The role of language connectedness in reducing home bias in trade, investment, information, and people flows

Article (Accepted Version)

The role of language connectedness in reducing home bias in trade, investment, information, and people flows

Abstract

This study introduces the concept of a country's language connectedness (LC), namely, the extent to which the country is connected to the rest of the world in terms of the number of potential communicative partners. LC depends on the extent to which the country's languages are spoken outside that country. Operationalizing and constructing an index capturing LC, I empirically show that a country's LC is strongly associated with its globalization level. This effect is particularly strong in cross-border trade and investment and information flows. I also find that countries with languages belonging to large linguistic families (i.e., countries with greater linguistic connectedness) are more globalized. This study presents language barriers as a key contributor to home bias, that is, the tendency toward more within-border than cross-border interactions.

JEL classification: F2; F21; F23; F15

Keywords: language; globalization; home bias; economic integration; linguistic distance

1. Introduction

In spite of the hype about the "death of distance" and the world becoming a "global village," research demonstrates that national borders continue to limit international economic integration, and the extent of globalization differs significantly from country to country. Factors that affect a country's degree of participation in cross-border activities are of great interest to policymakers. A country's participation in cross-border activities is affected by the ability of its people—or at least those directly involved in cross-border activities—to speak the languages of other countries. This study examines the concept of language connectedness (LC), which is the extent to which a country is connected to the rest of the world in terms of the number of potential communicative partners. LC depends on the languages used in the focal country and the extent to which these languages are spoken outside the focal country. I operationalize and construct an index of LC using the recently released Ethnologue Global Dataset (Lewis et al., 2014), which characterizes the population of each of 234 countries and their 7,479 languages. The LC index considers each country's language repertoire and captures the extent to which all these languages are spoken outside that country. Based on this index, I empirically show that a country's LC is a key enabler of its globalization, and I thus posit language barriers as a key contributor to home bias, that is, the tendency to have more within-border than cross-border interactions (sometimes also referred as the border effect).

This study makes several important contributions to the literature. In contrast to previous studies that have investigated language differences among countries at either a unilateral or bilateral level (Tenzer et al., 2017), this study takes a multilateral perspective by evaluating each country's language repertoire against the language composition of the rest of the world. From this perspective, the idea of LC is novel, as is the LC index. Second, this study brings language to the forefront of discourses that attempt to explain the spatial dynamics of globalization. Scholars have looked at a variety of ways in which places are connected across space, the most prominent of which are distance, relative location, accessibility, and situation (Sheppard, 2002); however, it is rather surprising that the language aspect has received less attention here, particularly given the crucial role of language in facilitating cross-border interpersonal contact and interactions. Third, national competitiveness, a concept that has attracted wide academic interest, has been examined in various dimensions (Schueth, 2011; Schwab, 2014; Snowdon and Stonehouse, 2006); however, the linguistic dimension has been largely neglected. Moreover, national competitiveness depends on a country's ability to integrate itself within the world economy (Snowdon and Stonehouse, 2006). In examining how LC affects this

ability, this study makes another contribution. One can view the country of origin of various actors/entities, such as multinational enterprises (MNEs), engaged in cross-border activities as either a resource or a liability. Although researchers have looked at various aspects of home country–specific advantages (or disadvantages), such as the home country's institutional context, infrastructure, technological development, and product/market development (Moeller et al., 2013; Önsel et al., 2008), language has received limited attention. When significant portions of a country's population share a common language with the citizens of other countries, this can be an advantage for engaging in cross-border activities. This linguistic advantage—or the lack of it—can have implications for firms, individuals, and products/services that cross national boundaries. As Rose (2008) suggests, a country can also enjoy a kind of soft power when its language(s) is widely spoken outside the country. LC provides a more comprehensive concept for gauging the relative linguistic advantage (or disadvantage) of countries. I believe various research and policy domains, particularly discourses in international business/economics and national language policies, will find this research useful.

The remainder of the paper is organized as follows. Section 2 examines the literature relevant to the concept of country-level globalization (and home bias) and its determinants. Section 3 introduces the concept of a country's LC and proposes how a higher LC can lead to greater integration with the rest of the world (or a smaller home bias). An index capturing LC is operationalized and constructed in Section 4. Section 5 presents the methodology and the empirical results of the econometric analysis to assess the impact of LC on country-level globalization. Finally, Section 6 provides discussions and a conclusion.

2. Background

Scholars of economics, international business, and other fields have all contributed to the scholarly discussion of globalization, but they do so by emphasizing different dimensions. Economists and scholars of international business usually define globalization as a process of greater integration within the world economy through the movements of goods and services, capital, technology, and (to a lesser extent) labor (Jenkins, 2004). This study takes an economic perspective. As depicted in many portrayals, such as in "the world is flat," global village, death of distance, "end of geography," "death of the nation state," and "hyperglobalization," the last few decades have long been considered a period of accelerated globalization (albeit with moderation and possible reversal recently) by both academia and the popular press. However, economic activity still remains largely within national borders. For example, Stulz (2005) demonstrates that the extent of financial globalization is

limited, even though explicit barriers to financial globalization have significantly fallen since the 1940s. In the *Depth Index of Globalization* report, Ghemawat and Altman (2013) show that a preponderance of nearly every type of activity they measured takes place within national borders, reflecting what they term home bias. They also reveal that the extent of globalization differs significantly from country to country. Carrieri et al. (2013) concur that the degree of market integration is considerably lower for emerging economies than for developed markets, and Ajide et al. (2019) emphasize that the African region has a low level of globalization.

Two streams of research have extensively examined the causes of home bias in investment and trade, that is, why trade and investment disproportionately take place within the home country rather than in a foreign country (e.g., Daly and Vo (2013); Evans (2003); Mishra and Ratti (2013); Yi (2010)). In the *Depth Index of Globalization* report, Ghemawat and Altman (2013) focus on measuring the degree of globalization of countries, but they have statistically analyzed the relative explanatory power of five factors: the gross domestic product (GDP) per capita, remoteness, population, whether the country is landlocked, and linguistic commonality. They find that the first three factors explain more than 62 percent of the variation in their globalization index. Their linguistic commonality variable measures the proportion of the rest of the world's GDP sharing an official language with the country in question. The current study builds on and expands on Ghemawat and Altman's understanding of linguistic commonality. It does so by first considering the number of language speakers. LC also recognizes the complexity of language composition in countries. LC is more representative in terms of languages, since Ghemawat and Altman (2013) have only considered whether countries share an official language, let alone considering major languages. They also ignore the fact that people in different countries can communicate via a non-native language or a lingua franca. For example, English has emerged as the dominant lingua franca in international business and facilitates cross-border activities not only between English-speaking countries, but also between countries that have different languages, in what is known as an interlanguage effect, or lingua franca effect (Hejazi and Ma, 2011). Since English is spoken as a second language in many countries, consideration of only the official language, ignoring second language capabilities, can also introduce considerable bias.

3. A country's LC and its role in reducing home bias

Ample evidence suggests that language differences between two countries are one of the key barriers that obstruct trade, investment, and other types of cross-border transactions/flows between them. A range of studies

find a positive and significant relation between the presence of a common language between two countries and bilateral trade between them. In a review of studies on this effect, Anderson and van Wincoop (2004) estimate that language barrier amount to the equivalent of a tariff of 5–12 percent. Egger and Lassmann (2012) also find that a common language between two countries directly increases trade flows between them by 44 percent.

Similarly, language can play an important role in foreign direct investment (FDI). Effective interactions through a common language enable investors to distinguish good from bad investments and improve post-investment management and returns, thus increasing subsequent investments (Goldberg et al., 2005). Several studies have found strong evidence of a negative relation between the language differences between two countries and bilateral FDIs between them (Goldberg et al., 2005; Hejazi and Ma, 2011; Oh et al., 2011). Even in foreign portfolio investment (FPI), a mode of investment that demands relatively little foreign investor involvement, language appears to be a factor that determines investors' origin. Mishra (2007) finds that a common language between two countries increases equity investment between them by 35 percent, and Grinblatt and Keloharju (2001) show that investors prefer to hold and trade the securities of firms that publish their annual reports in the investors' native language.

Anecdotal evidence suggests that language also influences cross-border human interaction via telecommunications and Internet networks. For example, followers who retweet on Twitter are likely to share a common language with the original tweeter, and people who communicate in MSN Messenger are 27 times more likely to speak the same language as people who do not (Ghemawat and Altman, 2013). It is commonly accepted that language is one of the most important skills that determines immigrants' access to employment and earning levels (Danzer and Dietz, 2014; Toomet, 2011). Bartz and Fuchs-Schündeln (2012) show that language borders are one of the major reasons for the lack of labor market integration across borders within the EU-15 countries. Similarly, the ability to communicate in the destination country's language can be quite useful for international tourists and students.

