

Donado, Alejandro

**Conference Paper**

## Foreign Languages and their Impact on Income and Unemployment

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik  
- Session: Norms and Culture, No. F10-V3

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Donado, Alejandro (2014) : Foreign Languages and their Impact on Income and Unemployment, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Norms and Culture, No. F10-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/100288>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Foreign Languages and their Impact on Income and Unemployment

ALEJANDRO DONADO<sup>1</sup>

Using a large European data set, I investigate the impact of knowing foreign languages on income and unemployment. The focus is on natives (not immigrants) in the labor force. I find that (1) knowing a foreign language increases the probability of having a high income by at least 5.8 percentage points and reduces the probability of being unemployed by at least 3.2 percentage points; (2) English and German tend to have a larger and more robust impact on income and unemployment than French, Spanish, and Italian; (3) but the impact of all these 5 languages varies considerably across countries.  
(JEL: Z10, J24, O52)

<sup>1</sup> University of Heidelberg, Department of Economics, Bergheimer Straße 58, 69115 Heidelberg, Germany. (e-mail: alejandro.donado@awi.uni-heidelberg.de). I would like to thank Frank Stähler, Harald Baayen, Marcel Smolka, and many seminar and conference participants for helpful comments and suggestions.

## 1. Introduction

In many countries in the world, foreign languages occupy an important place in schools' curricula. In Europe, for instance, school students typically spend more than 12% of their compulsory education learning foreign languages (see Table 1). By this criterion, foreign languages is the third most important subject in European schools after literacy learning and math.

Interestingly, even after school graduation, individuals' demand for language courses continues to be relatively high, as the thousands of foreign language teaching centers worldwide testify. For instance, the two leading French teaching organizations (Alliance Française and Institut Français) claim to have together almost 1 million French learners at their 577 teaching centers worldwide.<sup>2</sup> Other important organizations like the British Council (for English), the Goethe Institut (for German), and the Instituto Cervantes (for Spanish) also have thousands of learners in their many teaching centers located around the world.

But why do individuals spend so much time learning languages? When asked about the reasons for young people for learning foreign languages at school or university, 73% of the respondents from a 2005 survey conducted in 29 European countries gave "*to improve their job opportunities*" as one of the reasons (see Table 2). The other reasons were mentioned appreciably less often. It thus appears

<sup>2</sup> Ministère des Affaires Étrangères et Européennes (2012).

that among the several potential reasons for learning foreign languages, it is better labor market opportunities what people prioritize.

But do we really know that proficiency in foreign languages constitute an advantage in the labor market? To answer this question, I construct a repeated cross-section of more than 124,000 native residents from 31 European countries covering the period from 1990 to 2012. My data come from various series of the Eurobarometer survey that included questions on foreign language knowledge and other individual characteristics. Using these data, I estimate the conditional impact of knowing foreign languages on two specific labor market outcomes: relative incomes and the likelihood of unemployment. The data allow me to control for several individual characteristics and to include country and year effects, as well as country-year interaction dummies.

My results provide supporting evidence of the advantages of languages in the labor market. In particular, my benchmark estimates indicate that speaking at least one foreign language increases the probability of having a high income (defined as the top 25% of the income distribution) by at least 5.8 percentage points and reduces the probability of being unemployed by at least 3.2 percentage points.

However, when looking at the estimates for specific languages and disaggregated by countries, three interesting results emerge: First, the impact of English and German on income and unemployment tends to be stronger and more robust than that of French, Spanish, or Italian. Thus, the labor market does not reward all for-

eign languages. Second, there is considerably heterogeneity across countries in the magnitude and significance of the impact of languages. In particular, languages that have a large impact in one country might have a very moderate or no impact in another country. Third, it appears that foreign languages are more useful for increasing incomes than for reducing the probability of unemployment.

To help identify causal effects, I employ two instruments for foreign language ability. The first one is based on the linguistic distance between the respondent's mother tongue and the foreign language in question. This instrument exploits the fact that individuals are, *ceteris paribus*, less likely to learn languages that are linguistically more different or distant to their mother tongue because they consider them more difficult. The second instrument is the geographical distance between the respondent's region of residence and the closest region in which the foreign language is a primary language. Here, the rationale is that individuals are more likely to learn a foreign language if they live closer to a region in which the language is natively spoken. My instrumental variable estimates confirm the robustness of English and German in improving labor market outcomes.

Language proficiency has received considerable attention in the literature on immigration. The standard result is that immigrants who have a good command of their host country's language have higher earnings and are more likely to be em-

ployed.<sup>3</sup> However, the focus of my paper is exclusively on non-immigrants, and there are not many contributions that concentrate on this group of individuals. Altonji (1995) was probably the first one to study the returns to foreign languages for non-immigrants. One of his main objectives was to quantify the effects of specific high-school courses on wages in the US. Despite the fact that foreign languages were not the main focus of his paper, he estimates a 1.7% increase in wages for every additional year of foreign languages at high school. Other more recent papers have investigated this topic directly and have found wage premia of foreign languages for non-immigrants in the US (Saiz and Zoido 2005), South Africa (Levinsohn 2007), Israel (Lang and Siniver 2009), India (Azam et al. 2013), Turkey (Di Paolo and Tansel 2013), and Germany (Stöhr 2013). In an attempt to extend this literature, Ginsburgh and Prieto-Rodriguez (2011) and Williams (2011) include several West European countries in their analysis. Both contributions find positive rewards for individuals that use foreign languages in the workplace.

My paper advances the existing literature on non-immigrants in at least three respects. First, my analysis considers the impact of foreign languages *on unemployment* for the first time. All previous papers ignored this important labor market outcome and focused exclusively on the impact on earnings. Second, to

<sup>3</sup> See Chiswick and Miller (2007) and the references therein. Some recent contributions include Dustmann and van Soest (2001), Dustmann and Fabbri (2003), and Bleakley and Chin (2004).

account for the potential endogeneity of foreign language proficiency, I exploit two instrumental variables that have not been considered before in the literature of non-immigrants. Third, I show that many of the results from the literature still hold for a much larger group of countries and a considerably larger frame of time.

The remainder of the paper is organized as follows. The next section describes the empirical approach and data set. Section 3 presents the benchmark estimates on the impact of foreign languages on income and unemployment. Section 4 describes the instruments and reports the IV estimates. The last section concludes.

## **2. Empirical approach and data**

In order to assess the impact of foreign languages on income and unemployment, I estimate a regression of the form

$$(1) \quad Y_i = \alpha + \beta FL_i + \gamma X_i + u_i,$$

where  $Y$  is a measure of either income or unemployment,  $FL$  is an indicator of the individual's  $i$  foreign language ability,  $X$  is a vector of control variables, and  $u$  is the error term. The coefficient of interest is  $\beta$  which measures the impact of foreign languages on income or unemployment.

To estimate the regression coefficients, I employ data from the Eurobarometer surveys. These surveys started in 1970, have been performed several times per

year on behalf of the European Commission, and were designed to monitor attitudes toward issues related to the European Union. The number of participating countries has increased with the expansion of the European Union. European citizens aged 15 and over answer questions that are in many cases comparable across countries and over time. Some of the series have included questions on foreign language ability, income, and unemployment. These are the questions on which I base my investigation.

The question that I use to construct the foreign language indicators is “*which languages can you speak well enough to take part in a conversation, apart from your mother tongue?*” I construct a total of six indicators of foreign language ability. The first one is a dummy equal to one if the person is able to speak at least one foreign language. I call this indicator *FLONE*. The other five indicators are dummies for the main foreign languages in Europe: English, French, German, Spanish, and Italian.

To construct the income variable, I use a question in which respondents were prompted to choose their household income group from a predetermined list. The exact question is:

“Please count the total wages and salaries per month of all members of this household; all pensions and all social insurance benefits; child allowances and any other income like rents, etc... [...] Please give me the letter of the income group your household falls into before tax and other deductions.



B:	Up to 1,500 DM	L:	2,751 to 3,000 DM
T:	1,501 to 1,750 DM	N:	3,001 to 3,500 DM
P:	1,751 to 2,000 DM	R:	3,501 to 4,000 DM
F:	2,001 to 2,250 DM	M:	4,001 to 4,500 DM
E:	2,251 to 2,500 DM	S:	4,501 to 5,000 DM
H:	2,501 to 2,750 DM	K:	5,001 DM or more”

These income groups are from the German questionnaire from the Eurobarometer survey 55.1 that was conducted in 2001. The question was the same in all countries, but (for some reason that might appear incomprehensible to an economist) the income groups were always different, that is, they are not comparable or equivalent either across countries or over time. In order to make them comparable, I create a measure of *relative* income. In particular, I divide the respondents in every country and for every year into four approximate income quartiles as illustrated in Table 3.

