



3118

REPRINT

Victor Ginsburgh and Shlomo Weber

The economics of language

Journal of Economic Literature, , 348-404, 2020



CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

[https://uclouvain.be/en/research-institutes/
lidam/core/reprints.html](https://uclouvain.be/en/research-institutes/lidam/core/reprints.html)

The Economics of Language[†]

VICTOR GINSBURGH AND SHLOMO WEBER*

This paper brings together methodological, theoretical, and empirical analysis into the framework of linguistic diversity. It reflects both historical and contemporary research by economists and other social scientists on the impact of language on economic outcomes and public policies. We examine whether and how language influences human thinking (including emotions) and behavior, and analyze the effects of linguistic distances on trade, migrations, financial markets, language learning, and its returns. The quantitative foundations of linguistic diversity, which rely on group identification, linguistic distances as well as fractionalization, polarization, and disenfranchisement indices are discussed in terms of their empirical challenges and uses. We conclude with an analysis of linguistic policies and examine the trade-offs between the development of labor markets and the social costs that they generate in various countries. (JEL N30, Z13)

1. Introduction

The role of languages in human development can hardly be ignored. Language makes information operational. One can argue that, as a social technology, it defines the human species. Language affects cultural identity, effective communication in business and international trade, employment

opportunities, and to some extent our economic and social behavior.

The origin of the economics of language as a separate discipline is usually credited to Marschak (1965), who explicitly introduced economic concepts such as costs and benefits into linguistic analysis. It is possible that Marschak's own background, which included the command of a dozen of languages and the service as Secretary of Labor of the

* Ginsburgh: ECARES, Université Libre de Bruxelles, and CORE, Université catholique de Louvain. Weber: New Economic School. We are grateful to the Editor of the journal, Steven Durlauf, for his tremendous patience, encouragement and useful suggestions, and three anonymous referees who have followed our various missteps and made us to produce a better paper. We thank Kristina Butaeva, Christine Defrise, Edward Gibson, Jacques Melitz and Farid Toubal for their thoughtful comments, and Sheila Weyers who corrected countless versions of the paper and gave us invaluable advice. The authors are

grateful to the Center for Diversity and Social Interactions at the New Economic School for its support. The second author is grateful to the Domodedovo group of companies for financial support of the NES Center for the Study of Diversity and Social Interactions in 2018–19. This paper benefited from grant No.14.U04.31.0002 of the Ministry of Education and Science of the Russian Federation administered through NES CSDSI in 2013–17.

[†] Go to <https://doi.org/10.1257/jel.20191316> to visit the article page and view author disclosure statement(s).

multilingual Terek Soviet Republic in North Caucasus in 1919 sustained his interest in the economics of language. But studying diversity to understand human experience has a long tradition that goes back to the philosopher and linguist, Von Humboldt (1988 [1836]). Somewhat later, John Mills and Karl Marx pointed out that culture and cultural diversity are central to economic interactions.¹

The contributions by Fishman (1968), Pool (1972), Hočevár (1975), Bretton (1976), and certainly many others, stressed the importance of language on social, political, and economic outcomes to political scientists, sociologists, anthropologists, psychologists, and sociolinguists. Still, *The Economics of Language*, edited by Lamberton (2002), contains 20 papers of which only three were written before 1980: Marschak's (1965) five-page "Economics of Language," published in *Behavioral Science*; Pelikán's (1969) on "Language as a Limiting Factor for Centralization," published in the *American Economic Review*; and Thompson (1977) on "Information Technology and the Linguistic Constraint," published in *Telecommunications Policy*.

Linguist Joseph Greenberg was obviously aware of the possible connections with economics. He is the one who suggested constructing diversity indices with and without linguistic distances that could be "[correlated] with political, economic, geographic, historic, and other non-linguistic factors" (Greenberg 1956, p. 109). This idea was picked up much later by Mauro (1995), Easterly and Levine (1997), Alesina, Baqir, and Easterly (1999), and many others. A large number of articles soon followed that connected language, economics, and business. These touched upon an even larger number of topics, including trade, migrations,

consumer choices, earnings, language acquisition, choices of multinational corporations in various linguistic environments, and most importantly, the possible effects on growth, institutional quality, redistribution, regional and national development.

This paper provides a consistent and comprehensive overview of linguistic diversity.² It identifies common traits of languages as a part of culture and, using economic tools, addresses specific characteristics of diversity and its consequences. We take seriously Bretton's (1976) sentence that language is "the most explosive issue universally and over time, [t]his mainly because language alone, unlike all other concerns associated with nationalism and ethnocentrism, is so closely tied to the individual self."

In section 2, we start with a brief description of the long march that transformed the (possibly) unilingual "people" into today's multilingual universe and its 7,000 spoken languages. The scope of the existing language menu immediately brings up the question of incentives for learning foreign or local languages. Historically, the main incentive was trade. Transportation, either by walking and later by boats or donkeys, required crossing various cities and ports, where communication and negotiations without the command of local languages must have been problematic, inefficient, and costly.

Learning foreign languages was (and still is) also driven by migrations and wars. Some people were forced to leave their country and had to learn the language of their new homeland on the spot, while voluntary migrants were and are, to some degree, able to choose the country of their destination and possibly learn its language before leaving. The emerging multilingual world unleashed two different forces: (i) the need

¹This was picked up in a theoretical paper by Lazear (1999).

²See also Guiso et al. (2006) for an early survey.

for translation and interpretation,³ as well as the shifting *linguae francae* (Greek and Latin in the ancient world; Arabic; Latin again, as a scholarly language during the Middle Ages and Renaissance in Europe; French, for diplomacy; German, for science; and finally, English), and (ii) the interplay between two tendencies of preserving and strengthening local languages versus uniformity and standardization.

Von Humboldt's ideas were further developed during the twentieth century by Boas (1940), Mandelbaum (1949), and Whorf (1956), whose contributions were synthesized and came to be called the *Sapir-Whorf hypothesis*: language and culture are interdependent, and language influences behavior and decisions. This statement was challenged in 1955 by Noam Chomsky (published 1975), who suggested that the same grammar was wired into all human's brains and could not influence behavior in different ways. This led to confrontations between linguists. We deal with this problem at some length, trying to clarify whether, and if so, how, language may influence human behavior. Though this may look far from being an economic issue, economists recently found it interesting to analyze whether grammars had an influence on economic behavior. They also started to run experiments as do psycholinguists.

We end the section by discussing linguistic emotions, which emerge when individuals find themselves in a linguistic environment different from the one they are used to. This could happen in two ways. One is migration, voluntary or not, when people have to live, work, and write in a language different from their native tongue. We describe the pain, anguish, and insecurity in some cases and the elation and satisfaction in others. The other way is due to changes in the linguistic

environment resulting from government policies. These may include a shift of official languages and/or the languages of instruction in schools, and may create linguistic disenfranchisement for several population groups if their linguistic rights are restricted or denied.

Section 3, on the multilingual world, introduces the quantitative foundations of linguistic diversity, which rely on group identification, linguistic proximities, and the construction of fractionalization, polarization, and disenfranchisement indices. This rules out the notion of a "representative citizen" and forces researchers to cope with patterns of linguistic proficiency (who speaks what), varying aptitudes in learning other languages (some are more skilled or motivated), and costs and returns to learning foreign languages.

We first discuss the difficulties encountered when dividing a society into distinct ethno-linguistic groups. As Lewis (2009) points out, "not all scholars share the same criteria for distinguishing a language from a dialect." Is Venetian an Italian dialect or a language, and should we consider Venetians as being distinct from Romans or Neapolitans? To mitigate the effects of possible ambiguities or errors, it is useful to introduce the notion of proximity, or distance, between languages and linguistic groups. We describe a wide variety of such distances. They include lexicostatistical distances (based on similarities and supposed common roots of words in the vocabularies of various languages), Levenshtein distances (the distance between the word in one language and the same meaning in the other one is determined by the number of insertions, deletions, or substitutions of single characters required to convert one word into the other), cladistic distances (based on linguistic trees which take into account many descriptors), phonetic distances, and others.

³See for example Lewis (2004) in the context of the Middle East.

Next, we show how one can formalize fractionalization, polarization, and disenfranchisement indices. Fractionalization indices are based on the division of a society into distinct ethno-linguistic groups. Polarization indices take into account the strength of individuals' identification with their own group and their alienation toward other groups. Disenfranchisement indices generalize fractionalization in a different direction that accounts for the possibility that agents may be multilingual. These indices basically count the number of agents who speak each language in their repertoire (accounting or not for linguistic distances) and allow judging, for instance, which languages to retain as official languages.

Both tools, linguistic distances and indices, combined or not with linguistic distances, have been used by economists in formulating and estimating equations to explain international trade, language learning, migrations, or translations, as well as economic growth, regional development, institutional quality, redistribution, and conflicts. Examples will be discussed.

Section 4 analyzes some typical linguistic policies in multilingual settings and the possible inefficiency that may arise from using multiple languages in a given society. One of the most often used remedies to inefficiency is standardization, the foundations of which were laid down in Weber's (1968 [1910]) theory of rationalization. In our context, standardization implies using a small number of languages (often one) for official purposes in government, the press, education and trade.

While advantages of standardization are often evident, an inevitable outcome is disenfranchisement and the threat to the linguistic identity of groups. Indeed, standardizing may adversely impact, and sometimes even deny, the linguistic rights and sensibilities of various population groups. To address this issue, a certain number of drastic linguistic education policy shifts in various countries

(Spain, Rwanda, Morocco, South Africa, Kazakhstan, Latvia, India, and the United States) are briefly reported and analyzed. These changes were implemented as a result of a political change (Latvia, Kazakhstan, Spain), public demands (Morocco, India, and the United States) and changes in government policies for other reasons (South Africa before the abolition of apartheid, Rwanda in response to regional economic challenges). Switching the language of instruction to a native language is an important symbol of independence and national sovereignty, but the response of labor markets may generate substantial social costs.

To understand the specific effects of balancing efficiency and the social cost of linguistic disenfranchisement, three-language policies experimented with in India, Nigeria, the Soviet Union, and Kazakhstan are also reported. They amount to requiring all individuals to learn not only their own native language, but also the language of another region, as well as a world language, most often English. The reform failed in India and Nigeria, but has a fair chance to succeed in Kazakhstan since it benefits from the government's commitment, which is recognized by citizens.

Section 5 concludes the paper by returning to issues discussed and related to identifying groups, econometric problems met in the empirical work illustrated in sections 2.2, 3.3, and 3.4. We also suggest further research possibilities.

2. *Language and the Behavior of Economic Agents*

Whether there were several "speaking" human groups or only one, *Homo sapiens*, from whom we all descend, came from Africa—or as has been recently hypothesized, from the Middle East (HersHKovitz et al. 2010, 2016)—with one language. This number of languages increased over time,

as some subgroups of the first population(s) started their long march to other regions, and languages evolved, changed, and died over the last 150,000–400,000 years. Or perhaps, this happened because God dispersed the builders of the Tower of Babel from Shinar, Mesopotamia, which

resulted in mass lay-offs of workers, but none of them complained. Those, who like me, worked on a foreign construction site, surrounded by workers with different languages and colors, know how difficult it is to start objecting, even if everyone roughly speaks a common language, using improbable accents. I try to imagine what would have happened as we were being fired, if we, Turks, Yugoslavs, Algerians, Portuguese and one Italian could not understand each other; but we had a first draft of a common language in a foreign country which made it possible for us to protest, while this was impossible in the valley of Shinar. Therefore, the people left, and dispersed. And of course, they forgot their words, but also the high degree of specialization reached in their work and never again tried to build heaven (De Luca 1997).⁴

The quotation nicely summarizes the long march from one to the 7,097 (Eberhard, Simons, Fennig 2016) languages that still exist today. We all get born into (usually) one language, but we also face economic and cultural incentives to learn other languages to succeed in a globalized world (section 2.1). This is the reason to revert to a smaller number of languages and be able to communicate with each other.

Meanwhile, our native language probably also conditions part of our later behavior (section 2.2) and we certainly get emotional, if not infuriated, when it is mistreated (section 2.3).

2.1 *Economics and Language Learning*

Learning foreign languages (or refraining from doing so) is the result of several

economic incentives. The main is obviously trade. Even in the ancient world that is more familiar to us (the Near East, including Mesopotamia, Egypt, Iran, the Levant, Greece, and Rome), people were forced to communicate. Moore and Lewis (1999, p. 271, quoted by Holden 2016, p. 295) suggest that “their international business achievements were probably more impressive than those of modern business people.” And indeed, traveling from the Iranian border to Turkey through Mesopotamia⁵ probably took several weeks, needed transfers of goods between ships and donkeys, long walks in the desert, and speaking in a trustful way with a host of various and often hostile tribes. This required either interpreters or an ability to communicate in common languages. During the Middle Ages, traders would travel between China and Italy along the Silk Road. Somewhat later, ships sailed between Europe and the recently discovered Americas, as well as between Europe, India, China, and Japan, facing a long voyage since they had to make their way around Africa, where they often landed.

Another reason for learning is driven by migrations and wars. Voluntary migrants or people forced into exile for political, racial, or economic reasons have to learn the language of their new country. Some do so before leaving, and choose the country of destination accordingly. Others migrate and learn the language on the spot. This is not

⁵“Transportation was mainly ensured by boats, built in shipyards to be found in almost every city located along a river. Donkeys were used on land, when rivers were impossible to navigate, and, obviously, to cross deserts. Two main roads connected the Southern delta and the West. The first led from Sippar to Mari and then, crossing the desert, it reached Qatna and the Phenician ports on the Mediterranean sea. Though longer, the second one was safer as it did not go through a desert; it also started in Sippar, followed the Tigris to Nineveh and then turned west towards the sea, through Nagar and Harran. Because of the mountains, communications with the Eastern part of the country were much more problematic.” See Bossuyt, Broze, and Ginsburgh (2001, p. 375).

⁴Our translation into English from the French translation.

fully novel either. Phoenicians traveled and settled in several regions surrounding the Mediterranean Sea. Greeks moved to Asia Minor during the Trojan War and, according to Homer, it took Ulysses a couple of years to travel back home, visiting island after island, where he even had to speak the language of the mermaids. The Macedonian King Alexander the Great waged war to the East, as far as Persia and India, and to the southwestern border of today's Egypt at the oasis of Siwa. Greeks moved to the Roman Empire (Southern Italy and Sicily). Romans occupied Palestine and Constantinople. Christian crusaders traveled to "save" Jerusalem from the Muslims. Jews moved from Palestine to the whole Middle East, as well as to Spain and Portugal, from which they were expelled at the end of the fifteenth century and had to settle in Mediterranean countries, the Middle East (Iraq, Iran, and Syria), and Western Europe. Arabs invaded North Africa in the seventh century, and from there came to Spain in the early eighth century. All this forced or fostered invaders, as well as locals and migrants, to learn languages.

Interpretation between languages was always needed, but with the birth of writing in Mesopotamia, people also started translating. In the third century BC, the Jewish Torah was translated from Aramaic into Greek by seventy-two Jewish scholars working in the Ptolemaic Library of Alexandria, and this translation became the most important text from which the Old Testament traveled into other languages.⁶

On the other hand, there was also need for *linguae francae*. After the death of Muhammad in 632 AD, the Arab invasions in parts of the Middle East and North Africa turned Arabic into a lingua franca and it

still is so today. Latin was the lingua franca of scholars in the dark Middle Ages and the Renaissance. Isaac Newton's *Principia*, Baruch Spinoza's *Ethics*, Blaise Pascal's *Generatio conisectionum*, Carl Friedrich Gauss's *Disquisitiones Arithmeticae*, Galileo's *De Motu Antiquiora*, René Descartes's *Principia Philosophiae*, Pierre de Fermat's *Methodus ad disquirendam maximam et minimam* and *De tangentibus linearum curvarum*, Paracelsus's *Opus Chirurgicum*, and many other scientific works were written in Latin, since their authors were assured that other scientists would be able to read them.

Somewhat later, French became the language used by diplomats, and German the one used by scientists. Crystal (1999, p. 105) estimates that English is nowadays used by 1.5 billion people, while 400 million only speak it as their mother tongue.

2.2 *Language, Culture, Behavior, and the Contemporary Interpretation of the Sapir-Whorf Hypothesis*⁷

Wilhelm von Humboldt (1888, [1836]) was probably the first to emphasize that language was more than a means of communication. He suggested that it may also express the inner life of its speakers. A century later, Mandelbaum (1949) and Whorf (1956), influenced by Von Humboldt and anthropologist Franz Boas's (1940) *cultural relativism*, went a step further and hypothesized that language and culture were interdependent, and more importantly, that the structure of the language we use influences our way of

⁶Today the Bible is translated into over three thousand languages. Another four thousand languages are still waiting in line.

⁷Most of the work described in this section was carried out by (ethno)linguists, neuroscientists, and psychologists who started as early as 1940, and keep going. Economists came much later. The first paper that was close to what linguists also tried to understand was written by economists Roth et al. (1991). Then came papers by Chen (2013), Galor and Özak (2016), Galor, Özak, and Sarid (2016, 2018), Heblich, Lameli, and Riener (2015), Mavisakalyan (2015), Mavisakalyan and Weber (2018), Sutter et al. (2015), and Tabellini (2008), whose work is cited in this section.

thinking, observing, behaving and, eventually, making decisions. Here is what Sapir mentioned in one of his talks in 1929:⁸

The fact of the matter is that the “real world” is to a large extent unconsciously built up on the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are different worlds, not merely the same world with different labels attached ... Even comparatively simple acts of perception are very much more at the mercy of the social patterns called words than we might suppose. If one draws some dozen lines, for instance, of different shapes, one perceives them as divisible into such categories as “straight,” “crooked,” “curved,” “zigzag” because of the classificatory suggestiveness of the linguistic terms themselves. We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation.

Whorf (1956, p. 221) suggested that “users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world.”

The Sapir-Whorf (SW) hypothesis takes two forms. The *strong* form postulates that our native language *determines* our thinking and view of the world: We are “prisoners of our language” (Kramsch 1998, p. 13). This form, which was met with resistance by the scientific community fearing that it may introduce *relativism* in scientific discoveries, is no longer considered significant by most linguists. The *weak* form maintains only that language may *influence* our thinking, but may also constrain it.

In the early 1950s, Chomsky (1975) challenged the SW hypothesis, arguing that grammar was mostly inborn, and not acquired. He was quickly followed by many linguists. His deterministic *universal grammar* implies that all languages share the same underlying structure and potential to explain “the striking similarity of language learning in children all over the world, [which] captured the imagination of a generation of scholars” (Li and Gleitman 2002, p. 266).

One of Chomsky’s well-known followers, Pinker (1994, p. 232), claims that “according to Chomsky, a visiting Martian scientist would surely conclude that aside from their mutually unintelligible vocabularies, Earthlings speak a single language.” If this were true “there is less scope for the large differences among languages that the more extreme linguistic relativists had imagined” (Swyer, Baghrarian, and Carter 2014, p. 7).

Chomsky’s universal grammar is a set of principles that are common to all languages, though there may be parametric variations. The definite article, for instance, comes in front of a noun in French and English; it is attached to the end of a noun in Swedish and does not exist in Russian. In some sense, this is comparable to the principles of every spider producing its web. Spiders do not need to learn how to weave, they use their instinct or wired grammar, though there are some variations in building between say, Brazilian and European spiders. Children need to hear their parents or other people speaking, since otherwise they would not speak, though they may already have a grammar wired in their brain. The young spider needs no other spider to learn, but this is not necessarily the case for some mammals that are able to acquire nongenetically transmitted actions by looking at their own kind.

Nowadays, linguists support a weaker form of SW hypothesis. Dediu and Ladd (2007) accept that language is related to genes, but that the genetic determinant is not unique.

⁸Sapir (1949), cited by Leavitt (2011, pp. 138–39).

Evans and Levinson (2009) suggest that existing languages result from a combination of historical accidents and other influences, and not from linguistic universals. Leavitt (2011, p. 143) argues that one of the main contributions of relativism is that it “distinguished clearly between what is *possible* to think for speakers of any language, and what people *habitually* think, which may be strongly influenced by their language.”

This weaker SW version posits that while reflecting cultural preoccupations and constraining our way of thinking, language is also a carrier of culture: “There are cultural differences in the semantic associations evoked by seemingly common concepts” (Kramsch 1998, p. 13). The consequences 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” (Kramsch 1998, p. 12) or, in somewhat different terms, that “meaning is not fully encapsulated in lexicon and grammar” (Gumperz and Levinson 1991, p. 614). A new group of cognitivist linguists also started running experiments, the results of which are very close to the SW hypothesis: “The languages we speak profoundly shape the way we think, the way we see the world, the way we live our lives” (Boroditsky 2009).

The confirmation or rejection of the SW or *relativist* hypothesis is mostly examined on a case by case basis. But this also sets a strong limitation to accepting or rejecting the SW relativist or the Chomsky *universalist* hypothesis, since it seems impossible to prove or disprove their generality. It may well be that in some cases, one hypothesis is shown to be satisfied, while in others it is not, and that the finding (right or wrong) is often later turned down or rediscussed. But this does not mean that either of two hypotheses can be generalized. Many experiments have, for instance, been conducted with color terminologies,

or spatial expressions in different languages, but they do not seem to lead to generalizations. Moreover, it is not always obvious that some characteristics of language “cause” or are merely “correlated with” some economic facts and, though he is very Chomskian, Fabb (2016, p. 45) notes that causation may even go in the other direction: “Cultural values of a group of people (ancestors) caused linguistic forms to come into existence perhaps in a situation where the linguistic system offers alternatives, and selection of the chosen alternative is biased by cultural values,” which implies that culture and language may at best be correlated. This is even weaker than what Whorf claimed: “I should be the last to pretend that there is anything so definite as ‘a correlation’ between culture and language” (cited by Leavitt 2001, p. 142).

The debate is still going on, even in a case that looks quite simple—a reason for which we start with it. This case deals with colors such as red, green, or blue, which exist in all languages—though the number of terms used may be different. They are grounded on well-established physical concepts and therefore, there should be no linguistic relativism. This allows running reasonably simple experiments to analyze the debate between relativism and universalism. Such an analysis was carried out in a systematic way by Berlin and Kay (1969).

2.2.1 *Language and Colors*

The human eye can distinguish several millions of color differences, since the range is made of a three dimensional continuum defined by the wavelength of reflected light (called hue), the quantity of reflected light (saturation), and the dilution of the color with white (brightness) or its difference from black (for instance, light or dark blue). Though visible colors constitute a continuum, we categorize them using a small number of basic colors (red, green, yellow, blue, ...). Basic physics and optics consider visible

colors a continuum of wavelengths located between 400 and 700 nanometers (nms). Basic colors constitute intervals or bands in this continuum: we “see red” between 700 and 635 nms, this turns to orange between 635 and 590 nms, then to yellow, green, cyan, blue and violet or purple between 450 and 400 nms. Colors are invisible to the eye above 700 and below 400 nms. Chromatic colors (thus, excluding black and white⁹) can be described by their dominant wavelength, and one can produce a correspondence between the dominant wavelength and the term that describes it in different languages. Colors are thus much easier to compare across languages than other words or concepts used, for instance, to represent *time*, since their optic properties should be wired in our brain.

