XXX AEDEM Annual Meeting
Las Palmas de Gran Canaria 2016

The Influence of Language Similarity in International Trade: Evidence from Portuguese Exports in 2013

Sandra Ribeiro
sribeiro@autonoma.pt
Universidade Autónoma de Lisboa/OBSERVARE and ISCAL

Maria João Ferro
mjferro@iscal.ipl.pt
ISCAL and NOVA CLUNL

ABSTRACT

Communication plays a key role in today's globalized society, particularly when it comes to international trade, where agents need to understand each other in order to do business. Finding trading partners in distant parts of the world is now much easier but it also involves overcoming a number of barriers that contribute to increased trading costs. Working within the framework of the economics of language, in this paper we focus on communication costs, more specifically on those imposed by language barriers: although trading with a foreign partner that shares the same language decreases communication costs, trading with a foreign partner when no common language is available to both partners implies hiring some sort of intermediary, which will consequently increase those costs. Our findings suggest that Portuguese companies are effectively taking advantage of their collaborators' proficiency in English and Spanish to promote trade. On the other hand, we believe that the network of Portuguese speaking countries is underrepresented when we consider Portuguese exports.

KEY WORDS: Communication costs, Economics of language, International trade, Globalization, Gravity model, Language barriers.

RESUMEN

La comunicación juega un papel fundamental en la sociedad globalizada de hoy, sobre todo cuando hablamos de comercio internacional, donde los agentes tienen que entenderse entre sí si quieren hacer negocio. La búsqueda de socios comerciales en distintas partes del mundo es ahora mucho más fácil, pero también implica superar una serie de barreras que contribuyen a aumentar los costos de negociación. Desde el punto de vista de la economía de la lengua, en este trabajo nos centramos en los costos de comunicación, más específicamente en aquellos costos impuestos por las barreras del lenguaje: aunque el comercio con un socio extranjero que comparte el mismo idioma disminuye los costos de comunicación, la negociación con un socio extranjero cuando no hay un lenguaje común implica algún intermediario, que por lo tanto aumentará los costos. Nuestros hallazgos sugieren que las empresas portuguesas están aprovechando los conocimientos de inglés y de español de sus colaboradores para promover el comercio. Por otra parte, creemos que la red de países de habla portuguesa no está siendo aprovechada cuando consideramos las exportaciones portuguesas.

PALABRAS CLAVE: Barreras del lenguaje, Comercio internacional, Costos de comunicación, Economía de la lengua, Globalización, Modelo de gravedad.

INTRODUCTION

Communication plays a key role in today's globalized society, particularly when it comes to international trade, where agents need to understand each other in order to do business. Finding trading partners in distant parts of the world is now much easier but it also involves overcoming a number of barriers that contribute to increased trading costs, such as long distance transportation costs, information costs, customs costs, communication costs and many others. In this paper we focus on communication costs, more specifically on those imposed by language barriers: although trading with a foreign partner that shares the same language decreases communication costs, trading with a foreign partner when no common language is available to both partners implies hiring some sort of intermediary, which will consequently increase those costs. Fidrmuc & Fidrmuc (2016) stress the importance of sharing a similar language to make transactions smoother, with an effect similar to that of sharing a common culture or a common legal framework.

Our study was designed to understand the relationship between Portuguese exports and the language of the destination country. We wanted to know whether the language spoken in a given country would influence the choice of a foreign trading partner as far as Portuguese companies are concerned. Following our initial research, we raised two important questions: Do Portuguese companies tend to export more to countries where Portuguese is spoken? and In the absence of Portuguese as a common language, do Portuguese companies export more to countries where a similar language is spoken?

The remaining of our paper is structured as follows: first we provide an overview of the theoretical framework provided by the economics of language, which deals with the interplay between language and economy. We then suggest a methodology to analyse Portuguese exports taking language into consideration. In the next section we present and discuss the results of our study. Finally, we comment on the conclusions we derived from our findings and provide some future areas of research.

THE ROLE OF LANGUAGE IN THE ECONOMY

The economics of language is a field that studies the interconnection between language and the economy. This relationship was first established in the 1960s by Jacob Marschak (1965) who coined the term 'economics of language'. Studies on the economics of language mostly concentrate on immigration, foreign direct investment (FDI), and international bilateral trade. Before we focus on the influence of language in international trade, let us first consider the other two paths of research.

The influence of language in the choice of a host country

Several authors have argued that mastering the language of the country of destination is part of the human capital of each immigrant worker (Breton, 1978; Chiswick & Miller, 1995, 2014). Therefore, it is in the immigrants' best interest to acquire the linguistic capital

of the host country so that they are able to decrease the gap between their personal income and that of a native worker with similar characteristics. Language is necessary in all aspects of the professional activity of a worker — from finding a job to securing it, being able to communicate with their co-workers, their boss and even customers in some cases — as well as for the general well-being of the immigrant in the host society. All these activities can be carried out without mastering the host country's language, but that will imply some sort of dependency on someone who does master that language, which will usually have a negative impact on the immigrant's income.