Column 2 reports the number of respondents per income group, column 3 translates this number into percentages, and column 4 gives the cumulative percentages. It is this fourth column that allows me to identify the approximate income quartiles reported in column 5. For this example, the income “quartiles” were 23.63%, 54.12%, 76.96%, and 100%. For the other countries, the quartiles were also only approximate, but they are all reasonably close to the ideal 25%, 50%, 75%, and 100%. That the quartiles can only be approximated is clearly a disad-

vantage of my approach. However, the main advantage is that the respondents' relative income is comparable across countries and over time. In particular, the variable gives the relative income position of an individual in his/her country for a particular year. With this approach, it does not really matter if the person lives in a rich or in a poor country. It only matters the person's relative income position with respect to the other individuals in his/her country in that particular year. I can, therefore, estimate how learning foreign languages affects the relative income positions of individuals.

My second outcome variable, which I call "unemployed," is much simpler to construct. This variable is a dummy equal to one if the person reported being unemployed at the time of the survey.

To estimate regression (1), I restrict the sample in two ways. First, I exclude all individuals that are not part of the labor force (house persons, retired, and students). The reason is that, in this paper, I focus exclusively on the *labor market* rewards. Moreover, restricting the sample to respondents in the labor force makes my income variable more similar to the wage variable that has been used previously in the literature.

The second restriction is that I exclude all non-natives from the sample. The reason again is a matter of focus. My main objective is to understand how individuals living in their country of origin perform in the labor market if they learn foreign languages. As mentioned in the introduction, the performance of immi-

grants learning their host country's main language has already been investigated several times, and I ignore this important topic in this paper.

One difficulty with the Eurobarometer data set is how to identify the “natives” in the survey. Ideally, I would like to use information on the respondents' citizenship or place of birth, but unfortunately, many of the Eurobarometer surveys did not include these questions. I only know that the respondents are EU nationals, residents of the country where they took the survey. To circumvent this problem, I instead use information on the respondents' mother tongue, since this information was always available. Thus, according to the definition that I use, a person is a “native” if his/her mother tongue is one of the native languages in the country where he/she resides. For instance, all respondents in Portugal that declared Portuguese to be their mother tongue are “natives.” All others are “immigrants.” If a person reports more than one mother tongue, then I consider him/her to be a “native” if at least one of those mother tongues is native in his/her country of residence. My sources to identify the native languages in each of the countries in the sample are Central Intelligence Agency (2013) and Lewis et al. (2014). In Table A1 in the appendix, I summarize all countries with their native languages according to these sources. Finally note that all variables (including the income groups in Table 3) and all summary statistics in this paper are for my final sample of *natives in the labor force*.

Table 4 provides information on the 13 Eurobarometer surveys that included questions on foreign languages, income, and unemployment. The first survey was conducted in 1990 and the last in 2012 (see column 2). As the table shows, the number of respondents and countries have been increasing over time (although not monotonically). Adding all respondents from the 13 surveys in column 3 gives a total of 238,123 interviewees. However, when I exclude immigrants and respondents not in the labor force, I am left with the respondents in column 4 which add up to 124,395 interviewees. All of the 13 surveys included questions on foreign language ability in English, French, German, Spanish, and Italian, and information on unemployment. Unfortunately, as column 7 shows, the question on income was part of the questionnaires in only 8 of the 13 surveys.

Table 5 provides definitions of all the variables that I employ and Table 6 reports the proportion of respondents in my final sample (of natives in the labor force) for different characteristics (column 1) and also broken down by language knowledge (columns 2-7). All numbers are expressed as percentages.

As the table shows, the statistics on income quartiles in the first column make clear that my approach of grouping the respondents into 4 approximate income quartiles is not perfect but is close to the ideal distribution of 25% for each quartile (see column 1, rows 3 to 6). Also note in the same column (rows 8-9) that 87.6% of the respondents were employed while the other 12.4% were unemployed at the time of the survey.

Turning to foreign languages, the first row in column (2) indicates that 61.5% of the respondents in the final sample were able to speak at least one foreign language. English was by far the most spoken foreign language, followed by German, French, Spanish, and Italian. Rows 3 to 6 reveal that individuals with higher relative incomes were more likely to speak all languages. And rows 8 and 9 suggest a negative association between foreign language knowledge and unemployment. In the sample, females were more likely to speak all languages and, as expected, foreign language knowledge appears to increase with the education level. Married individuals were in general less likely to know foreign languages, while the households' main income earners appear to be more proficient in English, German, and Italian. There were also interesting differences in language knowledge by occupation with managers and other white collars being the most assiduous language learners.

Table 7 gives another perspective of the data by showing the shares of foreign language speakers for 12 of the 31 countries in the sample. It is interesting to see that these shares vary considerably across countries. Column 1, for example, reveals that in some countries like Luxembourg or the Netherlands, more than 90% of the natives in the labor force were able to speak at least one foreign language. In other countries, like the UK or Ireland, these shares were of around 30% or even lower. Not surprisingly, English and French were the most popular foreign languages, but in some countries, like Denmark and the Netherlands, German was

even more important than French. Spanish appears to be a popular in France and Portugal, while Italian is spoken by relatively many respondents in Luxembourg.

### **3. The estimates**

#### **3.1 Benchmark estimates**

In this section, I turn to the estimation of (1). My first dependent variable is “income.” Since the four outcomes of this variable have a natural ordering (the four income quartiles), I estimate the regression coefficients using an ordered probit specification. Columns 1 to 5 in Table 8 report the estimates. The columns in the table successively add covariates to explore the sensitivity of the estimates. The first column shows the results from the basic specification that only includes the main explanatory variable of interest (*FLONE*) in addition to the year and country dummies as well as year\*country interactions that are included in all models. The estimated coefficient is positive and highly significant, suggesting that individuals that speak at least one foreign language have higher relative incomes. As columns 2 to 5 show, adding other control variables reduces the size of the *FLONE* coefficient but its significance remains very high.

The estimates of the control variables yield reasonable results. In particular, males have higher relative incomes. Income increases with age but with a diminishing slope. Income is higher for the most educated, for married individuals, and for larger households, and is lower if the respondent reported being the main

earner in the household. Finally, managers are more likely to have higher relative incomes than individuals in the other occupations.

Since the income variable is not personal income but household income, in column 5, I replace the “married” control variable for “household size.” The *FLONE* estimated coefficient remains practically unaffected by this replacement. Only the number of observations is lower since the question on household size was not included in the Eurobarometer survey 55.1.

My second dependent variable is “unemployed.” Since this variable is a binary one, I estimate the regression coefficients using a traditional probit model. The results are shown in Table 8, columns 6 to 8. In column 6, the model does not include any control variables other than year and country dummies as well as year\*country interactions. The coefficient estimate of *FLONE* for this basic model is negative and highly significant, indicating that knowing at least one foreign language reduces the probability of being unemployed. As columns 6 and 7 illustrate, adding more control variables does not change the sign of the *FLONE* coefficient and the significance remains very high.

The coefficient estimates of the control variables are also according to expectations. In particular, males are less likely to be unemployed, and the coefficients of age and age squared suggest a convex relationship between unemployment and age. Also, more educated and married individuals are less likely to be unemployed.

As it is well known, the coefficient estimates of probit models are difficult to interpret. That is why, in Table 9, I report the average marginal effects (AMEs). In columns 1 to 4, the AMEs correspond to specification 4 in Table 8, while in column 5, the AMEs correspond to specification 8 also in Table 8. The interpretation of the AMEs of the *FLONE* coefficient in columns 1 to 4 is that, by learning at least one foreign language, an individual is able to increase the probability of having a 4<sup>th</sup> (top) quartile income by 5.8 percentage points and a 3<sup>rd</sup> quartile income by 1.4 percentage points. The individual is also able to decrease the probability of having a 2<sup>nd</sup> quartile income by 1.5 percentage points and a 1<sup>st</sup> quartile income by 5.7 percentage points. These AMEs are not negligible since, remember that by construction of the income variable, the unconditional probabilities of being in each of the 4 quartiles is of approximately 25%.

Also note that by construction of the income variable, the AMEs in columns 1 and 4 and in columns 2 and 3 are almost symmetric but with opposite signs. For this reason, in what follows, I will only report the AMEs for quartiles 4 and 3 (that is, for Q4 and Q3).

Finally, turning to the *FLONE* AMEs in column 5 of Table 9, the interpretation is that knowing at least one foreign language reduces the probability of being unemployed by 3.2 percentage points.



### 3.2 Estimates for specific languages

The results so far indicate that knowing at least one foreign language improves labor market outcomes. However, are there any particular languages that are better rewarded in the labor market? Table 10 presents the AMEs for specific languages. All the original regressions respectively include the same controls as in columns 4 and 8 from Table 8. The upper panel in Table 10 shows the AMEs when the foreign language dummies (English, French, German, Spanish, and Italian) are included in the regression one at the time, while the lower panel reports the AMEs when all these five language dummies are included in the regression simultaneously. According to the results in the upper panel, the highest “income premium” is for English followed by French, German, Spanish, and Italian. However, as the lower panel shows, this ranking is not preserved when all languages are included simultaneously in the regression. In this case, Italian comes before German, and Spanish comes last with an impact on income that is now insignificant and with the “wrong” sign. Williams (2011) also finds English to be the language with the highest returns, while Saiz and Zoido (2005: 531-2) find a higher return for German than for French for US college graduates.