The number of terms used to denominate basic colors varies across languages. English has 11 words that everyone knows—black, white, red, green, yellow, blue, brown, orange, pink, purple, grey—while the Bolivian Amazonian language, Tsimane, has only three—black, white, and red. The names of colors appear in a relatively fixed order across cultures and languages (Berlin and Kay 1969, Loreto, Mukherjee, and Tria 2012). Some populations have words for black and white only. A population that has a word for red, must also have words for black and white; the terms for brown or purple appear only after the distinction between green and blue.

In some languages, there is only one term for green and blue. This is so in one of the dialects of Amazigh, the language of Berbers in North Africa. In the Bantu language, Swahili, green and blue are described respectively by the color of grass *rangi ya mayani*, and the color of the sky *rangi ya*

angwa. In Pashto, an Indo-Iranian language spoken in Afghanistan and Pakistan, the term *shin* is used for both green and blue, but when it gets difficult to understand to which color *shin* relates, one may add “plants” or “sky” like in Swahili. Monolingual Zuni speakers and English speakers have the same color categories with the exception of yellow and orange for which Zunis have only one term, while English speakers have two (Lenneberg and Roberts 1956, p. 30). In contrast, Greek makes a difference between a darker shade of blue called *ble* and a light one called *ghalazio*. According to Thierry et al. (2009, p. 4567), this leads to “faster perceptual discrimination of these colors in native speakers of Greek than in native speakers of English ... which establishes an implicit effect of language specific terminology on human color perception.”

Brown and Lenneberg (1954) and Lenneberg and Roberts (1956) were among the first to investigate the idea that lexical differences result from cognitive differences, using experiments on colors. This idea was pursued and extended by Berlin and Kay (1969), who found that colors are hard-wired in our cognitive system, rejecting thus the *relativist* hypothesis. The debate switched from one extreme to the other and can be summarized by the following two questions formulated by Kay and Regier (2006, p. 52):

- (1) Is color naming across languages largely a matter of arbitrary linguistic conventions?
- (2) Do cross-language differences in color naming cause corresponding differences in color cognition?

In the standard rhetoric of the debate, a *relativist* argues that both answers are *Yes*, and a *universalist*, that both are *No*.

This generated hundreds of papers choosing one of the two directions or even creating the third: Gibson et al. (2017), for instance, show that the problem has much less to do with perception than with usefulness of a

⁹Formally, black and white are not colors; they result from the eye mixing wavelengths, and so does pink, for instance. To simplify, we will nevertheless call them colors.

color: “It is now established that native language affects one’s perception of the world. However, it is unknown whether this effect is merely driven by conscious, language-based evaluation of the environment or whether it reflects fundamental differences in perceptual processing between individuals speaking different languages.” And three years after Kay and Regier (2006, p. 52) formulated the two questions quoted above, in a paper subtitled “Whorf was half right,” Regier and Kay (2009, p. 439) changed the order of their names and responded:

With respect to question (1), color names do influence color perception—but primarily in the right visual field, and less so in the left. With respect to question (2), color naming across languages does reflect universal tendencies, as shown in earlier work—but also some degree of local linguistic convention.

If the answer to a relatively easy question with a small number of basic colors is that complicated, it should come as no surprise that answers to other more difficult questions are not obvious at all. The relation between language and behavior is far from being settled, and, as we shall see later, the linguists still argue about that.

We now concentrate on a couple of cases: time and its relation to space, gender of nouns, and politeness distinctions. Even though Fabb (2016, pp. 42–55) qualifies them as surface properties that “may not represent what people actually know when they know a language,” they have been studied by linguists, psychologists, cognitive scientists, and economists.¹⁰ This research may have been triggered by Gumperz and Levinson’s (1991, p. 622) claim that “the data against which to measure the linguistic relativity hypothesis still hardly exist,” or by Casasanto’s (2008) paper “Who’s afraid of the big bad Whorf,” who recommends

using “some sort of extralinguistic data to test” whether some form of the Sapir–Whorf hypothesis holds, since “otherwise, the only evidence that people who *talk* differently also *think* differently is that they *talk* differently” (Casasanto 2008, p. 67).

Before turning to this, it is worth pointing to a paper in the economics literature on whether it was language or culture that influenced economic decisions. It describes and analyzes an experiment run by Roth et al. (1991), probably without any intention to verify which linguistic theory was right or wrong. The authors ran two- and multiperson bargaining experiments (an ultimatum game and one-period market environment) in four different cultural and linguistic environments (Israel, Japan, the United States, and Yugoslavia). They concluded that the different behaviors in the four countries are probably attributable to cultural rather than linguistic dissimilarities. The question of whether language is or is not one of the facets of culture has obviously not lost its attraction even today.

2.2.2 *The Representation of Time*

Whorf studied the language spoken by Hopi Native Americans and found that it “contains no words, grammatical forms, constructions or expressions that refer directly to what we call time, or to past, present or future, or to enduring or lasting” (Whorf 1956, p. 57). This concept fell into disrepute, and the controversy was almost put to an end by Malotki (1983), who refuted Whorf’s views.¹¹ In a survey article of Malotki’s book, linguist Comrie (1984), an authority on the typology of tenses, concludes that “Malotki has performed the inestimable service of debunking one of the persistent myths of twentieth-century linguistics.” Not fully

¹⁰See Fabb (2016) for other examples.

¹¹Whorf was also criticized by claims that he did not know the Hopi language well enough, and that he gave no citations of Hopi sentences and examples in his book.

though, since cognitive scientists such as Boroditsky (2001), Boroditsky, Fuhrman, and McCormick (2011), and Casasanto and Boroditsky (2008) accept the idea that conceptions of time are different across languages. In particular, Mandarin speakers talk about time vertically (time has an upward or downward trajectory) and use vertical terms to describe it, while English speakers see it horizontally (behind and ahead). In Malagasy (Madagascar's native language), the "future comes from behind" (Dahl 1995). Kuuk Thaayorre, an Australian Aboriginal language, seems to represent time as moving from east to west (Boroditsky and Gaby 2010, Gaby 2012).

The paper by Boroditsky (2001), which claims that "English and Mandarin talk about time differently," was attacked by January and Kako (2007, p. 417) who experienced "six unsuccessful attempts at replicating [Boroditsky's finding]. In sum, we conclude that Boroditsky provides no support for the Whorfian hypothesis."

Economist Chen (2013) takes a different path in analyzing the grammar of languages, and, in particular, how languages mark the future (FTR for future time reference). He distinguishes the group of languages that *have to* mark the future in speaking, so-called strong-FTR languages, and those that *do not have to* mark it, called weak-FTR languages. Paraphrasing somewhat Chen, assume that you want to say "tomorrow it will rain" in French. Since the future tense of "to rain," "pleuvra" exists, you should say "demain, il pleuvra," and it would be strange if, instead you said "demain il pleut," where "pleut" is the present tense. In English, you have to add "will:" "tomorrow it will rain," and you can hardly say "tomorrow it rains." French and English are strong-FTR languages. In German, you can say "morgen regnet es," using the same form of the verb "regnen" (to rain) for tomorrow (morgen) as you would for today "heute regnet es." But you can also

use other forms such as "morgen wird es regnen" which can be translated as "tomorrow it is going to rain." But German is considered to be a weak-FTR language.

Using regression analysis, Chen shows that strong- and weak-FTR native speakers differ in the way they account for the future in their savings (and other types of) behavior. He finds that native speakers of weak-FTR languages "appear more future oriented in numerous monetary and non-monetary behaviors. [They] are 31 percent more likely to have saved in any given year, have accumulated 39 percent more wealth by retirement, are 24 percent less likely to smoke, are 29 percent more likely to be physically active, and are 31 percent less likely to be medically obese." Chen also finds that countries with weak-FTR native languages save some six percent more of their GDP than strong-FTR countries. The reason for this counterintuitive result is that strong-FTR languages make the future more distant, and thus less important (Dahl 2003). Chen (2013, p. 721) is very careful in discussing whether the grammar of the language causes the behavior of agents: it may be that "language is not *causing* but rather *reflecting* deeper differences that drive savings behavior," but he still claims that "much of the measured effects I find are causal," though he does not fully exclude that they may also reflect cultural values.

His paper came under strong attacks by linguists including Dahl, who used to be a convinced anti-Whorfian, but became more lenient in admitting some degree of Whorfianism (Dahl 2003). Dahl also happens to be at the source of the description of tenses (Dahl and Velupillai 2013) used by Chen. Dahl has three main worries:

- (i) there is not only one non-past form (such as the future tense) that refers to the future, such as "tomorrow, it is going to rain" instead of "it will rain," which makes strong- and weak-FTR distinction less dichotomous,

(ii) some strong-FTR languages such as French and Spanish also use the present tense for something we *intend* to do in the immediate future, and we probably more often make intention-based than prediction-based statements, which again blurs the distinction between strong- and weak-FTR,

(iii) there are many other ways to refer to the future than the existence or non-existence of the future tense.

Sutter et al. (2015) ran a controlled experiment to study the relationship between intertemporal choices among 860 six-to-eleven-year-old children raised speaking German, a weak-FTR or Italian, a strong-FTR language in a bilingual town in Northern Italy. The results confirm Chen's finding: German-speaking children are 46 percent more likely to delay gratification (toys, cookies, etc.) than their Italian-speaking peers.

Note that Galor and Özak (2016) and Galor, Özak, and Sarid (2016) offer an alternative explanation to Chen and Sutter. They ran probit models regressing a dummy variable defined by the presence or absence of the future tense on geographic characteristics (latitude, elevation, ruggedness, coast length), weather conditions (rain, temperature, and pre-1500 AD unproductive period), pre-1500 AD crop returns, as well as regional fixed effects, and find that a one standard deviation increase in crop returns leads to a 12 to 23 percentage point decline in the probability of having the strong future tense. Given their specification, Galor and his coauthors can obviously exclude reverse causality. More abundant crops are thus a reason for "caring less" about the future.

2.2.3 *The Gender of Nouns*

The influence of grammatical gender for nonbiologically gendered nouns has also generated some research. Though genders of nouns hardly exist in English,¹² they do

in many other languages, such as German where "Tisch" for table is masculine, while in French "table" is feminine. The question is whether this has consequences on cognition.

Boroditsky, Schmidt, and Phillips (2003) summarize the findings by psychologists that indeed gender of nouns affects thinking, but not very deeply. For instance, people who speak two languages in which genders are different for a common noun quickly realize that the assignment of genders is arbitrary, and has no deep meaning. Still, Boroditsky (2009), whose tests were administered in English, shows that the descriptions given by German and Spanish speakers about a word such as "key," which is masculine in German and feminine in Spanish, leads Germans to think of "hard," "heavy," "metal," or "useful," while Spanish speakers refer to "golden," "lovely," "shiny," and "tiny." Boroditsky also notes that abstract entities such as death, sin, victory, or time are represented as men or women in paintings or sculpture according to the grammatical gender the word takes in each language.

Economists have also entertained this idea. Mavisakalyan (2015) shows that in countries where the majority language is gender-intensive, women participate less in the labor force, and speakers of gender-intensive languages are more likely to think that women should not have equal access to jobs.¹³

In their recent paper, Galor, Özak, and Sarid (2018) again show that it may be that the agricultural determinants of gender gaps in agricultural productivity fostered the existence of sex-based grammatical genders.

even maritime authorities dislike the idea. See however, the very funny entry "Why is a ship she?" at <http://www.glossophilia.org/?p=1411>, last consulted June 1, 2017.

¹³See also Behle, Gorges, and Markowsky (2017).

¹²With some exceptions such as *ship*, which is feminine, though nowadays, the *Chicago Manual of Style* and

2.2.4 *Politeness Distinctions*

Tabellini (2008) distinguishes languages according to two rules that govern the use of pronouns :

- (i) the use of first and second person pronouns in conversations: In some languages, Italian, for instance, the use of first and second pronouns in conversations can be dropped, in English this is rare;
- (ii) in French you can address someone with “tu” if you know her or him well; otherwise, you would use “vous;” this is also the case in Spanish with the familiar “tu” and the more severe “usted.” Such differences do not exist in English unless you meet the Queen.

Tabellini considers these grammatical rules to be “deep” enough to be used to instrument cultural traits such as trust and respect, which are themselves used in the second-stage estimation to “explain” the quality of institutions.

Again, Galor, Özak, and Sarid (2018) suggest that the ecological origins of hierarchical societies may have contributed to the existence of politeness distinctions, and could well be the “deep” variables that serve explaining politeness distinctions.

2.2.5 *Additional Issues*¹⁴

Heblich, Lameli, and Riener (2015) find that, in Germany, regional accents affect individuals’ interactions. People tend to cooperate if they come from the same region and have the same accent. However they are more inclined to compete with those who speak with another accent. One must add, this could happen for other reasons than difference in dialects.

Costa et al. (2014) point out that moral decisions can be different when one uses one’s mother tongue or an acquired language. Their argument is that native languages elicit

more emotional responses than acquired ones, and that the lack of emotions lead to more utilitarian reactions.

2.2.6 *An Unfinished Story?*

The debate between pro- and anti-Whorfians will probably result in more writing and controversies. It is more than seventy years old, and keeps going. Fortunately, both linguists and, more recently, economists have given it a new life and cooperation between the two disciplines should become more frequent. The recent examples studied by economists point to mixed results. Comparing Galor and colleagues’ conclusions to those of Chen, Sutter, Mavisakalyan, and Tabellini, one may tentatively conclude that climatic and agricultural conditions and human interventions (farming) seem to influence language, which in turn seems to influence behavior. This and the fact that even a relatively simple (and old) issue of colors and language leads to the paper by linguists Regier and Kay entitled “Whorf was half right,” may, however, also make both relativists and universalists happy.

2.3 *Linguistic Emotions*¹⁵

Language is obviously a part of one’s self, and many philosophers, sociologists, essayists, and novelists discuss the subject and express their emotions. The German philosopher and sociologist Tönnies (1887) preferred “English for irreverence, German for metaphysical speculation or French for expressions of sensual pleasures,”¹⁶ while Charles V, the Holy Roman Emperor, is quoted as having said that he speaks Spanish to God, Italian to women, French to men, and German to his horse. Wismann (2012), a German-born philosopher who lives in France, claims that the French language allows for more complicity between speakers,

¹⁴ See also the recent paper by Mavisakalyan and Weber (2018).

¹⁵ This section is based on Bond and Ginsburgh (2016).

¹⁶ See Bond and Ginsburgh (2016, p. 234).

which is precluded in German because the syntax imposes to finish a sentence with a verb.¹⁷

Some famous writers who had to leave their native country against their will, such as Elias Canetti (Bulgaria), Paul Celan (Romania), Imre Kertész (Hungary), and Norman Manea (Romania) kept writing in their native languages.¹⁸ Joseph Brodsky (Russia) translated his own works into English. Others who went into exile for similar reasons switched to the language spoken in their country of adoption because they could no longer tolerate their native language (Aharon Appelfeld¹⁹), some did so because they probably felt more comfortable writing in the language of the country to which they had migrated (Samuel Beckett,²⁰ Joseph Conrad, and Eugène Ionesco). Franz Kafka was unhappy writing in his native German, longed for Yiddish, but could neither speak nor write it. And so was philosopher Jacques Derrida, a French Jew, born and educated in Algeria. He was fascinated by Berber, Ladino, and Hebrew, but wrote in French.

Some people have no difficulty forsaking their native language, feeling that communication with others who speak foreign languages is important. Most literary writings from the Middle Ages were in Latin, “the language of nobody” (Zink 2014, pp. 8–19), and certainly not the language of the poor. Nowadays, novels are written in English in many countries all over the world, and

writers from Nigeria, India, South Africa, Sri Lanka, Zanzibar, Pakistan, Rhodesia, Hong Kong, the West Indies, or Egypt have been awarded the British Man–Booker prize.

Others, who used to write in the language of their former colonizers, felt compelled to go back to their native language to “decolonize their mind” (wa Thiong’o 1986). But they were an exception. Many wanted “to identify with that which is the furthest removed from themselves; for instance, with other peoples’ languages rather than their own” (wa Thiong’o 1986, p. 3). One such case is Léopold Sédar Senghor, a Senegalese poet, who was the first African member elected by the Académie Française, and became the first President of the Republic of Senegal in 1960. He even wrote the text of the Senegalese national anthem in French, though it contains the following verse “Stand up, brothers, here is Africa assembled,”²¹ and may be played by typical Senegalese music instruments. Here is why he preferred French to his mother tongue and made French Senegal’s official language:

What does using French represent for me as a black author? [...] I think in French; I express myself better in French than in my mother tongue [...] French is one of those great organs in which you can pull all the stops, creating every possible effect, from the most suave sweetness to the ferocity of the storm. It can alternate or simultaneously be the flute, the oboe, the trumpet, the tam tam and even the canon. And French has given us its words for the abstract—which are so rare in our mother tongues in which tears become gems. Our words bear the aura of sap and blood; words in French shed the rays of a thousand lights, like diamonds. Rockets which set alight our night (Senghor 1956 and Senghor 1962, p. 842).

In the wake of independence gained by British, French, and Portuguese colonies in

¹⁷ See Bond and Ginsburgh (2016, p. 236).

¹⁸ See Canetti (1977) and Manea (2012), for example.

¹⁹ See Appelfeld (2017).

²⁰ This is one of Beckett’s specialist Edwards (1992) stories: “One of the replies that he gave to the perennial question, ‘Why did you decide to write in French?’ was ‘Pour faire remarquer moi’ ... [T]he answer is perfectly serious ... He speaks pidgin here as if to indicate ... that French written by a foreigner ... is not the same thing as French written by a Frenchman.” In fact, Beckett’s French is not correct. He should have said “Pour me faire remarquer.”

²¹ Debout, frères, voici l’Afrique rassemblée.

the 1960s and 1970s, many African politicians educated in western Europe reacted in the same emotional way, and imposed the language of the colonizer as official language in their country. This had as obvious consequence that the “majority of Africans are governed in a language that they do not understand” (Phillipson and Skutnabb-Kangas 1995, cited by Spolsky 2004, p. 182).

This is obviously not an isolated episode. Various decisions and policies tend to disenfranchise individuals or groups whose linguistic, cultural, and historical values and sensibilities are perceived to be under threat. Linguistic disenfranchisement (Ginsburgh and Weber 2005, 2011) occurs when linguistic rights are restricted or even denied. It is unavoidable that linguistic policies may entail social sacrifices, and one has to balance the common interest and those of the various groups. But the results could be painful, as shown by the following two examples.

Sri Lanka is populated by two major ethnic and linguistic communities: Sinhalese, predominantly Buddhist, and Tamil, predominantly Hindu. They had peacefully coexisted for some 2,000 years. After one hundred fifty years under the British rule, the island attained self-governance in 1948, which triggered attempts by both groups to promote their language to replace English as the official language. Fearing Tamil dominance, the Sinhalese majority forced the signature of the Sinhala-only Act, that prompted an emotional and violent reaction from the Tamil minority. A long civil war ensued, which claimed thousands lives and devastated the country.

The other interesting case is Rwanda, where the official language was French after the Belgian colonizers had left in 1962, while both the Tutsi and the Hutu used to speak the same native language, Kinyarwanda. In 1994, during the dramatic events between the Tutsi and Hutu, many Tutsi who had for

some time found refuge in English-speaking Uganda, came back to Rwanda and ruled the country from that time on. French was the language of education since 1962. English was added in 1994 and became the unique language of education in 2008, in a country where the majority of inhabitants live in the countryside and speak Kinyarwanda. Two main reasons come to mind in explaining this change to English. First, since the new Rwandan leaders are Tutsi who lived in exile in Uganda, they may have thought that English would be deemed less of a colonial language than French that was spoken in the country before its independence. But the decision may also be more political, since the Hutu had been supported, and some say, protected, by France during the 1994 slaughter.²²

Emotions are obviously brought on and are expressed by the languages that one speaks. Emotions have effects on economic, social, and political decisions. To quantify these effects, some formalization, represented by using notions such as ethnolinguistic fractionalization, polarization, and disenfranchisement, is examined in section 3.2.2.

3. *The Multilingual World*

3.1 *Identification of Groups*

While recognizing the role of linguistic, religious, or ethnic differences in studying diversity and its potential impact on economic outcomes, one has to tackle two issues. The first concerns identifying the boundaries of groups in order to partition one or several countries and compute the resulting fractionalization of the society. The second is how individuals identify themselves with a community to which they presumably belong.

²²An economic reasoning for the switch to English is discussed in section 4.1.

It is indeed difficult to identify cultural groups, especially within countries created by artificial constructions. In those countries with large numbers of linguistic, ethnic, and religious varieties, their ethnic spreads often do not coincide with the official borders. As Laitin (2000, p.143) points out:

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 various aspects of culture are hard to describe and for the sake of simplicity, language may be, and is often, used as a proxy for culture and/or ethnicity. It is thus not surprising that the first and most influential country-by-country identification widely known as ethnolinguistic fractionalization (ELF) conducted in the Soviet Union some fifty years ago was based mainly on linguistic and historic origins of various groups. It was undertaken by Soviet ethnographers of the Miklouho-Maclay Institute of Anthropology, Archeology and Ethnography in Moscow, and published in *Atlas Narodov Mira* (Bruk and Apenchenko 1964). This remarkable data set was used by Western scholars, starting with Rustow (1967) and Taylor and Hudson (1972) and, for almost fifty years, played a crucial role in analyzing the impact of linguistic diversity on growth, investment in public goods, quality of government services, corruption, and so on. Fearon (2003), Alesina et al. (2003), Alesina and Zhuravskaya (2011), and Desmet, Ortuño-Ortín, and Weber (2009), among others, developed more advanced fractionalization data sets.

Partitioning into linguistic groups is, however, not fully failsafe. First, members of the same group may exhibit behavioral intra-heterogeneity and even fight each other. Tutsis and Hutus share almost the same language (Kinyarwanda in Rwanda and

Kirundi in Burundi), but this prevented neither the Rwandan genocide nor the conflict between the two populations that still goes on in Burundi. Secondly, there is the issue of how fine the linguistic partition should be. Consider, for example, Nigeria with its 527 spoken languages, not including the vehicular ones. Hausa is spoken by 27 percent of the country's population, Yoruba by 17 percent, Igbo by 15 percent, and Fulfulde by 9 percent.²³ Should the fractionalization index account for all 527 groups or should it be limited to five groups (Hausa, Yoruba, Igbo, Fulfulde, and others)? The answer has a direct impact on the level of the country's diversity: with 527 linguistic groups it would be very large, but it reduces dramatically if the population is divided into five linguistic groups.

