most likely due to the desire to encourage their closer integration into the majority community. The fact that – unlike the indigenous and established minorities – immigrants are usually more widely spread out certainly played a role too. Nevertheless, the main immigrant groups in Western Europe often reach sufficiently high concentrations in the major urban and/or industrial areas to warrant a change in the official linguistic policy.

[11] *Avantage pour les sciences d'une langue dominante et laquelle des langues modernes sera nécessairement dominante au XXe siècle* (On the advantage for science of a language that dominates and on which modern language will necessarily be dominating in the 20th century). It is interesting to point out that Candolle corresponded with Darwin (see the website on The Darwin Correspondence Online Database) and was active in the nature-nurture controversy. On this issue, see R. E. Francher, Alphonse de Candolle, Francis Galton, and the early history of the nature-nurture controversy, *Journal of the History of the Behavioral Sciences* 19 (1983), 341-352.

to Latin in Europe. French was spoken by a large group of learned people. It had the advantage of being close to Latin, which was well-known at the time. An Englishman or a German knew half of the French language if he knew Latin. A Spaniard or an Italian knew 75 percent of it. Everyone could read a paper or follow a discussion in French. Today, civilization has moved to the North of France and population has become larger there than in the South. Because of America, the use of English has increased dramatically. Science is more and more present in Germany, England, Scandinavia and Russia.

Given these new conditions, a language can become dominant only if two conditions are satisfied: (1) it contains enough Germanic and Latin words and forms to be understood by those who know a Germanic or a Latin language; (2) it is spoken by a large majority of civilized (sic) people... English is the only language that satisfies both conditions. The future supremacy of the Anglo-American language is evident: it will be imposed by the growth of populations in both hemispheres."

Table 1 gives indications on the use of languages in Europe and worldwide, in various domains. With the exception of the number of web pages (columns 1-2) where English lost some ground between 1997 and 2000, English not only dominates, but its use even increases over time. Columns 3-4 provide an example of what is happening in the sciences. In 1998, more than 82 percent of the 559,000 articles in chemistry abstracted by *Chemical Abstracts* were written in English. This number was 62 percent twenty years before, at a time when Russian was still in the game. Column 5 shows that though French, Spanish and German are used as official languages in a large number of international organizations, English again leads. Despite the fact that the United Kingdom joined the EU in 1973, French still dominated in 1986 and 1992 as the first language in which EU texts were drafted before being translated. However, English took over after the 1995 enlargement and has kept increasing its lead since then. Following the latest enlargement, the share of English, with 62%, clearly dwarfs not only German but also French (columns 6-9). Columns 10-12 show how this trend translates into language learning. Though Central Europe was traditionally closer to Continental Europe, at present it has more students learning English than French or German. The same trend is

prominent in the rest of the world, as is illustrated by the enrollment of Berlitz language schools.

Insert Table 1

The use of English is on the rise also in Africa and in Asia. Though linguistic policies in African countries are still influenced by their former colonial status and by the language used by the colonial powers (essentially English, French or Portuguese), the situation has been changing swiftly during recent years. The end of apartheid brought South Africa to the forefront in the continent, and this is likely to foster the role of English. The 1998 Constitution in the Democratic Republic of Congo (formerly Belgian Congo and later Zaire) states that its official languages are French and English. Rwanda (another former Belgian protectorate) adopted English as its official language after the 1994 genocide, and English has been introduced in schooling. In Algeria, English is replacing French as the main foreign language thought at school.[12] French is thus losing ground to English in Africa, the only continent in which it still had a chance. In India, Pakistan, Singapore and the Philippines (970 million inhabitants) English has an official status in parallel with indigenous languages. English is also spreading in Japan; the Nissan-Renault alliance, which adopted English as its working language, is an example of this trend.[13]

2.3 Linguistic Standardization in the European Union

The linguistic practices implemented in the individual countries of the EU stand in stark contrast with the linguistic regime adopted by the EU itself, which is proud to claim that its "policy of official multilingualism as a deliberate tool of government is unique in the world. The EU sees the use of its citizens' languages as one of the factors which make it more transparent, more legitimate and more efficient."[14] Deciding which languages are used at the EU level can have wide-ranging implications. Whether or not a language is recognized as an official language affects the transparency of decision-making and may

[12] See Maurais (2003, pp. 14-15).
[13] See Kaiser (2003, p. 200).
[14] See the EU web portal "Languages and Europe" at http://europa.eu.int/languages/en/home.

prevent EU citizens not speaking that language from taking part in the political process. It has also other, less tangible but nevertheless very important implications. For instance, having its language recognized and used by the EU enhances the country's international prestige and recognition, and boosts the feelings of national pride and self-esteem.

Allowing multiple languages is costly. The EU 15 was spending some EUR 686 million annually[15] on translating and interpreting services. In the wake of enlargement, this cost is envisaged to rise to EUR 1.236 million. The EU employs an army of translators and interpreters (estimated to number 6 thousand, not counting external and free-lance collaborators) and is translating one and a half million pages annually. The number of translators is expected to rise further to accommodate fully the languages of the new member countries.

The non-monetary costs are equally important. At the outset of the European integration process, meetings involving six countries with four languages were relatively simple and manageable. With each enlargement, the combinations of languages requiring translations grew exponentially. At present, with EU membership having grown to 25 and the number of languages to 20 (and soon to 21), providing translations and interpretation is not an easy task.[16] In practice, the increasing costs of providing translating and interpreting services have been kept in check by scaling down the scope of services provided. For example, EU bodies increasingly make use of relay translations (that is, translating a text or speech first into one of the core languages and then translating the same text again into the target language). The downside is that relay translations more often result in incorrect translation. The issue of validity of legal documents is also important; national delegations may agree on a text prepared in a single language such as English, even though it is the translated text that is eventually

[15] Unofficial estimates are even larger. *Le Monde*, November 30, 1999, put the cost at 1.8 billion euros!

[16] The original Treaty of Rome recognized Dutch, French, German and Italian as the official languages of the Common Market. Danish, English, Finnish, Greek, Portuguese, Spanish and Swedish were added at later stages. The latest enlargement in 2004 resulted in the addition of Czech, Estonian, Hungarian, Latvian, Lithuanian, Maltese, Polish, Slovak, Slovene. Irish was given the same status in 2005 but the decision is to be implemented only as of 2007. All these languages enjoy the same privileges as the first four. Without a reform, the list of official EU languages is likely to grow even further. In a few years, Bulgarian, Romanian and possibly Croatian will be added. Turkish can be next, either because of Turkey's accession to the EU, or because of re-unification of Cyprus. Furthermore, as has happened for Irish, languages that currently enjoy national or regional official status in their own countries without being used at the EU level can eventually also become official EU languages. A number of other languages such as Luxembourgish, Catalan, Basque, Welsh or Russian, may thus follow suit.

incorporated into national law and becomes legally binding. As the example of Maltese illustrates, sometimes even finding enough qualified translators and interpreters for some languages may be a problem. Translation takes time too: the EU currently has a backlog of 60,000 pages awaiting translation and in some cases important decisions could not take effect because they were not translated in time.[17]

Though any change in EU's linguistic policy requires unanimity (Article 217 of the Treaty of Rome), in practice not all languages are equally often used by the various EU bodies. The argument here is that Article 6 of the same Treaty states that "the institutions of the Community may stipulate in their rules of procedure which of the languages are to be used in specific cases." This allowed each institution to adopt its own internal rules, which typically favor English, French and German as so-called *procedural* languages.[18] These are used for day-to-day communication within the EU bureaucracy and for preparing drafts of official documents. The vast majority of all EU documents are prepared in one of them (see Table 1), with the remaining languages accounting for less than 9 percent of all inputs. In February 2005, the Commission went one step further by suggesting to limit automatic translation of press conferences to English, French and German, which raised immediate protest by Italian and Spanish officials and journalists.

Until the last EU enlargement in May 2004, full multiligualism and simultaneous interpretation was the rule in the European Council, the Economic and Social Committee, and during the plenary sessions of the European Parliament. In preparatory meetings of the Council, a system of interpretation upon request has recently been implemented. While simultaneous interpretation is still used in the Parliament, its members were asked to use simple sentences and to avoid jokes. Full multilingualism is also used in contacts between the EU and its citizens, and all official documents are translated into all the member states' languages. But, ministerial meetings on topical issues and diplomatic meetings are interpreted into the three procedural languages only (Truchot, 2003, p. 102).

[17] In May 2004, the implementation of new directives on financial regulation and transparency of securities information had to be delayed because they were not translated in time (see Wall Street Journal Europe, "A Welcome Break," Editorial. May 17, 2004, p. A8). As the EU has expanded in the meantime, the directives now have to be translated into nine additional languages, necessitating a delay of six months. In 2003, the EU along with other rich countries agreed to allow developing countries to import cheap generic medication to treat diseases such as HIV, malaria and tuberculosis. The implementation of this decision was delayed by more than a year because of the need to translate it into all 20 official languages (see The Guardian, "EU Language Barrier Costing Lives," 28 July 2004)..

Of the approximately 4,000 meetings held every year (before the enlargement), 75 percent did not require simultaneous translation (Truchot, 2003, p. 102). See Appendix 2 for a more detailed description of the rules governing the use of languages in various EU institutions.

Other international organizations tend to be more restrictive with respect to the languages that they endorse. The official languages at the United Nations have been Chinese, English, French, Russian, Spanish, and Arabic since 1973, but its bureaucracy uses mainly English and French.[19] Speeches given in one of the official languages are translated simultaneously into the remaining official languages only. Delegates who wish to address the UN Assembly in any other language can do so only if they arrange translation into one of the official languages.[20] English is the language used by OECD, NATO, IMF, the Word Bank and other international organizations. But these examples are not necessarily relevant for the EU, since none of these organizations has the ambition of achieving political integration.

In conformity with the subsidiarity principle, each country should however be allowed to decide what happens "at home," though even this is not so obvious, as is shown by the following recent example reported by Truchot (2003, p. 107). French regulations state that insurance contracts should be written in French. This was found to be in violation of EU law. A compromise was finally reached so that contracts could be written in French, unless subscribers wanted to use another language, a decision that could not be made without the Commission's consent.

3 Suggestions for a Multilingual Europe

In this section we discuss several approaches which could be followed to help reduce the number of official and working languages in order to avoid the various dangers that were discussed in Section 2. The current status-quo in the EU is that over 90 percent of the written documents are drafted in English, French or German (see Table 1) and are

[18] The European Court of Justice and the European Central Bank, which use French and English, respectively, as their working languages are the main exceptions to this practice.
[19] In 2001, official representatives were asked in which language (English, French or Spanish) they wanted to receive their emails. Out of the 185 members who replied, 126 chose English (including 14 French speaking countries), 39 went for French and 20 for Spanish. See Calvet (2002, p. 154).
[20] See http://www.un.org/Depts/DGACM/faq_languages.htm.

subsequently translated into all the remaining languages. This includes languages which have a small number of speakers, such as Maltese, or languages of populations that would be able to understand a language other than their own. This suggests that the choice of official and working languages should take into account the number of citizens who speak each language, its spread to other countries besides those in which it is the native language as well as its linguistic proximity to other languages.

3.1 Power of Languages: The Shapley and Banzhaf Values

Ever since Shapley (1953) introduced the notion of value, it has been used as a method for an a-priori evaluation of individual player's expectations in a game with multiple participants. In the context of simple games, Shapley and Shubik (l954) used the value concept to estimate the power of voters in various contexts. The Shapley and Shubik contribution generated a large literature that produced a variety of alternative power indices, the most prominent among which is the Banzhaf value (Banzhaf, 1965).

We apply the approach in order to evaluate the relative (and absolute) power of existing languages. It is important to stress that in our framework the role of "players" will be assigned to languages whose power is evaluated on the basis of number of speakers. Let $\Lambda=\{1,\ldots,L\}$ and $N=\{1,\ldots,N\}$ respectively be the sets of languages and of countries. Consider a single country $j\in N$, and for every subset of languages $T\subset\Lambda$ denote by $v^j(T)$ the number of citizens of j who speak at least one language in T. For every $l\in\Lambda$ let $P^l = \{p^l{}_T : T\subset\Lambda\setminus\{l\}\}$ be the distribution of weights over all subsets of languages T that do not contain l, denoted Λ_{-l}, where the weights can be interpreted as the likelihood of adding language l to set T. Then, following Weber (1988), the value $\varphi(l)$ for every $l\in\Lambda$ is a *probabilistic value* if

$$\phi^j(l) = \sum_{l\in\Lambda_{-l}} p_T^l\left[v^j(T\cup l) - v^j(T)\right]. \tag{3.1}$$

This value is a weighted average of the marginal contribution of language l to the set of languages T in country j measured in terms of citizens who speak l but none of the languages in T. If $p^l{}_T$ is a subjective weight of language l when it is added to T, the probabilistic value is simply the expected contribution of language l.

