ECARES

Culture, Languages, and Economics

Victor Ginsburgh
ECARES, Université Libre de Bruxelles, CORE, Université catholique de Louvain and ECORE

Shlomo Weber
Southern Methodist University, Dallas and New Economic School, Moscow

Culture, Languages, and Economics[1]

Victor Ginsburgh
ECARES, Brussels, CORE, Louvain-la-Neuve and ECORE

and

Shlomo Weber
Southern Methodist University, Dallas and
New Economic School, Moscow

> "If you talk to a man in a language
> he understands, that goes to his head.
> If you talk to him in his language,
> that goes to his heart."
> (Attributed to Nelson Mandela)

1   Introduction

In his book *Decolonizing the Mind* (1986), Ngũgĩ wa Thiong'o, a Kenyan writer who is currently Distinguished Professor of English and Comparative Literature as well as Director of the International Center for Writing and Translation at the University of California (Irvine), makes it clear why language is an essential expression of culture. Thus he goes (pp. 3-5):

> "[T]he biggest weapon wielded and actually daily unleashed by imperialism against that collective defiance is the cultural bomb. The effect of a cultural bomb is to annihilate a people's belief

---

in their names, in their languages, in their environment, in their heritage of struggle, in their unity, in their capacities and ultimately in themselves. It makes them see their past as one wasteland of non-achievement and it makes them want to distance themselves from that wasteland. It makes them want to identify with that which is the furthest removed from themselves; for instance, with other peoples' languages rather than their own.

The choice of language and the use to which language is put is central to a people's definition of themselves in relation to the entire universe. Hence language has always been at the heart of two contending social forces in the Africa of the twentieth century.

Berlin in 1884 saw the division of Africa into the different languages of the European powers. African countries, as colonies and even today as neo-colonies, came to be defined and to define themselves in terms of the languages of Europe: English-speaking, French-speaking or Portuguese-speaking African countries. [W]riters also came to be defined and to define themselves in terms of the languages of imposition. Even at their most radical and pro-African position in their sentiments and articulation of problems they still took it as axiomatic that the renaissance of African cultures lay in the language of Europe.

I should know!"

An indeed, he knows. In 1977, he stopped writing plays, novels, and short stories in English and turned to Gikuyu and Kiswahili, two languages used in his native Kenya.[2] *Decolonizing the Mind* published in 1986, and from which the long quote that precedes is taken, was Ngũgĩ's last text written in English.

---

[2]Kiswahili is the *lingua franca* in almost all countries of the African East.

As African, Ngũgĩ writes about the continent where he was born.[3] But similar stances and ideas hold for all other continents, be they North or Latin America, the Australasian world, some parts of Asia, and even Europe, which *volens, nolens* is turning to a language that consists of a one-thousand-word-English."

It would be easy to suggest that Ngũgĩ's behavior is fueled by strong political and so-called "anti-imperialistic" views generated by colonialism. This is not necessarily so, as Shakespeare makes it obvious in *Richard II*. In the play, Thomas Mowbray was exiled by King Richard to Venice (where he died shortly after he was exiled). He does not lament about the loss of land or status but rather about the inability to speak his native language in exile:

> "A heavy sentence, my most sovereign liege,
> ...
> Have I deserved at your Highness' hands.
> The language I have learn'd these forty years,
> My native English, now I must forego;
> And now my tongue's use is to me no more
> ...
> Within my mouth you engaol'd my tongue,
> Doubly portcullis'd with my teeth and lips;
> And dull, unfeeling, barren ignorance
> Is made my gaoler to attend me.'
> ...
> I am too old to fawn upon a nurse,
> Too far in years to be a pupil now.[4]
> What is thy sentence, then but speechless death,

---

[3] In this respect, and since the misfortunes of Sub-saharan Africa will often be discussed in this chapter, we found worth quoting Isak Dinesen (also known as Karen Blixen) who wrote the following premonitory sentences in her short biographical notes *Shadows on the Grass* (1984), first published in 1960:

> "We Nations of Europe, I thought, who do not fear to floodlight our own inmost mechanisms, are here turning the blazing lights of our civilization into dark eyes, essentially different to ours. If for a long enough time we continue in this way to dazzle and blind the Africans, we may in the end bring upon them a longing for darkness, which will drive them into the gorges of their own, unknown mountains and their own, unknown minds."

[4] At the time, Mowbray was forty years old.

Which robs my tongue from breathing native breath?"

Though King Richard's decision was driven neither by anti-imperialism, nor by considerations of linguistic policy, Thomas Mowbray's plight is not an isolated episode. As we shall see, linguistic policies often tend to alienate groups of individuals whose cultural, societal and historical values and sensibilities are perceived to be threatened by what we call linguistic disenfranchisement, as linguistic rights are restricted or even denied.

Why does man have such an intimate relation with his language? What is it that makes the famous Portuguese writer Fernando Pessoa write that "[m]y homeland is my language" or contemporary French linguist Claude Hagège (2000) claim that "[l]anguages are the flags of dominated people," or that "[the] fight for French is a fight of the mind?" Why does political scientist Henry Bretton (1976) suggest that the "fear of being deprived of communicating skills seems to rise political passion to a fever pitch?"

The translation of Nabokov's (1955) first lines of *Lolita*, as well as what Nabokov writes about his own translation of the book into his native Russian are a good illustration of Mandela's words that appear as epigraph to this chapter. Here are *Lolita*'s first lines:

> "Lolita, light of my life, fire of my loins. My sin, my soul. Lo-lee-ta: the tip of the tongue taking a trip of three steps down the palate to tap, at three on the teeth. Lo. Lee. Ta."

and their translation into French, in which the English rhythm and alliterations are missing:[5]

> "Lolita, lumière de ma vie, feu de mes reins. Mon péché, mon âme. Lo-li-ta: le bout de la langue fait trois petit bonds le long du palais pour venir, à trois, cogner contre les dents. Lo. Li. Ta."

The lines are of course perfectly translated and understandable, but are far from conveying Humbert Humbert's love for Lolita and Nabokov's teasing look at Humbert that transpires in the English original version. The Russian translation by Nabokov himself goes as follows:

> "Lolita, svet moej zhizni, ogon moih chresel. Greh moj, dusha moja. Lo-li-ta: konchik jazika sovershaet put v tri shazhka vniz po nebu, chtoby na tretem tolknutsja o zuby. Lo. Li. Ta."

---

[5]Translation by by E.H. Kahane, Gallimard, 1959, approved by Nabokov who was also fluent in French.

It is worth quoting what Nabokov wrote about his own translation of the book:[6]

> "The history of this translation is the story of a disillusionment. Alas, that 'wondrous Russian tongue' that, it seemed to me, was waiting for me somewhere, was flowering like a faithful spring-time behind a tightly locked gate, whose key I had held in safe-keeping for so many years, proved to be nonexistent, and there is nothing behind the gate but charred stumps and a hopeless autumnal distance, and the key in my hand is more like a skeleton key. I console myself, fits of all, with the thought that the fault for the clumsiness of the translation offered here lies not only with the translator's loss of touch with his native speech but also with the *spirit* of the language."[7]

Different languages express different feelings and perceptions of feelings. To quote again from Nabokov's *Postscript* to his novel in Russian:

> "Everything tenderly human, but also everything coarse and crude, juicy and bawdy, comes out no worse in Russian than in English; but the subtle reticence so peculiar to English, the poetry of thought, the instantaneous resonance between the most abstract concepts in Russian become clumsy, prolix and often repulsive in terms of style and rhythm."

The suggestion that language is more than a means of communicating, has been first emphasized in the very beginning of the 19th century by Wilhelm von Humboldt (1988, [1836]). It was further reinforced by anthropologist Franz Boas (1940), linguists Edward Sapir (1949) and Benjamin Whorf (1956), and later came to be known as the Sapir-Whorf hypothesis: language and culture are interdependent, and the structure of the language that one uses (often as native language) influences the way of thinking and behaving. The consequences, summarized by Kramsch (1998, p. 12), imply that despite the possibility of translating from one language to another, "there will always be an incommensurable residue of untranslatable culture associated with the linguistic structures of any given language." Whorf illustrates this by a couple of examples, such as the notion of "time" in Hopi Indian and in English. This notion, he writes, is perceived so differently

---

[6] *Postcript to the Russian Edition of Lolita*, translated by Earl D. Sampson.
[7] Our italics.

that a Hopi and an English physicist may have a hard time understanding each other.

This bold hypothesis (language determines the way we think) was rejected by the scientific community, since it would have led to the "relativity" of scientific discoveries. Nowadays, a weaker hypothesis is thought to hold, namely that "there are cultural differences in the semantic associations evoked by seemingly common concepts. The way a given language encodes experience semantically, makes aspects of that experience not exclusively accessible, but just more salient for the users of that language" (Kramsch, 1998, p. 13). In short, language reflects cultural preoccupations and constrains the way we think, and culture is expressed through the use of the language (Kramsch, 1998, p. 14). Language is thus not only a means of communication, but also a carrier of culture. Interestingly enough, this seems also to hold true to some extent for writing. According to Carrière and Eco (2009), in Western movies, traveling shots usually move from left to right, while in Iranian movies, they often go from right to left, following the way Persian is written. More generally, according to Steiner (1992, pp. xii-xiv),

> "Each human language maps the world differently. There is life-giving compensation in the extreme grammatical complication of those languages (for example, among Australian Aboriginals or in the Kalahari) whose speakers dwell in material and social contexts of deprivation and barrenness. Each tongue – and there are no 'small' or lesser languages – construes a set of possible worlds and geographies of remembrance. It is the past tenses, in their bewildering variousness, which constitute history."

While recognizing the role of cultural differences in studying cultural diversity, one has to tackle two important empirical issues: (i) how does one measure cultural differences, and (ii) how does one identify cultural groups (the so-called group identification issue) and the resulting fractionalization of the society.

The various aspects of cultures are hard to describe and so are the distances between cultures, though several attempts have been and are still made by scientists to measure them.[8] In most cases, language may be used as a proxy for culture and/or ethnicity. There are of course exceptions. As Storti (2001) shows, there exist large cultural differences that may

---

[8] See Shalom Schwartz's Chapter 20 in this volume, and also Hofstede (2001).

lead to sometimes dramatic misunderstandings even between Americans and Britons, though they speak (almost) the same language. The same is true for Flemish speaking Belgians and Dutch people, though again, there are some small linguistic differences. On the other hand, it seems difficult to find two peoples, or countries that share the same culture, but speak different languages. Thus, broadly speaking, there is a one-to-one correspondence between languages and cultures, but linguistic diversity may be easier to apprehend than cultural diversity, since it is reasonably easy to check which native language is spoken, and measure distances between languages (see Section 2.1).

It is more difficult to identify cultural groups with countries that often result from artificial constructions and may host large numbers of linguistic groups, whose geographical spreads do not coincide with official borders.[9] As Laitin (2000, p.143) points out, it is also true that "people have multiple ethnic heritages, and they can call upon different elements of those heritages at different times. Similarly, many people throughout the world have complex linguistic repertoires, and can communicate quite effectively across a range of apparently diverse cultural zones." The importance of using native languages as identifiers of cultural groups is reinforced by Bretton (1976) who suggests that "[l]anguage may be the most explosive issue universally and over time. This mainly because language alone, unlike all other concerns associated with nationalism and ethnocentrism, is so closely tied to the individual self." And in fact, the first and most influential country-by-country identification widely known as ELF (Ethnolinguistic Fractionalization) conducted in the Soviet Union some fifty years ago (see Section 2.2), was based mainly on linguistic and historic origins of various groups, which, in many cases, are based on their native languages.