In fact, the answer to the question of whether to partition speakers of various languages and dialects in distinct groups or not should be provided by linguists and not by economists, though the answers matter for economic analysis. While economists are not qualified to determine whether Serbian and Croatian are different, they can use the notion of linguistic proximity and take advantage of the fact that Serbian and Croatian are indeed very close to each other. Thus, the notion of linguistic distance, which will be extensively discussed in section 3.2.1, allows mitigating the effect of dichotomous decisions.

One has to stress, however, that the link between identity and linguistic distances is not straightforward. Kinyarwanda and Kirundi are good examples. The conflict between Russia and Ukraine, its length and severity, show that being linguistically, culturally, and religiously close are by no means a guarantee of peaceful coexistence. Thus, a study of identity should not be reduced to the mere discussion of linguistic differences

²³The data are approximately for 2000. See *Ethnologue's* website, consulted on November 23, 2017.

or ethnic groups. Fearon (2003, p. 197) illustrates this problem in the context of the United States:

What are [the United States'] ethnic (or racial) groups? Let us make things much easier by restricting attention to groups with at least 1 percent of country population. If we consult official census categories, we get three "races"—white, African American, and Asian—and an additional group, Hispanic, which the government emphatically declares is "not a race." Is this the right list for the United States? Why not disaggregate Hispanic into Puerto Rican Americans, Cuban Americans, Mexican Americans, and so on, or likewise for Asian? Why not distinguish between Arab Americans, Irish Americans, Italian Americans, German Americans, and so on? And why should we use the current census categories, when earlier censuses formulated the categories quite differently.

To study linguistic diversity in the United States, Baugh (2009) suggests dividing the population into three groups: those who learn standard English as their first language, those who learn a nonstandard dialect of English natively, and those who do not learn English as their mother tongue.

It should also be pointed out that even the divisions mentioned by Fearon are not etched in stone. Individuals may try to signal a different identity than theirs by "out-migrating" from their group. By endogenously choosing their perceived identity, individuals may change their language, accent, names, social networks, etc. Kim and Loury (2019), who coined the term "out-migrating," analyze the economic and welfare implications of such choices and argue that they could lead to rising societal inequality.

How does one associate oneself with one or another group? How strong is the association of individuals in American cities with African American Vernacular English (AAVE) or New York Latino English (NYLE)? Grogger (2011) finds that African Americans whose accents are distinctively black earn about 12 percent less than those who do not use

the distinct dialect. The economic incentives of speaking standard English at the workplace rather than AAVE are thus quite obvious. Identification is driven by the fear of being rejected by one's own community, but the societal "stigma" (Besley and Coate 1992) may also lessen the ties with one's own community. As pointed out by Austen-Smith and Fryer (2005, p. 552), "the key idea is that individuals face a two-audience signaling quandary: Behaviors that promote labor market success are behaviors that induce peer rejection."

The extent of self-identification and the strength of affiliation with one's own group do not paint the whole story of a multilingual society. It is crucial, especially in the context of civil wars and conflicts, to address the relationship between groups that embodies two notions: identification with one's own group and alienation toward other groups.

The question is what linguistic diversity is really proxying for. Does it represent preference diversity? Does it relate to trust between groups? Or is language just related to communication? The empirical and theoretical aspects of identification that use linguistic diversity indices are often viewed as a proxy for societal divisions more generally: Easterly and Levine (1997) use ethnolinguistic diversity as proxy for ethnic diversity. In some parts of the world, such as sub-Saharan Africa, linguistic diversity is a proxy for ethnic diversity, whereas in other regions, such as the Americas, it is less the case. In that sense, in countries like the United States, linguistic diversity may miss part of the overall ethnolinguistic diversity.

Alesina et al. (2003) examine linguistic, ethnic, and religious diversity separately. In section 5.1 we get back to these issues and suggest merging the various measures into a unique index. Some of those questions are also examined by Melitz and Toubal (2014) and Desmet, Ortuño-Ortín, and Weber (2017) and are discussed later.

3.2 *Linguistic Distances and Ethnolinguistic Indices*

3.2.1 *Measuring Distances*

In order to measure proximity between languages, linguists use two basic methods: *synchronic* linguistics and *historical* or *diachronic* linguistics. Synchronic linguistics deals with languages at one point in time and leads, in particular, to computing *distances* between languages say, today: How far is English from Swedish? Can an American communicate with a Swede if each of them speaks or writes in his native language, or will they have to learn a common language (or use an interpreter) if they want to communicate or trade? Historical linguistics thinks in terms of reconstructing the earlier stages of one or several languages. This leads to linguistic trees (*cladistics*). However it does not tell us at what time a language branched off from a common stem and created a new language. French, Spanish and Italian have a common root. When did they separate? This is the object of *glottochronology*.

Economists use the results of synchronic and diachronic linguistics, and may even combine the two approaches, but are not (yet) interested in glottochronology.

Several elements contribute to languages being related or not. *Vocabulary* (or lexical terms) is the most obvious. *Phonetics* or the sound of words comes next. *Syntax*, the “way in which linguistic elements (as words) are put together to form constituents (as phrases or clauses),”²⁴ illustrates another difference or relation between languages. Verbs, for instance, change place: they are often in the middle of a sentence in English, and end a sentence in German. *Grammar* contributes to heterogeneity as well. German has declensions, but only very few of them remain in

English, such as the so-called possessive genitive. Some languages use genders (French), some do not or only seldom (English) and not all languages have the same tenses, as we shall see later. These are just a few examples to illustrate that computing the relatedness of languages is difficult, and this is without going into the interesting issue of whether languages have a common structure.²⁵ Both synchronic and diachronic methods are based on comparing languages. Diachronic methods go deeper, since they try to determine “common ancestry, and descent through time with gradual divergence from [a] common source” (McMahon and McMahon 2005, p. 3), but do not study intercommunication of people with different linguistic repertoires. They use morphological resemblances discussed above, such as lexical items, and identify groups and subgroups of languages according to their similarity. This leads to linguistic trees, with roots, branches, sub-branches, and final twigs, each of which corresponds to a unique language that may still be spoken or already extinct.²⁶

Table 1 gives an example of such a comparison of words for numbers *one* to *five* in 15 languages.²⁷ Linguists have to find languages that are related and partition them into groups. The first language is English; the reader is invited to guess the others and imagine the burden if she were faced with 1,000 words and 200 languages.

Table 2 gives the solution to this reasonably easy puzzle. The first four groups are Indo-European languages, subdivided into Germanic, Romance, Celtic, and Slavic families. The next ones are Hungarian, which belongs to the Uralic family, and Swahili, one of the Bantu languages spoken in Africa.

²⁵This still generates controversies, see Harris (1993), and Baker (2001).

²⁶See Nakhleh et al. (2005), who describe the construction of the Indo-European language tree.

²⁷This example is borrowed from Ginsburgh and Weber (2016).

²⁴Merriam-Webster Dictionary.

TABLE 1
WORDS FOR NUMBERS 1 TO 5 IN 15 LANGUAGES

Language	one	two	three	four	five
1. English	one	two	three	four	five
2.	um	dois	três	quatro	cinco
3.	uno	due	tre	quattro	cinque
4.	odin	dva	tri	chetyre	piat'
5.	eins	zwei	drei	vier	fünf
6.	moja	mbili	tatu	nne	tano
7.	een	twee	drie	vier	vijf
8.	unan	daou (m)	tri (m)	pevar (m)	pemp
9.	en	to	tre	fire	fem
10.	jeden	dwa	trzy	czetry	pieć
11.	en	två	tre	fyra	fem
12.	un	deux	trois	quatre	cinq
13.	un	dau (m)	tri (m)	pedwar (m)	pump
14.	uno	dos	tres	cuatro	cinco
15.	egy	kettő	három	négy	öt

Note: (m) is for masculine

TABLE 2
GROUPING WORDS FOR NUMBERS 1 TO 5 IN SOME INDO-EUROPEAN LANGUAGES, HUNGARIAN, AND SWAHILI

Language	one	two	three	four	five
<i>Germanic languages</i>					
1. English	one	two	three	four	five
5. German	eins	zwei	drei	vier	fünf
7. Dutch	een	twee	drie	vier	vijf
9. Danish	en	to	tre	fire	fem
11. Swedish	en	två	tre	fyra	fem
<i>Romance languages</i>					
2. Portuguese	um	dois	três	quatro	cinco
3. Italian	uno	due	tre	quattro	cinque
12. French	un	deux	trois	quatre	cinq
14. Spanish	uno	dos	tres	cuatro	cinco
<i>Celtic languages</i>					
8. Breton	unan	daou (m)	tri (m)	pevar (m)	pemp
13. Welsh	un	dau (m)	tri (m)	pedwar (m)	pump
<i>Slavic languages</i>					
4. Russian	odin	dva	tri	chetyre	piat'
10. Polish	jeden	dwa	trzy	czetry	pieć
<i>Uralic languages</i>					
15. Hungarian	egy	kettő	három	négy	öt
<i>Bantu languages</i>					
6. Swahili	moja	mbili	tatu	nne	tano

Note: (m) is for masculine

The words in the first five languages (Danish, Dutch, English, German, Swedish) are all related and belong to the family of Germanic languages. This is also the case for the next four (French, Italian, Portuguese, Spanish) which are Romance languages. The third group (Celtic) also shows that the first three words in all three families are related, but this is no longer the case for digits *four* and *five*. Next come Slavic languages, which are clearly related to each other. They also have a relation for digits *two* and *three* with the previous groups. The two remaining languages, Hungarian and Swahili have nothing in common with Indo-European groups, but may nevertheless have a faraway ancestor.²⁸

The comparative method can be made loose by using as much information as possible (*mass comparison*), or very tight by using a unique characteristic of a language, its *lexicon* (which leads to *lexicostatistics*). It can also use median situations, which are reproducible (or falsifiable) by other researchers. We now turn to some of the most common methods used in synchronic and diachronic analyses.

Mass Comparison.—Mass comparison was used by Stanford linguist Joseph Greenberg (1955, 1963, 1987, 2000, 2002) to classify Native American, Indo-Pacific, African, as well as Eurasiatic languages. It is, claim McMahon and McMahon (2005, p. 19, 22),

so straightforward and non-technical that in the eyes of many historical linguists it scarcely qualifies as a method at all.

Greenberg himself may not need such pedantry; his intuitive sense for linguistic affinity is the subject of some renown. But other linguists may. And science is supposed to be a game anyone can play,

which implies that what Greenberg did is difficult to repeat or to test statistically.

²⁸See Ruhlen (1994) for an entertaining, deep, and instructive exposition.

Lexicostatistics.—Lexicostatistical methods are based on the similarities and supposed common roots of words across the vocabularies of various languages, as is done in tables 1 and 2. Languages can be related, and their similarities can be explained by the following three mechanisms: (i) there are words that look common by accident—this is so for onomatopoeias; (ii) languages borrow words from other languages—some 30 percent of English words were borrowed from French after the Norman conquest in 1066;²⁹ and (iii) the words in different languages may descend from a common older language. These are the ones that matter and are used to compute lexical distances.³⁰

Since it would be daunting to compare long lists of words for each couple of languages, linguists rely on a small selection of carefully chosen words, a so-called *list of meanings*. Such a list was compiled by Swadesh (1952), who chose meanings that were basic enough to exist in most languages and cultures (such as animal, bad, bite, black, child, die, eat, eye, hunt, and digits from *one* to *five*),³¹ on which deductions can be based. Swadesh's list contains 200 meanings, which were later reduced to 100. Both lists are still used nowadays.

Distances between couples of languages can be computed in several ways. We shortly examine (i) *lexicostatistical distances*, based on the percentage of *cognate* words (that is, words based on historical links via an

²⁹According to Janson (2002, pp. 157–58), “around 90 per cent of the words in an English dictionary are of French, Latin or Greek origin.” This however does not imply that English and French are close, since “if one counts words in a text or in a recording of speech [in English], the proportion of Germanic words is much higher, for they are the most frequent ones, while most of the loans that figure in a dictionary are learned, rare items.”

³⁰Ignoring borrowed words may rule out factors that influence the closeness of two languages, and the ease of learning the other if one knows one of them.

³¹See Kessler (2001, pp. 199–257) for the lists of meanings chosen by Swadesh. His list has been slightly changed to accommodate Southeast Asian and Australian languages.

older language) in each couple of languages; (ii) *Levenshtein distances*, based on analyzing and comparing words (or their phonetical representations) character by character; (iii) distances defined on *linguistic trees*; (iv) alternative methods.

Lexicostatistical Distances.—Dyen, Kruskal, and Black (1992) were the first to use Swadesh's list of meanings to compute the distances between 84 Indo-European speech varieties.³² The lexicostatistical method consists of three steps: (i) collecting the words used for each meaning in Swadesh's list for each speech variety; (ii) making cognate decisions on each word in the list for each pair of speech varieties, according to whether they have a common ancestor,³³ do not have one, or a decision is unclear; and (iii) calculating lexicostatistical percentages, i.e., the percentages of non-cognates among all cognate and non-cognate links for each pair of languages.³⁴ In extreme cases the percentage is one if there are no cognate pairs and zero if all links are cognate.

If the only meanings to be compared were the five digits in tables 1 and 2, then the distances between each pair of the five Germanic languages would all be equal to 0. The same would be true for the pairwise distances within the two other families. Things get a little more difficult across the three families, since the words for digits *four* and *five* (compare the English *four* with the French *quatre* or the Welsh *pedwar*) are certainly further apart. Assume that as linguists, we decide to classify them as non-cognate. Then, the distance between, say English and

French, would be 2/5. Finally, the distance between any Indo-European language and Hungarian or Swahili would be equal to 1.

Levenshtein Distances.—Levenshtein (1966) suggested using an algorithm that measures the distance between strings, such as those formed by words. His idea was to transform the word of one language into the word of the other one by inserting, deleting, or substituting alphabetic (or phonetic) characters. The minimal number of such transformations divided by the maximum number of characters between the two words is the Levenshtein—also called *edit*—distance between the two words. The distance between two languages is the average of the distances over all the meanings that are considered.

Calculating Levenshtein distances is much easier to handle than lexicostatistical distances, which rely on decisions made by linguists. It may, however, lead to problems since obvious non-cognate words may be considered cognate by the computer.³⁵

According to McMahon and McMahon (2005, pp. 212–214), the technique is applicable to dialects, but it would “compromise the method if it were extended to comparisons between languages or across considerable spans of time, since it would then be more likely that changes in the order of segments [within words] would have taken place.”³⁶

Cladistic Distances.—An alternative way to compute distances is to use linguistic trees, based on world classifications of languages such as the ones in *Ethnologue*, and compute

³²The conjecture that Aryan languages spoken in parts of India and European languages may have a common ancestor had already been made by William Jones, in 1786, and is still holding today. See Gray and Atkinson (2003), Gamkrelidze and Ivanov (1990), and Ruhlen (1994) for a general overview.

³³See Warnow (1997) for further technical details.

³⁴Undecided cases are discarded.

³⁵An illuminating example is the tiny Levenshtein distance between *kitten* and *mitten* which only needs the substitution of the first character in *kitten* by an *m*, though the words mean very different things.

³⁶This approach was pioneered by Goebel (1982), and Kessler (1995) for Irish dialects, and more recently by Nerbonne, Heeringa, and associates for Dutch dialects. See Nerbonne and Heeringa (1997).

cladistic distances. These lie somewhere between mass comparison and lexicostatistical or Levenshtein distances, as they account for various language characteristics, including the lexicon, but also, syntax, phonology, and grammar. Such trees are available for almost all languages in the world.

Fearon and Laitin (1999), Laitin (2000), and Fearon (2003), who suggested this approach,³⁷ use the distances between the branches of a linguistic tree as a proxy for distances between linguistic groups. In the original Fearon and Laitin (1999) index, the score takes the level of the first branch at which the languages break off from each other for every pair of languages. The higher the number, the higher the similarity of languages. For example, the break-off between Greek and Turkish happens immediately at the first stage, yielding the highest distance between the two languages. On the contrary, Czech and Slovak go a long distance together, facing the break-off only at the last stage, yielding the smallest possible distance between the two. Fearon (2003) produced a data set for 822 ethnic groups in 160 countries. This approach is now used and refined by many researchers, since it covers most languages and computations can be stopped at various levels of branching or aggregation.³⁸

Intercommunication Distances.—Melitz (2008) and Melitz and Toubal (2014) construct various types of bilateral distances between each pair of countries. They include common official language (COL), common native language (CNL), and common spoken language (CSL). COL is obviously a dichotomous variable, CNL identifies the probability that two randomly chosen individuals, one in

each country, would be able to communicate in a common native language.³⁹ The same probability, but with respect to common spoken language is measured by CSL. It is important to point out that CNL and CSL play a different role in patterns of communication between two countries. CNL is rather associated with common ethnicity, and thus trust, whereas CSL is indicative of a mere technical efficiency of intercommunication that acts beyond ethnicity.

Analyzing Sound.—In its 2012 report the Functional Phylogenies Group (FPG) analyzes the properties of phonetic sounds “that include pulse, intensity, sound wave components, spectrum, and/or duration of the examined sound segment, as well as fundamental frequency, [which is what the] listener identifies as pitch, and relates to how fast the vocal folds of the speaker vibrate during speech” (Hadjipantelis, Aston, and Evans 2012, p. 4652) as well as speech sound evolution. The idea is to treat speech sounds as continuous functions, rather than discrete variables, and apply statistical methods (such as principal components,⁴⁰ and regression models). FPG intends to use this approach to construct cladistic trees.

Distances Based on Learning Scores.—The approaches described so far have served various purposes, and their link with genetics, migrations patterns, and archeology are of particular importance.⁴¹ But economists are also interested in today’s world and the study of current trades, migrations, translations, or language learning. An alternative method, which takes into account all characteristics including borrowed words,

³⁷According to McMahon and McMahon (2005, p. 125), a similar method had been suggested some years earlier by Poloni et al. (1997).

³⁸See Desmet, Ortuño-Ortín, and Wacziarg (2012) and section 5.1.

³⁹See section 3.2.2 for details.

⁴⁰See Aston, Chiou, and Evans (2010).

⁴¹See Cavalli-Sforza (1997, 2000), Cavalli-Sforza and Cavalli-Sforza (1995), Cavalli-Sforza, Menozzi, and Piazza (1994), Renfrew (1987), and Michalopoulos (2012) among many others.

would be to compute the speed at which people with different native languages learn foreign languages. This was envisaged by Hart-Gonzalez and Lindemann (1993) using a sample of Americans born in the United States who had acquired a variety of languages.⁴² Such distances encompass most of the difficulties encountered in acquiring a language, and borrowed words would also find their place. This seems to be a rare set of consistent data, and one can imagine the amount of money, time, and effort required to set up a coordinated project that would use this method.

3.2.2. Fractionalization, Polarization, and Disenfranchisement Indices

The building blocks described earlier, namely ethnolinguistic group identification and linguistic distances, are combined to define fractionalization, polarization, and disenfranchisement indices in a given society. In each case, we distinguish *dichotomous* indices, which take into account the dimension of groups only, and *distance-weighted* indices.

Fractionalization Indices.—Dichotomous fractionalization indices are defined for a multilingual society divided into distinct groups, each member of which speaks the same native language.⁴³ Let society S consist of K groups, $k = 1, \dots, K$, where the population of group k is given by N_k and $\sum_{k=1}^K N_k = N$. Let also $n_k = N_k/N$ be the fraction of group k in the society.

The A-fractionalization index⁴⁴ is defined as the probability that two individuals, randomly

picked from the entire society, belong to two different groups. It is formalized as $A = 1 - \sum_{k=1}^K n_k^2$.

Another important dichotomous index is Shannon's (1948) and Wiener's (1948) entropy measure $E = -\sum_{k=1}^K n_k \log n_k$, which is more often used in biology, statistics, and information science than in the social sciences.

Both the A - and the E -indices have similar mathematical properties. Their formulation was unified through common axioms by Davydov and Weber (2016),⁴⁵ who suggested the following general form $A^\alpha = 1 - \sum_{k=1}^K n_k^\alpha$, where α is a positive parameter different from one. Obviously, the value of the A -index coincides with A^α for $\alpha = 2$. It is also quite easy to verify that A^α approaches E when α tends to one.

The value of α in the above formulation can be interpreted as the degree of societal sensitiveness toward diversity. Countries, regions, or cities may differ with respect to their own valuation of fractionalization. Some could feel threatened by diversity, others may welcome it. These attitudes may be correlated with profound economic and political outcomes. The differences in "perceived diversity," which signal how people feel about the impact of globalization—channeled through employment prospects and the presence of immigrants in their own communities—are much less obvious. In other words, different societies may choose different α s, and estimating this parameter could be an interesting topic of research.

The A -index may, however, produce unexpected results, as shown by Desmet, Ortuño-Ortín, and Weber (2009), who compare its value in two European countries, Andorra and Belgium. In Andorra, half of the

⁴²These distances were used by Chiswick and Miller (2005).

⁴³Here one disregards proficiency in other languages.

⁴⁴This index is often mistakenly called ELF in the economics literature. The term ELF (ethnolinguistic fractionalization) had been used to denote the ethnolinguistic map of the world constructed by Russian ethnographers that was published in *Atlas Narodov Mira* (Bruk and Apenchenko 1964). What many scholars in the field called

the ELF-index, is, in fact, the Greenberg A -index applied to the partition of the world published in 1964.

⁴⁵See also Hill (1973) and Simovici and Jaroszewicz (2002).

population speaks Catalan, while the other half speaks Spanish. In Belgium, roughly 60 percent speak Dutch and 40 percent speak French. This leads to an A -index of 0.5 in Andorra, and 0.48 in Belgium, with Andorra seeming more diverse than Belgium. This is so because the A -index does not account for linguistic distances, which are much smaller in Andorra (two Romance languages) than in Belgium (one Germanic and one Romance language). Incorporating linguistic distances shows that Belgium is more diverse than Andorra.

To address this issue, Greenberg (1956) introduced a monolingual weighted index, which, in simple words, accounts for the average linguistic distance between randomly chosen individuals within a society. In addition to the previously defined notation, let d_{ki} denote the linguistic distance between two groups $i, k = 1, \dots, K$, which can be any of those described in section 3.2.1. Greenberg defined his index as $B = \sum_{k=1}^K \sum_{i=1}^K n_k n_i d_{ki}$, which generalizes the A -index. It is easy to check that if dichotomous distances are introduced in the B -index, it turns into the A -index.⁴⁶

The B -index obviously imposes more data requirements than the A -index, but the effort is rewarding. This is a clear outcome of the results by Desmet, Ortuño-Ortín, and Weber (2009) in their cross-country analysis of redistribution patterns, as well as in the definition of societal indices by Castañeda Dower, Ginsburgh, and Weber (2017) in the context of the Sri Lankan conflict.

In many applications such as the provision of public goods, redistribution, or the intensity of conflicts within a country, one may be led to account for the special status of

some regions, in particular a center c and all other $K - 1$ peripheral regions.⁴⁷ This leads to a special case of the B -index in which only the distances between center and periphery have to be accounted for, leading to an index $CP = n_c \sum_{k=2}^K n_k d_{kc}$.

