



Working Paper 03-34  
Economics Series 10  
July 2003

Departamento de Economía  
Universidad Carlos III de Madrid  
Calle Madrid, 126  
28903 Getafe (Spain)  
Fax (34) 91 624 98 75

**THE CATALAN PREMIUM:  
LANGUAGE AND EMPLOYMENT IN CATALONIA \***

Sílvio Rendon<sup>1</sup>

***Abstract***

This paper measures the contribution of knowing Catalan to finding a job in Catalonia. In the early eighties a drastic language policy change (**normalització**) promoted the learning and use of Catalan and managed to reverse the falling trend of its relative use versus Castilian (Spanish), thereby recovering its economic value. Using census data for 1991 and 1996, I estimate a significant positive Catalan premium: the probability of being employed increases between 3 and 5 percentage points if individuals know how to read and speak Catalan; it increases between 1 and 4 percentage points for writing Catalan. This premium is substantially higher for women than for men, and decreasing in schooling and in age.

**Keywords:** Language, Immigration, Skill Premium.

**JEL Classification:** J61, J70, J31, I20.

<sup>1</sup>The University of Western Ontario and Universidad Carlos III de Madrid, Spain;  
E.mail: srendon@uwo.ca

\* I thank Manuel Arellano, Jordi Galí, Albert Marcet, Núria Quella, Ernesto Villanueva and participants of seminars at Universitat Pompeu Fabra, Universidad Carlos III, Centra, CEMFI, and FEDEA for helpful discussions. I also thank the Institut d'Estadística de Catalunya (IDESCAT) for providing the data. Financial support of the Spanish Ministry of Science and Technology (Grant SEC 2001-0674) is gratefully acknowledged. Marta González provided excellent research assistance in an early stage of this project. The usual disclaimer applies.

# 1 Introduction

What is the economic value of knowing Catalan in Catalonia for an individual? During most of the twentieth century Catalan, although official during most of Catalonia's history, was given an informal status, thus losing its economic value. Public policies supported Castilian (Spanish<sup>1</sup>) against Catalan until the mid-seventies, but they reversed at the beginning of the eighties, and Catalan has since been given priority over Castilian. This switch in language policy (*normalització*) implied not only that immigrants<sup>2</sup> had to learn the local language to assimilate to a new economy, but also that locals had to become more proficient in the use of their own language. It is clear that in current times Catalan has recovered its economic value, but it is not clear how much this value represents.

In this paper I measure the contribution of knowing Catalan to the probability of being employed in Catalonia. I use a sample of census data for the last two available census years, 1991 and 1996. The data consist of socioeconomic variables, employment status, and knowledge of Catalan language. Accounting for the determinants of selection into knowing Catalan and verifying the sensitivity of results to specification used, I estimate the probability of being employed conditional on a level of Catalan knowledge by gender and census year. Speaking and reading Catalan increases the probability of being employed by between 3 and 5 percentage points, whereas writing Catalan increases the probability of being employed by between 1 and 4 percentage points. This premium is higher for women than for men and decreasing in years of schooling and in age. It is also positive both for individuals who know Catalan and for those who do not know it, although it is higher for the former than for the latter.

The economic incentive for learning Catalan is thus crucial, but not sufficient to explain differences in Catalan knowledge; non-economic considerations are also im-

---

<sup>1</sup>Castilian is generally known as "Spanish" outside the Iberian Peninsula, except in parts of Latin-America.

<sup>2</sup>In Catalonia the term "immigrant" is used indistinctly for all born outside Catalonia, be they born in the rest of Spain or in another country.

portant: being born in Catalonia, arriving at a young age and having spent more time in Catalonia increase the probability of knowing Catalan. In particular, externalities, such as living in a community where people know Catalan and are Catalan born, increase also the probability of knowing Catalan.

The economic literature treats language mostly as a particular form of human capital which, as such, has an economic value. As Chiswick & Miller (1995) explain, language has all the characteristics of human capital: it is embodied in the person, it is productive in the labor market and/or in consumption, and it is created at a sacrifice of time and out-of-pocket resources. Most of the literature on language has focused on measuring the effect of its acquisition on the absorption of immigrants in several countries: the USA, the UK, Canada, Australia, Israel, Germany (McManus 1985, Kossoudji 1988, Dustmann 1994, Rivera-Batiz 1990, Chiswick & Miller 1995). In this sense, language acquisition is a chapter in the study of immigrants' assimilation to a recipient economy (Borjas 1987).

Language, however, also has an economic value for the native population. Most societies are composed of different speech communities, which communicate using related but different verbal and nonverbal languages (Lang 1986). Thus, economic life is segregated by speech community to minimize transaction costs that are ultimately paid by the minority group, which tends to learn the language of the majority. As Lazear (1999) remarks, countries in which different languages are used, as a result of immigration or not, may wonder about the convenience of converging toward one language. In this line, there are several cases in which the official language of a country was changed. For example, Angrist & Lavy (1997) study the effect of arabization, that is, the switch from French to Arabic instruction, in the local population of Morocco, and find there that was a decrease in the returns to schooling. In Europe, there are also several cases in which the official language was reversed; some examples are: from German to Hungarian in Hungary, or from Russian to Estonian in Estonia (with similar changes in other Baltic countries). These language changes, which are

similar to the change experienced in Catalonia in the eighties, involve not only the ‘immigrants’ but also the ‘natives.’

The importance of Catalan language is illustrated in Figure 1, based on Reixach (1997) . It shows the percentage of people who understand, speak, read, and write Catalan in 1991 in Catalonia by *comarques*, a political subdivision smaller than the province. Only in Barcelona the percentage of people who understand Catalan is below 90%, but still above 75%. The percentage of people in each county that speak or read Catalan is at least 70%. For written Catalan these percentages can fall to 30% even in counties where Catalan is widely spoken, showing the big regional disparity in the different levels of Catalan proficiency.

Figure 2 portrays the evolution of the percentage of the population that speaks Catalan in Catalonia, Valencia, and the Balearics.<sup>3</sup> Until 1975, the percentage of Catalan speakers went down in all three areas. In Valencia and the Balearics this percentage has continued to go down. Only in Catalonia is this trend strongly reversed: the percentage of Catalan speakers has increased up to the level it used to have in the thirties. This change appears to be the result of the implementation of the Catalan government’s public policies oriented to promote the learning and the use of Catalan.

Because of the coexistence of Castilian and Catalan, Catalonia is sometimes perceived to be a bilingual society. Several sociolinguistic studies rather show that Catalonia is characterized by diglossia rather than bilingualism (Vallverdú 1970, Strubell i Trueta 1981, Tudela 1986, Woolard 1989). Bilingualism means that two languages coexist and both are used regularly and indistinctly for similar functions and with equal level of proficiency. On the contrary, diglossia happens when one language is reserved for education, formal communication and literature and the second or diglossic

---

<sup>3</sup>These are the three main areas where Catalan is spoken. It should be noted that in Valencia census data refer to “Valencian,” which is generally recognized as Catalan in Catalonia, but not always in Valencia. Figure 1 is based on the data presented in Table 1, pp 130-131, of Vallverdú (1990). I updated the percentages for 1991 using census data, which are available for the three areas. For 1996 there are data only for Catalonia.

language is used for family and ordinary life, popular culture and folklore.

The economic value of a diglossic language, expressed in the wage or employment probability premium, is not expected to be very large. Given this situation, one consensual objective in Catalonia has been to eliminate diglossia and move on to bilingual society, for which active language policies increased the economic value of knowing and using Catalan in Catalonia. This paper is the first attempt to measure the individual's economic value of Catalan. At this moment, data availability makes only possible to compute this economic value in terms of employment rates, not in terms of wages, as it is usually done. Further research will be possible, if future employment and wage surveys in Catalonia include information on language variables.

This paper is organized as follows. Section 2, presents a basic historical review of language policies in Catalonia. Section 3 describes a model of language choice and employment status. Section 4 discusses the data used in the estimation. Section 5 presents the estimation results: the estimated coefficients, the predicted and counterfactual Catalan knowledge and employment probabilities, the premium and its marginal effects, and a comparison of these results under different specifications. Section 6 summarizes the main conclusions of this paper.

## **2 Language Policy in Catalonia**

There are four official languages in Spain: Castilian, Catalan, Basque and Galician. Presently, Castilian is official in the totality of the territory, and each of the three other languages shares co-officiality with Castilian in its own territory.<sup>4</sup> However, after the Spanish Civil War (1936-39), and up to the death of Franco in 1975, all public uses of any local language other than Castilian were abolished; and, until the Spanish Constitution was passed in 1978, the only official language anywhere in Spain was Castilian. The interested reader can see the articles on language of the Spanish

---

<sup>4</sup>Altogether, around forty percent of the population of Spain lives in areas where there are two official languages.

Constitution and of the Autonomy Statute for Catalonia in Appendix A.1.

In 1980 the Generalitat, the autonomous Catalan government, was re-established and gradually recovered its pre-war competencies. At that time, Catalan language was clearly a diglossic language. The massive immigration flows from the rest of Spain received in the sixties, together with the decades-long absence of all public and formal use of Catalan, had allowed diglossia to institutionalize. Moreover, a very high proportion of the generations schooled in the period from 1939 to 1975 was not able to read or write in Catalan, even if it was their native language. Table 1 shows a basic chronology of some language-related historical facts.

Thus, the Generalitat actively promoted the learning and the use of Catalan at the expense of Castilian in Catalonia. To this object, it implemented a set of regulations and incentives with the long-term goal of having Catalan become the normal language of public and private use in Catalonia. This is known as the “Normalization policy” (*normalització*), which can be divided into two main periods:

1. From 1980 to 1990, characterized by the sensitization of the population and an increase in the knowledge of the language. This stage depended primarily on schools for implementation.
2. From 1990 up to today, in which the objective is to extend the social use of Catalan.

During the first stage, a General Directorship of Linguistic Policy was created in 1980, and in 1983 the Linguistic Normalization Bill was passed unanimously in the Catalan Parliament. This law extended the use of Catalan, particularly in the fields of education, public administration and public media. It was the legal instrument which allowed the progressive replacement of Castilian by Catalan as the main language of instruction in primary, middle and secondary schools.<sup>5</sup> The municipal Catalan

---

<sup>5</sup>The language of instruction in Catalan universities is not clearly determined by law. Therefore, it is, in practice, left to the instructor’s choice.

language services acted as the main instruments of language diffusion among the adult population, as did the new Catalan language mass media created by the Generalitat. See Table 2 for a description of Normalization measures.

The second stage has seen the territorial and sectorial expansion of the language policy. In 1998, the Parliament approved a new law, which extends the presence of Catalan in previously not regulated domains, such as privately owned media, cultural industries or the socioeconomic sphere. Measures in each of these respective domains are: a minimum percentage of broadcast time in Catalan, quotas on cinematographic distribution and showings, and specific regulations to promote the use of Catalan in publicly owned enterprises, public service enterprises, customer service, consumer information, publicity and professional activity.

In sum, these public policies, absent or weaker in other Catalan areas, seem to account for the recovery of Catalan language in Catalonia depicted in Figure 2. In the next section, I set up a model to determine an individual's Catalan knowledge and to what extent this knowledge influences his or her employment status.

### 3 Model

The model proposed in this section is tailored to the available data.<sup>6</sup> Knowing a language means an economic reward, but it also implies directly a satisfaction, conditional on contextual factors like family background and language most used in the community. As Willis & Rosen (1979), I proceed in two stages: one in which selection into knowing the language is determined depending on economic and non-economic incentives, and a second one in which economic rewards are realized conditional on individuals' characteristics.