The mere presence of distinct linguistic groups is not necessarily a "bad thing." A more diversified environment attracts creative individuals, ventures, businesses and capital. The complementarities of workers' skills, can sometimes easily offset the costs of cross-cultural interactions. The Euro-

---

[9]We will avoid the denomination "ethnolinguistic groups" which is very often used in this literature and stick to the definition provided by Hechter (1974), who suggests that ethnicity rather refers "to the sentiments which bind individuals into solidary groups on some cultural basis. Ethnicity therefore alludes to the quality of relations existing between individuals sharing certain cultural behaviors." Language may bind, but not necessarily so. Fearon (2003) and Labar (2010) also warn about the borderline and arbitrary decisions used to define ethnic groups. Again, it is easier to define linguistic groups than ethnic groups.

pean Union (EU), which consists of 27 countries whose peoples speak many languages, has its problems, but it eliminated the threats of, as well as, the actual military conflicts that were hovering in Europe for many centuries. EU countries live in peace, and so do Australia with 207 languages and the United States with 364.

Linguistic as well as ethnic *fractionalization* or *diversity* could, however, also breed institutional wastefulness, bureaucratic inefficiency and corruption, reduce political stability and hamper economic growth, as shown by the sad and painful example of post-colonial Africa.

How can a country (a region, or a group of countries such as the EU) deal with these negative aspects? A natural response often leads a fractionalized society to introduce a certain degree of *standardization* by adopting a smaller number of languages, including one language that would be common to all.

Standardization, however, creates *disenfranchisement* by depriving sometimes large segments of a society of their linguistic rights. In some cases, the disenfranchisement sentiment turns out to be relatively soft, as it was in France after the 1992 change of the Constitution which included an article imposing French as "the language of the Republic." Linguistic discontentment in Sri Lanka, however, resulted in a long and bloody war.

This chapter is organized as follows. Section 2 is devoted to ethnolinguistic diversity and fractionalization, their measurement and their social and economic consequences. In Section 3 we turn to standardization policies which aim at reducing the economic losses due to the excessive diversity. These policies may however lead to the linguistic disenfranchisement of large groups of a society, which could yield negative social and economic consequences, examined in Section 4. We conclude, in Section 5, by a discussion of the delicate balance between the goods and the ills of both diversity and standardization and try to address the crucial question about how to sustain a manageable degree of diversity while implementing various standardization policies that may disenfranchise certain groups.

2   Linguistic Diversity and Fractionalization

The interdependence of language and culture captured by the Sapir-Whorf hypothesis is related to the concept of a *native* language that plays a crucial role in defining the patterns of human reasoning and behavior. One may speak several languages (and many people do) but they appear not to be fully substitutable, and usually, only one is the mother tongue (probably

even in Nabokov's case). The importance of using one's native languages is underlined in Mandela's quotation in epigraph to this chapter. Serious conflicts due to misunderstandings are common even if the same language is used by some whose native tongues, and thus cultures, differ. Storti (2001) illustrates this by collecting cross-cultural dialogs and analyzing how people from different cultures (French, Germans, Americans and British) attach different interpretations to the same words and sentences, according to their native cultures.

Lolita's translation from English to Russian by Nabokov himself did not make him happy, but had otherwise no big consequences. These may, on the contrary, be extremely important as is shown by Lewis (2004) in *From Babel to Dragomans: Interpreting the Middle East*, where he describes the famous example of Article 1(i) of UN's resolution 242 of November 22, 1967, drafted several months after the war between Israel and its Arab neighbors (Egypt, Syria, and Jordan) ended. The resolution, initially drafted in English, requires the "withdrawal of Israeli Armed Forces from territories occupied during the recent conflict." Lewis (2004, p. 194) suggests that

> "it has been argued, and generally, conceded, that the omission of the article 'the' before 'territories' in the English text means that not all the territories are intended. The Russian text, which has no article since none exists in Russian, leaves, like the English text, the question open. In the Arabic version – unofficial since at that time Arabic was not yet an official UN language – the article is present as a stylistic if not a grammatical necessity."

In French the text is concerned with a "retrait des forces armes israéliennes *des* territoires occupés." In this sentence, the article *des* comes from the contraction of two words: the preposition *de* meaning *from* and the article *les* meaning *the*, and could be interpreted as "all the territories." It is impossible to deny the historical importance of that linguistic discrepancy.

Bellos (2011, pp. 24-236) has a beautiful story that did not cost as much blood as the previous one. He starts by arguing that law "is written in a language of its own that is almost impossible to understand, and what can't be understood can't be translated." He then describes the problem created by the French expression *droits de l'homme*, in English *human rights* and in German *Menschenrechte*. The only correct form is the last one, since in German, there exists *Mann* for *man*, *Frau* for *woman*, but there also exists *Mensch* which has no gender-flavor. The French word *homme* means *man*

and at some point, feminists insisted to also benefit from *droits de la femme*. Since the word *humain* does not have the exact same meaning as *human*, the French decided to re-christen the old expression (which goes back to the French revolution of 1789) and call it *droits de l'Homme*, with a capital *H*, pretending that the capital makes it lose its gender and means *Mensch*.

These arguments have a somewhat surprising practical impact on he definition of the concept of diversity and its measurement. Indeed, it would be natural to partition a society into groups of speakers who share the same native language, since in most cases, it is easier to identify individuals on the basis of their sole native language than of other characteristics. Then however comes the question about how different languages can be, to which we turn now.

2.1 Distances Between Languages

The last 1,250 pages thick 2009 edition of *Ethnologue* is a mine of information on the 6,909 languages that are currently spoken in the world, on where they are spoken, as well as on the number of speakers in each country. Whether 6,909 is large or small is open for discussion, and so is the number itself, since it results from a rather subjective count, and does not include dialects.

When does one consider two languages as being different? Are Venetian and Italian the same language? Are Serbian and Croatian? Flemish and Dutch? English in the UK and in the US? This is a crucial issue for "group identification" of speakers of a given language.

Even if languages are distinct, should we consider speakers of Serbian and Croatian as members of different groups in the same way as speakers of Greek and Turkish? After all, these two pairs are drastically different: Serbian and Croatian are very similar to each other whereas Greek and Turkish belong to entirely different linguistic families – Greek is a Indo-European language while Turkish belongs to the Altaic group.

To address these questions, one has to recognize some degree of distinctiveness between languages. Is it vocabulary that makes them different, or pronunciation, phonetics, phonology, syntax, grammar, among others, and this is even without going into the fundamental issue of whether languages have a common structure. In what follows, we discuss alternative ways of measuring distances between languages.

Language Trees

The need for classifying languages started in the early 19th century, with the birth of linguistics as a scientific discipline, though naïve "linguistic trees" had been drawn before that. These work like genealogical trees, which go from a common shared ancestor (the root of a family) to successive generations. Table 1 illustrates a tree with a common origin, the assumed "root" or "proto-language" of all our languages.[10] This root gave birth to several important level 1 families including Eurasiatic. In Table 1, we only illustrate the various subfamilies (level 2) of Eurasiatic (from Eskimo-Aleut to Indo-European), and then go into level 3 details that have as root Indo-European (from Germanic to Indo-Iranian).[11] In Table 2, we further go into the descendants of Germanic, with 6 additional levels: Today's English, for instance, is at level 6 together with Scots and Yinglish, and descends from West Germanic (level 4) and English (level 5), while Standard German appears at level 9. These trees are constructed using a host of different characteristics of each language, as well as possible migrations.

Fearon and Laitin (1999), Laitin (2000), and Fearon (2003) suggested calculating distances between the branches of linguistic trees as a proxy for distances between linguistic groups.[12] In the original Fearon and Laitin (1999) index (LANGFAM), for every pair of languages, the score takes the level of the first branch at which the languages break off from each other. The higher the number, the higher the similarity of languages.

To give an idea of how distances are calculated, consider German and Hungarian that come from structurally unrelated linguistic families: German is an Indo-European language, while Hungarian belongs to the Uralic family. Therefore, the two languages share no common branches and break off on the first branch: their score is 1. German and Italian share one common level since they are both Indo-European, but separate immediately after that into Germanic and Italic groups, making their score equal to 2. German and Danish share two classifications: they are both Indo-European

---

[10]Not all linguists agree that all languages have a unique root (monogenesis). Some think that languages appeared more or less simultaneously in different parts of the world, and can thus not be related to a common ancestor (polygenesis). See for example Ruhlen (1994).

[11]The numbers between brackets represent the number of languages in the (sub)family. For example, there are 449 Ind-European languages, of which 53 are Germanic, 48 Italic, etc. both Tables 1 and 2 are constructed on data provided by *Ethnologue* (2009).

[12]Fearon (2003) produces an impressive dataset for 822 ethnic groups in 160 countries. A variant of Fearon's formalization is used by Desmet et al. (2009).

and Germanic, and break off from the third branch, as German belongs to the Western branch of the Germanic group, while Danish to the Northern branch. Thus, their score is 3. German and English share three common levels: in addition to being Indo-European and Germanic, both belong to the Western branch of the Germanic group, and their score is 4. Similarly, German and Yiddish pick up a score of 5 and Afrikaans and Dutch, a score of 6. Distances are derived by properly normalizing the scores.

Insert Tables 1 and 2

Lexicostatistical Distances

Lexicostatistical distances are based on similarities and supposed common roots of words in the vocabularies of various languages. Following Ruhlen (1994, p. 12), languages can be related or similar, and these similarities can be explained by three mechanisms only: (a) there may be words that look common for accidental reasons; (b) languages may borrow words from other languages, as English, for example, which contains many French words; and finally, (c) two languages may descend from a common, older language. This is the case for French, Italian, Spanish and Portuguese, which belong to the same branch, and have Latin as ancestor. Lexicostatistical distances are built on so-called *cognate* words, occurring in languages with a historical chain linking them via an earlier language, thus ignoring not only borrowings and accidental similarities, but also syntax and grammar.

Since it would be a daunting task to compare long lists of words for each couple of languages, linguists are forced to rely on a small selection of carefully chosen words, a so-called "list of meanings." Morris Swadesh (1952) introduced some rigor in the choice of meanings to be basic enough to exist in all languages and cultures (such as animal, bad, bite, black, child, die, eat, eye, hunt, digits from one to five), on which deductions can be based. The list we are interested in consists of 200 basic meanings and is still in use nowadays.

Greenberg (1956) was the first to explicitly introduce the notion of distances between languages by using: "an arbitrary but fixed basic vocabulary," to compute "the proportion of resemblances between each pair of languages to the total list" and then use this proportion as a distance between each pair of languages. Dyen, Kruskal and Black (1992) followed on this idea using Swadesh's basic list of 200 words to classify 84 Indo-European speech varieties. They describe the lexicostatistical method as consisting of

12

three[13] steps:

(a) Collecting for each meaning the words used in each speech variety under consideration;
(b) Making cognate decisions on each word in the list for each pair of speech varieties, that is, deciding whether they have a common ancestral word, or not, or whether no clear-cut decision can be made;[14]
(c) Calculating the lexicostatistical percentages, i.e., the percentages of cognates shared by each pair of lists; these percentages lie between one (all words are cognate) and zero (no cognate).

The numbers given in Table 3 are equal to one minus the percentage of cognates. They concern the distances between 25 European languages[15] and the six European languages with the largest number of speakers in the European Union: two are Germanic (English and German), three are Romance (French, Italian and Spanish) and one is Slavic (Polish). It is easy to check that Danish, Dutch, English, German, Icelandic, Norwegian and Swedish are related. So are the Romance languages Catalan, French, Italian, Portuguese, Romanian and Spanish, the Slavic, Bulgarian, Czech, Russian, Serbo-Croatian, Slovak, Slovene, Ukrainian, and the Baltic, Latvian and Lithuanian. Albanian and Greek are distant from any language belonging to the three previous families.