Polarization Indices.—Both the A - and the B -indices rely on a preexisting partition into distinct ethnolinguistic groups. Polarization and the indices that it generates add self-identification, which comes through in two ways: strong identification with individuals in one's own group and alienation toward others. Esteban and Ray (1994) combined both identification and alienation, using an axiomatic approach. The functional form of their polarization index $P = \sum_{k=1}^K \sum_{i=1}^K n_k^{1+\alpha} n_i d_{ki}$ is very close to index B , where now the d_{ki} represent income differences between groups k and i and α is a positive parameter with values between 1 and 1.6.

Esteban and Ray (1994) were interested in income polarization, where distances between groups are naturally represented by income differences. Though one loses the elegance of their index, which is derived from axioms, one can think of incorporating linguistic distances instead, assuming that some additional factors such as linguistic proximity or some historical path could play an important role in determining the degree of identification and alienation.⁴⁸

Reynal-Querol (2002) considers a dichotomous version with 0-1 distances between groups and $\alpha = 2$, which reads $RQ = \sum_{k=1}^K \sum_{i=1}^K n_k^2 n_i$.⁴⁹ It is worth pointing

⁴⁷ See for example Desmet, Ortuño-Ortín, and Weber (2009, 2017).

⁴⁸ See Montalvo and Reynal-Querol (2005), Desmet, Ortuño-Ortín, and Weber (2017), and Castañeda Dower, Ginsburgh, and Weber (2017).

⁴⁹ Using additional axioms, Geng (2012) later showed that the range of α s could be shrunk to a single point $\alpha = 1$ to obtain the Reynal-Querol functional form.

⁴⁶ Bossert, d'Ambrosio, and La Ferrara (2011) offer an axiomatic foundation for a variant of the B -index, relying on primitives of individuals that are not pre-assigned and do not exogenously determine groups within a society in the way ethnic groups do.

out the intuitive difference between A and RQ . Recall that A is determined by the probability that two randomly chosen individuals belong to two different groups. Thus, the value of A is the sum of the terms $n_k(1 - n_k)$, each identifying the probability that one individual belongs to group k , while the other does not. The value of RQ is determined by the probability that among three randomly chosen individuals, two belong to the same group, while the third belongs to another one. Thus, RQ is the sum of the terms $n_k^2(1 - n_k)$, in which two individuals belong to group k , while the third does not.

Disenfranchisement Indices.—The notion of linguistic disenfranchisement, formally introduced by Ginsburgh and Weber (2005),⁵⁰ provides a rigorous framework for examining linguistic discrimination. In terms of Skutnabb-Kangas and Phillipson's (1989, p. 455) formulation, linguistic discrimination and disenfranchisement amount to introducing "ideologies and structures which are used to legitimate, effectuate, and reproduce unequal division of power and resources (both material and non-material) between groups defined on the basis of language." Denying linguistic rights used to be, and still is, a common problem on all continents.⁵¹ The most severe example is probably Africa, since in many countries, the language of the former colonial power replaced the native language as official language.

To provide a formal response, consider the set L of all spoken language in a country. Note, L may coincide with all K native languages, but may also include some other languages (such as English) that are not native in the country. The construction of such indices needs additional information

consisting of proficiency data for all or for a sample of individuals in each of the L languages. These can be found in censuses or specific surveys.

Consider the language(s) contained in subset $T \in L$ as a potential set of official language(s), to be used for official documentation, communication, media and educational purposes. For an individual m , let $k(m)$ denote the native tongue and $L(m)$ the set of all other languages spoken by m (his linguistic repertoire). Individual repertoires are then aggregated to evaluate the societal disenfranchisement level. There are four ways to measure individual m 's disenfranchisement:

- (i) *Index DN: Dichotomous distances and native languages only.* For each individual m in S , disenfranchisement is equal to 1 if his native language is not included in T . Otherwise, he is not disenfranchised and his disenfranchisement is set to 0. Then, DN represents the total population of the linguistic groups k whose language k is not included in the set T . That is, $DN = \sum_{k:k \notin T} n_k$.
- (ii) *Index DA: Dichotomous distances and all languages.* Individual m 's disenfranchisement is equal to 1 if she speaks no language in T . It is equal to 0, if m 's linguistic repertoire includes at least one language from T . Then $DA = \sum_{m:L(m) \cap T \neq \emptyset} 1$.
- (iii) *Index NN: Linguistic distances and native languages only.* For an individual m in society S , the level of disenfranchisement is set at the distance between his native language and the one closest to it in T . Denote this distance by $d^{nat}(m, T)$. Then, $NN = \sum_{m \in S} d^{nat}(m, T)$.

⁵⁰ See also Ginsburgh, Ortuño-Ortín, and Weber (2005).

⁵¹ See Ginsburgh and Weber (2011) for a number of cases.

- (iv) *Index NA: Linguistic distances and all languages.* For any $m \in S$, disenfranchisement is determined by the minimal distance between his linguistic repertoire $L(m)$ and the set of languages T . Denote this distance by $d^{all}(m, T)$. Then, $NA = \sum_{m \in S} d^{all}(m, T)$.

3.3 Using Linguistic Distances in Gravitational Models

Hägerstrand (1957) was the first to use the metaphor of the gravity model in his study of migrations in Sweden. Tinbergen (1962) popularized the model by applying it to study trade flows between countries. In these very early applications, both Hägerstrand and Tinbergen used geographical distances as representing frictions and costs, but somewhat later, linguistic distances were added as control variables and the idea was extended to model migratory decisions, translations of literary works and industrial patents, money laundering and, to some extent, financial transactions.

The simplest form of the equation (where left- and right-hand side variables are expressed in logarithms) is:

$$x_{ij} = \alpha + \beta z_i + \gamma z_j + \delta d_{ij},$$

where i and j are countries, regions, or languages in the case of translations, x_{ij} is a flow from i to j , z_i and z_j are vectors of variables describing i and j , and d_{ij} is a vector of frictions, including linguistic distances, that may restrain flows between i to j (δ should thus be negative). Finally, α is a scalar and β, γ , and δ are vectors of parameters.

3.3.1 International Trade

Tinbergen's idea was followed by hundreds of trade papers, both applied and theoretical, in order to elicit the assumptions and structural tenets that may lead to various forms of the original gravity equation. Those

papers provided a bridge from a partial to a general equilibrium framework in different economic environments (perfect competition, monopolistic competition, heterogeneous goods, etc.).⁵²

In the basic trade equation, x_{ij} represents the value of exports from country i to country j , the vectors z_i and z_j contain exporter- and importer-specific variables such as their GDPs, and d_{ij} are distances. We focus on linguistic distances, but many models also contain dummy variables representing adjacency, trade agreements, former colonial ties, etc. between couples of countries, as well as geographic or other types of distances. To represent linguistic distances, the first papers introduced a dummy variable with value 0 if the language was a common official, common native, or common spoken language in the two countries, and 1 otherwise. Later on, and as more detailed statistics on languages spoken in countries became available, specifications became more subtle. It would be frightening (and probably add little to the fact that linguistic distances have an effect on trade) to discuss in detail this large strand of literature. We therefore limit our review to a paper by Egger and Lassmann (2012) who produced a meta-analysis of the effect of languages on trades, and to one recent study by Melitz and Toubal (2014) that is more interested in language than in trade.

Egger and Lassmann (2012) analyze the effect of 701 coefficients picked up by linguistic distances in 81 articles published between 1970 and 2011 in 24 refereed journals.⁵³ The average coefficient is equal to -0.49 (with a standard deviation of 0.37). Ignoring the standard deviation, this implies

⁵²See Anderson and Van Wincoop (2004), Baier, Kerr, and Yotov (2017), Costinot and Rodríguez-Clare (2014), and Head and Mayer (2014) for recent surveys. See also Hanson and Xiang (2011) for heterogeneous goods.

⁵³The list of articles can be found in their working paper available at <http://hdl.handle.net/10419/54919>, last accessed on April 20, 2017.

that a reduction of 10 percent in the linguistic distance increases trade by almost 5 percent. Egger and Lassmann also run a meta-analysis using the estimated δ s as left-hand side variable, which they regress on 18 variables, and find two interesting and robust conclusions: Former colonial ties decrease the (absolute) value of the distance coefficient, while the language effect decreases over time.⁵⁴ In most trade papers, language is represented by a unique variable, since they focus on trade rather than on language, and language is a control variable among many others. The large variety of the size of the language effect is partly due to the fact that authors use different data sets and different linguistic distances, which may have different effects on trades.

Melitz and Toubal (2014) use the bilateral trade model with the precise objective of disentangling these many effects.⁵⁵ They utilized the COL, CNL, and CSL indices introduced in subsection 3.2.1, in addition to linguistic distances (LD), which could be either lexicostatistical or cladistic. They point out that if, in the presence of CNL and CSL, COL is also significant, then the contribution of official translations also helps communicating. Moreover, if LD is significant in the presence of the three other measures, translation and interpretation can be of use if official and native languages differ. Their panel consists of over 200,000 observations on trades over ten years (1998–2007). The following right-hand side variables are included: the four linguistic variables discussed above; GDPs; prices; geographic distances; contiguity; former colonial ties; common religion; common legal system; history of wars; and exporter, importer, and year fixed effects. Regression results (restricted to languages) are summarized in table 3. As

can be checked in columns 1 to 3, estimated parameters are highly significantly different from 0 if included one by one. They remain significant, though to a lesser degree, when they are included simultaneously, which shows that each of them has a specific role. It is worth noting that the aggregate effect, equal to -1.11 , is much larger (in absolute value) than what is found by Egger and Lassman (2012) as average effect in other studies.

Melitz and Toubal (2014) also study whether particular languages such as English or other international languages (French, Spanish, German, and Portuguese) could make a difference. They find that, finally, “all that really matters is a common language, whatever the language may be.”

3.3.2 *Migration Decisions*

The standard approach in analyzing the trade-offs of a decision to migrate is based on the difference between benefits and costs. The prospects of higher wages (times the probability of finding a job) or other benefits are contrasted with the monetary and psychological costs, adjustment to a new culture, a new language, and a possible uprootedness or even collapse of the family.

Though the specification of the typical immigration equation is very close to the trade equation (here x_{ij} represents migratory flows between source and destination countries i and j divided by the population of the source country), one of the most important right-hand side variables represents existing networks of source country's previous immigrants who create positive externalities for those who migrate later, since they reduce their cost and facilitate their assimilation. Linguistic distances between the language(s) spoken by possible migrants and the country to which they intend to migrate are also present.

Many papers have studied migrations decisions. The list includes the one by

⁵⁴ Egger (2008, p. 660) attributes this to the “changes in both prices and varieties.”

⁵⁵ See also Melitz (2008).

TABLE 3
EFFECTS OF LANGUAGES ON BILATERAL TRADES

Variable	(1)	(2)	(3)	(4)
Common official language (COL)	−0.51 (13.5)			−0.35 (7.6)
Common spoken language (CSL)		−0.77 (14.6)		−0.40 (4.9)
Common native language (CNL)			−0.86 (11.2)	−0.28 (2.3)
Linguistic distance (LD)				−0.08 (4.2)

Source: Melitz and Toubal (2014, p. 357).

Notes: The dependent variable is the log of bilateral trade. Student *t*-statistics shown in parentheses; 209,276 observations. The paper looks at two measures of linguistic distances: lexicographic and cladistic. Results are almost the same. The table reports on the result with lexicographic distances.

Adserà and Pytliková (2015), who analyze the role of language on migration decisions from 223 source countries to 30 OECD countries during the period 1980–2010, but it is by no means the only one.⁵⁶ The authors use several measures of linguistic distances (lexicographic, Levenshtein, cladistic, a self-constructed one using a combination of the previous ones) as well as GDPs per capita and unemployment rates in the source and destination countries, public expenditure/GDP (as measure of welfare) in the destination country, existing linguistic networks in the destination country, and political freedom indices in the source country. They find that migration rates are higher between countries whose first official languages are closer to each other. Identical results are obtained with other measures, but the importance of language is smaller than that of ethnic networks, a result already obtained by Beine, Docquier, and Özden (2011). Bredtmann, Nowotny, and Otten

(2017), who look at the interaction between linguistic distance and networks, show that the effect is positive, and that the negative effect of distance decreases if networks are larger. Note also that the effect of linguistic proximity increases with education.

Falck et al.'s (2012) paper is particularly interesting since it shows that such results also hold within a nation, Germany, where several dialects are spoken in some 440 districts. Their dependent variable consists of current migrations between districts observed during the period 2000–2006 (divided by the population living in the source district). The distances between dialects are based on phonological and grammatical differences that were collected between 1879 and 1888 (thus more than 100 years before the migrations took place) in 45,000 schools where students had to translate 40 German sentences into their local dialect. The authors find that indeed, “there are intangible cultural borders within a country that impede economic exchange,” even if today, all German citizens share a common language, which was not the case in the late nineteenth century.

⁵⁶ See Adserà and Pytliková (2016) for other references.

Chiswick and Miller (2007) suggested that distances based on learning scores could be positively correlated with barriers to inter-comprehension. They used them as a proxy for distances between American English and languages spoken by immigrants. Isphording (2014) estimates an equation in which the literacy skills of immigrants are influenced by linguistic distances between source and destination countries' languages. Immigrants with a distant language face higher costs to reach the same level of command of the destination language than close languages, especially in the case of migrants who arrive at destination when they are more than 12 years old.

3.3.3 *Translations*

Translations are sometimes viewed as a tool leading to a form of cultural domination or "unethical imbalance" by some languages, in particular English. Pym and Chrupala (2004) were the first to challenge this statement by pointing out that languages producing many books because many people speak and write novels in that language (English, obviously) translate less into their own language, while their production gets translated into languages that are spoken by fewer people.

Ginsburgh, Weber, and Weyers (2011) develop a microeconomic model that leads to a demand equation in which the number of novels translated from a (source) language to a (destination) language is determined by the sizes of the populations that speak both languages as first language (a proxy for the numbers of books written and read, which are unfortunately unavailable); the distance between the two languages, which also proxies cultural distances⁵⁷; the literacy rate; the average income of the population speaking the language into which the title is translated;

as well as source language dummies. The resulting model leads again to a Newton gravity equation, in which the elasticity of the number of translations with respect to linguistic distance is -1 . It also shows that, though the ratio of translations from English to other languages is equal to 22, while the same ratios for French and German are 0.87 and 0.37, there exists no hegemonic effect of English. Indeed, the source dummies picked up by English, French, and German are not significantly different at the 5 percent probability level.

Harhoff et al. (2016) study the effect of translations on patenting strategies in Europe. The system is structured in such a way that once a patent is granted by the European Patent Organization (EPO), the applicant has to validate the patent in each country where he or she wishes to be protected. This requires translations, which happen to be very expensive in some countries, and linguistic distances may, therefore, affect patenting decisions. The authors test this hypothesis on a data set that includes all patents granted by EPO in 2003. The results show that, once approved by EPO, a 1 percent increase in the linguistic distance between the source language in which the patent was written, and the one in which it has to be translated reduces the probability of validation by almost 16 percent.

3.3.4 *Money Laundering*

Walker and Unger (2009) make use of common languages between countries in estimating a gravity-type model of international money laundering flows. They find that money laundering is more significant between countries sharing a common language, which should not come as a surprise.

3.3.5 *Financial Transactions*

Grinblatt and Keloharju (2001) deal with data on the investment behavior by Finnish

⁵⁷A thriller that features New York is more likely to be translated from English into French than a Chinese or an Estonian thriller that unfolds in Shanghai or Tallinn.

investors and companies (whose native tongues may be Swedish or Finnish). They show that the language spoken has an effect on an individual's investment decisions. They do not use distances explicitly since there is a unique distance, the one between the two languages and the model is not of the gravitational type, but the results show that languages matter in exchanges. The authors had access to two years of the daily trades as well as to the language spoken by each investor. Finnish is spoken by some 93 percent of the population, Swedish by 6 percent only, but both languages are official and are very distant from each other.⁵⁸ Swedish speakers are more active in trading and hold 23 percent of household share owner wealth. Among the 97 firms whose stocks are traded, 12 report in Finnish only, two report in Swedish only, and 83 reports are multilingual. Grinblatt and Keloharju also define the culture of the firm according to the name of its CEO: 83 are of Finnish and 14 of Swedish culture. Their results indicate that both language and culture have an effect on the investment behavior of households: Swedish speakers invest more in Swedish firms that publish their annual report in Swedish, and speakers of Finnish rather opt for Finnish firms. A report in Swedish increases the fraction of Swedish shareholders by eight percent, with respect to Finnish shareholders, and Swedish-speaking households prefer to hold and trade the shares of companies whose CEO is of Swedish origin. Languages seem to exert a stronger influence than financial returns, though their effects are "smaller for financially savvy institutions than for households or less savvy institutions as well as for more sophisticated household investors" (Grinblatt and Kelohardju 2001, p. 1053).

⁵⁸The two languages belong to very different branches of the linguistic tree: Swedish is an Indo-European language, Finnish is an Uralo-Altaic language.

Bellofatto (2017) concludes in a similar way for Belgium, where both French and Dutch are spoken. French-speaking investors overweight their portfolios in France and other countries that speak French. Similarly, Dutch-speaking investors invest relatively more in the Netherlands. However, this may partly be due to geographical proximities as Dutch-speaking investors are closer to the Netherlands, while those who speak French are closer to France.

3.4 *Ethnolinguistic Indices and Economic Outcomes*

Greenberg (1956, p. 109) was probably the first linguist to suggest "developing quantitative measures of diversity in order to render impressions more objective, allow the comparing of disparate geographical areas, and eventually to correlate varying degrees of linguistic diversity with political, economic, geographic, historic, and other non-linguistic factors." Greenberg's challenge was addressed several years later, though it took time to use his *A* and *B* indices in applied work.

Indeed, Fishman (1968) and Pool (1972) came back with the idea that linguistic diversity could have an impact on economic activities, but proposed to measure diversity by the share of speakers of the most used language in each country or region. Some 30 years later, Nettle (2000) pointed out that the Fishman and Pool index did not fully address the existence of several languages. He suggested the use of the number of languages divided by the population. This may, however, also lead to distorted conclusions. Consider the Soviet Union (SU), whose breakup left the Russian Federation (RF) with about half of the population of the former union. The mass movements of populations in the region during the twentieth century contributed to the use of almost all the languages spoken in the SU on the

territory of the new RF as well. The value of Nettle's linguistic diversity index in the RF is therefore almost twice as large as it used to be during the SU. This is so despite the fact that the size and the power of the dominant group of Russian speakers in the RF is much larger than in the SU.

The indices discussed in section 3.2.2 avoid such problems, and are used in equations that link the possible effect of fractionalization or polarization and economic or sociological outcomes y such as growth, redistribution, provision of public goods, corruption and conflicts. The standard equation reads:

$$y = \alpha IND + \sum_k \beta_k z_k + \epsilon,$$

where IND represents any of the indices described above, z_k , $k = 1, 2, \dots, m$ are control variables, α and the β_k s are parameters to be estimated, and ϵ is an error term. The parameter of interest is α , since its sign informs us whether fractionalization or polarization are positively or negatively correlated with outcome y . Whether IND "causes" y is debated, though the content of IND (Greenberg's A or B index) is based on history and predates the measurement of y , so that, if there is causality, it can hardly go from y to IND .⁵⁹ In almost all cases, the equation is estimated using cross-sectional data, often countries, and the dichotomous A -index. Experiments carried out by Desmet, Ortuño-Ortín, and Weber (2009) show that the B -index performs better.

3.4.1 Fractionalization

Impact of fractionalization on growth.—Most papers (essentially written by economists) assert that linguistic fragmentation has an adverse impact on economic development and growth. Easterly and Levine

(1997), who coined the "Africa's growth tragedy" expression, point to the negative impact of diversity on economic growth. They compare African and East Asian 1969 to 1990 growth rates and argue that about 40 percent of the annual growth differential of 3.5 percent between the two regions can be attributed to the effects of fractionalization. Easterly and Levine's conclusions were reinforced by the cross-country analyses run by Annett (2001) and Alesina et al. (2003), who show that fractionalization generates political instability and excessive government consumption or reduces the quality of government that may, in turn, slow down growth. Collier, Honohan, and Moene (2001) and Alesina and La Ferrara (2005) find that the adverse effect of diversity on growth is mitigated in democratic societies, while, according to Lian and Oneal (1997) it remains strong under dictatorships. Laitin and Ramachandran (2018) further show that, in addition to linguistic diversity, linguistic policies that hold back colonial languages⁶⁰ played an important role in the failures supported by postcolonial Africa.

Campos, Saleh, and Kuzeyev (2011) detect no significant correlation of diversity and economic growth in 26 former communist countries of Eastern and Central Europe, Central Asia, and Mongolia between 1989 and 2007, that is after the collapse of the Soviet Union. They however make the important observation that while the absolute values of the fractionalization levels are not that important, their dynamic fluctuations definitely are. This conclusion should be taken up in further examining the impact of the dynamics of ethnolinguistic diversity and shifts in disenfranchisement.

Impact of fractionalization on other outcomes.—Mauro (1995) finds that

⁵⁹See also section 5 for further discussion.

⁶⁰This is further discussed in section 4.

ethnolinguistic fractionalization increases institutional inefficiency, arguing that multiple competing groups tend to advocate for their interests, which increases the amount of lobbying and leads to inefficiency and corruption. La Porta et al. (1999) find that fractionalization has a negative impact on various public services and goods, education, and infrastructure. Alesina, Baqir, and Easterly (1999) point out that ethnically fragmented communities run larger deficits and spend less on basic public goods, including provision of education, roads, and sewage systems.⁶¹ Merkel and Weiffen (2012) offer empirical evidence that fractionalization complicates democratic consolidation. Alesina, Michalopoulos, and Papaioannou (2016) stress the importance of both linguistic diversity and economic inequality, which are correlated with regional development. Desmet, Ortuño-Ortín, and Weber (2017) focus on the ethnolinguistic tension between the center and periphery and show that higher levels of peripheral diversity, represented by the *CP*-index, are associated with providing less public goods, but more transfers. Meanwhile, higher levels of overall ethnolinguistic diversity detected by the *B*-index have a negative effect on transfers.

Though cross-country results that associate fractionalization and economic outcomes are often negatively correlated, the conclusions are different at city or firm level, where diversity could, on the contrary, be a driving force for progress. The success of Silicon Valley in the late 1990s is often attributed to the diversity of backgrounds of foreign scientists, engineers, and entrepreneurs who landed in California from many parts of the world. Saxenian (1999) points out that over 30 percent of businesses in the Valley have a foreign-born cofounder. Lazear (1999)

and Prat (2002) argue that to be successful, teams should show some cultural and linguistic diversity. Florida (2002), and Florida and Gates (2001) show that metropolitan regions with a higher degree of diversity in terms of education, cultural background, sexual orientation, and country of origin correlate positively with a higher level of economic development. Ottaviano and Peri (2005) investigate whether and how linguistic diversity affects wage rates in 160 US cities. They find that more diversity is often associated with higher productivity and hourly wages.