Based on a large-scale study that gathered data for thirty OECD countries between 1980 and 2010, Adserà & Pytliková (2015) show that sharing the same or a similar language can influence the choice of a destination country. In fact, people tend to choose a host country whose language they either already speak or is easy for them to learn. Portuguese migration flows provide a good example of this strategy: until the 1970s, Portuguese workers mostly chose France and Luxembourg to work, in part due to the similarity between Portuguese and French, while currently they tend to choose Brazil, the United Kingdom, and Switzerland (Marques, 2010; Peixoto, 2012). These different choices in different moments in time provide a good illustration of our findings: Portugal shares the same language with Brazil and consequently moving to that country implies virtually no costs in terms of language acquisition. On the other hand, English is currently (and has been for some time since it replaced French as the first foreign language taught at school) the most widely studied foreign language in Portugal — according to Eurostat (2015), 34.9% of the Portuguese children learn English at primary level and 93.4% learn it at lower secondary level. The introduction of English in the school curriculum at an early stage clearly facilitates the move to the UK given that these workers will most probably already speak English if not very well, at least well enough to make themselves understood while they develop their knowledge of the language.

Language and Foreign Direct Investment

When it comes to choosing a country to invest in, language also seems to play a role. Focusing on four widespread languages — English, French, Spanish, and Arabic — Oh, Selmier & Lien (2011) show that a common language is determinant for FDI, and even more so than in international trade. Kim et al. (2015) also find a significant correlation between language and FDI, but they argue that it is not simply a matter of easiness of communication, but also the fact that language is a vehicle for culture. This argument had already been suggested by Guiso, Sapienza, & Zingales (2009), who explored cultural biases, using the commonality between two languages as a proxy for common culture. Lameli et al. (2014) tested dialect interference in the choice of trading partners in Germany, a country with only one *de jure* and *de facto* language (German), whose ancient dialectal differences still matter today in terms on domestic trade flows. The authors interpret this correlation as a measure of cultural ties, implying that language is an expression of culture.

Basing their study on Switzerland, a multilingual country with very specific characteristics, given that the speakers of the four official languages — German, French, Italian, and Romansh (by order of speakers) — are located within well-established boundaries and are usually proficient in the other languages besides their own native one, Egger and Lassmann (2015) correlate common native language with an expression of common culture and find that it has an effect on extensive margins of trade rather than on intensive ones. In a recent study involving Portuguese exporters and Angolan distributors, Alves, Raposo, & Antunes

(2012) have found that sharing the language and the culture it entails is a determinant for bilateral trade between these two countries, since the respondents to their survey explicitly mention 'same language' and 'same culture' as important factors for the success of the relationship.

Language in international trade

Half a century after Marschak first introduced the concept of the 'economics of language', language is now widely found in bilateral trade studies. Given that trade depends on the interaction between individuals it follows that the ease of communication between them will have a significant impact. When faced with the need to communicate with a trading partner, five strategies can be employed:

- sharing the same language and obviously using it to communicate;
- speaking in your own language and being understood by your trading partner who, in turn, will speak in his/her own language (i.e., intercommunication);
- choosing one of your languages as long as the other partner knows how to speak it well enough for the interaction to occur;
- choosing a foreign language, possiblt a *lingua franca* in their sector of activity or geographical location (e.g., Spanish in Latin America or English in the financial markets);
- hiring an intermediary (i.e., translator or interpreter).

With the widespread dissemination of the new information technologies, mostly the Internet, communication costs have significantly decreased and potential trading partners who would have been too distant in the past to be aware of each other's existence now have the possibility of conducting business. However, as we have mentioned before, there is still the language barrier to overcome. Similarly to the effect of sharing the same currency explored by Rose (2000), a common shared language can decrease the fixed costs of trade and thus influence the choice of country to which a company will export, although Helpman, Mélitz, & Rubinstein (2008) have found that it does not influence the volume of exports once that decision has been made.

Gravity models employed to explain bilateral trade flows frequently include some sort of language measure, which usually corresponds to the country's official language. Some authors have found a correlation between common language and trade volumes (see, for example, Helliwell, 1998; Mélitz, 2008; Egger & Lassmann, 2012), while others have shown that defining common language based solely on the country's official language is not enough and therefore some sort of measure of the linguistic similarity between languages must be found (notably Mélitz & Toubal, 2014).