It is important to point out that both the Shapley and Banzhaf values are probabilistic values. The Shapley value is built on the assumption that a language l will

be added with equal probability to any group of languages of a given size and all sizes are equally likely. That is, for the Shapley value

$$p_t^l = \frac{(L-t-1)!\,t!}{L!} \tag{3.2}$$

and the Shapley value of language l in country j is given by:

$$\phi_S^j(l) = \sum_{t=0}^{L-1} \sum_{\{T \in \Lambda_{-l} : |T|=t\}} \frac{(L-t-1)!\,t!}{L!} \left[v^j(T \cup l) - v^j(T) \right], \tag{3.3}$$

where t is the cardinality of T. Aggregation over all countries leads to:

$$\phi_S(l) = \sum_{j=1}^{N} \phi_S^j(l) \tag{3.4}$$

Unlike the Shapley value, the Banzhaf value is generated from equal weights being assigned to all sets of languages that do not contain l. Thus

$$p_t^l = \frac{1}{2^{L-1}}, \tag{3.5}$$

and the Banzhaf value of language l in country j is given by:

$$\phi_B^j(l) = \frac{1}{2^{L-1}} \sum_{T \in \Lambda_{-l}} \left[v^j(T \cup l) - v^j(T) \right]. \tag{3.6}$$

Again, aggregation over all countries yields:

$$\phi_B(l) = \sum_{j=1}^{N} \phi_B^j(l) \tag{3.7}$$

Note that the Banzhaf value does not distinguish among sizes of linguistic groups, whereas the Shapley value favors the contribution of extreme groups, whether large and small. More generally, one can consider probabilistic values that favor various groups of languages, but we leave this for future research.[21]

In order to use these expressions for specific calculations of Shapley and Banzhaf values, denote by $r^j(T)$ the number of individuals in country j who speak all languages from the set $T \subset \Lambda$. Then it can be shown that:

$$\phi_S^j(l) = r^j(l) - \frac{1}{2} \sum_{\{|T|=2 : l \in T\}} r^j(T) + \frac{1}{3} \sum_{\{|T|=3 : l \in T\}} r^j(T) - \ldots + (-1)^{L-1} \frac{1}{L} r^j(\Lambda) \tag{3.8}$$

[21] There are many other approaches to defining the power of a language. See Greenberg (1956), Laitin (2000), De Swaan (2001), Alesina et al. (2003), Fearon (2003), Desmet, Ortuno-Ortin and Weber (2005) among others.

and

$$\phi_B^j(l) = r^j(l) - \frac{1}{2}\sum_{\{|T|=2:l\in T\}} r^j(T) + \frac{1}{4}\sum_{\{|T|=3:l\in T\}} r^j(T) - \ldots + \left(-\frac{1}{2}\right)^{L-1} r^j(\Lambda). \qquad (3.9)$$

The two expressions are very similar. Differences appear in the third and higher-order terms only, which contain the number of speakers who speak all three, four, etc. languages in *T*. These are obviously small numbers, and the Shapley and Banzhaf values lead to very similar results. Therefore, we limit our empirical evaluation to the Shapley value only.

Results of our calculations for EU's six main languages (English, French, German, Italian, Spanish and Polish) appear in Table 2, for three groups of countries: the EU 15, that is the European Union before the 2004 enlargement, AC 10, the group of ten countries admitted in 2004 and the EU 25. The table also reports the native populations of the six languages within the EU 25 as well as the number of persons who speak each language (either because they are native speakers or because they learned it) in the EU15 and EU25.[22]

Insert Table 2

German is the largest language in the EU when considering only the native speakers. When looking at all people who speak the six languages, German ranks second after English and just slightly ahead of French. English clearly is the most widely spoken language, before and after the enlargement. It is also more powerful than the remaining five languages, as shown by the Shapley values.[23] While French and German were equally powerful in the EU 15, German has gained from the enlargement, and is now the second most powerful languages in the Union. This is clearly very different from the numbers that appear in Table 1 concerned with the languages used in the Commission's

[22] These figures are based on two Eurobarometer surveys commissioned by the European Commission: Eurobarometer 54: Special Survey on Languages, carried out in December 2000 in all 15 countries that were members of the EU at the time, and Candidate Countries' Eurobarometer 2001.1, carried out in October 2001 in all 13 countries that were candidates for EU membership then. The surveys were constructed to be nationally representative and contain detailed information about respondents' mother's tongues as well as other languages that they know. We used the survey results to estimate the number of persons who speak each language in the various countries and in the EU as a whole.

[23] Note that the values are normalized so that the value of English is equal to 100.

day-to-day business, where German plays a marginal role. Note also that Polish appears almost as important as Italian and Spanish (and, within the group of new member countries, it even takes precedence over English).

3.2 Measures of Disenfranchisement

Shapley and Banzhaf values describe the power of individual languages, and cannot be used for groups of languages, a valid issue if a country or a union is faced with several official or working languages. A better measure for comparing alternative policy scenarios then is *disenfranchisement* which, for a given group of languages, quantifies the number of citizens who lose their ability to understand and communicate if their language does not belong to the selected group. For any given subset T of Λ, Ginsburgh and Weber (2005) define the disenfranchisement $d^j(T)$ in country j as:

$$d^j(T) = n^j - v^j(T) \tag{3.10}$$

where n^j is the population of country j and $v^j(T)$ is, as before, the number of citizens in country j who speak at least one of the languages in T. When the set T consists of a single language l, the expression above reduces to the evaluation of disenfranchisement for an individual language:

$$d^j(l) = n^j - v^j(l) \tag{3.11}$$

It is interesting to note that the difference n^j - $v^j(T)$ is actually a probabilistic value for the special case where p^l_T=1 when T is an empty set and zero otherwise.

When comparing disenfranchisement across countries, it is more convenient to express it in terms of *disenfranchisement rates* :

$$D^j(T) = \frac{n^j - v^j(T)}{n^j} \tag{3.12}$$

Table 3 exhibits the disenfranchisement rates for individual EU 25 countries.[24] The results allow us to make several observations. First, English again dominates, though, as appears from column (2), if English were the only language to be adopted, 48 percent of the EU 25 population would nevertheless remain disenfranchised. But this share rises to 69 percent if English were replaced by French or German only (see columns (3)-(4)), and

[24] The numbers in the table differ from those reported in Ginsburgh and Weber (2005), and Ginsburgh, Ortuno-Ortin and Weber (2005), since the inclusion of AC 10 countries led us to consider four foreign languages instead of two. This applies to all further computations.

becomes even worse if Italian or Spanish were chosen (85 and 84 percent, respectively, see columns (5)-(6)).[25] Hence, the numbers given in the table show that even though English is the most commonly known language, the size of the disenfranchised population would be unacceptably large if English were to be chosen as the sole official language of the EU. In particular, such a policy would lead to the disenfranchisement of more than one half of the populations of France, Italy, Poland and Spain, four out of the six largest countries in the EU. Disenfranchisement would also be very high in Belgium, Greece, Portugal as well as in all the countries that joined the Union in 2004, with the exception of Cyprus and Malta.

Insert Table 3

Lower disenfranchisement rates could be achieved by choosing English, French and German as official languages. The average rate would drop to 18 percent in EU 15 and 26 percent in EU 25, but Greece, Italy, Portugal, Spain and many Central European countries would still experience excessive disenfranchisement. Adding Italian, Spanish and Polish would reduce disenfranchisement mainly in Italy, Spain and Poland, with only limited benefits for other countries as these languages are not widely spoken elsewhere. The average rate, nevertheless, drops to 9 percent for EU 25.

So far, we have discussed disenfranchisement and linguistic standardization under rather 'naïve' scenarios: choosing these six languages simply because they are the six largest languages in the EU at present, but we show in Section 3.4 that the cluster English, French, German, and the cluster of all six languages discussed above, would be part of an "optimal" set of languages for the EU if this set were constrained to have at least three or at least six languages, respectively.

An alternative approach is to choose languages that are linguistically very different, in order to increase the likelihood that each EU citizen can at least partially understand one of the official languages. Externalities may occur when choosing official languages because of proximity between some languages. To understand why such externalities occur, however, it is important to introduce the concept of linguistic distances.

[25] For Polish, disenfranchisement can only be computed for AC 10, since Polish was not included in the EU 15 survey. However, it is very unlikely that Polish is spoken by large numbers of citizens in EU 15, so

3.3 Distances between languages

An important criterion for finding the optimal set of languages in a multilingual society such as the EU is their distances one from the other. Indeed, if two languages are very close, such as Czech and Slovak or Danish and Swedish, then it may be optimal to choose only one of them as an official language, since the members of the other linguistic group would either have no difficulty in understanding and speaking that language or would be able to learn it easily. Similarities between languages therefore may be important and should not be ignored when analyzing linguistic policies. Disenfranchisement can be reduced not only by choosing a language that is spoken by many but also by choosing one that would be understood by many because of linguistic proximity, even if they have never learned that language formally.

Most European languages have common Indo-European roots, though they may have branched of at different points in time and taken different routes. Indo-European languages have been the object of close scrutiny for very long, leading to the construction of phylogenetic-like language trees, and to dating of separations between languages and divergence times.[26] Distances between all pairs of languages have also been computed by linguists. Here we use the distance matrix calculated by Dyen, Kruskal and Black (1992) to represent the Indo-European languages used today in EU 25 in a tree form based on a clustering algorithm. This tree ignores Hungarian, Finnish, Estonian and Maltese, which do not belong to the Indo-European group.

The tree, represented in Figure 1, identifies clearly the main groups of Indo-European languages spoken in Europe: Romance languages (Italian, French, Spanish and Portuguese), Germanic languages (German, Dutch, Swedish, Danish and English), Slavic languages (Slovak, Czech, Slovenian and Polish) and, somewhat isolated, Greek, and Baltic Languages (Lithuanian and Latvian). Within the first three groups, there are also sub-groups formed by languages that are closer to each other. This is shown on the vertical axis which measures dissimilarity. Within the Romance group itself, Italian and French are closer to each other than to Spanish, which in turn is closer to Portuguese than

---

that the disenfranchisement rate would be of the order of 90 percent for EU 25.

[26] It is thought that the Indo-European peoples originate from Central Russia, with the earliest evidence of their presence dating back to the 5th millennium BC. The break-up into the present-day linguistic families is estimated to have been completed by 3000 BC. See Diamond (1992) and the references cited therein.

to either French or Italian. The Germanic group splits into three sub-groups formed by smaller clusters (German-Dutch, Swedish-Danish and, somewhat more distant, English). A similar picture emerges in the Slavic group. Estonian, Finnish and Hungarian belong to the group of Ugro-finnic languages. The last language, Maltese, has Semitic origins. In summary, eight distinct groups of languages could thus be identified, the first six of which are Indo-European: (1) Romance languages, (2) Germanic languages, (3) English, (4) Slavic languages, (5) Baltic languages, (6) Greek, (7) Ugro-finnic languages and (8) Maltese.

Insert Figure 1

3.4 Sequencing of Optimal Sets of Official Languages

Determining the optimal set of official languages for a multilingual society entails, implicitly or explicitly, a cost-and-benefits analysis. In particular, the society must weigh the benefits of multilingualism (reducing linguistic disenfranchisement) against the costs. The costs go beyond the financial costs of maintaining several parallel languages: there are transactions costs when speakers of different languages interact with each other and there are also costs due to delays caused by the need to translate official documents and costs due to misunderstandings or erroneous translations. Ultimately, the process boils down to deciding what measure of disenfranchisement is acceptable given the costs of implementing a multi-language regime. The analysis that follows is concerned with choosing the optimal subsets of official languages, using disenfranchisement, or disenfranchisement along with linguistic distance, as criteria. The consideration of costs is relegated to Sections 4 and 5.