Consider an individual whose utility depends on the expected economic value  $EV_c$ , if she shows Catalan and on  $EV_n$ , if she does not. Let  $\alpha$  be an individ-

---

<sup>6</sup>For a model which incorporates marriage markets, language choice, and earnings outcomes see Robinson (1988).

ual's non-economic reward for knowing Catalan. She chooses to know Catalan to  $\max \{EV_c + \alpha, EV_n\}$ . After deciding whether to know Catalan or not, the individual goes to the job market where she can become employed or unemployed. Employers only hire individuals whose value exceeds a certain threshold, which is normalized to zero, that is, employers  $\max [V_j, 0]$ ,  $j = \{c, n\}$ , where  $V_c$  is the individual's value if she know's Catalan and  $V_n$  is the value if she does not know Catalan. Let the vector of attributes  $X$  determine the economic value for each language knowledge:

$$V_c = X\beta_c + u_c, u_c \sim N(0, \sigma_c^2), \quad (1)$$

$$V_n = X\beta_n + u_n, u_n \sim N(0, \sigma_n^2), \quad (2)$$

where  $\text{cov}(u_c, u_n) = 0$ . The non-economic value of Catalan knowledge depends on a vector of attributes  $Z$ :

$$\alpha = Z\delta + \varepsilon, \varepsilon \sim N(0, \sigma^2),$$

where, in general,  $\text{cov}(\varepsilon, u_j) \neq 0$ ,  $j = \{c, n\}$ . Define the difference between the total expected value of knowing and of not knowing Catalan as  $I = EV_c + \alpha - EV_n$ . Substituting the above definitions into this expression yields the following reduced form:

$$I = X(\beta_c - \beta_n) + Z\delta + \varepsilon, \quad (3)$$

This selection rule determines whether the individual knows Catalan or not. Let  $y_1$  be an indicator variable assuming the values:

$$y_1 = \begin{cases} 1, & \text{if } I > 0, \text{ the individual knows Catalan, and} \\ 0, & \text{if } I \leq 0, \text{ the individual does not know Catalan.} \end{cases}$$

Let  $X_1 = [X|Z]$  and  $\beta'_1 = [\beta'_c - \beta'_n | \delta']$ ,  $\Phi(\cdot)$  is the univariate cumulative normal function, then a simple probit estimation consists in maximizing the likelihood



function

$$\ln \mathcal{L}(\beta_1|X_1) = \sum [y_1 \ln \Phi(X_1\beta_1) + (1 - y_1) \ln \Phi(-X_1\beta_1)],$$

thereby obtaining the reduced form parameters of the selection equation  $\widehat{\frac{\beta_1}{\sigma_\varepsilon}}$ .

To perform the second stage estimation, the employment equation, let the difference in predicted value be  $\hat{I} = X_1\hat{\beta}_1$ . Estimating the probability of being employed directly from Equation (1) and from Equation (2) would produce a selection bias, because the subset of individuals who know and do not know Catalan is not a random sample of the population. I use the selection equation (3) in a conditional likelihood estimation. Let  $y_2$  be an indicator variable assuming the values:

$$y_2 = \begin{cases} 1, & \text{if } V_j > 0, \text{ the individual is employed, and} \\ 0, & \text{if } V_j \leq 0, \text{ the individual is not employed.} \end{cases}$$

The log-likelihood functions for employment status conditional on knowing Catalan and on not knowing Catalan are

$$\begin{aligned} \ln \mathcal{L}(\beta_c, \rho_c|X) &= \sum [y_2 \ln \Phi_2(X\beta_c, \hat{I}, \rho_c) + (1 - y_2) \ln \Phi_2(-X\beta_c, \hat{I}, -\rho_c) - \ln \Phi(\hat{I})], \text{ and} \\ \ln \mathcal{L}(\beta_n, \rho_n|X) &= \sum [y_2 \ln \Phi_2(X\beta_n, -\hat{I}, -\rho_n) + (1 - y_2) \ln \Phi_2(-X\beta_n, -\hat{I}, \rho_n) - \ln \Phi(-\hat{I})], \end{aligned}$$

where  $\Phi_2(\cdot)$  is the bivariate cumulative normal function. Parameters  $\hat{\beta}_c$  and  $\hat{\rho}_c$  and  $\hat{\beta}_n$  and  $\hat{\rho}_n$  are the respective maximizers of these two conditional functions, constructed over the two different samples. With these estimated parameters I compute the relevant counterfactuals and with them the employment probability premium for knowing Catalan for each individual:

$$\Delta = E(y_2|X, X_1, y_1 = 1) - E(y_2|X, X_1, y_1 = 0) = P_c - P_n,$$

where  $P_c = \frac{\Phi_2(X\hat{\beta}_c, X_1\hat{\beta}_1, \hat{\rho}_c)}{\Phi(X_1\hat{\beta}_1)}$  and  $P_n = \frac{\Phi_2(X\hat{\beta}_n, -X_1\hat{\beta}_1, -\hat{\rho}_n)}{\Phi(-X_1\hat{\beta}_1)}$ . The average Catalan premium is then  $\bar{\Delta} = N^{-1}\Sigma\Delta = \bar{P}_c - \bar{P}_n$ , where  $\bar{P}_c = N^{-1}\Sigma P_c$  and  $\bar{P}_n = N^{-1}\Sigma P_n$ . The

premium's variance<sup>7</sup> is

$$Var(\bar{\Delta}) = N^{-1} \left( N^{-1} \Sigma \Delta^2 - \bar{\Delta}^2 \right) + \frac{\partial \bar{\Delta}'}{\partial \beta_1} Var(\hat{\beta}_1) \frac{\partial \bar{\Delta}}{\partial \beta_1} + d'_c Var(\hat{\beta}_c, \hat{\rho}_c) d_c + d'_n Var(\hat{\beta}_n, \hat{\rho}_n) d_n,$$

where  $d'_c = \left[ \frac{\partial \bar{P}_c'}{\partial \beta_c}, \frac{\partial \bar{P}_c'}{\partial \rho_c} \right]$  and  $d'_n = \left[ -\frac{\partial \bar{P}_n'}{\partial \beta_n}, -\frac{\partial \bar{P}_n'}{\partial \rho_n} \right]$ . Given that employment parameters by Catalan knowledge are computed separately, they are independent. Marginal effects of a given covariate in the premium are computed as a difference-in-differences of the predicted conditional probabilities of employment. Hence, marginal effects are defined as the effect of a given (not an infinitesimal) increase in a determined variable on the conditional probability of being employed. Further details on this computation are given in Appendix A.2.

## 4 Data

Data come from the Catalan and Spanish National Statistical Institutes (IDESCAT-INE). This is the only data source in Spain that includes a question on language knowledge. I use two samples of 250 000 randomly selected individuals each from census data for years 1991 and 1996. Available variables are personal attributes such as gender, age, marital status, schooling, place of residence, place of birth, number of years in Catalonia, occupational status and knowledge of Catalan. Unfortunately, in this sample there is no way relate individuals across time or to use family level information, that is, to link two individuals of the same family, which can be interesting for analyzing the influence of parents' Catalan knowledge on their children. I select the main household members, that is, only parents and children, who are between 16 and 60 years old, born in Spain and who are in the labor force. Notice that this last condition is particularly important for women, as their participation rate is lower than men's. Individuals for whom there is no information on their level of Catalan

---

<sup>7</sup>I thank Manuel Arellano for his suggestions on computing this variance.

knowledge or for whom their arrival year is not available, if they were born outside of Catalonia, are excluded from the sample. The selected sample consists of 96 863 individuals for year 1991, and of 96 985 for 1996. Appendix A.3. gives further details on the sample selection.

In the census data the level of Catalan knowledge is self-assessed;<sup>8</sup> the respondent has to cross with an X the statement that applies:

1. R. (respondent) understands it.
2. R. does not understand it.
3. R. understands and reads it
4. R. speaks it.
5. R. speaks and reads it.
6. R. speaks, reads and writes it.

In the survey's instructions the interviewer has the following definitions to orient the respondent:

1. Somebody 'understands' Catalan when s/he is able to understand a conversation about any topic in Catalan.
2. Somebody 'reads' Catalan when s/he is able to read texts such as signs, newspapers' news, etc.
3. Somebody 'speaks' Catalan when s/he is able to maintain a conversation about a current topic.
4. Somebody 'writes' Catalan when s/he is able to write notes, letters, etc. with enough, not necessarily total, accuracy.

This variable is, thus, subjective and only informative about the potential use of Catalan language by the individual. I divide the level of Catalan knowledge in three levels:

1. Basic level: Understanding, reading or speaking.
2. Reading and Speaking.
3. Writing.

---

<sup>8</sup>Unfortunately, as Charette & Meng (1994) point out, self-assessed data can lead to biased inference. Future research on the subject may be based on language tests and on earnings data.

For each year I construct two subsamples according to gender for the following variables: Catalan knowledge and employment status, the dependent variables; the explanatory variables in the language and in the employment estimation, the  $X$  variables, and the explanatory variables only in the language estimation, the  $Z$  variables.

Table 1 presents descriptive statistics by subsample and by variable. Appendix A.4. explains the definition of these variables in further detail. From 1991 to 1996 both men and women appeared to have improved their level of Catalan knowledge, being women in both years were more proficient in Catalan than men. While in 1991 38% of men and 30% of women knew Catalan at a basic level, in 1996 these percentages dropped to only 26% for men and 19% for women. This descriptive evidence should be taken cautiously, because these are different samples. As it is shown in the next section, an improvement in Catalan knowledge is observed, even after accounting for the variation in the samples.

This general improvement in Catalan proficiency from 1991 to 1996 coincided with a worsening of labor market conditions: employment rates dropped for everybody, ranging the decline between 4 and 5 percentage points. However, looking at the employment rates by levels of Catalan knowledge, one can see that the basic shape is persistent. In both years employment rates are substantially higher for men who read and speak Catalan than for those who do not. Men who write Catalan have slightly lower employment rates than men who only read and speak Catalan. For women, in contrast, employment rates are monotonically increasing in the level of Catalan knowledge. Notice also that individuals who were not born in Catalonia do not exhibit substantially different employment rates than average employment rates.

The  $X$  variables are schooling, age, a dummy indicating whether the individual was affected by the Normalization policy, marital status, residence variables and Municipal variables. For both genders, years of schooling increased from year 1991 to 1996; women, however, show more schooling and are younger than men. The average age in the sample is 38 for men and around 35 for men. A dummy variable indicates

whether the individual was affected by the ‘Normalization’ process: if the individual was younger than 12 years old in 1984, this dummy takes the value of one and is zero otherwise. Understandably, there is a higher percentage of individuals affected by the Normalization in 1996 than in 1991. In the sample there is a higher percentage of married men than women; however, from 1991 to 1996 marriage rates of both genders decrease. Most of the population of Catalonia resides in the province of Barcelona. The provinces of Lleida, Girona and Tarragona concentrate, respectively, 5.6, 8.5, and 8% of individuals in the sample.

I also use average data at the district area called *municipi*, a smaller unit than the *comarques*. These data are assigned to each individual in the samples: out of 945 *municipis* in Catalonia, the sample of 1991 covers 943 *municipis* while the sample of 1996 covers 944. There is, however, a high concentration of individuals, almost one fourth of each sample, in the *municipi* of Barcelona. Municipal variables are the percentage of employed in the total population (PartM), the percentage of employment in services (ServM), the percentage of people who write Catalan (WriteM), the percentage of people born in Catalonia (CbornM). At the *municipi* the average employment rate is around 36%, while the percentage of employment in services is 52% in 1991 and 58% in 1996. These variables capture the externality effect of the community on individuals’ employment status, and as such are classed as *X* variables. The percentage of people who write Catalan in the *municipi* is 40% in 1991 and 46% in 1996. People born in Catalonia amount around two thirds of the Municipal population. These two variables reflect the externality effect of the community on Catalan knowledge, and are therefore *Z* variables.

