# Learning foreign languages: Theoretical and empirical implications of the Selten and Pool model

Victor Ginsburgh [a,1], Ignacio Ortuño-Ortín [b,2], Shlomo Weber [c,d,*]

[a] *ECARES, Université Libre de Bruxelles and CORE, Louvain-la-Neuve, Belgium*
[b] *Department of Economics and IVIE, University of Alicante, Spain*
[c] *Department of Economics, SMU, Dallas, USA*
[d] *CORE, Université Catholique de Louvain, and CEPR, Belgium*

## Abstract

In this paper we adopt the Selten–Pool [Selten, R., Pool, J., 1991. The distribution of foreign language skills as a game equilibrium. In: Selten, R. (Ed.), Game Equilibrium Models, vol. 4. Springer-Verlag, Berlin, pp. 64–84] framework of language acquisition based on "communicative benefits" derived from the ability to communicate with other speakers of an acquired language, and "learning costs" incurred by acquiring a foreign language. We show that, under some mild conditions, there exists a unique interior linguistic equilibrium. We then derive demand functions for foreign languages, that we estimate for English, French, German and Spanish in 13 European countries and demonstrate that the properties of these functions are consistent with our theoretical results.

* Corresponding author at: Department of Economics, SMU, 3300 Dyer str , Dallas, TX 75275-0496, USA Tel : +1 2147683577; fax: +1 2147683911

*E-mail addresses:* vginsbur@ulb ac be (V Ginsburgh), ortin@merlin fae ua es (I Ortuño-Ortín), sweber@mail smu edu (S Weber)

[1] ECARES C P 114, Université Libre de Bruxelles, 50 Avenue F Roosevelt, 1050 Brussels, Belgium Tel : +32 2 6503846; fax: +32 2 6504012

[2] Departamento Fundamentos Análisis Económico, Universidad de Alicante, Campus de San Vicente, 03071 Alicante, Spain Tel : +34 965 90 34 00x3228; fax: +34 965 90 38 98

# 1. Introduction

The reasons that induce inhabitants of a country to learn other languages can be analyzed by examining the benefits and the costs that learning generates. The benefits of acquiring an additional language are often linked with the increased earning potential,[3] especially in the case of immigrants who acquire the native language of the country in which they live. A different approach has been pioneered by Selten and Pool (1991) who introduce a general model of language acquisition that does not limit the analysis to "earnings as a mechanism and to firms as a milieu of the incentive to learn languages" (Selten and Pool, p. 66). In their model every individual derives a gross benefit from the knowledge of a foreign language and incurs a cost of learning it. The "gross communicative benefit" is positively correlated with the number of other individuals with whom an individual can communicate by sharing at least one common language. Naturally all languages are assumed to be "communicative substitutes," and communication between two individuals can take place in any common language they share. Selten and Pool show that an equilibrium of the multi-country multi-lingual language acquisition model exists. The characterization of an equilibrium is studied by Church and King (1993). They examine "corner" equilibria in a bilingual setting with two populations, $i$ and $j$, where all citizens of one population, say $i$, learn the language of $j$, whereas no citizen in population $j$ learns the language of $i$. Gabszewicz et al. (2005) examine the case of linear communicative benefit functions that do not distinguish between communication in the native and the non-native language. Assuming that learning costs are individual-dependent, their paper provides a characterization of linguistic equilibria that include interior equilibria.

The focus of our paper is two-fold. One is a characterization of a linguistic equilibrium in a non-linear setting and the second is the estimation of demand functions for languages. We consider a variant of the Selten–Pool model with two populations $i$ and $j$. Each population speaks its native language and may learn the other one. If the population that speaks $i$ is large relative to the other one, the incentive of an $i$-citizen to learn the other language is likely to be quite low since she can trade and communicate with enough citizens in her own country, but a large population that speaks $j$ may also attract citizens who speak $i$.[4] The substitution between the languages, however, is imperfect, and we assume that the benefits of communication are larger when the languages are native for both sides. We also assume that the benefits of learning the other language are positively correlated with its linguistic proximity to the individual's native language. Indeed, for a native speaker of Portuguese the benefits of learning Spanish is quite limited given the fact a native Portuguese can understand some Spanish without actually learning it. Thus, for any individual $t$ we represent the gross communication benefit by means of an increasing function with three arguments: the number of individuals who share a common native language with $t$, those who speak a language known by $t$ but do not share her native language, and the linguistic proximity between the two languages. We assume that the benefit functions are *supermodular* in the first two variables. This condition implies that an increase in the size of one of the populations raises the marginal communicative benefit of members of the other population to learn the foreign language. We also show that supermodularity is indispensable for our results to hold. Finally, we impose cross-country "learning cost heterogeneity": the difficulty, and thus the cost, of learning a new

---

[3] See e.g., MacManus et al. (1978), Grenier (1985), Lang (1986), and Chiswick (1998).