The number of languages that should be included in the optimal set depends on the rate of disenfranchisement in the EU as a whole. Let m, the cardinality of the set, be given and T_m be a subset of languages that minimizes disenfranchisement over all m-element sets. The sequence ($T_1, T_2, ..., T_{20}$) then is the sequence of optimal sets for all possible values of m=1,…20. More formally, for any number $m \leq L$, an ordered subset $T_m = \{ l_1, l_2, ..., l_m\}$ of Λ, which consists of different languages, is called m-optimal if for any other ordered subset $\{ l'_1, l'_2, ..., l'_m \}$ of Λ the following inequality holds

$$d(\{l_1, l_2, ..., l_m\}) \leq d(\{l'_1, l'_2, ..., l'_m\}). \quad (3.13)$$

where $d(\{l_1, l_2, ..., l_m\})$ is the disenfranchisement attributable to $T_m \subset \Lambda$ in the EU as a whole. Note that such a sequence may fail to exist. It may also happen that (3.13) is satisfied with equality for some sequences, in which case there are two or more sequences which are all m-optimal. As discussed below, this problem does not arrise in our analysis for relatively small values of m. Columns (1) through (3) of Table 4 report one such sequence.[27] The table is constructed in such a way that each row indicates which language should be added to the optimal subset formed by the languages reported in the preceding rows so as to maintain optimality. The optimal subset of one language, T_1, therefore contains English, T_2 contains English and German, T_3 is formed by English, German and French, and so on, until T_{20} adds Maltese and thereby collects all 20 languages.

As before, a person is considered not disenfranchised if she speaks one or more of the languages introduced into the set so far, whether it is her mother's tongue or an acquired language. The analysis therefore only considers knowledge of or proficiency in a language and ignores the fact that some languages are so close to each other that they are partially or fully intelligible. Though this calculation is conceptually simple, in practice it requires a large number of computations if we want to carry it out for all values of m. However, since European languages differ considerably in the numbers of people who speak them, the scope of the analysis can be narrowed down substantially. For instance, it is clear that English should be introduced first, followed by French or German, then the other large languages (Italian, Spanish and Polish) and so on. In this way, the choice of the most suitable combination if often easy and at any stage in the analysis the number of possibilities to be considered is relatively small.

[27] Detailed results for individual countries can be obtained from the authors. Note that many countries appear to show non-zero residual disenfranchisement even their national languages has been introduced into the sequence. This presumably reflects the presence of ethnic or linguistic minorities. In most cases, this kind of residual disenfranchisement is negligible, with the notable exception of Estonia, Latvia (both with large Russian minorities) and Ireland (given that we did not consider the Irish language). Second, introducing a language reduced disenfranchisement also outside the country where the language is officially spoken. The most obvious examples are English, French, German and Spanish which are spoken in most European countries. Interestingly enough and much less obviously, the residual disenfranchisement in the UK and Ireland fall further after the introduction of German, French, Spanish and Dutch, and German and Austrian residual disenfranchisement rates fall after the introduction of Danish.

As pointed out before, given our definition of optimality, it is possible that there exist two or more sequences for which (3.13) is met with equality for some value of *m*. If that happens, then instead of there being a unique sequence of optimal sets, there are two or more such sequences. In the specific case of the EU, the marginal benefit (in terms of reducing disenfranchisement) from adding an additional language falls with the position within the sequence (i.e. the value of *m*). Once *m* exceeds 10, the marginal contribution becomes as low as a fraction of a percent of the EU 25 population per language (for example, the disenfranchisement rates that prevail after the addition of Latvian and Danish only differ at the second decimal point).[28] Correspondingly, the level of confidence that one can put into the specific ordering of languages within the sequence of optimal set falls with the value of *m*.

Insert Table 4

English is clearly the first language in any sequence as it is spoken by 52 percent of EU population. German and French are in close race for the second position; German, with a 36 percent disenfranchisement rate, fares slightly better than French with 37 percent. The three languages together result with a 26 percent disenfranchisement rate. Italian, Polish or Spanish would each make almost the same contribution to reducing disenfranchisement, with Italian just slightly ahead with 19.9 percent (compared to 20 and 20.5 percent for Polish and Spanish, respectively). And with all six largest EU languages included, only 9 percent of EU population would still remain disenfranchised.

Of course, important differences across countries remain, with several countries facing more than 50 percent disenfranchisement rates: Greece, Portugal, Estonia, Hungary, Latvia, Lithuania and Slovakia. The most dramatic case is Hungary where only 26 percent of the population can speak one of the first six languages. Not surprisingly, Hungarian becomes the seventh language in the sequence. In addition to eliminating disenfranchisement in Hungary, this has a positive impact also on Slovakia, with its disenfranchisement rate declining from 66 percent to 51. The gain from adding Czech, in the eighth place, similarly extends to Slovakia (disenfranchisement falls further to 26

[28] The margin of error that is associated with surveys of the size as the Eurobarometer surveys that we use is +/- 2-3 percentage points or 3-4 percentage points at the 95% and 99% confidence level, respectively.

percent). After adding also Greek, the EU with nine official languages would leave less than 5 percent of its population disenfranchised; only in Portugal, Estonia, Latvia and Lithuania would the fraction of disenfranchised population exceed one half. With adding further languages, the marginal gain in terms of the reduction in disenfranchisement falls progressively and the benefits apply typically to a single country. The sequence is concluded, not surprisingly, by Danish, Slovene and Maltese.

Columns (5) to (7) of Table 4 report the results of an analogous exercise that takes into consideration the distance between languages. There are several pairs of languages that are close to each other: Danish and Swedish, Spanish and Portuguese, Dutch and German, Czech and Slovak are the most notable examples (see Figure 1). Since these languages are so similar, the speakers of either one would benefit from the introduction of the other language even if their own language remains left out. This consideration is now added to the analysis. Specifically, columns (1) through (3) of Table 4 reported the fractions of those who understand at least one of the languages included in the set T. The figures in columns (5) to (7), in contrast, correspond to

$$D(T_m) + (1 - D(T_m)) * \rho^2 \qquad (3.14)$$

where $D(T_m)$ is the overall EU 15 or EU 25 disenfranchisement rate of citizens who speak no language in T and ρ is the coefficient of linguistic proximity (Dyen, Kruskal and Black, 1992[29]) to the nearest of the languages already included in the set.[30] Effectively, this implies that, for example, the average Italian would understand 64 percent of a French text (the linguistic proximity is 0.803) but only 6 percent of an English text (linguistic proximity: 0.247).[31] In this case, adding a language translates into two types of gains. First, those individuals who speak this language are no longer disenfranchised.

[29] Dyen et al. only consider Indo-european languages. Therefore, we assume that the distance between all Indo-european languages and the other languages – Hungarian, Finnish, Estonian (Ugro-finnic) and Maltese (Semitic) is zero, that the similarity between Hungarian and either Finnish or Estonian is 0.35 and the similarity between Finnish and Estonian is 0.5.

[30] Ginsburgh, Ortuno-Ortin and Weber (2005) calculations are similar, though they do not square linguistic distance. See also Desmet, Ortuno-Ortin and Weber (2005).

[31] A slightly different approach, whereby the residual disenfranchisement rate is multiplied by the similarity coefficient rather than its squared value, produces very similar results.

Second, disenfranchisement falls also for those who do not speak this newly-introduced language but whose own language is similar to it.

In the single-language scenario, accounting for linguistic proximity makes little difference: the EU-wide disenfranchisement rate drops from 48 to 42 percent. Adding French reduces disenfranchisement also in all Romance-language countries, bringing the EU-wide rate to 26 percent. German still comes in the third position, followed now by Polish and then Spanish. After adding Hungarian and Greek, the disenfranchisement rate drops to 3 percent and the gains from adding further languages are correspondingly limited. Accounting for linguistic proximity lowers also country specific disenfranchisement rates. With four languages (English, French, German and Polish), only Estonia, Hungary and Latvia are left with more than half of their populations disenfranchised. The order at the end of the sequence changes somewhat too, with Slovak, which is very close to Czech, taking the penultimate position now.

3.5 Suggestions for a Set of Official Languages

The EU was formed as a union of independent countries and therefore its policies often give precedence to interests of nations rather than individual citizens. As a result, Malta and Estonia have the same weight as Italy and Poland, despite their vastly different populations. Similarly, Maltese and Estonian, at least in theory, enjoy the same status within the EU as Italian and Polish. While this emphasis on national interests is understandable (and indeed unavoidable) given the institutional framework adopted by the EU, it is also inherently undemocratic. In the context of linguistic policies, it implies that an individual Maltese or Estonian citizen weighs in more heavily than a Pole or Italian. If the EU is to avoid becoming overwhelmed with dozens of languages, it may have to shift the emphasis from national concerns to those of individual citizens; this would also enhance the democratic legitimacy of EU policies.

The tools introduced in the preceding sub-sections can be used to shed some light on what would constitute a reasonable set of official languages. A closer examination of the Shapley values, disenfranchisement rates, distances between languages and optimal sets show that not all languages play an equally important role within the EU. At the same time, it is clear that a unique *lingua franca* will hardly be sufficient as it would result in too high an extent of disenfranchisement, leaving nearly half of the EU

population ‘in the dark’. Similarly, a solution based on English, French and German, would still leave 26 percent of EU population (18 percent if we consider linguistic similarities) disenfranchised, which clearly many would consider unacceptably high. The case for an intermediate solution, with (at least) five or six official languages, appears much more appealing, with the resulting disenfranchisement below 10 percent. The status quo of 20 languages, on the other hand, is, to say the least, not very efficient.

The decision on the set of official languages is inevitably a political one, and boils down to deciding what extent of disenfranchisement is tolerable. As the discussion in Section 2 revealed, essentially all European countries tolerate a certain degree of linguistic disenfranchisement and it would be natural for the EU to do likewise. Whether the optimal set should contain five, six or more languages, however, is difficult to predict.

The inclusion of English, French, German, Polish and Spanish (and possibly also Italian) results in relatively low disenfranchisement and would ensure that four of the eight language groups identified in Section 3.3 are represented. The addition of Hungarian and Greek, which share few similarities with the aforementioned languages, would result in further benefits. The gains from including further languages, however, will be much more modest. This is especially the case with languages that are close to one of the aforementioned languages. For example, Czech and Portuguese are similar in size to Hungarian and Greek – but the rationale for adding them is more limited once Polish and Spanish are in the set of official languages. Furthermore, languages whose principal populations are highly proficient in other languages similarly make only limited contribution to reducing disenfranchisement. Thus, Dutch may be the seventh largest language in the EU but most of native Dutch speakers are proficient in at least one other language. In the case of the Maltese, who are few in numbers and who mostly speak English, the justification for giving their language an official status is even more questionable.

It should be stressed that it may be also necessary to apply a different linguistic regime to the European Parliament. As the Members of the European Parliament are elected by EU citizens rather than appointed, one can hardly impose a linguistic

prerequisite for the job.[32] Nonetheless, the EU could implement a linguistic reform that would maintain multilingualism in the sessions of the Parliament (and possibly also in other meetings of national representatives such as the Council of Ministers) but would at the same time reduce the number of official languages used at other meetings and in the preparation of official documents.

3.6 Dynamics

The disenfranchisement rates in Table 3 give a picture of the situation at the time of the survey (2000-01). However, the knowledge of languages changes over time, mainly because younger generations are usually more proficient than the older ones. And this is indeed the case for the main European languages considered in the paper. To see how this affects the dynamics of disenfranchisement rates, we carry our an analysis whereby generations are shifted forward in time. We make the following simplifying assumptions for each country:

(a) disenfranchisement rates are calculated for four age groups (generations): 15 to 25 years old (young); 26 to 40 (middle 1); 41 to 60 (middle 2); 61 and older (old);

(b) a "newborn" young generation is endowed with the same proficiency in languages as the young generation in 2000-01: there is thus no "additional" learning;

(c) there is no learning after the age of 25, and what was once acquired is not forgotten; this is a convenient assumption, since little is known about language retention over time;

(d) the old generation disappears once it is 81 or older;

(e) generations are simply shifted in time with their observed disenfranchisement rates;

(f) for the sake of simplicity, country populations as well as populations in each age group are assumed to remain stationary between 2000 and 2045; it is quite unlikely that any more sophisticated calculations taking account of demographic trends would fundamentally change the results.

The most crucial assumption is (b). There are good reasons to believe that younger generations become more and more proficient in languages, since this is what observation of the recent past leads us to. Hence, our calculations are likely to underestimate rather than overestimate future disenfranchisement rates, though there are some indications, at

[32] Although, as the example of the United Nations illustrates, it would be possible to give an opportunity to MEPs to address the sessions in their own language if they arrange interpretation at their own expense.

least for France, that the knowledge of foreign languages among young students is deteriorating over time (see French Ministry of Education, 2004).