Additional *Z* variables are indicators of origin and of migration. As shown above, people born outside Catalonia do not exhibit notorious differences in their employment rates with the average employment rates. However, as it will be clear in the next section, origin has an important effect in Catalan knowledge. Approximately one third of the sample consists of people born outside Catalonia. Around one third

of them arrived when they were no older than 10 years, and more than two thirds of the people born outside of Catalonia come from Andalusia. A very reduced proportion of individuals in the sample were born in Valencia, in the Balearics and in La Franja (three provinces of Aragon where Catalan is widely used), areas outside Catalonia where Catalan is also official. For individuals born outside of Catalonia an important variable is the time that they have spent in Catalonia, which is captured by the variable YSM (Years since migration). One can see that in 1991 they spent on average between 23 and 25 years in Catalonia, reflecting that on average people arrived to Catalonia at the end of the sixties.

## **5 Estimation Results**

This section discusses the results of the estimations in two stages and the computation of the premium and marginal effects described in Section 3. I perform two different estimations of the probability of knowing Catalan. The first one defines that an individual ‘knows’ Catalan, if she at least reads and speaks Catalan, otherwise she is defined as not knowing it; in a second definition, knowing Catalan means being able to write it.

### **5.1 Catalan Knowledge**

Table 4 reports the coefficients of the selection equation into reading and speaking Catalan on the left side and into writing Catalan on the right side. The covariates used in the estimations are the  $X$  and  $Z$  variables discussed in the previous section. The  $X$  variables include square terms for schooling and age and an interaction term for schooling and age; and the  $Z$  variables include an interaction term for A9 and YSM.

There are some clear common features in the estimated coefficients for the four

subsamples.<sup>9</sup> For both skills, the effect of years schooling is positive and significant but with decreasing returns. Age tends to have a positive effect on reading and speaking Catalan, but a negative effect on writing Catalan, that is, younger people are less likely to read and speak Catalan but more likely to write Catalan. The interaction of age and schooling has a negative effect on the probability of knowing Catalan. These negative interaction is shown to be stronger for writing skills than for reading and speaking skills. The coefficient of the Normalization dummy is positive for all groups, greater in 1991 than in 1996 for both genders, greater for men than for women, and greater for reading and speaking than for writing skills. This positive coefficient means that the shape of Catalan knowledge in age is first decreasing and then increasing.

Married individuals, of both genders and for both census year, are in general less likely to know Catalan. The coefficients for the three provinces other than Barcelona are positive and significant and greater for reading and speaking than for writing skills. The effects of the local employment rate are positive for reading and speaking but negative for writing Catalan, while the percentage of services at the Municipal level is in general negative. This result may reflect that both variables may have high values in relatively more urban areas where Catalan is relatively less used than in rural areas.

The percentage of people who write Catalan and the percentage of people born in Catalonia in the *municipi* are indicators of social interactions between individuals at the local level, which play also an important role in explaining Catalan knowledge. The more people know Catalan in the *municipi*, the more likely it is that one individual knows it. A higher percentage of people born in Catalonia in the *municipi* increases the probability that an individual reads and speaks Catalan, but, interestingly, it decreases the probability that an individual writes Catalan.

Not being born in Catalonia has a negative effect on both Catalan skills, especially

---

<sup>9</sup>I also estimated a probit with fixed effects for the *municipis* and an ordered probit with three categories for each proficiency level: basic knowledge, reading and speaking, and writing Catalan. Results are essentially the same as those shown here.

on reading and speaking Catalan. However, having arrived to Catalonia at a young age has a significant positive effect. For individuals who are not born in Catalonia, years since migration (YSM) has an important positive effect, which is increasing from 1991 to 1996 for both reading and speaking and for writing skills. More exposure to the local culture makes language assimilation more likely. The interaction between YSM and having arrived at a young age shows a negative coefficient. Individuals who were born in Andalusia are less likely to know Catalan, especially women for whom the effect is relatively big and strongly significant; for men this effect is relatively small and non-significant. On the contrary, individuals born in areas where Catalan is strongly used, such as Valencia, Balearics and La Franja, are more likely to know Catalan.

## **5.2 Gender Gap and Improvement in Catalan Knowledge**

After performing these estimations, I evaluate two counterfactual experiments: switching years, which allows me to measure to what extent the observed increase in the probability of knowing Catalan is driven by an actual improvement in its knowledge. Given that the sample of 1991 is different than the sample of 1996, one can measure the probability, say, that an individual from the sample of 1991 speaks Catalan in 1996. I perform a similar counterfactual experiment for gender, thus measuring to what extent the gender differences in Catalan knowledge are driven by attributes or by the gender-specific estimated coefficients.

Table 5 reports the predicted probability of knowing Catalan by proficiency level, census year, and gender with the two mentioned experiments. One can see that for men there is no significant increase Catalan knowledge from 1991 to 1996, although in the descriptive statistics one can see an increase of more than 10 percentage points: from 62% to 74% for reading and speaking and from 37% to 46% for writing. This means that this large increase is mainly explained by changing individual attributes in the sample. For women there is a significant increase in Catalan knowledge, how-



ever not as large as a first inspection of the data suggests. By the same type of decomposition, controlling for individual attributes, it is clear that women are more likely to read and speak and to write Catalan than men. This gender difference is greater for writing skills than for reading and speaking skills. In the absence of panel data, these results remarks the need for distinguishing between attributes and estimated coefficients when discussing differences in Catalan knowledge over time or across genders.

### **5.3 Employment by Catalan Knowledge**

The second stage consists of estimating the probability of finding a job conditional on a given Catalan proficiency level using the parameters of the selection equation previously estimated. Tables 6 and 7 show the conditional probit estimations of the probability of being employed for the subsamples characterized by census year, origin, gender and reading-speaking and writing skills.

These estimations use only  $X$  variables as covariates. Schooling and age tend both to have a positive and decreasing effect on employment for most groups. The coefficient of the interaction between these two variables has in general a negative effect on employment. The Normalization dummy, shown to be important in the Catalan knowledge equation, is non-significant for all subsamples.

The effect of being married in the probability of being employed is positive and significant for men, whereas it is smaller and even negative for women. Residents of Lleida, Girona and Tarragona exhibit in general higher employment rates than those who reside in Barcelona. A higher percentage of employed people in the *municipi* has a positive effect on the individual probability of being employed, whereas the percentage of people working in services in the *municipi*, probably capturing urbanization, has a negative effect.

Notice also that the correlation coefficient is generally not zero, showing that the employment equation is not independent from the selection equation. Interest-

ingly, the estimations exhibit negative correlation for individuals who know Catalan and positive correlation for those who do not know it. I also perform simple probit estimations for the employment equation and compute likelihood ratios which are significant for most subsamples, that is, one can reject the null hypothesis that the selection and the employment equation are independent.

## **5.4 Premium**

Table 8 shows the predicted probabilities of being employed by Catalan reading and speaking and writing skills based on the previous estimations. Notice that predictions are very close to the actual values. This table also reports counterfactual probabilities: the probability of being employed for individuals who do not know Catalan had they known it, and the probability of being employed for individuals who know Catalan had they not known it. With the counterfactuals, I compute the premium for knowing Catalan, that is, its contribution to being employed for the different subsamples. I also compute the counterfactual premia resulting of switching attributes by gender and by year. This table reveals that

1. the premium for reading and speaking Catalan is around 3% for men and around 5% for women; the premium for writing is around 1% for men and around 4% for women;
2. for both genders, for both years, and for both proficiency levels, the premium is greater for people who know Catalan than for people who do not know it;
3. for both genders, the difference in the premium between people who know Catalan and people who do not know Catalan is greater in 1996 than in 1991;
4. for both genders and for both years, the premium is greater for speaking and reading than for writing Catalan.
5. for both years and for both proficiency levels, the premium is greater for women than for men;
6. for both genders, except for men who read and speak Catalan and for both proficiency levels, the premium falls from 1991 to 1996.

The first finding shows that the premium is significant and relatively large, although it is smaller than the one computed by a descriptive direct comparison across samples. The second finding indicates that individuals that have a higher economic reward of knowing Catalan do actually know Catalan; those having a lower economic reward, do actually not know it. Note, however, that the economic reward for both groups is positive. The theory of comparative advantage predicts that individuals with positive rewards for knowing Catalan will actually know Catalan, if individuals only care for economic rewards. That means that the economic incentive is not enough to explain individuals' decisions on language knowledge. Additionally, given that the premium is larger for individuals who know Catalan, one can argue that this premium is measured in a transitional phase, when individuals are increasing their level of Catalan knowledge. One would expect that measuring the premium at a later stage of the Normalization process conforms closer to the theory of comparative advantage. The third finding gives evidence in that direction: in 1991 the difference in the premium for reading and speaking Catalan between people who know and who do not know Catalan is around 1% for men and around 2.5% for women; in 1996 the difference in the premium for the sample of 1991 would be 2% for men and around 5% for women.

The fourth finding can be interpreted as evidence for decreasing returns in the level of Catalan knowledge: the premium is higher for a reading and speaking than for a writing proficiency level. The fifth finding means that women, besides having a higher probability of knowing Catalan, have also a larger premium. This result should be taken cautiously because, as mentioned in Section 4, the sample includes only individuals who participate in the labor force. Women's estimated premium may be large because their participation rate is lower than men's. Accounting for participation decisions is beyond the scope of this paper and it is left for further research. The sixth finding means that the premium is decreasing over time, which can be also interpreted as evidence for decreasing returns in Catalan knowledge. As

people are more proficient in Catalan over time, knowing Catalan makes less of a difference for being employed.

## 5.5 Marginal Effects on the Premium

The premium computed above is an average for each census year and gender, but it is obviously different for each individual. A way of accounting for different individual attributes is to compute the marginal effects of each covariate on the premium. A variable's direct effect influences the premium through the employment equation, whereas an indirect effect affects the premium through the selection equation. Only  $X$  variables have both direct and indirect effects.

Table 9 and Table 10 report the direct and indirect marginal effects of each variable on the Catalan premium. As in the previous subsection, the average premium of reading and speaking Catalan is around 3% for men and around 5% for women; the premium for writing is around 1% for men and around 4% for women. Direct effects of age and of schooling on the premium are negative, while indirect effects are positive, being the direct effects predominant. For the Normalization dummy, the direct effect is negative but not significant, while the indirect effect is positive and significant; the total effect of this variable is negative. The marriage variable exhibits a marked difference by gender: for men the direct effect is negative and significant, while the indirect effect is not significant; for women the direct effect is not significant, but the indirect effect is negative and significant. Individuals who live in Lleida, Girona and Tarragona exhibit negative direct effects and positive indirect effects on the premium, being the indirect effect predominant. The employment rate and the percentage of services in employment have both a negative direct effect and a positive indirect effect, being the total effect negative.

The  $Z$  variables, which only enter in the selection equation, have only indirect effects. Individuals who live in *municipis* with a higher percentage of Catalan-writers and Catalan-born individuals have a significant and larger positive premium than

those who do not. Individuals who were not born in Catalonia have a significantly smaller premium than those who were born in Catalonia. Among those individuals who were not born in Catalonia, those who arrived a young age, who have more time in Catalonia, and who were born in Catalan-speaking areas have a higher premium. Individuals born in Andalusia have a smaller premium.