With respect to the impact of specific languages on unemployment, only English and German turn out to be significant in both panels in Table 10. Knowing either language seems to significantly reduce the probability of being unemployed, but the impact of English appears to be double the size of that of German.

### **3.3 Estimates by country**

Are specific foreign languages rewarded differently across countries? Table 11 reports the estimates for 12 countries. These are the countries with the shortcut “C12” in Table 4 and are the 12 countries for which I have the most observations since they appear in most Eurobarometer surveys (see Table 4). Column 1 shows that knowing at least one foreign language is significantly associated with a higher income in almost every country. Interestingly, the change in the probability of having a top income due to foreign language knowledge varies considerably across countries. In Ireland, for instance, there appears to be no significant income improvement, while in Germany knowing at least one foreign language increases the probability of a top income by 9.8 percentage points.

The impact of specific foreign languages on income also appears to be very heterogeneous across countries. For example, English increases income significantly in all countries except in Spain, but its change in the probability of a high income varies from 4.4 percentage points in Italy to 13.3 percentage points in Luxembourg. French and German are not significant in every country, but when they are significant, their impact also varies considerably. Moreover, Spanish and Italian turn out to be significant in only a few countries, and it is in Germany and the Netherlands where these two languages appear to be most valued.

Also in Table 11, it is clear that the impact of foreign languages is more robust on income than on unemployment. For instance, while knowing at least one foreign language increases income significantly in Denmark, France, Greece, and Italy, it does not seem to significantly affect the likelihood of unemployment in any of these 4 countries.

Finally, although English, French, and Germany appear to significantly reduce the likelihood of unemployment in several countries, the impact of Spanish and Italian is almost negligible in most countries.

#### **4. IV estimates**

##### **4.1 Discussion**

One important challenge when estimating regression (1) is that the foreign language indicator might be endogenous. There are three reasons why endogeneity can arise in this context: omitted variables, measurement error, and simultaneity.

First, *omitted variables* might be a problem since due to data unavailability the regression might not include all relevant control variables. A good example of such a variable in the context of foreign languages and earnings is omitted ability (Saiz and Zoido 2005: 526). The challenge is that more able individuals are both more likely to learn foreign languages and to have higher earnings.

A second way in which endogeneity might arise is *measurement error*. As mentioned above, the question that I use to construct the foreign language indicators

is: “*which languages can you speak well enough to take part in a conversation, apart from your mother tongue?*” This question is asked to people with different backgrounds and from different countries, and the interpretation of “*well enough to take part in a conversation*” might vary across these individuals. Thus, since language ability is self-reported and not based on an objective test, measurement error is likely to be a problem in this context.

Finally, *simultaneity* can also arise since being able to speak a foreign language might also be partly determined by the individual’s income.

In order to take into account these problems, I also report results based on instrumental variables estimation. As a source of exogenous variation in foreign language knowledge, I employ two instruments: linguistic distance and geographical distance.

## **4.2 The instruments**

### **4.2.1 Linguistic distance**

My first instrument was originally proposed by Ishphording (2013) and is a measure of the linguistic distance between the respondent’s mother tongue and the foreign language in question. The intuition is that individuals are, *ceteris paribus*, less likely to learn foreign languages that are very different or very distant to their mother tongue. For instance, English native speakers might find it easier to learn German than to learn Russian, but a Polish native speaker might prefer Rus-

sian. The reason is that English is linguistically closer to German than to Russian, and Polish is closer to Russian than to German.

Interestingly, mother languages (and their distances to all other languages) are exogenously given to individuals; and there is no reason to believe that the linguistic distance to a foreign language may directly affect income or unemployment. Thus, linguistic distance appears to be a potentially good instrument.

Ideally, a measure of linguistic distance should consider the many dimensions that characterize languages, such as differences in grammar, vocabulary, and pronunciation. However, accounting for all these dimensions in one single measure has been a difficult challenge for linguists. Following Isphording's (2013) approach, I therefore employ a measure from the Automated Similarity Judgment Program<sup>4</sup> (ASJP) that compares lists of words across languages based on their vocabulary and pronunciation.<sup>5</sup> Although simple, this measure has been used by experts to classify languages and to construct language family trees. Moreover, Isphording and Otten (2014) show that the ASJP measure is highly correlated with alternative measures that highlight other dimensions of linguistic distance.

Figure 1 illustrates the ASJP linguistic distances between English and other languages in the sample. In the figure, the languages that are linguistically less distant to English are Dutch and Swedish and the most distant languages are Es-

<sup>4</sup> See <http://email.eva.mpg.de/~wichmann/ASJPHomePage.htm>.

<sup>5</sup> More details on this measure can be found in Bakker et al. (2009) and Brown et al. (2008).

tonian and Finnish. Table 12 summarizes the linguistic distances to English from Figure 1 and also to the other 4 foreign languages that I study.

One challenge when using linguistic distances as an instrument in my context is that in countries in which there is only one dominant native language, most individuals have the same mother tongue and thus the same linguistic distance to other foreign languages. For example, in Germany, (according to my definition) all natives in the labor force have German as their mother tongue and for all of them the linguistic distance to French is the same. Therefore, in order to exploit this instrument, I restrict the sample to those countries that have at least a second native language that is spoken by at least 5% of the population.<sup>6</sup> These countries are summarized in Table 13. The idea is that for each of the language groups, the linguistic distance to other languages is different. In Belgium, for instance, it is easier to learn English for the Dutch-speaking population than for the French-speaking population. The reason is that English is linguistically closer to Dutch than to French. In what follows, I will call the group of countries in Table 13 “*the IV sample.*”

Also note that in the Eurobarometer surveys some respondents reported more than one mother tongue. For those individuals, I defined the linguistic distance as the shortest distance between either mother tongue and the foreign language in question.

<sup>6</sup> In Spain, I also included Basque and Asturian because these two languages are clearly native in that country.

#### **4.2.2 Geographical distance**

My second instrument is the geographical distance between the respondent's region of residence and the closest region in which the foreign language is a native language. The idea is that an individual should be more likely to know a foreign language if he/she lives closer to a region in which that language is a primary language. For instance, Spanish people who live closer to a French region should be, *ceteris paribus*, more likely to learn French than other Spanish people.

This instrument is similar in spirit to the well-known instrument from Card (1995). Card uses proximity to a college to instrument for schooling in a regression explaining earnings. He argues that geographic differences in the accessibility of college can be used as a source of exogenous variation in schooling. In my approach, it is geographic differences in opportunities for learning foreign languages that I use as a source of exogenous variation in foreign language proficiency. In France, for example, in virtually every school, students can learn English as a first foreign language. However, the availability of school classes for learning a second foreign language is typically influenced by the proximity to other neighboring countries. In particular, students in schools closer to the Spanish border are typically more likely to be offered Spanish as a foreign language. If the school is closer to the German border, it is German that is instead more likely

to be offered; and if the school is closer to the Italian border, it is Italian that is offered (see French Senate 2003).

To construct this instrument, I first defined the regions in each country using the NUTS classification system from the European Commission. NUTS is the acronym for the French name *Nomenclature d'Unités Territoriales Statistiques* (Nomenclature of Territorial Units for Statistics). This system divides the European countries into regions for statistical and policy purposes. In many cases (but not always), the NUTS regions correspond to the countries' own administrative divisions. Table 14 summarizes the NUTS regions for the countries in my IV sample.

The advantage of using the NUTS regions is that European Commission also provides the geographical coordinates of each region's centroid (see Eurostat 2014). The region's centroid is the region's geographical center and is typically the region's capital or main city. Using the same approach from the international trade literature, I first computed the pairwise great-circle distances between all regions. I then determined the closest regions in which one of the foreign languages was spoken as a native language. Table 15 and the map in Figure 2 illustrate the approach for the Spanish regions.

Consider for example the Spanish region of Navarra. This is region number 14 in Table 15 and in Figure 2. Based on my approach, the closest French-speaking region to Navarra is the French region of Aquitaine. The distance between the two



regions is 146 kilometers. This distance is measured as the great-circle distance between the centroids of the two regions, which are respectively the cities of Pamplona and Bordeaux. The idea of this instrument is that the inhabitants of Navarra should be more likely to speak French as a foreign language than the inhabitants of, for example, Extremadura (region 10), which are 673 kilometers away from the closest French-speaking region. The map also illustrates the closest Italian-speaking region to Navarra (which is Piemonte in Italy) and the closest German-speaking region (which is the canton of Bern in Switzerland<sup>7</sup>).

Figure 3 provides some evidence for geographical distance as a determinant of foreign language proficiency. The figure gives the share of individuals in the sample that were able to speak one of the five foreign languages for different geographical distances. For instance, the share of people able to speak German as a foreign language is highest in regions that are within 500 kilometers away from a German-speaking region. The share decreases for regions that are between 500 and 1000 kilometers away. And the lowest share is for regions that are more than 1000 kilometers away. As the figure shows, this pattern is the same for the other 4 languages.

<sup>7</sup> The map in Figure 2 only shows the 7 regions in Switzerland. For the calculations of the distances involving Switzerland, I actually used the 26 cantons which are even finer subdivisions of that country.