Therefore, if language plays a crucial role in determining counterparts in cross-border activities/interactions, then, based on the same logic, the degree to which a country's languages are spoken abroad should influence the tendency to locate an activity within the home country, as opposed to outside it. For example, because of the status of English in the world, English-speaking countries could have a greater potential to integrate with the rest of the world. In contrast, countries dominated by languages not widely spoken outside could experience a larger home bias. For example, differences in languages can increase the liability of foreignness

of actors/entities engaging in cross-border activities and could restrict the extent of their cross-border activities (increasing home bias).

Every language has a different degree of reach around the globe. Given the network externality effects associated with languages, the more people use a particular language, the more useful and powerful that language becomes (Church and King, 1993; Grin, 2002). This effect can be contrasted with the utility–scarcity relation in commodities, resources, and other skills/competencies, which tend to gain in value and utility with scarcity. The network externality effect of language is likely applicable to both individuals and countries. Therefore, if a language(s) spoken in a country is widely spoken outside the country, then that country has an advantage in terms of connecting with the outside world. While researchers generally describe the network externality effect in relation to individuals, the current research tests this effect in relation to countries.

In sum, a country of people who can communicate with a larger fraction of the world will have a relative advantage in terms of engaging in cross-border activities and, therefore, a higher potential to engage with the outside world. Individuals and firms originating from such countries will be able to engage with a larger fraction of the world with ease and could thus have more opportunities in a wider geographical region outside their home country. Such countries can also attract the inbound activities from many potential communicative partners outside the country. In contrast, individuals and firms originating from countries where fewer residents can communicate with people in other countries could face greater liability of foreignness when they engage in cross-border activities. For example, a firm originating from a country that uses a vernacular language that is not spoken outside the home country is likely to be linguistically bound and is, therefore, likely to face significant challenges in expanding its activities beyond its national border. Therefore, countries with higher LC are likely to be more globalized (or to have a lower home bias), ceteris paribus.

4. Operationalization of LC

To construct a measure of LC that is useful for explaining the extent of globalization, we need to understand how a country's language composition can influence its potential for a cross-border interaction, as opposed to a within-country interaction. Accordingly, I attempt to capture the proportion of potential communicative partners outside the country as the share of total potential communicative partners within and outside the country.

To operationalize LC, I denote the world by the subscript *w*, the languages in the world by the subscript *j (j = 1, ..., n)*, the countries in the world by the subscript *i (i = 1, ..., k)*, and the population that speaks language *j* in country *i* by P_{ij}. For a given language *j*, all the speakers in the world, P_{wj}, are distributed both inside the country (P_{ij}) and outside the country ($P_{wj} - P_{ij}$), where

$$P_{wj} = \sum_{i=1}^{k} P_{ij} \quad (1)$$

Individuals in country *i* who speak language *j* have the opportunity to interact with a population P_{ij} inside the country and a population $P_{wj} - P_{ij}$ outside the country. Of the total number of potential communicative partners, the proportion of potential communicative partners outside the country is then:

$$\frac{P_{wj} - P_{ij}}{P_{wj}} \quad (2)$$

For further illustration, if we assume all the speakers of a language are distributed inside and outside the country according to a 20:80 ratio, then the distribution of potential communication partners will be 20 percent within the country and 80 percent outside the country. If the individual has no preference between choosing an outsider and an insider (ceteris paribus), then that person has a propensity of 80 percent of interacting with someone outside the country.

Given all the people in the country and extending the same construct to all languages *(j = 1, ..., n)*, the country's LC index is

$$LC_i = \sum_{j=1}^{n} \frac{(P_{wj} - P_{ij})}{P_{wj}} \times \frac{P_{ij}}{P_i} \quad (3)$$

where P_i is the total population in country *i* and P_{ij}/P_i represents the fraction of the population who speaks language *j* in country *i*. In essence, *LC* is constructed by calculating $(P_{wj} - P_{ij})/P_{wj}$ for each language and weighting this ratio by the fraction of the population in that country who speaks that language, P_{ij}/P_i, and, finally, taking the aggregate over all languages.

I illustrate the index with three simple examples. First, let us take a country with only one language. If no one outside the country speaks the language, then *LC* takes the value of zero, indicating a 0 percent potential of the population in that country communicating with the outside world, ceteris paribus. Second, if speakers of this language are distributed between inside and outside the country on a 50:50 basis, then *LC* takes the value

of 0.5, indicating a 50 percent potential of the population in the country communicating with the outside world, ceteris paribus. Third, let us take a country with two languages, where 60 percent of the population speaks the first language and 40 percent speak the second. If speakers of the first language are distributed inside and outside the country on a 10:90 basis and speakers of the second language on an 80:20 basis, then *LC* can be computed as follows:

$$LC = (0.60 \times 0.90) + (0.40 \times 0.20) = 0.54 + 0.08 = 0.62$$

Note that the contribution of each language to the LC index is different: this country's LC is mainly higher because of the first language. In this construction, the LC index can be larger than 1.0 because of the presence of bilingual/multilingual speakers. Compared to monolingual speakers, bilingual/multilingual speakers have greater linguistic capital and could have a broader linguistic reach, thus contributing to their countries' linguistic capital. For example, as per the Ethnologue Global Dataset, in Singapore about 61 percent speak English, 41 percent speak Mandarin, and 8 percent speak Malay, with many other minority languages also spoken. This multilingualism significantly increases Singapore's population's potential to connect with the outside world, for example, with countries where Mandarin, Chinese dialects, Malay, or English is spoken. Therefore, Singapore's LC index is above 1.0, at 1.248. Of course, the linguistic capital of bilingual and multilingual speakers depends on their languages being spoken by significant numbers of people outside their country. For instance, India has many bilingual and multilingual speakers, but it has an LC index of 0.23 because the international reach of languages such as Hindi is very limited (Maurais and Morris, 2003; McConnell, 2003).

I use the recently released Ethnologue Global Dataset to calculate the LC index. *Ethnologue: Languages of the World* is widely regarded as the most comprehensive reference volume on language currently available and is widely used (e.g., Bartz and Fuchs-Schündeln (2012)). Ethnologue released its Ethnologue Global Dataset for the first time in 2014, allowing researchers to directly use the data in their own analyses. It contains statistics on country-level language speakers for 234 countries and 7,479 languages. These data allow our calculations to encompass bilingual and multilingual speakers, a significant advantage over previous language-related studies that consider only first-language speakers, particularly since many internationally important languages are widely spoken as a second language. For example, of all the English language speakers in the Ethnologue dataset (840 million), only 40 percent speak it as a first language.

Table 1 presents the top five and bottom five countries in the LC index, while Appendix A reports the LC index for all the countries included in this study. Luxembourg leads the index, largely due to its multilingualism (Education First, 2015; Fettes, 2003). European countries dominate the top part of the table; seven of the top 10 countries are European. This could be attributed to the fact that most European countries explicitly promote multilingualism and intercultural communications among their citizens (Education First, 2015; Fettes, 2003; Rose, 2008). Japan is ranked last, reflecting the fact that Japanese is scarcely spoken outside of Japan and few Japanese speak English (Kaiser, 2003).

The LC index does not consider the fact that some languages are more closely related than others. Languages with a common origin tend to have similar syntaxes, and people find it easier to understand and learn languages that share a common origin with their first languages than those belonging to different families (Chiswick and Miller, 2005; Cuypers et al., 2015). To account for this, I construct another index to capture the extent to which the country's language(s) are closer to the languages outside the country. I use the genetic classification in the Ethnologue (Lewis et al., 2014), which classifies all the languages into 141 different language families (top-level genetic groups). Some language families are more extensive than others. For example, the top six language families (each comprising at least 5 percent of the world's languages) together account for nearly two-thirds of all languages and five-sixths of the world's population (Lewis et al., 2014). Countries with languages that belong to a major language family are likely to be more linguistically connected with the outside world than countries with languages that belong to a minority language family.

To account for the role of language families, I measure linguistic connectedness (LIC). The process is the same as for the LC index, but I consider the 141 different language families reported in Ethnologue instead of the 7,479 languages. In essence, to construct the index for a country, I consider each language family and measure the proportion of speakers of the languages of that language family who are outside the country; I then weight this quantity by the fraction of the population that speaks the languages of that language family; finally, I take the aggregate over all the language families.

Table 2 presents the top five and bottom five countries in the LIC index, while Appendix A reports the LIC index for all the countries included in this study. As for the LC index, European countries dominate the LIC index, largely due to the presence of multilingualism and because European languages belongs to the Indo-European language family, the most widespread language family in the world (Lewis et al., 2014). Some countries that have very low connectedness in terms of LC have better connectedness in terms of LIC. A

notable exception is Japan. Japan remains the most linguistically isolated country, with a very low LIC index (0.02), since the Japanese language belongs to the Japonic language family, which consists of languages that are largely spoken in Japan.