Insert Table 3

Distances based on trees are coarser than lexicostatistical distances, but are easy to construct on the basis of existing linguistic trees, and can be established for all language families.

More About Linguistic Distances

In discussing various ways of measuring distances between languages, one can mention the Chiswick and Miller (2007, Chapter 20) method based on the difficulty speakers of one language face in acquiring another (non-native) language. Such a measure was established by Hart-Gonzalez and Lindemann (1993) using a sample of native Americans who were taught a variety of

---

[13]Actually, four steps, the last one consisting in partitioning languages into a tree.

[14]See Warnow (1997) for further technical details.

[15]Basque, Estonian, Finnish, Hungarian and Turkish (spoken in Cyprus) are excluded, since they do not belong to the Indo-European family.

languages, and whose proficiency was measured at different moments of time. The scores are assumed positively correlated with distances.[16] If such distances were available for a large number of language pairs (and measured according to the same criteria), they would certainly be a very good alternative to the other two types of distances as they encompass most of the difficulties encountered in acquiring a language. Practical challenges, however, stand in the way of expanding the existing dataset.

To close the discussion on distances one has to point out the chasm between the versions of the same language used in different countries and regions. There is, for example, a large (and ever increasing) number of meanings that are represented by different words in the United States and Great Britain: "garbage" and "rubbish", "cell phone" and "mobile", "subway" and "tube", "faucet" and "tab", "janitor" and "caretaker", "eraser" and "rubber", "truck" and "lorry", to cite a few. Canadians do not always use similar words as Britons, and identical words can as is well known produce different meanings.

We now turn to the issue of ethnolinguistic fractionalization, where linguistic distances will play an important role.

2.2 Ethnolinguistic Fractionalization and its Measurement

As suggested above, the impact of linguistic diversity on economic outcomes is ambiguous. Diversity facilitates the process of creation and innovation. This is studied by Lazear (1999), whose "global team" with members from different cultures or countries contributes to the success of the project the team is engaged in. While matching individuals from different cultures and languages could impose additional costs on the organization, the complementarity of workers' varied skills, may easily offset the cost of cross-cultural interaction. Linguistic and ethnic fractionalization or diversity could however also lead to dysfunctional government institutions, bureaucratic inefficiency, corruption, excessive lobbying and rents, and misguided economic and social policies. The tragic example of post-colonial Africa, where Cameroon has 279 languages, the Democratic Republic of Congo has 217, Nigeria 521, Sudan, 134, Tanzania, 129, makes it well-known but nevertheless painful. The positive and negative empirical aspects of ethnolinguistic fractionalization are covered in Chapter 18 of this volume. For the sake of completeness

---

[16]This distance is used by Chiswick and Miller (2007, Chapter 1), as well as by Hutchinson (2003) and Ku and Zusmann (2008).

of this chapter, we nevertheless discuss some general features, which make some overlap with chapter 18 unavoidable.

The numbers of distinct communities or ethnic groups have often led to argue that linguistic fractionalization has to be reduced to enhance economic progress. To make quantitative statements one has to partition a country into distinct linguistic communities. This group identification problem is not always an easy task.

Group Identification

The first comprehensive country-by-country worldwide study of ethnolinguistic divisions was undertaken by a group of about 70 Soviet ethnographers from the Miklukho-Maklai Research Institute in Moscow, then a division within the Department of Geodesy and Cartography at the State Geological Committee of the USSR. Their construction, widely known as ELF (Etholinguistic Fractionalization), was based mainly on the linguistic and historic origins of various groups. The findings of this remarkable project, conducted long before the birth of the internet and even computers, were published in 1964 in *Atlas Narodov Mira (Atlas of Peoples of the World)*. To a large extent, ELF was not influenced by the political pressures of those days.[17] There is a strong similarity between ELF and other datasets, such as those of Roberts (1962) and Muller (1964). Moreover, Taylor and Hudson (1972) point out that ELF contains no systematic differences between countries from the Western and Eastern blocs. After the almost immediate introduction of the study to the Western literature by Rustow (1967) and Taylor and Hudson (1972), this dataset still remains the most influential source in research concerned with diversity.

ELF data were later expanded by Alesina et al. (2003), who disentangle the linguistic and ethnic aspects of fractionalization and construct separate datasets determined by linguistic, ethnic and religious affiliation. The impressive Alesina et al. datasets cover some two hundred countries, 1,055 major linguistic and 650 ethnic groups. Alesina and Zhuravskaya (2008) went a step further and, by using census data, extended the previous dataset to cover about 100 countries on a sub-national (regional) level. Desmet et al. (2009) constructed an alternative dataset using distances based on *Ethnologue* (2009).[18]

---

[17] "Politically delicate" countries such as South and North Vietnam, South and North Korea, and Taiwan were nevertheless omitted.

[18] It is worth pointing out that almost all researchers on diversity refer to ELF as an

These datasets are used in conjunction with measures or indices that summarize them.

Diversity Indices

In most of the empirical and theoretical literature on diversity and fractionalization, the societal fractionalization index is determined by the probability that two randomly chosen members of the society belong to *different* linguistic groups. In a monolingual society that consists of one group only, such probability, and thus, the index of fractionalization, is equal to zero. If, on the other hand, the society consists of a large number of distinct groups with small memberships, the probability that two randomly chosen individuals speak different native languages is quite large. The highest degree of fractionalization (that is 1) obtains if each individual speaks a distinct language, and the probability that two individuals share the same language is zero.

This fractionalization index has been derived independently (and, naturally, under different names) in many areas of research, including economics, political science, linguistics, sociology, genetics, biology and other disciplines. The first reference appears in the contribution by Gini (1912), who called it the *mutuality index*. It was followed by Simpson's (1949) index. In linguistics it was introduced by Joseph Greenberg (1956) as the *monolingual non-weighted index*, or A-index.[19] The formal representation of the A-index is given by

$$A = 1 - (s_1^2 + s_2^2 + \ldots + s_n^2),$$

where $s_1, s_2, \ldots, s_n$ are the population shares of the n linguistic groups comprising the entire society, the sum of which is equal to one. Note that the functional form of the A-index is equal to one minus the celebrated Hirschmann-Herfindahl Index (HHI), defined for an industry with multiple

---

index of ethnolinguistic fractionalization. In fact, ELF is merely a dataset. What the literature refers to as the ELF index is Greenberg's A-index (see below) based on the ELF dataset, which was first calculated by Taylor and Hudson (1972). Likewise, the measures computed by Easterly and Levine (1997) are values of Greenberg's A-index for the Muller and Rogers datasets, respectively, and have nothing to do with ELF. The same comment applies to recent studies of fractionalization that calculate various indices based on *Ethnologue*'s trees, the *Britannica*, the *CIA World Fact Book* (2009) or other datasets. Thus, an index that is *not* based on the Soviet 1964 dataset, could be given different names, but ELF should not be one of them.

[19]This notation will be kept throughout this chapter.

firms.[20] HHI is computed as:

$$HHI = s_1^2 + s_2^2 + \ldots + s_n^2,$$

where s_k stands for the market share of firm k.

The A-index is not the only possible functional form. Another form, called *entropy,* was suggested by Shannon (1948):

$$Ent = -(s_1 \log s_1 + s_2 \log s_2 + \ldots + s_n \log s_n).$$

Both indices satisfy two fundamental requirements of diversity formulated by Shannon:

(a) *Size uniformity*: For a given number of groups, the index reaches its maximum when all groups are of the same size,
(b) *Richness*: If all groups are of equal size, the diversity index of a society with a larger number of groups will be larger.

The A-index offers a rather coarse treatment of diversity since it takes into account only the size of different groups. In practice, defining whether a group is distinct from another can be difficult, unless one uses distances.

Before proceeding with the formulation of indices, based both on sizes of groups and their linguistic distinctiveness, consider the example of two West-European countries, Andorra and Belgium discussed in Desmet et al. (2009). In tiny Andorra roughly half of the population speaks Catalan and the other half speaks Spanish (two relatively similar Romance languages), whereas in Belgium about 60 percent speak Dutch, a Germanic language, and the other 40 percent speak French, a Romance language. Given the linguistic proximity of Spanish and Catalan, as opposed to French and Dutch, one would expect Belgium to be linguistically more diverse than Andorra, though the A-index is larger for Andorra. To overcome this odd property of the A index, Greenberg (1956) proposed a so-called B-index, that accounts for distances between groups:

$$B = 1 - \sum_{i=1}^{K} \sum_{j=1}^{K} s_i \times s_j \times (1 - d_{ij}).$$

To verify that the A-index is, indeed, a special case of B, note that for every term that includes different groups i and j, the A-index sets the distance

---

[20] HHI is often viewed as an indicator of the industry's degree of monopolization and is widely applied in competition and anti-trust law.

d_{ij} equal to 1, so that the term vanishes. If i and j are identical, $d_{ij} = 0$, so that the term boils down to $s_i \times s_i$. Therefore, the A index collects only $s_i \times s_i$ terms, all the others will be equal to 0.

By using some simple algebra, it is easy to transform this index into the following more useful form:[21]

$$B = \sum_{i=1}^{K} \sum_{j=1}^{K} s_i \times s_j \times d_{ij}.$$

Note that B has a nice intuitive interpretation: it represents the average linguistic distance between all pairs of members of the society.

Index B has an important advantage over A since it satisfies a "continuity" property: If the distance between two groups is close to zero, the diversity of the society with two similar groups (e.g., Andorra) is close to the extreme case of a homogeneous society where both groups are merged into a single one. In the Desmet et al. (2009) study of the linkage between societal diversity and the scope of redistribution in a sample of 218 countries, the distance-based index has an explanatory power that is far superior to that of the A-index.

Desmet et al. (2005) propose a variant of the B-index in the case of a dominant group (called *center*) as in Spain, Russia, Kyrgyzstan, Laos, Thailand, Iran, Saudi Arabia, Kuwait, among others.[22] Their PI index takes into account only the distances between the center and peripheral groups, but not between peripheral groups themselves. The functional form of the peripheral index PI is similar to B, except that the distance between every pair of peripheral groups is 0. Thus, in a society with a central group whose population share is s_c, the PI index contains only $s_i \times s_c \times d_{ic}$ terms where d_{ic} is the distance between language i and the language of the central group. Unlike the A-index, the B- and PI-indices fail to satisfy Shannon's size-uniformity and richness conditions.

---

[21]See Desmet et al. (2005, 2009), Bossert et al. (2006).

[22]Dominance is not always correlated with the relative size of the group. For example, Tutsis, who account for only 14 percent in Rwanda, represent a dominant minority group in the country, where almost everybody else is of Hutu ethnic origin. Laitin (2000) points to several cases of minority groups imposing their language on the majority, such as Spanish in South America, Amharic in Ethiopia, and Afrikaans in South Africa.

2.3 Distances: Economic Impact

In this section, we examine the economic impact of linguistic distances only. The reader will find a full discussion of the economic impact of diversity in Chapter 18 of this volume.

Impact on International Trade

The process of globalization has forced workers (essentially white collars) to learn foreign languages to break linguistic barriers. In most cases, Europeans have chosen English as a second language, and English has become the *lingua franca* in Europe. Worldwide, English is spoken by more than 1.5 billion people (Crystal, 2003), and is probably the language that is most often used in international contacts and trade.