3.4.2 Polarization

Hibbs (1973) was the first (political) scientist who made use of an *A*-index in his study of mass political violence. But later on, Fearon and Laitin (2003) as well as Collier and Hoeffler (2004) gave some evidence that the *A*-index did not “predict” well enough the likelihood of civil conflicts. Montalvo and Reynald-Querol (2002, 2005) show that their polarization index (*RQ*) performs in a better way and that larger values of *RQ* point to longer civil conflicts.

In their assessment of the Sri Lankan civil war, Castañeda Dower, Ginsburgh, and Weber (2017) make use of a polarization measure that includes historical factors driven by different patterns of English language acquisition during the colonial era. They use an *RQ* distance-weighted index for each Sri Lankan district *j* that takes the form $D^j = \sum_{k=1}^K \sum_{i=1}^K (n_k^j)^2 n_i^j d_{ki}$, where n_k^j and n_i^j denote the fraction of linguistic groups *k* and *i* in district *j*, and d_{ki} is the linguistic distance between *k* and *i*. Distances take into consideration the changes in English proficiency during the precolonial period. The authors find that larger shares of English speakers resulting from colonial times are positively correlated with the number of victims from terrorist attacks in all regions of the country.

⁶¹ See also Kuijs (2000), who confirms these results.

3.4.3 Disenfranchisement

The best-known example and the rich data resources that can be used to compute linguistic disenfranchisement are provided by the European Union, which carries 24 official languages chosen by 28 countries,⁶² including English. Though today, it still is the most widely spoken language with 37 percent of Europeans knowing it well or very well, changes are expected. Some 60 million Britons will have left the European Union, while German and French will gain in importance (see Ginsburgh, Moreno-Ternero, and Weber 2017).

Maintaining 24 languages costs some \$1.2 billion per year in translation and interpretation, but it is impossible to reduce their number since Regulation no. 1, adopted by the European Council in 1958, states that (one of) the official language(s) of a country joining the European Union becomes an official language at the EU level. Since decisions on languages have to be unanimous, no country would agree to delete its language from the list, though Luxembourg accepted foregoing Luxembourgish when the Common Market was created in 1958.

Ginsburgh and Moreno-Ternero (2018) suggest that the European Union could subsidize⁶³ each member country to encourage the learning of a common language,⁶⁴ though each country would keep its official language

in its own territory and in the European Union. The proposal is meant to facilitate communication among all Europeans, and if possible, strengthen the ailing union. It is based on the responses to the Special Eurobarometer (2006) survey commissioned by the European Commission,⁶⁵ which shows that the reactions of Europeans are quite positive. Here are the statements and questions that EU citizens had to answer, as well as their reactions:

- (i) The European institutions should adopt one single language to communicate with European citizens. *Answer:* In 16 out of 27 countries, 50 percent or more voted in favor of a unique language, with an aggregate EU result of 54 percent.
- (ii) Everyone in the European Union should be able to speak a common language. *Answer:* In all countries, with the exception of Bulgaria, 50 percent or more voted yes, with an aggregate EU result of 69 percent.
- (iii) Everyone in the European Union should be able to speak one language in addition to their mother tongue. *Answer:* In all cases, the idea was adopted by more than 70 percent. Aggregate result: 83 percent.
- (iv) Which two languages, apart from your mother tongue, do you think are the most useful to know for your personal development and career?⁹ (any 2 of the 24 official languages could be selected) *Answer:* English (67 percent), French (25 percent), German (22 percent), Spanish (15 percent). These are followed by Russian (3.4 percent), Italian

⁶²Each country can choose one official language, but some languages are official in more than one country, like German in Germany, Austria and Belgium, French in France and Belgium, Dutch in the Netherlands and Belgium.

⁶³As will be shown in section 4, subsidies are often needed to “persuade” citizens to acquire a new language.

⁶⁴They also compute how much each country should receive, using two adjudicating rules (Aristotle’s proportional rule and the so-called Talmud rule) to satisfy the claim of each country. The claim is assumed to be proportional to the number of young individuals (12–26 age bracket) who do not speak the chosen subsidized lingua franca, adjusted for the lexical distance between the official language of the country and the lingua franca.

⁶⁵This survey was analyzed in Fidrmuc, Ginsburgh, and Weber (2007).

(3.2 percent), and Chinese (1.5 percent). Beyond that, usefulness drops to less than one percent.

These answers show that 83 percent of EU citizens accept the idea that a common language would be useful, though it may just be wishful thinking. Their most cited choices are English (67 percent; note that the survey predates the Brexit decision), French (25 percent), and German (22 percent). This, however, would by no means imply that countries lose their identity and home language, and the current translation and interpretation system would remain as it is.

In what follows, we use disenfranchisement indices to check which language should be chosen as lingua franca within a reasonable number of years. In this exercise, we make the assumption that the whole population would be as fluent as today's 15-to-29-year-old generation, which is much more fluent than their elders.⁶⁶ Table 4 contains the results of the calculations for two indices, *DN* and *NN*.

The first three columns show *DN*-indices by country and for the whole European Union if English, German, or French were chosen as lingua franca. As is obvious, English would be the first language to choose. It is not known to 44.5 percent of the population that would have to be taught the language (but in many countries, especially in Eastern Europe and France, this percentage is much larger), and ends up being less expensive to implement than German (73.9 percent) and French (77.8 percent). And this will still be so after Brexit, though some 60 million English speakers will have left the European Union: The 44.5 percent of pre-Brexit disenfranchised citizens

increases to 50.5 after Brexit, but this is still much lower than the disenfranchisement rates generated by German (70.1 percent) and French (74.5 percent).

The last three columns give the values of *NN*-indices, which are in all cases lower than their *DN* counterparts that are corrected for the (lexicostatistical) distance between the chosen lingua franca and the one spoken in each country. But English will no longer be the best language to teach. It will be overtaken by French, which is much closer to Italian and Spanish, the official languages in Italy (58.5 million inhabitants) and Spain (43 million) and will obviously not need extra teaching in France (60.6 million inhabitants).

3.5 Foreign Language Acquisition

Learning is obviously more difficult if the acquired language is distant (say, Chinese for an English speaker) than if it is close (German for the same English speaker), but it is also driven by other factors that may be more important than linguistic distances such as trade relations with foreign countries and the decision of some large countries (France, Germany, Great Britain, and Spain) to promote their native language in foreign countries via their publicly financed language institutes, like the Alliance Française,⁶⁷ the British Council,⁶⁸ the Goethe Institute for

⁶⁶ Many more young people of that generation are more fluent than the rest of the population, and here we make the assumption that within 20 to 30 years the whole population in each country would be as fluent as the 15–29 year olds are today.

⁶⁷ Created in 1883 under the name of French alliance for the propagation of the national language in the colonies and abroad. Nowadays, it has a network of over one thousand schools in France and in 135 foreign countries, and teaches French to 450,000 students

⁶⁸ In his introduction to Graddol's (2006) report on the situation of English in the world, Lord Kinnock—member of the House of Lords in the United Kingdom, former vice president (1999–2004), and former member of the European Commission (1995–99)—mentions that “the English language teaching sector directly earns £1.3 billions for the UK as invisible exports and our education related exports earn up to £10 billion a year more.”

TABLE 4
DN- and NN-DISENFRANCHISEMENTS BY COUNTRY IN THREE LANGUAGES
 (IN PERCENT OF THE TOTAL POPULATION IN EACH EU COUNTRY)

	<i>DN-index</i>			<i>NN-index</i>		
	English	German	French	English	German	French
Austria	41	1	85	17	0	64
Belgium	39	90	25	19	36	11
Bulgaria	57	87	95	44	67	75
Croatia	34	76	99	26	58	77
Cyprus	18	98	91	15	80	77
Czech Republic	64	80	98	48	60	75
Denmark	9	68	98	4	20	75
Estonia	33	85	100	33	85	99
Finland	29	95	97	29	95	97
France	67	95	0	51	72	0
Germany	38	1	88	16	0	67
Greece	40	93	94	33	75	79
Hungary	76	82	99	76	82	99
Ireland	2	94	85	0	40	65
Italy	54	94	85	41	69	17
Latvia	55	96	99	44	76	79
Lithuania	49	93	99	39	72	77
Luxembourg	50	8	3	0	0	0
Malta	10	99	92	10	99	92
Netherlands	11	59	88	4	10	67
Poland	57	83	97	43	62	76
Portugal	62	99	87	47	74	25
Romania	69	97	82	53	73	34
Slovak Republic	57	66	98	43	49	75
Slovenia	22	72	97	16	52	76
Spain	65	98	92	50	73	24
Sweden	5	89	96	2	27	73
United Kingdom	2	95	90	0	40	68
Pre-Brexit European Union	44.5	73.9	77.8	31.4	49.2	33.7
Post-Brexit European Union	50.5	70.1	74.5	35.5	50.3	30.9

Source: Own calculations based on Fidrmuc, Ginsburgh, and Weber (2007).

Germany,⁶⁹ and the Instituto Cervantes for Spain.⁷⁰

⁶⁹There exist 149 Goethe Institutes in 49 countries; 126 offer linguistic education (Uebelmesser, Huber, and Weingarten 2018).

⁷⁰With 90 centers in 44 countries.

Most of the economic literature on language learning is devoted to migrants who have to acquire the language of their destination country to be integrated and eventually find a job—this is discussed above—and to the development of “new” global languages that replace eighteenth-century-used Latin, as well as French and

German.⁷¹ Note that this issue was already studied in the late nineteenth century by De Candolle (1987, [1873], pp. 298–299) who predicted that English would become the new lingua franca: “The coming preponderance of Anglo-American is obvious; movements of population in both hemispheres will impose their language. Here is an easy proof, using a very small number of words and data...” Canadian economists were among the first to become interested in this topic, since there are serious linguistic issues involving English and French in their country.⁷² Sociolinguists and language educators, whose work is generally more qualitative, also made important contributions, some of which are discussed in section 4.

A more encompassing approach was suggested by Selten and Pool (1991) in a game-theoretic model of communicative benefits, where the utility of an individual increases with the number of those she can communicate with by using a common language. This, in turn, can offer new cultural, social, and economic opportunities. Church and King (1993) used a simplified Selten and Pool set-up for a two-languages two-populations (regions or countries) model. Their assumption of homogenous learning costs (or aptitudes to learn) in each country yields corner solutions only, where one of the two populations learns the other language. Lazear (1999) and Gabszewicz, Ginsburgh, and Weber (2011) introduced heterogeneous aptitudes (or costs) of learning in the Church and King (1993) framework. This leads to the existence of interior Nash equilibria and closed-form solutions: Both populations learn, to some extent, the nonnative language, and the

share of learners depends on the number of individuals in each population and the cost of learning. These formulations lead to possible inefficiencies given that the equilibrium levels of learning fall below the socially optimal levels. Given the network externalities, some individuals in one community rely on members of the other community to learn. This implies that subsidizing learning could be essential to improving the common social welfare.

An alternative approach by Armstrong (2015) is based on asymmetric information. It relies on the fact that acquiring the other language is costly, and that the command of the other language may provide employers with a verifiable signal of the skills of a bilingual employee. In this model with heterogeneous productive skills, the utility functions of individuals are represented by their wage net of the language learning cost. Here, the welfare effects of linguistic policies, including subsidization of learning costs, could be ambiguous since they depend on market conditions and the wage structure.

The Selten and Pool (1991) model with communicative benefits is taken to data by Ginsburgh, Melitz, and Toubal (2017), who estimate an equation that considers the learning decisions of 13 of the most widespread world languages⁷³ by individuals who live in 193 countries. The equation reads:

$$x_{cj} = \alpha + \beta N_i + \gamma N_j + \delta d_{ij} + \tau T_{cj} + \phi I_c,$$

where c is a country whose native language is i while j is the target language of learners. The dependent variable x_{cj} is the share of the population in country c that learns language j .

Five factors are assumed to influence learning, three of which are: N_i , the world

⁷¹ See Crystal (2003), Graddol (2006), and Phillipson (1992, 2003) for English, and Maurais and Morris (2003) for English and other languages.

⁷² See, among others, Bloom and Grenier (1992) and the volume edited by Vaillancourt (1985).

⁷³ Chinese, English, Spanish, Arabic, Russian, French, Portuguese, German, Malay, Japanese, Turkish, Italian, and Dutch, in descending order of number of speakers.

population of speakers of language i (expected negative effect, since a native language spoken by many should lead to less learning); N_j , the population that speaks j (expected positive effect, since a language spoken by many attracts more people to learn it); d_{ij} , the linguistic distance between languages i and j (with a negative effect). This is very close to the gravity equation discussed earlier. There are two important additional variables: T_{cj} , the ratio of the total trade of country c with the j -speaking world (expected positive effect, since more trade should increase learning)⁷⁴ and I_c , the literacy rate in learning country c (expected positive effect). The trade between country c and acquired language j obviously introduces a reverse causality issue, since the acquisition of the language in this country may promote trade with those who speak it.⁷⁵ Thus, T_{cj} has to be instrumented.

The instrument used is borrowed from Frankel and Romer (1999), who faced a similar problem in their regressions of economic growth on trade ratios. Their solution was to instrument trade ratios on predetermined variables such as national land area, landlocked status, common border, geographic distance, and population size. This choice was criticized by Brock and Durlauf (2001, pp. 237–238) and Rodriguez and Rodrik (2001, p. 264) since the instruments are correlated with the dependent variable (growth rates), and therefore also with the omitted growth factors in the right-hand side of the equation. This is not the case here, because the instruments are unlikely to be

correlated with shares of populations that learn foreign languages.

Since there are many cases of no or very little learning (Finns do hardly learn Malay), the 2,300 observations on learning contain many zeros. Therefore, the authors use two estimation procedures: probit and IV probit for the full sample, ordinary least squares (OLS) and two-stage least squares (2SLS) for the much smaller sample of 240 observations of positive learning. The results show that the coefficient picked up by linguistic distances always has the expected (significantly negative) sign, and so do the populations of the acquired language (significantly positive in the full sample) and the native language (significantly negative). The effect of literacy is also positive, but not significantly so in the reduced sample.

Trade with countries that speak the target language always has a significant positive effect at the one percent probability level and exerts an important effect on learning. In the 2SLS estimation with 240 observations, a 10 percentage-point increase in the ratio of trade with native speakers of the destination language raises learning of this language by 14 percentage points.

The interesting effect that can be derived from the model and its estimated coefficients is that the incentives to learn English are similar to those of learning other languages. Since trade has a sizable effect on learning, growth in, say, Arabic–English trade, should promote the learning of Arabic in countries where English is the native language, as well as the learning of English in Arab countries. Which language will gain the battle is, however, difficult to predict. But the demographics of larger birth rates in Arabic and Spanish-speaking populations will prompt more people to learn these two languages, while the Arab and Spanish-speaking populations will learn fewer foreign languages. This goes in the same direction as Graddol's (2006) prediction that the number of speakers of Arabic and Spanish will increase.

⁷⁴Actually, the numerator of this ratio consists of a weighted sum of trades with countries that speak language j . The weights are shares of populations which know language j in these countries. The denominator is the population of country j .

⁷⁵Note that this may introduce reverse causality in trade equations that use acquired languages as right-hand side variables.

3.6 *Financial Returns to Language Learning*

Most papers on returns to language learning are concerned with immigrants who acquire the language of the host country for assimilation and/or job-enhancement purposes.⁷⁶ Several papers also look at the issue of nationals whose objective is to acquire foreign languages that are of use at the workplace.⁷⁷

The main issue here is to estimate the effect of language proficiency on wages, but this leads to the well-known problem that unobserved heterogeneity affects the attempt to estimate returns to languages as it does in the returns to education literature, where both education and earnings may depend on unobservable individual skills and talent, and the right-hand side variables may be correlated with the error term in the equation, requiring thus instrumentation techniques. See Chiswick and Miller (1995) for further details.

Results of the many studies on returns to languages for migrants are reported and summarized by Chiswick and Miller (2014). They vary between 5 and 35 percent, depending on data sets, source and destination countries and languages, and on gender. Adserà and Chiswick (2007) and Ginsburgh and Prieto-Rodriguez (2007) use the same database collected by the European Community Household Panel Survey 1994–2000 to study

immigrants and nationals respectively. Still, earnings differ greatly across countries, but in both cases, knowing the language of the immigration country or speaking and using a foreign language at the workplace have positive returns.

Within the same country, returns to the same language may differ across groups. Levinsohn (2007) finds that, in South Africa, returns to speaking English increased between 1993 and 2000 for whites but not for blacks. In Kazakhstan, Kazakh and Russian are both official languages, but returns to Russian are larger for people who *do not* speak Kazakh than for those who speak both languages. The reason seems to be that those who speak Russian and Kazakh signal that they attended a Kazakh school and not a Russian school with better schooling quality (Aldashev and Danzer 2014).

4. *The Multilingual World: Solutions and Policies*

The relevance of linguistic policies in many multilingual countries, regions, and international bodies has always been and still is a hot topic. Recognizing that the multiplicity of languages inevitably requires, explicitly or implicitly, some degree of standardization is an element of Max Weber's (1968, [1910]) rationalization theory. The requirements of calculability, efficiency, predictability, and control over uncertainties lead to creation of a common legal system and, hence, a common language used for administrative purposes. As Jain (2017, p. 475) indicates, "using uniform languages can lower the cost of communication, facilitate education and expand economic growth. However, attempts to impose official languages can meet considerable resistance, both because language is an important component of identity, and because learning a new language is difficult."

⁷⁶Australia (Chiswick and Miller 1995); Canada (Abbott and Beach 1992, Aydemir and Skuterud 2005, Chiswick and Miller 1995); Germany (Dustmann and Van Soest 2002); Israel (Beenstock, Chiswick, and Repetto 2001, Berman, Lang, and Siniver 2003, Chiswick and Miller 1995, Chiswick 1998); the United Kingdom (Leslie and Lindley 2001); and the United States (Bleakley and Chin 2004, Bratsberg, Ragan, and Nasir 2002, Chiswick and Miller 2002, Hellerstein and Neumark 2003).

⁷⁷Canada (Shapiro and Stelcner 1997), the European Union (Williams 2006, Ginsburgh and Prieto-Rodriguez 2007), Hungary before joining the European Union (Galasi 2003), Switzerland (Cattaneo and Winkelmann 2005), and the United States (Fry and Lowell 2003).

Linguistic rationalization and standardization can be implemented in different ways. One possibility is choosing the language of the majority group: French in France, Han Chinese in China, or Kuotsugo Japanese in Japan. Another route is to accept a lingua franca, which may also be a language spoken by the majority of the population without being the mother tongue of any large group in the society. This was the case of Swahili in Tanzania, and in fact in most East African countries, Bahasa in Indonesia, or even English in the United States.⁷⁸

Challenges remain important even in the case of two languages, one that is native and the “other.” In terms of education, some former colonies opted for native-language instruction in public schools (Morocco, Malaysia, Pakistan, and India) while others kept going with the colonial language (much of sub-Saharan Africa and the Philippines). Native-language instruction may indeed reinforce national identity and make schooling more accessible. But, since top jobs in government and business often continue to use the colonial language, native-language instruction may reduce economic opportunities for the poor.

In addition, it should come as no surprise that it also generates strong emotions. This is made clear by a unique sentence in Fernando Pessoa’s *Book of Disquiet*: “Minha pátria é a lingua portuguesa,” my homeland is the Portuguese language. As section 2.3 shows, being deprived of or persecuted because of one’s native language is difficult to accept. There exist many attempts to bring about the benefits of standardization, but the feeling of being disenfranchised must be respected. Though some policies, such as the so-called *three language formula* experimented with in India, Kazakhstan, and Nigeria or the 24 official language policy in the European Union,

respect native languages, their success was and still is, at best, mixed (Laitin 1994 and Ginsburgh and Weber 2011).

Many papers and books on linguistic policies have been written. It is probably the largest subject treated by sociolinguists. Spolsky’s (2004, 2012) *Language Policy and Handbook of Language Policy* give excellent overviews and many detailed descriptions of most issues, past and present. The examples in section 4.1 restrict the discussion to a few cases of language-shifting policies, for which the consequences of a reform on labor markets were observed and studied by economists. In this section we focus on dichotomous policies that entail a major shift of education from one language to another. Section 4.2 is devoted to the various policies that balance education in different languages. The formula failed in India and Nigeria, but seems to have succeeded in Kazakhstan.

4.1 Language-Shifting Policies

Linguistic policies and the changes they bring about indeed have effects on educational systems and labor markets. Angrist and Lavy (1997) point out that the “promotion of native languages suggests that language is an important symbol of independence and national sovereignty. Governments appear to be willing to tolerate the many social costs of a language transition in return for the perceived social benefits.” The effects of such transitions are of interest given the economic value of the human capital embodied in language skills, but even neglecting emotions, outcomes are unclear. School quality and the economic returns to schooling could either increase or decrease after a change in the language of instruction. They may also increase for some citizens but decrease for others.⁷⁹ In most examples that follow, the

⁷⁸See Laitin (1989, 1994) and Ginsburgh and Weber (2011).

⁷⁹The possibility that individuals with better language skills may earn more for reasons other than those skills is

short-run consequences indicate that language switches have limited influence on the performance of students.

Aspachs-Bracons et al. (2008) study the impact of the change in linguistic education in Catalonia and the Basque Country in 1983, a time in which Catalan and Basque⁸⁰ were given the status of official languages in addition to Spanish. Teaching Catalan or Basque to all children became the norm in the two provinces. It turns out that, in Catalonia, the new system had a much deeper effect on the formation of regional identity, since the reform was compulsory. In the Basque Country parents could choose the language in which they wanted their children to be educated. To strengthen the point about the impact of identity in Catalonia, Clots-Figueras and Masella (2013) show that those who were exposed to Catalan for a longer period have stronger Catalan feelings, even if their parents were not of Catalan origin.

In 1983, Morocco switched from French to Arabic as its language of instruction in grade six and above. Angrist and Lavy (1997) estimate the effect of French language skills on test scores and earnings and suggest that the elimination of compulsory French instruction led to a substantial reduction in the returns to schooling for those affected by the change. This reduction appears to be largely attributable to the loss of French writing skills. Students must fit into the existing economic and social system dominated by the French language and culture, but at the same time, the insistence of instruction in a foreign language constitutes a barrier for poor or rural students who seem to learn more effectively when schooling is given in their native language. Strong native language skills may prove to be of more enduring

value in local labor markets than French language skills. In the short run, however, the language reform contributed to the popular impression that Moroccan schools “are now turning out bilingual illiterates.”