[4] See Lazear (1999, p. 124). who points out that "the incentives are greater for any individual to learn the majority language when only a few persons in the country speak his or her native language".

language depends on its linguistic proximity with the individual's native language, as a native speaker of Portuguese would find it easier to learn Spanish than Swedish. The fact that learning a foreign language is easier if it is close will have an impact on the number of those who learn it. The net communicative benefit that determines the individuals' behavior is the difference between the gross communicative benefit and the cost of acquiring a new language.

The paper is organized as follows. The intuition on the expected properties of demand functions for foreign languages is confirmed by the theoretical model in Section 2. Section 3 describes the data that will be used to estimate such demand functions, while results are reported in Section 4. Section 5 is devoted to some concluding remarks.

## 2. Modelling the learning of a foreign language

Though one can examine a multi-lingual environment, the main features of the model can be captured in a two-language framework. As indicated in Section 1, we adopt the Selten–Pool paradigm of communicative benefits and assume that the utility of each individual $t$ is determined by the number of individuals she can communicate with. The quality of $t$'s communication in a given language depends on whether the language is native or not. Her utility is therefore represented by the benefit function $U_t(x, y, l)$, where $x$ is the (log of the) number of individuals who speak the same native language as $t$, $y$ the (log of the) number of individuals who share with $t$ a language that is not her native language, and $l$ is the (log of the) linguistic distance between the two languages.[5] For simplicity, we assume that the benefit functions are common to all individuals, so that $U_t(x, y, l) = U(x, y, l)$.

To make our framework more concrete, we consider two countries, $i$ and $j$, with $N_i$ and $N_j$ citizens, respectively. Citizens speak their own native language, denoted by $i$ and $j$ as well, and, for simplicity, we assume that all citizens are originally unilingual, but may consider learning the other language. We denote by $N_{ij}$ (respectively $N_{ji}$) the number of citizens of country $i$ (country $j$) who study language $j$ ($i$).

More specifically, the communicative benefit of an $i$-speaker who learns $j$ (and, thus, is able to communicate with all $j$-speakers) is given by a continuously differentiable and increasing function $U(n_i, n_j, l_{ij})$. (Here $n$ represents the logarithm of the corresponding value of $N$ and $l_{ij} = l_{ji}$ is the (log of the) linguistic distance $L_{ij}$ between languages $i$ and $j$.) The benefit of an $i$-speaker who does not learn language $j$ is $U(n_i, n_{ji}, l_{ij})$: she will communicate with those who know her language in country $j$. For $j$-speakers the levels of communicative benefit are $U(n_j, n_i, l_{ji})$ and $U(n_j, n_{ij}, l_{ji})$, respectively. An individual in country $i$ who learns language $j$ incurs a cost $C(l_{ij})$, determined by the linguistic proximity between the two languages, where the function $C$ is continuously differentiable and increasing. We impose the following additional assumptions.

**Assumption 1.** For every $l > 0$, the function $U(\cdot, \cdot, l)$ is *supermodular*; that is, for every two pairs of positive numbers $\bar{n}_i, n_i, \bar{n}_j, n_j$ with $\bar{n}_i > n_i$, $\bar{n}_j > n_j$, and every $l > 0$ the following inequality holds:

$$U(\bar{n}_i, \bar{n}_j, l) - U(\bar{n}_i, n_j, l) \geq U(n_i, \bar{n}_j, l) - U(n_i, n_j, l).$$

The supermodularity of $U$ implies that an increase in the population of one country raises the marginal communicative benefit of citizens of the other to learn the foreign language. If the

---

[5] Logarithms are used to link the model to the empirical results. This entails no loss of generality.

function $U$ is twice continuously differentiable in its two first variables, this condition amounts to the positivity of the cross-derivative $U_{ij}$ (see Topkis, 1979).