Studies on the impact of countries' linguistic differences on bilateral trade flows are often based on what is known as the "gravity model," whose name comes from its analogy with Newton's Law of Universal Gravitation. Newton's reasoning was that any two objects A and B in the universe exert gravitational attraction on each other with a force that is proportional to the product of their masses and inversely proportional to (the square of) the distance that separates the two objects. Tinbergen (1962) suggested to apply this law to the analysis of international trade flows between countries A and B, assuming that the attraction force represents the volume of exports from A to B, the masses are measures of wealth or income (population, gross domestic product, ...), and distance is represented by the geographic distance between the capitals of A and B. Both masses should have a positive effect on exports between A and B, while distance should have a negative effect.

Economists soon realized the importance of alternative distances, such as transaction costs caused by the search for trading partners, bilateral tariffs, non-tariff barriers (prohibiting imports of wines containing sulfites), countries' geography (adjacency, islandness, landlockedness, common sea), former colonial links, immigration stocks (that could foster trade between countries through the building of social and commercial networks), common markets and common currency zones, and of course linguistic (but also cultural and genetic) distances.

Melitz's (2008) study is of particular interest, since he uses two measures of linguistic distances between trading partners and tries to estimate their effect. Open-circuit communication (OCC) demands that the language be either official or widely spoken (at least 20 percent of the population knows

the language). Spanish, for instance, will be an OCC between Bolivia (where 44 percent of the population knows Spanish) and Mexico (where this percentage is 88). Likewise, Arabic will be an OCC between Mauritania (38 percent) and Iraq (58 percent). Melitz identifies fifteen such languages. A direct communication (DC) language is a language that is spoken by at least four percent in each country. The rationale for introducing this second measure is based on Melitz's claim that any linguistic distance measure works in explaining the intensity of trades, but not all of them tell the same story. He suggests to distinguish channels through which the effect takes place, and separates therefore OCCs that depend on translation (which can be produced as long as there are enough people who can provide it in both countries) and DCs which make possible direct communication between traders. His estimation results point to the following conclusions. Direct communication has obviously the largest positive effect on trades: A ten percent increase in the probability that two citizens, one in country A, the other in B, speak the same language increases their trades by ten percent. European OCCs also contribute, but somewhat less.[23] Interestingly enough, Melitz shows that English is not more effective than other European languages in promoting trade. But in all cases, distances matter, and the larger the linguistic distance between two countries, the less they trade.

Impact on Migrations

While migration decisions are to large extent based on the existence of networks of former immigrants in the country of destination, the decision to migrate is also influenced by a common language between the source and the destination country. A common language is also more likely to attract high-skilled than low-skilled workers.

The standard approach in analyzing the trade-offs of a decision to migrate is based on evaluating costs and benefits. The prospects of higher wages or other benefits[24] are contrasted with the monetary and psychological costs, adjustment to a new culture and possible uprooting of the family. The cultural and linguistic frictions in a new country can profoundly influence individual decisions. The degree of labor adjustment and complemen-

---

[23]Other OCC languages such as Arabic, Chinese, Hindi, Malay, Farsi and Turkish have no significant effect on trade.

[24]In migrations between developing countries prevention against risks rather than income maximization seems to be of major factor of the migration decision. See Guilmoto and Sandron (2001).

20

tarity to existing technological processes are also crucial to the well-being and the adaptation process. To illustrate this point, consider the flow of Chinese immigration to the United States and to Japan.[25] Chinese immigrants find it difficult to integrate into Japanese production processes that are characterized by an intensive level of interaction and communication within the labor force. The situation is more flexible in the United States, where immigrants can rapidly exhibit a reasonably high degree of labor complementarity. In addition, even though Chinese characters are used to some extent in Japanese writing, their pronunciation is completely different from the Chinese. More importantly, the Chinese and English language structures are very similar, while being quite distinct from Japanese.

The differences are even more profound in the case of immigration from India to the United States and Japan. In 1990, Japan modified its restrictive immigration policy (Immigration-Control Refugee-Recognition Act – ICRRA) to attract highly educated engineers and computer specialists from abroad, including India. Cultural and technological adjustment for Indians in Japan has, however, proved quite difficult and their satisfaction with working conditions is quite low compared with immigrants from other Asian countries.[26] On the contrary, the ability of a large number of educated Indians to speak English mitigates the degree of cultural friction they face in the United Sates.

The form of the typical immigration equation is very close to the trade equation, but its theoretical underpinnings are different. Beine et al. (2009), for example, examine the determinants of migration flows between 1990 and 2000 from 195 countries to 30 OECD countries. Since the incentives to migrate may be different they also distinguish flows of low-skilled (with less than upper secondary education) and high-skilled workers (with post-secondary education), studying how migrants sort themselves across destinations. Instead of estimating an equation for each skill level, they estimate how the ratio of high skilled to low-skilled emigrants is affected by a certain number of determinants, the most important of which are the sizes of the diasporas, wages and the generosity of immigration policies in the destination country. But a common language between the country of origin and destination also has a strong influence on the migratory flow.

Migrations are not only international and intercontinental, they happen also within a continent and even across regions within a country. Falck et al.

---

[25]See Fujita and Weber (2010).
[26]See Ota (2008).

(2009) examine the effect of various factors, including linguistic data on the variation of phonological and grammatical attributes across regions in Germany. They find that these have a significant effect on regional migrations, beyond what geographical distance only would suggest. They interpret the closeness of dialects as explaining cultural identity.

Impact on Literary Translations

Translation of literary works is an essential vehicle of transmitting culture, and as Susan Sonntag said,[27] "a passport to enter, a larger life, that is the zone of freedom." However, translations are often accused, mostly by sociologists, but also by economists, to be dominated by English, or more precisely, by translations "from English."[28] This is hardly surprising, since the population speaking English as a first language is, with the exception of Mandarin, the largest in the world. Moreover, English is spoken in countries with very distinct cultures (Great Britain, Ireland, Canada, the United States, former British colonies in Africa, South Africa, Australia, New Zealand and Tasmania, India, or the West Indies[29]). It can therefore be expected that more fiction is produced in English than in many other languages, and why publishers in these countries do not feel the same interest to translate from other languages, since the diversity that readers look for is available without translation.[30]

It is therefore true that English has a privileged (rather than a dominat-

---

[27]In her acceptance speech of the Freedom Prize of the German Book Trade, Frankfurt, 2003.

[28]See Ganne and Minon (1992), Heilbron (1999), Heilbron and Sapiro (2002, 2007), Janssen (2009), Sapiro (2010) Melitz (2007).

[29]To illustrate this diversity, Ginsburgh et al. (2011) sampled some authors who were shortlisted by, or winners of, the Booker (now Man Booker) prize since its inception in 1969. Here is a list of such English writing authors born elsewhere than in Great Britain, though some may have moved there, and whose native culture often impregnates their work and provides diversity: Chinua Achebe (Nigeria), Aravind Adiga (India), Margaret Atwood (Canada), André Brink (South Africa), John M. Coetzee (South Africa), Ashmat Dangor (South Africa), Anita Desai (India), Kiran Desai (India), Amitav Ghosh (India), Nadine Gordimer (South Africa), Romesh Gunesekera (Sri Lanka), Abdulrazak Gurnah (Zanzibar), Mohsin Hamid (Pakistan), Thomas Keneally (Australia), Doris Lessing (Rhodesia), Yann Martel (Canada), Timothy Mo (Hong Kong), V. S. Naipaul (West Indies), Ben Okri (Nigeria), Michael Ondaatje (Sri Lanka), Arundhati Roy (India), Salman Rushdie (India), Indra Sinha (India), Ahdaf Soueif (Egypt). This may also be the case for France, Spain, and other important languages.

[30]See also Pym (1999).

ing or hegemonic) position on the market for novels. Authors who write in English benefit from three advantages. They have a large market in their own language, they also have access to other large markets since their books get translated, and finally, they do not "need" spending on translation costs. This is also reflected in the small numbers of British (and probably American) high school students who learn foreign languages. A recent European survey shows that while some 98 to 99 percent of children in the European Union study at least one foreign language, the number is 81 percent in Ireland and only 48 percent in the United Kingdom. Some do indeed benefit from free lunches.

Ginsburgh et al. (2011) nevertheless show that, if account is taken of factors such as production in the source language, reception in the destination language[31] and proximity between cultures (represented by lexicographic distances), the number of titles translated from English, in fact, falls behind the number of titles translated from other languages, including Scandinavian ones and French.

Domination, therefore, does not seem to come from the number of titles translated from English, but from the number of books sold, of which Dan Brown's *Da Vinci Code* is one example. As of May 2006, sixty million copies of the book, published in early 2003 were in print or sold.[32] The *Code* was translated into 44 languages, and by October 2004, it had generated some sixteen titles supporting or debunking the *Code*.

3 Standardization

The negative aspects of fractionalization in a multi-lingual country or community are exacerbated by the substantial costs needed to maintain several (and sometimes a large number) of official languages. Even before the 2004 enlargement, the institutions of the EU were the largest recruiter of interpreters and translators in the world. In 1999 the total translation and interpretation costs for the Commission alone amounted to some 30 percent of its internal budget.[33] The burden of maintaining official languages is not limited to the direct costs of translation and interpretation. Communication constitutes an even more serious challenge in societies with a large number

---

[31]Both production in the source language and reception in the destination language are proxied by populations, since the number of titles published is not available for all the countries that they study.

[32]http://en.wikipedia.org/wiki/The_Da_Vinci_Code (accessed June 12, 2009).

[33]De Swaan (2001, p. 172).

of official languages. Errors as well as the delays caused by translations, may end up paralyzing multilateral discussions and negotiations.

The basic principles of political accountability and equality among citizens require, however, that all, or at least a substantial part of the full-fledged translation services, will have to be maintained in some of the EU institutions (Council of the European Union, European Council, European Parliament). Failing to provide translation services by the EU may simply shift the provision of the service to individual countries, leading to duplications that may raise the total cost of services,[34] as well as to divergent translations and interpretations.[35]

This will often lead a fractionalized society to introduce some degree of *standardization.* This concept is present in Max Weber's celebrated *rationalization theory* outlined in his 1914 essay (translated into English in 1968): the practices of state standardization include a common currency, a common legal system, and a common administrative language.

The modes of introducing standardization vary across countries and historical periods. An extreme and bloody way of imposing such a policy is what came to be called "ethnic cleansing," that was unfortunately practiced on various occasions. A less extreme and apparently more manageable way of standardizing is to impose a unique language, or to force citizens to speak a unifying language in addition to their native tongue. The central power can also abolish education in languages it does not favor.[36] French was imposed to all French provinces in 1539 by King François I, though in 1794, Henri Grégoire[37] showed that in 68 out of 83 local states (départements) another tongue was spoken next to French, and French was spoken by only 3 out of 28 million citizens. In 1972 French President George Pompidou claimed that there "was no place for regional languages in France, which is destined to play a fundamental role in Europe." In 1975 a new law stipulated that French must be used in commerce, by the media and by the public service. In 1994, the Toubon Act made French the compulsory medium in consumer affairs, employment, education and at congresses held in France (Spolsky, 2004, pp. 66-67).[38]

---

[34]Mamadouh and Hofman (2001).

[35]Such as the one discussed earlier on UN Resolution 242.

[36]In some cases, standardization emerges in a natural way, when states recognize an existing *lingua franca.* Laitin (2000) points out that such was the case with Swahili in Tanzania, Bahasa in Indonesia and English in both the United States and England.

[37]See De Certeau, Julia et Revel (1975).