## 5.6 Premium and Effects with Alternative Specifications

This section discusses how sensitive are these results to the  $Z$  variables. A recursive bivariate probit model, such as the one used here, is not identified without exogenous sources of variation, that is, only with  $X$  variables, which determine both employment status and language choice (Maddala 1983). The introduction of  $Z$  variables breaks the ambiguity and pins down the parameters of the language selection equation and the employment equation. Discussions on bivariate probit results are thus focused on the meaning and quality of these variables, for example, the discussion on the effect of Catholic schools on College attendance (Neal 1997, Evans & Schwab 1995, Altonji, Elder & Taber 2002). A matter of concern is that the estimated premium is a spurious result of unmeasured individual characteristics that are correlated with employment. This section follows the discussion of Altonji et al. (2002) who compare estimates of different specifications of a bivariate probit model. This comparison is basically heuristic, given that models of limited dependent variables, which do not have estimated sample errors, do not have established tests of instruments, as it is the case with linear models.

Table 11 presents the premium<sup>10</sup> for all groups by knowledge, proficiency level of Catalan and by exclusion of  $Z$  variables. One can compare the baseline estimation without any exclusion, the estimation with no Municipal variables (percentage of people born in Catalonia and percentage of people who write Catalan in the *municipi*),

---

<sup>10</sup>Variations in the premium come basically from variations in the counterfactual employment probability. The predicted employment probability is close to the actual one and does not vary more than 0.01% across specifications. It is thus enough to report the premium.

and the estimation with no origin variables (born in Catalonia, A9, YSM, A9×YSM, born in Andalusia, Valencia-Balearics, Franja).

The exclusion of Municipal variables means a reduction of the average premium for men, except for writing Catalan in 1996, and a reduction of the average premium for women in 1991 and an increase in 1996. For most cases, it also implies a reduction in the gap between people who know and who do not know Catalan, although it does not change the result that the former have a higher premium than the latter. Notice also that there is a mild increase in the standard deviation of the premium for all groups.

If origin variables are excluded, the average premium for reading and speaking Catalan decreases for men and increases for women, while it increases for both genders for writing Catalan. Variations introduced by this exclusion are substantially larger than those produced by the previous one, not only in the premium by Catalan knowledge, but also in the estimates' precisions.

Table 12 reports the marginal effects for two selected variables, schooling and age, by exclusion restriction. As in the previous table, larger variations in effects and larger standard deviations occur when origin variables are excluded. Note that direct and indirect effects exhibit large variations across specifications, but total effects do not. Understandably, changing the set of  $Z$  variables redistributes total marginal effects between direct and indirect effects. This is clearer for schooling, whose marginal effects exhibit large variations, than for age, whose indirect effects do not vary much. For both proficiency levels the exclusion of Municipal variables reduces slightly the indirect effects of schooling, whereas the exclusion of origin variables produces an important increase in the indirect effects of schooling.

In short, the estimated premium and its marginal effects are shown to be sensitive to exclusion restrictions, in particular, of origin variables, which are the primary source of identification of the premium.

## 6 Conclusions

In this paper I find evidence that the returns to reading, speaking and writing Catalan are positive for men and women for census years 1991 and 1996. Accounting for the selection into knowing Catalan, I find that speaking and reading Catalan increases the probability of being employed by between 3 and 5 percentage points, whereas writing Catalan increases the probability of being employed by between 1 and 4 percentage points. The premium is greater for women than for men, and decreasing in years of schooling and in age. It is positive both for individuals who know Catalan and for those who do not know it, although it is higher for the former than for the latter. This finding conforms partly to the theory of comparative advantage, that is, individuals with higher economic returns to knowing Catalan actually know it. However, it may also indicate that (i) non-economic incentives are very important and (ii) that the process of Normalization, along with the progressive Catalan recovery, was still going on in 1991 and 1996.

These results encourage research on language and labor market behavior, even when labor data are very scarce and the researcher is forced to rely on census data. In this paper returns to language are measured as a difference of employment probabilities, a premium; yet they are found to be significant and show sensible trends. Measurement of these returns in terms of wages, as it is done in other contexts, is left for future research, when employment and wage surveys in Catalonia include questions on language proficiency.

## Appendix

### A.1. Legal Status of Catalan

The following are translations of articles related to language in Catalonia from the Spanish Constitution and from the Autonomy Statute for Catalonia. They are taken from Webber & Strubell i Trueta (1991).

Spanish Constitution (1978). Article 3:

1. Castilian is the official language of the State. All Spaniards have the duty to know it and the right to use it.
2. The other Spanish languages will also be official in their respective Autonomous Communities, in accordance with their Statutes.
3. The richness of Spain's different linguistic modalities is a cultural patrimony which will be the object of special respect and protection.

Autonomy Statute for Catalonia (1979). Article 3:

1. Catalonia's own language is Catalan.
2. The Catalan language is the official one in Catalonia, as is Castilian, official throughout the Spanish State.
3. The Generalitat will guarantee the normal and official use of both languages, will take the necessary measures to ensure adequate knowledge of them and will create the conditions which will allow them to attain full equality with respect to the rights and duties of the citizens of Catalonia.

### A.2. Marginal Effects

I compute the marginal effects as a difference-in-differences of the conditional probabilities of being employed. Let  $X_2 = X$ , then the direct and indirect marginal effects of  $X_k$  variables are

$$\Delta_{kj} = \Delta|_{X_{kj}+\theta_j} - \Delta|_{X_{kj}} = (P_c - P_n)|_{X_{kj}+\theta_j} - (P_c - P_n)|_{X_{kj}},$$

where  $\Delta|_{X_{kj}+\theta_j} = (P_c - P_n)|_{X_{kj}+\theta_j}$  is the premium when the  $j$ -th variable of  $X_k$ ,  $X_{kj}$ , is replaced by  $X_{kj} + \theta_j$ , if it is continuous, and by  $X_{kj} = 1$ , if it is a dummy; and  $\Delta|_{X_{kj}} = (P_c - P_n)|_{X_{kj}}$  is the premium, if  $X_{kj}$  is continuous, and it is the premium when  $X_{kj}$  is replaced by  $X_{kj} = 0$ , if it is a dummy;  $k = \{1, 2\}$ . For  $k = 2$ , effects are direct; for  $k = 1$ , effects are indirect. The continuous variables are schooling, age, YSM, and local-level percentages. For the local-level percentages the increase is one percentage point, that is,  $\theta_j = 0.01$ ; for all the other variables the increase is one,  $\theta_j = 1$ . Given that in the estimation there are covariates which are linked together,



such as square and interaction terms, a particular marginal effect for one of these covariates is determined by several terms and coefficients.

The average marginal effects are then  $\bar{\Delta}_{kj} = N^{-1} \Sigma \Delta_{kj}$ , that is, I compute average marginal effects, not marginal effects at the average of the covariates, as it sometimes appears in the literature. The variances of the average direct marginal effects are

$$Var(\bar{\Delta}_{2j}) = N^{-1} (N^{-1} \Sigma \Delta_{2j}^2 - \bar{\Delta}_{2j}^2) + d'_{cj} Var(\hat{\beta}_c, \hat{\rho}_c) d_{cj} + d'_{nj} Var(\hat{\beta}_n, \hat{\rho}_n) d_{nj},$$

where  $d'_{aj} = \left[ \frac{\partial \bar{P}_a}{\partial \hat{\beta}_{aj}}, \frac{\partial \bar{P}_a}{\partial \hat{\rho}_a} \right] \Big|_{X_{2j} + \theta_j} - \left[ \frac{\partial \bar{P}_a}{\partial \hat{\beta}_{aj}}, \frac{\partial \bar{P}_a}{\partial \hat{\rho}_a} \right] \Big|_{X_{2j}}$ ,  $a = \{c, n\}$ . As said in the main text, these two sets of parameters were computed independently, so there is no correlation between them. The variances of the average indirect marginal effects are

$$Var(\bar{\Delta}_{1j}) = N^{-1} (N^{-1} \Sigma \Delta_{1j}^2 - \bar{\Delta}_{1j}^2) + d'_{1j} Var(\hat{\beta}_1) d_{1j}$$

where  $d_{1j} = \left( \frac{\partial \bar{P}_c}{\partial \hat{\beta}_{1j}} - \frac{\partial \bar{P}_n}{\partial \hat{\beta}_{1j}} \right) \Big|_{X_{1j} + \theta_j} - \left( \frac{\partial \bar{P}_c}{\partial \hat{\beta}_{1j}} - \frac{\partial \bar{P}_n}{\partial \hat{\beta}_{1j}} \right) \Big|_{X_{1j}}$ , that is, it is the derivative of a difference-in-differences of conditional probabilities.

The derivatives of the individual predicted probabilities of employment conditional on Catalan knowledge with respect to each coefficient - see Greene (1996, 1998) - are

$$\begin{aligned} \frac{\partial P_c}{\partial \hat{\beta}_{cj}} &= \frac{\phi(X\hat{\beta}_c) \Phi\left(\frac{X_1\hat{\beta}_1 - \hat{\rho}_c X\hat{\beta}_c}{\sqrt{1-\hat{\rho}_c^2}}\right)}{\Phi(X_1\hat{\beta}_1)} X_{2j}, \\ \frac{\partial P_n}{\partial \hat{\beta}_{nj}} &= \frac{\phi(X\hat{\beta}_n) \Phi\left(\frac{-X_1\hat{\beta}_1 + \hat{\rho}_n X\hat{\beta}_n}{\sqrt{1-\hat{\rho}_n^2}}\right)}{\Phi(-X_1\hat{\beta}_1)} X_{2j}, \\ \frac{\partial P_c}{\partial \hat{\rho}_c} &= \frac{\phi_2(X\hat{\beta}_c, X_1\hat{\beta}_1, \hat{\rho}_c)}{\Phi(X_1\hat{\beta}_1)}, \\ \frac{\partial P_n}{\partial \hat{\rho}_n} &= -\frac{\phi_2(X\hat{\beta}_n, -X_1\hat{\beta}_1, -\hat{\rho}_n)}{\Phi(-X_1\hat{\beta}_1)}, \\ \frac{\partial P_c}{\partial \hat{\beta}_{1j}} &= \left[ \Phi\left(\frac{X\hat{\beta}_c - \hat{\rho}_c X_1\hat{\beta}_1}{\sqrt{1-\hat{\rho}_c^2}}\right) - P_c \right] \frac{\phi(X_1\hat{\beta}_1)}{\Phi(X_1\hat{\beta}_1)} X_{1j}, \text{ and} \\ \frac{\partial P_n}{\partial \hat{\beta}_{1j}} &= \left[ P_n - \Phi\left(\frac{X\hat{\beta}_n - \hat{\rho}_n X_1\hat{\beta}_1}{\sqrt{1-\hat{\rho}_n^2}}\right) \right] \frac{\phi(-X_1\hat{\beta}_1)}{\Phi(-X_1\hat{\beta}_1)} X_{1j}. \end{aligned}$$

The average derivatives required to compute the previous variances are just the average of these expressions over all individuals.

### A.3. Sample Selection

The following table illustrates the importance of the selection criteria in constructing the sample.

	1991	1996
Total sample	250 000	250 000
Only main household members: parents and children	-17 654	-17 903
Only individuals between 16 and 60 years old	-82 297	-81 770
Only Spaniards	-5 740	-4 745
Only if arrival in Catalonia available		-3 788
Only individuals in the labor force	-47 421	-44 809
Only if Catalan language variable available	-25	
Selected sample	96 863	96 985

### A.4. Definition of the variables

The explanation on the construction of each variable is presented below.

Employment.- The census reports a variable called “relationship with the activity.” Somebody is employed if he or she reports to be working; somebody is unemployed, if he or she reports to be not working and looking for his/her first job or having worked before. Individuals who report other options (military service, retired, student, working at home) are not in the sample.