### 4.3 The estimates

Before moving to the second-stage estimates, I first report the results from the first-stage regression which can be expressed as

$$(2) \quad FL_i = \delta + \theta LD_i + \rho GD_i + \mu X_i + v_i,$$

where  $FL$  is either of the 5 foreign language dummies (English, French, German, Spanish, or Italian),  $LD$  is the linguistic distance between the respondents mother tongue and the foreign language in question,  $GD$  is the geographical distance between the respondent's region of residence and the closest region in which the foreign language is a native language,  $X$  is the vector of controls, and  $v$  is the error term. For example, if  $FL$  is the French dummy and  $i$  is a respondent from the region of Navarra that has Spanish as his mother tongue, then  $LD$  is the linguistic distance between French and Spanish and  $GD$  is the geographical distance between Navarra and the region of Aquitaine in France.

Table 16 presents the OLS estimates of a regression like (2). As expected, the coefficients of the linguistic distance variable are all negative, confirming that individuals are less likely to learn a foreign language that is linguistically more distant to their mother tongue. Concerning the magnitude, the interpretation of the coefficient estimate in column 1 is that increasing the linguistic distance between English and the individual's mother tongue by 1 unit reduces the probability that

the individual speaks English by 0.0075 percentage points. Although this and the magnitudes of the coefficients for the other languages are all small, they are highly significant for English, French, German, and Spanish. Thus, the impact of linguistic distance turns out to be small in magnitude but significant in most cases.

The coefficient estimate of the geographical distance variable is also negative in all columns in Table 16. The implied effect in column 1 is that increasing the geographical distance to an English-speaking region by 1,000 kilometers reduces the probability that the individual speaks English as a foreign language by 3.3 percentage points. Here too the magnitudes are small but are significant at the 5% level for French, German, and Italian, and at the 10% level for Spanish. However, the coefficient in the English regression is not significant. This is probably due to the importance of English as a global language, which means that individuals learn this language independently of whether they live close to an English-speaking region or not.

I now turn to the second-stage estimates that employ the instruments described above. These results are reported in rows (d) to (f) from Table 17. As a benchmark against which to compare these IV estimates, rows (a) to (c) show the results without instrumentation but that also restrict the sample to only include the group of countries summarized in Table 13 (my IV sample). In other words, the results in rows (a) to (c) are similar to those in the upper panel from Table 10 but they are only for the group of countries in my IV sample.

Interestingly, there are some differences between the results in the upper panel from Table 17 and those in the upper panel from Table 10. In particular, the impact of French on income and that of Italian on unemployment are not significant in Table 17. These differences are however not entirely surprising since in Table 17 the sample of countries is smaller and we already know (see Table 11) that the impact of foreign languages varies across countries. Thus, by restricting the sample, the group of countries change and this can lead to different results. For English and German, however, the impact turns out to be very similar in the two tables.

Now, the IV estimates at the bottom panel in Table 17 support the idea that English and German increase relative income and decrease the likelihood of unemployment. The results for Spanish are weaker and it seems that Spanish can only increase income in the 3<sup>rd</sup> quartile range. Finally, the impact of French and Italian on income and unemployment appears to be insignificant for the group of countries in my IV sample.

Also note that the IV estimates for English and German predict a larger impact than the estimates without instrumentation. Most studies also find a larger impact of foreign languages on earnings when implementing an IV strategy (see Dustmann and van Soest 2001, Bleakley and Chin 2004, Saiz and Zoido 2005 (p. 535), and Ginsburgh and Prieto-Rodriguez 2011 (p. 607)). In principle, it is not a priori clear whether the IV estimates should be larger or smaller than the estimates

without instrumentation. While omitted ability should lead to smaller IV estimates, measurement error should give rise to larger IV estimates (see Dustmann and van Soest 2001). My results for English and German suggest that measurement error is more important.

## **5. Conclusion**

Individuals spend a considerable amount of time learning foreign languages at school and after school graduation. The main reason for learning foreign languages appears to be to improve their labor market opportunities. In this paper, I employ a large European data set to explore whether proficiency in foreign languages indeed represents an advantage in the labor market. I focus on two labor market outcomes: relative incomes and the likelihood of unemployment.

My results indicate that individuals that speak foreign languages have higher incomes and are less likely to be unemployed. Interestingly, not all foreign languages affect labor market outcomes equally. English and German appear to be more important in increasing income and reducing unemployment than other major languages. Also, the impact of foreign languages across countries tends to be very heterogeneous. While a particular language might be highly rewarded in one country, the same language might have a very marginal or no impact in another country. In more general terms, I find that foreign languages have a more important impact in increasing incomes than in reducing the probability of unem-

ployment. My general conclusion is that foreign language knowledge can be a valuable asset in the labor market.

## **Appendix**

{Place Table A1 about here}

## **References**

- Altonji, Joseph G.** 1995. "The Effects of High School Curriculum on Education and Labor Market Outcomes." *Journal of Human Resources*. 30 (3): 409-438.
- Azam, Mehtabul, Aimee Chin, and Nishith Prakash.** 2013. "The Returns to English-Language Skills in India." *Economic Development and Cultural Change*. 61 (2): 335-367.
- Bakker, Dik, André Müller, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Pamela Brown, Dmitry Egorov, Robert Mailhammer, Anthony Grant, and Eric W. Holman.** 2009. "Adding Typology to Lexicostatistics: A Combined Approach to Language Classification." *Linguistic Typology*. 13: 169-181.
- Bleakley, Hoyt, and Aimee Chin.** 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants." *The Review of Economics and Statistics*. 86 (2): 481-496.
- Brown, Cecil H., Eric W. Holman, Søren Wichmann, and Viveka Valupillai.**

2008. "Automated Classification of the World's Languages: A Description of the Method and Preliminary Results." *STUF - Language Typology and Universals*. 61 (4): 285-308.
- Card, David E.** 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In *Aspects of Labor Market Behavior: Essays in Honor of John Vanderkamp*, ed. Louis Christofides, E. Kenneth Grant and Robert Swindinsky, 201-222. Toronto: University of Toronto Press.
- Central Intelligence Agency.** 2013. *The World Factbook 2013-14*. Washington, DC. <https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html#ar>.
- Chiswick, Barry R., and Paul W. Miller.** 2007. *The Economics of Language: International Analyses*. New York: Routledge.
- Di Paolo, Antonio, and Aysit Tansel.** 2013. "Returns to Foreign Language Skills in a Developing Country: The Case of Turkey." Working Paper no. 2349584, SSRN.
- Dustmann, Christian, and Francesca Fabbri.** 2003. "Language Proficiency and Labour Market Performance of Immigrants in the UK." *The Economic Journal*. 113 (489): 695-717.
- Dustmann, Christian, and Arthur van Soest.** 2001. "Language Fluency and Earnings: Estimation with Misclassified Language Indicators." *The Review of Economics and Statistics*. 83 (4): 663-674.

- European Commission.** 2006. *Europeans and their Languages*. Brussels.
- Eurostat.** 2014. *EuroGeographics for the administrative boundaries*.  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco\\_Geographical\\_information\\_maps/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/popups/references/administrative_units_statistical_units_1).
- Eurydice.** 2012. *Recommended Annual Taught Time in Full-time Compulsory education in Europe, 2011/12*. Brussels.
- French Senate. Committee on Cultural Affairs (Commission des Affaires Culturelles).** 2003. *L'enseignement des langues étrangères en France: Rapport d'information n° 63*. November 12. [http://www.senat.fr/rap/r03-063/r03-063\\_mono.html](http://www.senat.fr/rap/r03-063/r03-063_mono.html).
- Ginsburgh, Victor A., and Juan Prieto-Rodriguez.** 2011. "Returns to Foreign Languages of Native Workers in the European Union." *Industrial and Labor Relations Review*, 64 (3): 599-618.
- Isphording, Ingo E.** 2013. "Returns to Foreign Language Skills of Immigrants in Spain." *Labour*, 27 (3): 443-461.
- Isphording, Ingo E., and Sebastian Otten.** 2014. "Linguistic Barriers in the Destination Language Acquisition of Immigrants." *Journal of Economic Behavior & Organization*, 105: 30-50.
- Lang, Kevin, and Erez Siniver.** 2009. "The Return to English in a Non-English Speaking Country: Russian Immigrants and Native Israelis in Israel." *The B.E. Journal of Economic Analysis & Policy*. 9 (1): 1-28.



- Levinsohn, James.** 2007. "Globalization and the Returns to Speaking English in South Africa." In *Globalization and Poverty*, ed. Ann Harrison. Chicago: University of Chicago Press.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.).** 2014. *Ethnologue: Languages of the World, Seventeenth Edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Ministère des Affaires Étrangères et Européennes.** 2012. *La France et la Promotion du Français Dans le Monde*. <http://www.diplomatie.gouv.fr/fr/politique-etrangere-de-la-france/promotion-de-la-francophonie-et-de/>.
- Saiz, Albert, and Elena Zoido.** 2005. "Listening to What the World Says: Bilingualism and Earnings in the United States." *The Review of Economics and Statistics*. 87 (3): 523-538.
- Stöhr, Tobias.** 2013. "The Returns to Occupational Foreign Language Use: Evidence from Germany." Working Paper no. 1880, Kiel Institute for the World Economy.
- Toomet, Ott.** 2011. "Learn English, Not the Local Language! Ethnic Russians in the Baltic States." *American Economic Review*. 101 (3): 526-531.
- Williams, Donald R.** 2011. "Multiple Language Usage and Earnings in Western Europe." *International Journal of Manpower*. 32 (4): 372-393.