5. Empirical analysis

5.1 Data and methodology

Following Ghemawat and Altman (2013), I consider a country's integration with the rest of the world as manifested by its participation in international flows of products and services, capital, information, and people. I employ the country-level depth index of globalization (*DIG*) reported by Ghemawat and Altman (2013) to measure each country's degree of external integration with the rest of the world, the key dependent variable in this study. The variable *DIG* is a measure of the fraction of transactions that take place across borders out of all the transactions that take place within and across borders. This index is a multidimensional measure of globalization that is based on four pillars: trade (exports/imports of merchandise/services), capital (FDI and FPI), information (the Internet, phone calls, and printed publication trade), and people (migrants, tourists, and international students). For each activity, the extent of a country's cross-border flows is estimated compared to the overall (domestic and cross-border) measure in question or the domestic measure.[1] Unlike other globalization indexes, such as the KOF Index of Globalisation (Dreher, 2006), the index of Ghemawat and Altman (2013) distinguishes between cross-border flows and their enablers and only uses actual cross-border flows. The KOF Index of Globalisation includes such measures as import barriers, trade tariffs and taxes, the number of embassies, and international call tariffs, all of which are globalization enablers (Zinkina et al., 2013).

The key explanatory variable in this study is the LC index (*LC*). By construction, it corresponds well to the operationalization of *DIG* and is well suited for explaining it. The variable *DIG* measures the proportion of an economy's activities or flows that are international. In a similar analogy, the LC index captures the proportion of speakers of the country's language(s) outside the country. The LC index captures the potential of having a cross-border interaction via a common language versus a within-country interaction via a common language. If we control for other factors that can affect globalization, then we can separate the effect of LC on the globalization level. Therefore, I specify the following representation, controlling for country size,

[1] See Chapter 7 of Ghemawat and Altman (2013).

approximated by population; income level, approximated by the GDP per capita (*GDPPC*); economic growth rate (*GDPG*); geographical connectedness (*GC*); whether the country is landlocked; the extent of regional integration (*RTA Size*); human capital (*HC*); infrastructure; tax rate (*Tax*); trade tariffs (*Tariffs*) and the level of economic freedom (*EF*):

$$DIG = \beta_0 + \beta_1 LC + \beta_2 GC + \beta_3 Landlocked + \beta_4 RTA\ Size + \beta_5 Population + \beta_6\, GDPPC + \beta_7\, GDPG + \beta_8 Tax + \beta_9 Tariff + \beta_{10} Infrastructure + \beta_{11} HC + \beta_{12} EF \quad (4)$$

Appendix B provides the variable measurements and data sources. Geographical connectedness (*GC*), that is, physical closeness to the rest of the world, can facilitate globalization (Ghemawat and Altman, 2013). Controlling for geographical connectedness is particularly important, since countries with similar languages tend to be closely clustered. Based on a common approach to measure the geographical concentration of firms around a focal firm (e.g., Dai et al. (2013); Sorenson and Audia (2000), for each country i, I weight the GDP of all the others (where $j \neq i$ and $j = 1, \ldots, n$) by the inverse of their distance from the focal country[2] ($1/D_{ij}$) and then aggregate the weighted GDP measures:

$$GC_{it} = \sum_{j(\neq i)=1}^{N} \frac{GDP_{jt}}{D_{ij}} \quad (5)$$

The variable *Landlocked* is a binary variable that takes the value of one if the country is landlocked and zero otherwise. A landlocked country has national borders on all sides, which is a significant barrier to its integration with countries beyond its neighbors (Collier and Gunning, 1999). Regional integration can directly increase trade and facilitate cross-border investment and people flows (Medvedev, 2012; Moshirian, 2008). The variable *RTA Size* represents the extent to which the country is regionally integrated, operationalized by the cumulative market size that is linked through regional trade agreements (RTAs). Following Medvedev (2012), for each country, I aggregate the GDPs of all the countries that share an RTA with the focal country, as follows:[3]

$$RTA\ Size_{it} = \sum_{j(\neq i)=1}^{N} (GDP_{jt} \times RTA_{ijt}) \quad (6)$$

[2] The GDP data are from the World Development Indicators (WDI) and the distance data are from Rose and Spiegel (2011).

[3] The GDP data are from WDI and the data on RTAs are from Rose and Spiegel (2011) and the World Trade Organization.

where RTA_{ijt} is a binary variable that takes the value of one if country *i* and country *j* share an RTA at time *t*.

Population, *GDPPC* (GDP per capita), and *GDPG* (GDP growth rate) represent the size, income level, and economic growth rate, respectively.[4] Country size tends to be inversely related to globalization, while the country's income level and economic growth rate tend to be positively related to globalization (Ghemawat and Altman, 2013; Lane and Milesi-Ferretti, 2008; Moshirian, 2008). The variables *Tax* and *Tariff* are the country's total tax rate and average rate of trade tariffs, respectively.[5] Lower trade tariffs can facilitate both imports and exports, as well as attract export-oriented FDI. Lower tax rates can increase the competitiveness of exports and attract foreign investments (Tanzi, 1996). Various aspects of a country's domestic environment and competitiveness in terms of skills and resources are likely to offer incentives (or disincentives) for cross-border activities. In particular, a country's institutional environment, infrastructure facilities, and human capital skills, for example, can entice foreign firms to locate their value-added activities there (Globerman and Shapiro, 2002; Kinda, 2010; Moshirian, 2008) and facilitate the expansion and global competitiveness of native firms (Makhija et al., 1997). Therefore, I control for the quality of a country's overall infrastructure, human capital level, and economic freedom, and I expect these to have a positive effect on the globalization level. The variable *Infrastructure* is a multi-indicator measure of infrastructure.[5] It is a composite variable based on the quality of the country's overall infrastructure, roads, railroad infrastructure, port infrastructure, and air transport infrastructure; the number of available airline seat kilometers; the quality of its electricity supply; and the numbers of mobile telephone subscriptions and fixed telephone lines. The variable *HC* represents the level of human capital as measured by the gross secondary school enrollment rate. The variable *EF* is the Index of Economic Freedom,[6] which is based on the following 10 dimensions: business freedom, fiscal freedom, government spending, monetary freedom, trade freedom, investment freedom, financial freedom, property rights protection, freedom from corruption, and labor freedom. Finally, I include a series of year dummies to control for year-specific effects. The sample covers 114 countries[7] and the period 2006–2012.

Table 3 presents summary statistics and correlation coefficients. The mean value of the LC index is 0.66, with a standard deviation of 0.43, indicating considerable variation in the LC index across countries. We can

[4] Source: WDI.
[5] Source: Global Competitiveness Index.
[6] Source: Heritage Foundation.
[7] The choice of countries was based on data availability.

make a similar observation in terms of the LIC index. The LC and LIC indexes are positively correlated with the depth index of globalization, although we cannot ascertain the effect of the LC and LIC indexes without controlling for the other determinants of globalization.

Before the empirical results are presented, a few econometric concerns require attention. I cannot use country fixed effects, since this precludes the inclusion of time-invariant variables (such as *LC*). I use an exhaustive set of country-specific variables, thereby minimizing the chances of omitting country-specific variables that could be correlated with explanatory variables. I follow previous studies that use panel models with time-invariant variables (e.g., Degryse et al. (2012)). The second concern is the potential endogeneity between the dependent variable and some of the control variables. For example, the level of globalization could be endogenous to the economic growth rate, in that faster-growing economies could be more likely to become integrated, and integrated economies could be more likely to grow faster. To mitigate the effects of such potential endogenous relations and maintain the simplicity and uniformity of the analysis, I lag *Population*, *GDPPC*, *GDPG*, *HC*, and *EF* by one year. Lagging potentially endogenous explanatory variables is a common method used to mitigate endogeneity issues (Carrieri et al., 2013; Desbordes, 2007). Globalization could affect a country's composition of language speakers; for example, the number of people speaking English in a host country could increase because of incoming FDI from English-speaking countries. However, the impact of reverse causality is likely to be minimal, since these changes occur very slowly and their magnitude is unlikely to be considerable. Therefore, *LC* is considered an exogenous variable. I believe this assumption to be reasonable, given that language is usually considered a slowly changing variable. All estimations are estimated with robust standard errors to control for heteroscedasticity.

5.2. Results

The estimated results of the main specification are shown in column (1) in Table 4. The variable *LC* is positive and significant ($p < 0.001$), providing strong evidence of the positive impact of the LC of a country on its degree of globalization. Of the control variables, *GC* is positive and significant ($p < 0.05$), indicating that countries with greater geographical proximity to the rest of the world are more globalized. Similarly, *GDPG* ($p < 0.05$), *Infrastructure* ($p < 0.001$), and *HC* ($p < 0.05$) are positive and significant, indicating that higher economic growth and better infrastructure and human capital can facilitate a country's globalization. *Population* is negative and significant ($p < 0.01$), indicating that smaller countries tend to be more integrated.