We assume that communicative benefits are sufficiently high relatively to learning costs by requiring that if no $j$-speaker learns $i$, an $i$-speaker would get a positive net benefit from learning $j$. That is, her access to all $j$-speakers outweighs the language learning cost $C(l_{ij})$. If this assumption is violated, then no citizen of country $i$ learns the foreign language. Similarly, it is worthwhile for a $j$-speaker to study $i$ if no $i$-speaker learns $j$. This very mild condition is stated as follows.

**Assumption 2.**

$$U(n_i, n_j, l_{ij}) - U(n_i, 0, l_{ij}) > C(l_{ij}) \quad \text{and} \quad U(n_j, n_i, l_{ij}) - U(n_j, 0, l_{ij}) > C(l_{ij}).$$

Then we have the following proposition.

**Proposition 1.** *Under Assumptions 1 and 2, there exists a unique interior linguistic equilibrium where all individuals are indifferent between learning the foreign language and incurring the cost of learning it and not learning the language.[6] This equilibrium is a solution of the following system of two equations*:

$$U(n_i, n_j, l_{ij}) - C(l_{ij}) - U(n_i, n_{ji}, l_{ij}) = 0, \tag{1}$$

$$U(n_j, n_i, l_{ij}) - C(l_{ij}) - U(n_j, n_{ij}, l_{ij}) = 0. \tag{2}$$

**Proof.** The proof is straightforward. Indeed, Assumption 2 together with the continuity of $U$ in the second argument yields the unique $n_{ji}$ that satisfies (1), whereas Assumption 2 together with the continuity of $U$ in the first argument guarantees the uniqueness of $n_{ij}$ that satisfies (2). $\square$

The interior linguistic equilibrium yields functions $n_{ij}(n_i, n_j, l_{ij})$ and $n_{ji}(n_j, n_i, l_{ij})$, that identify the number of learners of the foreign language in countries $i$ and $j$, respectively. Denote by $\log(N_{ij}/N_i) = D_i(n_i, n_j, l_{ij})$, the equilibrium share (demand function) of individuals whose native language is $i$ and who learn language $j$. The properties of $D_i$ are described in the following proposition.[7] It states that the propensity of an individual to learn another language declines with the size of her own population, but increases with the size of the population of the other country. It is important to point out that the last conclusion hinges upon the supermodularity of the communicative benefit function. Finally, we show that the desire to learn the other language is negatively correlated with the cost of learning when the latter is increasing in the linguistic distance between the two languages.

**Proposition 2.** *Suppose that Assumptions 1 and 2 hold. Then*,

(a) *If $U$ is concave in the second variable, $D_i(\cdot, n_j, l_{ij})$ is decreasing in $n_i$.*
(b) *$D_i(n_i, \cdot, l_{ij})$ is increasing in $n_j$.*

**Proof.**

(a) We can rewrite Eq. (2) as

$$U(n_j, n_i, l_{ij}) - U(n_j, D_i(n_i, n_j, l_{ij})) - C(l_{ij}) = 0. \tag{3}$$

---

[6] There could also be "corner" equilibria, which are not examined here.
[7] Obviously, the same properties hold for $D_j$.

Consider two values, $\bar{n}_i > n_i$. Concavity of $U$ in the second variable implies that

$$U(n_j, \bar{n}_i, l_{ij}) - U(n_j, D_i(n_i, n_j, l_{ij})) - C(l_{ij}) < 0.$$

Since

$$U(n_j, \bar{n}_i, l_{ij}, ) - U(n_j, D_i(\bar{n}_i, n_j, l_{ij})) - C(l_{ij}) = 0,$$

the monotonicity of $U$ implies

$$D_i(\bar{n}_i, n_j, l_{ij}) < D_i(n_i, n_j, l_{ij}).$$

(b) Consider (3) and let $\bar{n}_j > n_j$. Assumption 1 implies that

$$U(\bar{n}_j, n_i, l_{ij}) - U(\bar{n}_j, D_i(n_i, n_j, l_{ij}), l_{ij}) - C(l_{ij}) > 0,$$

but since

$$U(\bar{n}_j, n_i, l_{ij}) - U(\bar{n}_j, D_i(n_i, \bar{n}_j, l_{ij}), l_{ij}) - C(l_{ij}) = 0,$$