TABLE 1. Minimum share of time per subject in compulsory school education

Country	School Subject					
	Literacy Learning	Math	Foreign Languages	Natural Sciences	Social Studies	Other
Luxembourg	0%	17%	<b>42%</b>	6%	1%	34%
Germany	13%	11%	<b>20%</b>	11%	3%	42%
France	26%	17%	<b>12%</b>	9%	11%	25%
Poland	11%	9%	<b>10%</b>	9%	4%	56%
Spain	14%	10%	<b>9%</b>	5%	6%	57%
Sweden	22%	14%	<b>7%</b>	12%	13%	32%
Average over 29 countries	19%	14%	<b>12%</b>	10%	8%	38%

*Notes:* This table shows the shares of minimum time per subject to total teaching time that the authorities in each country recommend for compulsory school education. The category “other” includes among others “physical education”, “arts education”, “religion, ethics and moral education”, and “technology.” The average at the bottom of the table is over the following 29 European countries: Austria, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lichtenstein, Lithuania, Luxembourg, Malta, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and Turkey. The following countries were not included because the allocation of the compulsory taught time across subjects is flexible: Belgium, the Netherlands, and the United Kingdom.

*Source:* Author calculations based on Eurydice (2012: 46-90).

TABLE 2. Reasons for young people to learn languages

<b>For what reasons do you think it is important that young people learn other languages at school or university? (Multiple answers possible)</b>	
To improve their job opportunities	73%
Because the language is widely spoken around the world	38%
To feel more comfortable when going on holiday to a region where the language is spoken	30%
To be multilingual	28%
Because the language is widely spoken in Europe	28%
To understand what life is like for people in other countries	27%
To be more tolerant and accepting towards people from other cultures	23%
To be able to communicate with family or friends in a region where the language is spoken	18%
Because of the culture associated with the language	14%
To feel more European	13%

*Notes:* The sample included 28,694 respondents from 29 European countries.

*Source:* European Commission (2006: 44).

TABLE 3. Constructing the approximate income quartiles

<b>Income Group</b>	<b>Frequency</b>	<b>Percent</b>	<b>Cumulative percent</b>	<b>Quartile</b>
(1)	(2)	(3)	(4)	(5)
B	71	7.05	7.05	
T	29	2.88	9.93	
P	39	3.87	13.8	<b>Q1</b>
F	53	5.26	19.07	
E	46	4.57	23.63	
H	42	4.17	27.81	
L	82	8.14	35.95	<b>Q2</b>
N	92	9.14	45.08	
R	91	9.04	54.12	
M	118	11.72	65.84	<b>Q3</b>
S	112	11.12	76.96	
K	232	23.04	100	<b>Q4</b>
<b>Total</b>	<b>1,007</b>	<b>100</b>		

TABLE 4. Eurobarometer Surveys

Survey number	Year	Number respondents	Final sample	Number countries	Countries	Question on income?
(1)	(2)	(3)	(4)	(5)	(6)	(7)
34.0	1990	13,883	7,587	13	C12, no	Yes
34.2	1990	7,706	3,931	12	C12	No
41.0	1994	15,406	7,607	13	C12, no	Yes
44.0	1995	16,641	8,642	15	C12, C3	Yes
50.0	1998	16,155	8,595	15	C12, C3	Yes
52.0	1999	16,071	8,713	15	C12, C3	Yes
55.1	2001	16,163	8,551	15	C12, C3	Yes
CC2001.1	2001	12,077	6,302	13	C10, C2, tr	Yes
CC2002.2	2002	12,147	6,162	13	C10, C2, tr	Yes
63.4	2005	29,328	15,393	30	C12, C10, C3, C2, hr, nctr, tr	No
64.3	2005	29,193	14,935	29	C12, C10, C3, C2, hr, tr	No
73.3	2010	26,602	13,802	27	C12, C10, C3, C2	No
77.1	2012	26,751	14,175	27	C12, C10, C3, C2	No

*Notes:* The first column gives the Eurobarometer (EB) survey number. CC2001.1 and CC2002.2 are from the so-called Candidate Countries EB surveys. The other surveys in the first column are the standard EB surveys. All these surveys can be downloaded at <http://zacad.gesis.org/webview/>. In EB 41.0 and in EB 64.3, I respectively excluded Finland and Northern Cyprus from the sample since in these countries the questions on foreign languages were not asked. In EB 41.0, the original number of respondents was larger than the 15,406 reported in column 3. The respondents that I excluded were part of an oversample of individuals that were not asked questions on foreign languages. The table employs the following shortcuts: C12 = [Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, United Kingdom], C10 = [Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia], C3 = [Austria, Finland, Sweden], C2 = [Cyprus, Malta], hr = [Croatia], no = [Norway], tr = [Turkey], nctr = [Northern Cyprus].

TABLE 5. Definitions

<b>Variable</b>	<b>Definition</b>
Income	Household income quartile on a scale from 1 to 4 (top).
Unemployed	1 if unemployed at the time of the survey.
<i>FLONE</i>	1 if able to have a conversation in at least one foreign language.
English	1 if able to have a conversation in English as a foreign language.
French	1 if able to have a conversation in French as a foreign language.
German	1 if able to have a conversation in German as a foreign language.
Spanish	1 if able to have a conversation in Spanish as a foreign language.
Italian	1 if able to have a conversation in Italian as a foreign language.
Male	1 if male.
Age	Age.
Age squared	Age squared.
Education	Age at which respondent stopped full-time education.
Married	1 if (re)married.
Household size	Number of people living in household.
Main earner	1 if respondent is the person who contributes most to the household income.
Manager	1 if manager.
Other white collar	1 if other white collar.
Self-employed	1 if self-employed.
Manual worker	1 if manual worker.
Geographical distance	Great-circle distance between respondent's region of residence and the closest region in which the foreign language is the primary language. The distances are between regions' centroids.
Linguistic distance	Linguistic distance between respondent's mother tongue and foreign language in question.

TABLE 6. Share of respondents by different characteristics (in %)

	All	FLONE	English	French	German	Spanish	Italian
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All		61.5	46.7	12.0	18.6	3.2	3.1
Income quartile							
4 <sup>th</sup> (top)	25.3	70.5	57.7	17.5	24.0	3.9	3.8
3 <sup>rd</sup>	24.6	59.5	46.8	11.1	19.2	2.5	2.7
2 <sup>nd</sup>	25.8	57.9	40.9	8.9	18.4	2.3	2.1
1 <sup>st</sup>	24.4	51.3	35.6	7.9	15.7	2.5	2.0
Employment status							
Employed	87.6	62.6	48.6	12.7	19.5	3.3	3.2
Unemployed	12.4	53.4	33.4	7.4	12.7	2.6	2.1
Sex							
Male	46.8	59.8	45.8	10.9	18.5	3.0	2.8
Female	53.2	63.3	47.8	13.3	18.9	3.4	3.4
Age							
15 ≤ age < 30	24.9	64.0	55.6	13.8	17.2	3.9	3.1
30 ≤ age < 40	26.7	64.9	51.4	12.1	18.9	3.6	3.4
40 ≤ age < 50	24.5	61.0	43.2	11.9	19.2	2.7	3.0
50 ≤ age	24.0	55.5	36.2	10.2	19.2	2.5	2.8
Education to age							
15 ≤ age	15.8	30.0	16.6	5.3	6.1	2.3	1.3
16 ≤ age < 20	49.8	57.8	39.7	9.0	14.5	2.4	2.9
20 ≤ age	34.4	77.1	64.8	18.1	27.6	4.5	4.0
Marital status							
Married	55.9	59.2	42.1	11.3	18.5	2.6	2.8
Other	44.1	64.4	52.7	13.0	18.8	4.0	3.4
Main income earner							
Yes	62.4	57.8	44.5	12.1	18.8	3.0	2.8
No	37.6	59.2	44.0	12.5	16.0	2.7	2.4
Occupation							
Manager	17.2	81.5	69.4	19.6	30.6	4.4	5.0
Other white collar	19.6	68.0	55.3	15.3	22.0	3.7	3.5
Self employed	14.8	56.6	42.0	12.5	16.5	3.9	3.6
Manual worker	36.0	53.1	37.6	7.9	14.2	2.3	2.1
Unemployed	12.4	53.4	33.4	7.4	12.7	2.6	2.1

Notes: The sample only includes native individuals in the labor force.

TABLE 7. Share of natives in the labor force able to speak (as a foreign language) ...