GDPPC and *EF* are positive, as expected (Ghemawat and Altman, 2013), and *Tax*, *Tariff*, and *Landlocked* are negative, as expected, but none of these measures is significant. Contrary to our expectations, *RTA Size* is nonsignificant. Countries that form trade blocks tend to be closely clustered (Kohl and Brouwer, 2014), suggesting that most of the effects of *RTA Size* could be already captured by *GC*. This could also be due to the recent proliferation of RTAs[8] (Subramanian and Kessler, 2013). Not all RTAs are equally effective (Arribas et al., 2011) and some RTAs are political gestures rather than effectively implemented integration agreements. Table 5 presents a summary of the expected and estimated effects of all the control variables.

The beta coefficients, BC,[9] can be used to compare the effects of each explanatory variable on the globalization index. The variable *LC* has the largest BC (0.29), followed by *GC* (0.24), *HC* (0.16), *Infrastructure* (0.16), *Population* (0.15), and *EF* (0.07). This result clearly shows that LC is strongly related to the extent of a country's globalization.

It would be interesting to know whether the importance of LC has changed over time, particularly since technological developments in communication could reduce the effect of language differences over time. Therefore, I interact the LC index with the discrete variable *Time*, which changes from zero to seven as the year changes from 2006 to 2012. The results are reported in column (2) of Table 4. The interaction term is negative and significant, which indicates that the effect of LC has reduced over time. However, the coefficient of the interaction term (-0.223) is much smaller compared to the coefficient of the LC index (8.033); therefore, the language effect is diminishing only very slowly.

Given the proliferation of the English language and its increasing use as a global lingua franca, one could argue that the empirical effect of LC captured here is largely a manifestation of the impact of the English language on globalization. To check this possibility, I deconstruct the LC index into two indexes, one for LC through the English language (*LC English*) and the other for LC through all the other languages (*LC Other*) and re-estimate the baseline model with them. The results are reported in column (3) in Table 4. Both indexes are positive, with *LC Other* having a higher significance level ($p < 0.001$) than *LC English* ($p = 0.12$), which is marginally nonsignificant. Therefore, we can be reassured that the effect of LC that we observe is not a manifestation of the effect of the English language alone.

[8] More than a half of the country pairs in the sample in this study shared an RTA.
[9] The beta coefficient of the explanatory variable is equal to the product of the estimated coefficient of the explanatory variable and the ratio of the standard deviation of the explanatory variable to the standard deviation of the dependent variable.

Some languages can be more powerful than others, such as those predominantly used in the business world and legal system. To address this aspect, I conduct a further analysis to account for the power of languages based on the Power Language Index (Chan, 2016a). This index is constructed based on 20 indicators to measure the world's most influential languages according to five basic opportunities offered by language: geography, economy, communication, knowledge and media, and diplomacy. Based on the Power Language Index, I divide all the languages into two groups: 1) the 10 most powerful languages and 2) the remainder of the languages (Chan, 2016b). I then construct two LC indexes, *LC Power 10* and *LC Power Rest*, similar to when I compared English (*LC English*) and the rest of the languages (*LC Other*). I then carry out another analysis based on these two LC indexes. The results are reported in column (4) in Table 4. Both *LC Power 10* and *LC Power Rest* are significant ($p < 0.01$ and $p < 0.001$, respectively), while the estimated coefficient and the significance level for *LC Power Rest* are larger than for *LC Power 10*. This result shows that that the effect of LC that we observe is not a manifestation of the effect of powerful/major languages alone.

I rely on the total population of countries to calculate the LC index; however, such country-level capabilities might not always represent the language capabilities of those directly involved in cross-border activities, such as when "international languages" are predominantly spoken by elites/groups that are primarily involved in cross-border activities. To mitigate this bias, I carry out another robustness test based on second-language speakers, instead of considering all the language speakers in a country. The logic behind using second-language speakers is that those operating in investing/invested firms and as importers/exporters are more likely to be bilingual. Ethnologue classifies language speakers into two categories: first-language (L1) users and second-language (L2) users. The total number of users of a language in a country is the sum of the numbers of L1 and L2 users. When constructing the LC index (equation (3)), I consider all the language speakers in the country; that is, P_{ij}/P_i in equation (3) represents the fraction of the population that speaks language *j* in country *i*. I replace this term with the fraction of the population that speaks language *j* as a second language in country *i* (L2) and construct another index, that is, *LC(L2)*. In essence, *LC(L2)* is constructed by calculating the value of *(Pwj - Pij)/ Pwj* for each language and weighting it by the fraction of the population in that country that speaks that language as a second language and, finally, taking the aggregate over all the languages. I then carry out another analysis based on this index. The results are reported in column (5) in Table 4. *LC(L2)* is positive and significant.

To make sure that the empirical effect of LC captured here is not a manifestation of the impact of multilingual countries—since countries in which the population speaks multiple languages are more globalized than other countries—I carry out a further robustness test by including the total number of second-language speakers (L2) in the country as a percentage of the total population (*L2 Speakers*) alongside the LC index. The results are reported in column (6) in Table 4. This variable *L2 Speakers* is nonsignificant, while *LC* remained significant. This result shows that being a "multilingual country" (rather than a country whose language is spoken elsewhere) alone is not enough to be more interconnected.

To make sure that the LC–globalization relation observed is not driven by a few types of cross-border activities and to investigate the impact of LC on different cross-border activities, I extend the analysis to the four pillars included in the depth index of globalization. In all the following estimations, the dependent variable measures the extent of international flows in that activity, compared to the overall (domestic and cross-border) measure in question or the domestic measure (see Appendix B for the operationalization of each dependent variable). The estimated results for the trade and investment pillars are reported in Table 6. First, *LC* is positive and significant for overall trade intensity ($p < 0.01$). In terms of the different trade categories, *LC* is positive for all of them and significant for all of them except merchandise exports and services imports. The estimated beta coefficient of LC is significantly higher for services trade (0.44) than for merchandise trade (0.25), showing that language is more important for services than for merchandise trade. Among the different trade categories (services/merchandise and exports/imports), the highest beta coefficient for *LC* is for services exports (0.53), indicating the crucial role of LC in attaining more services exports. In terms of the control variables,[10] *GC* remains positive and significant in all the estimates, clearly showing the importance of geographical connectedness in trade intensity. Except for *GDPPC* and *Tariff*, all the control variables perform similarly as in the previous results. The variable *GDPPC* has now become negative and significant, indicating that countries with lower incomes tend to have greater trade intensity. Unsurprisingly, *Tariff* is now significant for merchandise trade.

The variable *LC* is positive for both FDI and FPI, showing the importance of LC for cross-border investment flows. However, *LC* is significant only for FDI flows. This result could reflect the smaller role of language in FPI, which demands less foreign investor involvement and is less knowledge intensive than FDI. In contrast, language plays a crucial role in FDI in both the selection and the management of international ventures. The

[10] Due to space constraints, the results for the control variables are not tabulated here but are available upon request.

variable *GC* is only significant for FDI outflows, indicating that geographical connectedness is less crucial for FDI relative to trade.

Table 7 reports the estimated results for the information and people pillars. Again, *LC* is positive for all types of cross-border information flows and significant for all except outgoing telephone calls. The variable *GC* remains positive and significant in all cases, except for incoming telephone calls. The variable *LC* is positive for all types of cross-border people flows, but significant only for inbound tourists. Of the three types of people flows, the standard error of the estimated coefficient for long-term movement (migration) is much higher relative to that for medium-term (university students) and short-term (tourism) movements. These results could indicate that long-term movements are the least sensitive to language.

Finally, I re-estimate the results using the LIC index. Column (7) in Table 4 provides the results of the main specification. As with the LC index, *LIC* is positive and significant ($p < 0.05$). Therefore, countries in which people speak languages belonging to large linguistic families are more globalized. Table 8 reports the estimated impact of *LIC* on different cross-border activities.[11] Although these results are qualitatively similar to those for *LC*, they are generally less significant. A notable deviation from the previous set of results is that *LIC* is largely nonsignificant for international telephone usage, which could reflect the fact that linguistic proximity does not facilitate direct communication. In contrast, *LIC* remains very significant for trade intensity in printed publications, probably reflecting the ease of translation between linguistically close languages.

6. Conclusion

Due to many benefits attributed to globalization, the barriers and enablers of globalization are of interest to policymakers and firms. The findings of this study provide strong support for the premise that a country's globalization level depends on its LC to the rest of the world. LC is found to have a larger effect than other determinants of globalization. Its effect is also evident in all four pillars, particularly in trade, investment, and information flows. In essence, these results strongly indicate that a country's language profile is a key determinant of its destiny in the globalization process.

The impact of LC is not solely driven by the English language. Interestingly, LC through other languages has a more significant effect than LC through the English language. Although English is widely recognized as

[11] For brevity, only the estimated coefficient on LIC is reported, but all the specifications include all the control variables.

the dominant lingua franca in international business, these results indicate that other popular languages, such as French, Spanish, and Portuguese, can also play a significant role in the globalization process.