Total disenfranchisement rates for each country are computed for 2000-01 (the time the surveys were carried out), as well as for 2015, 2030 and 2045 by shifting one by one the four generations defined in (a), taking into account the assumptions (b) to (f). To avoid showing to many results, we concentrate on the dynamics of English, country by country (in Table 5) and the EU-wide dynamics of each of the six languages, as well as groups of languages (the cluster English, French and German and the cluster composed of all six languages).

Insert Table 5

Table 5 illustrates the dynamics for English. At present, 14 out of the 25 EU 25 countries experience disenfranchisement rates in excess of 50 percent. Their number drops to six (Czech Republic, Hungary, Latvia, Lithuania, Poland and Slovakia) by 2015, but the rates are still quite high (56 to 79 percent). In 2045, only five countries remain above the 50 percent threshold and rates are substantially smaller in all of them (54 to 66 percent). The EU 25 disenfranchisement rate drops from 48 percent in 2000 to 25 percent 45 years later, though the average for the 2004 newcomers remains above 50 percent. Finally, knowledge of English remains low in Bulgaria, Romania and Turkey.

This shows that English could indeed be chosen as a unique *lingua franca* by 2045. It would be extremely controversial to implement this kind of change at present, however, with more than one half of the EU 25 countries whose citizens would be disenfranchised.

Table 6 shows that the situation would be even worse if any other of the main European languages were chosen as a unique *lingua franca*. At present, disenfranchisement rates in EU 25 are as large as 69 percent for French and German, 84 percent for Italian, and 85 percent for Spanish.[33] In contrast to English, the disenfranchisement rates for these languages are set to undergo only very modest improvements over time. Choosing groups of language is obviously a much better

[33] Polish was not among the languages listed in the survey for EU 15, but it is probably almost unknown in all countries.

alternative. The combination of English, French and German would leave 26 percent of EU 25 disenfranchised at present, but this rate would gradually drop to 13 percent by 2045. Obviously, the combination of all six languages does even better, falling from nine percent at present to five percent by 2045.

Again, the implications of linguistic reform are much less favorable for the candidate countries. Given that French is well understood and close to Romanian, and that Bulgarian would to some extent benefit from the introduction of Polish, the main sticking point would be Turkey. This raises another important – and at the same time highly controversial – issue. Our analysis so far has been carried out on the assumption that there are 20 *eligible* languages. Without a linguistic reform soon, this will no longer be the case. Irish is scheduled to become an official EU language as of 2007, followed (or accompanied) by Bulgarian, Romanian, possibly Croatian, and eventually Turkish. In just a few years, there may be 25 eligible languages or even more. That would substantially change the balance of power between the languages that we considered. Most notably, Turkish could well become one of the top three EU languages. Therefore, procrastinating the linguistic reform is likely to make the status quo only more complex and the prospect of reform more contentions.

4 The Choice of Linguistic Regime under Externalities

When choosing an optimal set of languages, the society weighs the benefits of multilingualism – reducing linguistic disenfranchisement – against the costs. The costs go beyond the financial costs of maintaining several parallel languages and generate transactions costs when speakers of different languages interact with each other as well as costs due to delays caused by the need to translate official documents and costs due to misunderstandings or erroneous translations. These monetary and non-monetary costs represent negative externalities and hamper the functioning of a multilingual society. The choice of linguistic regime should reflect the interplay between the reduction of disenfranchisement and the negative externalities that it creates. In this section, we formulate a theoretical model that incorporates these aspects of linguistic-regime choice and discuss some of its implications.

Consider a Union comprising L distinct linguistic groups. Denote p^j, $j = 1, 2, ..., L$, the population of linguistic group j. Each group speaks its own language and there is one core language spoken as the native language by group 1, which is assumed to be the default option for all Union communication if no translation is provided. The ability to use one's language constitutes a public good, which can extend to various levels: interacting with family and friends, obtaining formal education, using the language at one's workplace and communicating with the authorities. We assume that all the citizens of group j speak their own (native) language and enjoy the benefits of their own public good. To extend similar privileges with respect to the other languages requires translation to and from the core language which is costly. Since there is no translation needed to the core language $j = 1$, from now on, the analysis covers linguistic groups $j = 2, ..., L$. The translation can be full or partial so that the public good enjoyed by linguistic group j, denoted henceforth by α_j, ranges between 0 (no translation from the core language) and 1 (a full translation regime). For all j, The translation cost $C_j(\alpha_j)$ per language depends on the extent of translation, and is increasing in α_j. Moreover, the cost function is identical for all groups, differentiable, and convex, so that

$$C_j(\alpha_j) = C(\alpha_j),\ C'(\alpha_j) > 0 \text{ and } C''(\alpha_j) > 0, \text{ with } C(0) = 0.$$

We also assume that translation costs, whatever the value of α_j, are shared equally among the members belonging to the linguistic group j.

The representative citizen of the linguistic group j derives the utility $U_j(\alpha_j)$ from the translation level α_j. The utility, which increases in α, reflects the proficiency of individuals in group j in the core language, and the attitude towards the language represented by the group's degree of disenfranchisement (if only partial translation to and from the core language is provided). The utility function is assumed to be differentiable, increasing and concave, that is $U'(\alpha_j) > 0$, $U''(\alpha_j) < 0$.

The costs involved in the translation to and from the core language may impose negative externalities on all members of the Union due to errors, delays and misunderstanding caused by the translation process. Specifically, for every L-tuple of translation levels $(\alpha_2, ...\alpha_L)$ we denote by $G(\alpha_2, ...\alpha_L)$ the expected loss due to miscommunication between groups and translation delays. Since the increased volume of

translation raises the probability of miscommunication and delays, G is an increasing and concave function in the volume of translation. For simplicity, we assume that the expected loss from miscommunication and delays depends only on the aggregated level of translation so that $G(\alpha_2, \ldots \alpha_L) = G(\Sigma_i \alpha_i)$.

We are now ready to consider several possible regimes for determining the levels of translation services.

Uniform Regime: The Union determines a uniform level of translation α

This is the current policy adopted in the EU, whereby the level of translation is equalized for all linguistic groups, $\alpha_j = \alpha$, and translation is also financed jointly. The level α represents the average level translation across the institutions of the Union, which may differ across institutions, as discussed in Section 2.

If the union adopts a welfare approach, it would wish to maximize the aggregate utility $W_U(\alpha)$ of all groups' representative consumers net of (per capita) translation miscommunication and delay costs, $C(\alpha)$ and $G[(L\text{-}1)\alpha]$, where

$$W_U(\alpha) = \Sigma_j U_j(\alpha) - (L\text{-}1)C(\alpha) - G[(L\text{-}1)\alpha] \qquad (4.1)$$

Note that since we assume the language of group 1 to be the core language, translation must take place into the remaining (L-1) languages.

The optimal level of translation, α^*, can be obtained from the first order condition:

$$\Sigma_j U'_j(\alpha^*) - (L\text{-}1)C'(\alpha^*) - G'[(L\text{-}1)\ \alpha^*] = 0 \qquad (4.2)$$

Dual Regime: The Union translates into only a subset of languages

The choice of optimal translation levels supported in this paper argues for a deviation from the uniform approach. Namely, it may be optimal to select a subset T of languages as official languages, for which the level of translation α is determined as in the previous scenario, but the level of translation is reduced to $\beta < \alpha$ for other languages. Which

languages should receive the reduced translation regime can be determined by a political compromise or using an optimization method as that outlined in sections 3.4-3.5. We further assume that the groups who speak the reduced-regime languages are compensated by a lump-sum transfer, $C(\alpha)$ - $C(\beta)$, the difference in translation costs between full and reduced translation.

A policy equilibrium is now constituted by the triple (T, α, β), with $\beta < \alpha$, where $\alpha_j = \alpha$ for all j in the subset T, and $\alpha_j = \beta$ for the other languages. Then the welfare of the Union is

$$W_U(T, \alpha, \beta) = \Sigma_{j \in T} U_j(\alpha) + \Sigma_{i \in T'} U_i(\beta) - (L-1)C(\alpha) - G[t\alpha + (L-t-1)\beta]. \qquad (4.3)$$

where t is the number of groups in the subset T and T' is the set of languages that do not belong to T.

Thus, for every subset T, one can choose the optimal translation levels $\alpha(T)$ and $\beta(T)$ that maximize the expression in (4.3). A simple but important observation is that, in general, there will exist a triple (T, α, β) that generates higher welfare than the uniform policy α^* discussed in the previous scenario. This is so despite the fact that the financial costs of the two regimes are the same, since the groups subject to the reduced-translation regime are compensated. The dual regime, however, results in lower losses due to miscommunication and delays. This supports the argument for selecting a reduced group of official languages.

This is considered in Ginsburgh, Ortuno-Ortin and Weber (2005) who examine optimal sets of official languages, which are determined by two parameters: the sensitivity of the society towards language disenfranchisement of its members and the degree of comprehensiveness of its language regime. The latter is represented by α in the theoretical model, which can take any intermediate form between full translation, and no translation. They also run simulations to derive the optimal sets of official languages over a wide range of the two parameters.

Dual Regime with Group Discretion: Some groups decide on the level of translation for their language

One can modify the previous scenario by allowing each of the L (> 2) linguistic groups that does not belong to subset T to choose its own level of translation, whereas keeping the level of translation for languages in T at the previous level α. It is natural to assume again that each language $j \in T'$ receives a lump-sum subsidy $C(\alpha)$.

Formally, let the subset of t linguistic groups in T and their level of translation α be given. We consider the non-cooperative game among language groups in T' that choose their own level of translation. Given the vector of choices $A = ((\alpha_i = \alpha)_{i \in T}, (\alpha_j)_{j \in T'})$ of all non-core linguistic groups, the representative citizen of group $j \in T$ derives the following payoff:

$$w_j(A) = U_j(\alpha) - C(\alpha)/p_j - G(t\alpha + \Sigma_{i \in T'}\alpha_i) \qquad (4.4)$$

where we assume that the cost from miscommunication and delays, G, is a non-rival public *bad* which applies to each individual equally. For every $j \in T'$, the payoff is:

$$w_j(A) = U_j(\alpha_j) - C(\alpha_j)/p_j - G(t\alpha + \Sigma_{i \in T'}\alpha_i), \qquad (4.5)$$

from which the following conclusions can be derived.

(a) In the traditional literature on public goods with externalities (Olson, 1965), the efficient level of contribution to the public good is higher than the equilibrium one. In our context, translation is a public *good* for the group itself, whereas the overall process of translation is a public *bad* because of the externalities due to miscommunication and delays. Thus, while there is a free riding phenomenon in contributions to public goods ("too little of a good thing"), in our framework we have an over-commitment for a *public burden*. The mathematical argument follows from the comparison of the individual first-order conditions derived from (4.4) and (4.5) with the condition for efficiency. By summing the expressions in (4.4)-(4.5) over all language groups, we derive the aggregate welfare of their representative consumers:

$$\Sigma_{j \in T}[U_j(\alpha) - C(\alpha)/p_j] + \Sigma_{j \in T'}[U_j(\alpha_j) - C(\alpha_j)/p_j] - (L\text{-}1)G(t\alpha + \Sigma_{i \in T'}\alpha_i) \qquad (4.6)$$

and derive the following first order conditions for j$\in T'$:

$$U'_j(\alpha_j) - C'(\alpha_j)/p_j - (L\text{-}1)G'[t\alpha + \Sigma_{i\in T}\alpha_j^e\ ] = 0, \qquad (4.7)$$

where α_j^e is a solution of (4.7). However, the first-order conditions derived from (4.5) yield:

$$U'_j(\alpha_j) - C'(\alpha_j)/p_j - G'[t\alpha + \Sigma_{i\in T}\alpha_i] = 0. \qquad (4.8)$$

Obviously, the choice of α_i^e in (4.7) takes account of externalities due to delays and miscommunication whereas the choice of α_j in (4.8) results from individual and non-cooperative optimization of each group in (4.5) does not. Therefore, the non-cooperative-equilibrium level of translation will be greater than α_j^e. That is, the groups translate too much in order to alleviate their own disenfranchisement, without internalizing the externalities that they impose on the other groups. The policy implication is that a complete decentralization would be wasteful and the Union should sustain a substantial degree of control over the translation channels.