In section 2.3, we already mentioned the even more dramatic shift of the medium of education from French to English in Rwanda in 2008. After the 1994 genocide, the return from Uganda of the anglophone Tutsi refugees and the influence of the United States, English has become more and more used by the population and the administration. Following the reforms, children now begin their study in English starting in grade one and take their secondary school entrance exams in grade six in English. The change, which was an important part of social and education reforms, also aimed at increasing Rwanda’s participation in global and regional economic development, especially with other members of the East African Community, which consists mostly of English-speaking neighboring countries such as Uganda, Kenya, and Tanzania (Samuelson and Freedman, 2010), as well as South Africa.

Eriksson (2014) examined the impact of the 1955 Bantu Education Act in South Africa, which required that the local Bantu mother tongue should be taught for six instead of four years, replacing English or Afrikaans in subsidized South African schools. Using the 1980 South African census, Eriksson estimates a difference-in-difference model and finds that the two years difference generated positive effects, contrary to what happened in Morocco. The effects could even have been larger in the absence of labor market discrimination against blacks under apartheid.

The former Soviet Republic of Latvia implemented a similar reform in 2004. Ivlevs and King (2014) examine the other side of the coin, namely the educational level in Russian-language schools. They estimate the effect of the switch from Russian (now

a particular concern in the labor economics literature, and are often difficult to ascertain.

⁸⁰As well as Galician.

a minority language) instruction to a composite of 60 percent Latvian and 40 percent Russian. Using data from centralized exam results for all Latvian secondary schools between 2002 and 2011, they observe that a significant deterioration of performance took place in minority schools during the early years following the reform.

In Spain and South Africa, bilingual or multilingual programs were implemented to protect ethnolinguistic minorities, while Morocco, Kazakhstan,⁸¹ and Latvia share the common feature of introducing the language of the ethnolinguistic majority that aimed to protect itself. South Africa and Kazakhstan subsidized the switch. The comparison seems to imply that introducing languages spoken by minorities to protect them is almost a necessary condition to make a reform successful, but subsidization may also be needed. Note also that the Bantu Education Act was not imposed in response to popular demand. This distinguishes it from the policy changes in Morocco and Puerto Rico that were driven by extensive national debates.

An interesting example of a shift and its reversal is given by the experience of the Indian state of West Bengal, where, in 1983, the communist government abolished English teaching at the primary level in public schools. Roy (2005) argues that this lowered academic standards, and despite its avowed objective of making education more accessible, there is no evidence of positive effects, even on the poorest income quartiles. In fact, Chakraborty and Kapur Bakshi (2008) point out that the effect of this policy was negative. They show that a 1 percent decline in the probability of learning English reduced weekly wages by 1.6 percent and that a substantial part of this decline was due to the policy change. Ironically, the program

increased the wage gap that it was supposed to close.

Jain (2017) investigated the impact of official language policies on education using state formation in India. Colonial provinces consisted of some districts where the official language matched the district's language and some where it did not. Linguistically mismatched districts experienced 18.8 percent lower literacy rates and 27.6 percent lower college graduation rates. Educational achievement caught up in mismatched districts after the 1956 reorganization of Indian states on linguistic lines, suggesting that political reorganization can mitigate the impact of mismatched language policies.

Lleras-Muney and Shertzer (2015)⁸² estimated the effect of US statutes requiring English as the language of instruction and compulsory schooling laws on school enrollment, literacy, and English fluency of immigrant children during the 1910–30 Americanization period. It turned out that the English-only statutes moderately increased literacy rates of certain foreign-born children, particularly of those living in cities or whose parents were not fluent in English. The laws had no impact on immigrants' labor market outcomes or measures of social integration.

And time does not help either, as the coming example shows. In 1898, Spain lost the war against the United States. Puerto Rico became American, though not yet an American state, and the language of instruction became English in most post-primary grades in public schools. Education switched back to Spanish 50 years later (in 1948–49). Though American policy makers claimed that English instruction increased English-speaking ability among Puerto Rican natives, Angrist, Chin, and Godoy (2008) show that the conclusion was supported by

⁸¹ See section 4.2 for further details.

⁸² See also Lleras-Muney (2002), Goldin and Katz (2011), and Clay, Lingwall, and Stephens (2012).

“naïve estimates” and that “more serious estimates” found no effect.

We now turn to linguistic policies that encompass simultaneous use of several languages in a country.

4.2 *Balanced Linguistic Policies*

One of the remarkable examples of a linguistic policy aimed at balancing efficiency, national pride, fairness, and equality of economic opportunity was introduced in India in 1965–66 and adopted by the parliament in 1968. It followed another recommendation made just after the country’s independence in 1948–49 by the University Education Commission, which was called the three-language formula (English, Hindi, and the local tongue), making Hindi the language of the government, while hoping that it would eventually replace English.

The 1965–66 recommendation, also called the three-language formula,⁸³ was initiated as a national response to bitter complaints against the previous one by Tamil Nadu and other Southern states. These regions claimed that the use of Hindi in government services imposed formidable barriers since they were required to become proficient in two nonnative languages, English and Hindi, whereas speakers of Hindi had to learn English only. Therefore, the new formula, which varied across states, required children in Hindi-speaking states to study Hindi, English, and one of the southern languages, whereas children in non-Hindi speaking states were supposed to learn their own regional language, Hindi and English. This masterful and well-crafted formula that seemed to foster group identity, preservation of mother tongues and traditions through the study of regional languages, national pride and unity by acquiring Hindi, and administrative efficiency and standardization

by using English, was not successful either. In addition to the inadequate and lukewarm backing by the regional administrations, the formula failed to generate wide public support both in the North and the South. In Hindi regions relatively little effort and resources were spent on studying English and even less so in learning other languages. There was more enthusiasm for English and Tamil in Tamil Nadu, but almost no interest in acquiring Hindi. The lack of public commitment and of resources needed to implement the recommendation caused its failure.

It is worth pointing out that several variants of the three-language formula were experimented with in Nigeria. There, “official regional” languages Hausa, Igbo, and Yoruba were suggested and were considered, like in India, a unifying device (Laitin 1986). The idea was killed by the same challenges that slowed down and often stopped the implementation of the formula in India: No government resources and lack of commitment among students, their families, and regional authorities.

Language policy in the Soviet Union that evolved over the years is also worth considering. The Soviet society was viewed as a new type of a super-ethnic unity after the Great Patriotic War (June 1941–May 1945). The convergence and fusion of nationalities was officially declared as a step in building the communist society and was adopted in 1962 (Kadochnikov 2016). The program guaranteed the development of all nationalities and languages and called for the study of Russian and other native languages. In practice, it led to the introduction of a variant of the three-language formula: Russian, a regional language, and English for international communication.⁸⁴

An interesting example of the recent implementation of that formula is Kazakhstan,

⁸³See Laitin (1989).

⁸⁴Note that English had already replaced German as international language after World War II.

the last Soviet republic to declare its independence from the Soviet Union in 1991, with its two large ethnic groups, Kazakhs and Russians. Seventy-four percent of the population of 12 million speak, or at least understand, Kazakh, while 94 percent speak, or at least understand, Russian (Smailov 2011). Under the constitution adopted in 1993, Kazakh became the “state” language, whereas Russian, spoken by most Kazakhs, was declared an “official” language, used routinely in business, government, and interethnic communication. Using two newly assembled data sets, Aldashev and Danzer (2014) uncover negative returns to speaking Kazakh and a negative wage premium in the first years of Kazakh independence. In spite of the official support of the language at that time, scholastic achievements were substantially lower for pupils taught in Kazakh, a probable consequence of the comparatively poor quality of schools that taught in Kazakh. It is important to point out that changing the official state language without appropriate investments is unlikely to cure the economic disadvantage of a previously marginalized language. However, in 2011 a presidential decree introduced a three-language formula that, in addition to Kazakh and Russian, included English as a “world” language. Contrary to India and Nigeria, Kazakhstan committed considerable resources to improving infrastructure, gave incentives to learn English, raised demand for all three languages in government programs, and preserved linguistic diversity. This required upgrading the low level of education in Kazakh schools during the first years of independence (Aldashev and Danzer 2014). An important factor that contributed to the success of the formula was the public willingness to buy into a program that strengthened the role of the national Kazakh language and preserved the current and future role of Russian as a vehicle of regional communication and support for social stability in

the country. It also highlighted the role of English as an indispensable factor of international communication and business cooperation. The Kazakh level of commitment to the three-language formula makes its prospects more promising than in India and elsewhere.

It is quite easy to notice that, in most cases, the application of extensive and balanced linguistic policies leads to levels of learning that exceed socially optimal levels.⁸⁵ This welfare cost can be considered the price to pay for national cohesiveness and cultural convergence, which should be followed by labor mobility and more efficient labor markets. This relates to the issue of reducing and eliminating disenfranchisement and to support the argument that including all population groups in the political and economic process is essential for a country’s progress in times of growing competition.

5. *Conclusions, Problems, and Further Research*

Though Marschak’s (1965) paper triggered the interest of economists, they focused essentially on linguistic policies and reforms. Much later, at the end of the twentieth century, the impact of languages on economic and social outcomes became a field of study that borrowed many ideas from linguistic theory and ethnolinguistics. To conclude, we discuss issues that deserve to be considered during the years to come. These are devoted to the identification of groups, as well as to some econometric problems, relativism versus universalism, and other research issues.

5.1 *Identification of Groups*

As discussed in section 2.1, the identification of groups in a nation is often based on language, since this is a characteristic that

⁸⁵ See Church and King (1993); Gabszewicz, Ginsburgh, and Weber (2011); and Davydov, Shapova, and Weber (2018).

is reasonably easy to observe. Still, as Laitin (2000) and Fearon (2003) suggest, language may not be sufficient to describe all aspects of a person. Alesina et al. (2003, tables 13, 14, and 15) computed fractionalization indices based on linguistic, but also on ethnic and religious diversity in some 190 countries. They found that these measures may provide different results:

Effects of Fractionalization on Growth.—The effect of ethnicity is negative though it fails to be significantly different from 0 at the 5 percent probability level in the presence of some additional controls such as financial depth, black market premium, and number of telephones. Religious diversity has no significant effect, while more linguistic diversity produces the same negative result as usual.

Effect of Fractionalization on the Quality of Government Policies and Institutions.—Ethnic diversity has varying (often not significant) effects on most indicators of government quality (business climate, corruption and bureaucratic quality, taxation, size of the public sector, size of government, provision of public goods, schooling and literacy, political rights).

The challenge would be to examine the combined effect of several dimensions on economic outcomes. This is discussed in some recent research by Desmet, Ortuño-Ortín, and Weber (2017), Davydov and Weber (2016),⁸⁶ and Desmet, Ortuño-Ortín, and Wacziarg (2012).

The very promising paper by Desmet, Ortuño-Ortín, and Wacziarg (2017) explores a new type of diversity based on cultural norms, values, and attitudes,⁸⁷ using surveys conducted by the World Values Surveys

(2009) in 76 countries between 1981 and 2008. They construct *cultural* fractionalization indices that assume the same functional form as the usual ethnolinguistic fractionalization index, say $CF = 1 - \sum_{j=1}^J (1 - v_j)^2$, where v_j represents the probability that a randomly chosen individual disagrees with a specific answer j to a question. Thus, CF gives the probability that two randomly drawn individuals from a given population or society give different answers to the set of J questions raised in the survey.

Desmet, Ortuño-Ortín, and Wacziarg (2017) also define a new index χ^2 that depends on the average distance between the observed share of answers to question j by ethnolinguistic group k and the expected share if the distributions of groups and answers were independent. If χ^2 is small, there is little overlap: the cultural answers of each ethnic group are close to the country's average answers. This index thus combines both cultural and ethnic diversities.

Davydov and Weber (2016) discuss a family of indices (and the axioms needed to construct this family) that includes both Greenberg's A-index and Shannon's entropy index: $A^\alpha = 1 - \sum_{k=1}^K n_k^\alpha$, where α is a positive parameter different from one, and not necessarily equal to the commonly used value of 2. Each society i (city, region, or country) could be characterized by its own diversity index $A^{\alpha_i} = 1 - \sum_{k=1}^K n_k^{\alpha_i}$. Consider, say, Los Angeles (LA) and some midwestern town T. LA consists of a large number of various linguistic groups (Mexicans, Koreans, Chinese, Iranians, Armenians, and others), while T is relatively homogenous. If the same diversity index were applied to both localities, LA's diversity would be much larger than T's. However, if both indices were adjusted to societal attitudes, or to cultural frictions (Fujita and Weber 2010), the diversity level in T may exceed the one in LA.

⁸⁶See also Ginsburgh and Weber (2016).

⁸⁷See also Hofstede (1980, 2001) and Schwartz (1992, 2014) for earlier approaches.

This idea of this family of fractionalization indices is close to the one introduced by Esteban and Ray (1994). Both families are indeed characterized by a unique parameter α that belongs to a certain interval. Changing the value of α could change the results obtained by using these indices in regressions that try to relate ethnolinguistic diversity to, say, countries' growth rates or to the incidence, severity, and length of civil wars and conflicts. This happened indeed when Montalvo and Reynal-Querol (2002, 2005) showed that a variation of Esteban and Rey's (1994) polarization index worked for civil wars, while Feron and Laitin (2003) reported their failure with an A-fractionalization index. Using Davydov and Weber *DW* indices instead of A-indices to compute, say, the relation between countries' growth rates and a *DW* index could reverse the negative sign in a certain number of cases.

The choice of an appropriate value of α_i could and, in some cases, should be based on relevant surveys that encompass cultural norms, values, and attitudes (such as the World Values Survey), and made "endogenous" as do Desmet, Ortuño-Ortín, and Wacziarg (2012, p. 322) who suggest to "compute measures of diversity at different levels of linguistic aggregation and let the data inform us which linguistic cleavages are most relevant for a range of political outcomes, rather than making ad hoc choices." They use a linguistic tree that has a root and over 6,000 final twigs representing existing languages, and study different group structures that depend on the level of aggregation on the tree. Coarse linguistic divisions, obtained at high levels of aggregation, describe cleavages that emerged thousands of years ago, while lower levels generate finer partitions that emerged more recently.

The authors argue that cleavages may differ according to the research question (economic growth, patterns of redistribution, provision of public goods, conflicts, ...).

They find that high levels give better results for redistribution and conflicts, since they are tied to solidarity, empathy, and trust, which go a long way back in time, while low levels are needed for economic growth, which is linked to more contemporaneous barriers such as coordination and communication between agents. Therefore, in the research question at hand (redistribution, growth, etc.), the left-hand side variable y should be regressed on index *IND* calculated at different aggregation levels of the tree. The "best" fit should thus determine the choice of the appropriate value of *IND*.

5.2 Econometric Problems

The empirical econometric literature surveyed in sections 2.2, 3.3, and 3.4 is based on *association*, *coevolution*, or *correlation* between variables.⁸⁸ Scholars usually avoid the term *causality*, though sometimes it accidentally slips into their writings in different forms (such as "explains"), which may give the impression of "causes" and the not-very-careful reader could take it for granted that in a regression $y = \alpha + \beta x + \epsilon$, x *causes* y (high ethnolinguistic fractionalization in a country *causes* slow economic growth, existence or absence of future tenses in languages *causes* saving behavior). Here are some examples.

- (i) The title of section 2 in Easterly and Levine's (1997, pp. 1207–1208) landmark paper on Africa's growth tragedy reads "[u]sing cross-country regressions to explain growth," in which the word "explain" may resonate as "causes," but in the section itself, they add: "this section shows that many

⁸⁸Note that reverse causality can usually be excluded: Ethnolinguistic fractionalization obviously precedes GDP growth rates observed during the twentieth century, and the grammar of future tenses precedes saving behavior in recent years.

indicators [appearing in the right-hand side] have a close association with growth and account for a substantial amount of the cross-country variation in growth rates over the last 30 years,” which plays down the term *explains* to *associates*.

- (ii) Alesina and La Ferrara (2005, p. 772) are also tempted to point to causality by writing “the estimates suggest that *ceteris paribus*, going from perfect homogeneity to maximum heterogeneity [that is, changing the value of the fractionalization index they use from 0 to 1] would reduce a country’s growth rate by 2 percentage points per year,” and (p. 776) “in any case, neither of these studies argues that ethnic fragmentation is the only cause of ‘poor quality of government’: La Porta et al. (1999), for instance, argue that legal origins are at least as important.”
- (iii) In their abstract, Galor, Özak, and Sarid (2016) write that their “research explores the economic causes and consequences of language structures,” but, somewhat later in the paper (p. 15), they mention that “still, the results might be biased due to omitted variables, precluding a causal interpretation of the estimated coefficients.”

The real question is, however, whether we need to make causal statements that, as Angrist and Pischke (2009, p. 3) point out, are “useful for making predictions about the consequences of changing circumstances or policies; it tells us what would happen in alternative (or ‘counterfactual’) worlds.” We may be satisfied with observing correlation, and not necessarily causality though, even if there is only correlation, the estimated coefficient β may be strongly biased if x is

correlated with the error terms ϵ , and at worst, the sign may be wrong. Therefore, the computed negative correlation between growth and fractionalization may not even be negative and lead to a wrong conclusion.

Still, we often interpret such results as causal, and recommend or, in some cases, impose, for example, less linguistic diversity. This may lead to mixed results, or even make the situation worse at least during the transition to a new “equilibrium,” and provoke unrests or civil wars, similar to those described in section 4.

But there are also other econometric issues encountered in section 2.2 on the Sapir–Whorf hypothesis that need further research:

- (i) Chen (2013) and Sutter et al. (2015) find that strong- and weak-future tense have an influence on savings (in the broad sense). Galor and Özak (2016) suggest that the type of tense can be explained by crop yields and growth cycles before 1500 AD.
- (ii) Mavisakalyan’s (2015) research points out that in countries where the language is gender intensive, women participate less in the labor force. In their recent paper, Galor, Özak, and Sarid (2018) again suggest that ecological origins of hierarchical societies may have contributed to the existence of sex-based gender in nouns.
- (iii) Tabellini (2008) considers that politeness rules could be used to instrument cultural traits such as trust and respect, which are then used in the second-stage estimation to “explain” the quality of institutions. Galor, Özak, and Sarid. (2018) point out that ecological origins of hierarchical societies may have contributed to the existence of politeness distinctions,

and may be the “deep” variables that explain politeness distinctions.

These three examples lead to the following questions:

- (i) Do Chen (2013), Sutter et al. (2015), Mavisakalyan (2015), and Tabellini (2008) use the “right” variable, that is, linguistic differences, to explain human behavior, which would lead us to accept a form of the Sapir–Whorf hypothesis that language is at least partly responsible for or even “causes” our behavior?
- (ii) Or are Galor and Özak (2016) and Galor, Özak, and Sarid (2018) right to suggest that crop yields and ecological origins of hierarchical societies have an influence on languages, which may in turn influence our behavior,
- (iii) or are Galor, Özak, and Sarid variables directly responsible for our behavior,
- (iv) or are linguistic differences and variables suggested by Galor, Özak, and Sarid (2016, 2018) both concurring to influence our behavior?

What is exogenous with respect to what? Can we decipher any causality, even if we are willing to relax somewhat Hume’s (2017 [1748], p. 38) famous sentence that the relation of cause and effect should be such that “where, if the first event hadn’t occurred, the second wouldn’t have occurred either”? Or is all this coevolution or correlation? At least, it shows that there are good reasons for linguists and economists to work together.

5.3 *Language, Culture and Behavior*

The preceding examples show that it is difficult to distinguish whether it is language or culture that has an effect on behavior. As

Kramsch (1998, p. 13) points out, “there are cultural differences in the semantic associations evoked by seemingly common concepts.” Obviously the word “socialism” and its concept had a very different meaning among Russians after the 1989 collapse of the USSR than it had in France at the same time. Language and culture are both part of a nation’s values. They may develop simultaneously, or within too small time intervals to enable historians to distinguish which one occurred before the other, and there is little hope for economists to run serious causality tests.

There are, however, clear contemporary cases for which one is certain that culture has an impact on language, and in which “our grammars follow our values” (Everett, 2012, p. 273). Feminization has eventually changed our way of thinking as well as our writing and speaking. Today we can hardly write “he” for a consumer or a producer in our models. Some authors change this to “(s)he,” “he/she” or “she/he.” Others switch between “he” and “she” in the same sentence, or decide that consumers would be feminine and producers masculine, or the other way around.

French is a gendered language. When a woman happens to be a university “professor,” it used to be acceptable to say and write “madame le professeur” where the definite article “le” is masculine. This changed to the feminine article “la” a couple of decades ago. Today, you have to say “madame la professeure,” where the usual feminine ending “e” is added to “professeur.” If one needs to include both feminine and masculine professors of a school, it was recently suggested to use the so-called “écriture inclusive” (inclusive writing) that follows: “mesdames et messieurs les professeur.e.s,” where “.r” and “.e” signal masculine or feminine, while the final “s” is for plural. For the time being, this is not enforced by the Académie Française, but becomes more and more frequent.

In contemporary French grammar, the masculine form takes precedence over the feminine form if both appear in the same sentence and there is a qualifying adjective, even if the qualifier can be feminized (which is very often the case in French). One should say and write “il et elle sont heureux” (he and she are happy), where “heureux” is the masculine plural qualifier. This was not so before the mid-seventeenth century. At the time, and following the so-called proximity rule used in ancient Greek and Latin, the sentence would have been “il et elle sont heureuses,” where “heureuses” is the feminine plural, since “elle” comes second and is closer to the adjective or “elle et il sont heureux” since here, “il” comes second. There are recent proposals to change the current rule of precedence, which was introduced for reasons that were probably not considered to be sexist at the time, but look pretty much so today.⁸⁹

5.4 Other Research Issues

The “returns to language learning” literature initiated by Chiswick in the early 1970s produced a large body of research,⁹⁰ but there exists no clear theoretical model of the relation between language learning and labor markets. We are able to infer empirically the wage premium of migrants who learn the language of the host country, as well as the one of those who were born in a country but acquire another language than the one spoken in that country. But we have no microeconomic foundations for making

those decisions. There is common work to do by economists and neuroscientists.

The issue of linguistic and social identity should attract the attention of economists. Both are obvious determinants of our individual or collective behavior, especially at times of social and economic critical situations. There is little doubt that one can observe strong signs of linguistic identity, as is illustrated by the attachment to African American Vernacular English and New York Latino English in the United States, or Catalan in Europe, as well as by the societal, economic, and even judicial backlash effects of identity choices. One also has to address the yet unexplored issue of measurement of identity, its construction, and transmission, including the possibility of multiple identities or the voluntary changes of identity in some cases. “Identity” languages are not only studied, but also get revived: Hebrew is of course the case par excellence, but, in the face of our globalizing world, Catalan, Basque, Welsh, Maori in New Zealand, or Inuktitut in Eastern Canada are also. As mentioned earlier, Aspachs-Bracons et al. (2008) stressed the compulsory nature of introducing Catalan in Catalonia; this generated the dramatic rise of identity that may have contributed to the tensions that we observe nowadays between Catalonia and the rest of Spain. The voluntary option of studying Basque in the Basque Country failed to produce the same result, and so did the many attempts to reintroduce local languages such as Breton or Provençal in French schools during the late 1960s.