The impact of LC is stronger for services trade than for merchandise trade. This finding aligns with anecdotal evidence that shows some countries with low LC have significant participation in global trade and FDI in the manufacturing sectors, but are less active in trade/FDI in the services sectors. For example, Japan is ranked last in the LC index (Appendix A) and most Japanese MNEs are largely product driven (Neeley, 2013). China is ranked fifth from the bottom in terms of the LC index, and the majority of Chinese exports and FDI has taken place in the manufacturing sectors. Services require simultaneity, inseparability, and perishability, characteristics that do not apply to product manufacture (Boddewyn et al., 1986). Therefore, service delivery requires simultaneous interpersonal communications—typically face-to-face contact—and active involvement between customers and service providers, which makes language traits more important for services trade than for manufacturing trade.

The results also show that a country's geographical connectedness to the rest of the world strongly influences its globalization level. This result corroborates those of Nachum et al. (2008), who have found that US MNEs make greater FDIs in countries that are proximate to the world's knowledge and markets. This study shows that similar dynamics manifest in other forms of inbound and outbound cross-border flows.

6.1 Implications of the study

This study has many implications for national policymakers (countries), firms, and individuals. First, by identifying the crucial enablers and barriers to globalization, it provides insights for policymakers who wish to foster globalization. Countries differ in terms of their linguistic endowments largely due to their different exposures to conquests, slavery, migrations, and colonization (Michalopoulos, 2012), and some countries are fortunate to have inherited widely spoken languages. However, deliberate language policy choices can also affect LC. For example, the Eastern European countries that were previously part of the Soviet Union changed their official language from Russian to national languages (Danzer and Dietz, 2014). In contrast, English (which was introduced during the British occupation) remained the official language of Singapore and India. Policymakers in countries that rely primarily on vernacular languages that are rarely spoken outside the country could follow suit, or otherwise support second-language competency in a widely spoken language in the world or a lingua franca. National policymakers need to make language learning central to their education

policies, and difficult choices could be necessary in terms of deciding what languages need to be promoted and which ones discouraged. The findings of this study also imply that a country being multilingual alone is not enough to have greater interconnectedness; therefore, choosing the right language(s) is particularly important. Of course, such a language choice should only be made after careful consideration of many other factors/concerns, such as pedagogical efficiency and linguistic democracy; however, the influence of LC on globalization documented here provides vital information for evaluating such language policy decisions. A less controversial approach could be to simply encourage additional languages (relevant business languages and widely spoken languages) to be taught in schools.

The LC index provides a unique gauge to measure the relative language capabilities (and barriers) of countries, particularly in terms of their capacity to engage in cross-border activities. The example of Japan, once called the factory of the world, is suggestive. The country ranks last on the LC and LIC indexes and 113rd out of 139 countries in the globalization index. Japan's linguistic isolation is likely to pose serious challenges to its process of globalization and its future economic prosperity, particularly due to the rising importance of services in globalization. A similar analogy can be made for China, which ranks very low in the LC and globalization indexes. With the growth of services and knowledge-intensive sectors and the increase in the complexity of spatial transactions requiring face-to-face communications and shorter lead times (Kohl and Brouwer, 2014), the problem associated with lower LC could become more severe for such countries.

This study also has implications for supranational institutions (both global and regional) that advocate greater integration among countries. If LC is the most crucial enabler of globalization, then the important question is, what is the best geolinguistic strategy to make the world more interlingual or linguistically integrated? Is it the promotion of universal multilingualism or an attempt to achieve hegemonic world monolingualism, probably through English? Or a mix of both strategies, or the promotion of a few auxiliary international languages? There is no easy answer to these questions, since each strategy has its own set of pros and cons; for example, hegemonic world monolingualism clashes with the notion of linguistic democracy, and multilingualism comes at a higher cost and has greater inefficiency (Ginsburgh and Weber, 2005; Maurais, 2003). Nevertheless, LC is largely structural in nature. The implementation of language policies can be quite challenging, since languages are deeply embedded in the roots of individuals and countries and changing them is quite difficult, particularly in the short/medium term. The real impact of language policy becomes visible

decades and sometimes generations after its implementation. Language barriers remain the stickiest hindrance to globalization, even as technology and open policies continue to encourage the flattening of the world.

Finally, this study has implications for firms (managerial implications) and individuals. Both firms and individuals that originate from countries with low LC could find it difficult to engage in cross-border activities due to the language barrier, and their destination choices will be limited due to linguistic boundaries. To surmount the problem, such firms might have to invest in training their staff in a widely spoken language or a lingua franca.

6.2. Limitations of the study

Despite its merits, this study is not without limitations. First, it should be noted that language statistics are by no means complete and up-to-date for all the countries, particularly when it comes to the statistics of second-language speakers. Nevertheless, since Ethnologue is the most comprehensive source of language statistics, and given its large coverage of languages and countries and good coverage of second-language speakers, I believe that the LC index calculated in this study is a reliable measure. It is also reassuring that the ratio of total language speakers in a country to its population is greater than 1.0 for most countries and greater than 2.0 for some, indicating good coverage of bilingual and multilingual speakers. Second, I rely on the total populations of countries to calculate the LC index, and such country-level capabilities might not always represent the language capabilities of those directly involved in cross-border activities. However, I try to mitigate this bias by using the number of second-language speakers instead of considering all the language speakers in the country. The third limitation relates to the LIC index. Although I use Ethnologue's genetic classification (top-level genetic groups) and assume that the languages within these groups are more closely related and the languages between these groups more distantly related, I do not account for differences between sublevel language groups. Additionally, there can be instances in which the language family affiliations are not that clear, particularly when it comes to hybrid languages that could have more than one origin. However, it is difficult to account for such nuances.

Appendix A: The LC and LIC indexes and the depth index of globalization

Rank	Country	LC index	LIC index	Depth index of globalization (2012)
1	Luxembourg	1.96	2.47	46
2	Belgium	1.58	1.96	43
3	Denmark	1.54	2.50	34
4	Sweden	1.51	2.19	35
5	Finland	1.38	2.10	30
6	Bolivia	1.37	1.61	15
7	Paraguay	1.33	1.37	16
8	Slovak Republic	1.28	1.99	32
9	Moldova	1.27	1.41	26
10	Latvia	1.25	2.27	32
11	Singapore	1.25	1.25	47
12	Lithuania	1.23	2.11	33
13	Honduras	1.17	1.19	24
14	Hong Kong	1.17	1.26	50
15	Panama	1.15	1.18	35
16	United Kingdom	1.14	1.23	28
17	Uruguay	1.11	1.12	15
18	Estonia	1.10	1.85	37
19	New Zealand	1.10	1.16	25
20	El Salvador	1.08	1.10	18
21	Cyprus	1.08	1.17	31
22	Slovenia	1.06	2.05	33
23	Ireland	1.05	1.27	43
24	Canada	1.03	1.06	28
25	Botswana	1.02	1.14	21
26	Argentina	1.02	1.09	9
27	Guatemala	1.00	1.13	15
28	Dominican Republic	0.99	1.01	16
29	Portugal	0.99	1.04	28
30	Croatia	0.99	1.59	26
31	Germany	0.98	1.74	30
32	Nicaragua	0.98	0.99	25
33	Australia	0.97	0.99	21
34	Costa Rica	0.96	0.97	23
35	Colombia	0.94	1.02	12
36	Ecuador	0.94	1.05	14
37	Lebanon	0.93	1.18	34
38	Spain	0.90	1.23	25
39	Chile	0.90	0.92	26
40	Peru	0.89	1.02	15
41	France	0.86	1.20	26
42	Mauritius	0.86	1.37	32
43	Czech Republic	0.81	1.63	33
44	Greece	0.80	1.61	19
45	Mexico	0.77	0.98	20
46	Switzerland	0.77	1.21	36
47	United States	0.77	0.96	18
48	Malaysia	0.71	1.31	37
49	Kazakhstan	0.70	0.93	25
50	Italy	0.66	1.83	24
51	Tunisia	0.66	1.53	24
52	Nepal	0.63	1.29	6
53	Austria	0.62	1.13	36
54	Armenia	0.56	1.11	21
55	Kyrgyz Republic	0.55	0.94	22
56	Jordan	0.55	0.98	30
57	Serbia	0.52	1.33	25
58	Bulgaria	0.49	1.35	31

Appendix A (continued)