(b) If one group exhibits a higher degree of sensitivity towards its own public good than the other, say, $U_j(\alpha) > U_i(\alpha)$ for all values of α, then group *j* will choose a higher degree of translation in equilibrium.

(c) If two groups have the same utility functions, then the larger group chooses a higher level of translation.

(d) Corner solutions: In equilibrium, every group chooses a positive level of translation α_j. However, there could be a corner solution, in which group *j* declines the option to translate and chooses $\alpha_j = 0$. In particular, if $U'_j(0) - C'(0)/p_j - G'(0) < 0$, group *j* will never choose a positive level of translation as its potential benefits are lower than the associated costs.

Clearly, to implement this model, data on costs of translations, as well as on those generated by delays and mistranslations are needed. The total cost of translation in the EU is known, and one can assume that costs are identical for all languages (ignoring the

fact that some EU languages are used to draft documents, and only a small number of languages is used by most institutions). But the two other costs are very difficult to estimate, since some mistranslations are innocuous, some are less so,[34] and so are the costs due to delays.

Fidrmuc and Ginsburgh (2005) propose a reform based on decentralization, whereby countries (or linguistic groups) are given control over the funds earmarked for translation. In that case, each country would be able to determine its optimal extent of translation and keep the savings (if any). Any proposal for a linguistic reform in the EU is unlikely to be successful given the current decision-making rules, especially the requirement of unanimous support among member countries. Such a reform would be more likely to enjoy sufficient political support to be successful.

5 The Costs and Benefits of Multilingualism

The monetary costs of multilingualism in the EU were discussed already briefly in sub-section 2.3. The non-monetary costs in terms of delays, misunderstandings and errors in the translation process, also discussed in sub-section 2.3, and in Section 4, are non-negligible either.[35] The sequence of optimal sets outlined in sub-section 3.3 yields further insights on the cost of multilingualism in the EU. In particular, the cost of adding another official language can be computed by dividing the average cost per language (which is assumed to be € 65.1 million, that is the total cost divided by the number of languages, minus one) by the number of people who are prevented from being disenfranchised, following the introduction into the optimal set of a new language. The costs are reported in columns (4) and (8) of Table 4. The per-person costs attributable to the first few languages within the sequence are modest but thereafter they increase progressively, reaching over €100 per person for Estonian and Slovene and exceeding €1,000 for Maltese. Accounting for linguistic proximity increases the cost even further, to hundreds

[34] Think of the costs associated with the two different interpretations of the UN Security Council Resolution 242 concerning the territories captured by Israel during the 1967 war.

[35] For example, the official EU web pages celebrating multiculturalism contain a short three-paragraph text translated into all 20 official languages. The example refers to the "increasingly interdependent world of the 21st century…". The Czech translation of this phrase, however, replaced the word 'interdependent' with 'independent'. See http://europa.eu.int/abc/european_countries/languages/index_en.htm

of euros per person in some cases. The exception to this pattern is Maltese for which the cost remains the same according to both methodologies; this is because Maltese, as a Semitic language, is assumed to be unrelated to all the other languages, and comes in as last language.

5.1 Opportunity costs of multilingualism

We now turn to the issue of opportunity costs that arise from multilingualism. Konrad and Thum (1993) point out that language fulfils two main functions: *medium of exchange* and *store of value*. The former means that we can communicate with other people, cooperate and coordinate our actions when pursuing a shared objective and transmit relevant information to each other. The latter means that knowledge and information can be stored over time and passed on to others or to future generations, in oral or written form. Developing the ability to communicate was an achievement of an enormous importance for mankind. In terms of productivity improvement that it facilitated, it easily ranks ahead of either the adoption of agriculture or the industrial revolution: Diamond (1992) even argues that it was the crucial ingredient that triggered the development of modern humans and gave them a crucial competitive advantage over more primitive humanoids such as the Neanderthals.[36]

All the benefits that language bestows on us, however, depend on our ability to understand each other. Two people who do not share a common language derive little benefit from either the medium of exchange or the store of value functions of their respective languages. Consequently, their ability to communicate, cooperate, engage in joint economic activities and to trade with each other is limited. In fact, so pervasive and persistent is linguistic diversity even in the face of close geographic proximity and/or long periods of trade and migration flows, peaceful cooperation or military domination, that *discrimination* (i.e. excluding strangers from communication) could well be identified as the third main function of languages.[37]

---

[36] Language allowed humans to engage to engage in sophisticated coordination of their hunting or food-gathering activities, to teach each other better how to make and use tools and to transmit knowledge from the elders to the young. This may explain why Neanderthals' tools remained very crude and changed little over millennia whereas tools produced by Homo Sapiens underwent dramatic improvements over time (see Diamond, 1992).

[37] Anyone who ever worked in a multi-lingual work environment will be familiar with the use of foreign languages to exclude others from one's private conversations. Other notable examples include the use of

The discrimination function of language implies that the productivity-enhancing benefits of language only apply to those who share the same language. The dependence on language for communication therefore imposes additional transaction costs whenever a common language is missing. This applies to a wide range of transactions, most notably international trade, migration and investment.

The parallel between languages and currencies with respect to trade is obvious. Both common currency and common language lower transaction costs, increase transparency of bilateral contracts and reduce uncertainty. One of the defining characteristics of the integration process in Europe has been monetary unification. A key argument in favor of the European common currency has been formulated concerning the expected impact of monetary integration on trade flows. Empirical evidence obtained by estimating gravity equations for bilateral trade, such as the various contributions by Andrew Rose with various co-authors (see, for example, Rose, 2000, Frankel and Rose, 2002, Glick and Rose, 2002, Rose and Stanley, 2005) tends to conclude that the introduction of a common currency increases trade flows between countries sharing the common currency approximately two-fold. This estimate is based on observing trade between countries that at present already use the same currency; it is assumed that the same trade-enhancing impact of common currency will be enjoyed also by the countries of the Eurozone. The same kind of analysis tends to find that common language increases bilateral trade by another approximately 50 percent (both results are obtained after controlling for a host factors such as geographical distance, economic size of countries, contiguity, common history and/or colonial legacy and the like). While falling short of the impact of common currency, this is approximately equivalent to the impact of EU membership on trade. Mélitz (2002) extends this analysis by pointing out that trade increases not only for countries sharing the same official language, but also for countries with sufficiently high numbers of common third-language speakers (see also Helliwell, 2000).

---

the Navajo language and Navajo Indians as radio operators in the US Navy to transmit and receive coded messages during the World War II operations in the Pacific (unlike the more traditional codes used by the US Army and Air Corps, the Navajo Code was never broken by the Japanese), and the Cockney Rhyming Slang, supposedly developed by convicts who wanted to be able to exclude the prison guards from their discussions.

Frankel and Rose (2002) argue one percent increase in a country's overall trade increases its long-term per-capita income by at least one third of a percent. The implications are quite powerful: by maintaining a multitude of languages, European countries are foregoing substantial gains in their mutual trade and also in terms of economic development. Adopting a common language, on the other hand, would increase trade and make Europe more prosperous.

Another important issue is the impact of language on migration flows. Europe is notorious for its labor being largely immobile, even in the wake of adverse regional shocks (see Bentivogli and Pagano, 1999, Puhani, 2001, and Fidrmuc, 2004). This has important implications for the ability of regions and countries to deal with asymmetric shocks. Thus, while US regions tend to absorb the adverse effects of asymmetric shocks primarily through out-migration of labor (Blanchard and Katz, 1992), in Europe the bulk of adjustment occurs through unemployed workers eventually dropping out of the labor force (Decressin and Fatas, 1995). While there may be other reasons for these differences in the pattern of regional adjustment to shocks, most notably the generosity of European welfare states, the linguistic diversity in Europe clearly is an important obstacle to international migration flows. As a consequence, workers in depressed regions who could potentially improve their employment and wage prospects by migrating to another region in another country tend to stay put with inferior labor-market outcomes.

Even when workers do move to another country, migrants are often subject to substantial occupational downgrading. For example, the UK received approximately 300 thousand immigrants from the new member countries of the EU in the wake of the latest EU enlargement. Although survey and anecdotal evidence suggests that many migrants are relatively highly skilled and educated (see, for example, Drinkwater, 2003), the vast majority of them tend to take up low-skilled occupations (see Home Office et al., 2005). Although further research on the causes of occupational downgrading of post-enlargement migrants is not yet available, insufficient linguistic skills may be an important reason explaining why migrants accept work which is less than commensurate with their qualifications.

5.2 The private returns to languages

The importance of linguistic skills for migrants' labor-market is confirmed by the literature on private return to a foreign language. An important challenge faced by that literature is that education is endogenous in earnings and therefore OLS regression of earnings on linguistic skills will yield biased results. The solution, used in the literature on private returns to education, is to use instruments for education, such as time of admission (Angrist and Krueger, 1991), or to rely on natural experiments, such as observing educational outcomes and earnings of identical twins (Ashenfelter and Krueger, 1994). The same holds for returns to language acquisition.

Not surprisingly, most papers deal with the language skills of immigrants in 'traditional' immigration countries such as Australia (Chiswick and Miller, 1995), Canada (Abbott and Beach, 1992, Aydemir and Skuterud, 2005, Chiswick and Miller, 1995), Germany (Dustmann and Van Soest, 2002), Israel (Beenstock, Chiswick and Repetto, 2001, Berman, Lang and Siniver, 2003, Chiswick and Miller, 1995, Chiswick, 1998), the United Kingdom (Leslie and Lindley, 2001), and the United States (Bratsberg, Ragan and Nasir, 2002, Chiswick and Miller, 1995, 2002, Hellerstein and Neumark, 2003, Bleakey and Chin, 2004). A few papers consider the case of natives in multilingual societies, such as Canada (Shapiro and Stelcner, 1997), Hungary (Galasi, undated), Luxemburg (Klein, 2003), Switzerland (Cattaneo and Winkelman, 2003) and the U.S. (Fry and Lowell, 2003). EU 15 countries are the subject of papers by Williams (2005) and Ginsburgh and Prieto (2006).

In the remainder of the section, we concentrate on multilingualism in EU countries. Galasi (undated) uses two Hungarian surveys of young career beginners who graduated from public higher education as full-time students in 1998 and 1999. His paper is devoted to the returns to education rather than to languages. Due to this, he merely corrects for the usual bias in estimating the returns to education equations but not for the similar bias with respect to the language dummies. He finds that speaking English or German increases wages by some 6 and 4 percent, respectively, although the results are not very robust (including obtaining negative, though insignificant, returns, in some cases) to using different specifications.

In his paper on the returns to languages in Luxemburg, Klein (2003) concludes that there is little advantage to be proficient in any of the official languages (French, German and Luxemburgish), but it pays to know English, the only truly "foreign" language.

Williams (2005) uses data from the European Community Household Panel Survey (ECHP) between 1994 and 1999, the only years in which information about languages used at work was introduced. The question used is: "Does your work involve use of a language other than (the official language in the country)." Williams runs ordinary least squares Mincer-type regressions where the explanatory variables include a broad array of socio-economic indicators. To capture the effect of language knowledge, a dummy is introduced for each of the following languages if it is used as the first foreign language used at work: English, French, German, Spanish, Italian, Dutch, All other. In an alternative specification, all languages are pooled, and only one coefficient is estimated. Equations are estimated for each EU 15 country separately (with the exception of Sweden) in 1966. The results indicate that in Austria, Finland, Italy, Spain and the Netherlands, English is the only language that yields a significant return. However, substantial returns are also found for French in Denmark, Luxemburg, Greece and Portugal, while German generates positive and significant returns in Belgium, Luxemburg and France, Spanish does so in France, Italian in Luxemburg and Portugal, and Dutch in Belgium. In the U.K., there is no return to using a second language. The returns vary between 5 and 20 percent of earnings, depending on the country.

In a paper in progress, Ginsburgh and Prieto (2006) use the same survey, but concentrate on the years 2000 and 2001 and on males only. There are two additional important differences compared to the study by Williams (2005). First, instead of using a dummy for each language, they pool all languages by creating a unique explanatory variable, which represents the countrywide disenfranchisement rate of the language used at work by each respondent.[38] Second, they estimate a Mincer-type equation with this language variable in 2001, and use as instrument the observed value in 2000. An equation is estimated for several EU 15 countries (Austria, Denmark, Finland, France, Greece,

[38] Using the disenfranchisement rates calculated in Ginsburgh and Weber (2005) for English, French, German, Italian, and Spanish. They also assume that disenfranchisement is equal to 0 for the national language (for example French in France), and to 100 for all other foreign languages.