Language similarity indexes

Language similarity is based on the assumption that any two languages may share a number of common traits that make it easier for the native speakers of one language to learn the other one. The literature shows that four methods to measure language similarity have gained significant support: the results of second language acquisition tests taken by American students (Chiswick & Miller, 2005), the Language Barrier Index (Lohman, 2011), Levenshtein's distance (Isphording & Otter, 2013), and open-circuit and direct-communication languages (Mélitz, 2008). For a comprehensive overview of these methods and a critique regarding their linguistic accuracy see Ferro & Costa (2016, forthcoming).

In order to find a measure of language similarity, Chiswick & Miller (2005: 1) define linguistic distance as 'the extent to which languages differ from each other' and base their research on the results of tests made by English-speaking American learners of forty-three different foreign languages that wrote two tests: the first one sixteen weeks after the course had started, and the second six weeks later. Based on the results of these assessment tests, the authors created a table of linguistic distances between English (the native tongue of the students assessed) and the forty-three foreign languages they were learning. According to the ranking presented by the authors, Afrikaans, Norwegian, Romanian and Swedish are the closest languages to English, while Korean and Japanese are the most distant ones. Besides purely linguistic issues that we will not go into here, Chiswick and Miller's classification is of limited applicability since it cannot be extrapolated to language pairs that do not include English (and even so only in relation to the forty-three languages surveyed).

Johannes Lohman created what he called the Language Barrier Index (LBI) that 'quantifies international language barriers by measuring the dissimilarity between the main languages of trading partners' (Lohman, 2011: 159). The LBI was built based on the perceived similarity between two languages using linguistic data obtained from the *World Atlas of Language Structure* (WALS) (Dryer & Haspelmath, 2013), which presently provides detailed data on 2,678 languages, including the description of up to 144 linguistic features for each language. Although based on linguistic criteria, and therefore more promising than Chiswick & Miller's proposal, the WALS is not an entirely reliable source when we look at the classification of Portuguese, due to the several inaccuracies we have found (Ferro & Costa, 2016 forthcoming).

Ishphording & Otten (2012) also resort to linguistic data. In their case, they use data developed for the reconstruction of language families by applying lexicostatistics, which provides the quantitative comparison of lexical cognates, i.e., a word that has the same linguistic derivation of another word. This methodology was developed by the Max Planck Institute for Evolutionary Anthropology and uses a specific software entitled Automatic Similarity Judgement Program (ASJP). The main purpose of the ASJP is the automatic reconstruction of relationships between languages (Bakker et al., 2009). The Levenshtein's distance measures the minimum number of additions, deletions, and substitutions to transform a word into another one (Wichman, Müller, & Velupillai, 2010), which, among other things, does not take into consideration diachronic change and does not account for loans, onomatopoeias, or any random changes.

Mélitz (2008) does not suggest a measure to calculate linguistic distance as such, but classifies languages according to their channels of influence, distinguishing direct-communication languages from open-circuit ones. Mélitz introduces the important notion of a 'widely spoken' language in a given country, i.e., a language used by at least 20% of the population, and does not base his calculations solely on the official language(s) of a given country.

We have created a method for classifying linguistic similarity based on linguistic criteria, specifically etymological ones, which allowed us to organize languages according to the linguistic family they belong to. Since our set of data refers to Portugal exports only, we found that the simple classification of languages into Romance, Germanic, and Other would be enough for our purposes, since languages from any other branches would be too removed from Portuguese to have any impact on language similarity (at this stage of our research, we posited that acquiring Chinese or Arabic would entail the same difficulty for a

Portuguese native speaker). Moreover, building on Mélitz's work, we also included in our study a common spoken language in Portugal, considering that cultural and other incentives for learning a language may be more important than its similarity to Portuguese.

METHODOLOGY

As mentioned before, we employed the gravity model to analyse the data we had collected. This model is the most widely used econometric tool for the study of international trade flows. This model has been used since the 1960s, initially by Tinbergen (1962), who found that bilateral trade flows between any two countries can be explained by a law called the 'gravity equation', which resembles the Newtonian theory of gravitation. This model has been subsequently improved over the years and expanded to include several variables that intend to explain bilateral trade flows.

Most studies in this field work with the analysis of trade volumes between pairs of countries in an NxN format and few apply the model to a specific country, since this type of study is fairly recent. Wall (1999) first studied trade flows between the United States and 85 countries from 1994 to 1996 in order to estimate the costs of protectionism. And Sohn (2005) later used the gravity model to explain South Korean trade.

The theoretical foundations for the gravity model were explored in the works of Anderson (1979), Helpman & Krugman (1985), and Kalirajan (1999). Over time, other variables were added to the basic variables considered by the model (GDP and distance), such as population, GDP per capita (Bergstrand, 1990), country size and set of countries (Azevedo, 2004). Binary variables that intend to represent the specific characteristics of the countries have also been added, such as cultural proximity, language (Endoh, 1999; Breuss, & Egger, 1999; Nitsch, 2000; Feenstra, 2002), cultural similarity, belonging to the same trade bloc — in the 1990s this effect was also considered as existence of preferential trade agreements (Breuss & Egger, 1999) — common borders, colonial relationship (Glick & Rose, 2002), among others.