Rank	Country	LC index	LIC index	Depth Index of globalization (2012)
59	Bangladesh	0.49	1.78	4
60	Mozambique	0.48	1.12	17
61	Philippines	0.48	1.45	13
62	Albania	0.47	1.18	24
63	Romania	0.46	1.24	21
64	Israel	0.45	1.14	29
65	South Africa	0.45	2.24	21
66	Poland	0.45	1.34	28
67	Venezuela	0.44	0.46	7
68	Hungary	0.42	0.76	36
69	Nigeria	0.42	1.13	16
70	Netherlands	0.40	1.17	42
71	Georgia	0.39	0.44	26
72	Korea, Republic	0.38	0.38	29
73	Namibia	0.37	1.06	24
74	Benin	0.36	0.98	10
75	Pakistan	0.35	1.37	5
76	Senegal	0.35	1.18	16
77	Macedonia	0.32	0.97	28
78	Morocco	0.32	1.14	22
79	Guinea	0.31	0.99	19
80	Russian Federation	0.29	1.05	17
81	Ukraine	0.28	0.89	27
82	Azerbaijan	0.28	1.70	20
83	Burkina Faso	0.28	0.86	10
84	Sri Lanka	0.28	1.14	10
85	Malawi	0.27	0.91	15
86	Chad	0.26	0.73	16
87	Oman	0.26	0.57	30
88	India	0.23	0.71	10
89	Mali	0.22	1.58	11
90	Bahrain	0.22	0.47	37
91	Kuwait	0.21	0.25	27
92	Cameroon	0.18	0.65	10
93	Jamaica	0.17	0.95	24
94	Kenya	0.16	0.97	13
95	Turkey	0.16	0.85	15
96	Iran	0.15	1.14	2
97	Uganda	0.15	0.85	10
98	Angola	0.14	0.82	17
99	Saudi Arabia	0.14	0.59	24
100	Cambodia	0.13	0.85	30
101	Rwanda	0.12	0.67	9
102	Qatar	0.12	0.12	27
103	Ghana	0.10	1.07	19
104	Egypt	0.10	0.81	11
105	Ethiopia	0.08	0.79	7
106	Guyana	0.07	0.89	33
107	Burundi	0.07	1.27	5
108	Mongolia	0.07	0.97	31
109	Madagascar	0.05	0.81	12
110	Norway	0.05	0.95	31
111	China	0.04	0.12	10
112	Thailand	0.04	0.46	31
113	Iceland	0.03	0.60	33
114	Indonesia	0.03	0.52	9
115	Japan	0.02	0.02	12

Appendix B: Variable descriptions, measurement, and data sources

Variable	Description/Measurement	Data Source
LC	Language connectedness index	Constructed using language data obtained from Lewis et al. (2014)
LIC	Linguistic connectedness index	
DIG	Depth index of globalization	Ghemawat and Altman (2013)
GC	Geographical connectedness, constructed using the GDP (WDI) and geographical distance (Rose and Spiegel (2011))	WDI, Rose and Spiegel (2011)
Landlocked	A binary variable that equals unity if the country is landlocked	Central Intelligence Agency World Factbook
RTA Size	The cumulative GDP of all the countries with which the focal country shares an RTA, constructed using the GDP (WDI) and the RTA (Rose and Spiegel (2011)) for 2006 and updated 2007–2012 from the World Trade Organization)	WDI, Rose and Spiegel (2011); World Trade Organization
Population	Population	WDI
GDPPC	GDP per capita	
GDPG	GDP growth rate	
HC	Human capital measured as secondary school enrollment (percent gross)	
Tax	Total tax rate	Global Competitiveness Index
Tariff	Average rate of trade tariffs	
Infrastructure	Multi-indicator measure of infrastructure based on the quality of the overall infrastructure, roads, railroad infrastructure, port infrastructure, and air transport infrastructure; number of available airline seat kilometers; the quality of the electricity supply; and the numbers of mobile telephone subscriptions and fixed telephone lines	
EF	Index of Economic Freedom	Heritage Foundation (http://www.heritage.org/index/)
Trade	Total trade as a percentage of the GDP	Ghemawat and Altman (2013)
Merchandise Trade	Total merchandise trade as a percentage of the GDP	
Services Trade	Total trade of commercial services as a percentage of the GDP	
Merchandise Exports	Total merchandise exports as a percentage of the GDP	
Merchandise Imports	Total merchandise imports as a percentage of the GDP	
Services Exports	Total exports of commercial services as a percentage of the GDP	
Services Imports	Total imports of commercial services as a percentage of the GDP	
FDI Outflows	FDI outflows as a percentage of gross fixed capital formation (average of the current year and the two previous years)	
FDI Inflows	FDI inflows as a percentage of gross fixed capital formation (average of the current year and the two previous years)	
Portfolio Equity Outflows	Equity securities assets (net) as a percentage of the GDP (average of the current year and the two previous years)	
Portfolio Equity Inflows	Equity securities liabilities (net) as a percentage of the GDP (average of the current year and the two previous years)	
Internet Bandwidth	International Internet bandwidth per Internet user	
Outgoing Telephone	Total outgoing telephone calling minutes per capita (TDM + VoIP).	
Incoming Telephone	Total incoming telephone calling minutes per capita (TDM + VoIP)	
Printed Publications Exports	Total exports of HS49 per capita, where HS49 includes printed books, newspapers, pictures, manuscripts, typescripts, and plans	
Printed Publications Imports	Total imports of HS49 per capita, where HS49 includes printed books, newspapers, pictures, manuscripts, typescripts, and plans	
Outbound Migrants	International outbound migrants' share of the population.	
Inbound Migrants	International inbound migrants' share of the population.	
Outgoing International Students	Total number of students studying abroad as a percentage of the total number of tertiary students in the country	

Incoming International Students	Total number of foreign students as a percentage of the total number of tertiary students in the country	
Inbound Tourists	Inbound tourists, that is, the number of arrivals of non-resident overnight visitors (tourists) at national borders as a percentage of the total population	
Outbound Tourists	Outbound tourists, that is, the number of departures of overnight visitors (tourists) as a percentage of the total population	

REFERENCES

Ajide, K.B., Raheem, I.D., Asongu, S.A., 2019. Dollarization and the "unbundling" of globalization in Sub-Saharan Africa. Research in International Business and Finance 47, 398-409.

Anderson, J.E., van Wincoop, E., 2004. Trade Costs. Journal of Economic Literature 42, 691-751.

Arribas, I., Pérez, F., Tortosa-Ausina, E., 2011. A new interpretation of the distance puzzle based on geographic neutrality. Economic Geography 87, 335-362.

Bartz, K., Fuchs-Schündeln, N., 2012. The role of borders, languages, and currencies as obstacles to labor market integration. European Economic Review 56, 1148-1163.

Boddewyn, J.J., Halbrich, M.B., Perry, A.C., 1986. Service multinationals: Conceptualization, measurement and theory. Journal of International Business Studies 17, 41-57.

Carrieri, F., Chaieb, I., Errunza, V., 2013. Do implicit barriers matter for globalization? Review of Financial Studies 26, 1694-1739.

Chan, K.L., 2016a. Power Language Index. http://www.kailchan.ca/wp-content/uploads/2016/12/Kai-Chan_Power-Language-Index-full-report_2016_v2.pdf (Accessed 09 October 2019)

Chan, K.L., 2016b. These are the most powerful languages in the world, World Economic Forum. https://www.weforum.org/agenda/2016/12/these-are-the-most-powerful-languages-in-the-world/ (Accessed 09 October 2019)

Chiswick, B.R., Miller, P.W., 2005. Linguistic distance: A quantitative measure of the distance between English and other languages. Journal of Multilingual and Multicultural Development 26, 1-11.

Church, J., King, I., 1993. Bilingualism and network externalities. The Canadian Journal of Economics / Revue canadienne d'Economique 26, 337-345.

Collier, P., Gunning, J.W., 1999. Why has Africa grown slowly? The Journal of Economic Perspectives 13, 3-22.

Cuypers, I.R.P., Ertug, G., Hennart, J.-F., 2015. The effects of linguistic distance and lingua franca proficiency on the stake taken by acquirers in cross-border acquisitions. Journal of International Business Studies 46, 429-442.

Dai, L., Eden, L., Beamish, P.W., 2013. Place, space, and geographical exposure: Foreign subsidiary survival in conflict zones. Journal of International Business Studies 44, 554-578.

Daly, K., Vo, X.V., 2013. The determinants of home bias puzzle in equity portfolio investment in Australia. International Review of Financial Analysis 27, 34-42.

Danzer, A.M., Dietz, B., 2014. Labour migration from Eastern Europe and the EU's quest for talents. JCMS: Journal of Common Market Studies 52, 183-199.

Degryse, H., Havrylchyk, O., Jurzyk, E., Kozak, S., 2012. Foreign bank entry, credit allocation and lending rates in emerging markets: Empirical evidence from Poland. Journal of Banking & Finance 36, 2949-2959.

Desbordes, R., 2007. The sensitivity of U.S. multinational enterprises to political and macroeconomic uncertainty: A sectoral analysis. International Business Review 16, 732-750.

Dreher, A., 2006. Does globalization affect growth? Evidence from a new index of globalization. Applied Economics 38, 1091-1110.

Education First, 2015. EF English Proficiency Index. Education First. https://www.ef.com/__/~/media/centralefcom/epi/downloads/full-reports/v5/ef-epi-2015-english.pdf (Accessed 26 November 2015)

Egger, P.H., Lassmann, A., 2012. The language effect in international trade: A meta-analysis. Economics Letters 116, 221-224.

Evans, C.L., 2003. The economic significance of national border effects. The American Economic Review 93, 1291-1312.