	<i>FLONE</i>	<b>English</b>	<b>French</b>	<b>German</b>	<b>Spanish</b>	<b>Italian</b>
	(1)	(2)	(3)	(4)	(5)	(6)
Belgium	68.2	51.8	58.4	20.9	4.0	4.2
Denmark	89.1	85.2	8.5	53.2	2.6	0.9
France	48.7	38.4	0.0	7.6	10.5	4.5
Germany	50.7	44.4	8.2	0.0	2.0	1.6
Greece	51.7	45.2	6.0	5.7	0.9	3.6
Ireland	30.5	2.7	14.4	4.2	2.4	0.8
Italy	44.3	30.8	18.3	3.9	4.5	0.0
Luxembourg	97.2	66.0	83.6	94.0	4.2	13.1
Netherlands	93.1	86.8	21.0	68.3	4.1	1.7
Portugal	39.7	26.8	21.4	2.2	7.9	1.4
Spain	41.7	20.9	10.5	1.6	7.9	1.8
UK	27.0	0.0	18.3	6.3	3.8	1.5



TABLE 8. Benchmark estimates

Dependent variable: Model:	Income Ordered probit					Unemployed Probit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>FLONE</i>	0.390*** (0.027)	0.406*** (0.023)	0.307*** (0.026)	0.219*** (0.025)	0.218*** (0.027)	-0.038*** (0.008)	-0.046*** (0.007)	-0.032*** (0.007)
Male		0.072*** (0.022)	0.264*** (0.033)	0.324*** (0.028)	0.302*** (0.031)		-0.030*** (0.007)	-0.030*** (0.007)
Age		0.066*** (0.009)	0.031*** (0.007)	0.023*** (0.007)	0.035*** (0.008)		-0.013*** (0.001)	-0.008*** (0.001)
Age squared (*10 <sup>3</sup> )		-0.766*** (0.093)	-0.383*** (0.079)	-0.311*** (0.078)	-0.372*** (0.082)		0.140*** (0.013)	0.087*** (0.012)
Education			0.067*** (0.006)	0.047*** (0.005)	0.047*** (0.005)			-0.007*** (0.001)
Married			0.557*** (0.045)	0.499*** (0.044)				-0.063*** (0.005)
Household size					0.172*** (0.029)			
Main income earner			-0.507*** (0.057)	-0.703*** (0.047)	-0.634*** (0.050)			
Occupation								
Manager				1.504*** (0.053)	1.560*** (0.059)			
Other white collar				1.068*** (0.046)	1.135*** (0.053)			
Self-employed				1.089*** (0.085)	1.129*** (0.091)			
Manual worker				0.766*** (0.043)	0.810*** (0.045)			
Year, country, and year*country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R <sup>2</sup>	0.017	0.023	0.077	0.116	0.114	0.054	0.073	0.093
Obs.	47,105	47,074	39,701	39,701	33,805	123,478	123,400	113,096

Notes: Columns 1-5 report ordered probit coefficients and columns 6-8 report probit coefficients. The omitted occupation category in columns 4 and 5 is “unemployed.” All standard errors in parenthesis are robust, clustered at the country level. \*\*\* Significant at the 1% level. \*\* Significant at the 5% level. \* Significant at the 10% level.

TABLE 9. Average marginal effects (AMEs)

Dependent variable: <i>Model:</i>	Income				Unemployed
	<i>Ordered probit (AMEs)</i>				<i>Probit (AMEs)</i>
Quartile:	Q4	Q3	Q2	Q1	
	(1)	(2)	(3)	(4)	(5)
<i>FLONE</i>	0.058***	0.014***	-0.015***	-0.057***	-0.032***
Male	0.087***	0.020***	-0.023***	-0.083***	-0.030***
Age	0.006***	0.001***	-0.002***	-0.006***	-0.008***
Age squared (*10 <sup>3</sup> )	-0.084***	-0.019***	0.022***	0.080***	0.087***
Education	0.013***	0.003***	-0.003***	-0.012***	-0.007***
Married	0.134***	0.030***	-0.036***	-0.129***	-0.063***
Main income earner	-0.189***	-0.043***	0.051***	0.181***	
Occupation					
Manager	0.404***	0.091***	-0.108***	-0.387***	
Other white collar	0.287***	0.065***	-0.077***	-0.275***	
Self-employed	0.292***	0.066***	-0.078***	-0.280***	
Manual worker	0.206***	0.047***	-0.055***	-0.197***	
Year, country, and year*country effects	Yes	Yes	Yes	Yes	Yes

*Notes:* All columns report average marginal effects that respectively correspond to the regressions in columns 4 and 8 from Table 8. All standard errors in the underlying regressions are robust, clustered at the country level. \*\*\* Significant at the 1% level. \*\* Significant at the 5% level. \* Significant at the 10% level.

TABLE 10. Average marginal effects for specific languages

		English	French	German	Spanish	Italian
		(1)	(2)	(3)	(4)	(5)
One at the time	Income: Q4	0.084***	0.060***	0.038***	0.029***	0.057***
	Income: Q3	0.019***	0.010***	0.008***	0.006***	0.010***
	Unemployed	-0.045***	-0.022***	-0.023***	-0.004	-0.022**
Simultaneously	Income: Q4	0.091***	0.057***	0.020***	-0.004	0.039**
	Income: Q3	0.022***	0.011***	0.005***	-0.001	0.008***
	Unemployed	-0.041***	0.001	-0.016***	0.004	-0.017

*Notes:* This table reports average marginal effects of regressions that respectively have the same specification to those in columns 4 and 8 from Table 8. The only difference is that the main dependent variable is not *FLONE* but dummies for specific foreign languages. The upper panel reports the results of specifications that include the foreign language dummies one at the time, while in the bottom panel the specification includes all five foreign languages simultaneously. All standard errors in the underlying regressions are robust, clustered at the country level. \*\*\* Significant at the 1% level. \*\* Significant at the 5% level. \* Significant at the 10% level.

TABLE 11. Estimates by country

		<i>FLONE</i>	English	French	German	Spanish	Italian
		(1)	(2)	(3)	(4)	(5)	(6)
Belgium	Income: Q4	0.078***	0.066***	--	0.046**	0.076	-0.012
	Income: Q3	0.018***	0.013***	--	0.007**	0.009***	-0.003
	Unemployed	-0.073***	-0.073***	--	-0.054***	-0.004	0.012
Denmark	Income: Q4	0.062***	0.072***	0.079***	0.043***	0.040	0.072
	Income: Q3	0.016***	0.019***	0.012***	0.009***	0.007	0.010
	Unemployed	-0.010	-0.031**	-0.022*	-0.014*	-0.007	-0.035
France	Income: Q4	0.049***	0.062***	--	0.075***	0.004	0.030
	Income: Q3	0.012***	0.014***	--	0.013***	0.001	0.007
	Unemployed	-0.003	-0.010	--	-0.015	0.002	0.009
Germany	Income: Q4	0.098***	0.121***	0.088***	--	0.097**	0.168***
	Income: Q3	0.021***	0.023***	0.012***	--	0.011***	0.009*
	Unemployed	-0.093***	-0.105***	-0.081***	--	-0.054***	-0.070***
Greece	Income: Q4	0.091***	0.102***	0.097**	-0.001	-0.028	0.060
	Income: Q3	0.021***	0.022***	0.013***	-0.000	-0.007	0.010**
	Unemployed	0.000	-0.011	0.009	-0.000	0.021	0.041
Ireland	Income: Q4	-0.008	--	0.015	0.042	-0.056	-0.056
	Income: Q3	-0.002	--	0.004	0.009	-0.019	-0.020
	Unemployed	-0.030***	--	-0.031**	-0.058***	0.003	0.031
Italy	Income: Q4	0.045***	0.044**	0.012	0.032	-0.010	--
	Income: Q3	0.011***	0.010**	0.003	0.006	-0.003	--
	Unemployed	-0.006	-0.009	0.017	-0.017	-0.010	--
Luxembourg	Income: Q4	0.077	0.131***	--	--	0.115	0.057
	Income: Q3	0.003	0.005	--	--	-0.001	0.001
	Unemployed	-0.008	-0.007	--	--	0.005	0.003
Netherlands	Income: Q4	0.075***	0.090***	0.049**	0.022*	0.074	0.131*
	Income: Q3	0.028***	0.033***	0.011***	0.006	0.014**	0.016***
	Unemployed	-0.027*	-0.023**	-0.002	-0.019**	-0.014	0.002
Portugal	Income: Q4	0.061***	0.070***	0.048**	-0.013	0.009	0.044
	Income: Q3	0.012***	0.012***	0.009***	-0.003	0.002	0.007
	Unemployed	-0.017*	-0.023**	0.004	-0.065***	-0.009	0.019
Spain	Income: Q4	0.034*	0.027	-0.030	0.021	--	0.063
	Income: Q3	0.004*	0.003	-0.004	0.002	--	0.005
	Unemployed	-0.039***	-0.033**	-0.031*	0.001	--	-0.021
UK	Income: Q4	0.086***	--	0.103***	0.127***	0.040	-0.018
	Income: Q3	0.008***	--	0.007***	0.003	0.004	-0.003
	Unemployed	-0.050***	--	-0.069***	-0.028*	-0.009	-0.059*

Notes: This table reports average marginal effects for particular countries of the foreign language indicator coefficients. The specification of the underlying regressions is identical to those in columns 4 and 8 from Table 8 except for the country and year\*country effects that are not included here. All standard errors in the underlying regressions are robust. \*\*\* Significant at the 1% level. \*\* Significant at the 5% level. \* Significant at the 10% level.