[38]Even nowdays *Ethnologue* identifies seven languages in France, each having more than

The russification of the territories of the Russian Empire, is another example. After quelling the Polish and Lithuanian uprisings in the 1860s, the Russian government, threatened by the prospect of Polonization expressed by the increasing influence of the Catholic Church and the Polish and Lithuanian languages, implemented the use of Russian as the only administrative, official and working language. Lithuanian and Polish were banned from usage in public places and later in schools and administrations. General Mikhail Muravyov, the Governor General of Lithuania, claimed that "what the Russian bayonet did not accomplish, the Russian school will."[39] All students were required to study Russian and social mobility was not possible without being proficient in Russian.

A similar standardization was imposed in China. After the 1911 Revolution began promoting as national language, a vernacular Chinese based on Mandarin. Within a few years it was understood that the Beijing pronunciation would be the new standard. In 1958 the government initiated a linguistic reform whose main goal was to simplify written Chinese and promoting a "general language" called Putonghua. It is interesting to point out a similar unifying effort to encourage the Singapore Chinese community to adopt Mandarin had a mixed effect and, according to the 2002 census, only 45 percent of Chinese families speak Mandarin at home, while the rest speak other Chinese dialects or English.[40]

The attempt at enforcing a unique national language led to disastrous consequences in Sri-Lanka (formerly Ceylon) which shows how emotional, explosive and dangerous the choice of official or national languages may become. Sri-Lanka has two major ethnic and linguistic communities, the Sinhalese majority, predominantly Buddhist, and the Tamil minority, mainly Hindu,[41] who had peacefully coexisted over a period of about two thousand years. After hundred and fifty years under the British rule, the island attained self-governance in 1948. The superior system of teaching English in northern Tamil regions allowed Tamils to have easier access and numerically disproportionate representation in university education and jobs in the pres-

---

half of million of speakers: Alsatian, Basque, Breton, Italian, Portuguese, Algerian Arabic and Kabyle (Berber). This is without mentioning Corsican, Gascon and Provençal, among others with hundreds of thousands of speakers.

[39] See O'Connor (2003, p.58).

[40] See Spolsky (2004, pp. 180-181).

[41] The *CIA World Factbook* (2009) estimates that the Sinhalese majority and the Tamil minority constitute, respectively, some 74 and 18 percent of Sri-Lanka's population of about 21 million people.

tigious government sector. According to deVotta (2004), in 1946, "Tamils made up 33 percent of the civil service and 40 percent of the judicial service. They also accounted for 31 percent of the students in the university system. In the medical and engineering fields, Tamils numerically equaled the Sinhalese."

The advancement of a larger number of educated Tamils and the desire for a larger piece of the national pie drew many Singhalese into supporting the Sinhala-Only (only Sinhalese) movement, which was led by Buddhist monks who claimed that not only the Sinhala language, but Buddhism itself would be threatened if parity between Sinhala and Tamil were sustained. Another important element in rejecting Tamil was the Sinhalese fear of being dominated by the well-developed Tamil literature and culture. The Sinhala-Only Act in 1956, considered by Tamil leaders as a form of apartheid, led to mass riots in which hundreds of people lost their lives. Brown and Ganguly (2003, p.11). note that "the passage of the Sinhala-Only Act was a turning point in the Sinhalese-Tamil relations. Tamil grievances subsequently grew, because in Sri Lanka as elsewhere, language policies had wide-ranging implications for educational and economic opportunities." The government took a range of conciliatory measures in favor of the Tamil minority. Following the 1977 riots, the government abolished controversial university-entrance policies and the 1978 Constitution officially awarded Tamil the status first of a *national* and then in 1988 of an *official* language. All this came too late and tens of thousands of lives were lost over the course of twenty six years of a full-fledged civil war, which only recently came to an uneasy halt.

Bellos (2011, pp. 202-223) points out that there exist also cases in which the language that gains prestige is the language of the conquered, which may even crowd out the language of the conqueror in some uses. This was so when the Akkadians conquered Sumer around 2250 B.C. They adopted the Sumerian script, and Sumerian became the mark of educated people. This also happened to Greek in the Roman Empire.

Standardization does not necessarily refer to the imposition of a unique language, but rather to restrict the number languages used for official, legal or educational purposes. An example of such a policy is the so-called three-language-formula adopted in India some fifty years ago. The formula, whose aim was to balance efficiency, national pride, sensitivity and economic well-being of multiple linguistic groups, was introduced as a national policy response to bitter complaints from Southern states. Since Hindi is not widely spoken there, the Southerners (mostly Tamil Nadu) felt discriminated against and claimed that the use of Hindi in government services forced

them to learn two languages (English and Hindi) whereas Hindi speakers had to learn only English. The three language formula (with some variations across states) implied that children in Hindi speaking states would study three languages: Hindi, English and one of the regional, preferably Southern, languages, whereas children in non-Hindi speaking states were to be taught Hindi, English and their regional language.[42] This masterful and well-crafted formula that seemed to achieve group identity, preservation of mother tongues (by sustaining proficiency in regional languages), national pride and unity (through spreading Hindi), and administrative efficiency and technological progress (by means of acquiring English), failed to achieve the success the formula's creators were hoping for. The reasons were insufficient funding, lack of teachers, inadequate support of the regional administration and little enthusiasm on the part of students and their families to undertake the required efforts to learn languages spoken in other regions. In Hindi regions, relatively little effort or resources were put in studying English and even less so in learning a third language. In Tamil Nadu, English and Tamil are studied quite extensively, whereas Hindi received lip service.

A variant of the three language formula was introduced in Nigeria, the most populous African country with 141 million inhabitants[43] who speak 527 languages[44] and are divided into 250 ethnic groups. Nigeria, like many other African countries, turned out into "a principal victim of God's wrath aimed against those who constructed the Tower of Babel" (Laitin, 1994, p. 623). English is the official language in Nigeria, used in government and education, but Hausa (spoken by 18.5 million in the north), Igbo (18 million in the south-east) and Yoruba (19 million in the south-west) are official regional languages. In addition, several other regional languages have some official status. The three language formula based on the use of Hausa, Yoruba, and Igbo, was considered a unifying device for this diverse country. However, like in India, its implementation was inhibited by the lack of qualified instruction and the resistance of linguistic groups, identified with one language, that were forced to learn another major language.

Linguistic standardization inevitably restricts the usage of some lan-

---

[42]There were some complaints suggesting that native speakers of minority languages in states where neither Hindi nor English are state languages, were forced to study three languages, in addition to their native one: Hindi, English and the state language. Since the Hindi-speaking states require learning of two languages, Laitin (1989) labeled the three-language formula as the 3 ± 1 arrangement.

[43]All the Nigerian data are from *Enthologue* (2009).

[44]By comparison, India with over one billion inhabitants has 'only' 452 languages.

guages, which are not included in the set of the official ones, but its effects go much beyond restricting access to information. Standardization may alienate groups of individuals whose cultural, societal and historical values and sensibilities are not represented by the official languages and consequently create *linguistic disenfranchisement*.[45] In the context of the European Parliament, for example,

> "the right of an elected Member to speak, read and write in his or her own language lies in the heart of Parliament's democratic legitimacy. The case for multi-lingualism is based not only on fairness to Members, from whichever country they are elected. It is necessary to ensure the support of citizens in all Member states; if Parliament does not recognize their language, it is less likely that citizens will recognize it as being *their* Parliament." (Report of Secretary General, document PE 305.269/BUR/fin, 2001.)

In short, standardization breeds disenfranchisement. But how does one balance the effects of standardization and disenfranchisement? In the next section, we examine several aspects of this important question.

4   Disenfranchisement

Before proceeding, we need to stress two important points. Firstly, linguistic or ethnic fractionalization studied in Section 2 was implicitly considered to be exogenous, and used as explaining the impact it had on economic outcomes such as trade, migrations, literary translations, growth, quality of governments, corruption or military conflicts. ELF, for instance, represents a snapshot description of linguistic partitions existing across the globe in the 1960s. Of course, fractionalization had evolved before that: some languages had disappeared, while others were born over time as consequences of trades, migrations, wars and others reasons, including voluntary or non-voluntary standardizations that are pervasive in our history and prehistory. Standardization gives way to new partitions that will at some point, in fifty or one hundred years, be considered as exogenous, though they are no more exogenous than the ones that we studied in Section 2. Second, the considerations on disenfranchisement, including the ways to measure the phenomenon, that

---

[45]See Ginsburgh and Weber (2005) and Ginsburgh et al. (2005).

will follow can be seen as *ex post* measures about who speaks what, but also as a way of studying or simulating the possible consequences of proposed standardization policies.

Linguistic disenfranchisement is essentially based on the fact that the languages of some linguistic groups are not included in the list of official languages. In describing the linguistic situation in Africa, Phillipson and Skutnabb-Kangas (1995, p. 335) offered the following very blunt and succinct assessment:

> "[t]he majority of Africans are governed in a language that they do not understand."

The statement however applies in many other circumstances and serves as basis for the quantitative analysis of linguistic disenfranchisement, which is presented below.

We now consider a multi-lingual society, which faces the challenge of selecting a subset of languages to be used in official documents, for communication between institutions and citizens, debates in official bodies, etc. Call these languages *core* languages. Their choice may have a major negative impact on the well-being of some members by limiting their access to laws, rules and regulations, and debates in their elected or legislative bodies. In some cases, these limitations could even violate the basic principles of the society. To determine the optimal set(s) of core languages, one has to weigh the costs and the benefits of linguistic standardization, but here we merely focus on the construction of disenfranchisement measures, and their use to simulate results that can help politicians to decide on a "satisfyicing" set of official languages.[46]

4.1 Measuring Disenfranchisement

Every member member of a society may be characterized by her linguistic repertoire, represented by the languages she is proficient in. Though, as argued above, the linguistic identity is usually associated with native languages, we extend our analysis to all languages spoken by an individual. Here, the argument is that in order to evaluate the functionality of linguistic policies, one may take into account all languages spoken by an individual and not her native language only. Thus, we distinguish between disenfranchisement indices based on native languages of an individual, and those

---

[46]See Ginsburgh et al. (2005) who compute the optimal number as well as the languages that should be included in the set of official languages.

based on *all* languages she is proficient in. Indeed, in terms of disenfranchisement, an individual may judge the set of core languages on the basis of one criterion only: does her native language belong to this set or not. But she may also be able to communicate in other non-native languages, and we therefore examine both possibilities.

In addition, we also follow Greenberg (1956) and the discussion of fractionalization indices, and construct *dichotomous* and *distance-adjusted* indices. An individual is disenfranchised under the dichotomous approach (Ginsburgh and Weber, 2005), if she speaks no core language; she is not disenfranchised if she speaks at least one core language. However, even if an individual speaks no core language, some of those languages may nevertheless have common roots with her native tongue and could be considered as reducing her disenfranchisement. Taking into account this argument, we also examine indices adjusted for linguistic proximity (Ginsburgh et al., 2005).

This leads us to consider four indices: dichotomous indices, based on native languages only, dichotomous indices based on all spoken languages, continuous (distance-adjusted) indices, based on native languages only, and continuous (distance-adjusted) indices, based on all languages spoken by an individual.

The distinction between native and non-native languages has an important implication for the analysis. If we consider native languages only, it is sufficient to disaggregate the society into groups according to native languages as is done in ELF. If disenfranchisement is determined on the basis of the entire linguistic repertoire of individuals, the division according to native languages is too coarse and we need to disaggregate into *clusters* of individuals with identical linguistic repertoires. For, example in a society with two spoken languages there are three clusters: monolingual speakers of each of two languages and bilingual individuals who speak both.

Since the discussion of how to construct these indices is rather technical, we refer the reader to Ginsburgh and Weber (2011, p. 130-133).