Schooling.- The census reports the maximum level of studies attained by the individual. To each level, I assign the number of years of schooling.

Age.- It is the census year, 1991 or 1996, respectively, minus the year of birth.

Normalization.- If the individual was younger than 12 years old in 1984, this dummy variable takes the value of one and zero otherwise.

Married.- This variable takes the value of one, if the respondent reports to be currently married; it is zero if the respondent reports to be a widow(er), separated, or divorced.

Residence variables.- The census reports the *municipi* and the Province of residence for each individual. With this information I construct dummies for Lleida, Girona and Tarragona.

Origin variables.-The census reports the *municipi* and the province of birth for each individual. With this information I construct dummies for people who are not born in Catalonia, Andalusia, Valencia, Balearics and La Franja.

YSM (Years since Migration).- The census reports the year of arrival to Catalonia. YSM is the census year minus this number. I also construct the dummy indicating if somebody arrived when s/he was no more than 9 years old.

Municipal variables.- I use the residence variable to assign to each individual the corresponding information of the *municipi*.

## References

- Altonji, J. G., Elder, T. E. & Taber, C. R. (2002), An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools. NBER WP. No. 9358.
- Angrist, J. D. & Lavy, V. (1997), 'The effect of a change in language of instruction on the returns to schooling in Morocco', *Journal of Labor Economics* **15**, 48–76.
- Borjas, G. (1987), 'Self-Selection and the Earnings of Immigrants', *American Economic Review* **77**(27), 21–37.
- Charette, M. & Meng, R. (1994), 'Explaining language proficiency. objective versus self-assessed measures of literacy', *Economic Letters* **44**, 313–321.
- Chiswick, B. R. & Miller, P. W. (1995), 'The endogeneity between language and earnings: International analyses', *Journal of Labor Economics* **13**, 246–88.
- Dustmann, C. (1994), Speaking fluency, writing fluency and earnings of migrants. CEPR WP. No. 905.
- Evans, W. N. & Schwab, R. M. (1995), 'Finishing High School and Starting College. do Catholic Schools make a Difference?', *Quarterly Journal of Economics* **110**, 947–974.
- Greene, W. (1996), Marginal effects in the bivariate probit model. Working paper number 96-11. Department of Economics, Stern School of Business, New York University.
- Greene, W. (1998), 'Gender Economics Courses in Liberal Arts Colleges: Further Results', *Journal of Economic Education* **29**(4), 291–300.
- Kossoudji, S. (1988), 'English language ability and the labor markets opportunities of Hispanic and East Asian immigrant men', *Journal of Labor Economics* **6**(2), 205–228.
- Lang, K. (1986), 'A Language Theory of Discrimination', *Quarterly Journal of Economics* **101**(2), 363–382.
- Lazear, E. (1999), 'Culture and Language', *Journal of Political Economy* **107**(6), S96–S126.
- Maddala, G. S. (1983), *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge.
- McManus, W. S. (1985), 'Labor Market Costs of Language Disparity: An Interpretation of Hispanic Earnings Differences', *American Economic Review* **75**(4), 818–827.
- Neal, D. (1997), 'The Effects of Catholic Secondary Schooling on Educational Attainment', *Journal of Labor Economics* **15**, 98–123.
- Reixach, M. (1997), *El coneixement del català*, Generalitat de Catalunya. Departament de Cultura, Barcelona.
- Rivera-Batiz, F. (1990), 'English language proficiency and the economic progress of immigrants', *Economic Letters* **34**, 295–300.
- Robinson, C. (1988), Language Choice: the distribution of language skills and earnings in a dual-language economy, in E. R. Ehrenberg, ed., 'Research in Labor Economics', JAI Press Inc., London, pp. 53–90.

- Strubell i Trueta, M. (1981), *Llengua i Població a Catalunya*, Edicions de la Magrana, Barcelona.
- Tudela, J. (1986), *El futur del català. Una radiografia sociolingüística*, Publicacions El Món S.A., Barcelona.
- Vallverdú, F. (1970), *Dues llengües: dues funcions? Per una històrica lingüística de la Catalunya contemporània.*, Edicions 62, Barcelona.
- Vallverdú, F. (1990), *L'ús del català: un futur controvertit*, Edicions 62, Barcelona.
- Webber, J. & Strubell i Trueta, M. (1991), *The Catalan Language. Progress towards Normalisation*, The Anglo-Catalan Society, Exeter.
- Willis, R. & Rosen, S. (1979), 'Education and self-selection', *Journal of Political Economy* **87**(5), S7–S36.
- Woolard, K. A. (1989), *Double Talk. Bilingualism and the politics of ethnicity in Catalonia*, Stanford University Press cop., Stanford, California.

Table 1: Historical Facts

---

---

- 1716. Decree of Nova Planta. Castilian comes to be the only official language.
- 1858. Law of Public Instruction. Castilian as compulsory language of instruction.
- 1860. Catalan is still the habitual language in the rural and urban areas.
- 1923. Primo de Rivera: Suppression of Catalan.
- 1931. Spanish Republic: Abolition of dictates against Catalan.  
Co-officiality of Catalan and Castilian. However, Catalan is the official language of the Catalan government, schools and media.
- Post Civil War years. Linguistic Repression.
  - Institutionalization of diglossia: dialectization of Catalan in the consciousness of its speakers.
  - Abolition of all public uses of Catalan.
  - Closing of schools and institutions.
  - Penalties, fines and loss of positions for Catalan speakers.
  - No mass media in Catalan (TV, radio and journals).
- Post Civil War years. Massive immigration of non-Catalan speakers.
  - In 1970, 40% of the population of Catalonia was non-native born.
  - In 1975, Catalan speakers represented 31.0% of high school dropouts; Castilian speakers, 68.8%.
  - In the same year 62.30% of the people with university studies were Catalan speakers; Castilian speakers, 35.7%.

---

Source: Woolard (1989)

Table 2: Law of 1983

---

---

Catalan and Castilian are both official in Catalonia.

Declared Goals:

1. to have Catalan be the most widely spoken language in private and public organisms, and in all those public usages;
2. respecting the rights of Castilian speakers.

The first phase of normalization focuses in 3 areas of action:

**(i) Official Use. Catalan government.**

- Catalan is the language of the Generalitat (together with Castilian). 2 year transition period.
- Official documents must be published both in Catalan and Castilian.
- Administrative procedures in Catalan are valid, as are in Castilian.

**(ii) Pre-college education.**

- Primary education in the usual language of the child.
- Catalan must be compulsorily taught in all pre-university studies.
- All teachers should know both official languages.

**(iii) Mass media.**

- Incentives for using Catalan in mass media (TV, radio).
- Subsidies to regular publications written totally or partially in Catalan.
- Incentives given to theater and books in Catalan.

---

Table 3: Summary Statistics

Census Year	1991		1996	
Gender	Men	Wom.	Men	Wom.
Catalan Knowledge				
% Basic level	37.7	29.5	26.3	18.5
% Speak&Read	25.8	22.9	27.5	22.4
% Write	36.5	47.6	46.2	59.1
Employment				
%Employed	89.9	78.5	85.1	74.5
%Employed by Catalan Knowledge				
Basic level	87.8	72.3	81.9	68.0
Speak & Read	92.0	78.5	87.3	73.3
Write	91.0	82.5	85.6	77.0
%Employed for not born in Catalonia	91.0	77.0	86.3	74.2
X variables				
Schooling (S)	7.7	8.3	8.6	9.2
Age	38.0	34.4	38.2	35.5
% Normalization (Norm)	3.9	5.0	13.0	15.9
% Married	66.6	58.2	62.8	57.4
% Lleida	5.7	5.0	5.9	5.2
% Girona	8.4	8.4	8.5	8.7
% Tarragona	8.9	7.7	9.4	8.5
% employed over population in <i>municipi</i> (PartM)	37.2	37.4	36.3	36.3
% work in Services in <i>municipi</i> (ServM)	51.7	52.5	57.4	58.2
Z variables				
% writes in <i>municipi</i> (WriteM)	39.7	40.1	45.6	45.9
% Catalan-born in <i>municipi</i> (CatbornM)	67.1	67.2	68.2	68.1
% not born in Catalonia (NbornC)	37.6	30.2	31.0	24.9
% arrived age $\leq 9$ (A9)	11.0	11.1	10.0	9.6
Years since migration (YSM) if not born in Cat	25.0	23.6	28.4	26.8
% born in Andalusia (And)	24.1	18.3	19.5	14.4
% born in Valencia (PV)	1.1	1.1	1.0	0.9
% born in Balearics	0.1	0.1	0.1	0.2
% born in Franja (Franja)	0.5	0.4	0.4	0.4
Nobs	60446	36417	57785	39200

Table 4: Reduced-form Catalan Knowledge Equation. Standard errors in small fonts

Catalan Skill Census Year Gender	Reading and Speaking				Writing			
	1991		1996		1991		1996	
	Men	Wom.	Men	Wom.	Men	Wom.	Men	Wom.
Const.	-3.98	-4.16	-4.17	-4.08	-1.56	-1.49	-0.73	-1.30
	0.17	0.24	0.22	0.30	0.16	0.21	0.19	0.24
Schooling ( $S$ )	3.54	3.85	4.39	4.81	3.75	4.13	4.42	4.84
$\times 10^{-1}$	0.10	0.15	0.15	0.20	0.10	0.13	0.13	0.18
$S^2$	-7.64	-9.02	-9.60	-11.36	-6.75	-8.76	-7.01	-9.70
$\times 10^{-3}$	0.34	0.49	0.57	0.77	0.31	0.44	0.55	0.70
Age	6.19	6.69	2.84	3.03	-3.68	-4.33	-10.26	-9.35
$\times 10^{-2}$	0.53	0.73	0.70	0.96	0.50	0.65	0.62	0.80
Age <sup>2</sup>	-3.96	-4.60	0.05	-0.34	6.91	7.18	13.99	12.29
$\times 10^{-4}$	0.59	0.83	0.78	1.09	0.57	0.76	0.71	0.92
Age $\times S$	-2.12	-2.00	-2.66	-2.56	-2.73	-2.08	-3.56	-2.72
$\times 10^{-1}$	0.18	0.27	0.23	0.32	0.17	0.24	0.20	0.26
Norm	3.48	3.00	3.10	2.96	1.81	1.66	0.03	1.05
$\times 10^{-1}$	0.38	0.45	0.36	0.46	0.35	0.41	0.30	0.37
Married	-8.76	-120.37	13.86	-83.05	-196.58	-202.38	-120.14	-182.60
$\times 10^{-3}$	17.09	19.87	17.98	21.12	16.30	18.17	15.79	17.56
Lleida	2.63	1.69	1.15	0.91	1.12	1.65	1.39	1.21
$\times 10^{-1}$	0.33	0.47	0.37	0.54	0.29	0.39	0.29	0.39
Girona	2.35	2.42	2.39	1.63	1.68	1.98	2.68	1.46
$\times 10^{-1}$	0.26	0.36	0.31	0.42	0.23	0.31	0.25	0.31
Tarragona	1.62	2.10	2.16	1.48	1.28	1.34	1.83	1.45
$\times 10^{-1}$	0.24	0.34	0.25	0.35	0.23	0.31	0.22	0.28
PartM	7.43	7.85	7.67	5.42	-7.75	-12.64	-9.74	-5.08
$\times 10^{-1}$	3.68	5.13	3.46	4.69	3.35	4.55	2.92	3.79
ServM	-7.75	-7.31	-3.78	-4.55	-9.76	-10.36	-9.69	-7.86
$\times 10^{-1}$	0.57	0.76	0.68	0.89	0.54	0.69	0.59	0.73
WriteM	2.71	2.41	2.74	2.30	3.90	3.64	4.01	3.48
	0.11	0.16	0.14	0.21	0.10	0.14	0.12	0.16
CatbornM	8.55	11.37	12.43	15.56	-8.25	-4.72	-5.46	-1.25
$\times 10^{-1}$	1.11	1.64	1.33	1.96	1.03	1.43	1.10	1.49
NbornC	-1.86	-1.71	-1.99	-1.77	-1.40	-1.34	-1.57	-1.39
	0.04	0.05	0.04	0.05	0.05	0.06	0.05	0.06
A9	8.14	8.41	9.83	9.93	4.49	5.39	8.36	8.37
$\times 10^{-1}$	0.65	0.84	0.79	1.03	0.75	0.93	0.86	1.04
YSM	1.56	1.46	1.94	1.60	0.88	1.03	1.40	1.32
$\times 10^{-2}$	0.13	0.19	0.14	0.19	0.18	0.24	0.18	0.22
A9 $\times$ YSM	-7.35	-9.28	-11.23	-13.91	-2.47	-6.49	-11.77	-14.04
$\times 10^{-3}$	2.13	2.99	2.34	3.19	2.58	3.49	2.66	3.40
Andalusia	-1.45	-10.58	-6.10	-8.16	-1.28	-5.58	-3.79	-9.69
$\times 10^{-2}$	2.11	2.93	2.30	3.08	2.80	3.57	2.95	3.53
PV-Balearics	2.31	1.59	1.87	3.39	2.86	0.62	2.49	1.54
$\times 10^{-1}$	0.55	0.75	0.64	0.87	0.61	0.79	0.64	0.79
Franja	3.04	5.83	2.44	3.80	3.02	3.46	1.54	0.74
$\times 10^{-1}$	0.85	1.27	1.02	1.41	0.91	1.17	1.03	1.21
$-10\overline{\ln \mathcal{L}}$	4.43	4.10	3.77	3.11	4.69	4.86	4.78	4.56