the monotonicity of $U$ yields

$$D_i(n_i, n_j, l_{ij}) < D_i(n_i, \bar{n}_j, l_{ij}). \quad \square$$

It is worth pointing out that assertion (b) of Proposition 2 does not necessarily hold if the benefit function is not supermodular. Indeed, consider the example where the benefit function $U(x, y, l)$ is given by $\log(x + y) + \log c$, where $c = e^{C(l_{ij})}$. Then Eq. (2) turns into

$$\log(n_j + n_i) - \log(n_j + n_{ij}) = \log c.$$

Thus,

$$\frac{n_j + n_i}{n_j + n_{ij}} = c \quad \text{and} \quad n_{ij} = \frac{n_i}{c} + \frac{(1 - c)n_j}{c}.$$

However, since $\log c > 0$ and $c > 1$, it follows that the number of learners of language $j$ in country $i$ declines in the population size of country $j$.

Our last proposition claims that a larger linguistic distance between the languages may reduce the number of those who learn foreign languages. In order to prove this assertion we assume that the marginal communicative benefit with respect to linguistic distance declines with the number of learners of one's own language in the other country.

**Assumption 3.** The function $\partial U(n, \cdot, l)/\partial l$ is decreasing for every $n, l > 0$.

Then we have the following proposition.

**Proposition 3.** *Under Assumptions 1–3, the demand function $D_i(n_i, n_j, \cdot)$ is decreasing in $l_{ij}$.*

Table 1
Main languages used in the European Union (millions)

| | Native speakers in the EU[a] | Mother tongue[b] | Woldwide use[c] |
|---|---|---|---|
| English | 62 3 | 341 | 1800 |
| French | 64 5 | 77 | 169 |
| German | 90 1 | 100 | 126 |
| Spanish | 39.4 | 340 | 450 |
| Italian | 57.6 | 62 | 63 |
| Dutch | 21 9 | 20 | 20 |

[a] English is the native language in Great Britain and Ireland. French is the native language in France and is spoken by 40 percent of Belgians. German is the native language in Germany and Austria. Spanish and Italian are the native languages in Spain and Italy, respectively. Finally, Dutch is the native language in the Netherlands and is spoken by 60 percent of Belgians. Note that for Dutch there is a discrepancy between speakers in the EU and the estimates in columns (2) and (3).

[b] Number of first language speakers as given by http://www.ethnologue.com. For Spanish, the number is the average between the two estimates by http://www.ethnologue.com.

[c] Dalby (2002, p. 31) for English and Spanish. The French diplomatic service (http://www.france.diplomatie.fr/francophonie/francais/carte.html) provides the estimation of 169 million people who use French. For Italian and Dutch, see Crystal (2001).

**Proof.** Consider Eq. (3) and two values, $\bar{l}_{ij} > l_{ij}$. Since $C'(\cdot)$ is positive, Assumption 3 implies that

$$\frac{\partial U(n_j, n_i, l_{ij})}{\partial l} - \frac{\partial U(n_j, D_i(n_i, n_j, l_{ij}), l_{ij})}{\partial l} - C'(l_{ij}) < 0$$

and

$$U(n_j, n_i, \bar{l}_{ij},) - U(n_j, D_i(n_i, n_j, l_{ij}), \bar{l}_{ij}) - C(\bar{l}_{ij}) < 0.$$

Then the monotonicity of $U$ in the second argument implies that

$$D_i(n_i, n_j, \bar{l}_{ij}) < D_i(n_i, n_j, l_{ij}).$$

That is, the number of $j$-learners in country $i$ declines if the linguistic distance is larger. $\square$

## 3. Data

We estimate the demand functions derived in Section 2 for English, French, German and Spanish by citizens from the European Union (EU) whose native languages are none of these. The data consist of knowledge of native and foreign languages in various EU countries and distances between languages.

Language proficiency was the topic of a survey on languages ordered by the Directorate of Education and Culture of the EU in 2000.[8] In each of the 15 then EU members, 1000 interviews (with some minor variations) were conducted on the use of languages. Table 1 gives a general overview of the six languages most extensively used by native speakers in the EU15. It also shows to what extent each language is spoken elsewhere than in its country of "origin." Column (1) shows the number of native speakers, in fact the population in each country.[9] Column (2) displays estimates of the worldwide use of each language as mother tongue. The last column gives estimates

---

[8] INRA (2001).

[9] To simplify, we assume that immigrants speak the language of the country to which they migrated.