Ireland, Italy, Spain, and Portugal[39]) using a generalized method of moments (GMM) estimator. The coefficient estimated for the disenfranchisement variable can then be translated for each language in each country. The expected sign for the coefficient should be positive, since the less a foreign language is known, that is, the higher the disenfranchisement rate, the higher is the expected return to the language. The instruments used are valid in all cases and the coefficient on language disenfranchisement is positive and significantly different from zero at the usual five percent probability level for Austria, Finland, France, Greece, Italy, Spain and Portugal. It is significantly different from zero but at the ten percent level only for Denmark, and negative, not significant for Ireland. Thus the return to languages is positive, with lower or no significance in the two countries where English is widely spoken (Denmark) or considered an official language (Ireland). The paper outlines that private returns decrease as disenfranchisement decreases. This raises an interesting question since it shows that there is a tradeoff between public investment in language education and private returns: public education will eventually crowd out private returns.

6 Conclusions

The choices of linguistic policies in multi-lingual countries, multi-lateral federations and multinational bodies have potentially far-reaching political, cultural and social implications. Language and linguistic issues stir up very strong passions in people, in a way that is perhaps paralleled only by religion and football. Linguistic disputes often lie at the root of social and inter-ethnic conflicts within countries as demonstrated by the examples of Belgium, Spain, and, more recently, Ukraine, or even disintegration of countries (e.g. the former Soviet Union, Yugoslavia and Czechoslovakia). Linguistic policies, however, can also help cement societies and foster their integration: France, the Italy, US or the UK, for instance, were much more linguistically heterogenous in the past than at present. Finally, recognizing a language at the regional level (e.g. Catalan in Spain or Welsh in the UK), national level (e.g. Irish in Ireland or Swedish in Finland) or

[39] No survey results are available for Sweden and the U.K in both years and for the Netherlands in 2001. For Germany, the questions changed between 2000 and 2001. Finally, Ginsburgh and Prieto drop Belgium and Luxemburg because several languages enjoy official status in these countries, which makes the results more difficult to interpret.

at the EU level can significantly strengthen the relative standing of the language. In Ireland, the recognition of Irish as an official language with equal status to English is largely credited with helping save the language from extinction. In Malta, the use of Maltese at the EU level apparently has lead to its greater use also by national authorities (which, traditionally, relied predominantly on English).

Linguistic policies have also important economic ramifications. While we recognize the gravity of political, cultural and social aspects of linguistic standardization, our analysis focuses on economic and political-economy considerations. While many of our arguments apply to any multi-lingual entity, linguistic policies in the EU are at the center of our focus for a number of reasons. Multilingual societies typically impose limits on the use of minority and regional languages. The EU, in contrast, takes pride in its commitment to multilingualism. Linguistic standardization is typically chosen and implemented in a top-down fashion whereby the political elite (which may be dominated by one of the linguistic groups) chooses and enforces the choice of the official language or languages. The EU chose a bottom-down approach whereby individual countries are allowed to nominate their national or state language as the official language of the EU, even when it serves little practical purpose (as is apparently the case with Irish and Maltese, both of which were nominated by countries that also have English as their official language). Last but not least, with each enlargement, the linguistic issues in the EU become more and more complex. Multilingualism may have been a perfectly rational and easy to implement solution in a community of six countries with very limited responsibilities. Yet, in a Union of 25 or more countries with 20+ languages, multilingualism increasingly takes on Babel-like appearance.

In virtually every multilingual society, some languages enjoy a more prominent status than others. This may be due to historical legacies: for example, the language of the elite is often imposed on the rest of the society, even if the elite (initially) accounts only for a small fraction of the society as a whole (e.g. English in Ireland). In other instances, the language of the largest linguistic group takes precedence (e.g. French in France, Castilian in Spain and Mandarin in China). The other languages may persist but their use is restricted, either formally or informally. In the EU, the same pattern has emerged too. Many more Europeans speak English than any other language. This is largely because they chose to learn English: for each native English speaker in the EU,

there are almost four non-native speakers. English plays a leading role not only in the EU but also globally: it is increasingly the language of choice of the internet, commerce, science and education.

If the EU were to select a *lingua franca*, it would clearly have to be English. A reform imposing a single official language, however, would be neither acceptable nor optimal in a multilingual union such as the EU. Every multilingual entity accepts, explicitly or implicitly, a certain level of disenfranchisement that is deemed tolerable. Choosing English as the sole official language of the EU, however, would linguistically disenfranchise nearly one EU citizen in two, which clearly would not be acceptable. Even embracing the three *procedural* languages, English, French and German, would be far from optimal, as the resulting degree of disenfranchisement would still be excessive.

Having said that, the status quo of 20 official languages (and still counting) is hardly optimal either. It is neither efficient nor particularly fair. It implicitly aims at completely preventing disenfranchisement, which makes it stand out among linguistic regimes adopted by other multilingual societies. It is also extraordinarily costly, not only because of the direct financial costs of accommodating 20 official languages but also in view of the errors and delays that it causes. Furthermore, although the EU approach to multilingualism is motivated by fairness, it effectively extends much greater privileges to the Maltese than to the Catalan, despite the latter being much more numerous.

An optimal linguistic reform, instead, would be one that occupies the middle ground. Selecting six official languages (English, French, German, Polish, Italian and Spanish) would leave 9 percent of the EU citizens disenfranchised. Accounting for linguistic distance brings this figure further down to 6 percent. This may still appear too high. However, we don't live in a static world. Young Europeans already possess much better linguistic skills than their older compatriots, especially with respect to speaking English. Therefore, even without any linguistic reform in the EU, disenfranchisement is bound to fall over time, and reducing the number of official languages would only help accelerate this process.

Compared to this six-language scenario, adding further official languages has only limited benefits in terms of reducing disenfranchisement, despite increasing the overall costs considerably. The direct financial costs thus differ substantially from language to language. The non-monetary costs due to delays and errors are non-negligible either. The

costs of multilingualism, however, extend far beyond the direct monetary costs of providing translation and interpreting services. Linguistically diverse Europe fails to realize the potential gains from trade and migration, which could be substantial. The parallel with the envisaged gains from monetary unification seems particularly fitting in this context; it is ironical that while Europe is moving ever closer towards economic, monetary and even political unity, it continues to encourage linguistic disunity. In the absence of common linguistic standards, those individuals who acquire foreign-language skills tend to enjoy a wage premium that is often substantial. The available evidence seems to suggest that the return to linguistic skills is particularly high for the English language, and in countries where few speak it.

Unlike in previous analyses[40] we introduce the notion of language externalities. In particular, we show that in the presence of externalities, decentralization may lead to lower overall welfare. Indeed individual countries (linguistic groups) determine their optimal extent of translation without taking account of the negative effects of miscommunication and delays in translations and interpretation imposed on the rest of the Union. Therefore, while maximization of the overall (Union-wide) welfare may require a reduction in the number of official languages and/or in the extent of translation, centralization leads to a higher welfare in the presence of externalities.

References

Abbott, M. and Ch. Beach (1992), Immigrant earnings differentials in Canada: A more general specification of age and experience effects, *Empirical Economics* 17, 221-238.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and R. Wacziarg (2003), Fractionalization, *Journal of Economic Growth* 8, 155-194.

Ammon, U. (2003), The international standing of the German language, in J. Maurais, ed., *Languages in a Globalising World*, Cambridge: Cambridge University Press.

Angrist, J. and A. Krueger (1991), "Does compulsary school attendance affect schooling and earnings?" *Quarterly Journal of Economics* 106, 979-1014.

Ashenfelter, O. and A. Krueger (1994), "Estimates of the Economic Return to Schooling From A New Sample of Twins," *American Economic Review* 84, 1157-1173.

[40] E.g. by Ginsburgh, Ortuno_Ortin and Weber (2005) and Fidrmuc and Ginsburgh (2005).

Aydemir, A. and M. Skuterud (2005), Explaining the deterioration entry earnings of Canada's immigrant cohorts, 1996-2000, *Canadian Journal of Economics* 38, 641-672.

Banzhaf, J. F. (1965), "Weighted voting does not work," *Rutgers Law Review* 19, 317-343.

Beenstock, M., B. Chiswick and G. Repetto (2001), The effect of language distance and country of origin on immigrant language skills: Application to Israel, *International Migration* 39, 33-62.

Bentivogli, C. and P. Pagano (1999), "Regional Disparities and Labor Mobility: The Euro-11 versus the USA," *Labour* 13 (3), 737-760.

Berman, E., K. Lang and E. Siniver (2003), Language skill complementarity: Returns to immigrant language acquisition, *Labour Economics* 10, 265-290.

Blanchard, O., and L. F. Katz (1992), "Regional Evolutions," *Brookings Papers on Economic Activity* 1, 1-61.

Bleakey, H. and A. Chin (2004), Language skills and earnings: Evidence from childhood immigrants, *The Review of Economics and Statistics* 86, 481-496.

Bratsberg, B., J. Ragan and Z. Nasir (2002), The effect of naturalization on wage growth: A panel study of young male immigrants, *Journal of Labor Economics* 20, 568-597.

Bretton, H. (1976), Political science, language, and politics, in W. M. O'Barr and J. F. O'Barr, eds*., Language and Politics*, The Hague: Mouton.

Calvet, J.-L. (2002), *Le marché aux langues*, Paris: Plon.

Candolle, A. de (1987, [1873]), *Histoire des Sciences et des Savants depuis deux siècles*, Paris: Odile Jacob.

Cattaneo, A. and R. Winkelmann (2003), Earning differentials between German and French speakers in Switzerland, Working Paper 0309, University of Zürich.

Chiswick, B. (1998), Hebrew language usage: Determinants and effects on earnings among immigrants in Israel, *Journal of Population Economics* 15, 253-271.

Chiswick, B. and P. Miller (1995), The endogeneity between language and earnings: international analyses, *Journal of Labor Economics* 11, 246-288.

Chiswick, B. and P. Miller (2002), Immigrant learning: Language skills, linguistic concentrations and the business cycle, *Journal of Population Economics* 15, 31-57.

Crawford, J. (1997) The official language question, *Issues in U.S. Language Policy*.

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

Decressin, J. W., and A. Fatas (1995), "Regional labor market dynamics in Europe." *European Economic Review* 39 (9), 1627-1655.

De Swaan, A. (2001), *Words of the World*, Cambridge: Polity Press.

Desmet, K., I. Ortuño-Ortin and S. Weber (2005), Peripheral linguistic diversity and redistribution, CEPR Discussion Paper.

Diamond, J. (1992), *The Third Chimpanzee: The Evolution and Future of the Human Animal*, Harper Perenial.

DG Press and Communication (2003), Applicant Countries Eurobarometer 2001: Public Opinion in the Countries Applying for European Union Membership, European Commission, March 2002.

DG Translation (2005), *Translating for a Multilingual Community*, European Commission, April 2005.

Drinkwater, S. (2003), *Go West? Assessing the Willingness to Move from Central and Eastern European Countries*, FLOWENLA Discussion Paper No. 5, Hamburg Institute of International Economics, Hamburg.

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

Dustmann, C. and A. Van Soest (2002), Language and the earnings of immigrants, *Industrial and Labor Relations Review* 55, 473-492.

Easterly, W. and R. Levine (1997), Africa's Growth Tragedy: Policies and Ethnic Divisions, *The Quarterly Journal of Economics* 112, 1203 – 1250.

Fearon, J.D. (2003), Ethnic and cultural diversity by country, *Journal of Economic Growth* 8, 195-222.

Fidrmuc, J. (2004), "Migration and Regional Adjustment to Asymmetric Shocks in Transition Economies," *Journal of Comparative Economics* 32, 230-247.

Fidrmuc, J. and V. Ginsburgh (2005), Languages in the European Union: The quest for equality and its cost, CEPR Discussion Paper 4795.

Fodor, F. and S. Peluau (2003), Language geostrategy in Eastern and Central Europe: Assessments and Perspectives, in Jacques Maurais, ed., *Languages in a Globalising World*, Cambridge: Cambridge University Press.

Frankel, J. and A. K.Rose (2002), "An Estimate of the of Common Currencies on Trade and Income," *Quarterly Journal of Economics* 117 (2), 437-466.