Fettes, M., 2003. The geostrategies of interlingualism. In: Maurais, J., Morris, M.A. (Eds.), Languages in a Globalising World. Cambridge University Press, Cambridge. pp. 37-46.

Ghemawat, P., Altman, S.A., 2013. Depth index of globalization: And the big shift to emerging economies. IESE Business School. https://www.ghemawat.com/wordpress/wp-content/uploads/2016/07/Depth-Index-of-Globalization_2013-1.pdf (Accessed 29 September 2014)

Ginsburgh, V., Weber, S., 2005. Language disenfranchisement in the European Union. JCMS: Journal of Common Market Studies 43, 273-286.

Globerman, S., Shapiro, D., 2002. Global foreign direct investment flows: The role of governance infrastructure. World Development 30, 1899-1919.

Goldberg, M.A., Heinkel, R.L., Levi, M.D., 2005. Foreign direct investment: The human dimension. Journal of International Money and Finance 24, 913-934.

Grin, F., 2002. Using Language Economics and Education Economics in Language Education Policy. Language Policy Division, Out-of-School and Higher Education - DGIV, Council of Europe, Strasbourg.

Grinblatt, M., Keloharju, M., 2001. How distance, language, and culture influence stockholdings and trades. The Journal of Finance 56, 1053-1073.

Hejazi, W., Ma, J., 2011. Gravity, the English language and international business. The Multinational Business Review 19, 152-167.

Jenkins, R., 2004. Globalization, production, employment and poverty: Debates and evidence. Journal of International Development 16, 1.

Kaiser, S., 2003. Language and script in Japan and other East Asian countries: Between insularity and technology. In: Maurais, J., Morris, M.A. (Eds.), Languages in a Globalising World. Cambridge University Press, Cambridge. pp. 188-202.

Kinda, T., 2010. Investment climate and FDI in developing countries: Firm-level evidence. World Development 38, 498-513.

Kohl, T., Brouwer, A.E., 2014. The development of trade blocs in an era of globalisation. Environment and Planning A 46, 1535-1553.

Lane, P.R., Milesi-Ferretti, G.M., 2008. The drivers of financial globalization. The American Economic Review 98, 327-332.

Lewis, M.P., Simons, G.F., Fennig, C.D., 2014. Ethnologue: Languages of the World, 17th ed. SIL International, Dallas, Texas.

Makhija, M.V., Kim, K., Williamson, S.D., 1997. Measuring globalization of industries using a national industry approach: Empirical evidence across five countries and over time. Journal of International Business Studies 28, 679-710.

Maurais, J., 2003. Towards a new global linguistic order, In: Maurais, J., Morris, M.A. (Eds.), Languages in a Globalising World. Cambridge University Press, Cambridge. pp. 13-36.

Maurais, J., Morris, M.A., 2003. Introduction, In: Maurais, J., Morris, M.A. (Eds.), Languages in a Globalising World. Cambridge University Press, Cambridge. pp. 1-10.

McConnell, G.D., 2003. Towards a scientific geostrategy for English, In: Maurais, J., Morris, M.A. (Eds.), Languages in a Globalising World. Cambridge University Press, pp. 298-312.

Medvedev, D., 2012. Beyond trade: The impact of preferential trade agreements on FDI inflows. World Development 40, 49-61.

Michalopoulos, S., 2012. The origins of ethnolinguistic diversity. The American Economic Review 102, 1508-1539.

Mishra, A.V., 2007. International investment patterns: Evidence using a new dataset. Research in International Business and Finance 21, 342-360.

Mishra, A.V., Ratti, R.A., 2013. Home bias and cross border taxation. Journal of International Money and Finance 32, 169-193.

Moeller, M., Harvey, M., Griffith, D., Richey, G., 2013. The impact of country-of-origin on the acceptance of foreign subsidiaries in host countries: An examination of the “liability-of-foreignness”. International Business Review 22, 89-99.

Moshirian, F., 2008. Globalisation, growth and institutions. Journal of Banking & Finance 32, 472-479.

Nachum, L., Zaheer, S., Gross, S., 2008. Does it matter where countries are? Proximity to knowledge, markets and resources, and MNE location choices. Management Science 54, 1252-1265.

Neeley, T., 2013. Language and Globalization: "Englishnization" at Rakuten (A). Harvard Business School Case 412-002, August 2011. (Revised April 2013.).

Oh, C.H., Selmier, T.W., Lien, D., 2011. International trade, foreign direct investment, and transaction costs in languages. The Journal of Socio-Economics 40, 732-735.

Önsel, Ş., Ülengin, F., Ulusoy, G., Aktaş, E., Kabak, Ö., Topcu, Y.İ., 2008. A new perspective on the competitiveness of nations. Socio-Economic Planning Sciences 42, 221-246.

Rose, A.K., Spiegel, M.M., 2011. The Olympic effect. Economic Journal 121, 652-677.

Rose, R., 2008. Political communication in a European public space: Language, the Internet and understanding as soft power. JCMS: Journal of Common Market Studies 46, 451-475.

Schueth, S., 2011. Assembling international competitiveness: The republic of Georgia, USAID, and the doing business project. Economic Geography 87, 51-77.

Schwab, K., 2014. The Global Competitiveness Report 2014 - 2015. World Economic Forum. http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf (Accessed 25 November 2015)

Snowdon, B., Stonehouse, G., 2006. Competitiveness in a globalised world: Michael Porter on the microeconomic foundations of the competitiveness of nations, regions, and firms. Journal of International Business Studies 37, 163-175.

Sorenson, O., Audia, P.G., 2000. The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940–1989. American Journal of Sociology 106, 424-462.

Stulz, R.M., 2005. The limits of financial globalization. The Journal of Finance 60, 1595-1638.

Subramanian, A., Kessler, M., 2013. The Hyperglobalization of Trade and Its Future. Peterson Institute for International Economics, Washington, DC.

Tanzi, V., 1996. Globalization, tax competition and the future of tax systems. IMF working paper.

Tenzer, H., Terjesen, S., Harzing, A.-W., 2017. Language in international business: A review and agenda for future research. Management International Review 57, 815-854.

Toomet, O., 2011. Learn English, not the local language! Ethnic Russians in the Baltic States. The American Economic Review 101, 526-531.

Yi, K.-M., 2010. Can multistage production explain the home bias in trade? The American Economic Review 100, 364-393.

Zinkina, J., Korotayev, A., Andreev, A.I., 2013. Measuring globalization. Campus-Wide Information Systems 30, 321-339.

Table 1: The top five and the bottom five countries in the LC index

Rank	Country	LC index
1	Luxembourg	1.96
2	Belgium	1.58
3	Denmark	1.54
4	Sweden	1.51
5	Finland	1.38
111	China	0.04
112	Thailand	0.04
113	Iceland	0.03
114	Indonesia	0.03
115	Japan	0.02

Table 2: The top five and the bottom five countries in the LIC index

Rank	Country	LIC index
1	Denmark	2.5
2	Luxembourg	2.47
3	Latvia	2.27
4	South Africa	2.24
5	Sweden	2.19
111	Korea, Republic	0.38
112	Kuwait	0.25
113	Qatar	0.12
114	China	0.12
115	Japan	0.02

Table 3: Descriptive statistics and correlation matrix

Variable		Mean	S.D.	Min	Max	Correlation Coefficients														
						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*DIG*	22.47	10.35	1	50															
2	*LC*	0.66	0.43	0.02	1.96	0.41														
3	*LC English*	0.12	0.25	0	0.99	0.22	0.39													
4	*LC Other*	0.53	0.41	0.02	1.96	0.3	0.82	-0.21												
5	*LIC*	1.17	0.48	0.02	2.5	0.32	0.66	0.31	0.51											
6	*GC*	2.07	0.7	0.87	4.13	0.29	-0.13	-0.06	-0.1	0.18										
7	*Landlocked*	0.22	0.42	0	1	-0.08	-0.05	-0.17	0.05	0.06	0.16									
8	*RTA Size*	20.54	12.91	4.55	52.68	0.03	0.16	-0.07	0.21	0.05	-0.04	-0.27								
9	*Population*	0.05	0.18	0	1.34	-0.27	-0.22	0	-0.24	-0.25	0	-0.13	0.29							
10	*GDPPC*	16.46	20.28	0.16	112.03	0.56	0.31	0.43	0.06	0.19	0.17	-0.12	0.08	-0.1						
11	*GDPG*	4.05	4.5	-17.95	34.5	-0.2	-0.18	-0.19	-0.07	-0.17	-0.02	0.11	-0.06	0.16	-0.25					
12	*Tax*	45.55	22.95	10.6	286.7	-0.29	-0.01	-0.09	0.04	0.03	-0.17	0.09	0.06	0.16	-0.14	0.02				
13	*Tariff*	5.4	5.57	0	55.8	-0.51	-0.32	-0.2	-0.21	-0.18	-0.01	0.07	0.11	0.18	-0.4	0.23	0.16			
14	*Infrastructure*	3.99	1.28	1.47	6.77	0.66	0.34	0.39	0.12	0.17	0.17	-0.29	0.16	0	0.75	-0.29	-0.18	-0.46		
15	*HC*	82.79	26.51	13.15	147.62	0.62	0.38	0.31	0.21	0.21	0.13	-0.25	0.27	-0.11	0.59	-0.24	-0.14	-0.48	0.71	
16	*EF*	63.74	9.38	37.1	90	0.66	0.44	0.49	0.17	0.18	0.04	-0.14	0	-0.18	0.65	-0.25	-0.3	-0.51	0.73	0.59