Table 2
Knowledge of languages in the European Union (millions and percent)

| Country (1) | Native language known by (millions) (2) | Percentage who know | | | |
|---|---|---|---|---|---|
| | | English (3) | French (4) | German (5) | Spanish (6) |
| Austria (G) | 100.0 | 46 | 11 | 100 | 1 |
| Denmark (Dk) | 5.3 | 75 | 5 | 37 | 1 |
| Finland (Fi) | 6.0 | 61 | 1 | 7 | 1 |
| France (F) | 77.0 | 42 | 100 | 8 | 15 |
| Germany (G) | 100.0 | 54 | 16 | 100 | 2 |
| Greece (Gr) | 12.0 | 47 | 12 | 12 | 5 |
| Italy (I) | 62.0 | 39 | 29 | 4 | 3 |
| Ireland (E) | 341.0 | 100 | 23 | 6 | 2 |
| The Netherlands (D) | 20.0 | 70 | 19 | 59 | 1 |
| Portugal (P) | 176.0 | 35 | 28 | 2 | 4 |
| Spain (S) | 340.0 | 36 | 19 | 2 | 100 |
| Sweden (Sw) | 9.0 | 79 | 7 | 31 | 4 |
| United Kingdom (E) | 341.0 | 100 | 22 | 9 | 5 |

The native language in each country is given between parentheses (G, German; Dk, Danish; Fi, Finnish; F, French; G, German; Gr, Greek; I, Italian; E, English; D, Dutch; P, Portuguese; S, Spanish; Sw, Swedish). The numbers in the first column are from http://www.ethnologue.com. The percentages of people who know English, French, German and Spanish in each country are from Ginsburgh and Weber (2005).

of worldwide knowledge as mother tongue and otherwise. English, French and German are the languages most widely spoken in the EU. Though Italy has a larger population than Spain, and the number of native speakers of Italian is larger, the language is hardly spoken outside of Italy,[10] which is less so for Spanish. Dutch is spoken only in the Netherlands and Belgium. Therefore, in Table 2 we restrict our attention to the "knowledge" of four non-native languages (English, French, German, and Spanish) in 13 members of EU15 that have a dominant language.[11] Column (2) contains the world population that speaks the language of the country listed in column (1) as first language.[12] The other four columns give for each EU country the share of the total population that (claims to) know English, French, German and Spanish. Table 3 displays distances between languages derived by Dyen et al. (1992).[13]

## 4. Estimation results

The general idea is to estimate the following demand function for languages $j$ ($j$ = English, French, German, Spanish) by those whose native language is $i$ ($i$ = Austria, Denmark, Finland, France, Germany, Greece, Italy, Ireland, the Netherlands, Portugal, Spain, Sweden and the United Kingdom):

$$\log\left(\frac{N_{ij}}{N_i}\right)_{EU} = \alpha_0 + \alpha_1 n_i + \alpha_2 n_j + \alpha_3 l_{ij} + u_{ij}, \tag{4}$$

---

[10] Italian is known by 9 percent of the population in Austria, 6 percent in Belgium and Greece, 5 percent in France, and 1 percent in other countries of the EU.

[11] Belgium and Luxembourg are excluded with no major impact on our results.

[12] Following http://www.ethnologue.com.

[13] See also Ginsburgh et al. (2005) where these distances are used to compute optimal sets of languages in the EU before the last two enlargements.

Table 3
Distances between languages (1000 ×)

|  | English | French | German | Spanish |
|---|---|---|---|---|
| Danish | 407 | 759 | 293 | 750 |
| Dutch | 392 | 756 | 162 | 742 |
| English | 0 | 764 | 422 | 760 |
| Finnish | 1000 | 1000 | 1000 | 1000 |
| French | 764 | 0 | 756 | 266 |
| German | 422 | 764 | 0 | 747 |
| Greek | 838 | 843 | 812 | 833 |
| Italian | 753 | 197 | 735 | 212 |
| Portuguese | 760 | 291 | 753 | 126 |
| Spanish | 760 | 266 | 747 | 0 |
| Swedish | 411 | 756 | 305 | 747 |

*Source*: Dyen et al. (1992) for further details. See also Ginsburgh et al. (2005).