TABLE 12. Linguistic distances

Mother tongue	Linguistic distance to ...				
	English	French	German	Italian	Spanish
Asturian	93.12	80.92	93.82	61.96	42.45
Basque	100.95	100.55	101.40	102.65	101.71
Bulgarian	89.69	91.67	91.56	88.38	93.80
Catalan	86.26	75.45	87.95	65.17	72.12
Dutch	60.73	92.96	49.35	87.40	91.82
Estonian	101.50	98.06	98.20	98.11	97.13
Finnish	104.24	98.81	100.26	103.57	97.38
French	91.55	0	97.01	78.65	84.03
Galician	90.41	79.86	92.02	50.27	54.82
German	69.04	97.01	0	86.61	93.57
Hungarian	95.19	96.93	98.16	101.66	101.97
Latvian	96.83	96.56	93.35	95.51	98.45
Lithuanian	95.55	94.41	93.15	93.09	94.50
Luxembourgish	66.32	96.58	41.94	90.39	91.14
Polish	95.02	94.20	95.93	91.13	91.04
Romanian	86.64	78.89	88.15	54.10	73.71
Russian	92.61	93.84	91.18	95.52	95.73
Slovak	90.60	91.94	92.79	91.83	91.06
Spanish	93.34	84.03	93.57	58.31	0
Swedish	64.15	94.85	69.35	92.48	92.63
Turkish	100.95	98.71	99.73	97.43	98.51

TABLE 13. “Native” language groups

<b>Country</b>	<b>Language groups (share of the total population)</b>
Belgium	Dutch (60%), French (40%)
Bulgaria	Bulgarian (76.8%), Turkish (8.2%)
Estonia	Estonian (68.5%), Russian (29.6%)
Finland	Finnish (94.2%), Swedish (5.5%)
Latvia	Latvian (56.3%), Russian (33.8%)
Lithuania	Lithuanian (82%), Russian (8%), Polish (5.6%)
Luxembourg	Luxembourgish (77%), French (6%), German (5%)
Romania	Romanian (85.4%), Hungarian (6.3%)
Slovakia	Slovak (78.6%), Hungarian (9.4%)
Spain	Spanish (74%), Catalan (17%), Galician (7%), Basque (2%), Asturian (<1%)

*Sources:* Central Intelligence Agency (2013), Lewis et al. (2014).

TABLE 14. NUTS regions

Country	Regions
Belgium	1 Hainaut, 2 Limburg, 3 Namur, 4 Oost-Vlaanderen, 5 West-Vlaanderen, 6 Liège, 7 Luxembourg, 8 Vlaams-Brabant, 9 Antwerpen, 10 Bruxelles, 11 Brabant Wallon
Bulgaria	1 Severozapaden, 2 Severen tsentralen, 3 Severoiztochen, 4 Yugoiztochen, 5 Yugozapaden, 6 Yuzhen tsentralen
Estonia	1 Tallinn, 2 North, 3 West, 4 Tartu area, 5 South, 6 North-East
Finland	1 Uusimaa, 2 Varsinais-Suomi, 3 Satakunta, 4 Kanta-Häme, 5 Pirkanmaa, 6 Päijät-Häme, 7 Kymenlaakso, 8 Etelä-Karjala, 9 Etelä-Savo, 10 Pohjois-Savo, 11 Pohjois-Karjala, 12 Keski-Suomi, 13 Etelä-Pohjanmaa, 14 Pohjanmaa, 15 Keski-Pohjanmaa, 16 Pohjois-Pohjanmaa, 17 Kainuu, 18 Lappi
Latvia	1 Rīga, 2 Pierīga, 3 Vidzeme, 4 Kurzeme, 5 Zemgale, 6 Latgale
Lithuania	1 Vilnius, 2 Utena, 3 Kaunas, 4 Alytus, 5 Marijampolė, 6 Panevėžys, 7 Šiauliai, 8 Tauragė, 9 Telšiai, 10 Klaipėda
Luxembourg	1 Luxembourg
Romania	1 North-East, 2 South-East, 3 South, 4 South-West, 5 West, 6 North-West, 7 Central, 8 Bucharest
Slovakia	1 Bratislavský kraj, 2 Západné Slovensko, 3 Stredné Slovensko, 4 Východné Slovensko
Spain	1 Andalucía, 2 Aragón, 3 Asturias, 4 Illes Balears, 5 Canarias, 6 Cantabria, 7 Castilla-León, 8 Castilla-La Mancha, 9 Cataluña, 10 Extremadura, 11 Galicia, 12 Madrid, 13 Murcia, 14 Navarra, 15 La Rioja, 16 Comunidad Valenciana, 17 País Vasco

TABLE 15. Geographical distance to Spanish regions

Spanish Region	Closest French-speaking region	Distance (km)	Closest German-speaking region	Distance (km)	Closest Italian-speaking region	Distance (km)
1 Andalucía	Aquitaine (FR)	734	Bern (CH)	1429	Sardegna (IT)	1205
2 Aragón	Aquitaine (FR)	202	Bern (CH)	885	Piemonte (IT)	787
3 Asturias	Aquitaine (FR)	469	Bern (CH)	1126	Piemonte (IT)	1110
4 Illes Balears	Languedoc-R. (FR)	434	Bern (CH)	885	Sardegna (IT)	521
5 Canarias	Aquitaine (FR)	2170	Bern (CH)	2883	Sardegna (IT)	2601
6 Cantabria	Aquitaine (FR)	319	Bern (CH)	1012	Piemonte (IT)	974
7 Castilla y León	Aquitaine (FR)	456	Bern (CH)	1176	Piemonte (IT)	1099
8 Castilla-La Mancha	Aquitaine (FR)	515	Bern (CH)	1210	Sardegna (IT)	1039
9 Cataluña	Midi-Pyrénées (FR)	209	Bern (CH)	727	Piemonte (IT)	605
10 Extremadura	Aquitaine (FR)	673	Bern (CH)	1392	Sardegna (IT)	1293
11 Galicia	Aquitaine (FR)	647	Bern (CH)	1313	Piemonte (IT)	1296
12 Madrid	Aquitaine (FR)	484	Bern (CH)	1196	Sardegna (IT)	1086
13 Murcia	Aquitaine (FR)	610	Bern (CH)	1242	Sardegna (IT)	939
14 Navarra	Aquitaine (FR)	146	Bern (CH)	866	Piemonte (IT)	803
15 La Rioja	Aquitaine (FR)	223	Bern (CH)	941	Piemonte (IT)	880
16 Com. Valenciana	Aquitaine (FR)	436	Bern (CH)	1054	Sardegna (IT)	807
17 País Vasco	Aquitaine (FR)	212	Bern (CH)	927	Piemonte (IT)	871

*Notes:* The shortcuts FR, CH, and IT respectively correspond to the countries France, Switzerland, and Italy.

*Source:* Author calculations using the geographical coordinates provided by Eurostat 2014.



TABLE 16. First-stage estimates

	English (1)	French (2)	German (3)	Spanish (4)	Italian (5)
Linguistic distance (*10 <sup>3</sup> )	-0.075*** (0.019)	-0.051*** (0.009)	-0.046*** (0.003)	-0.005*** (0.001)	-0.007 (0.010)
Geographical distance (*10 <sup>3</sup> )	-0.033 (0.039)	-0.066*** (0.014)	-0.069** (0.029)	-0.043* (0.024)	-0.035** (0.013)
Male	-0.006 (0.006)	-0.030*** (0.005)	0.002 (0.007)	-0.005* (0.003)	-0.008*** (0.002)
Age	-0.013*** (0.003)	-0.002 (0.001)	-0.006*** (0.002)	-0.001 (0.001)	-0.001* (0.000)
Age squared (*10 <sup>3</sup> )	0.062** (0.028)	0.017 (0.014)	0.058*** (0.018)	0.007 (0.006)	0.007 (0.005)
Education	0.023*** (0.003)	0.012*** (0.002)	0.009*** (0.002)	0.003*** (0.001)	0.002*** (0.001)
Married	-0.037*** (0.009)	-0.021*** (0.004)	0.004 (0.005)	-0.010*** (0.002)	-0.007*** (0.002)
Main income earner	-0.011* (0.005)	-0.001 (0.004)	0.015*** (0.004)	0.003 (0.002)	0.004 (0.003)
Occupation					
Manager	0.202*** (0.015)	0.075*** (0.013)	0.080*** (0.011)	0.010*** (0.003)	0.012** (0.004)
Other white collar	0.127*** (0.012)	0.030** (0.011)	0.035** (0.012)	0.007*** (0.002)	0.002 (0.003)
Self-employed	0.101*** (0.017)	0.036*** (0.011)	0.032*** (0.009)	0.013*** (0.004)	0.012*** (0.003)
Manual worker	0.002 (0.013)	-0.012 (0.009)	-0.019** (0.009)	-0.004 (0.004)	-0.003 (0.002)
Year, country, and year*country effects	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.348	0.105	0.287	0.038	0.148
Obs.	37,887	38,022	36,010	41,726	41,660
First-stage <i>F</i>	8.98***	31.52***	129.24***	12.44***	3.83**

Notes: The table reports the OLS coefficients estimates of regressions like (2). All standard errors in parenthesis are robust, clustered at the country level. \*\*\* Significant at the 1% level. \*\* Significant at the 5% level. \* Significant at the 10% level.