French Ministry of Education (2004), Evaluation des compétences en anglais des élèves de 15 à 16 ans dans sept pays européens, http://www.education.gouv.fr/stateval (March 2004)

Fry, R. and B. L. Lowell (2003), The value of bilingualism in the U.S. labor market, *Industrial and Labor Relations Review* 57, 128-140.

Galasi, P. (no date), Estimating wage equations for Hungarian higher education graduates, manuscript.

Ginsburgh, V. and S. Weber (2005), Language disenfranchisement in the European Union, *Journal of Common Market Studies* 43, 273-286.

Ginsburgh, V., I. Ortuño-Ortin and S. Weber (2005), Disenfranchisement in linguistically diverse societies. The case of the European Union, *Journal of the European Economic Association* 3, 946-965.

Ginsburgh, V. and J. Prieto (2006), The returns to languages in EU 15 countries, draft.

Glick, R. and A. K.Rose (2002), "Does a Currency Union Affect Trade? The Time-series Evidence," *European Economic Review* 46, 1125-1151.

Greenberg, J.H. (1956), The measurment of language diversity, *Language* 32, 109-115.

Hagège, C. (2000), *Le souffle de la langue*, Paris: Odile Jacob.

Hellerstein, J. and D. Neumark (2003), Ethnicity, language, and workplace segregation: Evidence from a new matched employer-employee data set, *Annales d'Economie et de Statistique* 71/72, 19-78.

Helliwell, J. F. (2000), *Language and Trade*, The Department of Canadian Heritage, Official Languages Support Program Branch, published on-line at http://www.canadianheritage.gc.ca/progs/lo-ol/perspectives/english/explorer/page_01.html.

Home Office, Department for Work and Pensions, HM Revenue and Customs, and Office of the Deputy Prime Minister (2005), *Accession Monitoring Report: May 2004 – September 2005*, available on-line at http://www.ind.homeoffice.gov.uk/ind/en/home/about_us/reports/accession_monitoring.html.

Huntington, S. P. (2004), *Who Are We? The Challenges to America's National Identity*. New York: Simon & Schuster.

INRA (2001), Eurobaromtre 54 Special, Les Européens et les langues, Février.

Kaiser, S. (2003), Language and script in Japan and other East Asian countries: Between insularity and technology, in J. Maurais, ed., *Languages in a Globalising World*, Cambridge: Cambridge University Press.

Klein, C. (2003), La valorisation des compétences linguistiques sur le marché du travail luxembourgeois, CEPS/INSTEAD Paper, Luxemburg.

Konrad, K. A., and M. Thum (1993), "Fundamental Standards and Time Consistency," *Kyklos* 46 (4), 545-568.

Laponce, J. (1992), Language and politics, in Mary Hawkesworth and Maurice Hogan, eds., *Encyclopedia of Government and Politics*, London: Routledge.

Laponce, J. (2003), Babel and the market: Geostrategy for minority languages, in J. Maurais, ed., *Languages in a Globalising World*, Cambridge: Cambridge University Press.

Leslie, D. and J. Lindley (2001), The impact of language ability on employment and earnings of Britain's ethnic communities, *Economica* 68, 587-606.

Longman, C. (2004), The convention as communicative environment: The challenge of multilingual deliberation, mimeo.

Maurais, J. (2003), Towards a new linguistic world order, in Jacques Maurais, ed., *Languages in a Globalising World*, Cambridge: Cambridge University Press.

Mélitz, J. (2002), *Language and Foreign Trade*, CEPR Discussion Paper No. 3590, Centre for Economic Policy Research, London.

Ministre de l'Education Nationale (2004), Evaluation des compétences en anglais des éves de 15-16 ans dans sept pays européens, http://www.education.gouv.fr/stateval (Mars 2004).

Olson, M. (1965), *The Logic of Collective Action*, Harvard University Press, Cambridge MA.

Ortega, J., and T. P. Tangeras (2004), Unilingual versus Bilingual Education System: A Political Economy Analysis, Université de Toulouse and IUI, Stockholm, mimeo.

Pool, J. (1991), The official language problem, *American Political Science Review* 85, 495-514.

Puhani, P. (2001), "Labor Mobility – An Adjustment Mechanism for Euroland?" *German Economic Review* 2 (2), 127-140.

Rose, A. K. (2000), "One Market, One Money: Estimating the Effect of Common Currencies on Trade," *Economic Policy* 30, 7-46.

Rose, A. K. and T.D. Stanley (2000), "A Meta-Analysis of the Effect of Common Currencies on International Trade," *Journal of Economic Surveys* 19 (3), 347-366.

Shapiro, D.M. and M. Stelcner (1997), Language earnings in Quebec: Trends over twenty years, 1970-1990, *Canadian Public Policy* 23.

Shapley, L.S. (1953), A value of n-person game, *Annals of Mathematics Study* 28, 307-318.

Shapley, L.S. and M. Shubik (1954), A method for evaluating the distribution of power in a committee system, *American Political Science Review* 48, 787-792.

Sridhar, K. (1989), *English in Indian Bilingualism*, New Delhi: Manohar.

Tonkin, H. (2003), The search for a global liguistic strategy, in Jacques Maurais, ed., *Languages in a Globalising World*, Cambridge: Cambridge University Press.

Truchot, C. (2003), Languages and supranationality in Europe: The linguistic influence of the European Union, in J. Maurais, ed., *Languages in a Globalising World*, Cambridge: Cambridge University Press.

Van Parijs, P. (2003), Europe's three language problems, in R. Bellamy, D. Castiglione and C. Longman, eds., *Multilingualism in Law and Politics*, Oxford: Hart, forthcoming.

Weber, R. J. (1988) , Probabilistic values for games, in A. E. Roth, ed., *The Shapley Value*, Cambridge University Press, 129-136.

Williams, D. (2005), The economic returns to multiple language usage in Europe, CEPS/INSTEAD Paper, Luxemburg.

Table 1 Languages in Europe and in the World, Selected Indicators

	Web pages		Indexation of articles		Internat. Organiz.	Primary texts EU				Students Centr Eur.	Students Berlitz	
	1997	2000	1978	1998		1986	1992	1997	2004	(1994-95)	1970	1995
	%	%	%	%	(units)	%	%	%	%	millions	%	%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	
Selected EU languages												
English	84	68.4	62.3	82.5	181	26	35.1	45.4	62	1.27	42	65
German	4.5	5.8	5	1.6	58	11	6.2	5.4	3.1	0.8	12	8
French	1.8	3	2.4	0.5	165	58	46.9	40.4	26	0.69	25	8
Italian	1	1.6	...	...	15	...	...	...	...	...	...	...
Spanish	2.4	...	...	...	77	...	...	...	...	...	12	9
Other important languages												
Arabic	...	...	...	...	30	-	-	-	-	...	...	...
Chinese	3.9	0.3	5.9	...	...	-	-	-	-	...	...	...
Japanese	3.1	5.9	4.7	4.5	...	-	-	-	-	...	...	...
Russian	2.4	19.5	19.5	3.1	35	-	-	-	-	...	...	...
Other	5.2	6.9	1.9	5	-	5	8.8	8.7	8.8	...	9	10
Total	100	100	100	100		100	100	100	100		100	100

Sources: Web pages: Maurais (2003, pp. 20-22); Indexation of articles: Laponce (2003, p. 60); International Organizations: Ammon (2003, p. 241); Primary texts EU: Truchot (2003, p. 104) for 1986 and DG Translations (2005) for the other years; Students in Central Europe (Bulgaria, Hungary, Poland, Czech Republic, Romania, Slovakia): Fodor and Peluau (2003, p. 96); Students in Berlitz Schools, worldwide: Ammon (2003, p. 246).

...: not availabe; -: does not apply to the case.

Table 2 Number of Speakers and Power of EU Main Languages

	Native population	Population knowing the language		Shapley Value		
	EU25	EU15	EU25	EU 15	AC10	EU 25
English	62.3	208.6	224.3	100	100	100
French	64.5	127.8	130.0	59	10	56
German	90.1	118.3	132.6	60	90	62
Italian	57.6	65.2	65.2	35	10	33
Spanish	39.4	56.3	56.3	29	0	27
Polish	38.6		40.8	0	330	25

Notes: English is considered the native language in the UK and Ireland, French is native in France and for 40 percent of Belgians, German is native in Germany and Austria, and Dutch is native in the Netherlands and 60 percent of Belgians. Spanish, Italian and Polish are native in Spain, Italy and Poland respectively. Shapley values are normalized so the English = 100.

Table 3 Disenfrachisement Rates for Main Languages by Country
(in % of country population)

	ENG	FR	GER	IT	ES	PL	EFG	All 6
Austria	50	88	1	91	98	na	0	0
Belgium	59	21	81	93	95	na	16	16
Cyprus	39	93	98	98	99	100	38	38
Czech Republic	76	97	70	98	98	93	54	49
Denmark	22	86	56	100	96	na	16	15
Estonia	69	99	86	99	100	100	62	61
Finland	33	92	74	99	97	na	32	32
France	54	0	2	90	82	na	0	0
Germany	48	84	2	97	96	na	1	1
Greece	55	91	92	95	99	na	51	50
Hungary	85	98	86	99	100	100	75	74
Italy	58	70	94	1	94	na	45	0
Ireland	5	72	92	99	96	na	4	3
Latvia	77	99	86	100	100	96	67	64
Lithuania	80	98	88	100	100	77	70	54
Luxemburg	29	3	16	75	90	na	0	0
Malta	17	91	99	55	99	100	12	12
Netherlands	18	62	30	95	91	na	13	13
Poland	79	97	83	98	99	0	66	0
Portugal	62	69	95	98	90	na	55	54
Slovakia	87	98	79	100	99	93	71	66
Slovenia	48	96	57	85	98	100	27	23
Spain	62	81	97	97	1	na	55	0
Sweden	16	86	64	99	92	na	15	15
United Kingdom	2	76	89	98	93	na	1	1
Bulgaria	87	95	94	99	99	100	80	79
Romania	80	83	95	96	99	100	68	67
Turkey	84	98	95	100	100	100	81	81
EU15	42	64	66	82	83	na	18	5
AC10	78	97	81	98	99	47	64	28
EU25	48	69	69	85	84	na	26	9
Candidates	83	94	95	99	99	100	78	78

Sources: Eurobarometer 54: Special Survey on Languages (December 2000) and Candidate Countries' Eurobarometer 2001.1 (October 2001) for the EU14 and AC10 and candidate countries, respectively. The figures aggregate native speakers of each language with those who report to know or have learned the language.

Table 4 Sequence of Optimal Sets of Languages

Disenfranchisement only				Accounting for linguistic distance			
Sequence (1)	EU 15 (2)	EU 25 (3)	Cost (4)	Sequence (5)	EU 15 (6)	EU 25 (7)	Cost (8)
No language	100	100		No language	100	100	
English	42.3	48.2	0.27	English	36.1	42.4	0.25
German	30.3	36.0	1.16	French	16.7	25.9	0.86
French	18.3	25.7	1.38	German	9.1	17.5	1.69
Italian	11.5	19.9	2.42	Polish	9.1	11.0	2.18
Polish	11.5	14.2	2.49	Spanish	5.7	8.2	5.02
Spanish	5.3	9.0	2.75	Italian	3.3	6.1	6.92
Hungarian	5.3	7.2	7.79	Hungarian	3.2	4.3	7.66
Czech	5.3	5.9	10.36	Greek	1.9	3.1	12.38
Greek	3.9	4.6	11.15	Czech	1.9	2.5	22.46
Portuguese	2.4	3.4	11.46	Lithuanian	1.9	2.1	32.05
Dutch	1.4	2.5	17.06	Dutch	1.4	1.7	39.05
Lithuanian	1.4	2.1	34.14	Finnish	1.1	1.4	42.61
Finnish	1.0	1.7	38.77	Portuguese	0.7	1.1	52.31
Slovak	1.0	1.4	48.68	Swedish	0.5	0.9	72.04
Swedish	0.7	1.2	48.54	Latvian	0.5	0.7	83.05
Latvian	0.4	0.7	53.76	Estonian	0.5	0.6	155.23
Danish	0.4	0.7	69.18	Danish	0.4	0.5	186.58
Estonian	0.4	0.6	105.34	Slovene	0.5	0.6	261.90
Slovene	0.4	0.5	145.67	Slovak	0.4	0.5	304.05
Maltese	0.4	0.4	1068.15	Maltese	0.4	0.4	1068.15