4.2 Simulating the effects of standardization in the EU

The European Union's official linguistic policy is based on Regulation 1/1958 of the Treaty of Rome drafted in 1958, recognizing Dutch, French, German and Italian as official languages. But this number went from four in 1958 to 23 in 2011. The regulation basically says that all official languages should have equal treatment. Reality, including in the Parliament, is however very

different, as discussed by Ginsburgh and Weber (2011, pp. 162-177) and it is unavoidable that at some point, the EU which counts 27 countries and 23 official languages, will have to consider a certain degree of linguistic standardization.

We now show how the various indices can be used to judge the impact on disenfranchisement of various choices of core languages. We formulate a procedure for selecting subsets of languages among all eligible 23 official languages so as to minimize the EU-wide disenfranchisement rate, which measures the share of citizens (in the EU as a whole, but also in each member country) who would be unable to communicate under a particular restricted set of languages.[47] Our assumption is that the members of the Council who are responsible for the language policy would take such information into account when casting their vote for or against a specific standardization proposal.[48] We implement this procedure for different subsets consisting of one, two, three etc. languages, focusing on the obvious subset consisting of the following six languages: English, French, German, Italian, Spanish and Polish.

In order to deal with the language repertoires, one has to use census or survey data that include questions on the entire repertoire of each individual and not only on the native language or on the language that is spoken at home. In addition, it is useful to have some idea on the proficiency of the languages, and this is rarely dealt with in censuses, more often in surveys.

The Special Eurobarometer 243 (2006) survey on which the simulations are based was carried out in November 2005 in 29 countries, including all 25 member countries of the European Union, Bulgaria and Romania (who were not yet members in 2005) and two candidate countries, Croatia and Turkey. In most countries, 1,000 citizens were interviewed.

Who Does Not Speak English?

Table 4 exhibits disenfranchisement rates[49] for the seven most widely spread languages in each EU member country. The results lead to several observations. Firstly, even though English is the most widely spoken language, it

---

[47]In our definition of disenfranchisement, a citizen is considered disenfranchised in a language (a) if he does not speak it (that is, if he does not cite it among the languages that he "knows") or (b) if, when asked how proficient he is in a language that he "knows" he responds that his knowledge is only basic.

[48]According to EU's rules, votes on linguistic matters have to be unanimous.

[49]As defined in footnote 55.

would nevertheless disenfranchise 62.6 percent of EU citizens if it were the only official language. Moreover, there are only seven countries were less than 50 percent of the population would be disenfranchised. But the EU-wide disenfranchisement rate rises to 75.1 and 80.1 percent if English were replaced by German or French, respectively, and it would be even worse if Italian or Spanish were chosen. Secondly, with the exception of English, German, French, Italian (and Russian), no language is spoken by more than five percent of the population in more than two European countries. Finally, though Russian is not an EU-official language, it disenfranchises less people in the EU than many official languages (Bulgarian, Czech, Danish, Estonian, Finnish, Greek, Hungarian, Irish, Latvian, Lithuanian, Maltese, Portuguese, Slovak, Slovenian and Swedish).

Insert Table 4

Young people often speak foreign languages more readily, but English is the only language for which disenfranchisement rates are significantly lower among younger generations. If English were the only official language, EU's disenfranchisement would drop from 62.6 to 44.6 percent, if the proficiency level of English in the whole population were equal to that of the 15-29 years old generation. Therefore, one could expect that some 30 to 40 years from now, English would be spoken by more than half of the EU population. A similar calculation for German or French would lead to global disenfranchisement rates of 73 and 77 percent. Italian, Spanish and Dutch would even do worse.

A linguistic reform in which English would be the only official language would, nevertheless, not be satisfactory.

More Languages Than English

We now address the question of whether a subset of official languages could do better. Calculating disenfranchisement for every subset of 23 languages would be a cumbersome task, and nobody could read the millions of tables that would be the outcome of such a computation. Therefore, we chose a procedure which selects the subsets of languages that minimize disenfranchisement in the EU for every given number of languages. Let m take the values 1, 2, 3, ..., 23. Then, for every m, denote by T_m the subset of the 23 languages that minimizes the disenfranchisement rate over all sets with m languages, ending up with a set T_m for every m between 1 and 23.

Though this calculation is conceptually simple, it would, in practice, still require a very large number of computations if the number of languages among which core languages are chosen is large.[50] However, since European languages differ considerably in the number of their speakers, 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, identifying the most suitable combination is often easy and at any stage in the analysis the number of possibilities to be considered is relatively small.

In principle, a language may enter into the optimal set (given the number of languages) and then move out when we add one or two more languages. This does not happen in our case, and once a language is in the optimal set containing m languages, it would stay there, whatever the number of additional languages in optimal sets containing more languages.

The results of these computations are reproduced in Tables 5 and 6, respectively, for all respondents and respondents who are less than 30 years old.[51] Each column indicates which language should be added to the subset formed by the languages reported in the preceding columns so as to minimize EU's disenfranchisement rate. Consider first the results where all respondents are taken into account. The optimal one language set is English. For two languages, the optimal set contains English and German, and so on.[52]

Insert Table 5

The marginal contribution of each additional language to reducing disenfranchisement falls under one percent of the EU population once the number of languages exceeds 13 and the differences between marginal contributions attributable to further candidate languages are often minute. To save space, we only report on the first 11 languages.

English is clearly the first language in any sequence as it is spoken well or very well by one third of the EU population. German and French are in close race for the second position; German, with a 49.3 percent disenfranchisement rate, fares better than French with 50.6 percent. The bundle of

---

[50]For example, if $m = 6$, the number of arrangements of 6 languages among the 23 is equal to $18 \times 19 \times 20 \times 21 \times 22 \times 23$, which is of the order of 72.6 million.

[51]For further details on the calculation and additional results, see Fidrmuc, Ginsburgh and Weber (2007).

[52]Note that there are instances where two languages result in approximately the same reduction in disenfranchisement at a particular step in the sequence. For example, the tenth language could be Czech or Greek.

33

three languages leads to a disenfranchisement rate of 37.8 percent. Italian, Spanish or Polish would each make almost the same contribution to reducing disenfranchisement further. Spanish, in turn, performs only marginally better than Polish. With the six largest languages included, 16 percent of the EU population would still remain disenfranchised. Adding Romanian brings the residual disenfranchisement rate further down to 13 percent. Of course, important differences across countries remain. The most dramatic case is Hungary, where only 16 percent of the population can speak one of the first seven languages. Not surprisingly, Hungarian becomes the eighth language in the sequence. This also has a positive impact on Slovakia whose disenfranchisement rate declines from 70 to 57 percent. Portuguese is the ninth language, followed by Czech and Greek tied in the tenth position.

The disenfranchisement rates in Table 5 are a snapshot of the situation at the time of the survey (end of 2005). However, the knowledge of languages changes over time. In particular, the pattern of learning foreign languages may change both with respect to languages that are popular and the frequency with which people learn other languages. Therefore, we calculated a sequence of optimal sets based on the disenfranchisement rates of the youngest generation (15 to 29 years old) only. This sequence is presented in Table 6.

Insert Table 6

The first difference is that German which was second to enter in Table 5 (whole population), is replaced by French. This is due to the fact that among the younger generation in Germany and in Austria, 60 percent of the population knows English, so that German becomes less necessary. Beyond the first two languages, the sequence is essentially the same as before, and includes English, French, German, Italian, Spanish, Polish, Romanian, Hungarian, Portuguese, Czech, Greek and Bulgarian, Dutch, and Finnish, Slovak, Lithuanian and Latvian (the last four languages, along with Russian, are all in a tie for the fourteenth position). The criterion used before (a language's contribution to reducing disenfranchisement should be at least 1 percent) now results in ten languages. The disenfranchisement rate that would prevail among the youngest generation with ten official languages is 3.9 percent. This percentage is even likely to decrease further as more and more children in upper secondary education study languages (essentially English, but to some extent, French and German also).

Ginsburgh and Weber (2011, pp. 159-160) also report results with continuous indices in which distances are accounted for. In the single-language

34

(English-only) scenario, accounting for linguistic proximity reduces the EU-wide disenfranchisement considerably, from 62.6 to 43.1 percent. Adding French reduces disenfranchisement also in all Romance-language countries, bringing the EU-wide rate to 24 percent. A deviation from the two sequences reported above is that Polish now comes in the third position ahead of German that becomes fourth. Italian is the fifth language followed by Hungarian and Spanish. Greek ties with Romanian for the eighth position. The requirement of at least 1 percent contribution to reducing disenfranchisement cuts off the sequence at nine languages with the resulting disenfranchisement rate of 2.9 percent. Adding further languages (Czech, Finnish, Bulgarian, Swedish and Portuguese) brings the residual disenfranchisement rate to 0.9 percent. The gains from adding the remaining languages (Danish, Dutch, Estonian, Irish, Latvian, Lithuanian, Maltese, Slovak, and Slovene) are negligible.

These sequences of sets which minimize EU's global rate of disenfranchisement can now be used to simulate the political feasibility of linguistic reforms and have the European Council (or the Parliament) casting votes on their preferred set.[53] If an agreement can be reached on a subset of today's 23 official languages, one can consider different possibilities that will compensate those countries whose languages are not included in the official set, or make countries whose languages are included to pay for the support of the core set.[54]

5   Concluding Remarks

The role of culture on economic outcomes has recently become the issue of intensive interest in various branches of economics.[55] Our main focus was linguistic diversity as one of the important aspects of cultural heterogeneity.

The issue of linguistic diversity can be traced back to the well-known mythological attempt of a monolingual "people" to build a tower in Shinar (Babylonia) to be closer to the sky. God disliked the idea, and confused the builders speach so that one person could no longer understand another. Today most countries are (still) multilingual (with notable exceptions, such as North Korea),and the challenges of multilingual societies persist everywhere.

---

[53]See Fidrmuc et al. (2009).

[54]See Fidrmuc and Ginsburgh (2007), and Ginsburgh and Weber (2011, pp. 187-200) for further extensions.

[55]See Fernández (2010) for an extensive review of the literature.

It is often argued that the linguistic diversity represents an impediment to economic progress and institutional development of all sorts. To address this question, one needs to quantify linguistic fractionalization. This, in turn, requires a way of dividing a society into distinct groups, and possibly taking into account "how much" distinct they are, by using linguistic (or other types of) distances.To mitigate the negative impact of linguistic fractionalization, societies has often chosen to standardize by reducing the number of official languages. The numerous examples of such standardization policies (in the Russian Empire, India, Sri Lanka, among many others) often lead to the feeling of disenfranchisement experienced by some population groups. The American-Chilean professor and writer Ariel Dorfman (2002, p.92) describes those feelings in the context of the ascendency of English by arguing that it restricts the access of many to her or his own language, leaving

> "too many invisible losers, too many people silenced. Do you come from a place that does not control a language that commands respect? Do you reside in a language whose existence does not have the kind of value in the marketplace that can get you a good job and help you in everyday's struggle to survive?"

The search for a compromise between efficiency and the sentiment of being disenfranchised represents a serious challenge for any multi-lingual country or union. In this chapter, we consider a formal analysis of linguistic disenfranchisement and describe it can help in analyzing linguistic policies, using as example the European Union which has to cope with 23 official languages.

The aim of this chapter was to formally examine two opposing forces, standardization and efficiency on the one hand, and cultural attachment and linguistic disenfranchisement, on the other, and to outline ways of bringing them to balance each other.