Table 4: Estimated results for the depth index of globalization

	Dependent variable: Depth index of globalization (*DIG*)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LC	6.944***	8.033***				7.299***	
	(1.812)	(2.021)				(2.006)	
LC × Time		-0.223*					
		(0.126)					
Time		0.262**					
		(0.127)					
LC English			5.108				
			(3.292)				
LC Other			7.280***				
			(1.791)				
LC Power 10				5.733***			
				(1.987)			
LC Power Rest				13.18***			
				(3.069)			
LC(L2)					7.072***		
					(1.924)		
L2 Speakers						-0.00784	
						(0.0160)	
LIC							3.622***
							(1.355)
GC	3.592**	3.496**	3.545**	2.626*	2.414*	3.708***	2.608**
	(1.402)	(1.397)	(1.405)	(1.490)	(1.245)	(1.423)	(1.267)
Landlocked	-1.204	-1.326	-1.336	-1.883	-1.024	-1.207	-1.057
	(1.538)	(1.563)	(1.540)	(1.643)	(1.576)	(1.536)	(1.650)
RTA Size	-0.0409	-0.0427	-0.0436	-0.0384	-0.0336	-0.0408	-0.0341
	(0.0311)	(0.0313)	(0.0318)	(0.0312)	(0.0302)	(0.0311)	(0.0305)
Population	-8.837***	-9.230***	-8.563**	-8.326***	-11.04***	-8.677***	-9.715***
	(3.373)	(3.449)	(3.335)	(3.169)	(3.507)	(3.351)	(3.658)
GDPPC	0.00340	0.0111	0.00622	0.00773	0.00435	0.00353	0.0134
	(0.0254)	(0.0256)	(0.0258)	(0.0256)	(0.0233)	(0.0255)	(0.0236)
GDPG	0.0643**	0.0639**	0.0641**	0.0636**	0.0632**	0.0643**	0.0632**
	(0.0280)	(0.0275)	(0.0279)	(0.0280)	(0.0281)	(0.0280)	(0.0281)
Tax	-0.00633	-0.00496	-0.00637	-0.00589	-0.00607	-0.00633	-0.00639
	(0.00450)	(0.00438)	(0.00448)	(0.00465)	(0.00475)	(0.00451)	(0.00482)
Tariff	-0.0269	-0.0305*	-0.0274	-0.0273	-0.0295*	-0.0267	-0.0311*
	(0.0170)	(0.0163)	(0.0171)	(0.0170)	(0.0166)	(0.0170)	(0.0169)
Infrastructure	1.294***	1.349***	1.315***	1.269***	1.312***	1.296***	1.371***
	(0.378)	(0.373)	(0.378)	(0.378)	(0.388)	(0.379)	(0.388)
HC	0.0633**	0.0498*	0.0645**	0.0641**	0.0759***	0.0625**	0.0750***
	(0.0285)	(0.0299)	(0.0284)	(0.0282)	(0.0272)	(0.0286)	(0.0275)
EF	0.0756	0.0724	0.0786	0.0741	0.0892*	0.0746	0.0911*
	(0.0534)	(0.0528)	(0.0540)	(0.0530)	(0.0540)	(0.0535)	(0.0543)
Constant	-3.442	-3.094	-3.640	-2.096	-0.211	-3.569	-3.653
	(3.936)	(3.855)	(3.912)	(3.983)	(3.822)	(3.931)	(4.092)
N	660	660	660	660	660	660	660
Countries	114	114	114	114	114	114	114
R^2	0.580	0.576	0.587	0.601	0.581	0.580	0.584
χ^2	572.8***	552.8***	588.3***	600.9***	510.3***	569.6***	478.5***

Notes: Of the countries listed in Appendix A, Singapore could not be included in the regressions due to unavailability of data for the HC variable. Year-specific fixed effects are not reported here for brevity. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5: Estimated effects for the control variables

Control variable	Expected effect	Estimated effect
GC	Positive	Positive
Landlocked	Negative	Nonsignificant
RTA Size	Positive	Nonsignificant
Population	Negative	Negative
GDPPC	Positive	Nonsignificant
GDPG	Positive	Positive
Tax	Negative	Nonsignificant
Tariff	Negative	Negative in some estimations
Infrastructure	Positive	Positive
HC	Positive	Positive
EF	Positive	Nonsignificant

Table 6: Estimated results for the trade and investment pillars

	Trade Pillar							Investment Pillar			
	(3.1) *Trade*	(3.2) *Merchandise Trade*	(3.3) *Services Trade*	(3.4) *Merchandise Exports*	(3.5) *Merchandise Imports*	(3.6) *Services Exports*	(3.7) *Services Imports*	(3.8) *FDI Outflows*	(3.9) *FDI Inflows*	(3.10) *Portfolio Equity Outflows*	(3.11) *Portfolio Equity Inflows*
LC	44.29***	25.39**	19.26**	8.572	16.82***	14.43**	4.778	15.86*	15.92*	3.571	18.03
	(13.71)	(10.50)	(9.486)	(5.758)	(5.588)	(6.077)	(3.411)	(8.830)	(8.982)	(2.522)	(15.80)
N	658	658	658	658	658	658	658	638	659	575	566
Countries	114	114	114	114	114	114	114	110	114	98	96
R^2	0.291	0.264	0.166	0.239	0.252	0.203	0.121	0.393	0.120	0.283	0.125
χ^2	217.3***	232.8***	50.99***	170.6***	217.3***	49.57***	79.16***	66.27***	43.46***	59.82***	9.258

Notes: Control variables and year-specific fixed effects are not reported here for brevity. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 7: Estimated results for the information and people pillars

	Information Pillar					People Pillar					
	(4.1) *International Internet Bandwidth*	(4.2) *Outgoing Telephone*	(4.3) *Incoming Telephone*	(4.4) *Printed Publications Exports*	(4.5) *Printed Publications Imports*	(4.6) *Outbound Migrants*	(4.7) *Inbound Migrants*	(4.8) *Outgoing International Students*	(4.9) *Incoming International Students*	(4.10) *Inbound Tourists*	(4.11) *Outbound Tourists*
LC	130,604*	114.4	91.41**	30.02***	24.16**	0.146	1.199	13.13	3.484	0.289**	0.663
	(71,456)	(76.76)	(40.66)	(8.854)	(9.897)	(2.110)	(3.405)	(11.78)	(3.528)	(0.140)	(0.515)
N	660	564	564	651	651	467	467	441	528	536	423
Countries	114	112	112	112	112	110	110	104	92	107	78
R^2	0.243	0.375	0.359	0.393	0.582	0.0189	0.137	0.0488	0.287	0.372	0.223
χ^2	119.7***	54.33***	115.9***	93.81***	117.3***	56.42***	50.01***	33.98***	32.08**	88.68***	62.40***

Notes: Control variables and year-specific fixed effects are not reported for brevity. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 8: Estimated coefficients for the LIC index for the trade, investment, information and people pillars

Trade pillar	*Trade*	*Merchandise Trade*	*Services Trade*	*Merchandise Exports*	*Merchandise Imports*	*Services Exports*	*Services Imports*
	19.91**	9.017	11.16	1.267	7.716**	8.216*	2.904
	(9.811)	(6.149)	(7.228)	(3.874)	(3.451)	(4.733)	(2.493)
Investment pillar	*FDI Outflows*	*FDI Inflows*	*Portfolio Equity Outflows*	*Portfolio Equity Inflows*			
	7.693	5.977	2.172	10.43			
	(6.066)	(6.515)	(1.601)	(9.687)			
Information pillar	*International Internet Bandwidth*	*Outgoing Telephone*	*Incoming Telephone*	*Printed Publications Exports*	*Printed Publications Imports*		
	80,510	-46.81	6.802	22.83***	15.18**		
	(50,389)	(70.95)	(33.29)	(4.836)	(7.308)		
People pillar	*Outbound Migrants*	*Inbound Migrants*	*Outgoing International Students*	*Incoming International Students*	*Inbound Tourists*	*Outbound Tourists*	
	0.470	-5.422	6.939	-0.669	0.0395	0.130	
	(1.349)	(4.183)	(8.123)	(3.073)	(0.180)	(0.0993)	

Notes: The estimated results for the control variables are qualitatively similar to the previous sets of results with the LC index and are not tabulated here for brevity. Year-specific fixed effects are included but not reported for brevity. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$