where $(N_{ij}/N_i)_{\text{EU}}$ represents the proportion of inhabitants of EU country $i$ who are proficient in language $j$ (columns (3)–(6) in Table 2); $n_i$ and $n_j$ represent respectively the (log of the) world populations whose native languages are $i$ and $j$ (column (2) in Table 2) and $l_{ij}$ is the (log of the) distance between languages $i$ and $j$ (Table 3). Proposition 2 suggests that the signs of coefficients should be $\alpha_1 < 0, \alpha_2 > 0$. The sign of $\alpha_3$ results from two opposite effects, a positive one through the utility function (a distant language is more interesting and beneficial to learn), and a negative one due to the cost of learning a new language, which is higher if the distance is larger. The two effects cannot be identified separately, and only the net effect can be estimated. Proposition 3 indicates that under reasonable conditions the cost effect outweighs the benefit advantage, yielding $\alpha_3 < 0$.

We first estimate a demand function for each foreign language $j$ separately:

$$\log \left( \frac{N_{ij}}{N_i} \right)_{\text{EU}} = \alpha_0 n_j + \alpha_1 n_i + \alpha_3 l_{ij} + u_{ij}.$$

Note that in each case, the intercept $\alpha_0$ is multiplied by the world population that practices language $j$ and can be interpreted as $\alpha_2$. This normalization (which has no consequence on other coefficients) will make it possible to give a first insight into the attraction power of individual foreign languages. Estimation results are reproduced in Table 4. They show that the the fit is excellent and consistent with theory for English and German. This is not the case for French and Spanish, though distance always picks a negative sign. The coefficient for the country of origin, which should also be negative, is so for English and German only. Finally, observe that the (population-weighted) intercept terms are all positive, but their magnitudes differ widely, giving a first (expected) indication that the four languages exercise different attraction powers.[14]

The number of observations for single languages is rather small (11 or 12), and pooling makes it possible to estimate the basic equation using more observations. Pooling also checks whether the relations that govern the learning of a foreign language hold more generally than for each language individually.

Estimation of Eq. (4) on the full set of 46 observations leads to poor results that are not reported. The results given in the last column of Table 4 are for an equation that has the same form, except

---

[14] We also computed such equations using native populations in each country instead of the number of speakers of each language given in column (2) of Table 2. The results are qualitatively similar.

Table 4
Estimation results

| | English | French | German | Spanish | All four | Trade shares |
|---|---|---|---|---|---|---|
| Population speaking language $i$ ($\alpha_1$) | −0.153*(0.021) | 0 355*(0.138) | −0.361*(0.072) | 0.032(0.168) | −0.058(0.069) | −0.055(0.070) |
| Population speaking language $j$ ($\alpha_2$) | | | | | 0.625*(0.057) | 0.600*(0.067) |
| Distance between languages $i$ and $j$ ($\alpha_3$) | −0.408*(0.082) | −0 512(0.416) | −1.362*(0.214) | −0.560(0.385) | −0.954*(0.200) | −0.789*(0.205) |
| Intercept ($\alpha_0$) | 0.733*(0.016) | 0 193(0.121) | 0.586*(0.077) | 0.091(0.109) | 0.080(0.100) | 0.070(0.096) |
| French speaking population ($\alpha_{0F}$) | | | | | −0.112(0.062) | |
| German speaking population ($\alpha_{0G}$) | | | | | −0.233*(0.061) | |
| Spanish speaking population ($\alpha_{0S}$) | | | | | −0.514*(0.050) | −0.340*(0.044) |
| Trade share | | | | | | 0.249(0.134) |
| $R^2$ | 0.919 | 0 599 | 0.910 | 0.232 | 0.758 | 0.712 |
| No. of observations | 11 | 12 | 11 | 12 | 46 | 46 |

Standard errors are given between brackets, next to the coefficients. Starred coefficients are significantly different from 0 at the 5 percent (or 1 percent) probability level. The number of observations is equal to 12 for France and Spain, since French and Spanish is a foreign language in 12 of the 13 countries. This number is 11 for English (spoken in the UK and Ireland) and German (spoken in Germany and Austria).

that it contains an $\alpha_{0j}\delta_j n_j$ term for each language $j =$ French, German and Spanish, where $\delta_j$ is a dummy variable equal to 1 for a language $j$ that is acquired, and 0 otherwise. As can be checked, the $\alpha_{0j}$ parameters take significantly different values and explain why the fit of the general equation is poor. The results are consistent with the theoretical model ($\alpha_1$ is negative, but fails to be significantly different from 0 at the usual 5 percent level), but the four languages have different attraction powers. The coefficients can be considered to be elasticities. The "distance elasticity" is not significantly different from $-1$, implying that a 1 percent increase in the distance between two languages decreases the number of its learners by 1 percent. This is far from being negligible. The "origin elasticity" is small (5 percent), but the "destination elasticity" is large for English and French (63 and 52 percent), somewhat smaller for German and Spanish.