Many problems remain open, in particular the speed at which people acquire languages, forgetting their own mother tongue , as well as the speed at which languages change and become "new" languages in our globalizing world. The chapter on the dynamics of cultural changes by Bisin and Verdier (2013) in this *Handbook* sheds light on how this issue can be tackled.

6   References

Alesina, Alberto, Reza Baqir, and William Easterly (1999), Public goods and ethnic divisions, *Quarterly Journal of Economics* 114, 1243-1284.

Alesina, Alberto, Arvind Devleeschouwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003), Fractionalization, *Journal of Economic Growth* 8, 155-194.

Alesina, Alberto and Ekaterina Zhuravskaya (2008), Segregation and the quality of government in a cross-section of countries, NBER Working Paper 14316.

Atlas Narodov Mira (1964), The Miklucho-Maklai Ethnological Institute at the Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union.

Beine, Michel, Frédéric Docquier and Çaglar Özden (2009), Diasporas, manuscript, Department of Economics, Université catholique de Louvain.

Bellos, David (2011), *Is That a Fish in Your Ear*, London and New York: Penguin Books.

Bisin, Alberto and Thierry Verdier (2013), Trade and cultural diversity, this volume.

Boas, Franz (1940), *Race, Language and Culture*, Chicago, IL.: University of Chicago Press.

Bossert, Walter, Conchita D'Ambrosio, and Eliana La Ferrara (2006), A generalized index of ethno-linguistic fractionalization, IGIER Working Paper 313, Università Bocconi, Milan.

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

Brown, Michael and Sumit Ganguly (2003), *Fighting Words: Language Policy and Ethic Relations in Asia*, Cambridge, MA: MIT Press.

Carrière, Jean-Claude and Umberto Eco (2009), *N'espérez pas vous débarrasser des livres*, Paris: Grasset & Fasquelle.

Chiswick Barry and Paul Miller (2007), *The Economics of Language, International Analyses*, London and New York: Routledge.

CIA World Fact Book (2009), https://www.cia.gov/library/publications/the-world-factbook/

Crystal, David (2003), *English as a Global Language*, Cambridge: Cambridge University Press.

De Certeau, Michel, Dominique Julia et Jacques Revel (1975), *Une politique de la langue: la Révolution française et les patois. L'enquête Grégoire*, Paris: Gallimard.

Desmet, Klaus, Ignacio Ortuño-Ortín and Shlomo Weber (2005), Peripheral diversity and redistribution, CEPR Discussion Paper 5112.

Desmet, Klaus, Ignacio Ortunño-Ortín and Shlomo Weber (2009), Linguistic diversity and redistribution, *Journal of the European Economic Association* 7, 1291-1318.

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

deVotta, Neil (2004), Ethnic domination, violence and illiberal democracy, in Muthiah Alagappa, ed., *Civil Society and Political Change in Asia*, Stanford, CA: Stanford University Press.

Dinesen, Isak (1984, [1960]), *Shadows on the Grass*, London: Penguin Books.

Dorfman, Ariel (2002), The nomads of language, *American Scholar* 71, 89-94.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black (1992), An Indo-European classification: A lexicostatistical experiment, *Transactions of the American Philosophical Society* 82(5), Philadelphia: American Philosophical Society.

Ethnologue (2009), *Languages of the World*, M. Paul Lewis, ed., Dallas, TX: SIL International.

Easterly, William and Ross Levine (1997), Africa's growth tragedy: policies and ethnic divisions, *Quarterly Journal of Economics* 112, 1203-1250.

Ethnologue (2009), *Languages of the World*, M. Paul Lewis, ed., Dallas, TX: SIL International.

Falck, Oliver, Stephan Heblich, Alfred Lameli and Jens Südekum (2009), Dialects, cultural identity, and economic exchange, manuscript.

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

Fearon, James and David Latin (1999), Weak states, rough terrain, and large ethnic violence since 1945, Paper presented at the annual meetings of the American Political Science Association, Atlanta, GA.

Fearon, James and David Laitin (2003), Ethnicity, insurgency and civil war, *American Political Science Review* 97, 75-90.

Fernandez, Raquel (2010), Does culture matter? NBER Working Paper 16277, August 2010.

Fidrmuc, Jan and Victor Ginsburgh (2007), Languages in the European Union: The quest for equality and its cost, *European Economic Review* 51, 1351-1369.

Fidrmuc, Jan, Victor Ginsburgh, and Shlomo Weber (2007), Ever closer union or Babylonian discord? The official-language problem in the European Union, CEPR Discussion Paper 6367, Center for Economic Policy Research, London.

Fidrmuc, Jan, Victor Ginsburgh, and Shlomo Weber (2009), Voting on the choice of core languages in the European Union, *European Journal of Political Economy* 25, 56-62.

Fujita, Masahisa and Shlomo Weber (2010), Immigration quotas in the globalized economy, *Journal of the New Economic Association* 7, 10-23.

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

Gini, Corrado (1912), Variabilità e mutabilità, in *Studi Economico-Giuridici della R. Universitadi Cagliari* 3, 3-159. Reprinted in Enrico Pizetti and Tomasso Salvemini, eds., *Memorie di metodologica statistica*, Roma: Libraria Eredi Virilio Vechi (1955).

Ginsburgh, Victor, Ignacio Ortuño-Ortin and Shlomo Weber (2005), Disenfranchisement in linguistically diverse societies. The case of the European Union, *Journal of the European Economic Association* 3, 946-964.

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

Ginsburgh, Victor and Shlomo Weber (2011), *How Many Languages Do We Need? The Economics of Linguistic Diversity*, Princeton, NJ.: Princeton University Press.

Ginsburgh, Victor, Shlomo Weber and Sheila Weyers (2011), The economics of literary translation. A simple theory and evidence, *Poetics* 39, 228-246.

Greenberg, Joseph (1956), The measurement of linguistic diversity, *Language* 32, 109-115.

Guilmoto, Christophe and Frederic Sandron (2001, The international dynamics of migrations networks in developing countries, *Population: An English Selection* 13, 135-164.

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

Hart-Gonzalez, Lucinda and Stephanie Lindemann (1993), Expected achievement in speaking proficiency, School of Language Studies, Foreign Service Institute, Department of State.

Hechter, Michael (1974), The political economy of ethnic change, *American Journal of Sociology* 79, 1151-1178.

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

Heilbron, Johan and Gisèle Sapiro (2002), La traduction litteraire, un objet sociologique, *Actes de la Recherche en Sciences Sociales* 144, 35.

Heilbron, Johan and Gisèle Sapiro (2007), Pour une sociologie de la traduction: bilan et perspectives, *La traduction comme vecteur des échanges*

culturels internationaux. Circulation des livres de liiérature et de sciences sociales et évolution de la place de la France sur le marcheé mondial de l'édition (1980-2002).

Hofstede, Geert (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*, Beverly Hills, CA: Sage.

Humboldt, Wilhelm von (1988, [1836]), *The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind*, Cambridge, UK: Cambridge University Press.

Hutchinson, William (2003), Linguistic distance as determinant of bilateral trade, Working Paper 01-W30R, Department of Economics, Vanderbilt University.

Janssen, Susanne (2009), Foreign literature in national media, *Arcadia International Journal for Literary Studies* 44, 352375.

Kramsch, Claire (1998), *Language and Culture*, Oxford: Oxford University Press.

Ku Hyejin and Asaf Zussman (2008), The role of English in international trade, Manuscript, Department of Economics, Cornell University.

Labar, Kelly (2010), What is hidden behind the indicators of ethno-linguistic fragmentation, Working Paper, Fondation pour les Etudes et Recherches sur le Développement International.

Laitin, David (1989), Language policy and political strategy in India, *Policy Studies* 22, 415-436.

Laitin, David (1994), The Tower of Babel as a coordination game: political linguistics in Ghana, *American Political Science Review* 88, 622-634.

Laitin, David (2000), What is a language community?, *American Journal of Political Science* 44, 142-155.

Lazear, Edward (1999), Globalization and the market for team-mates, *Economic Journal* 109, 15-40.

Lewis, Bernard (2004), *From Babel to Dragomans: Interpreting the Middle East*, Oxford: Oxford University Press.

Mamadouh, Virginie and Kaj Hofman (2001), The Language Constellation in the European Parliament, 1989-2004, Report for the European Cultural Foundation, Amsterdam.

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

Melitz, Jacques (2008), Language and foreign trade, *European Economic Review* 52, 667-699.

Muller, Siegfried (1964), *The World's Living Languages*, New York: Frederick Ungar.

Nabokov, Vladimir V. (1955), *Lolita*, Paris: Olympic Press.

Ngũgĩ wa Thiong'o (1986), *Decolonizing the Mind. The Politics of Language in African Literature*, Oxford: James Currey.

O'Connor, Kevin (2003), *The History of Baltic States*, Westport, CT: Greenwood Press.

Ota, Hitoshi (2008), Indian IT software engineers in Japan: a preliminary Analysis, Discussion Paper 168, Institute of Developing Economies.

Phillipson, Robert and Tove Skutnabb-Kangas (1995), Language rights in postcolonial Africa, in Robert Phillipson, Mart Rannut and Tove Skutnabb-Kangas, eds., *Linguistic Human Rights: Overcoming Linguistic Discrimination*, Berlin and New York: Mouton De Gruyter.

Pym, Anthony (1999), Scandalous statistics. A note on the percentages of translation in English, Source, *The Newsletter of the Literary Division of the American Translators Association* 29, 19.

Roberts, Janet (1962), Sociocultural change and communication problems, in Frank Rice, ed., *Study of the Role of the Role of Second Languages in Asia, Africa, and Latin America*, Washington, DC: Center for Applied Linguistics of The Modern Language Association of America.

Rustow, Dankwart (1967), *A World of Nations: Problems of Political Modernization*, Washington, DC: Brookings Institution.

Ruhlen, Merritt (1994), *The Origin of Language*, New York: John Wiley and Sons.

Sapir, Edward (1949), *Selected Writings of Edward Sapir in Language, Culture, and Personality*, Berkeley, CA: University of California Press.

Sapiro, Gisèle (2010), Globalization and cultural diversity in the book market: the case of literary translations in the US and in France, *Poetics* 38, 419-439.

Schwartz, Shalom (2011), National culture as value orientations: Consequences and cultural distance, Chapter 20, this volume.

Shannon, Claude (1948), A mathematical theory of communication, *Bell Systems Technical Journal* 27, 379-423 and 623-656.

Simpson, Edward (1949), Measurement of diversity, *Nature* 163, 688.

Special Eurobarometer 243 (2006), *Europeans and their languages*, Brussels: European Commission.

Spolsky, Bernard (2004), *Language Policy*, Cambridge: Cambridge University Press.

Steiner, George (1992), *After Babel*, Second Edition, Oxford: Oxford University Press.

Storti, Craig (2001), *Old World, New World: Bridging Cultural Differences*, Boston, MA: Intercultural Press.

Swadesh, Morris (1952), Lexico-statistic dating of prehistoric ethnic contacts, *Proceedings of the American Philosophical Society* 96, 121-137.

Taylor, Charles and Michael Hudson (1972), *World Handbook of Social and Political Indicators*, Ann Arbor, MI: ICSPR.

Tinbergen, Jan (1962), *Shaping the World Economy: Suggestions for an International Economic Policy*, New York: The Twentieth Century Fund.

Warnow, Tandy (1997), Mathematical approaches to comparative linguistics, *Proceedings of the National Academy of Sciences of the USA* 94, 6585-6590.

Whorf, Benjamin (1956), *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*, Cambridge, MA: MIT Press.