TABLE 17. Probit estimates for the IV sample (Average marginal effects)

		English	French	German	Spanish	Italian
	<i>Model</i>	(1)	(2)	(3)	(4)	(5)
(a) Income: Q4	<i>Ordered probit</i>	0.080***	0.028	0.039***	0.071***	0.056*
(b) Income: Q3	<i>Ordered probit</i>	0.014***	0.005	0.007***	0.011***	0.008**
(c) Unemployed	<i>Probit</i>	-0.058***	-0.027***	-0.041***	-0.004	-0.009
(d) Income: Q4	<i>IV ordered probit</i>	0.110***	0.017	0.084***	0.064	0.073
(e) Income: Q3	<i>IV ordered probit</i>	0.018***	0.003	0.012***	0.010**	0.010
(f) Unemployed	<i>Bivariate probit</i>	-0.249***	0.072	-0.203***	0.413	0.184

*Notes:* This table reports average marginal effects of regressions that have the same specification to those in columns 4 and 8 from Table 8. There are however two main differences: First, the main dependent variable is not *FLONE* but dummies for specific foreign languages that enter the regression one at the time. Second, the sample in all regressions only includes the group of countries summarized in Table 13. The upper panel reports the results without instrumentation, while the bottom panel reports the IV results. All standard errors in the underlying regressions are robust, clustered at the country level. \*\*\* Significant at the 1% level. \*\* Significant at the 5% level. \* Significant at the 10% level.

TABLE A1. Native languages

<b>Country</b>	<b>Native languages</b>
Austria	German
Belgium	Dutch, French
Bulgaria	Bulgarian, Turkish
Croatia	Croatian
Cyprus (Republic)	Greek
Northern Cyprus	Turkish
Czech Republic	Czech
Denmark	Danish
Estonia	Estonian, Russian
Finland	Finnish, Swedish
France	French, Alsatian, Basque, Breton, Occitan
Germany	German
Greece	Greek
Hungary	Hungarian
Ireland	English, Irish
Italy	Italian
Latvia	Latvian, Russian
Lithuania	Lithuanian, Russian, Polish
Luxembourg	Luxembourgish, French, German
Malta	Maltese, English
Netherlands	Dutch, West Frisian
Norway	Norwegian
Poland	Polish
Portugal	Portuguese
Romania	Romanian, Hungarian
Slovakia	Slovak, Hungarian
Slovenia	Slovenian
Spain	Spanish, Catalan, Galician, Basque, Asturian
Sweden	Swedish
Turkey	Turkish
UK	English, Scottish Gaelic, Welsh

*Sources:* Central Intelligence Agency (2013), Lewis et al. (2014).

FIGURE 1. Linguistic distance to English

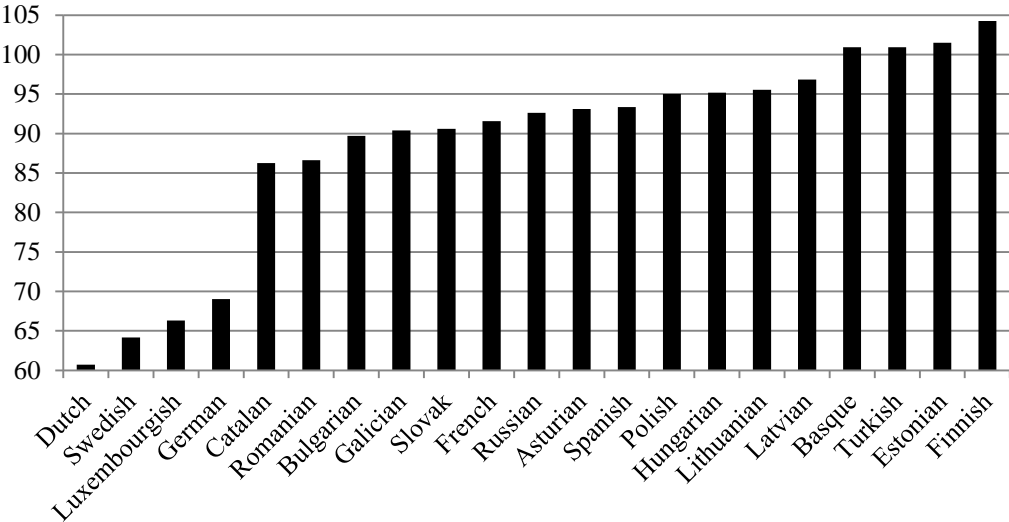
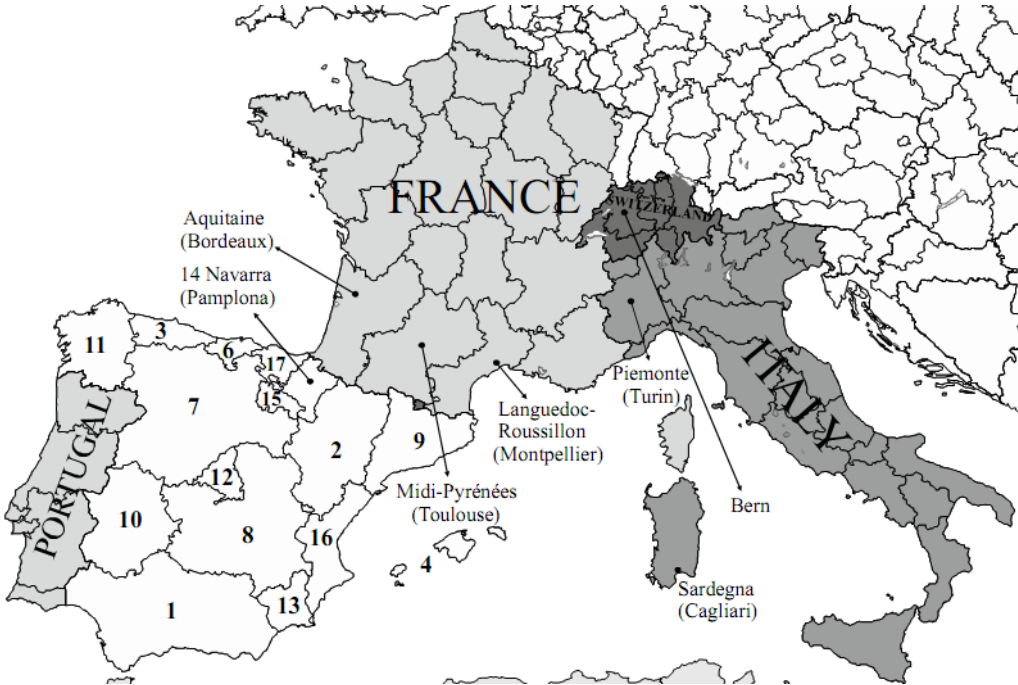
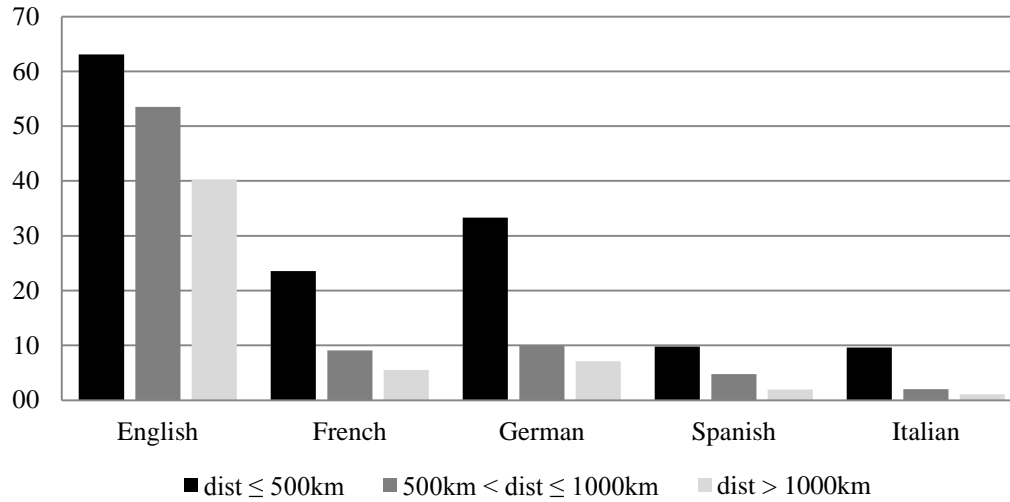


FIGURE 2. Spanish regions and geographical distance



Notes: The numbers in the map correspond to the numbers of the Spanish regions listed in Table 15, column 1.

FIGURE 3. Geographical distance matters



Notes: The figure reports the share of foreign language speakers for different geographical distances.