The last equation shows the effect of trade shares[15] (which do not appear in the theoretical model) as possible determinants (or incentives) to learn foreign languages. The parameter picked by the trade shares variable is almost significant at the 5 percent probability level, but hardly changes the values of the other parameters, though now English, French and German benefit from similar attraction power (the null hypothesis that they are equal cannot be rejected).

## 5. Concluding comments

Our results show that, in conformity with the theoretical model, three variables explain reasonably well the share of people who learn a foreign language, without taking into account the incentives every individual has to acquire a language. The larger the native population that speaks the language, the fewer speakers are prone to learn another language; the more the foreign language is spoken, the more it attracts others to learn it; and the larger the distance between two languages, the smaller the proportion of people who will learn it. However, our results also show that the attraction powers of the four foreign languages are significantly different, and that other determinants, mostly historical, must be at play. Spanish, for instance, should attract Europeans much more than it currently does. With the exception of France, there is no country in which more than 5 percent of the population knows the language. The isolation of Spain until 1975 is a partial explanation of this phenomenon, but the large population of native Spanish speakers in Mexico and South America, and increasingly in the United States, does not seem to generate large incentives to learn the tongue. France, on the other hand, is quite restrictive on the use of foreign languages in the public domain. For instance, it issued regulations that make it difficult for certain categories of civil servants to use languages other than French in national and even international meetings.

Dynamics, past and current cultural relations, common borders (Germany and the Netherlands), non-centrality of a country (Spain, and to some extent, France) can be represented by the trade shares. This does not improve the estimation results in any major way. Therefore our model provides only partial answers to the questions of why English is becoming the *lingua franca* in Europe (and probably in the world), and Spanish is relatively less spoken in Europe.

## References

Chiswick, B R., 1998. Hebrew language usage: determinants and effects on earnings among immigrants in Israel. Journal of Population Economics 11, 253–271.

---

[15] Value of exports and imports of country $i$ from and to English, French, German or Spanish speaking countries.

Church, J., King, I., 1993. Bilingualism and network externalities. Canadian Journal of Economics 26, 337–345.

Crystal, D., 2001. A Dictionary of Language. Chicago University Press, Chicago.

Dalby, A., 2002. Languages in Danger. Allen Lane-The Penguin Press, London.

Dyen, I., Kruskal, J.B., Black, P., 1992. An Indo-European classification: a lexicostatistical experiment. Transactions of the American Philosophical Society 82 (part 5).

Gabszewicz, J., Ginsburgh, V., Weber, S., 2005. Bilingualism and Communicative Benefits. Mimeo.

Ginsburgh, V., Ortuño-Ortín, I., Weber, S., 2005. Language disenfranchisement in linguistically diverse societies. The case of the European Union. Journal of the European Economic Association 3, 946–965.

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

Grenier, G., 1985. Bilinguisme, transferts linguistiques et revenus du travail au Québec, quelques éléments d'interaction. In: Vaillancourt, F. (Ed.), Economie et Langue. Editeur officiel, Québec, pp. 243–287.

INRA, 2001. Eurobaromètre 54 Special, Les Européens et les Langues.

Lang, K., 1986. A language theory of discrimination. Quarterly Journal of Economics 100, 363–381.

Lazear, E., 1999. Culture and language. Journal of Political Economy 107, S95–S126.

MacManus, W., Gould, W., Welsch, F., 1978. Earnings of Hispanic men: the role of English language proficiency. Journal of Labor Economics 1, 101–130.

Selten, R., Pool, J., 1991. The distribution of foreign language skills as a game equilibrium. In: Selten, R. (Ed.), Game Equilibrium Models, vol. 4. Springer-Verlag, Berlin, pp. 64–84.

Topkis, D., 1979. Equilibrium points in non-zero sum $n$-person submodular games. SIAM Journal of Control and Optimization 17, 773–787.