Table 1. Indo-European Language Tree (General)

0. Root
   ...
   1. Eurasiatic
      2. Eskimo-Aleut
      2. Chukotian
      2. Gilyak
      2. Korean-Japanese-Ainu
      2. Altaic
      2. Uralic-Yukaghirc
      2. Etruscan
      2. Indo-European (449)
         3. Germanic (53)
            4. East (1)
            4. North (11)
            4. West (41)
         3. Italic (48)
            4. Latino-Faliscan (1)
            4. Romance (47)
         3. Slavic (19)
            4. East (4)
            4. South (7)
            4. West (8)
         3. Albanian (4)
            4. Gheg (1)
            4. Tosk (3)
         3. Armenian (1)
         3. Baltic (3)
            4. Eastern (2)
            4. Western (1)
         3. Celtic (7)
            4. Insular (7)
         3. Greek (6)
            4. Attic (5)
            4. Doric (1)
         3. Indo-Iranian (308)
            4. Indo-Arian (219)
            4. Iranian (87)
            4. Unclassified (2)

Table 2. Indo-European Language Tree (Germanic languages)

0. Root
    1. Eurasiatic
        2. Indo-European (449)
            3. Germanic (53)
                4. East (1)
                    5. *Gothic* [got] (Ukraine)
                4. North (11)
                    5. East Scandinavian (6)
                        6. Danish-Swedish (6)
                            7. Danish Bokmåll (1)
                              8. **Norwegian, Bokmål** [nob] (Norway)
                            7. Danish-Riksmal (2)
                              8. Danish (2)
                                9. **Danish** [dan] (Denmark)
                                9. *Jutish* [jut] (Denmark)
                            7. Swedish (3)
                              8. *Dalecarlian* [dlc] (Sweden)
                              8. *Scanian* [scy] (Sweden)
                              8. **Swedish** [swe] (Sweden)
                    5. West Scandinavian (5)
                        6. *Faroese* [fao] (Denmark)
                        6. **Icelandic** [isl] (Iceland)
                        6. *Jamtska* [jmk] (Sweden)
                        6. **Norwegian, Nynorsk** [nno] (Norway)
                        6. *Norn* [nrn] (United Kingdom)
                4. West (41)
                    5. English (3)
                        6. **English** [eng] (United Kingdom)
                        6. *Scots* [sco] (United Kingdom)
                        6. *Yinglish* [yib] (USA)
                    5. Frisian (3)
                    5. High German (20)
                        6. German (18)
                            7. *Frankish* [frk] (Germany)
                            7. Middle-German (9)
                              8. East Middle German (3)
                                9. **German, Standard** [deu] (Germany)
                                9. *Silesian, Lower* [sli] (Poland)
                                9. *Saxon, Upper* [sxu] (Germany)
                              8. West Middle German (6)
                            7. Upper German (8)
                              8. Alemannic (4)
                              8. Bavarian-Austrian (4)
                        6. Yiddish (2)
                    5. Low Saxon-Low Franconian (15)
                        6. *Frisian, Eastern* [frs] (Germany)
                        6. Low Franconian (4)
                            7. **Afrikaans** [afr] (South Africa)
                            7. **Dutch** [nld] (Netherlands)
                            7. *Vlaams* [vls] (Belgium)
                            7. *Zeeuws* [zea] (Netherlands)
                        6. Low Saxon (10)

Table 3. Distances Between Selected Indo-European Languages
(value times 1,000)

	English	French	German	Italian	Spanish	Polish
Albanian	883	878	870	877	883	871
Bulgarian	772	791	769	769	782	369
Catalan	777	286	764	236	270	784
Czech	759	769	741	753	760	234
Danish	407	759	293	737	750	749
Dutch	392	756	162	740	742	769
English	0	764	422	753	760	761
French	764	0	756	197	291	781
German	422	756	0	735	747	781
Greek	838	843	812	822	833	837
Icelandic	454	772	409	755	763	758
Italian	753	197	735	0	212	764
Latvian	803	793	800	782	794	668
Lithuanian	784	779	776	758	770	639
Norwegian	452	770	367	754	761	762
Polish	761	781	754	764	772	0
Portuguese	760	291	753	227	126	776
Romanian	773	421	751	340	406	784
Russian	758	778	755	761	769	266
Serb-Croatian	766	772	764	755	768	320
Slovak	750	765	742	749	756	222
Slovene	751	782	733	760	772	367
Spanish	760	291	747	212	0	772
Swedish	411	756	305	741	747	763
Ukrainian	777	781	759	774	782	198

Source: Dyen et al. (1992, pp. 102-117).

Table 4. Disenfranchisement in European Languages
All Respondents (in %)

	English	German	French	Italian	Spanish	Polish	Dutch
Austria	55	1	94	95	98	100	100
Belgium	59	87	29	97	97	99	32
Bulgaria	84	94	96	99	99	100	100
Cyprus	49	98	95	99	99	100	100
Czech R.	84	81	98	100	100	98	100
Denmark	34	73	97	99	98	100	100
Estonia	75	92	100	100	100	100	100
Finland	69	95	99	100	100	100	100
France	80	95	1	95	93	100	100
Germany	62	1	92	99	98	98	100
Greece	68	94	95	98	100	100	100
Hungary	92	91	100	99	100	100	100
Ireland	1	98	91	100	99	99	100
Italy	75	96	90	3	97	100	100
Latvia	85	97	100	100	100	99	100
Lithuania	86	96	99	100	100	87	100
Luxembourg	61	12	11	95	99	100	99
Malta	32	99	95	65	99	100	100
Netherlands	23	43	81	100	97	100	1
Poland	82	90	99	99	100	2	100
Portugal	85	98	91	99	96	100	100
Romania	86	97	90	98	99	100	100
Slovak R.	83	82	99	100	100	98	100
Slovenia	59	79	98	91	99	100	100
Spain	84	98	94	99	2	100	100
Sweden	33	88	97	99	99	100	100
Un. Kingdom	1	98	91	99	98	100	100
EU	62.6	75.1	80.1	86.7	88.9	91.6	95.1

Source: Fidrmuc et al. (2007).

Table 5. Disenfranchisement in the Sequence of Optimal Language Sets
All respondents (in %)

| Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10a | 10b | 11 |
Languages	EN	GE	FR	IT	SP	PL	RO	HU	PT	CZ	GR	CZ&GR
Austria	55	0	0	0	0	0	0	0	0	0	0	0
Belgium	59	56	18	18	18	18	18	18	18	18	18	18
Bulgaria	84	81	79	79	78	78	78	78	78	77	77	77
Cyprus	49	49	49	48	48	48	48	48	48	48	0	0
Czech R.	84	69	69	69	69	67	67	66	66	0	66	0
Denmark	34	31	31	31	31	30	30	30	30	30	30	30
Estonia	75	70	70	70	70	69	69	69	69	69	69	69
Finland	69	67	67	67	67	67	67	67	67	67	67	67
France	80	77	1	0	0	0	0	0	0	0	0	0
Germany	62	1	1	1	1	1	1	1	1	1	1	1
Greece	68	64	63	63	63	63	63	63	63	63	0	0
Hungary	92	85	85	85	85	85	84	0	0	0	0	0
Ireland	1	1	1	1	1	1	1	1	1	1	1	1
Italy	75	74	69	1	1	1	1	1	1	1	1	1
Latvia	85	83	83	83	83	82	82	82	82	82	82	82
Lithuania	86	82	82	82	82	71	71	71	71	71	71	71
Luxembourg	61	8	1	1	1	1	1	1	1	1	1	1
Malta	32	31	31	31	31	31	31	31	31	31	31	31
Netherlands	23	18	18	18	18	18	18	18	18	18	18	18
Poland	82	77	76	76	76	1	1	1	1	1	1	1
Portugal	85	84	81	81	79	79	79	79	0	0	0	0
Romania	86	85	81	80	79	79	1	1	1	1	1	1
Slovak R.	83	72	72	72	72	70	70	57	57	44	57	44
Slovenia	59	50	50	45	45	45	45	45	45	45	45	45
Spain	84	84	81	80	1	1	1	1	1	1	1	1
Sweden	33	33	33	33	33	33	33	33	33	33	33	33
Un. Kingdom	1	1	1	1	1	1	1	1	1	1	1	1
EU	62.6	49.3	37.8	29.5	22.4	16.4	12.9	10.9	9.2	7.7	7.7	6.2

Source: Fidrmuc et al. (2007). One language is added to the previous ones in each column. In columns 10a, and 10b, two languages result in the same percentage reduction in disenfranchisement. In column 11, they are both added to the set. Languages are abbreviated as follows: Czech (CZ), English (EN), French (FR), German (GE), Greek (GR), Hungarian (HU), Italian (IT), Spanish (SP), Polish (PL), Portuguese (PT), Romanian (RO).

Table 6. Disenfranchisement in the Sequence of Optimal Language Sets
Young Respondents (in %)

Number	1	2	3	4	5	6	7	8	9	10	11a	11b	12
Languages	EN	FR	GE	IT	SP	PL	RO	HU	PT	CZ	GR	BG	GR&BG
Austria	40	40	1	1	1	1	1	1	1	1	1	1	1
Belgium	39	8	8	8	8	8	8	8	8	8	8	8	8
Bulgaria	56	56	53	53	53	53	52	52	52	52	51	3	3
Cyprus	18	18	18	18	18	18	17	17	17	17	1	17	1
Czech R.	64	64	52	52	52	50	50	50	50	0	0	0	0
Denmark	9	8	6	6	6	5	5	5	5	5	5	5	5
Estonia	33	33	29	29	29	29	29	29	29	29	29	29	29
Finland	29	29	29	29	29	29	29	29	29	29	29	29	29
France	67	0	0	0	0	0	0	0	0	0	0	0	0
Germany	38	38	0	0	0	0	0	0	0	0	0	0	0
Greece	40	40	36	36	36	36	36	36	36	36	0	36	0
Hungary	76	76	64	64	64	64	63	0	0	0	0	0	0
Ireland	2	2	2	2	2	2	2	2	2	1	1	1	1
Italy	54	51	49	1	1	1	1	1	1	1	1	1	1
Latvia	55	55	54	54	54	53	53	53	53	53	53	53	53
Lithuania	49	49	45	45	45	36	36	36	36	36	36	36	36
Luxembourg	50	2	1	1	1	1	1	1	0	0	0	0	0
Malta	10	10	10	10	10	10	10	10	10	10	10	10	10
Netherlands	11	11	10	10	10	10	10	10	10	10	10	10	10
Poland	57	56	50	49	49	1	0	0	0	0	0	0	0
Portugal	62	60	60	60	59	59	59	59	0	0	0	0	0
Romania	68	62	62	61	59	59	2	0	0	0	0	0	0
Slovak Rep.	57	57	39	39	39	38	38	31	31	23	23	23	23
Slovenia	22	22	17	15	15	15	15	15	15	15	15	15	15
Spain	65	63	63	62	1	1	1	1	1	1	1	1	1
Sweden	5	5	5	5	5	5	5	5	5	5	5	5	5
UK	2	2	2	2	2	1	1	1	1	1	1	1	1
EU	44.6	34.5	25.8	19.9	14.4	10.4	7.8	6.3	5.1	3.9	3.1	3.1	2.3

Source: Fidrmuc, Ginsburgh and Weber (2007). One language is added to the previous ones in each column. In columns 10a, and 10b, two languages result in the same percentage reduction in disenfranchisement. In column 11, they are both added to the set. Languages are abbreviated as follows: Bulgarian (BG), Czech (CZ), English (EN), French (FR), German (GE), Greek (GR), Hungarian (HU), Italian (IT), Spanish (SP), Polish (PL), Portuguese (PT), Romanian (RO).