Incorporating linguistic diversity into production functions may also lead to interesting research. This was suggested by Alesina and La Ferrara (2005) and pursued by Fujita and Weber (2010), Osang and Weber (2017), and Desmet, Ortuño-Ortín, and Weber (2017). They introduced as a variable in societal utility functions, the number of distinct linguistic groups, or the degree of cultural

⁸⁹The reason was invoked by the French grammarian Claude de Vaugelas (2009 [1647], p. 83) in his *Remarques sur la langue française*. He suggested that “since the masculine gender is more noble, it should always prevail if a feminine and a masculine words are together” (“le genre masculin étant le plus noble, [il] doit prédominer toutes fois que le masculin et le féminin se trouvent ensemble”).

⁹⁰See Adserà and Pytlíková (2016) for an excellent survey.

and linguistic distinctions between natives and immigrants. More specifically, in their models individuals are endowed with a utility function $u(G, c)$ where G and c represent the public and the private good. The production of G is financed by a proportional tax and its cost is increasing in the level of diversity (for example, the average ethnolinguistic distance between individuals), as more diversity, that is more players with diverse interests, makes it more difficult to reach a decision on taxation. Introducing various layers of diversity in economic calculus may probably be useful and promising.

Last but not least, Esteban, Mayoral, and Ray (2012) developed a new line of research on the impact of ethnic divisions on conflict. They link conflict intensity within a country to three indices of ethnic distribution: polarization, fractionalization, and linguistic diversity represented by Greenberg's B -index. Their empirical analysis supports the assertion that all three distributional measures are significantly correlated to conflict. They also introduce country-specific measures of group cohesion based on linguistic distances and the importance of public goods, combining them with the distributional measures, and point out that though linguistic distances are exogenous to conflict, they can be expected to drive or at least influence antagonisms across groups. Taking this argument even further, one may probably show that combining linguistic distances and income inequality could be a deciding factor of the emergence and length of conflicts.

REFERENCES

- Abbott, Michael G., and Charles M. Beach. 1992. "Immigrant Earnings Differentials in Canada: A More General Specification of Age and Experience Effects." *Empirical Economics* 17 (2): 221–38.
- Adserà, Alicia, and Barry R. Chiswick. 2007. "Are There Gender and Country of Origin Differences in Immigrant Labor Market Outcomes across European Destinations?" *Journal of Population Economics* 20: 495–526.
- Adserà, Alicia, and Mariola Pytliková. 2015. "The Role of Language in Shaping International Migration." *Economic Journal* 125 (586): F49–81.
- Adserà, Alicia, and Mariola Pytliková. 2016. "Language and Migration." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 342–72. London: Palgrave Macmillan.
- Aldashev, Alisher, and Alexander M. Danzer. 2014. "Economic Returns to Speaking the Right Language(s)? Evidence from Kazakhstan's Shift in State Language and Language of Instruction." Institute of Labor Economics (IZA) Discussion Paper 8624.
- Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (4): 1243–84.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8 (2): 155–94.
- Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43 (3): 762–800.
- Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou. 2016. "Ethnic Inequality." *Journal of Political Economy* 124 (2): 428–88.
- Alesina, Alberto, and Ekaterina Zhuravskaya. 2011. "Segregation and the Quality of Government in a Cross-Section of Countries." *American Economic Review* 101 (5): 1872–911.
- Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature* 42 (3): 691–751.
- Angrist, Joshua, Aimee Chin, and Ricardo Godoy. 2008. "Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?" *Journal of Development Economics* 85 (1–2): 105–28.
- Angrist, Joshua D., and Victor Lavy. 1997. "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco." *Journal of Labor Economics* 15 (1): S48–76.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton and Oxford: Princeton University Press.
- Annett, Anthony. 2001. "Social Fractionalization, Political Instability, and the Size of the Government." *IMF Staff Papers* 46 (3): 561–92.
- Appelfeld, Aharon. 2017. *The Man Who Never Stopped Sleeping*. New York: Schocken Books.
- Armstrong, Alex. 2015. "Equilibria and Efficiency in Bilingual Labour Markets." *Journal of Economic Behavior & Organisation* 112: 204–20.
- Aspachs-Bracons, Oriol, Irma Clots-Figueras, Joan Costa-Font, and Paolo Masella. 2008. "Compulsory Language Educational Policies and Identity Formation." *Journal of the European Economic Association* 6 (2–3): 434–44.
- Aston, John A. D., Jeng-Min Chiou, and Jonathan P. Evans. 2010. "Linguistic Pitch Analysis Using Functional Principal Component Mixed Effect Models."

- Journal of the Royal Statistical Society. Series C (Applied Statistics)* 59 (2): 297–317.
- Austen-Smith, David, and Roland G. Fryer, Jr. 2005. "An Economic Analysis of 'Acting White.'" *Quarterly Journal of Economics* 120 (2): 551–83.
- Aydemir, Abdurrahman, and Mikal Skuterud. 2005. "Explaining the Deterioration Entry Earnings of Canada's Immigrant Cohorts, 1996–2000." *Canadian Journal of Economics* 38 (2): 641–72.
- Baier, Scott L., Amanda Kerr, and Yoto V. Yotov. 2017. "Gravity, Distance, and International Trade." Center for Economic Studies (CESifo) Working Paper 6357.
- Baker, Mark C. 2001. *The Atoms of Language*. New York: Basic Books.
- Baugh, John. 2009. "Econolinguistics in the USA." In *Language and Poverty*, edited by Wayne Herbert, Sally McConnell-Ginet, Amanda Miller, and John Whitman, 67–77. Bristol: Multilingual Matters.
- Beenstock, Michael, Barry R. Chiswick, and Gaston L. Repetto. 2001. "The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel." *International Migration* 39 (3): 33–60.
- Behle, Miriam, Louise Gorges, and Eva Markowsky. 2017. "Speaking of Gender: Does Language Affect Labor Market Outcomes?" Paper presented at the Colloquium on Language Skills for Economic and Social Inclusion, Berlin.
- Beine, Michel, Frédéric Docquier, and Çağlar Özden. 2011. "Diasporas." *Journal of Development Economics* 95 (1): 30–41.
- Bellofatto, Anthony. 2017. "How Does Language Impact Foreign Investing in a Multilingual Country?" Available on SSRN at 3001825. 00000
- Berlin, Brent, and Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley, Los Angeles, and Oxford: University of California Press.
- Berman, Eli, Kevin Lang, and Erez Siniver. 2003. "Language-Skill Complementarity: Returns to Immigrant Language Acquisition." *Labour Economics* 10 (3): 265–90.
- Besley, Timothy, and Stephen Coate. 1992. "Understanding Welfare Stigma: Taxpayer Resentment and Statistical Discrimination." *Journal of Public Economics* 48 (2): 165–83.
- Bleakley, Hoyt, and Aimee Chin. 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics* 86 (2): 481–96.
- Bloom, David E., and Gilles Grenier. 1992. "Economic Perspectives on Language: The Relative Value of Bilingualism in Canada and the United States (1992)." In *Language Loyalties: A Source Book on the Official English Controversy*, edited by James Crawford, 445–52. Chicago: University of Chicago Press.
- Boas, Franz. 1940. *Race, Language and Culture*. Chicago: University of Chicago Press.
- Bond Niall, and Victor Ginsburgh. 2016. "Language and Emotion." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 231–60. London: Palgrave Macmillan.
- Boroditsky, Lera. 2001. "Does Language Shape Thought? Mandarin and English Speakers' Conceptions of Time." *Cognitive Psychology* 43 (1): 1–22.
- Boroditsky, Lera. 2009. "How Does Our Language Shape the Way We Think?" *Edge Foundation*, June 11, Accessed March 31, 2020. <https://www.edge.org/conversation/leraboroditsky-how-does-our-language-shape-the-way-we-think>.
- Boroditsky, Lera, Orly Fuhrman and Kelly McCormick. 2011. "Do English and Mandarin Speakers Think about Time Differently?" *Cognition* 118 (1): 123–29.
- Boroditsky, Lera, and Alice Gaby. 2010. "Remembrance of Times East: Absolute Spatial Representations of Time in an Australian Aboriginal Community." *Psychological Science* 21 (11): 1635–39.
- Boroditsky, Lera, Lauren A. Schmidt, and Webb Phillips. 2003. "Sex, Syntax, and Semantics." In *Language in Mind: Advances in the Study of Language and Thought*, edited by Dedre Gentner and Susan Goldin-Meadow, 61–80. Cambridge: MIT Press.
- Bossert, Walter, Conchita d'Ambrosio, and Eliana La Ferrara. 2011. "A Generalized Index of Fractionalization." *Economica* 78 (312): 723–50.
- Bossuyt, Audrey, Laurence Broze, and Victor Ginsburgh. 2001. "On Invisible Trade Relations between Mesopotamian Cities during the Third Millennium B.C." *Professional Geographer* 53 (3): 374–83.
- Bratsberg, Bernt, James F. Ragan, Jr., and Zafar M. Nasir. 2002. "The Effect of Naturalization on Wage Growth: A Panel Study of Young Male Immigrants." *Journal of Labor Economics* 20 (3): 568–97.
- Bredtmann, Julia, Klaus Nowotny, and Sebastian Otten. 2017. "Linguistic Distance, Networks and Migrants' Regional Location Choice." Institute of Labor Economics (IZA) Discussion Paper 11171.
- Bretton, Henry. 1976. "Political Science, Language, and Politics." In *Language and Politics*, edited by William O'Barr and Jean O'Barr, 431–48. Berlin: de Gruyter.
- Brock, William A., and Steven N. Durlauf. 2001. "Growth Empirics and Reality." *World Bank Economic Review* 15 (2): 229–72.
- Brown, Roger W., and Eric H. Lenneberg. 1954. "A Study in Language and Cognition." *Journal of Abnormal and Social Psychology* 49 (3): 454–62.
- Bruk, Solomon Il'ich, and V. S. Apenchenko. 1964. *Atlas Narodov Mira*. Moscow: Gosudarstvennogo Geologicheskogo Komiteta SSSR.
- Campos, Nauro F., Ahmed Saleh, and Vitaliy Kuzeyev. 2011. "Dynamic Ethnic Fractionalization and Economic Growth." *Journal of International Trade and Economic Development* 20 (2): 129–52.
- Canetti, Elias. 1977. *Die Gerettete Zunge: Geschichte einer Jugend*. München: Carl Hanser Verlag.
- Casasanto, Daniel. 2008. "Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought." In *Time to Speak: Cognitive and Neural Prerequisites for Time in Language*, edited by Peter Indefrey and Marianne Gullberg, 63–79. Ann Arbor: University of Michigan.

- Casasanto, Daniel, and Lera Boroditsky. 2008. "Time in the Mind: Using Space to Think about Time." *Cognition* 106: 579–93.
- Castañeda Dower, Paul, Victor Ginsburgh, and Shlomo Weber. 2017. "Colonial Legacy, Polarization and Linguistic Disenfranchisement: The Case of the Sri Lankan War." *Journal of Development Economics* 127: 440–48.
- Cattaneo, Alejandra, and Rainer Winkelmann. 2005. "Earning Differentials between German and French Speakers in Switzerland." *Swiss Journal of Economics and Statistics* 141 (2): 191–212.
- Cavalli-Sforza, Luigi Luca. 1997. "Genes, Peoples and Languages." *Proceedings of the National Academy of Sciences of the United States of America* 94 (15): 7719–24.
- Cavalli-Sforza, Luigi Luca. 2000. *Genes, Peoples, and Languages*. Berkeley: University of California Press.
- Cavalli-Sforza, Luigi Luca, and Francesco Cavalli-Sforza. 1995. *The Great Human Diasporas: The History of Diversity and Evolution*. Cambridge: Perseus Books.
- Cavalli-Sforza, Luigi Luca, Paolo Menozzi, and Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Chakraborty, Tanika, and Shilpi Kapur Bakshi. 2016. "English Language Premium: Evidence from a Policy Experiment in India." *Economics of Education Review* 50: 1–16.
- Chen, M. Keith. 2013. "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." *American Economic Review* 103 (2): 690–731.
- Chiswick, Barry R. 1998. "Hebrew Language Usage: Determinants and Effects on Earnings among Immigrants in Israel." *Journal of Population Economics* 11 (2): 253–71.
- Chiswick, Barry R., and Paul W. Miller. 1995. "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics* 13 (2): 246–88.
- Chiswick, Barry R., and Paul W. Miller. 2002. "Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle." *Journal of Population Economics* 15 (1): 31–57.
- Chiswick, Barry R., and Paul W. Miller. 2005. "Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages." *Journal of Multilingual and Multicultural Development* 26 (1): 1–11.
- Chiswick, Barry R., and Paul W. Miller. 2007. *The Economics of Language: International Analyses*. London and New York: Routledge.
- Chiswick, Barry R., and Paul W. Miller. 2014. "International Migration and the Economics of Language." In *Immigrants*. Vol. 1A, *Handbook of the Economics of International Migration*, edited by Barry R. Chiswick and Paul W. Miller, 211–70. Amsterdam: North-Holland.
- Chomsky, Noam. 1975. *The Logical Structure of Linguistic Theory*. New York: Plenum Press.
- Church, Jeffrey, and Ian King. 1993. "Bilingualism and Network Externalities." *Canadian Journal of Economics* 26 (2): 337–45.
- Clay, Karen, Jeff Lingwall, and Melvin Stephens, Jr. 2012. "Do Schooling Laws Matter? Evidence from the Introduction of Compulsory Attendance Laws in the United States." NBER Working Paper 18477.
- Clots-Figueras, Irma, and Paolo Masella. 2013. "Education, Language and Identity." *Economic Journal* 123 (570): F332–57.
- Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers* 56 (4): 563–95.
- Collier, Paul, Patrick Honohan, and Karl Ove Moene. 2001. "Implications of Ethnic Diversity." *Economic Policy* 16 (32): 129–55.
- Comrie, Bernard. 1984. "Book Review of Ekkehart Malotki, Hopi Time: A Linguistic Analysis of the Temporal Concepts in the Hopi Language." *Australian Journal of Linguistics* 3 (1): 131–33.
- Costa, Albert, Alice Foucart, Sayuri Hayakawa, Melina Aparici, Jose Apesteguia, Joy Heafner et al. 2014. "Your Morals Depend on Language." *PLOS ONE* 9 (4): e94842.
- Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economics*, Vol. 4, edited by Gita Gopinath, Elhanan Helpman and Kenneth Rogoff, 197–261. Amsterdam: North-Holland.
- Crystal, David. 1999. *A Dictionary of Languages*. Chicago: University of Chicago Press.
- Crystal, David. 2003. *English as a Global Language*. 2nd ed. Cambridge: Cambridge University Press.
- Dahl, Östen. 2003. "Stuck in the Futureless Zone." *Diversity Linguistics Comments*, March 9. <http://dlc.hypotheses.org/360> (accessed February 22, 2016).
- Dahl, Östen, and Viveka Vellupillai. 2013. "Tense and Aspect." In *World Atlas of Language Structures Online*, edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dahl, Øyvind. 1995. "When the Future Comes from behind: Malagasy and Other Time Concepts and Some Consequences for Communication." *International Journal of Intercultural Relations* 19 (2): 197–209.
- Davydov, Denis, Alexander Shapoval, and Shlomo Weber. 2018. "Linguistic Equilibrium with Local and World Languages: Challenges of Globalization." *World Economy* 41 (7): 1790–811.
- Davydov, Denis, and Shlomo Weber. 2016. "A Simple Characterization of the Family of Diversity Indices." *Economics Letters* 147: 121–23.
- Dediu, Dan, and D. Robert Ladd. 2007. "Linguistic Tone Is Related to the Population Frequency of the Adaptive Haplogroups of Two Brain Size Genes, ASPM and Microcephalin." *Proceedings of the National Academy of Sciences of the United States of America* 104 (26): 10944–49.
- de Candolle, Alphonse. 1987. *Histoire des Sciences et*

- des Savants Depuis Deux Siècles; Suivie d'Autres Études sur des Sujets Scientifiques, en Particulier sur la Sélection dans l'Espèce Humaine." Paris: Arthème Fayard. [1873].
- De Luca, Erri. 1997. *Ora Prima*. Monastero di Bose: Qiqajon.
- Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg. 2012. "The Political Economy of Linguistic Cleavages." *Journal of Development Economics* 97 (2): 322–38.
- Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg. 2017. "Culture, Ethnicity, and Diversity." *American Economic Review* 107 (9): 2479–513.
- Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber. 2009. "Linguistic Diversity and Redistribution." *Journal of the European Economic Association* 7 (6): 1291–318.
- Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber. 2017. "Peripheral Diversity: Transfers versus Public Goods." *Social Choice and Welfare* 49 (3–4): 787–823.
- de Vaugelas, Claude F. 2009. *Remarques sur la Langue Française*. Geneva: Droz [1647].
- Dustmann, Christian, and Arthur Van Soest. 2002. "Language and the Earnings of Immigrants." *ILR Review* 55 (3): 473–92.
- Dyen, Isidore, Joseph B. Kruskal, and Paul Black. 1992. "An Indoeuropean Classification: A Lexicostatistical Experiment." *Transactions of the American Philological Society* 82 (5): iii–iv.
- Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112 (4): 1203–50.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2016. *Ethnologue: Languages of the World*. 19th ed. Dallas: SIL International Publications.
- Edwards, Michael. 1992. "Beckett's French." *Translation and Literature* 1 (1): 68–83.
- Egger, Peter. 2008. "On the Role of Distance for Bilateral Trade." *World Economy* 31 (5): 653–62.
- Egger, Peter H., and Andrea Lassmann. 2012. "The Language Effect in International Trade: A Meta-analysis." *Economic Letters* 116 (2): 221–24.
- Eriksson, Katherine. 2014. "Does the Language of Instruction in Primary School Affect Later Labour Market Outcomes? Evidence from South Africa." *Economic History of Developing Regions* 29 (2): 331–35.
- Esteban, Joan, Laura Mayoral, and Debraj Ray. 2012. "Ethnicity and Conflict: An Empirical Study." *American Economic Review* 102 (4): 1310–342.
- Esteban, Joan-María, and Debraj Ray. 1994. "On the Measurement of Polarization." *Econometrica* 62 (4): 819–51.
- European Commission. 2006. "Europeans and Their Languages." Special Eurobarometer 243.
- Evans, Nicholas, and Stephen C. Levinson. 2009. "The Myth of Language Universal: Language Diversity and Its Importance for Cognitive Sciences." *Behavioral and Brain Sciences* 32: 429–92.
- Everett, Daniel L. 2012. *Language: The Cultural Tool*. New York: Vintage Books.
- Fabb, Nigel. 2016. "Linguistic Theory, Linguistic Diversity and Whorfian Economics." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 17–60. Basingstoke: Palgrave Macmillan.
- Falck, Oliver, Stephan Heblich, Alfred Lamelli, and Jens Südekum. 2012. "Dialects, Cultural Identity, and Economic Exchange." *Journal of Urban Economics* 72 (2–3): 225–39.
- Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8 (2): 195–222.
- Fearon, James D., and David D. Laitin. 1999. "Weak States, Rough Terrain, and Large-Scale Ethnic Violence Since 1945." Accessed March 31, 2020. <https://web.stanford.edu/group/ethnic/workingpapers/insurg1.pdf>.
- Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (1): 75–90.
- Fidrmuc, Jan, Victor Ginsburgh, and Shlomo Weber. 2007. "Ever Closer Union or Babylonian Discord? The Official-Language Problem in the European Union." Centre for Economic Policy Research Discussion Paper 6367.
- Fishman, Joshua A. 1968. *Readings in the Sociology of Language*. The Hague and Paris: Mouton.
- Florida, Richard. 2002. *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community, and Everyday Life*. New York: Perseus Book Group.
- Florida, Richard, and Gary Gates. 2001. *Technology and Tolerance: The Importance of Diversity to High-Tech Growth*. Washington, DC: Brookings Institution Center on Urban and Metropolitan Policy.
- Frankel, Jeffrey A., and David H. Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89 (3): 379–99.
- Fry, Richard, and Lindsay Lowell. 2003. "The Value of Bilingualism in the U.S. Labor Market." *ILR Review* 57 (1): 128–40.
- Fujita, Masahisa, and Shlomo Weber. 2010. "Immigration Quotas in the Globalized Economy." *Journal of the New Economic Association* 7: 10–23.
- Functional Phylogenies Group. 2012. "Phylogenetic Inference for Function-Valued Traits: Speech Sound Evolution." *Trends in Ecology and Evolution* 27 (3): 160–66.
- Gabszewicz, Jean, Victor Ginsburgh, and Shlomo Weber. 2011. "Bilingualism and Communicative Benefits." *Annals of Economics and Statistics* 101/102: 271–86.
- Gaby, Alice. 2012. "The Thaayorre Think of Time Like They Talk of Space." *Frontiers in Psychology* 3: Article 300.
- Galasi, Peter. 2003. "Estimating Wage Equations for Hungarian Higher-Education Graduates." Available on SSRN at 465882.
- Galor, Oded, and Ömer Özak. 2016. "The Agricultural