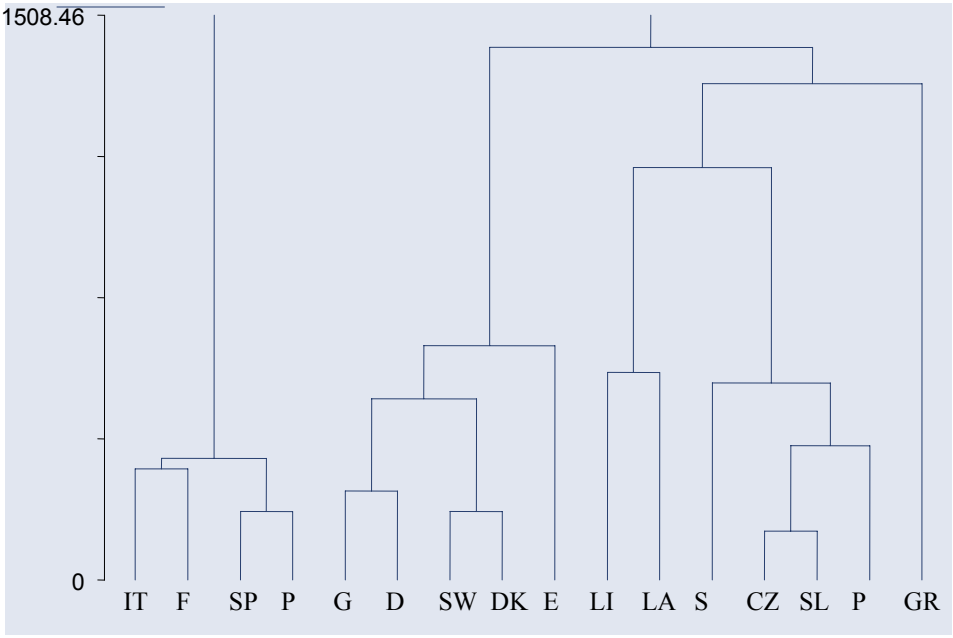
Table 5 Country Disenfranchisement Rates for English, 2000-2045
(in % of population)

	2000	2015	2030	2045
Austria	50	37	30	24
Belgium	59	48	45	42
Cyprus	39	26	16	13
Czech Republic	76	66	56	49
Denmark	22	15	13	12
Estonia	69	56	47	41
Finland	33	21	11	9
France	54	41	34	30
Germany	48	33	24	17
Greece	55	38	28	22
Hungary	85	77	70	66
Italy	58	43	33	27
Ireland	5	5	5	5
Latvia	77	67	60	54
Lithuania	80	7	65	59
Luxemburg	29	25	22	20
Malta	17	9	7	8
Netherlands	18	12	9	8
Poland	79	70	60	52
Portugal	62	46	36	30
Slovakia	87	79	72	66
Slovenia	48	30	21	16
Spain	62	45	34	27
Sweden	16	8	4	1
United Kingdom	2	2	2	2
Bulgaria	87	78	70	64
Romania	80	69	59	49
Turkey	84	82	80	79
EU15	42	31	24	20
AC10	78	69	60	53
EU25	48	37	30	25
Candidate countries	83	79	74	71

Table 6 Aggregate Disenfranchisement Rates for Main European Languages, 2000-2045
(in percent of population)

		2000	2015	2030	2045
English	EU15	42	31	24	20
	AC10	78	69	60	53
	EU25	48	37	30	25
	Candidates	83	79	74	71
French	EU15	64	61	59	567
	AC10	97	96	96	95
	EU25	69	67	65	63
	Candidates	94	94	94	93
German	EU15	66	65	65	64
	AC10	81	79	77	74
	EU25	69	68	67	65
	Candidates	95	95	95	95
Italian	EU15	82	81	81	81
	AC10	98	98	97	97
	EU25	84	84	84	84
	Candidates	99	99	98	98
Spanish	EU15	83	81	80	79
	AC10	99	99	99	99
	EU25	85	84	83	82
	Candidates	100	99	99	99
Polish	EU15	na	na	na	na
	AC10	46	46	46	47
	EU25	na	na	na	na
	Candidates	100	100	100	100
English, French and German	EU15	18	13	10	8
	AC10	64	55	45	39
	EU25	26	20	16	13
	Candidates	78	74	71	67
All six languages	EU15	5	4	3	2
	AC10	28	24	20	18
	EU25	9	7	6	5
	Candidates	78	73	70	67

Figure 1
The philogenetic tree of Indo-European languages used in EU 25



Note. IT = Italian, FR = French, SP = Spanish, PR = Portuguese, GE = German, DU = Dutch, SW = Swedish, DK = Danish, EN = English, LI = Lithuanian, LA = Latvian, SO = Slovene, CZ = Czech, SL = Slovak, PL = Polish, GR = Greek.

Appendix 1

Table A1 Official, regional and other minority languages in the EU

Country	Population [millions]	Official language	Regional official languages	Other important languages	Living languages[3]
Austria	8.2	German	Slovene[1], Croatian[2]	Czech, Hungarian, Slovak	8
Belgium	10.4	Dutch, French	German[2]		8
Denmark	5.4	Danish		German	7
Finland	5.2	Finnish, Swedish		Saami	11
France	60.6	French		Basque, Breton, Catalan, Corsican, Dutch (French Flemish), German, Occitan	28
Germany	82.5	German	Sorbian[2]	Danish, Frisian,	26
Greece	11.1	Greek		Albanian, Bulgarian, Macedonian, Turkish,	13
Ireland	4.1	English, Irish			4
Italy	58.5	Italian	German[1], Slovene[2], Sardinian[2], Ladin[2]	Albanian, Catalan, Croatian, Franco-Provençal, Friulian, Greek, Occitan, Slovene	32
Luxembourg	0.5	Letzeburgesh, German, French			3
Netherlands	16.3	Dutch	Frisian[1]		14
Portugal	10.5	Portuguese		Mirandese	6
Sweden	9.0	Swedish		Finnish, Saami	14
Spain	43.0	Spanish (Castilian)	Basque[1], Catalan[1], Galician[1], Occitan (Aranese)[2]	Aragonese, Asturian, Berber, Portuguese	11

Table A1 Official, regional and other minority languages in the EU (continued)

Country	Population	Official language	Regional official languages	Other important languages	Living languages[3]
United Kingdom	60.0	English	Welsh[1],	Scottish Gaelic, Irish, Cornish	11
Cyprus	0.7	Greek	Turkish[1]	English	4
Czech Republic	10.2	Czech		German, Polish, Romani, Slovak	7
Estonia	1.3	Estonian		Russian, Finnish, Polish, Swedish, Ukrainian	1
Hungary	10.1	Hungarian	German[2], Croatian[2], Slovak[2], Romanian[2], Serbian[2], Slovene[2]	Romani, German, Croatian, Ukrainian	11
Latvia	2.3	Latvian		Russian, Polish, Belarussian, Ukrainian, Lithuanian,	4
Lithuania	3.4	Lithuanian	Russian[2], Polish[2]	Belarussian, Ukrainian	3
Malta	0.4	Maltese, English			2
Poland	38.2	Polish		German, Belarusian, Ukrainian, Ruthenian, Lithuanian, Kashubian	10
Slovakia	5.4	Slovak	Hungarian[2]	Romani, German, Ruthenian, Ukrainian, Czech	9
Slovenia	2.0	Slovene	Hungarian[2], Italian[2]	Bosnian, Croatian, Serbian, Romani	3
Bulgaria	7.8	Bulgarian	Turkish[2]	Romani	9
Croatia	4.4	Croatian	Italian[1]		5
Romania	21.7	Romanian		Hungarian, German, Romani, Turkish	14

Table A1 Official, regional and other minority languages in the EU (continued)

Turkey	71.6	Turkish	Kurdish, Arabic, Azerbaijani, Georgian, Armenian, Bulgarian, Serbian, Romani, Albanian,	33
Iceland	0.3	Icelandic		2
Norway	4.6	Norwegian	Danish, Saami	10
Switzerland	7.5	German, French, Italian, Romansch	Swizerduetsch	9

Notes:
[1] Language enjoys official status at the regional or provincial level: Slovene is an official language in the Austrian province of Carinthia. Turkish is nominally an official language in Cyprus in parallel with Greek; in reality, it is used only sporadically within the territory effectively controlled by the Cypriot government. German is an official language in parallel with Italian in the province of Bolzano (South Tyrol). Frisian is used as an official language in parallel with Dutch in the province of Friesland. While Spanish is an official status in the whole of Spain, Basque has an official status in the provinces of Basque Country and Navarre, Catalan in Catalonia, Balearic Islands and Valencia (where it is often referred to as Valencian and is professed to be distinct from Catalan), and Galician in Galicia. Welsh is used in parallel to English in Wales. Italian is used as an official language in parallel to Croatian in the region of Istria.
[2] Language can be used locally in contact with judicial and/or public authorities but not at the national or provincial level.
[3] Number of living languages, as reported by the Ethnologue database, not counting extinct languages, artificial languages without native speakers (such as Esperanto), and deaf sign languages. Only indigenous languages are included in this figure, omitting languages of recent immigrants (e.g. Russian in Estonia).
Primary Source: *The Euromosaic Study*, European Commission, http://europa.eu.int/comm/education/policies/lang/languages/langmin/euromosaic/index_en.html. Where information was missing or unavailable, the following sources were used: European Bureau for Lesser-Used Languages, http://www.eblul.org/; and Gordon, Raymond G., Jr. (ed.), 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International. Online version**:** http://www.ethnologue.com/. The exception is the number of living languages (last column), which is based only on the Ethnologue web database.

Appendix 2. Rules governing the use of languages in EU institutions

Article 1 of Council Regulation (EC) no 920/2005 of June 13, 2005 amending Regulation no 1 of April 15, 1958 determining the language to be used by the EEC specifies that the official and working languages of the institutions of the European Union are the 20 languages discussed in our paper, plus Irish. Article 2 adds that regulations and other documents of general application are drafted, and the *Official Journal of the European Union* is published in the 21 official languages, and all the versions are authentic.

The Constitution does not set any rule regarding the usage of languages, but empowers the Council to "adopt unanimously a regulation laying down the rules governing the languages of the Union's Institutions languages...".The internal use of languages in the institutions is set through secondary legislation, and the decision is thus left to the Council, but has to be reached unanimously.

However, under article 6 of Council Regulation no 1 of 15 April 1958, each institution may stipulate in its rules of procedure "which of the languages are to be used in specific cases." The result is as follows.

The Parliament. Documents should be drafted in all official languages. Speeches delivered in one of the official languages shall be simultaneously interpreted into the other official languages

The Council. "Except as otherwise decided unanimoulsy by the Council on grounds of urgency, the Council shal deliberate and take decisions only on the basis of documents and drafts drawn up in the languages specified in the rules in force governing languages [that is the official languages]." (Article 14). If the document is not available in a certain official language, a delegation may oppose its discussion. In practice, documentation is often only drafted in English, French and, sometimes, German. However, a text may be adopted only if it is available in all languages.

The Commission. The Commission is given a wide degree of freedom in internal linguistic use: "the Commission shall, as necessary, lay down rules to give effect to these Rules of Procedure [and] may adopt supplementary measures relating to the functioning of the Commission and of its departments..."

"Internally, when the European Commission staff hold meetings, no interpretation is provided: Officials are expected to be able to do without. The weekly meeting of the College of Commissioners has interpretation in English, French and German."

The Court of Justice. Cases can be dealt with in any of the 20 languages. Publications shall be issued in all 20 languages, though some judgments have appeared only in the language of the case.

The European Central Bank. Only one "working" language is guaranteed, English. Only when guidelines and instructions have to be officially published will all official languages be used.

The European Ombudsman. Any of the Treaty languages may be used in communications.

There is some imprecision in the use of the words "official," "working" and "procedural" language. Article 1 of Council Regulation (EC) no 920/2005 of June 13, 2005 amending Regulation no 1 of April 15, 1958 uses both "official" and "working" without distinction. The addition of Irish has added some confusion, since from now on, the list of "languages of the Constitution" or "Treaty languages" includes Irish, while the list of "official and working languages" does not. In the literature concerning EU langauges, the terms "official" and "working" are often used as synonyms. We will do the same, and reserve the term "procedural" for the language(s) used in practice in an institution of the EU.

See Bernat Pujadas, The rules governing the languages of the European Union: which languages and to what extent? A practical guide, www.ciemen.org/mercatorbutlettins/60-48.htm, corroborated by other sources.