The initial model was represented by:

$$T_{ij} = f\left[\frac{(GDP_i\ GDP_j)}{D_{ij}}\right] \qquad (1)$$

$$T_{ij} = \beta 0\ (GDP_i\ .\ GDP_j)^{\beta}{}_1 \quad .\ D_{ij}{}^{\beta 2} \quad .\ e^{\varepsilon} \qquad (2)$$

This model initially considered that exports between two countries are positively associated with the size of their economies and negatively related with factors that indicate the existence of barriers to trade, most prominently the distance between them. Thus, the basic gravity model relates the volume of exports between two countries T_{ij} with the economic weight of those two countries, measured using the GDP of exporter and importer (GDP_iGDP_j) and the cost of trade between them, represented by the distance between them D_{ij}— models (1) and (2) — where i and j indicate countries. With the evolution of research in this field, the need to consider more variables in order to study the impact of new situations in the volume of exports of the countries became apparent. Thus, the base model was increased when new variables were considered.

Most international trade models based on the generalized gravity model study existing exports between two countries as a function not only of the variables considered on the basic model but also of their income (measured by GDP), their population, the distance between them, as well as a set of dummy variables in order to measure 'qualitative factors'. Most of those variables, such as the analysis of some qualitative factors, are considered dummy variables (binary variables that assume value 1 when the factor is true and 0 otherwise).

Considering that the equation always implies a log-log transformation, we present the possible representation for the augmented gravity model:

$$Ln(T_{ij}) = \beta_0 + \beta_1 Ln\,(GDP_i\ GDPj) + \beta_2 LnD_{ij} + \beta_3 Lang_{ij} + \beta_4 Cont_{ij} + \beta_5 RTA_{ij} + \beta_6 ComCol_{ij} + \varepsilon_{ij} \quad (3)$$

Where i and j indicate countries and the variables are defined as follows:

T – trade volume (either imports or exports and imports) between two countries:
GDP – real GDP
D – Distance
Lang – dummy variable that is 1 when i and j share a common language and 0 otherwise
Cont – dummy variable that is 1 when i and j share a common land border and 0 otherwise
RTA – dummy variable that is 1 when i and j belong to a free trade agreement area and 0 otherwise
Comcol – dummy variable that is 1 when i and j had a colonial relationship and 0 otherwise

In line with the authors mentioned above that apply the gravity model to a single country, we based our study on the data available for international trade flows from 2014, i.e., real data regarding 2013. We used the multiple linear regression of the gravity model since it has been extensively used in the past forty years and has shown to have empirical robustness and explanatory power (Kepaptsoglou, Karlaftis, & Tsamboulas, 2010). The Ordinary Least Squares (OLS) method is the most usual technique for estimating the coefficients of the gravity model specification in its log-log form. This model allows for an easy interpretation of the estimated parameters because being in its logarithmic form its parameters represent elasticities. For example, the estimated parameter for GDP in a gravity equation estimated in logarithms is the elasticity of trade to GDP, indicating the percentage variation in trade volume following a variation of 1% in GDP.

The present study uses the OLS method considering the variable explained as an economic variable translating the logarithm of export volume between Portugal and a trading partner. The variables used in the regression are:

- economic (such as export volume between Portugal and a trading partner),
- linguistic (official language of the country, language family and/or language proximity), and
- geographical (distance between Portugal and a trading partner).

Based on these data we have studied the relationship between the volume of exports from Portugal to its 56 main trading partners worldwide in 2013 and the language family/language used in those countries in order to answer the questions stated before. We

chose the 56 main trading partners because we wanted to analyse solely the relations that Portugal establishes with its main trading partners. Since we needed a sufficient number of observations to obtain a fair degree of quality of the model and since we wanted to analyse only the main partners we chose 56 out of the 211 that represent the total number of countries to which Portugal exports.

We would like to highlight that the main aim of the present study was to analyse the relationship between the language spoken in the destination country and the volume of Portuguese exports to that country. If we were to use only one (or even two) explanatory variable, given that it is a dummy variable econometrically we would not obtain models with sufficient quality. Therefore, and so as to avoid that drawback, we also considered distance in order to add more quality to the analysed relationship, which is represented by the kilometres that separate Lisbon and the capital of the country. Since we consider the variable logarithm, this will represent the elasticity of trade regarding an absolute geographical distance. We expect the coefficient of this variable to be negative since it constitutes a barrier to trade given that the larger the distance between the countries the larger the barrier to the commercial relationship. Building on Lohmann (2011), we expect to find empirical evidence of the inverse effect of language commonality or similarity.