Table 5: Predicted Probability of knowing Catalan (in %) Standard errors in small fonts

Catalan Skill Census Year Gender	Reading and Speaking				Writing			
	1991		1996		1991		1996	
	Men	Women	Men	Women	Men	Women	Men	Women
Actual	62.31	70.54	73.72	81.54	36.51	47.63	46.27	59.10
	48.46	45.59	44.02	38.79	48.14	49.94	49.86	49.17
Predicted	62.36	70.57	73.79	81.59	36.57	47.70	46.33	59.16
	0.67	0.15	0.65	0.12	0.40	0.16	0.42	0.15
Counterfactuals								
Switch year	62.30	72.52	73.19	79.82	36.36	49.87	47.64	58.29
	0.67	0.15	0.65	0.12	0.40	0.17	0.43	0.14
Diff. year	-0.06	1.95	-0.60	-1.77	-0.21	2.17	1.31	-0.87
	0.95	0.21	0.92	0.17	0.57	0.23	0.60	0.21
Switch gender	65.68	67.14	77.50	78.10	41.05	42.95	52.48	53.03
	0.70	0.15	0.69	0.13	0.45	0.15	0.48	0.15
Diff. gender	3.32	-3.44	3.71	-3.49	4.48	-4.74	6.15	-6.12
	0.97	0.21	0.95	0.17	0.60	0.21	0.64	0.22



Table 6. Employment Equation by Reading and Speaking Skills  
Standard errors in small fonts

Census Year	1991				1996			
Gender	Men		Women		Men		Women	
Knowledge	Knw.	D.knw	Knw.	D.knw	Knw.	D.knw	Knw.	D.knw
Const.	-1.98	-3.07	-2.44	-3.33	-2.66	-4.28	-1.89	-2.58
	0.26	0.29	0.28	0.38	0.23	0.38	0.24	0.52
Schooling ( $S$ )	1.25	13.76	6.09	0.28	4.07	20.17	3.38	7.11
$\times 10^{-2}$	1.72	1.93	1.92	2.97	1.87	2.78	2.03	4.03
$S^2$	-5.13	-28.99	-5.46	28.24	-13.95	-25.04	2.08	18.55
$\times 10^{-4}$	5.82	7.68	6.18	10.79	7.44	10.21	7.71	13.29
Age	1.06	1.04	0.63	0.63	1.16	1.27	0.64	0.81
$\times 10^{-1}$	0.08	0.08	0.08	0.12	0.08	0.13	0.09	0.18
Age <sup>2</sup>	-1.37	-1.20	-0.60	-0.51	-1.51	-1.34	-0.63	-0.60
$\times 10^{-3}$	0.09	0.09	0.10	0.13	0.09	0.14	0.10	0.19
Age $\times S$	7.64	-5.04	4.32	-0.99	9.12	-20.89	7.27	-8.56
$\times 10^{-4}$	2.65	3.76	3.21	5.31	2.35	4.65	2.70	6.73
Norm	11.72	76.92	-36.93	123.81	-73.06	91.73	-33.51	124.71
$\times 10^{-3}$	45.39	63.34	44.33	77.34	34.32	69.48	33.66	89.42
Married	5.14	5.04	-0.20	-0.73	4.71	4.76	0.44	0.35
$\times 10^{-1}$	0.26	0.29	0.23	0.32	0.21	0.31	0.19	0.37
Lleida	2.91	2.45	2.96	2.81	3.26	2.54	2.13	1.82
$\times 10^{-1}$	0.44	0.73	0.44	0.88	0.39	0.83	0.37	1.09
Girona	2.56	0.77	1.24	-0.78	1.93	1.55	2.36	2.63
$\times 10^{-1}$	0.39	0.55	0.36	0.62	0.33	0.67	0.32	0.82
Tarragona	1.82	2.00	1.89	1.44	1.66	0.95	1.48	2.32
$\times 10^{-1}$	0.35	0.47	0.35	0.58	0.28	0.50	0.29	0.66
PartM	3.41	4.32	4.14	6.37	3.96	4.73	2.40	2.15
	0.49	0.66	0.48	0.75	0.36	0.60	0.35	0.75
ServM	-7.00	-4.56	-3.87	-1.48	-8.29	-3.09	-4.28	-5.02
$\times 10^{-1}$	0.84	1.13	0.81	1.30	0.73	1.27	0.72	1.56
$\rho$	-0.05	0.16	-0.05	0.01	-0.08	0.13	-0.19	0.01
	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.04
LRT ( $\rho = 0$ )	2.03	27.14	1.66	0.07	6.44	15.79	24.32	0.11
$-10\ln \mathcal{L}$	2.72	3.38	4.58	5.66	3.57	3.57	5.17	6.00
Nobs	37660	22786	25687	10730	42597	15188	31965	7235

Table 7. Employment Equation by Writing Skills  
Standard errors in small fonts

Census Year	1991				1996			
Gender	Men		Women		Men		Women	
Knowledge	Knw.	D.knw	Knw.	D.knw	Knw.	D.knw	Knw.	D.knw
Const.	-1.75	-2.90	-1.94	-3.49	-2.77	-3.76	-1.93	-2.81
	0.33	0.24	0.34	0.28	0.29	0.27	0.28	0.35
Schooling ( $S$ )	-2.85	11.70	2.96	2.86	-2.69	17.35	-0.52	8.15
$\times 10^{-2}$	2.26	1.71	2.50	2.24	2.44	2.07	2.54	2.86
$S^2$	-8.43	-19.24	-6.62	24.55	-5.83	-17.08	4.06	10.28
$\times 10^{-4}$	7.39	6.07	7.99	7.63	9.75	7.22	9.84	9.00
Age	1.03	1.14	0.69	0.78	1.45	1.25	0.94	0.81
$\times 10^{-1}$	0.10	0.07	0.11	0.09	0.10	0.09	0.11	0.12
Age <sup>2</sup>	-1.45	-1.31	-0.70	-0.68	-1.91	-1.34	-1.07	-0.61
$\times 10^{-3}$	0.12	0.08	0.13	0.10	0.12	0.10	0.13	0.13
Age $\times S$	17.87	-5.78	9.97	-4.91	17.58	-18.14	12.58	-9.06
$\times 10^{-4}$	3.49	2.88	4.09	3.89	3.05	3.34	3.23	4.60
Norm	-44.70	110.29	-84.31	158.52	-51.21	73.65	-22.40	133.67
$\times 10^{-3}$	52.04	54.38	50.33	63.97	39.30	51.09	37.22	62.83
Married	5.36	5.07	-0.03	-0.41	4.44	5.07	0.70	0.38
$\times 10^{-1}$	0.36	0.24	0.29	0.25	0.28	0.23	0.23	0.25
Lleida	1.72	3.85	2.83	2.96	2.68	3.58	1.58	2.83
$\times 10^{-1}$	0.54	0.50	0.55	0.57	0.47	0.52	0.43	0.62
Girona	2.36	1.73	1.06	0.12	1.52	1.94	1.97	2.72
$\times 10^{-1}$	0.49	0.42	0.43	0.44	0.41	0.45	0.38	0.50
Tarragona	1.71	2.10	1.46	1.99	1.22	1.67	1.01	2.25
$\times 10^{-1}$	0.46	0.35	0.44	0.41	0.35	0.34	0.34	0.42
PartM	3.81	3.64	3.26	5.93	3.95	3.97	2.03	2.33
	0.63	0.50	0.59	0.56	0.45	0.43	0.41	0.50
ServM	-6.86	-5.68	-3.30	-1.85	-6.85	-6.28	-4.59	-3.07
$\times 10^{-1}$	1.12	0.87	1.04	0.96	0.94	0.88	0.87	1.02
$\rho$	-0.11	0.10	-0.14	-0.01	-0.21	-0.01	-0.22	-0.05
	0.05	0.04	0.05	0.04	0.04	0.04	0.05	0.04
LRT ( $\rho = 0$ )	5.56	5.67	8.22	0.02	23.94	0.12	21.10	1.55
$-10\ln \mathcal{L}$	2.82	3.05	4.35	5.38	3.59	3.59	5.01	5.75
Nobs	22073	38373	17348	19069	26725	31060	23166	16034

Table 8: Probability of employment and Premium (in %) by Catalan skills  
Standard errors in small fonts