- Origins of Time Preference." *American Economic Review* 106 (10): 3064–103.
- Galor, Oded, Ömer Özak, and Assaf Sarid. 2016. "Geographical Origins and Economic Consequences of Language Structures." Center for Economic Studies (CESIFO) Working Paper 6149.
- Galor, Oded, Ömer Özak, and Assaf Sarid. 2018. "Geographical Roots of the Coevolution of Cultural and Linguistic Traits." NBER Working Paper 25289.
- Gamkrelidze, Thomas V., and V. V. Ivanov. 1990. "The Early History of Indo-European Languages." *Scientific American* 262 (3): 82–89.
- Geng, Difei. 2012. "Identifying the Unique Polarization Index: A Mean-Preserving Axiomatic Approach." *Journal of Public Economic Theory* 14 (5): 791–812.
- Gibson, Edward, Richard Futrell, Julian Jara-Ettinger, Kyle Mahowald, Leon Bergen, Sivalogeswaran Ratnasingam et al. 2017. "Color Naming across Languages Reflects Color Use." *Proceedings of the National Academy of Sciences of the United States of America* 114 (40): 10785–90.
- Ginsburgh, Victor, Jacques Melitz, and Farid Toubal. 2017. "Foreign Language Learning and Trade." *Review of International Economics* 25 (2): 320–61.
- Ginsburgh, Victor, and Juan D. Moreno-Ternero. 2018. "Compensation Schemes for Learning a Lingua Franca in the European Union." *World Economy* 41 (7): 1775–89.
- Ginsburgh, Victor, Juan D. Moreno-Ternero, and Shlomo Weber. 2017. "Ranking Languages in the European Union: Before and after Brexit." *European Economic Review* 93: 139–51.
- Ginsburgh, Victor, Ignacio Ortuno-Ortín, and Shlomo Weber. 2005. "Disenfranchisement in Linguistically Diverse Societies. The Case of the European Union." *Journal of the European Economic Association* 3 (4): 946–64.
- Ginsburgh, Victor, and Juan Prieto-Rodriguez. 2007. "Returns to Foreign Languages of Native Workers in the EU." *ILR Review* 64 (3): 599–618.
- Ginsburgh, Victor, and Shlomo Weber. 2005. "Language Disenfranchisement in the European Union." *Journal of Common Market Studies* 43 (2): 273–86.
- Ginsburgh, Victor, and Shlomo Weber. 2011. *How Many Languages Do We Need? The Economics of Linguistic Diversity*. Princeton: Princeton University Press.
- Ginsburgh, Victor, and Shlomo Weber. 2016. "Linguistic Distances and Ethnolinguistic Fractionalization and Disenfranchisement Indices." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 137–73. Basingstoke: Palgrave MacMillan.
- Ginsburgh, Victor, Shlomo Weber, and Sheila Weyers. 2011. "The Economics of Literary Translation: Some Theory and Evidence." *Poetics: Journal of Empirical Research on Culture, the Media and the Arts* 39 (3): 228–46.
- Goebel, Hans. 1982. *Dialektometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*. Wien: Austrian Academy of Sciences.
- Goldin, Claudia, and Lawrence F. Katz. 2011. "Putting the 'Co' in Education: Timing, Reasons, and Consequences of College Coeducation from 1835 to the Present." *Journal of Human Capital* 5 (4): 377–417.
- Graddol, David. 2006. *English Next*. London: British Council.
- Gray, Russell D., and Quentin D. Atkinson. 2003. "Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin." *Nature* 426: 435–39.
- Greenberg, Joseph H. 1955. *Studies in African Linguistic Classification*. Branford: Compass Publishing Company.
- Greenberg, Joseph H. 1956. "The Measurement of Linguistic Diversity." *Language* 32 (1): 109–15.
- Greenberg, Joseph H. 1963. *The Languages of Africa*. Bloomington: Indiana University Press.
- Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press.
- Greenberg, Joseph H. 2000. *Grammar*. Vol. 1, *Indo-European and Its Closest Relative: The Eurasiatic Language Family*. Stanford: Stanford University Press.
- Greenberg, Joseph H. 2002. *Lexicon*. Vol. 2, *Indo-European and Its Closest Relative: The Eurasiatic Language Family*. Stanford: Stanford University Press.
- Grinblatt, Mark, and Matti Keloharju. 2001. "How Distance, Language, and Culture Influence Stockholdings and Trades." *Journal of Finance* 56 (3): 1053–73.
- Grogger, Jeffrey. 2011. "Speech Patterns and Racial Wage Inequality." *Journal of Human Resources* 46 (1): 1–25.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives* 20 (2): 23–48.
- Gumperz, John J., and Stephen C. Levinson. 1991. "Rethinking Linguistic Relativity." *Current Anthropology* 32 (5): 613–23.
- Hadjipantelis, Pantelis Z., John A. D. Aston, and Jonathan P. Evans. 2012. "Characterizing Fundamental Frequency in Mandarin: A Functional Principal Component Approach Utilizing Mixed Effect Models." *Journal of the Acoustical Society of America* 131 (6): 4651–64.
- Hägerstrand, Torsten. 1957. "Migration and Area. Survey of Swedish Migration Fields and Hypothetical Consideration of their Genesis." In *Migration in Sweden: A Symposium*. Vol. 13, *Lund Studies in Geography*, edited by David Hannerberg, Torsten Hägerstrand, and Bruno Odeving, 27–158. Lund: Royal University of Lund.
- Hanson, Gordon, and Chong Xiang. 2011. "Trade Barriers and Trade Flows with Product Heterogeneity: An Application to US Motion Picture Exports." *Journal of International Economics* 83 (1): 14–26.
- Harhoff, Dietmar, Karin Hoisl, Bruno van Pottelsberghe de la Potterie, and Charlotte Vandepuut. 2016. "Languages, Fees and the International Scope of Patenting." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 403–22. London: Palgrave Macmillan.

- Harris, Randy. 1993. *The Linguistic Wars*. Oxford: Oxford University Press.
- Hart-Gonzalez, Lucinda, and Stephanie Lindemann. 1993. *Expected Achievement in Speaking Proficiency*. Unpublished.
- Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*, Vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 131–95. Amsterdam: North-Holland.
- Heblich, Stephan, Alfred Lameli, and Gerhard Riener. 2015. "The Effect of Perceived Regional Accents on Individual Economic Behavior: A Lab Experiment on Linguistic Performance, Cognitive Ratings and Economic Decisions." *PLOS ONE* 10 (2): e0113475.
- Hellerstein, Judith, and David Neumark. 2003. "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set." *Annales d'Economie et de Statistique* 71/72: 19–78.
- Hershkovitz, Israel et al. 2010. "Middle Pleistocene Dental Remains from Qesem Cave (Israel)." *American Journal of Physical Anthropology* 144 (4): 575–92.
- Hershkovitz, Israel et al. 2016. "New Middle Pleistocene Dental Remains from Qesem Cave (Israel)." *Quaternary International* 398: 148–58.
- Hibbs, Douglas A. 1973. *Mass Political Violence: A Cross-National Causal Analysis*. New York: Wiley.
- Hill, Mark O. 1973. "Diversity and Evenness: A Unifying Notation and Its Consequences." *Ecology* 54 (2): 427–32.
- Hočevar, Toussaint. 1975. "Equilibria in Linguistic Minority Markets." *Kyklos* 28 (2): 337–57.
- Hofstede, Geert. 1980. *Culture's Consequences: International Differences in Work-Related Values*. Thousand Oaks, CA: Sage.
- Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. Thousand Oaks, CA: Sage.
- Holden, Nigel. 2016. "Economic Exchange and Business Language in the Ancient World: An Exploratory Review." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 290–311. Basingstoke: Palgrave MacMillan.
- Hume, David. 2017. "Enquiry Concerning Human Understanding." Accessed January 5, 2018. <https://www.earlymoderntexts.com/assets/pdfs/hume1748.pdf>. [1748].
- Ispording, Ingo E. 2014. "Disadvantage of Linguistic Origin—Evidence from Immigrant Literacy Scores." *Economics Letters* 123 (2): 236–39.
- Ivlevs, Artjoms, and Roswitha M. King. 2014. "Minority Education Reform and Pupil Performance in Latvia." *Economics of Education Review* 38: 151–66.
- Jain, Tarun. 2017. "Common Tongue: The Impact of Language on Educational Outcomes." *Journal of Economic History* 77 (2): 473–510.
- Janson, Tore. 2002. *Speak: A Short Story of Languages*. Oxford: Oxford University Press.
- January, David, and Edward Kako. 2007. "Re-evaluating Evidence for Linguistic Relativity: Reply to Boroditsky (2001)." *Cognition* 104 (2): 417–26.
- Kadochnikov, Denis V. 2016. "Languages, Regional Conflicts and Economic Development: Russia." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 538–80. London: Palgrave Macmillan.
- Kay, Paul, and Terry Regier. 2006. "Language, Thought and Color: Recent Developments." *Trends in Cognitive Sciences* 10 (2): 51–54.
- Kessler, Brett. 1995. "Computational Dialectology in Irish Gaelic." In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, 60–66. San Francisco: Morgan Kaufmann Publishers.
- Kessler, Brett. 2001. *The Significance of Word Lists*. Stanford: Center for the Study of Language and Information.
- Kim, Young-Chul, and Glenn C. Loury. 2019. "To Be or Not to Be: Stereotypes, Identity Choice and Group Identity." *Journal of Public Economic* 174: 36–52.
- Kramsch, Claire. 1998. *Language and Culture*. Oxford: Oxford University Press.
- Kuijs, Louis. 2000. "The Impact of Ethnic Heterogeneity on the Quantity and Quality of Public Spending." International Monetary Fund Working Paper WP0049.
- Laitin, David D. 1986. *Hegemony and Culture: Politics of Religious Change among the Yoruba*. Chicago: University of Chicago Press.
- Laitin, David D. 1989. "Language Policy and Political Strategy in India." *Policy Studies* 22 (3–4): 415–36.
- Laitin, David D. 1994. "The Tower of Babel as a Coordination Game: Political Linguistics in Ghana." *American Political Science Review* 88 (3): 622–34.
- Laitin, David D. 2000. "What Is a Language Community?" *American Journal of Political Science* 44 (1): 142–55.
- Laitin, David D., and Rajesh Ramachandran. 2018. "Linguistic Diversity, Official Language Choice and Nation Building: Theory and Evidence." Accessed October 22, 2018. https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SAEe2018&paper_id=239.
- Lamberton, Donald M. 2002. *The Economics of Language*. Vol. 150, *International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar Publishing.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1999. "The Quality of Government." *Journal of Law, Economics and Organization* 15 (1): 222–79.
- Lazear, Edward P. 1999. "Culture and Language." *Journal of Political Economy* 107 (S6): S95–127.
- Leavitt, John. 2011. *Linguistic Relativities. Language Diversity and Modern Thought*. Cambridge: Cambridge University Press.
- Lenneberg, Eric H., and John M. Roberts. 1956. *The Language of Experience: A Study in Methodology*. Vol. 13, *Indiana University Publications in Anthropology and Linguistics*. Baltimore: Waverly

- Press.
- Leslie, Derek, and Joanne Lindley. 2001. "The Impact of Language Ability on Employment and Earnings of Britain's Ethnic Communities." *Economica* 68 (272): 587–606.
- Levenshtein, V.I. 1966. "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals." *Cybernetics and Control Theory* 10 (8): 707–10.
- Levinsohn, James. 2007. "Globalization and the Returns to Speaking English in South Africa." In *Globalization and Poverty*, edited by Ann Harrison. Chicago: Chicago University Press.
- Lewis, Bernard. 2004. *From Babel to Dragomans: Interpreting the Middle East*. Oxford: Oxford University Press.
- Lewis, Bernard. 2009. *Islam: The Religion and the People*. Upper Saddle River: Prentice Hall.
- Li, Peggy, and Lila Gleitman. 2002. "Turning the Tables: Language and Spatial Reasoning." *Cognition* 83 (3): 265–94.
- Lian, Brad, and John R. Oneal. 1997. "Cultural Diversity and Economic Development: A Cross-National Study of 98 Countries, 1960–1985." *Economic Development and Cultural Change* 46 (1): 61–77.
- Lleras-Muney, Adriana. 2002. "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939." *Journal of Law and Economics* 45 (2): 401–35.
- Lleras-Muney, Adriana, and Allison Shertzer. 2015. "Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schooling Laws on Immigrants." *American Economic Journal: Economic Policy* 7 (3): 258–90.
- Loreto, Vittorio, Animesh Mukherjee, and Francesca Tria. 2012. "On the Origin of the Hierarchy of Color Names." *Proceedings of the National Academy of Sciences* 109 (18): 6819–24.
- Malotki, Ekkehart. 1983. *Hopi Time: A Linguistic Analysis of the Temporal Concepts in the Hopi Language*. Berlin: Mouton.
- Mandelbaum, David G., ed. 1949. *Selected Writings of Edward Sapir in Language, Culture, Personality*. Berkeley, CA: University of California Press.
- Manea, Norman. 2012. *The Fifth Impossibility: Essays on Exile and Language*. New Haven: Yale University Press.
- Marschak, Jacob. 1965. "Economics of Language." *Behavioral Science* 10 (2): 135–40.
- Maurais, Jacques, and Michael A. Morris, eds. 2003. *Languages in a Globalising World*. Cambridge: Cambridge University Press.
- Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110 (3): 681–712.
- Mavisakalyan, Astghik. 2015. "Gender in Language and Gender in Employment." *Oxford Development Studies* 43 (4): 403–24.
- Mavisakalyan, Astghik, and Clas Weber. 2018. "Linguistic Structures and Economic Outcomes." *Journal of Economic Surveys* 32 (3): 916–39.
- McMahon, April, and Robert McMahon. 2005. *Language Classification by Numbers*. Oxford: Oxford University Press.
- Melitz, Jacques. 2008. "Language and Foreign Trade." *European Economic Review* 52 (4): 667–99.
- Melitz, Jacques, and Farid Toubal. 2014. "Native Language, Spoken Language, Translation and Trade." *Journal of International Economics* 93 (2): 351–63.
- Merkel, Wolfgang, and Brigitte Weiffen. 2012. "Does Heterogeneity Hinder Democracy?" *Comparative Sociology* 11: 387–421.
- Michalopoulos, Stelios. 2012. "The Origins of Ethnolinguistic Diversity." *American Economic Review* 102 (4): 1508–39.
- Montalvo, José G., and Marta Reynal-Querol. 2002. "Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth." Available on SSRN at 394926.
- Montalvo, José G., and Marta Reynal-Querol. 2005. "Ethnic Polarization, Potential Conflict, and Civil Wars." *American Economic Review* 95 (3): 796–816.
- Moore, Karl, and David Lewis. 1999. *Birth of the Multinational: 2000 Years of Ancient Business History—From Ashur to Augustus*. Copenhagen: Copenhagen Business School Press.
- Nakhleh, Luay, Tandy Warnow, Don Ringe, and Steven N. Evans. 2005. "A Comparison of Phylogenetic Reconstruction Methods of an Indo-European Dataset." *Transactions of the Philological Society* 103 (2): 171–92.
- Nerbonne, John, and Wilbert Heeringa. 1997. "Measuring Dialect Difference Phonetically." In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*, edited by John Coleman, 11–18. Madrid: Special Interest Group of the Association for Computational Linguistics.
- Nettle, Daniel. 2000. "Linguistic Fragmentation and the Wealth of Nations: The Fishman-Pool Hypothesis Reexamined." *Economic Development and Cultural Change* 48 (2): 335–48.
- Osang, Thomas, and Shlomo Weber. 2017. "Immigration Policies, Labor Complementarities, Population Size and Cultural Frictions: Theory and Evidence." *International Journal of Economic Theory* 13 (1): 95–111.
- Ottaviano, Gianmarco I.P., and Giovanni Peri. 2005. "Cities and Cultures." *Journal of Urban Economics* 58 (2): 304–37.
- Pelikán, Pavel. 1969. "Language as a Limiting Factor for Centralization." *American Economic Review* 59 (4): 625–31.
- Phillipson, Robert. 1992. *Linguistic Imperialism*. Oxford: Oxford University Press.
- Phillipson, Robert. 2003. *English-Only Europe? Abingdon*: Routledge.
- Phillipson, Robert, and Tove Skutnabb-Kangas. 1995. "Language Rights in Postcolonial Africa." In *Linguistic Human Rights: Overcoming Linguistic Discrimination*, edited by Robert Phillipson, Mart Rannut, and Tove Skutnabb-Kangas, 336–46. Berlin: De Gruyter.
- Pinker, Steven. 1994. *The Language Instinct*. New York: Harper Perennial Modern Classics.

- Poloni, Estelle S., Ornella Semino, Giuseppe Passarino, A. Silvana Santachiara-Benerecetti, Isabelle Dupanloup, André Langaney et al. 1997. "Human Genetic Affinities for Y-chromosome P49a,f/TaqI Haplotypes Show Strong Correspondences with Linguistics." *American Journal of Human Genetics* 61 (5): 1015–35.
- Pool, Jonathan. 1972. "National Development and Language Diversity." In *Advances in the Sociology of Language*. Vol. 2, *Selected Studies and Applications*, edited by Joshua Fishman, 86–101. The Hague: Mouton.
- Prat, Andrea. 2002. "Should a Team Be Homogeneous?" *European Economic Review* 46 (7): 1187–207.
- Pym, Anthony, and Grzegorz Chrupala. 2004. "The Quantitative Analysis of Translation Flows in the Age of an International Language." In *Less Translated Languages*, edited by Albert Branchadell and Lovell Margaret West, 27–38. Amsterdam: John Benjamins Publishing Company.
- Regier, Terry, and Paul Kay. 2009. "Language, Thought, and Color: Whorf Was Half Right." *Trends in Cognitive Sciences* 13 (10): 439–46.
- Renfrew, Colin. 1987. *Archeology and Language: The Puzzle of Indo-European Origins*. Cambridge: Cambridge University Press.
- Reynal-Querol, Marta. 2002. "Ethnicity, Political Systems, and Civil Wars." *Journal of Conflict Resolution* 46 (1): 29–54.
- Rodriguez, Francisco, and Dani Rodrick. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In *NBER Macroeconomics Annual 2000*, Vol. 15, edited by Ben S. Bernanke and Kenneth Rogoff, 261–338. Cambridge: MIT Press.
- Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. 1991. "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review* 81 (5): 1068–95.
- Roy, Joydeep. 2005. "Redistributing Educational Attainment: Evidence from an Unusual Policy Experiment in India." Available on SSRN at 630141.
- Ruhlen, Merritt. 1994. *The Origin of Language: Tracing the Evolution of the Mother Tongue*. New York: John Wiley and Sons.
- Rustow, Dankwart A. 1967. *A World of Nations: Problems of Political Modernization*. Washington, DC: Brookings Institution.
- Samuelson, Beth Lewis, and Sarah Warshawer Freedman. 2010. "Language Policy, Multilingual Education, and Power in Rwanda." *Language Policy* 9 (3): 191–215.
- Saxenian, AnnaLee. 1999. *Silicon Valley's New Immigrant Entrepreneurs*. San Francisco: Public Policy Institute of California.
- Schwartz, Shalom H. 1992. "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries." *Advances in Experimental Social Psychology* 25: 1–65.
- Schwartz, Shalom H. 2014. "National Culture as Value Orientations: Consequences of Value Differences and Cultural Distance." In *Handbook of the Economics of Art and Culture*, Vol. 2, edited by Victor Ginsburgh and David Throsby, 257–86. Amsterdam: North-Holland.
- Selten, Reinhard, and Jonathan Pool. 1991. "The Distribution of Foreign Language Skills as a Game Equilibrium." In *Game Equilibrium Models*, Vol. 4, edited by Reinhard Selten, 64–87. Berlin: Springer.
- Senghor, Léopold Sédar. 1956. *Ethiopiques, Oeuvre Poétique*. Paris: Editions du Seuil.
- Senghor, Léopold Sédar. 1962. "Le Français, Langue de Culture." *Esprit* 311 (11): 837–44.
- Shannon, Claude E. 1948. "A Mathematical Theory of Communication." *Bell Systems Technical Journal* 27: 379–423, 623–56.
- Shapiro, Daniel M., and Morton Stelcner. 1997. "Language and Earnings in Quebec: Trends over Twenty Years, 1970–1990." *Canadian Public Policy* 23 (2): 115–40.
- Simovici, Dan A., and Szymon Jaroszewicz. 2002. "An Axiomatization of Partition Entropy." *IEEE Transactions on Information Theory* 48: 2138–42.
- Skutnabb-Kangas, Tove, and Robert Phillipson. 1989. "Mother Tongue: The Theoretical and Sociopolitical Construction of a Concept." In *Status and Function of Languages and Language Varieties*, edited by Ulrich Ammon, 450–77. Berlin: De Gruyter.
- Smailov, Arman. 2011. *Results of the 2009 National Population Census of the Republic of Kazakhstan: Analytical Report*. Astana, Kazakhstan: Agency on Statistics of the Republic of Kazakhstan.
- Spolsky, Bernard. 2004. *Language Policy: Key Topics in Sociolinguistics*. Cambridge: Cambridge University Press.
- Spolsky, Bernard, ed. 2012. *The Cambridge Handbook of Language Policy*. Cambridge: Cambridge University Press.
- Sutter, Matthias, Silvia Angerer, Daniela Glätzle-Rützler, and Philipp Lergetporer. 2015. "The Effect of Language on Economic Behavior: Experimental Evidence from Children's Intertemporal Choices." Center for Economic Studies (CESifo) Working Paper 5532.
- Swadesh, Morris. 1952. "Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos." *Proceedings of the American Philosophical Society* 96 (4): 121–37.
- Swoyer, Chris, Maria Baghramian, and Adam Carter. 2014. "Relativism: The Linguistic Relativity Hypothesis." In *Stanford Encyclopaedia of Philosophy Archived*, edited by Edward Zalta. Stanford: Stanford University Press.
- Tabellini, Guido. 2008. "Institutions and Culture." *Journal of the European Economic Association* 6 (2–3): 255–94.
- Taylor, Charles L., and Michael C. Hudson. 1972. *World Handbook of Political and Social Indicators*. New Haven: Yale University Press.
- Thierry, Guillaume, Panos Athanasopoulos, Alison Wigggett, Benjamin Dering, and Jan-Rouke Kuipers.

2009. "Unconscious Effects of Language-specific Terminology on Preattentive Color Perception." *Proceedings of the National Academy of Sciences of the United States of America* 106 (11): 4567–70.
- Thompson, Gordon B. 1977. "The World Turned Upside Down: Information Technology and the Linguistic Constraint." *Telecommunications Policy* 1 (2): 153–57.
- Tinbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund.
- Tönnies, Ferdinand. 1887. *Gemeinschaft und Gesellschaft*. Leipzig: Fues's Verlag.
- Uebelmesser, Silke, Matthias Huber, and Severin Weingarten. 2018. "The German Language Worldwide: A New Data Set on Language Learning." *CESIFO Economic Studies* 64 (1): 103–21.
- Vaillancourt, Francois, ed. 1985. *Economie et Langue: Recueil de Textes*. Québec: Editeur officiel du Québec.
- von Humboldt, Wilhelm. 1988. *On Language: The Diversity of Human Language-Structure and Its Influence on the Mental Development of Mankind*. Cambridge: Cambridge University Press. [1836].
- Walker, John, and Brigitte Unger. 2009. "Measuring Global Money Laundering: The Walker Gravity Model." *Review of Law and Economics* 5 (2): 821–53.
- Warnow, Tandy. 1997. "Mathematical Approaches to Comparative Linguistics." *Proceedings of the National Academy of Sciences of the United States of America* 94 (13): 6585–90.
- wa Thiong'o, Ngũgĩ. 1986. *Decolonizing the Mind: The Politics of Language in African Literature*. London: James Currey.
- Weber, Max. 1968. *Economy and Society: An Outline of Interpretive Sociology*. Berkeley, CA: University of California Press. [1910].
- Whorf, Benjamin Lee. 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.
- Wiener, Norbert. 1948. *Cybernetics: or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.
- Williams, Donald R. 2006. "The Economic Returns to Multiple Language Usage in Western Europe." Available on SSRN at 1028758.
- Wismann, Heinz. 2012. *Penser entre les Langues*. Paris: Albin Michel.
- World Values Survey. 2009. "WVS 1981–2008 Official 5-Wave Aggregate v.20090902." World Value Survey. Accessed December 19, 2017. www.worldvaluessurvey.org.
- Zink, Michel. 2014. "Ouverture. Quelle Langue est Mienne." In *D'Autres Langues que la Mienne*, edited by Michel Zink. Paris: Odile Jacob.