RESULTS AND DISCUSSION

At the beginning, we asked whether Portuguese exports were higher to countries that share the same language. In order to provide an answer to that question, we studied the following model:

$$Ln(T_{ij}) = \beta_0 + \beta_1 P_j + \beta_2 LnD_{ij} + \varepsilon_{ij} \qquad (4)$$

Where T_{ij} represents exports between Portugal and country j, P_j is a dummy variable that is 1 when country j has Portuguese as an official language and D_{ij} is the distance between Portugal and country j.

Table 1 - Results of the estimation of model (4)

Explanatory variables	Ln Exports	
	OLS Coefficient	Standardized coefficient (Beta)
Constant	18.667 (1.601)	------
P_j	0.403 (0.519)	0.095
LnD_{ij}	-0.764 (0.197)	-0.476
F = 7.599		

$R^2 = 0.22$

Notes:
Numbers in parentheses are standard deviations
Significance level 5%

Table 1 shows that although the model is explanatory overall, only the variable D_{ij} has explanatory capacity. This result is consistent with the fundamental hypotheses of the gravity model highlighting a decrease in trade with a given country due to an increase in distance. In this case, for each 1% increase in the number of kilometres, exports will decrease by 0.764%. Thus, we concluded that the variable P_j has no explanatory capability, i.e., there is no direct relationship between the volume of Portuguese exports and the fact that the destination country has Portuguese as an official language. This conclusion is hardly surprising in part, we believe, because companies might not be taking full advantage of the network of Portuguese-speaking countries to expand their business.

Extending the analysis and organizing countries according to the language family of the official language of the country, we studied model (5) below. Combining a threefold approach to the influence of language in trade, we grouped the 56 countries according to their language families. The criteria underlying our classification were:

- linguistic criteria: languages were classified according to their language family, based on etymological criteria;
- language similarity: given that Portuguese is a Romance language we included the languages that belong to this family in our analysis since they have a high degree of similarity between them;
- foreign language: the most common foreign language studied in Portugal is currently English, a Germanic language, followed by two Romance languages (French and Spanish) and then another Germanic language, German; this led us to include Germanic languages in our analysis as well.

Since at this stage we were interested in isolating these two language families (Romance and Germanic languages), we classified all the remaining languages as belonging to a group called 'Other'. We considered the dummy variables R_j, G_j, and O_j to identify respectively Romance, Germanic and Other language families. Our aim was to analyse whether belonging to each of these language families has a direct impact on Portuguese exports for that country. Thus, the variable R_j is 1 when country j has a Romance official language (and 0 otherwise) and G_j when country j as a Germanic official language. In case one of these variables is 1, the variable O_j is 0, conversely the latter would be 1 when the country has an official language that does not belong to any of these two families. However, econometrically this cannot be used together with the other two variables because it would entail multicollinearity given the linear relation that exists between the three independent variables.

$$Ln(Tij) = \beta 0 + \beta 1\ Rj + \beta 2\ Gj + \beta 3\ LnDij + \varepsilon ij \qquad (5)$$

Table 2 - Results of the estimation of model (5)

Explanatory variables	Ln Exports	
	OLS Coefficient	Standardized coefficient (Beta)

Constant	17.999 (1.611)	------
R_j	0.682* (0.391)	0.243
G_j	0.447 (0.382)	0.163
LnD_{ij}	-0.723 (0.192)	-0.450
F = 6.026		
R^2 = 0.258		

Notes:
Numbers in parentheses are standard deviations
Significance level 5%
* Significance level 10%

Table 2 shows that the model is explanatory overall and only variable G_j does not have explanatory capacity. Thus we concluded that Portuguese exports are higher to countries that share a similar language, given that R_j has explanatory capacity, i.e., there is a direct relationship between the Portuguese volume of exports and the fact that the destination country has a Romance official language. Since this is also the language family to which Portuguese belongs, this result was expected given that when the countries share the same language the linguistic barrier is erased and when they share a similar one communication costs tend to be lower.

In order to deepen our analysis, we introduced a new variable named ProxLing that intends to capture the language proximity between two countries. We defined this variable taking into consideration the official language of the destination country, namely Portuguese, Spanish, or English. Our aim was to capture a threefold effect: with this variable we identified the countries that share a common language with Portugal but also included those that have Spanish as their official language to reflect language similarity, and those that have English as their official language to capture the effect of the most widely studied and spoken foreign language in Portugal.

Based on a report made by the European Commission (2012), French is the second most widely spoken foreign language in Portugal and therefore we initially included French in our model as well. However, the analysis did not provide any statistical relevance to that fact, which we understood as being a consequence of the shift that happened some decades ago from French into English as the first foreign language studied by Portuguese children. Nowadays, although older generations still speak French, younger generations, those currently in charge of establishing commercial relationships with foreign partners, will most probably be fluent in English and not French.