Census Year	1991				1996			
Gender	Men		Women		Men		Women	
Knowledge	Knw.	D.knw	Knw.	D.knw	Knw.	D.knw	Knw.	D.knw
READING AND SPEAKING								
Actual	91.05	87.97	81.16	72.28	86.21	81.91	75.99	68.04
	28.55	32.54	39.11	44.76	34.48	38.49	42.72	46.63
Pred: Know	91.05	90.00	81.16	75.76	86.21	83.70	75.99	69.05
	0.35	0.31	0.86	0.86	0.68	0.24	1.50	0.33
Pred: DKnow	88.09	87.97	75.27	72.29	82.25	81.91	70.93	68.05
	0.35	0.24	0.70	0.36	0.45	0.21	0.71	0.24
<b>Premium</b>	2.96	2.03	5.89	3.48	3.96	1.78	5.06	1.01
	0.50	0.31	1.10	0.51	0.81	0.30	1.65	0.38
Counterfactual: Switch year								
Pred: Know	87.39	84.13	75.04	66.10	90.24	89.59	81.17	76.99
	0.41	0.20	0.98	0.26	0.57	0.32	1.33	0.49
Pred: DKnow	83.29	82.18	69.56	65.39	87.05	87.55	75.15	73.10
	0.43	0.23	0.76	0.27	0.38	0.30	0.66	0.38
<b>Premium</b>	4.10	1.95	5.48	0.71	3.20	2.05	6.02	3.89
	0.62	0.42	1.26	0.58	0.66	0.23	1.47	0.34
Counterfactual: Switch gender								
Pred: Know	82.17	77.08	89.98	89.19	76.60	70.05	85.78	83.44
	0.52	0.19	0.63	0.51	0.92	0.17	1.11	0.61
Pred: DKnow	76.69	74.17	87.25	86.96	71.64	69.58	82.46	81.54
	0.54	0.26	0.51	0.52	0.59	0.21	0.55	0.56
<b>Premium</b>	5.47	2.92	2.73	2.23	4.96	0.47	3.32	1.90
	0.80	0.54	0.72	0.30	1.13	0.42	1.20	0.27
WRITING								
Actual	90.40	89.59	82.45	74.99	85.57	84.66	77.02	70.91
	29.46	30.54	38.04	43.31	35.14	36.04	42.07	45.42
Pred: Know	90.40	90.69	82.45	78.75	85.58	84.99	77.02	73.40
	0.25	0.36	0.57	0.58	0.45	0.40	0.98	0.55
Pred: DKnow	88.56	89.59	76.89	75.00	85.08	84.65	72.98	70.91
	0.32	0.64	0.66	0.96	0.41	0.60	0.76	0.74
<b>Premium</b>	1.84	1.09	5.56	3.75	0.59	0.43	4.04	2.48
	0.40	0.73	0.87	1.11	0.60	0.73	1.24	0.91
Counterfactual: Switch year								
Pred: Know	86.05	84.33	75.67	69.88	89.68	90.53	82.77	80.97
	0.25	0.48	0.62	0.65	0.41	0.35	0.87	0.53
Pred: DKnow	85.49	84.76	71.03	68.07	88.10	89.43	77.42	75.74
	0.37	0.83	0.73	1.05	0.36	0.49	0.70	0.71
<b>Premium</b>	0.55	-0.43	4.64	1.81	1.58	1.09	5.36	5.23
	0.50	0.98	1.01	1.28	0.49	0.55	1.08	0.81
Counterfactual: Switch gender								
Pred: Know	83.37	79.89	89.56	90.03	77.28	74.00	85.32	85.08
	0.26	0.54	0.49	0.48	0.55	0.52	0.79	0.52
Pred: DKnow	77.77	76.56	87.98	88.61	72.94	72.35	85.58	84.26
	0.47	1.12	0.51	0.65	0.55	0.85	0.57	0.62
<b>Premium</b>	5.59	3.33	1.59	1.42	4.34	1.65	-0.27	0.81
	0.65	1.29	0.57	0.65	0.84	1.05	0.89	0.62

Table 9: Direct and Indirect Marginal Effects on Premium for Reading and Speaking Catalan  
Standard errors in small fonts

	Men 1991			Women 1991			Men 1996			Women 1996		
	Dir	Ind	Tot	Dir	Ind	Tot	Dir	Ind	Tot	Dir	Ind	Tot
Premium (%)			2.61			5.17			3.39			4.31
			0.81			1.61			1.10			2.02
School	-0.86	0.38	-0.48	0.31	0.14	0.45	-0.90	0.52	-0.39	-0.57	0.43	-0.14
	0.49	0.09	0.49	0.72	0.03	0.72	0.70	0.15	0.70	0.99	0.11	0.99
Age	-0.05	0.04	-0.01	-0.18	0.02	-0.16	-0.01	0.02	0.01	-0.21	0.02	-0.19
	0.04	0.49	0.49	0.08	0.22	0.22	0.06	0.79	0.79	0.12	0.86	0.86
Norm	-1.19	0.86	-0.33	-4.64	0.22	-4.42	-3.66	0.90	-2.76	-4.94	0.61	-4.32
	1.23	0.10	1.23	2.49	0.03	2.49	1.70	0.11	1.70	2.93	0.09	2.93
Married	-1.62	-0.02	-1.64	1.66	-0.09	1.57	-1.98	0.04	-1.94	0.19	-0.18	0.01
	0.40	0.04	0.40	1.13	0.02	1.13	0.57	0.05	0.57	1.24	0.05	1.24
Lleida	-0.23	0.64	0.41	-0.75	0.13	-0.62	0.25	0.33	0.58	0.40	0.19	0.59
	1.02	0.08	1.02	2.18	0.03	2.18	1.45	0.11	1.45	3.05	0.11	3.05
Girona	2.05	0.57	2.62	5.62	0.18	5.80	0.15	0.69	0.84	-1.34	0.34	-1.00
	0.91	0.07	0.91	1.90	0.03	1.90	1.35	0.09	1.35	2.32	0.08	2.32
Tarragona	-0.83	0.39	-0.44	0.54	0.16	0.70	0.99	0.63	1.62	-2.84	0.31	-2.53
	0.81	0.06	0.81	1.68	0.03	1.68	1.09	0.07	1.09	2.00	0.07	2.00
PartM	-0.27	0.02	-0.25	-0.81	0.01	-0.80	-0.32	0.02	-0.29	0.03	0.01	0.04
	0.13	0.01	0.13	0.24	0.00	0.24	0.17	0.01	0.17	0.26	0.01	0.26
ServM	-0.02	-0.02	-0.04	-0.06	-0.01	-0.06	-0.10	-0.01	-0.11	0.03	-0.01	0.02
	0.05	0.00	0.05	0.05	0.00	0.05	0.05	0.00	0.05	0.06	0.00	0.06
WriteM		0.06	0.06		0.02	0.02		0.08	0.08		0.05	0.05
		0.00	0.00		0.00	0.00		0.00	0.00		0.01	0.01
CbornM		0.02	0.02		0.01	0.01		0.04	0.04		0.03	0.03
		0.00	0.00		0.00	0.00		0.00	0.00		0.01	0.01
NbornC		-4.15	-4.15		-1.51	-1.51		-5.68	-5.68		-5.06	-5.06
		0.04	0.04		0.06	0.06		0.10	0.10		0.24	0.24
A9		1.97	1.97		0.51	0.51		2.68	2.68		1.36	1.36
		0.23	0.23		0.27	0.27		0.51	0.51		1.26	1.26
YSM		0.04	0.04		0.01	0.01		0.05	0.05		0.03	0.03
		0.00	0.00		0.00	0.00		0.00	0.00		0.00	0.00
Andalusia		-0.03	-0.03		-0.08	-0.08		-0.18	-0.18		-0.18	-0.18
		0.05	0.05		0.02	0.02		0.07	0.07		0.07	0.07
PV-Balea		0.56	0.56		0.12	0.12		0.55	0.55		0.67	0.67
		0.14	0.14		0.06	0.06		0.19	0.19		0.16	0.16
Franja		0.75	0.75		0.41	0.41		0.72	0.72		0.74	0.74
		0.22	0.22		0.08	0.08		0.31	0.31		0.24	0.24

Table 10: Direct and Indirect Marginal Effects on Premium for Writing Catalan  
Standard errors in small fonts

	Men 1991			Women 1991			Men 1996			Women 1996		
	Dir	Ind.	Tot	Dir	Ind.	Tot	Dir	Ind.	Tot	Dir	Ind.	Tot
Premium (%)			1.36			4.61			0.50			3.40
			1.13			1.99			1.32			2.16
School	-0.75	0.34	-0.42	-0.23	0.45	0.22	-1.05	0.47	-0.58	-0.82	0.53	-0.28
	0.50	0.06	0.50	0.60	0.07	0.60	0.58	0.10	0.58	0.68	0.13	0.68
Age	-0.11	-0.01	-0.12	-0.07	-0.02	-0.09	0.10	-0.04	0.06	-0.07	-0.04	-0.10
	0.04	0.39	0.39	0.08	0.59	0.59	0.06	0.72	0.72	0.10	0.89	0.89
Norm	-2.46	0.35	-2.11	-6.71	0.36	-6.36	-2.61	0.01	-2.60	-4.75	0.25	-4.49
	1.19	0.07	1.19	2.24	0.09	2.24	1.36	0.08	1.36	2.19	0.08	2.19
Married	-0.32	-0.38	-0.70	1.14	-0.45	0.69	-1.81	-0.30	-2.11	0.89	-0.44	0.45
	0.46	0.03	0.46	1.08	0.04	1.08	0.53	0.04	0.53	1.02	0.05	1.02
Lleida	-2.90	0.22	-2.68	-1.32	0.35	-0.96	-1.57	0.35	-1.22	-3.78	0.28	-3.50
	0.95	0.06	0.95	1.80	0.08	1.80	1.15	0.07	1.15	2.03	0.09	2.03
Girona	0.51	0.33	0.84	2.35	0.42	2.77	-0.86	0.64	-0.22	-2.49	0.34	-2.15
	0.88	0.05	0.88	1.65	0.06	1.65	1.11	0.06	1.11	1.69	0.07	1.69
Tarrago	-0.80	0.25	-0.55	-1.91	0.29	-1.62	-0.91	0.44	-0.47	-3.78	0.34	-3.44
	0.81	0.04	0.81	1.52	0.06	1.52	0.93	0.05	0.93	1.53	0.06	1.53
PartM	-0.03	-0.02	-0.04	-0.87	-0.03	-0.90	-0.03	-0.02	-0.05	-0.13	-0.01	-0.14
	0.14	0.01	0.14	0.24	0.01	0.24	0.15	0.01	0.15	0.20	0.01	0.20
ServM	-0.01	-0.02	-0.03	-0.03	-0.02	-0.05	-0.01	-0.02	-0.03	-0.04	-0.02	-0.06
	0.05	0.00	0.05	0.05	0.00	0.05	0.04	0.00	0.04	0.04	0.00	0.04
WriteM		0.08	0.08		0.08	0.08		0.10	0.10		0.08	0.08
		0.00	0.00		0.00	0.00		0.01	0.01		0.01	0.01
CbornM		-0.02	-0.02		-0.01	-0.01		-0.01	-0.01		0.00	0.00
		0.00	0.00		0.00	0.00		0.00	0.00		0.00	0.00
NbornC		-2.73	-2.73		-3.34	-3.34		-4.66	-4.66		-4.46	-4.46
		0.09	0.09		0.19	0.19		0.23	0.23		0.29	0.29
A9		0.83	0.83		0.94	0.94		1.39	1.39		1.08	1.08
		0.37	0.37		0.91	0.91		1.06	1.06		1.65	1.65
YSM		0.02	0.02		0.02	0.02		0.03	0.03		0.03	0.03
		0.00	0.00		0.00	0.00		0.00	0.00		0.00	0.00
Andalusia		-0.03	-0.03		-0.12	-0.12		-0.10	-0.10		-0.24	-0.24
		0.05	0.05		0.08	0.08		0.08	0.08		0.09	0.09
PV-Balearics		0.56	0.56		0.13	0.13		0.59	0.59		0.36	0.36
		0.12	0.12		0.17	0.17		0.14	0.14		0.17	0.17
Franja		0.59	0.59		0.71	0.71		0.41	0.41		0.18	0.18
		0.18	0.18		0.22	0.22		0.24	0.24		0.28	0.28

Table 11: Premium (in %) computed excluding municipal and origin variables  
Standard errors in small fonts