Thus, the dummy variable $ProxLing_j$ is 1 when country j has Portuguese, Spanish, or English as its official language and 0 otherwise.

$$Ln(Tij) = \beta 0 + \beta 1\ ProxLingj + \beta 2\ LnDij + \varepsilon ij \qquad (6)$$

After reaching a conclusion regarding the impact of language on Portuguese exports and solely with the aim of improving the quality of our model, we added the fact that a country might belong to the European Union (EU). Both effects together form the variable UEProxLing, which intends to combine the effect of a country simultaneously having one of the three aforementioned languages as its official language and belonging to the EU. We would like to highlight the fact that although Norway and Switzerland do not belong to the EU, they were considered as such given the free trade agreements that exist between Portugal and these countries.

$$Ln(Tij) = \beta 0 + \beta 1\ UEProxLingj + \beta 2\ LnDij + \varepsilon ij \qquad (7)$$

Table 3 presents data on the analysis of these models.

Table 3 - Results of the estimation of models (6) and (7)

Explanatory variables	Ln Exports (6)		Ln Exports (7)	
	OLS Coefficient	Standardized coefficient (Beta)	OLS Coefficient	Standardized coefficient (Beta)
Constant	18.107 (1.546)	------	15.578 (1.835)	-------
$ProxLing_j$	0.726 (0.318)	0.243	--------	------
$UEProxLing_j$	------	-------	0.757 (0.271)	
LnD_{ij}	-0.722 (0.187)	-0.450	-0.450 (0.210)	
F = 10.475			F = 12.144	
$R^2 = 0.283$			$R^2 = 0.314$	

Notes:
Numbers in parentheses are standard deviations
Significance level 5%

After analysing Table 3, we can conclude that both models are explanatory overall and that every variable has explanatory capacity. We would like to stress that the quality of the model increases with the introduction of the fact that the country belongs to the EU together with what we called language proximity. Thus, we strengthened the positive effect of language proximity on Portuguese exports, which increases when both countries belong to the EU.

We also concluded that the impact of variable ProxLing in the volume of exports is slightly higher than that of variable R. Although both cases — when the official language of the destination country is a Romance language or when there is 'language proximity' as argued above — have a positive impact on Portuguese exports, the impact of language proximity is slightly higher. When considered together with the possibility of the destination country

belonging to the EU, the effect of language proximity is even higher (when compared with all other cases) and the effect of distance is consequently lower. We can thus conclude that if a country belongs to the EU (given the free trade between member-states and their relative proximity), language proximity is more important when a company is choosing a potential trading partner than the distance between both respective countries.

We would like to highlight that all the models have a low R^2. This was not unexpected since there are many other variables with explanatory capacity that we did not consider in these models since our only aim was to analyse the impact of language on exports within the framework of the economics of language.

CONCLUSION

In a highly globalized economy, companies will find trading partners wherever they can, which implies dealing with language issues. Trade costs are an important determinant of a country's capacity to take advantage of regional and global production and distribution networks, and thus are highly important from a policy perspective (Arvis et al., 2013). Language barriers can impose significant costs on bilateral trade between countries that do not share some sort of common language, either an official language or a widely spoken foreign language. Kim et al. (2015) argue that governments have the capacity to manipulate the population's linguistic skills and thus favour the introduction of new foreign languages in the school curriculum or even virtually erase the native tongue from formal education in favour of a given foreign language that they intend to promote, usually for economic or political reasons.

Our findings suggest that Portuguese companies are taking advantage of their collaborators' proficiency in English and Spanish to promote trade. However, we believe that the network of Portuguese speaking countries is underrepresented when we consider Portuguese exports. We are aware that different factors are at stake, such as the distance to those countries — much higher than between most EU countries — or the size of their economies, but if we consider the impact of language barriers on bilateral trade flows, sharing a common language should work as an incentive to trade between all members of the Community of Portuguese Language Countries (CPLP).

In the future, we shall extend our analysis to all the 211 countries to which Portugal exports. We also intend to focus on the role played by shared foreign languages, specifically *linguae francae* that can act as facilitators of trade between groups of countries such as CPLP or in geographical settings such as Spanish in Latin America.

Although we have started our research based solely on the exports of goods, we intend to include total exports and imports of goods and services and thus provide a comprehensive analysis of Portuguese commercial relations in the framework of the economics of language.

ACKNOWLEDGEMENTS

This research has been partially financed by Portuguese national funding through FCT – Fundação para a Ciência e a Tecnologia as part of the project Centro de Linguística da Universidade Nova de Lisboa - UID/LIN/03213/2013.

REFERENCES

Adserà, A., & Pytlikova, M. (2015). The role of language in shaping international migration. *The Economic Journal, 125(586)*, F49-F81.

Alves, G., Raposo, M., & Antunes, A. (2012). The influence of relationship quality and culture in the relations between exporter and distributor: The case of Portuguese exports to Angola. DOI: http://hdl.handle.net/10400.22/1186

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review, 69*, 106-116.

Arvis, J. F., Duval, Y., Shepherd, B., & Utoktham, C. (2013). Trade costs in the developing world: 1995-2010. World Bank Policy Research Working Paper, No. 6309. Available at SSRN: http://ssrn.com/abstract=2198265 [25/03/2016].

Azevedo, A. (2004). O efeito do Mercosul sobre o comércio: Uma análise com o Modelo Gravitacional. *Pesquisa e Planejamento Econômico, 34*, 307-339.

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., ... & Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology, 13(1)*, 169-181.

Bergstrand, J. (1990). The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-industry Trade. *Economic Journal, 100(403),* 1216-29.

Breton, A. (1978). Nationalism and language policies. *The Canadian Journal of Economics, 11(4)*, 656-668.

Breuss, F., & Egger, P. (1999). How Reliable Are Estimations of EastWest Trade Potentials Based on Cross-Section Gravity Analyses? *Empirica, 26(2)*, 81-95.

Chiswick, B. R., & Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics, 13(2)*, 246-288.

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)*, 1-11.

Chiswick, B. R., & Miller, P. W. (2014). International migration and the economics of language. *IZA Discussion Paper 7880*. Bonn: Institute for the Study of Labor (IZA).

Dryer, M. S., & Haspelmath, M. (Eds.) (2013). The world atlas of language structure online. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info [07-10-2015].

Egger, P. H., & Lassmann, A. (2012). The language effect in international trade: A meta-analysis. *Economics Letters, 116(2)*, 221-224.

Egger, P. H., & Lassmann, A. (2015). The causal impact of common native language on international trade: Evidence from a spatial regression discontinuity design. *The Economic Journal, 125(584)*, 699-745.

Endoh, M., 1999. Trade Creation and Trade Diversion in the EEC, the LAFTA and the CMEA: 1960-1994. *Applied Economics, 31*, 207-216.

European Commission (2012). *Europeans and their languages*. Special Eurobarometer 386. Directorate-General for Communication. Available at http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf [29-03-2016].

Eurostat (2015). *News release, 164/2015*. Eurostat Press Office. Available at http://ec.europa.eu/eurostat/documents/2995521/7008563/3-24092015-AP-EN.pdf [25/03/2016].

Feenstra, R. (2002). Border Effects and the Gravity Equation: Consistent Methods for Estimation. *Scottish Journal of Political Economy, 49(5),* 491–506.

Ferro, M. J., & Costa, R. (2016, forthcoming). The economic value of language (Ch. 10). In O. Gomes & H. F. Martins (Eds.). *Advances in applied business research: The L.A.B.S. Initiative*. New York: Nova Publishers.

Fidrmuc, J., & Fidrmuc, J. (2016). Foreign languages and trade: Evidence from a natural experiment. *Empirical Economics, 50(1)*, 31-49.

Glick, R., & Rose, A. (2002). Does a currency union affect trade? The time-series evidence *European Economic Review 46*

Guiso, L., Sapienza, P., & Zingales, L. (2009). Cultural biases in economic exchange? *The Quarterly Journal of Economics, 124(3)*, 1095-1131.

Helliwell, J. F. (1998). *How much do national borders matter?* Washington: Brookings Institution.

Helpman, E., & Krugman, P., (1985). *Market Structure and Foreign Trade – Increasing Returns, Imperfect Competition, and the International Economy* MIT Press

Helpman, E., Mélitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics, 123(2)*, 441-487.

Isphording, I. E., & Otten, S. (2013). The costs of Babylon: Linguistic distance in applied economics. *Review of International Economics, 12(2),* 354-369.

Kalirajan, K. (1999). Stochastic varying coefficients gravity model: An application in trade analysis. *Journal of Applied Statistics*, *26(2)*, 185-193.

Kepaptsoglou, K., Karlaftis, M., & Tsamboulas, D. (2010). The gravity model specification for modeling international trade flows and free trade agreement effects: A 10-year review of empirical studies. *The Open Economics Journal, 3*, 1-13.

Kim, M., Liu, A. H., Tuxhorn, K. L., Brown, D. S., & Leblang, D. (2015). Lingua mercatoria: Language and foreign direct investment. *International Studies Quarterly, 59(2)*, 330-343.

Lameli, A., Nitsch, V., Südekum, J., & Wolf, N. (2015). Same Same But Different: Dialects and Trade. *German Economic Review, 16(3)*, 290-306.

Lohmann, J. (2011). Do language barriers affect trade? *Economics Letters, 110(2)*,159-162.

Marques, J. C. (2010). A emigração portuguesa em tempos de imigração. *Polígonos, 20*, 115-129.

Marschak, J. (1965). The Economics of Language. *Behavioral Science, 10(2)*, 135-140.