Group	Men 1991			Women 1991			Men 1996			Women 1996		
Knowledge	Kn.	D.Kn	Avr	Kn	D.Kn	Avr	Kn	D.Kn	Avr	Kn	D.Kn	Avr
READING AND SPEAKING												
Baseline	2.96	2.03	2.61	5.89	3.48	5.17	3.96	1.78	3.39	5.06	1.01	4.31
	0.50	0.31	0.81	1.10	0.51	1.61	0.81	0.30	1.10	1.65	0.38	2.02
No <i>municipi</i>	2.59	2.00	2.37	5.55	3.55	4.96	3.72	2.02	3.28	5.02	1.79	4.42
	0.52	0.34	0.86	1.14	0.52	1.66	0.85	0.31	1.15	1.73	0.40	2.11
No origin	2.19	2.24	2.21	6.13	3.79	5.44	2.73	2.06	2.55	4.80	2.41	4.36
	0.87	0.57	1.44	2.19	0.96	3.16	1.63	0.57	2.21	3.60	0.76	4.36
WRITING												
Baseline	1.84	1.09	1.36	5.56	3.75	4.61	0.59	0.43	0.50	4.04	2.48	3.40
	0.40	0.73	1.13	0.87	1.11	1.99	0.60	0.73	1.32	1.24	0.91	2.16
No <i>municipi</i>	1.61	1.07	1.27	5.13	3.96	4.52	0.62	0.70	0.67	4.11	2.99	3.65
	0.47	0.86	1.33	0.99	1.30	2.30	0.67	0.82	1.49	1.36	1.02	2.39
No origin	1.95	1.60	1.73	6.15	4.89	5.49	0.49	1.60	1.08	4.15	3.92	4.06
	0.55	0.98	1.52	1.33	1.68	3.02	0.95	1.06	2.01	2.05	1.35	3.41

Table 12: Direct and Indirect Marginal Effects of Schooling and Age on Premium (in %) computed excluding municipal and origin variables. Standard errors in small fonts

	Men 1991			Women 1991			Men 1996			Women 1996		
	Dir	Indir	Tot	Dir	Indir	Tot	Dir	Indi	Tot	Dir	Indi	Tot
READING AND SPEAKING												
Schooling												
Baseline	-0.86	0.38	-0.48	0.31	0.14	0.45	-0.90	0.52	-0.39	-0.57	0.43	-0.14
	0.49	0.09	0.49	0.72	0.03	0.72	0.70	0.15	0.70	0.99	0.11	0.99
No muni	-0.82	0.35	-0.47	0.41	0.04	0.45	-0.88	0.50	-0.38	-0.51	0.35	-0.15
	0.49	0.08	0.49	0.71	0.02	0.71	0.71	0.14	0.71	1.02	0.09	1.02
No origin	-0.96	0.52	-0.43	-0.14	0.61	0.47	-0.91	0.55	-0.37	-0.76	0.65	-0.11
	0.51	0.10	0.51	0.73	0.14	0.73	0.60	0.10	0.60	0.91	0.17	0.91
Age												
Baseline	-0.05	0.04	-0.01	-0.18	0.02	-0.16	-0.01	0.02	0.01	-0.21	0.02	-0.19
	0.04	0.49	0.49	0.08	0.22	0.22	0.06	0.79	0.79	0.12	0.86	0.86
No muni	-0.05	0.04	-0.01	-0.18	0.00	-0.18	0.00	0.03	0.03	-0.21	0.02	-0.19
	0.04	0.45	0.45	0.08	0.08	0.09	0.06	0.75	0.75	0.12	0.69	0.69
No origin	-0.07	-0.01	-0.08	-0.18	0.00	-0.19	-0.03	-0.03	-0.06	-0.20	-0.03	-0.22
	0.04	0.54	0.54	0.10	0.78	0.78	0.07	0.75	0.75	0.15	1.11	1.11
WRITING												
Schooling												
Baseline	-0.75	0.34	-0.42	-0.23	0.45	0.22	-1.05	0.47	-0.58	-0.82	0.53	-0.28
	0.50	0.06	0.50	0.60	0.07	0.60	0.58	0.10	0.58	0.68	0.13	0.68
No muni	-0.76	0.33	-0.43	-0.01	0.20	0.19	-1.10	0.49	-0.61	-0.79	0.47	-0.32
	0.48	0.05	0.48	0.57	0.06	0.57	0.58	0.10	0.58	0.68	0.12	0.68
No origin	-0.95	0.56	-0.39	-0.78	1.00	0.22	-1.13	0.50	-0.63	-1.00	0.70	-0.31
	0.51	0.10	0.51	0.71	0.21	0.71	0.52	0.11	0.52	0.62	0.18	0.62
Age												
Baseline	-0.11	-0.01	-0.12	-0.07	-0.02	-0.09	0.10	-0.04	0.06	-0.07	-0.04	-0.10
	0.04	0.39	0.39	0.08	0.59	0.59	0.06	0.72	0.72	0.10	0.89	0.89
No muni	-0.11	-0.01	-0.12	-0.10	0.01	-0.09	0.11	-0.04	0.07	-0.08	-0.02	-0.10
	0.04	0.41	0.41	0.08	0.36	0.36	0.06	0.75	0.75	0.11	0.79	0.79
No origin	-0.09	-0.06	-0.14	-0.03	-0.11	-0.14	0.13	-0.05	0.08	-0.03	-0.06	-0.10
	0.04	0.51	0.51	0.10	0.99	0.99	0.08	0.71	0.71	0.14	1.02	1.02

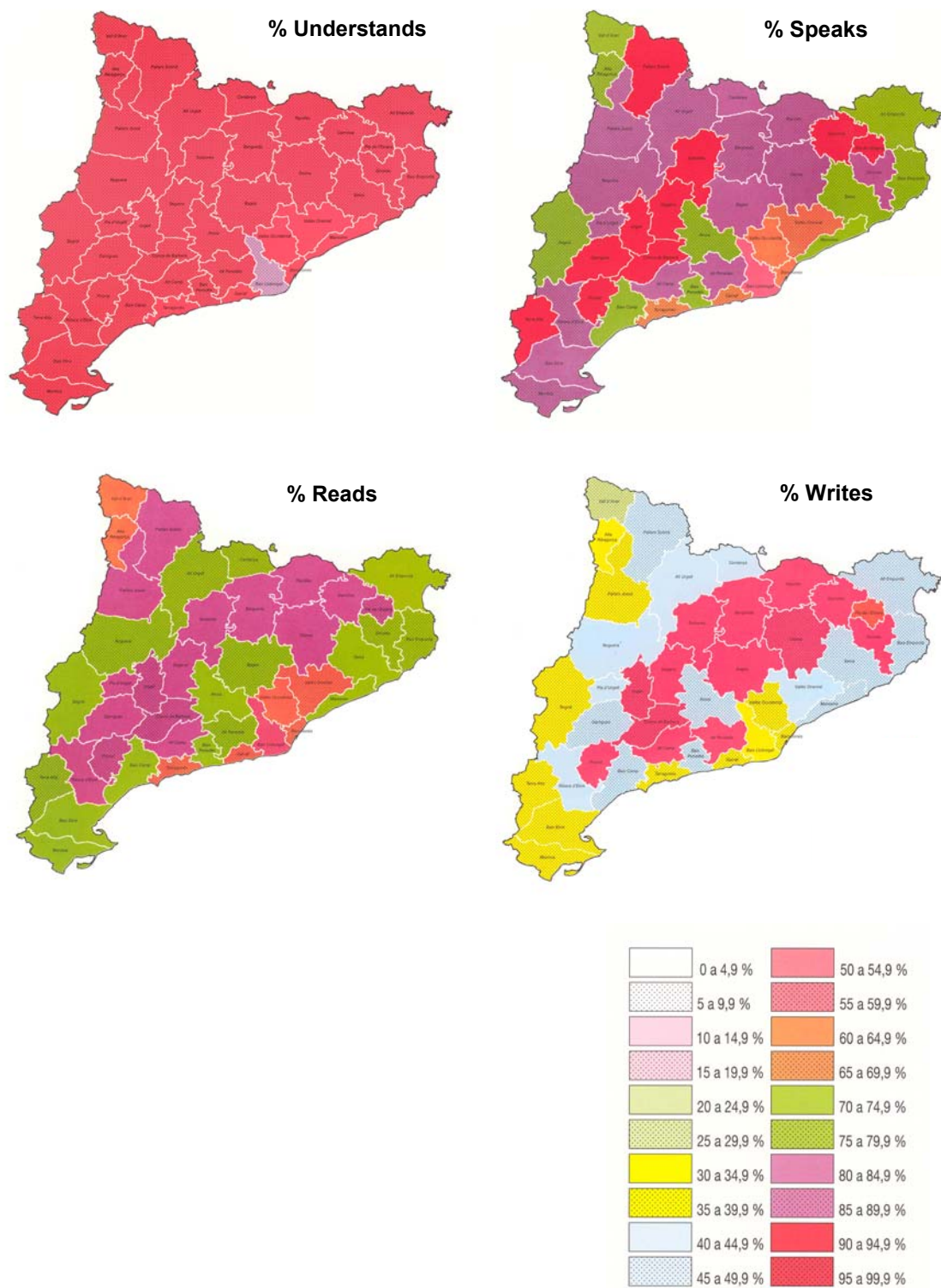


Figure 1. Percentage of the population that knows Catalan by language skill  
(Source: Reixach 1997)



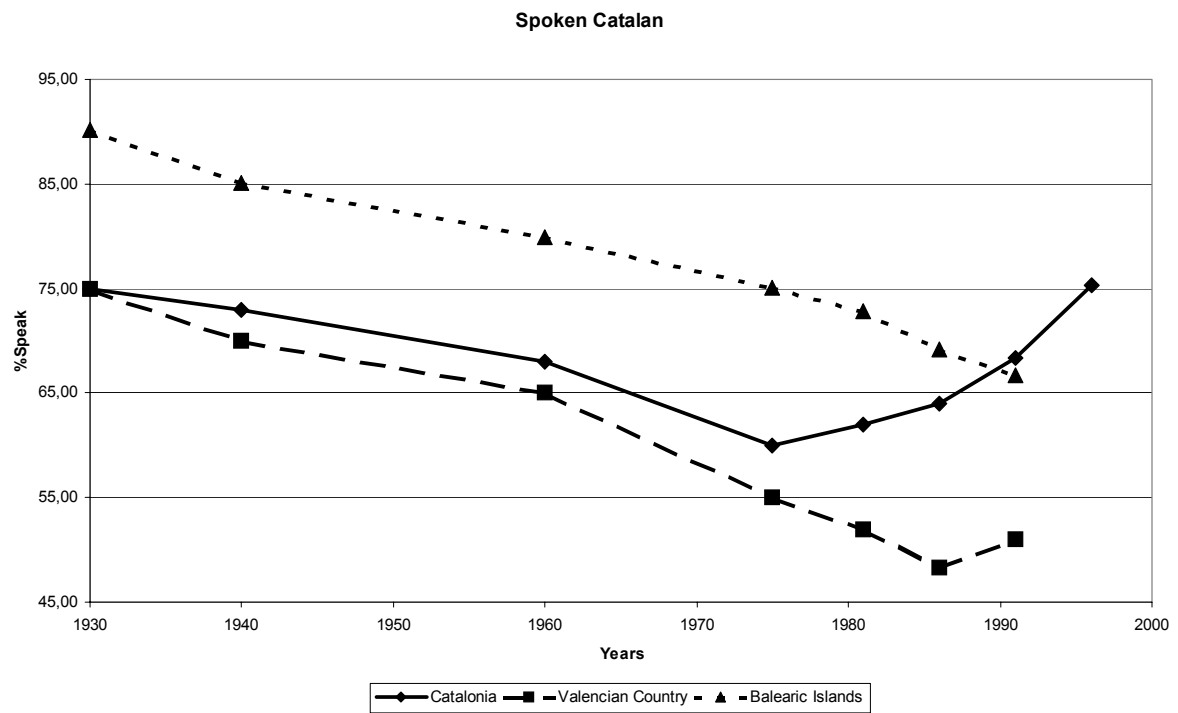


Figure 2: Spoken Catalan in Catalonia, Valencia and Balearics (Source: Vallverdú 1990)