Mélitz, J. (2008). Language and foreign trade. *European Economic Review, 52(4)*, 667-699.

Mélitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics, 93(2)*, 351-363.

Nitsch, V. (2000). National borders and international trade: Evidence from the European Union. *Canadian Journal of Economics, 33*, 1091-1105.

Oh, C. H., Selmier, W. T., & Lien, D. (2011). International trade, foreign direct investment, and transaction costs in languages. *The Journal of Socio-Economics, 40(6)*, 732-735.

Peixoto, J. (2012). A emigração portuguesa hoje: O que sabemos e o que não sabemos. *SOCIUS Working Paper 5/2012*. Lisbon: Centro de Investigação em Sociologia Económica e das Organizações.

Rose, A. (2000). One money, one market: The effect of common currencies on trade. *Economic Policy, 15(30)*, 8-45.

Sohn, C. (2005). Does the gravity model explain South Korea's trade flows? *Japanese Economic Review, 56 (4)*, 417-430

Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. New York: The Twentieth Century Fund.

Wall, H. J. (1999). Using the gravity model to estimate the costs of protection. *Federal Reserve Bank of St. Louis Review*, *81(1)*, 33-40.

Wichmann, S., Müller, A., & and Velupillai, V. (2010). Homelands of the world's language families: A quantitative approach. *Diachronica, 27(2)*, 247-276.

APPENDIX

Table A1 Data for the Gravity Estimation

Exports (thousands euros)	2013	LnExport	Distance (km)	Ln Distance
Algeria	527404	13.1757221	2285	7.7341213
Angola	3112688	14.9509972	6240	8.7387354
Argentina	89533	11.4023625	10329	9.2427107
Australia	90067	11.4083091	16150	9.6896753
Austria	257034	12.4569636	2037	7.6192334
Belgium	1343285	14.1106286	1583	7.3670770
Brazil	738946	13.5129801	7486	8.9207898
Bulgaria	55343	10.9213054	2831	7.9483852
Canada	213133	12.2696716	6918	8.8418819
Cape Verde	201995	12.2159982	3020	8.0130121
Chile	77421	11.2570133	10570	9.2657750
China	657484	13.3961757	9157	9.1222738
Czech Republic	285491	12.5619657	2190	7.6916568
Denmark	314982	12.6602707	2280	7.7319307
Egypt	65520	11.0901107	3859	8.2581633
Equatorial Guinea	65758	11.0937366	5000	8.5171931
Finland	216751	12.2865045	3391	8.1288801
France	5496752	15.5196679	1138	7.0370276
Germany	5508688	15.5218370	1951	7.5760973
Gibraltar	342334	12.7435421	442	6.0913098
Greece	192696	12.1688691	2576	7.8539930
Guinea-Bissau	69787	11.1532030	3144	8.0532511
Hong Kong	130726	11.7808588	10904	9.2968849
Hungary	181123	12.1069316	2389	7.7786301
India	116801	11.6682269	8339	9.0286985
Ireland	154050	11.9450325	1558	7.3511582
Israel	98986	11.5027337	3982	8.2895394
Italy	1564826	14.2632851	1771	7.4792996
Japan	139006	11.8422723	10993	9.3050139
Korea	87058	11.3743298	10448	9.2541658
Kuwait	57577	10.9608784	5152	8.5471402
Luxemburg	67562	11.1208009	1617	7.3883278
Mexico	196456	12.1881937	8717	9.0730304
Morocco	732595	13.5043483	852	6.7475865
Mozambique	327778	12.7000918	7894	8.9738582
Netherlands	1892131	14.4532142	1786	7.4877337
Nigeria	61390	11.0250022	3766	8.2337687
Norway	106897	11.5796210	2613	7.8682542
Poland	440110	12.9947799	2518	7.8312202

Romania	301725	12.6172712	2787	7.9327210
Russia	263046	12.4800842	7310	8.8969985
São Tomé and Príncipe	50344	10.8266347	4614	8.4368504
Saudi Arabia	151612	11,9290799	5250	8.5659833
Senegal	53133	10.8805534	2835	7.9497972
Singapore	57092	10.9524192	11782	9.3743282
Slovakia	89151	11.3980868	2440	7.7997533
South Africa	160894	11.9885010	8419	9.0382463
Spain	11176719	16.2293435	300	5.7037824
Sweden	440625	12.9959494	2964	7.9942949
Switzerland	419110	12.9458886	1565	7.3556411
Tunisia	166195	12.0209170	1631	7.3969486
Turkey	381111	12.8508459	3709	8.2185175
United Arab Emirates	101711	11.5298907	6067	8.7106195
United Kingdom	2612563	14.7758422	2100	7.6496926
USA	1997743	14.5075286	5974	8.6951720
Venezuela	190114	12.1553791	6863	8.8338999